



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Enhancing the Energy Efficiency of Radio Base Stations

Hauke Andreas Holtkamp



THE UNIVERSITY
of EDINBURGH

Thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy
to the
University of Edinburgh — 2013

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

Hauke Andreas Holtkamp

November 2013

Abstract

This thesis is concerned with the energy efficiency of cellular networks. It studies the dominant power consumer in future cellular networks, the Long Term Evolution (LTE) radio Base Station (BS), and proposes mechanisms that enhance the BS energy efficiency by reducing its power consumption under target rate constraints. These mechanisms trade spare capacity for power saving.

First, the thesis describes how much power individual components of a BS consume and what parameters affect this consumption based on third party experimental data. These individual models are joined into a component power model for an entire BS. The component model is an essential step in analysis but is too complex for many applications. It is therefore abstracted into a much simpler parameterized model to reduce its complexity. The parameterized model is further simplified into an affine model which can be applied in power minimization.

Second, Power Control (PC) and Discontinuous Transmission (DTX) are identified as promising power-saving Radio Resource Management (RRM) mechanisms and applied to multi-user downlink transmission. PC reduces the power consumption of the Power Amplifier (PA) and is found to be most effective at high traffic loads. DTX mostly reduces the power consumption of the Baseband (BB) unit while interrupting transmission and is better applied in low traffic loads.

Joint optimization of these two techniques is found to enable additional power-saving at medium traffic loads and to be a convex problem which can be solved efficiently. The convex problem is extended to provide a comprehensive power-saving Orthogonal Frequency Division Multiple Access (OFDMA) frame resource scheduler. The proposed scheduler is shown to reduce power consumption by 25-40% in computer simulations, depending on the traffic load.

Finally, the thesis investigates the influence of interference on power consumption in a network of multiple power-saving BSs. It discusses three popular alternative distributed uncoordinated methods which align DTX mode between neighboring BSs. To address drawbacks of these three, a fourth memory-based DTX alignment method is proposed. It decreases power consumption by up to 40% and retransmission probability by around 20%, depending on the traffic load.

Lay Summary

This thesis is about reducing the power consumption of mobile phone networks. Base Stations, the antennas on rooftops everywhere, currently require a large amount of electricity which is expensive. This thesis studies base stations and proposes mechanisms which reduce their power consumption while providing an unchanged connection quality. This is achieved by reducing the base station power when only few people use their mobile phones.

First, the thesis describes how much power the individual electronic elements of a BS consume and what parameters this consumption depends on. These individual models are joined into a component power model for an entire base station. With this power model, one can study a base station in theory without having an expensive laboratory. The component model is very complex and is therefore further simplified into a so-called parameterized model.

Second, two techniques are studied which reduce the power consumption of base stations when it is transmitting data to mobile phones. One is Power Control (PC), which reduces the power at the antenna when it is not needed, because users are standing near the base stations. The other is Discontinuous Transmission (DTX). DTX turns off the base station for a very short time, which is not noticeable for mobile phones. Using the parameterized power model, PC is found to work very well when many people use their phones, for example, at daytime.

DTX is more effective, when few people use their phones, for instance at night-time. It is found that we can mathematically combine both techniques to save even more power. Using a mathematical technique called convex optimisation, this can be done extremely fast. Using this technique, the power consumption could be reduced by 25-40% in computer simulations, depending on how many people use their phones at one time.

Finally, the thesis investigates how much power an entire network consumes when all base stations are transmitting. When all base stations transmit at the same time, the network becomes noisy. It is better for base stations to take turns. The thesis discusses three popular alternative methods which try to organize when the different base stations in the network transmit. To address disadvantages of these three, a fourth method is proposed which provides a trade-off between user satisfaction and power-saving. A small decrease in network quality could reduce power consumption by 20%, depending on how many people use their phone at one time.

Acknowledgements

This document could not have been produced without help. I owe deep gratitude to my academic supervisor, Professor Harald Haas of the University of Edinburgh, for believing in me and providing me with guidance and support. This work is in its current shape only through his teachings, insistence and attention to detail. Furthermore, I am grateful to the thesis examiners, Dr James R. Hopgood of the University of Edinburgh, and Dr Emad Alsusa of the University of Manchester.

Extended thanks goes to DOCOMO Euro-Labs GmbH in Munich. There, Dr. Gunther Auer was the best industry supervisor any PhD student could wish for. He pulled me back on track when I drifted. Guido Dietl helped me through extended technical discussions. The entire DOCOMO team over the years were my friends, sharpened my technical understanding and provided support when it was needed. Thanks, Emmanuel, Iwamura-san, Marwa, Matthias, Patrick, Petra, Samer, Serkan, Shinji, Toshi, and Zubin.

At the University of Edinburgh, Bogomil, Harald, Nikola, and Stefan have helped me to get through the university jungle and lead the way to a PhD. Go Jacobs/IUB alumni!

My friends have provided me with the indispensable social support. Most of all I am thankful to my parents without whom I would not be where I am today.

Contents

Abstract	v
Lay Summary	vii
Contents	xi
List of Tables	xv
List of Figures	xvii
List of Acronyms	xix
Nomenclature	xxi
1 Introduction	1
1.1 Overview	1
1.2 Thesis Context	1
1.3 Thesis Contributions	4
1.4 Thesis Structure	4
1.5 The EARTH Project	6
2 Motivation and Background	9
2.1 Overview	9
2.2 Energy Efficient Base Stations	10
2.3 Quantifying Energy Efficiency	13
2.4 Green Radio in Literature	15
2.5 Technical Background	19
2.5.1 LTE	20
2.5.2 Multi-carrier Technology	22
2.5.3 Multiple Antenna Technology	25
2.5.4 Convex Optimization	28
2.5.5 Network Simulation	31
2.6 Summary	35

3	Power Saving on the Device Level	37
3.1	Overview	37
3.2	Challenges in Power Modeling	39
3.3	Existing Power Models	40
3.4	The Component Power Model	41
3.4.1	Remarks	41
3.4.2	The Components of a BS	42
3.4.3	BS Power Consumption	56
3.5	The Parameterized Power Model	58
3.6	The Affine Power Model	64
3.7	Summary	67
4	Power Saving on the Cell Level (Single-cell)	71
4.1	Overview	71
4.2	Power-saving RRM in Literature	72
4.3	PC and TDMA	74
4.4	Power and Resource Allocation Including Sleep (PRAIS)	81
4.5	Resource allocation using Antenna adaptation, Power control and Sleep modes (RAPS)	90
4.5.1	Problem Formulation	91
4.5.2	Step 1: Antenna Adaptation (AA), DTX and Resource Allocation	94
4.5.3	Step 2: Subcarrier and Power Allocation	97
4.5.4	Results	102
4.6	Summary	110
5	Power Saving on the Network Level (Multi-cell)	113
5.1	Overview	113
5.2	Channel Allocation in Literature	115
5.3	System Model and Problem Formulation	116
5.4	DTX Alignment Strategies	118
5.4.1	Sequential Alignment	118
5.4.2	Random Alignment	119
5.4.3	P-persistent Ranking	119
5.4.4	Distributed DTX Alignment with Memory	120
5.5	Results	123
5.6	Summary	130
6	Conclusions, Limitations and Future Work	133
6.1	Summary and Conclusions	133
6.2	Limitations and Future Work	137
	Appendices	139

A Appendix	139
A.1 Proof of Convexity for Problem (4.12)	139
A.2 Proof of Convexity for Problem (4.15)	140
A.3 Margin-adaptive Resource Allocation	140
B List of Publications	143
B.1 Published	143
B.2 Accepted	144
B.3 Submitted	145
B.4 Project Reports	145
B.5 Contributions	145
C Attached Publications	147
Publications	201
Literature References	203

List of Tables

3.1	Model parameters assumed for figures of Chapter 3	42
3.2	Reference power consumption values of RF transceiver blocks . . .	48
3.3	Complexity of BB operations	49
3.4	Parameter breakdown	64
3.5	Summary of affine power model parameters	67
3.6	Comparison of required input parameters for different power models	69
4.1	Simulation parameters	80
4.2	Power model parameters used in Section 4.4	87
4.3	System parameters	104
5.1	Simulation parameters	124

List of Figures

1.1	Mobile traffic forecast 2012-2017	2
2.1	The carbon footprint for an average subscriber in 2007	10
2.2	European daily traffic pattern	12
2.3	Adoption of LTE technology	21
2.4	Mobile subscriptions by technology, 2009-2018	21
2.5	An illustration of TDM	23
2.6	An illustration of FDM	24
2.7	Resource allocation in a combined OFDMA/TDMA system	25
2.8	Example of multi-user diversity	26
2.9	Comparison of Multiple-Input Multiple-Output (MIMO) capacities	28
2.10	Geometric interpretation of a simple convex problem	30
2.11	A network in the network simulator	33
2.12	Sample simulation flowchart	34
3.1	The components of the modelled BS	43
3.2	The power-added efficiency over the maximum output power	45
3.3	P_{PA} as a function of bandwidth used	46
3.4	P_{RF} as a function of bandwidth used	47
3.5	P_{BB} as a function of bandwidth used	51
3.6	DC conversion loss function	52
3.7	P_{DC} as a function of bandwidth used	53
3.8	AC conversion loss function	54
3.9	P_{AC} as a function of bandwidth used	54
3.10	P_{COOL} as a function of bandwidth used	56
3.11	P_{supply} as a function of bandwidth used	57
3.12	P_{supply} per component in percent	59
3.13	Load-dependent power model for an LTE BS	63
3.14	Comparison of the parameterized with the complex model	65
4.1	Illustration of two possible power/time trade-offs	75
4.2	Comparison of the effect of load dependence	80
4.3	Illustration of PRAIS for two links	81
4.4	Supply power consumption for a target spectral efficiency	84
4.5	Fundamental limits for power consumption in BSs	89

4.6	OFDM frame structure	91
4.7	Illustration of margin-adaptive power allocation over three steps .	102
4.8	Outline of the RAPS algorithm.	103
4.9	Performance comparison in step 1	105
4.10	Outage probability in step 1 and the BA benchmark	106
4.11	Average number of DTX time slots over increasing target link rates	108
4.12	Supply power consumption for different RRM schemes overall . .	108
4.13	Supply power consumption for different RRM schemes in step 2 .	109
4.14	Energy efficiency as a function of sum rate	111
5.1	Illustration of DTX alignment in a network	114
5.2	Illustration of sequential alignment	119
5.3	BS power consumption over different cell sum rates	126
5.4	BS power consumption over OFDMA frames at 1 Mbps per mobile	127
5.5	BS power consumption over OFDMA frames at 2 Mbps per mobile	128
5.6	Retransmission probability over targeted rate.	129

List of Acronyms

3GPP	3rd Generation Partnership Project
AA	Antenna Adaptation
AC	Alternating Current
ADC	Analog-to-Digital Converter
BA	Bandwidth Adaptation
BB	Baseband
BS	Base Station
CDM	Code Division Multiplexing
CSI	Channel State Information
DAC	Digital-to-Analog Converter
DC	Direct Current
DTX	Discontinuous Transmission
EARTH	Energy Aware Radio and neTwork tecHnologies
FDM	Frequency Division Multiplexing
FPGA	Field-Programmable Gate Array
GOPS	Giga Operations Per Second
GSM	Global System for Mobile communications
HetNet	Heterogeneous Network
ICT	Information and Communication Technologies
IQ	In-phase/Quadrature
LNA	Low-Noise Amplifier

LTE	Long Term Evolution
M2M	Machine-to-Machine
MIMO	Multiple-Input Multiple-Output
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PA	Power Amplifier
PC	Power Control
PRAIS	Power and Resource Allocation Including Sleep
QoS	Quality of Service
RAPS	Resource allocation using Antenna adaptation, Power control and Sleep modes
RF	Radio Frequency
RCG	Rate Craving Greedy
RRH	Remote Radio Head
RRM	Radio Resource Management
SIMO	Single-Input Multiple-Output
SINR	Signal-to-Interference-and-Noise-Ratio
SISO	Single-Input Single-Output
SNR	Signal-to-Noise-Ratio
SotA	State-Of-The-Art
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
VCO	Voltage-Controlled Oscillator

Nomenclature

$B_{\text{target},k}$	Target rate for user k in bps, page 101
C	Ergodic capacity in bps/Hz, page 27
D	Number of BS antennas/radio chains, page 44
E	Energy efficiency in bit/Joule, page 110
G_k	Channel gain on link k , page 75
I_{BB}	Set of subcomponents of the BB unit, page 49
K	Number of users/links, page 76
M_{R}	Number of receiver antennas, page 27
M_{sec}	Number of sectors on a macro BS, page 56
N	Number of subcarriers, page 24
N_{Tx}	Number of resource blocks scheduled for transmission, page 117
P_0	Idle BS power consumption (at zero transmission power) in W, page 63
P_1	BS power consumption at maximum load in W, page 63
$P_{\text{BB}}^{\text{ppm}}$	Power consumption of the BB units in the parameterized model, page 61
$P_{\text{RF}}^{\text{ppm}}$	Power consumption of the Radio Frequencys (RFs) transceivers in the parameterized model, page 61
P_k	Transmission power on link k in W, page 75
$P_{0,D}$	Idle BS power consumption (at zero transmission power) in the affine model in W, page 66
$P_{1,D}$	BS power consumption at maximum load in the affine model in W, page 66
$P_{\text{AC,out,max}}$	Maximum output power of the Alternating Current (AC)-Direct Current (DC)-converter in W, page 53

$P_{AC,S}$	Power consumption of AC-DC-conversion in sleep mode in W, page 53
P_{AC}	Power consumption of AC-DC-conversion in W, page 53
$P_{BB,S}$	Power consumption of the BB unit during sleep mode in W, page 50
P_{BB}	Power consumption of the BB unit in W, page 50
$P_{COOL,S}$	Power consumption of active BS cooling in sleep mode in W, page 55
P_{COOL}	Power consumption of active BS cooling in W, page 55
$P_{DC,out,max}$	Maximum output power of the DC-DC-converter in W, page 51
$P_{DC,S}$	Power consumption of DC-DC-conversion in sleep mode in W, page 52
P_{DC}	Power consumption of DC-DC-conversion in W, page 51
P_N	Thermal noise power in W, page 75
$P_{PA,S}$	Power consumption of all PAs in a BS in sleep mode in W, page 45
$P'_{PA,S}$	Power consumption of a PA in sleep mode in W, page 45
$P_{PA,limit}$	Maximum output power of the PA in W, page 62
P_{PA}	Power consumption of all PAs in a BS in W, page 45
P_{PA}^{ppm}	Power consumption of the PAs in the parameterized model, page 61
$P_{RF,S}$	Power consumption of all RF transceivers in sleep mode in W, page 47
P_{RF}	Power consumption of all RF transceivers in W, page 47
P''_{RF}	Sum of RF blocks power consumption in W, page 48
P'_{RF}	Power consumption of an RF transceiver in W, page 47
$P_{S,0}$	Reference value for the single antenna BS sleep mode power consumption in W, page 62
$P_{S,D}$	BS power consumption in sleep mode in the affine model in W, page 66
P_S	BS power consumption in sleep mode in W, page 63
P_{Tx}	Transmission power of a BS at the input of the antenna interface in W, page 27
P_{max}	BS maximum sum transmission power in W, page 61
P_{out}	The power output of a single PA in W, page 44

P_{supply}	BS supply power consumption in W, page 56
$P_{a,e}$	Transmission power per resource in W, page 92
$P_{i,\text{BB}}^{\text{ref}}$	Reference power consumption of a BB subcomponent i in W, page 50
Q	Ranking tuple, time slots in order of capacity, page 120
Q_0	Time slot with highest virtual capacity, page 120
R_k	Capacity on link k in bps, page 75
R_t	Virtual sum capacity of time slot t in bits, page 118
T	Number of time slots, page 24
T_{Active}	Number of transmission time slots, page 93
T_{Sleep}	Number of DTX time slots, page 93
T_S	Number of time slots in which for all subcarriers no user is allocated, page 117
V	Priority tuple, time slots in order of score, page 120
W	Total system bandwidth in Hz, page 24
\mathbf{H}	Channel state matrix, page 27
$\mathbf{H}_{n^c,t^c,k}$	Matrix of channel states of the OFDMA frame's central resource, page 94
$\mathbf{H}_{n,t,k}$	Matrix of channel states on subcarrier n , time slot t , link k , page 91
\mathbf{r}	Vector of target rates, length K , page 92
Δ_w	Subcarrier spacing in Hz, page 24
Δ_p	Power consumption load factor, page 63
Ω_k	Set of channels which contribute positive power in margin adaptive power allocation, page 101
Π	Function mapping resource blocks to users, page 117
$\Upsilon_u, \Upsilon_{uu}$	Set of used/unused time slots, respectively, page 120
$\boldsymbol{\mu}^*$	Solution vector to the time share problem, page 78
χ	Transmission load share, page 63

ϵ_i	i -th eigenvalue, page 27
$\eta_{\text{PA,max}}$	Maximum PA efficiency in the parameterized model, page 62
η_{PA}	PA efficiency in the parameterized model, page 61
η_{ld}	Load dependence of a BS, page 79
γ_k	Signal-to-Noise-Ratio (SNR) on link k , page 75
$\mathbb{E}_{\mathbf{H}}(\cdot)$	Expected value operator, page 27
$\mathcal{A}_{k,t}$	Set of subcarriers on time slot t assigned to user k , page 99
\mathcal{E}_k	Vector of spatial channel eigenvalues scheduled to user k , page 95
$\mathcal{E}_{a,k}$	Vector of spatial channel eigenvalues on resource block a scheduled to user k , page 91
\mathcal{U}_k	Set of resources scheduled to user/link k , page 91
Φ	Share of time spent transmitting instead of idling or DTX, page 82
μ_k	Normalized transmission time on link k , page 76
ν	Normalized time spent in DTX mode, page 84
ψ	Scoring map, assigning scores to users, page 120
$\psi_{\text{ul}}, \psi_{\text{ll}}$	Upper/lower limit of score, respectively, page 120
σ_{AC}	Loss factor for AC-DC-conversion in the parameterized model, page 61
σ_{COOL}	Loss factor for active cooling in the parameterized model, page 61
σ_{DC}	Loss factor for DC-DC-conversion in the parameterized model, page 61
σ_{feed}	Feeder loss, page 44
τ_k	Transmission time on link k in s, page 76
τ_{S}	Time spent in DTX in s, page 84
τ_{frame}	Frame duration in s, page 76
θ	PA efficiency decrease, page 62
v	Water-level in margin-adaptive power allocation, page 101
ϑ	Operating temperature in K, page 75
ξ_{COOL}	Cooling loss factor, page 55

ζ_{AC}	Ratio of used to maximum output power by the AC-DC-converter, page 53
ζ_{DC}	Ratio of used to maximum output power by the DC-DC-converter, page 51
a_{TECH}	Technology scaling for RF transceiver, page 48
f	Share of available bandwidth used for transmission, page 44
k	Link or user index, page 75
l_{AC}	Efficiency of AC-DC-conversion; a function of ζ_{AC} , page 53
l_{DC}	Efficiency of DC-DC-conversion; a function of ζ_{DC} , page 51
l_{PA}	PA efficiency, page 45
m_k	Number of resources scheduled to user k , page 97
m_{rem}	Remaining resources during quantization, page 98
$m_{k,t}$	Number of resources in time slot t scheduled to user k , page 98
$m_{t,rem}$	Remaining resources in time slot t during quantization, page 98
n	Subcarrier index, page 91
q_B	Boltzmann constant in $m^2kgs^{-2}K^{-1}$, page 75
$r_{n,t,k}$	Capacity of subcarrier n , time slot t when scheduled to user k , page 117
t	Time slot index, page 91
w	Subcarrier bandwidth in Hz, page 92
y_i^D	Scaling exponent for BB power consumption which depends on the number of antennas, page 50
y_i^f	Scaling exponent for BB power consumption which depends on the share of available bandwidth used, page 50
z_s	Sleep mode switch for the calculation of BB power consumption, page 50

Chapter 1

Introduction

1.1 Overview

In the first section of this chapter, it is argued why energy efficiency in wireless networks is an important research topic and what goals this thesis sets out to achieve. The second section introduces the contributions of this thesis towards achieving these goals. The third section outlines the thesis structure. In the last section, the Energy Aware Radio and neTwork tecHnologies (EARTH) project is briefly introduced which provided laboratory results for the contents of Chapter 3.

1.2 Thesis Context

Since the emergence of mobile communications, both the number and the density of mobile devices have constantly increased [1]. They are predicted to rise further, fueled by innovations such as Machine-to-Machine (M2M) communication and the ‘Internet of Things’ [2, 3]. It is predicted that a trillion devices will be connected

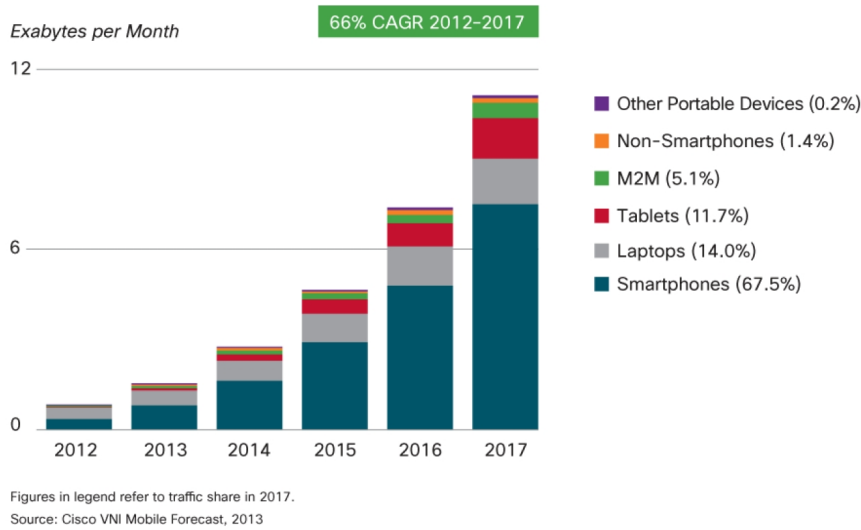


Figure 1.1: Mobile traffic forecast 2012-2017 with a Compound Annual Growth Rate of 66% [5, p. 7].

to the Internet by the end of 2013 [4], with a growing share of these on mobile networks. Fig. 1.1 shows a forecast of the resulting mobile traffic until 2017. To serve this growing traffic load, the capacity, size and density of the infrastructure network are continually upgraded by network operators.

While for the past 20 years network capacity, reliability and deployment were the main concerns during these upgrades, new factors like rising energy prices, increasing consumer attention to the emission of CO₂, the deployment of Base Stations (BSs) in remote off-grid locations, and disaster recovery have come into play. To respond to these new factors, the energy efficiency of the current and future mobile networks needs to be enhanced. Delivery of bits through the cellular network has to become cheaper. CO₂ emissions have to be reduced. BSs in remote locations must become self-sustaining. And maintaining connectivity after disasters is critical as the dependence of people and services on mobile communication in a disaster aftermath grows.

Lowering the power consumption of the BSs in cellular networks addresses all

of these problems. It reduces the network operational cost and the required CO₂ emission. Installing grid-independent BSs becomes more feasible when less power is needed for their operation. Furthermore, connectivity can be provided for a longer period of time in case of disastrous power interruptions.

However, the causes of power consumption in BSs have not been studied in depth. Whether and by how much this power consumption can be lowered through operational modifications is an open problem. The answer to this problem affects the design of future transmission schemes and hardware. Once the most effective mechanisms for reducing power consumption are understood, they can be used to operate BSs with lower power consumption and, thus, with higher efficiency. Radio resource schedulers need to be developed which enhance the efficiency of operation in both individual BSs and a network of BSs.

As such, this thesis sets out to achieve the following goals:

- Identify how the power consumption of a BS is comprised in terms of hardware components. Evaluate what operating parameters have the strongest effect on the BS power consumption. Model this consumption in equations such that it can be reused and modified.
- Identify promising operational techniques to reduce a BS's power consumption without affecting the Quality of Service (QoS). Generate resource schedulers which employ such techniques.
- Understand how power consumption can be reduced in a network of multiple BSs. Provide solutions to address power consumption on the network level.

1.3 Thesis Contributions

This thesis contributes towards the enhancement of energy efficiency in wireless networks by reducing the power consumption of radio BS through Radio Resource Management (RRM) without affecting the QoS. In particular, it provides the following contributions:

- First, the power consumption of State-Of-The-Art (SotA) Long Term Evolution (LTE) BSs is studied and presented in detail. The number of antennas, the transmission power and the lowered consumption through Discontinuous Transmission (DTX) are derived as the most relevant operating parameters. A general, practical power model for cellular BSs is constructed.
- Second, the attainable power savings of power control, DTX and Antenna Adaptation (AA) are quantified. The three techniques are jointly applied to a new energy efficient and spectrum efficient Orthogonal Frequency Division Multiple Access (OFDMA) scheduler.
- Third, the alignment of DTX time slots between neighbouring BSs is identified as a multi-cell power-saving mechanism. Three alternative alignment schemes are studied with regard to their power consumption. The findings lead to a novel time slot alignment scheme which is shown to overcome the limitations of the other techniques.

1.4 Thesis Structure

Chapter 2 first addresses the issue why this work is concerned with the supply power consumption of cellular BSs in particular. Next, it provides an overview

of Green Radio research alongside a review of relevant literature. Finally, a technical background for Chapters 3, 4, and 5 is provided about the topics LTE, multi-carrier technology, multi-antenna transmission, convex optimization, and computer simulation techniques.

Chapter 3 begins with an analysis of the power consumption of a SotA BS in order to identify opportunities for power saving. Each subcomponent is individually inspected and described with regard to its power consumption. A comprehensive BS power model is established as the sum of each subcomponent's power consumption. Relevant and promising saving mechanisms are identified. A parameterized power model is proposed which encompasses these saving mechanisms and abstracts architectural details in favor of simplicity. Finally, an affine power model is derived, which is used through the remainder of this thesis. The affine mapping this model provides between transmission power and supply power is advantageous for its application in optimization.

Chapter 4 applies the knowledge on BS power consumption from Chapter 3 by proposing power-saving RRM mechanisms. On the basis of the affine power model, Power Control (PC) and DTX are studied. The potential of their joint optimization is identified. The power savings achieved by this method over the SotA are quantified in simulation. The joint optimization of PC, DTX and AA is applied in a comprehensive scheduler for power-saving in OFDMA downlink transmission within a cell.

Chapter 5 adds the consideration of intercell interference. When multiple BSs are considered, the use of sleep modes in each BS results in significant fluctuations of interference. Constructive alignment of interference is proposed to exploit this effect and can be employed for a decrease of power consumption. To address multi-cell power saving resource allocation, *distributed DTX with memory* is proposed

and compared to alternative alignment mechanisms. Network simulations are used to estimate achievable savings.

Finally, conclusions of the above research, a discussion of limitations and an outlook for future research is provided in Chapter 6.

Regarding the format of this document: There are two non-overlapping bibliographies at the end of this document. One contains publications by the author of this thesis. References to these publications are made with alphabetic indices, such as ‘[a]’ or ‘[b]’. References to other literature are written with numeric indices such as ‘[1]’ or ‘[22]’.

1.5 The EARTH Project

The research presented in Chapters 3 and 4 of this thesis has received funding and experimental data from the Energy Aware Radio and neTwork tecHnologies (EARTH) project. Initiated by the European Union’s Framework Programme (FP) 7, the EARTH project aligned cooperation between 15 industry and academic institutions towards the common goal of driving up energy efficiency in current and future cellular networks. The project lasted from January 2010 to June 2012. It was founded to work on

- deployment strategies,
- network architectures,
- network management,
- adaptation to load variations with time,

-
- innovative component designs with energy efficient adaptive operating points,
 - and new radio and network resource management protocols for multi-cell cooperative networking.

The most prominent outcomes of the project were

- the cellular network life cycle analysis [6],
- the energy efficiency evaluation framework [7, 8],
- the BS power model [e],
- hardware implementations of Power Amplifiers (PAs) with improved dynamics [9] and a sleep mode enabled small cell [10],
- and numerous individual techniques which reduce the power consumption of cellular networks such as the ones presented in Chapter 4 [g].

All project deliverables are available on the project website for reference [11].

Chapter 2

Motivation and Background

2.1 Overview

This chapter first outlines the motivation why reducing the power consumption of cellular Base Stations (BSs) constitutes a large steps towards energy efficient wireless networks and how this reduction can be achieved. Second, it discusses the difficulties of quantifying energy efficiency and provides an overview of Green Radio research topics in literature. Finally, technical concepts which are applied in Chapters 4 and 5 are briefly introduced for the unfamiliar reader.

The work in Section 2.2 has been previously published by the author of this thesis in [a]. The technical background in Section 2.5 is sourced from literature as referenced.

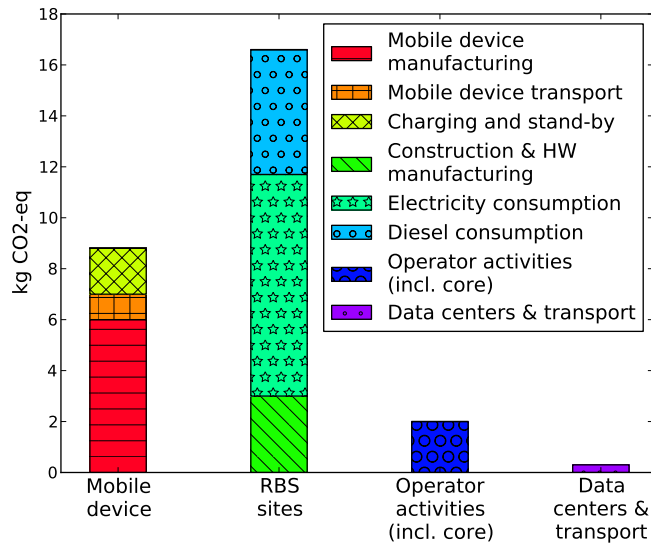


Figure 2.1: The carbon footprint (CO₂-equivalent emissions, see Section 2.3) for an average subscriber in 2007 [12], with a focus on radio base stations (RBS).

2.2 Energy Efficient Base Stations

Life cycle analyses of mobile networks and their equipment have shown that the access infrastructure (BSs, data centers, controllers) generates significantly more CO₂ than the connected mobile devices [6]. They also reveal that mobile devices cause the majority of CO₂ emissions during manufacturing due to their battery-optimized operation and short life times. In contrast, BSs tend to be less power optimized with longer life times leading to the majority of CO₂ emitted during operation. This is illustrated in Fig. 2.1 by the detailed carbon footprint of the average mobile network subscriber in 2007, the most recent data available. It shows that only 20% of the CO₂ emissions over the life of a mobile device were caused during operation. On the contrary, 82% of BS emissions are owed to operation, either as diesel or electricity consumption.

While the power consumption of mobile devices has always been a topic of research and development due to the constraints imposed by battery operation,

BSs were usually installed in urban locations with good connections to the power grid providing little incentive for efforts in energy efficiency. However, as briefly mentioned in Chapter 1, the energy efficiency of cellular BSs is receiving more attention due to several factors [13]:

- Rising energy prices [14] combined with decreasing profit margins [15] require operators to optimize their operational expenses by decreasing power bills.
- In countries with incomplete or unreliable power grids, BSs are operated with grid-independent power sources like diesel generators or renewable energy sources [15]. The costly deliveries of diesel and size of renewable energy sources like wind engines or solar panels make low consumption of the BS desirable [16].
- In the face of global warming, customers have developed an understanding and sensibility for power consumption. Companies aim to apply *green* labels to their products [17, 18].
- Disasters like the 2011 Tohoku earthquake and tsunami have shown that the installed battery backups were insufficient for long power grid interruptions [19]. After the disaster, BSs (and, thus, most communication) were unavailable for days until repair units had reconnected them to the power grid. Efforts in disaster recovery include low-power emergency modes and a generally lower power consumption to enable BSs to operate longer on battery power [20].

For these reasons, reducing the operating power consumption of Long Term Evolution (LTE) BSs is the goal of the work presented in this thesis. Recent analyses provided information on how this goal could be achieved. In particular,

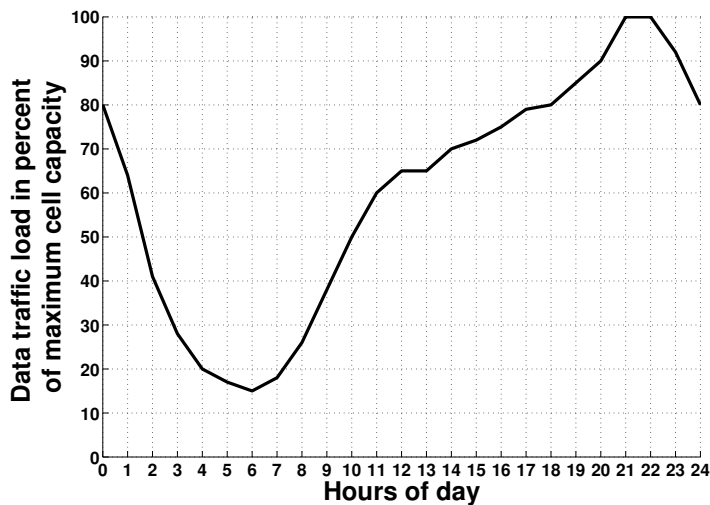


Figure 2.2: European average wireless daily traffic in 2010 [21].

it was found that while the cell traffic load fluctuates significantly, BSs are mostly unable to adapt to these fluctuations.

BSs have been designed and deployed to provide peak capacities to minimize outages. However, in practical scenarios, traffic is rarely at its peak. While BSs may be designed for crowded midday downtown situations, there are strong variations over time and location with regard to the density of mobile traffic. For example, see the typical mobile traffic in Europe over the course of a day in Fig. 2.2. It varies over a range of more than 80% and has an average of only 60% load [21]. Yet, although load varies so strongly, the power consumption of cellular BSs was found to vary as little as 2% over the course of a day [22]. BS power consumption is thus not yet sufficiently adaptive to the traffic demand. Hence, the times of day when the traffic load is below peak provide room for exploitation. During such times, the capacity of a BS can be reduced to achieve lower power consumption without affecting the user satisfaction or Quality of Service (QoS). How this could be achieved is the contribution of this thesis.

Note here an important differentiation between Green Radio technology proposals

being either in design or in operation. Design proposals are made to affect the architecture or hardware of a BS, such as changing the type of Power Amplifier (PA). Design changes may take a long time to reach the market due to specification, approval or manufacturing. In contrast, operation proposals, like turning a BS off when it is unused, can reach the market much faster. They may be applicable to existing hardware and can be applied via software upgrades.

Although design and operation changes can be made independently, they affect each other. For example, a change in design—like the introduction of a secondary low power wake-up controller—may enable new types of operation like a cognitive wake-up functionality. In turn, the popular application of a certain type of operation, like sleep modes, may have consequences for future design decisions by providing incentive for producing sleep-mode-enabled hardware. In this thesis, this intertwining of design and operation is taken into account by first identifying the structure and potential of the hardware and then exploiting this potential in operational improvements to the BS.

2.3 Quantifying Energy Efficiency

Enhancing the energy efficiency of communication networks has led to a field of research popularly labelled *Green Radio*, which this thesis is part of. This field encompasses all efforts taken to increase the ratio of network quality over energy spent or CO₂ emitted. Making ‘radios greener’ can be achieved by either increasing network quality at unchanged energy expenditure, by providing equal quality at a lower cost, or both. Here, typical network quality indicators to be improved are capacity, fairness, latency, reliability, and range. Cost is often measured in CO₂ emission, energy consumption, or power consumption. Note that

energy consumption is equal to the product of the average power consumption and a certain duration. In other words, power is energy normalized over time.

When discussing Green Radio research and results, it is important to be aware of some intricacies and necessary differentiations. Energy efficiency is not easily compared when considering different network quality indicators. For example, at equal energy consumption, an increase in capacity may come at the cost of increased latency, which cannot be easily weighted against each other. Alternatively, a decrease in power consumption through reduced network quality may trigger users to change their behavior which negates the initial savings. These problems are present in many fields of research and have led to discussions about QoS, which tries to assess the alternative a user may be more satisfied with [23]. But since an industry standard for the QoS does not yet exist, the common ground is to measure capacity as the most basic metric while stating the considered scenario assumptions such as geographic location, transmission duration, latency, *etc.* [8].

Just as there are several ways to measure network quality, there are several alternatives to measuring the cost incurred. The simplest is to only consider *transmission power* at the antenna of a wireless transmitter. This number is usually well-known as it is regulated and targeted in hardware design. A more comprehensive approach is to consider total power spent by a transmitter which includes the heat generated while producing or processing a wireless signal. This number can usually be measured at the mains supply or deduced from battery durations. It is called the *supply power*. But even the supply power required by a device during operation may not display the full picture. Aside from operation, a wireless device requires production before and disposal after. These steps can consume significant energy, particularly in electronic devices. They are composed of resource acquisition and mining, transportation, design, facility construction, production, and recycling. When widening the inspection

of power or energy consumption to this global level, the consideration of watts or joules is insufficient. Rather, the global cost of the steps is measured in terms of CO₂ emission. Comparing CO₂ emissions instead of power consumptions allows taking into account that the CO₂ emitted by generating a watt of power varies depending on many factors. For instance, a BS operated in a remote location on imported diesel may cause much higher CO₂ emissions than a solar-powered BS on an urban rooftop. As energy generation usually does not lead to the emission of pure CO₂, but rather a mixture of global warming affecting gases, the effects of these actions are typically normalized to CO₂-equivalent (or CO₂-eq). For example, the Information and Communication Technologies (ICT) industry is estimated to have caused about 1.5% of global CO₂-eq emissions in 2007 [12]. From 2007 to 2020, the CO₂-eq emission caused by mobile networks is predicted to grow from 0.2% to 0.4% [6]. Thus, a doubling of the energy efficiency of mobile communications would allow keeping CO₂-eq emissions on the 2007 level which is a common research and development target [24].

As described above, both measuring network quality as well as measuring cost are very intricate. Therefore, the metric of watts of supply power per bit alongside clear network definitions has been chosen as the reference for this thesis. When applying techniques described here to larger scenarios, the additional definitions provide the information necessary for converting watts to CO₂-eq over a regional expanse and duration of operation.

2.4 Green Radio in Literature

Historically, the early works on energy efficiency for wireless networks have been triggered by a desire to extend battery lives in cellular, ad hoc or sensor networks, for example see [25, 26, 27, 28]. By the early 2000s, growing environmental

concerns had lead to extensive life cycle analyses of cellular networks [29, 30]. These analyses collected all steps in a devices' life, quantified them with regard to energy expenditure and CO₂ emissions and joined them with the number of units sold and installed. The resulting statistics revealed that the power consumption of mobile network infrastructure was significant and that the CO₂ emissions of the ICT industry were as high as those of international air traffic [6, 15]. Through these findings, interest grew with regard to the energy efficiency of transmitters which are powered by the mains grid.

As a consequence, in 2008, the collaborative *Green Radio* project was initiated in the UK which coined the term [31]. The efforts were extended in 2009 by the European Energy Aware Radio and neTwork tecHnologies (EARTH) project [32]. Both projects encouraged specifically the enhancement of the energy efficiency of cellular access networks by improving their architecture and operation. The works proposed since then can be distinguished by the network aspects they consider.

For one, there are works which address the hardware of the radio access network. A general overview of such hardware improvements is provided in [33]. Specifically, the improvement of transceiver units is considered in [9, 34]. New PA architectures are proposed in [35]. The EARTH project has summarized proposed hardware improvements to LTE BSs in [10].

Another group of works is the field of Heterogeneous Networks (HetNets). These works are concerned with the power consumption of future networks which are predicted to consist of a very large number of low power small BSs, such as micro, pico, or femto, in addition to macro BSs. The study in [36] finds the optimal density of small cells to match a traffic requirement from an energy perspective. In [37], the area power consumption metric is proposed to enable comparison between different HetNets. Badic *et al.* [38] formalize the trade-off that is present between capacity and power consumption in HetNets. In [39], it is suggested

that the overlap between macro and small cells in HetNets should be used in combination with sleep modes when the capacity increase provided by small cells is not needed. The authors in [40] argue that due to the complexity of HetNets, they should be self-organizing and self-adapting to the traffic situation.

As a third topic, several kinds of sleep modes are proposed. Long sleep modes with duration of minutes or hours are very effective in reducing power consumption, but require a wake-up mechanism and reduce coverage. Short sleep modes in the order of microseconds to seconds, which are also called dozing, micro sleep or Discontinuous Transmission (DTX), do not pose these problems but promise smaller reductions of the power consumption. As a consequence, long sleep modes are proposed for situations in which coverage can be maintained through other means. For example, when network densities are sufficient such that neighbouring BSs can take over the coverage of sleeping BSs [41, 42]. Short sleep modes can be applied independent of coverage. The authors in [43] propose to use DTX when a cell is completely empty, posing a simple but very inflexible mechanism. In [44] it is assumed that a BS consists of multiple independent transmitters, some of which can be deactivated according to traffic requirements, thus increasing the adaptivity to load. With regard to an entire network of sleep capable cells, Abdallah *et al.* [45] study the alignment of sleep modes between neighbouring cells and conclude that sleep modes should be synchronized and orthogonal for maximum efficiency.

Related to sleep modes is the field of cell breathing, cell shrinking or cell zooming through the adjustment of the transmission power [46, 47, 48]. By introducing flexibility with regard to the transmission power of a BS, a network can flexibly adjust to coverage requirements by increasing or decreasing cell sizes. For example, an increase of a cell's size may be needed when neighbouring cells enter sleep mode.

An extension of sleep modes is the topic of adapting multi-antenna transmission. Rather than using all installed antennas, a BS could deliberately reduce its Multiple-Input Multiple-Output (MIMO) degree to deactivate some antennas when high capacities are not required [49, 50, 51, 52]. These approaches are labeled MIMO adaptation, antenna adaptation or MIMO muting. An alternative approach to exploit multiple antennas for energy efficiency is proposed by Wu *et al.* [53] and Stavridis *et al.* [54], who, through—spatial modulation—use all antennas, but never at the same time.

Resource scheduling is a diverse group of works in which different approaches are proposed. Relaxing the delay requirement can be exploited for opportunistic power saving [55, 56]. Spreading signals over a larger bandwidth allows reducing the transmission power [57, 58]. Approaches such as game theory promise gains by providing trading mechanisms for a self-organized allocation of resources for enhanced energy efficiency [59].

In order to exploit the typically deployed overlapping radio access network generations, it is proposed in [60] to combine a deactivation of one network technology, *e.g.* Global System for Mobile communications (GSM), with the fall-back to another such as LTE. However, this generates issues with backward compatibility and QoS.

To exploit spare backhaul capacity, it is suggested in [61] to use coordinated multi-point transmission between multiple BSs while deactivating parts of each BS. However, the benefits strongly depend on the additional power consumption caused by the backhaul link. When combined with sleep modes, it was found in [62] that this approach only reduces power consumption in the presence of high data rate users.

In [63] it is proposed to switch sectorized macro BSs to omni-directional operation

when the capacity requirements are low. While this promises to save two thirds of the power consumption in typical three-sector setups, it also requires the installation of additional omni-directional antennas at BS sites.

The most far-reaching proposals for energy efficient cellular networks break with the cellular assumptions presumed in GSM, Universal Mobile Telecommunications System (UMTS) or LTE. By separating data and control planes from one another, future networks could have large, long-range ‘umbrella’ control transmitters which take care of discovery and coverage. These would overarch small data plane BSs which are only activated when they are needed. Such a network setup would provide a very high capacity at a very high degree of load flexibility. These proposals are referred to in literature as ‘ghost’ or ‘phantom’ cell concepts [64, 65], since the small data plane BS are ‘invisible’ to the User Equipment (UE).

For good surveys on Green Radio topics, the reader is referred to [17, 21, 66, 67, 68].

Note that detailed literature backgrounds are again provided for each individual chapter in Chapters 3, 4 and 5.

In this wide field of research, this thesis can be placed in the group of works dealing with resource scheduling.

2.5 Technical Background

The following sections outline some fundamental concepts which reappear in Chapters 3, 4, and 5 of this thesis, namely, LTE, multi-carrier technology, multiple antenna transmission, convex optimization, and the usage of simulations to model complex cellular systems.

2.5.1 LTE

LTE [69] is a wireless access standard superseding the GSM and UMTS for increased network capacities. It is predicted to be adopted in 93 countries by the end of 2013 [70]. See Fig. 2.3 for an illustration on the world map. To this extent, the number of LTE subscriptions is rising steadily as it replaces older network technologies, as shown in Fig. 2.4. LTE is being continually standardized and developed by the 3rd Generation Partnership Project (3GPP). It is designed to achieve the following set of goals [71]:

- faster connection establishment and shorter latency;
- higher user spectral efficiencies of more than 5 bps/Hz on the downlink using link bandwidths of up to 20 MHz;
- uniformity of service provision independent of mobile position in a cell;
- improved cost per bit;
- flexibility in spectrum usage to accommodate to national band allocations;
- simplified network architecture;
- seamless switching between different radio access technologies;
- reasonable power consumption for the mobile terminal.

As LTE is a wireless technology on a shared medium, it is required to handle resource sharing and interference. These two important challenges are addressed in Chapters 4 and 5, respectively. Note that while LTE also consists of non-radio aspects such as the System Architecture Evolution (SAE), this thesis is only concerned with its BS and radio access aspects.

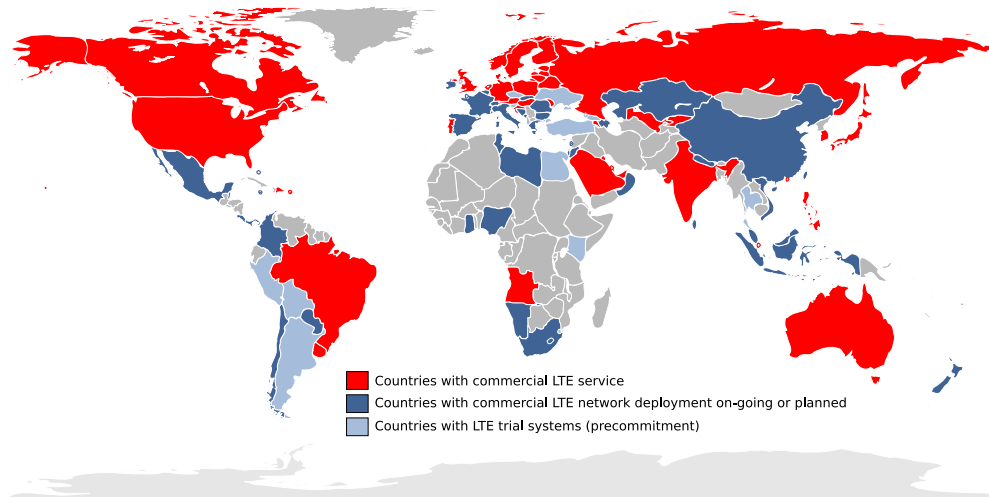


Figure 2.3: Adoption of LTE technology as of May 8, 2012 [70].

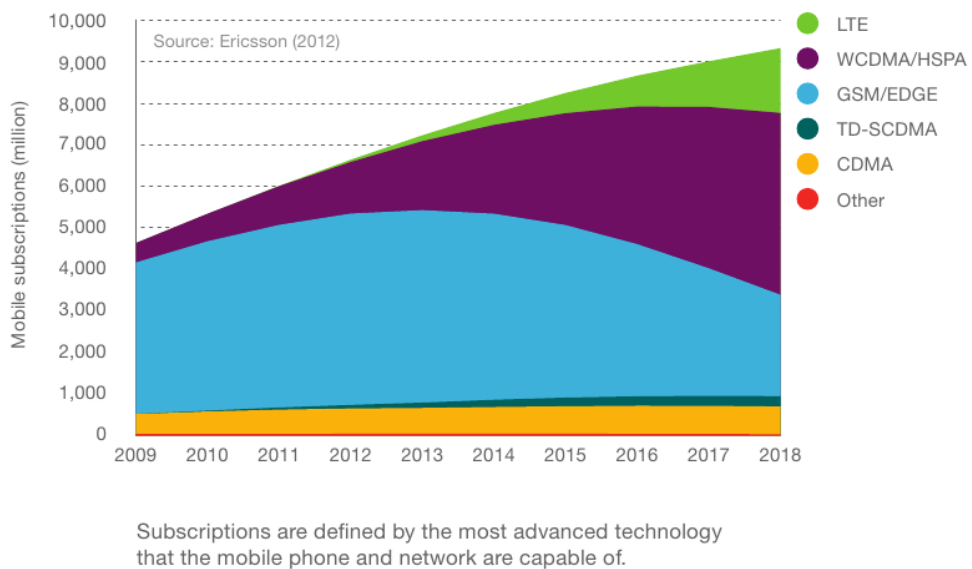


Figure 2.4: Mobile subscriptions by technology, 2009-2018 [72, p. 7].

2.5.2 Multi-carrier Technology

The wireless medium is inherently shared. Wireless transmissions occur simultaneously in resources such as the frequency spectrum, time, and location. Therefore, transmissions can negatively affect one another through interference. The technique of dividing a resource into chunks to avoid overlap between multiple transmissions is called *multiplexing*. Time is typically shared via Time Division Multiplexing (TDM), frequency via Frequency Division Multiplexing (FDM), and location via installing BSs a significant distance apart from one another. Multiplexing allows providing multiple independent links carrying independent information streams on the wireless medium. These streams can be allocated to different users or scheduled to the same user for increased sum data rates, thus increasing the scheduling flexibility. In TDM, the shared resource is the time domain in which only one link can be active at any point in time and all links are spread over the entire available bandwidth. This concept is illustrated in Fig. 2.5 for four users. In FDM, the shared resource is the frequency domain. Here, links only take up a portion of the available frequency spectrum. See Fig. 2.6 for an illustration of FDM with multiple carriers and four users. Both of these technologies have been applied commercially in wireless technologies for decades. For example, TDM is the access scheme of choice in the GSM standard [73], whereas FDM has been employed in cordless phones [74]. A third popular multiple access scheme, Code Division Multiplexing (CDM)—*e.g.* used in the UMTS [75]—is not discussed here as it is not related to this thesis.

Orthogonal Frequency Division Multiplexing (OFDM) extends FDM to provide a very flexible high capacity multiple access scheme. Similar in notion to FDM, OFDM subdivides the bandwidth available for signal transmission into a multitude of narrowband subcarriers. The channels affecting these subcarriers are sufficiently narrow in the frequency domain to be considered flat-fading [69]

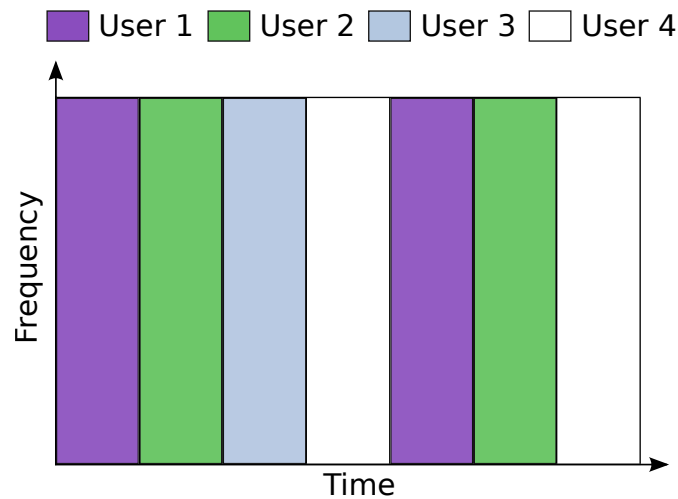


Figure 2.5: An illustration of TDM. Different users are assigned non-overlapping shares of the transmission time spanning a wide bandwidth.

which is important for link adaptation. Unlike in FDM where frequency bands are non-overlapping, OFDM subcarriers are packed tightly and do overlap. Yet, through selection of the right pulse shape, the signals on each subcarrier remain orthogonal and can be transmitted in parallel with little to no mutual interference.

In addition to this fine separation in frequency, in OFDM, the transmission duration is divided into short slots to create an *OFDM block* with flat-fading characteristics in which one modulation symbol is transmitted (Multiple antenna technology allows the transmission of multiple symbols, as described in the following section). The modulation symbols can be from any modulation alphabet such as Quadrature Phase Shift Keying (QPSK), 16-Quadrature Amplitude Modulation (QAM), or 64-QAM. In LTE, such an OFDM block is labeled a *resource element*. The available bandwidth and a duration called *frame* are divided into a number of resource elements. Multiple resource elements are combined to constitute *resource blocks*. When OFDM is used to grant multiple users access to the shared transmission medium, it is called Orthogonal Frequency Division Multiple Access (OFDMA). An illustration of a resulting OFDMA frame consisting of a set of resource blocks scheduled to different users is shown in

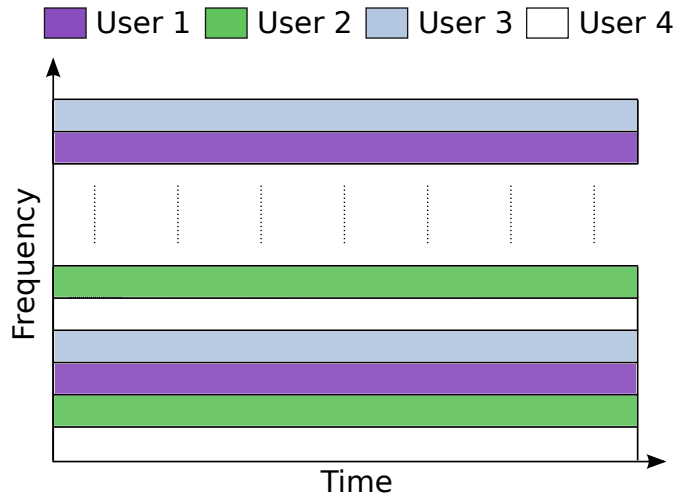


Figure 2.6: An illustration of FDM. Different users are assigned non-overlapping shares of the transmission bandwidth over a significant duration.

Fig. 2.7. For a typical LTE system with a subcarrier spacing $\Delta_w = 15$ kHz, a system bandwidth $W = 10$ MHz, 12 subcarriers per resource block and a 10% guard band, the number of available resource blocks over the frequency domain is

$$N = \frac{0.9W}{12\Delta_w} = 50. \quad (2.1)$$

Although a resource block consists of multiple subcarriers constituting resource elements, it is the smallest unit which can be individually scheduled, according to the LTE standard [76]. For this thesis, it is thus assumed that a resource block consists of one subcarrier and stretches over the duration of one time slot. A frame is comprised of T time slots. For example, with $N = 50$ and $T = 10$, there are

$$NT = 500 \quad (2.2)$$

resource blocks available for scheduling in an OFDMA frame.

The option to schedule each resource block to any user in the system enables

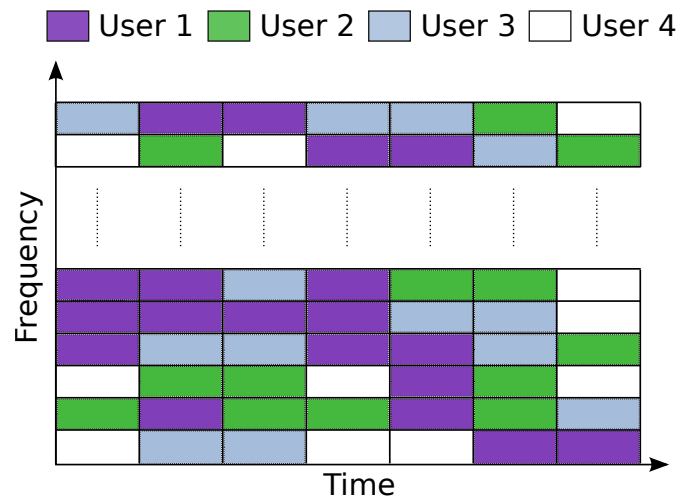


Figure 2.7: Example of resource allocation in a combined OFDMA/TDMA system constituting an OFDMA frame [71].

exploitation of an effect called *multi-user diversity*. As the channels on each resource block between each user and the BS is different, resource blocks can be scheduled such that only good channels are used. Statistically, over a certain bandwidth, number of time slots, and number of users, it is very probable that a good channel exists. This high probability results in an increase of the transmission data rate compared to a system which does not exploit this effect. An example of scheduling frequency by multiuser diversity is illustrated in Fig. 2.8. Opportunistic scheduling between multiple users in a system is an important technique employed in Chapter 4.

2.5.3 Multiple Antenna Technology

The transmission and reception of wireless signals with multiple antennas is a standard technique in modern communication systems [77]. Exploiting multipath scattering, MIMO techniques deliver significant performance enhancements in terms of data transmission rate and interference reduction. The term *multi-path* refers to the arrival of transmitted signals at an intended receiver through

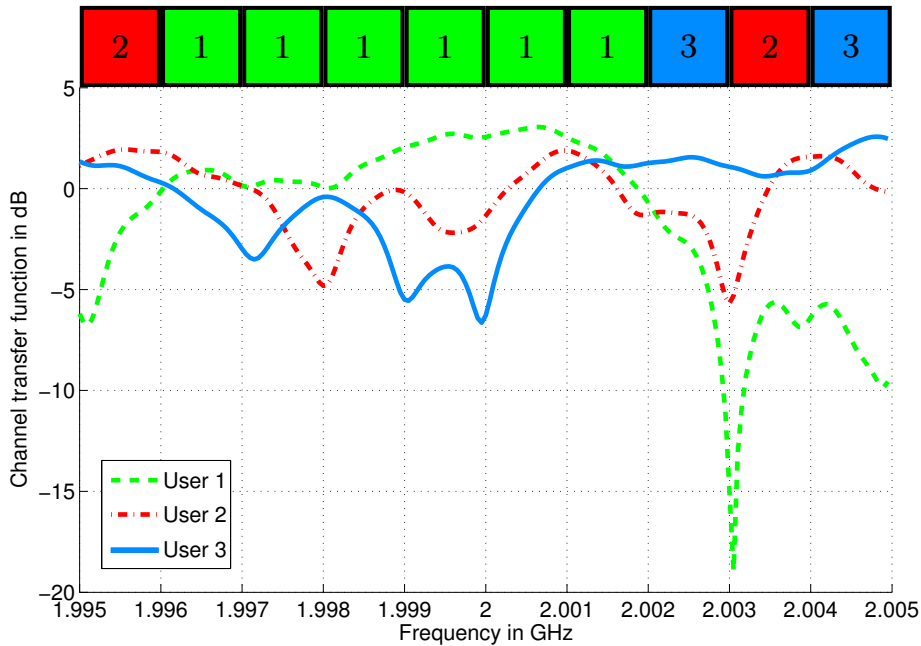


Figure 2.8: Example of multi-user diversity scheduling. For each frequency chunk the user with the best channel is selected.

differing angles, time delays and/or frequency shifts due to the scattering of electromagnetic waves in the environment, *i.e.* over multiple paths. Consequently, the received signal power fluctuates through the random superposition of the arriving multi-path signals. This random fluctuation allows the recognition and separation of the paths signals take between the multiple transmit antennas and multiple receive antennas, allowing to benefit from two important effects. First, like with multi-user diversity described in the previous section, having multiple signal paths reduces the probability of a deep fade. Hence, the reliability of the transmission is increased. The benefit of this effect is called the spatial diversity gain. Second, under suitable conditions such as rich scattering in the environment, separate data streams can be transmitted on the multiple paths. Thus, at the same time, on the same frequency spectrum, and in the same place, multiple data streams can be transmitted, increasing the data rate compared to a system with

fewer antennas [78]. The benefit of this effect is called the spatial multiplexing gain.

Under the assumption of statistical Channel State Information (CSI) at the transmitter, full CSI at the receiver and equal power allocated to each antenna, the ergodic MIMO link capacity is given by

$$C = \mathbb{E}_{\mathbf{H}} \left[\log_2 \det \left(\mathbf{I}_{M_R} + \frac{P_{Tx}}{D} \mathbf{H} \mathbf{H}^H \right) \right], \quad (2.3a)$$

$$= \sum_{i=1}^{\min(D, M_R)} \log_2 \left(1 + \frac{P_{Tx}}{D} \epsilon_i \right) \quad (2.3b)$$

with the expected value operator $\mathbb{E}_{\mathbf{H}}(\cdot)$, transmit power P_{Tx} , channel state matrix \mathbf{H} , the number of transmit and receive antennas D and M_R , respectively, identity matrix of size M_R , \mathbf{I}_{M_R} , and ϵ_i as the i -th singular value of $\mathbf{H} \mathbf{H}^H$ [79]. The gains in capacity through the use of multiple antennas can be charted by generating a number of Rayleigh distributed channel coefficients \mathbf{H} and plotting the achieved capacities. Figure 2.9 shows an illustration of achievable capacities, revealing two important characteristics of the MIMO capacity. First, it can be seen that capacity grows linearly with Signal-to-Noise-Ratio (SNR) at high SNRs. Second, the capacity grows linearly with $\min(D, M_R)$. Thus, more antennas yield a higher capacity. In this thesis, up to four receive and transmit antennas are considered due to limitations detailed in Section 4.5.2.

While raising the number of antennas increases the capacity, it also increases the complexity of transmitters and receivers. As will be shown in Chapter 3, this increase in complexity is accompanied by a larger power consumption. It is studied in Chapter 4 how to trade off power consumption with capacity for energy efficiency by changing the MIMO mode through Antenna Adaptation (AA).

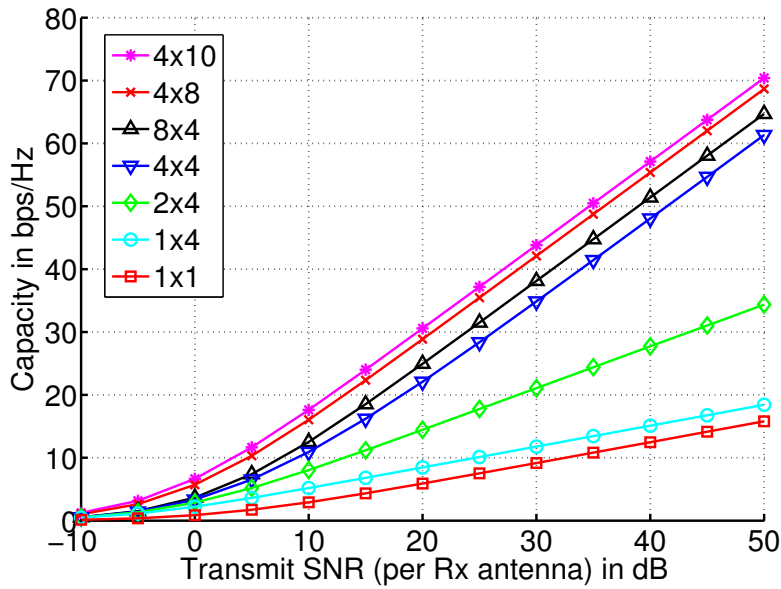


Figure 2.9: Power-fair MIMO capacities at perfect CSI at the receiver and statistical CSI at the transmitter.

2.5.4 Convex Optimization

An important topic employed in Chapter 4 is convex optimization which is briefly introduced here.

A general optimization problem is often written as

$$\begin{aligned}
 & \text{minimize} && f_0(\mathbf{x}) \\
 & \text{subject to} && f_i(\mathbf{x}) \leq b_i, \quad i = 1, \dots, m \\
 & && h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p.
 \end{aligned} \tag{2.4}$$

The problem consists of the optimization variable $\mathbf{x} = (x_1, \dots, x_n)$, the objective function $f_0 : \mathbb{R}^n \mapsto \mathbb{R}$, the inequality constraint functions $f_i : \mathbb{R}^n \mapsto \mathbb{R}$, $i = 1, \dots, m$ and equality constraint functions $h_i : \mathbb{R}^n \mapsto \mathbb{R}$, $i = 1, \dots, p$, and the constraints b_1, \dots, b_m . \mathbf{x}^* is the solution of the problem (2.4) if for

any z , $f_1(\mathbf{z}) \leq b_1, \dots, f_m(\mathbf{z}) \leq b_m$ and $h_1(\mathbf{z}) = 0, \dots, h_p(\mathbf{z}) = 0$, it holds that $f_0(\mathbf{z}) \geq f_0(\mathbf{x}^*)$ [80].

This form of optimization problems is convex, if the functions f_0, \dots, f_m satisfy

$$f_j(\alpha \mathbf{x} + \beta \mathbf{y}) \leq \alpha f_j(\mathbf{x}) + \beta f_j(\mathbf{y}) \quad (2.5)$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, all $\alpha, \beta \in \mathbb{R}$ with $\alpha + \beta = 1, \alpha \geq 0, \beta \geq 0$, and $j = 0, \dots, m$. In other words, it is convex if the objective function as well as all constraints are convex. Furthermore, any equality constraints h_i have to be affine as any

$$h(\mathbf{x}) = 0 \quad (2.6)$$

can be represented as

$$h(\mathbf{x}) \leq 0, \quad (2.7a)$$

$$-h(\mathbf{x}) \geq 0. \quad (2.7b)$$

If $h(\mathbf{x})$ were convex, $-h(\mathbf{x})$ would be concave. The only way for both (2.7a) and (2.7b) to be convex is when $h(\mathbf{x})$ is affine.

The important property of convex optimization problems is that they can be solved by interior point methods and that the local optimum is always also the global solution [81]. Furthermore, convex optimization problems can often be solved very efficiently enabling their application even in real-time systems such as resource schedulers. A one-dimensional convex optimization problem without constraints can be graphically illustrated. An example is shown in Fig. 2.10 for some $t = f_0(x)$.

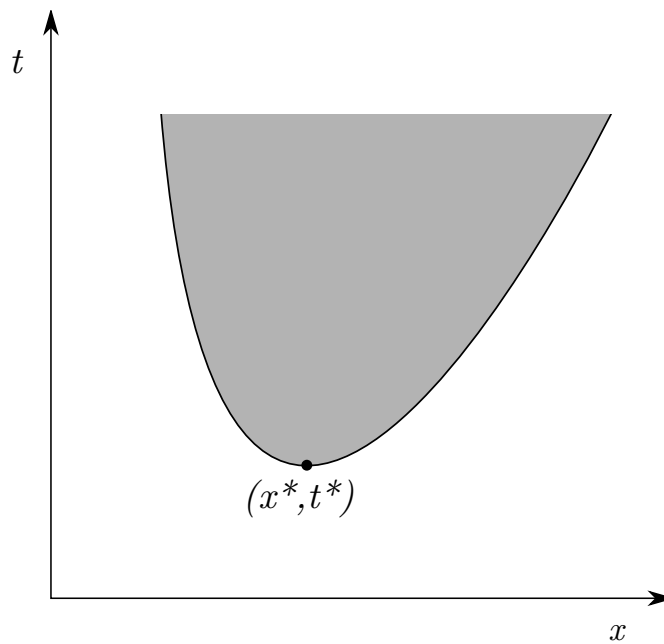


Figure 2.10: Geometric interpretation of a simple convex problem [80].

A simple example of a two-dimensional convex optimization problem is

$$\begin{aligned}
 &\text{minimize} && f_0(\mathbf{z}) = z_1^2 + z_2 + 1 \\
 &\text{subject to} && f_1(\mathbf{z}) = -1 - z_1 \leq 0 \\
 &&& h_1(\mathbf{z}) = z_2 - 1 = 0.
 \end{aligned} \tag{2.8}$$

The second gradient of $f_0(\mathbf{z})$ is

$$\nabla^2 f_0(\mathbf{z}) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.9}$$

The objective function and its inequality constraints are convex as $\nabla^2 f_0(\mathbf{z})$ is greater or equal to zero on all entries. The equality constraint is affine. The sample problem is thus a convex optimization problem. By inspection of (2.8), it can be found that the solution is

$$\mathbf{z}^* = (0, 1). \tag{2.10}$$

Either proving convexity of a problem or transforming a problem into a convex problem are ways to efficiently solve optimization problems.

2.5.5 Network Simulation

For many problems in communications research, analytical solutions have not yet been found. To still be able to evaluate such problems, researchers resort to repeated independent trial executions of a given problem. The statistical outcomes of these trials provide insight into the problems and potential solutions. These trials are often performed as computer simulations—rather than in hardware—to avoid costly and time consuming prototyping. Such computer simulations allow to model specific problems, vary their nature and generally experiment on them. They can also be used to test solutions and assess whether inventions deliver the anticipated results. As they prove to be such a versatile tool, they are used also in this thesis to assess the proposed solutions, when analytical solutions cannot be found.

More specifically, cellular networks are simulated which are as large as the computational constraints allow. For example, see Fig. 2.11 for the layout of a typical network simulation consisting of UEs and BSs distributed over an area.

The network simulator which produced the results shown in this thesis was developed from the ground up by the author and contains the following elements:

- A configuration interface via configuration files,
- Network generation consisting of:
 - Geometric modules for the distribution of physical entities onto hexagonally arranged three-cell BS sites,

- Path loss and shadowing modules for the association of UEs to BSs as defined by the 3GPP [82],
- OFDM MIMO Signal-to-Interference-and-Noise-Ratio (SINR) calculation,
- A frequency-selective fading model calibrated according to the Wireless world INitiative NEw Radio (WINNER) air interface configuration [83],
- Margin-adaptive resource allocation,
- Several modules for the algorithms described in Chapters 4 and 5,
- Interior point convex optimization via the Interior Point OPTimizer (IPOPT) [84],
- Data collection, analysis and plotting tools,
- A unit test suite,
- Detailed event logging,
- Parallelization via the high-throughput parallelization system HTCondor [85].

In addition to the list of simulator elements, it is also important to note what is not modelled. For example, capacity is calculated based on SINR by the Shannon limit. Bit errors, retransmission, error correction or protocols are not included; neither are pilots, channel sounding or LTE signaling and, thus, imperfect CSI. UE mobility is modelled as a fast-fading component rather than a change in position on the network map.

A typical simulation consists of a large number of independent trials executed in Monte Carlo fashion, *i.e.* in parallel, thus yielding statistical outcomes over all trials. A flowchart of the most important steps involved in a simulation run is

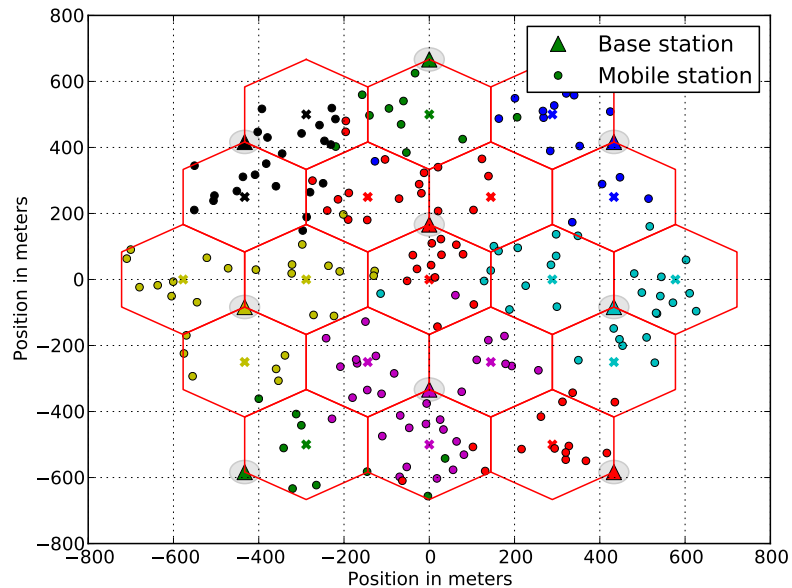


Figure 2.11: A sample network generated by the network simulator.

provided in Fig. 2.12. To ensure sufficient consideration of interference, data is only collected from cells which are surrounded by at least two layers (tiers) of cells causing interference.

The simulators which produced the results presented in Chapters 3 and 4 were written as MATLAB [86] scripts. Required packages are the Optimization and Statistics Toolboxes. An optional package is the Parallel Computing Toolbox. The network simulator employed in Chapter 5 was written in python [87] using numpy, scipy, matplotlib, virtualenv, IPOPT, ATLAS, LAPACK, MKL and ACML libraries. Depending on the network size, extensive simulation runs can take up to several days on a parallelized system. Unlike the highly optimized convex optimization processes, computation of SINRs and data handling are very time consuming, causing long simulation run times.

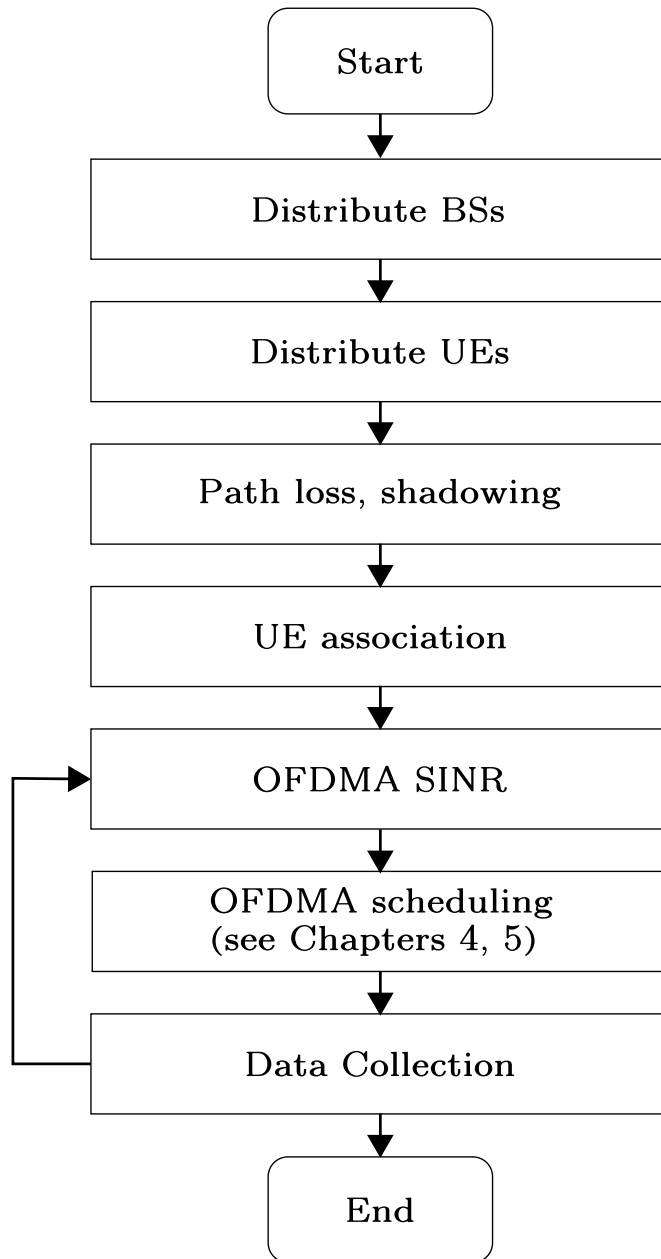


Figure 2.12: Sample simulation flowchart.

2.6 Summary

This chapter served to provide the motivation for BS power saving, establish the literature context, and briefly introduce key technical concepts. To begin with, it was explained that cellular BSs contribute significantly to the power consumption and CO₂ emission of mobile networks and that they do so during operation. Low load situations, such as night time, were identified as opportunities for reducing BS power consumption. The necessity of observing BS design along with BS operation for power saving was emphasized which is reflected in the structure of this thesis. Next, the Green Radio literature was discussed. In the last part of this chapter, key technical concepts were introduced which reappear in the following chapters. LTE, multicarrier and multiple antenna technology, and convex optimization are all important topics applied in this thesis. Furthermore, the computer simulation of cellular networks was described in detail as it was used for producing numerical results where analytical solutions could not be found.

Chapter 3

Power Saving on the Device Level

3.1 Overview

Power models describe how much power a device consumes in different configurations or scenarios. Depending on their depth, power models may also reveal architectural details and allow comparisons between different scenarios or assist the exploration of configurations which are not available for measurement in laboratories.

In energy efficiency research, power models are the basis of any power consumption analysis. Solid models are required to produce reliable results, since the model assumptions significantly affect the energy efficiency solutions proposed. For example, if a model were to falsely describe heat dissipation as a major contributor to power consumption, research efforts with regard to the reduction of heat dissipation might be ineffectively spent. Therefore, it is immensely important to have reliable power models such as the ones presented in this chapter.

In this chapter, three levels of detail of a power model for Long Term Evolution

(LTE) Base Stations (BSs) are derived. The reason is that higher levels of detail allow ensuring the reliability of the model, manipulating specific aspects and parameters and recognizing the underlying architecture. In contrast, lower levels of detail abstract the model specifics in favor of applicability. The goal of this chapter is to allow the reader to first follow the construction of the component model and later the simplification towards the affine model.

The remainder of this chapter is structured as follows. Section 3.2 summarizes the challenges encountered when modeling BS power consumption. Existing power models are discussed in Section 3.3. Section 3.4 describes the *component power model* in detail. This model reveals the sources of power consumption within cellular BSs. In Section 3.5, the complex component model is parameterized into a *parameterized power model* such that only those parameters which are often manipulated for power saving remain, greatly reducing the model complexity. In Section 3.6, an *affine power model* approximation is defined. The affine mapping contained in this model allows integrating it into analytical problems in Chapters 4 and 5. Section 3.7 summarizes the chapter and compares the three models with regard to their complexity.

The component model was developed in collaboration with the Energy Aware Radio and neTwork tecHnologies (EARTH) project. Hardware manufacturers have provided the experimental data on individual components. The author of this thesis has contributed to this power model by capturing it in MATLAB source code and aligning its concurrent development between EARTH partners. The model has been published and applied in parts in [8, 7, 21, e]. The detailed description of the model in this chapter is an original contribution. The parameterized model has been submitted for publication in [h].

3.2 Challenges in Power Modeling

Generic modeling of BS power consumption is not a trivial task for multiple reasons. First, in order to protect their industrial designs, equipment vendors and manufacturers are hesitant to reveal the architecture of their products. They go so far as to contractually prohibit the opening of device casings for customers. Consequently, it is difficult to understand the contributing components to power consumption within cellular BSs. Second, vendors and manufacturers have differing designs, such that a power model may either only be valid for a single brand of devices or that multiple models would have to be developed. Third, architectures and technologies continually evolve, weakening the long-term applicability of power models. As such, a particular model may only be valid for a certain generation of BSs.

All of these problems have been addressed in the component power model as follows: By the combined effort of experts from multiple manufacturers, general common architectures were agreed upon without revealing competitive technical differences. These common architectures also helped bridge the design differences present between competing manufacturers. Finally, to ensure the lasting applicability of the model, foreseen technical advances were included in the model. As LTE is the prominent cellular network technology of the coming decade (see Section 2.5), the component power model describes an LTE BS.

The data sets provided in this chapter were all obtained through measurements within the EARTH project.

3.3 Existing Power Models

In the literature, three distinct power models for wireless transmitters are proposed and applied.

While in need of a power model for the comparison of the power efficiency of Multiple-Input Multiple-Output (MIMO) with Single-Input Single-Output (SISO) transmission, Cui *et al.* [26] derived a first technology-independent model for the total power consumption of a generic radio transceiver. It was established that power consumption can be divided into the power consumption of the Power Amplifier (PA) and the power consumption of all other circuit blocks. The power consumption of the PA was described to vary with transmission settings like the number of antennas, the energy per bit, the bit rate, and the channel gain, while the consumption of circuit blocks was assumed a constant. Note that this power model is not a power model for cellular BSs, but for a generic wireless transceiver. Yet, it represents the first steps towards modeling the power consumption of wireless devices.

Arnold *et al.* [22] derive the power consumption of cellular BSs by considering that a typical BS consists of multiple sectors, has multiple PAs, a cooling unit, and a power supply loss. The remaining components' consumption is considered a constant. This model is accurate in its assumptions. But it only has data available for Global System for Mobile communications (GSM) and Universal Mobile Telecommunications System (UMTS) BSs. For LTE, the model is based on predictions and differs significantly from the values presented in this chapter.

Deruyck *et al.* [88] used extensive experimental measurements to determine the power consumption of several types of State-Of-The-Art (SotA) BSs, like the macro cell, micro cell and femto cell. The power consumption of a BS is assumed to be a constant, independent of the traffic load. Thus, this model is not suitable

for the exploration of BS operating power consumption as is required in Chapters 4 and 5.

The model described in the following has partly been published as an element of an EARTH project deliverable [8]. It has been published in abbreviated form in [7, 21, e].

3.4 The Component Power Model

Before the description of the model, it is necessary to make some introductory remarks.

3.4.1 Remarks

First, this model describes the supply power consumption of a LTE cellular BS and is derived as the sum of the consumption of its subcomponents. The consumption of each subcomponent is derived from experimental measurements. The prediction of the dependence on operating parameters is derived from their architecture.

Second, the model is an instantaneous model. It does not model effects which occur over time, such as power up or power down. Rather it models instantaneous power consumption at a certain configuration. When estimating power consumption over a time duration consisting of different configurations, a weighted averaging of the instantaneous values needs to be performed.

Third, this model encompasses the architecture of an LTE BS entering the market in the year 2012. For figures in this chapter, the system parameters in Table 3.1

Parameter	Value
Feeder loss, σ_{feed}	0.5
Number of radio chains, D	2
Transmitted power, P_{Tx}	40 W
Number of BS sectors, M_{sec}	3
System bandwidth, W	10 MHz

Table 3.1: Model parameters used in figures of this chapter.

are used. It represents the architecture found in most BSs which is foreseen to remain unchanged for the coming five to ten years. Note that this chapter is not concerned with how the hardware could be improved. Such hardware improvements were published in [10].

3.4.2 The Components of a BS

The modelled LTE macro BS consists of a transceiver with multiple antennas. The transceiver comprises one or more lossy antenna interfaces, PAs and Radio Frequency (RF) transceiver sections as well as one Baseband (BB) interface including both a receiver section for uplink and transmitter section for downlink, one Direct Current (DC)-DC power supply regulation, one active cooling system and one mains Alternating Current (AC)-DC power supply for connection to the electrical power grid. Fig. 3.1 shows an illustration.

In practice, a BS may be serving multiple cells by sectorization. In that case, there may be overlap between the components and the sectors. For example, a single cooling cabinet may be used to house the electronics for all sectors. However, for purposes of modeling, it is assumed that a BS only serves a single cell. Thus,

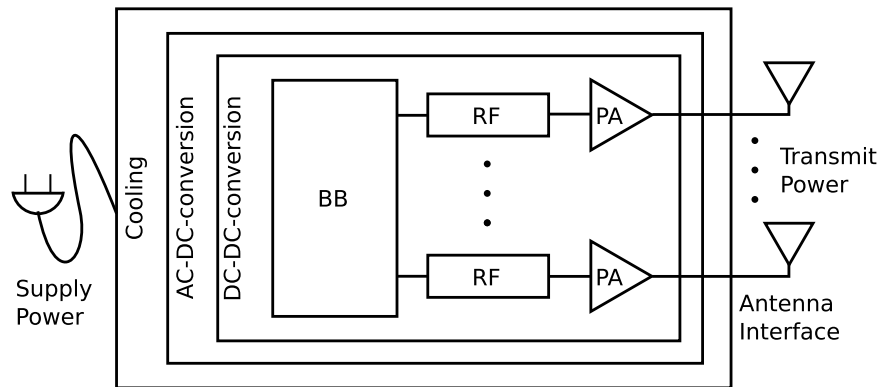


Figure 3.1: The components of the modelled BS.

the power consumption of a BS for multiple cells can be found as the sum of the individuals.

In currently prevailing architecture as portrayed in Fig. 3.1, the BS contains one or multiple so-called *radio chains* each consisting of one RF transceiver, one PA and one antenna. Thus, the number of radio chains, antennas, PAs, and RF transceivers is identical and terminology can be used interchangeably.

In the following, the power consumption of each individual component is inspected to generate the component power model for the entire BS. Components are treated from right to left as depicted in Fig. 3.1.

Antenna Interface

This passive component is listed here for completeness, although it has no power consumption characteristics. For macro BS with long feeder cables between the PA and the antenna, a feeder loss has to be compensated for, which typically reduces the signal power by around 3 dB. For the model, this feeder loss is included in the modeling of the PA.

PA

The PA amplifies the signal before it is transmitted by the antenna. Due to low achievable efficiencies, PAs contribute significantly to the overall power consumption if high powers are targeted at the antenna, as described in Section 2.2.

Note that when comparing transmission powers in this study, it is referred to as the sum measured at the inputs of all antenna interfaces. Thus, feeder cable losses are included and BSs with differing numbers of antennas can be compared fairly with regard to their transmission power.

Due to its non-linearity, the PA modeling involves measurement results of the power-added efficiency. The power-added efficiency of the PA is looked up during modeling via the function $l_{\text{PA}}(\cdot)$. The efficiency function is shown in Fig. 3.2. The efficiency of the PA can reach up to 54% at very high transmission powers. However, these high efficiencies can rarely be achieved in operation due to the strong transmission power fluctuation of Orthogonal Frequency Division Multiple Access (OFDMA) signals and power control. At low transmission powers, PAs have very low efficiencies. This effect cancels the power saving of power control to some degree.

At constant power spectral density, the power output of each PA, P_{out} in W, is a function of the transmission power, P_{Tx} in W, the number of antennas, D , the share of the maximum bandwidth used for transmission, $f \in [0, 1]$, and an adjustment for feeder losses, $\sigma_{\text{feed}} \in [0, 1]$ is given as

$$P_{\text{out}} = \frac{P_{\text{Tx}} f (1 - \sigma_{\text{feed}})}{D}. \quad (3.1)$$

The power consumption of the PAs in W is then determined by the efficiency and

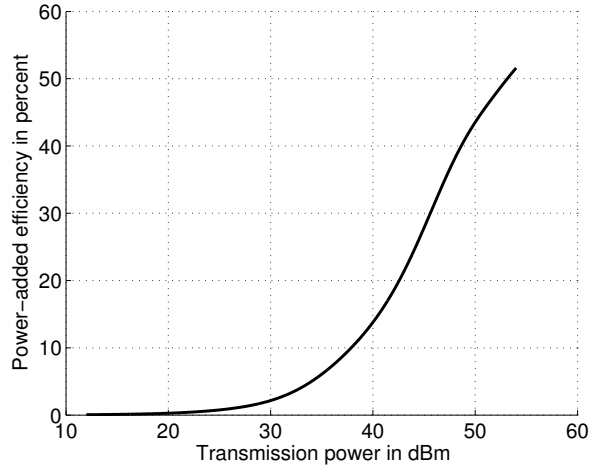


Figure 3.2: The power-added efficiency, $l_{\text{PA}}(\cdot)$, over the maximum output power as measured within the EARTH project.

the number of PAs, such that

$$P_{\text{PA}} = D \frac{P_{\text{out}}}{l_{\text{PA}}(P_{\text{out}})}. \quad (3.2)$$

The power consumption in sleep mode, $P_{\text{PA,S}}$ in W, scales with the number of PAs. The power consumption of one PA in sleep mode, $P'_{\text{PA,S}}$, was found by experimental measurement, such that

$$P_{\text{PA,S}} = DP'_{\text{PA,S}}, \quad (3.3)$$

with $P'_{\text{PA,S}} = 27.75 \text{ W}$.

The power consumption of the PA over f is shown in Fig. 3.3. The effect of the reduced PA efficiency at lower transmission powers results in an almost linear relationship of bandwidth/transmission power to power consumption. At a maximum of 250 W, the PAs can consume a significant amount of power. However, this can be mitigated by reducing the transmission bandwidth. Also,

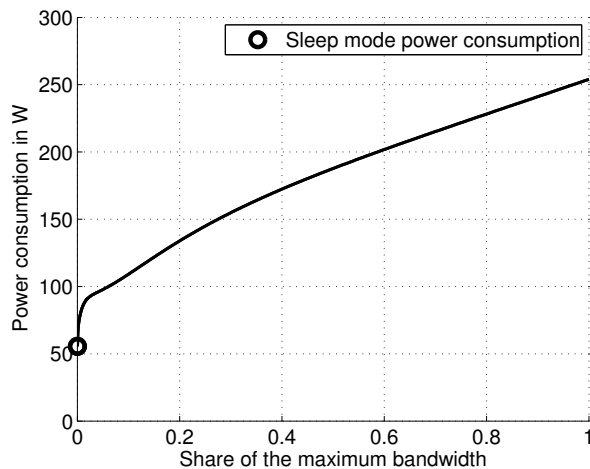


Figure 3.3: P_{PA} as a function of bandwidth used.

as an analog component, a PA can be switched to a sleep mode with low power consumption quickly which has just above 50 W power consumption.

RF Transceiver

The RF transceiver provides the signal conversion between the digital and analog domain as well as frequency conversion between the BB and RF frequencies. It consists of a large number of elements such as In-phase/Quadrature (IQ)-Modulators, voltage-controlled oscillators, mixers, Digital-to-Analog Converters (DACs) and Analog-to-Digital Converters (ADCs), low-noise and variable gain amplifiers and clocks. Power consumption of each of these elements is independent of external factors in measurement. The values and the resulting sum consumption are shown in Table 3.2. Thus, the power consumption of the RF transceiver is modelled as a constant. As one RF transceiver is installed in each antenna chain, RF power consumption scales with the number of antenna chains.

The power consumption of all RF transceivers, P_{RF} in W, is modelled as a linear

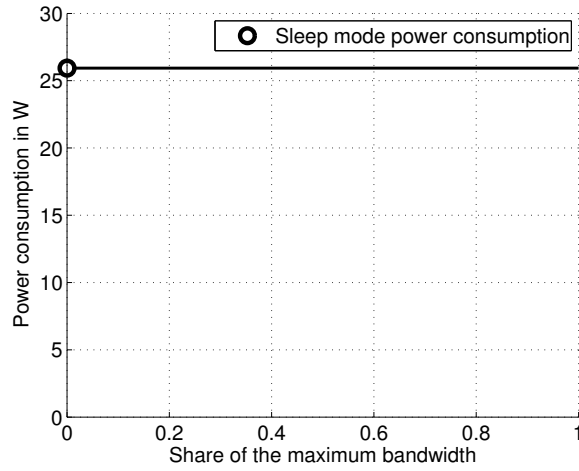


Figure 3.4: P_{RF} as a function of bandwidth used.

function of an individual transceiver, P'_{RF} , with

$$P_{\text{RF}} = DP'_{\text{RF}}, \quad (3.4)$$

where P'_{RF} in W is comprised of reference values for commercially available components as shown in Table 3.2. A technology scaling factor, a_{TECH} , is included to account for different generations of RF transceivers.

The RF transceiver as present in currently manufactured BSs is incapable of sleep mode operation. Therefore, its power consumption in sleep mode is equal to its power consumption when active.,

$$P_{\text{RF,S}} = P_{\text{RF}}. \quad (3.5)$$

The power consumption of the RF transceiver is shown as a function of f in Fig. 3.4. Since it is constant, it can only be manipulated through the number of RF transceivers in operation. Sleep mode operation is not yet foreseen for RF transceivers.

Block element	Power consumption in mW
IQ Modulator, DL	1000
Variable attenuator, DL	10
Buffer, DL	300
Voltage-Controlled Oscillator (VCO), DL	340
Feedback mixer, DL	1000
Clock Generation, DL	990
DAC, DL	1370
ADC, DL	730
Low-Noise Amplifier (LNA) 1, UL	300
Variable Attenuator, UL	10
LNA 2, UL	1000
Dual mixer, UL	1000
Variable Gain Amplifier, UL	650
Clock generation, UL	990
ADC 1, UL	1190
Sum P''_{RF}	10880
Technology scaling a_{TECH}	1.19
$P'_{\text{RF}} = a_{\text{TECH}}P''_{\text{RF}}$	12940

Table 3.2: Reference power consumption values of RF transceiver blocks, where UL and DL refer to blocks which are part of the uplink or downlink operation, respectively.

Type of processing, i	GOPS	$P_{i,\text{BB}}^{\text{ref}}$ in W	y_i^D	y_i^f	z_s
Time Domain Processing	360	9.0	1	0	0
Frequency Domain Processing	60	1.5	2	1	0
Forward Error Correction	60	1.5	1	0	1
Central Processing Unit	400	10.0	1	0	1
Common Public Radio Interface	300	7.5	1	0	1
Leakage	118	3.0	1	0	1

Table 3.3: Complexity of BB operations and their scaling with the number of transmit antennas and the transmission bandwidth.

BB Unit

The BB unit generates and processes the digital signal before it is passed to the RF transceiver. The BB unit of macro BSs is typically a stand-alone system-on-chip based on flexible Field-Programmable Gate Array (FPGA) architecture. Since this component is a digital processor, it is best modelled by its operations, namely by Giga Operations Per Second (GOPS). As the power cost of GOPS can be estimated at 40 GOPS/W and a set of different functions of the BB unit, I_{BB} , can be associated with the number of GOPS required, the power consumption can be modelled. Unlike in other BS components, the BB unit scales non-linearly with D and f . For example, the frequency domain processing complexity grows exponentially with the number of antennas. The exact scaling exponents along with the reference complexities for the modeling of BB operations are shown in Table 3.3.

These reference values and scaling exponents are used to find the BB unit power

consumption in W , with

$$P_{\text{BB}} = \sum_{i \in I_{\text{BB}}} P_{i,\text{BB}}^{\text{ref}} D^{y_i^D} f^{y_i^f}, \quad (3.6)$$

where $P_{i,\text{BB}}^{\text{ref}}$ in W is the power consumption per sub-component, y_i^D is the scaling exponent for the number of radio chains, and y_i^f is the scaling exponent for the share of bandwidth used, each as provided in Table 3.3.

The sleep mode power consumption of the BB units, $P_{\text{BB,S}}$ in W , is lower than when in active mode, as time domain and frequency domain processing can be disabled as indicated by the sleep mode switch, z_s . Like the active power consumption, sleep mode power consumption is found by the combination of reference values and scaling exponents,

$$P_{\text{BB,S}} = \sum_{i \in I_{\text{BB}}} P_{i,\text{BB}}^{\text{ref}} D^{y_i^D} f^{y_i^f} z_s. \quad (3.7)$$

When drawn over the transmission bandwidth, see Fig. 3.5, the power consumption of the BB unit is found to be only lightly affected by load, *i.e.* the bandwidth used. However, sleep modes are available for the BB which can reduce its power consumption by more than 30%. MIMO operation has a significant effect on the BB power consumption, as it is modelled with a quadratic scaling factor due to the added computational complexity of MIMO processing.

DC-DC-Conversion

To supply the required DC voltages for the various components mentioned above, a BS is equipped with several DC-DC-converters. As the architectures of these converters may vary across vendors, the conversion is modelled as a single

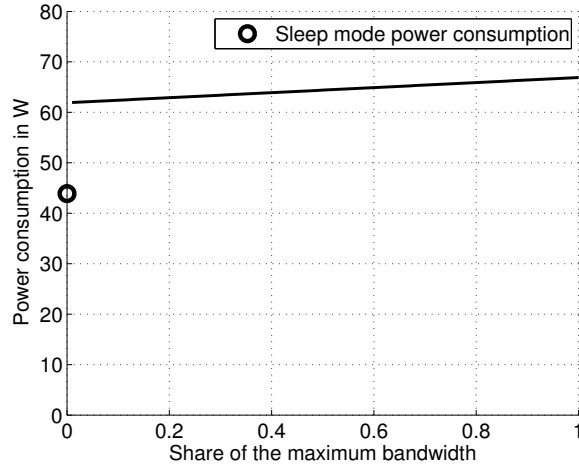


Figure 3.5: P_{BB} as a function of bandwidth used.

component. Laboratory measurements have been taken from one representative DC-DC-converter during the EARTH project as a function of the required load. Conversion has an efficiency of less than one, which can also be represented as a loss. With the DC-DC-converter built for a maximum power output of $P_{\text{DC,out,max}}$ in W, the loss is larger the further the actual required power by the PA, RF transceiver and BB unit is from $P_{\text{DC,out,max}}$. See Fig. 3.6 for the measured losses function, $l_{\text{DC}}(\zeta_{\text{DC}})$, of the ratio of maximum power output to actual power output, ζ_{DC} , with

$$\zeta_{\text{DC}} = \frac{P_{\text{DC,out,max}}}{P_{\text{PA}} + P_{\text{RF}} + P_{\text{BB}}}. \quad (3.8)$$

The power consumption caused by DC-DC-conversion in W is then found as

$$P_{\text{DC}} = l_{\text{DC}}(\zeta_{\text{DC}}) (P_{\text{PA}} + P_{\text{RF}} + P_{\text{BB}}). \quad (3.9)$$

SotA DC-DC-converters are assumed to possess no capability for sleep modes and, thus, remain activated during BS sleep mode, with power consumption during

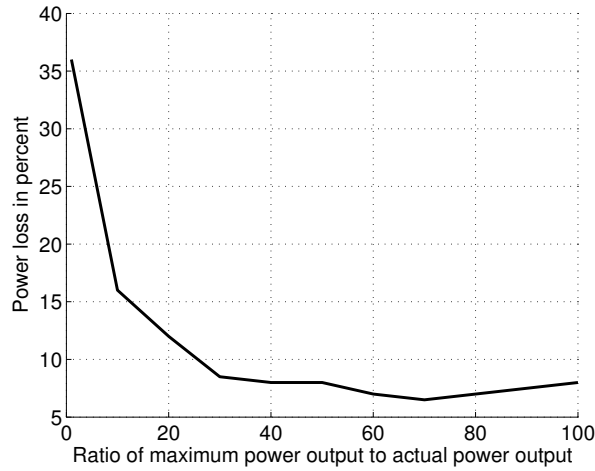


Figure 3.6: DC conversion loss function $l_{\text{DC}}(\zeta_{\text{DC}})$ as measured within the EARTH project.

sleep mode, $P_{\text{DC},s}$ in W,

$$P_{\text{DC},s} = P_{\text{DC}}. \quad (3.10)$$

The power consumption of the DC-DC-converter is shown as a function of f in Fig. 3.7. It is directly dependent on the components the DC-DC-converter powers and varies by less than 20% over all bandwidths. The reduced power consumption in sleep mode is only a passive effect and not an active switching process.

Mains Supply/AC-DC-Conversion

Power from the AC mains grid is converted to DC by the mains power supply unit. This unit is measured and modelled in the same fashion as DC-DC-conversion with conversion loss. See Fig. 3.8 for the measured losses as a function of the ratio of maximum power output to actual power output, ζ_{AC} , with

$$\zeta_{\text{AC}} = \frac{P_{\text{AC},\text{out},\text{max}}}{P_{\text{PA}} + P_{\text{RF}} + P_{\text{BB}} + P_{\text{DC}}}. \quad (3.11)$$

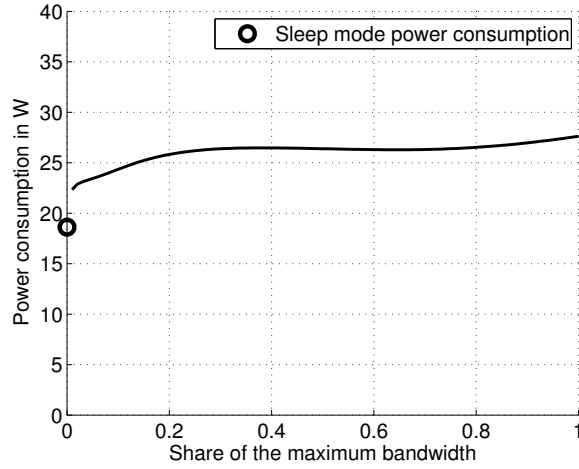


Figure 3.7: DC-DC-converter power consumption as a function of bandwidth used.

The power consumption caused by AC-DC-conversion is then found as

$$P_{AC} = l_{AC}(\zeta_{AC}) (P_{PA} + P_{RF} + P_{BB} + P_{DC}). \quad (3.12)$$

SotA AC-DC-converters are assumed to possess no special adaptation to sleep modes and, thus, remain activated during BS sleep mode, with power consumption during sleep mode, $P_{AC,S}$ in W,

$$P_{AC,S} = P_{AC}. \quad (3.13)$$

The power consumption of the AC-DC-converter is shown as a function of f in Fig. 3.9. Although the converter has non-linear efficiencies, these seem to compensate for the non-linearities of the components it powers. As a result, the power consumption curve is close to a straight line.

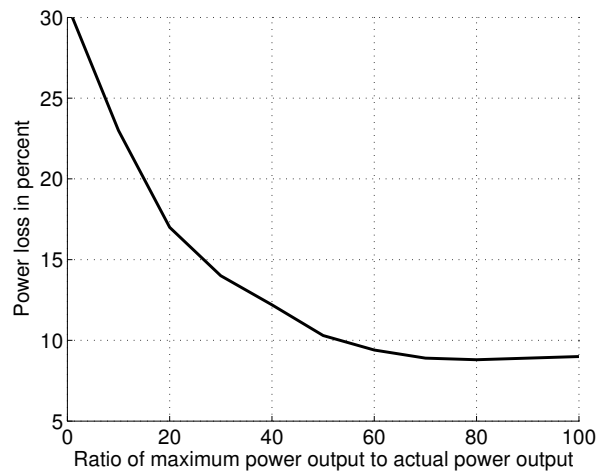


Figure 3.8: AC conversion loss function $l_{AC}(\zeta_{AC})$ as measured within the EARTH project.

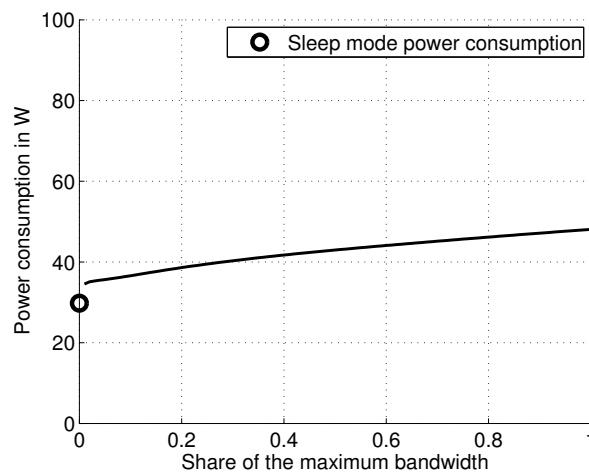


Figure 3.9: AC-DC-converter power consumption as a function of bandwidth used.

Cooling

Macro BSs are typically housed in a cooled cabinet. This cooling is a very difficult effect to model. Cooling requirements vary greatly between different geographic locations, positioning in or on buildings and the size of the storage cabinet. Different vendors recommend different operating temperatures for their devices and these recommendations may not be followed by the network operators. Furthermore, cooling is generally a very slow process compared to the orders of timing in a BS. Thus, considering cooling at its instantaneous power consumption may be misleading. As there is no simple solution to this problem, cooling is modelled as a fixed power loss.

The heat dissipation of all other components needs to be compensated for by active cooling. This heat dissipation can be assumed to be proportional to the power consumed. Thus, cooling power consumption is modelled to be proportional to the consumption of all other components. With cooling loss, $\xi_{\text{COOL}} = 0.12$, the power consumption of the cooling unit in W is

$$P_{\text{COOL}} = \xi_{\text{COOL}} (P_{\text{PA}} + P_{\text{RF}} + P_{\text{BB}} + P_{\text{DC}} + P_{\text{AC}}). \quad (3.14)$$

In this model, there are no further assumptions with regard to cooling during sleep mode. The power consumption of the cooling unit during sleep mode, $P_{\text{COOL,S}}$ in W, is therefore identical, with

$$P_{\text{COOL,S}} = P_{\text{COOL}}. \quad (3.15)$$

The power consumption of the cooling unit as a function of f is shown in Fig. 3.10.

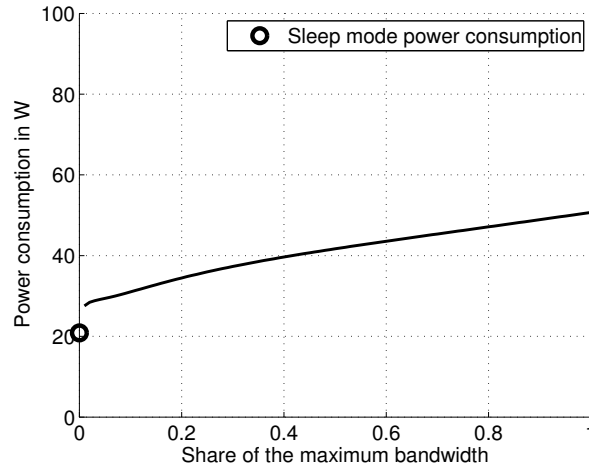


Figure 3.10: Cooling power consumption as a function of bandwidth used.

As it depends on all other components, it follows their power consumption behavior and is almost affine.

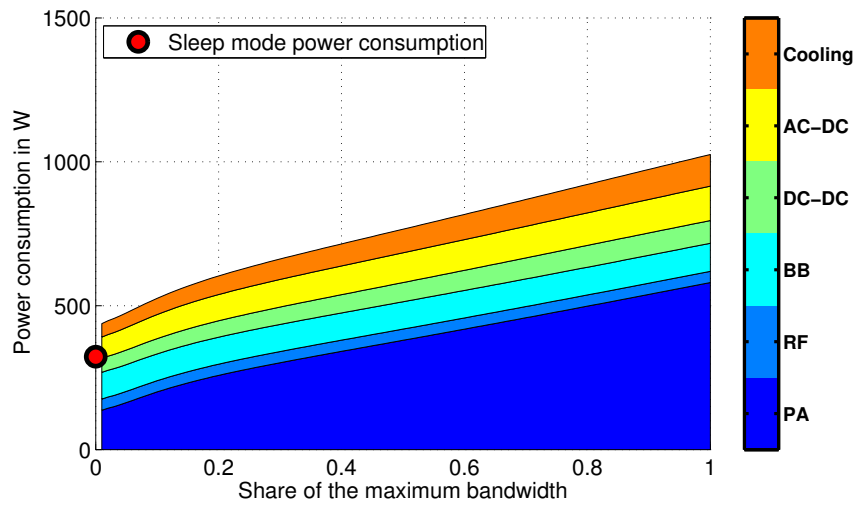
3.4.3 BS Power Consumption

The power consumption, P_{supply} in W, of an entire LTE BS is found as the sum of its components. With M_{sec} the number of sectors,

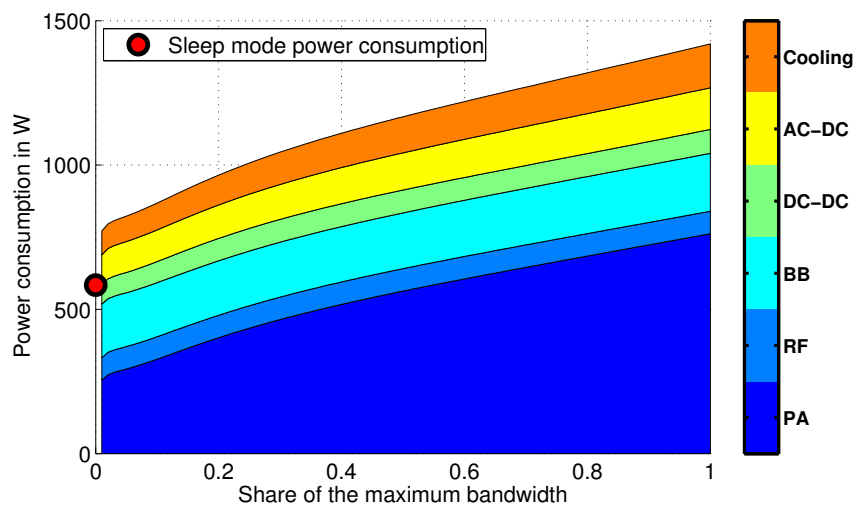
$$P_{\text{supply}} = M_{\text{sec}} (P_{\text{PA}} + P_{\text{RF}} + P_{\text{BB}} + P_{\text{DC}} + P_{\text{AC}} + P_{\text{COOL}}). \quad (3.16)$$

It is illustrated in Fig. 3.11 for $D = \{1, 2\}$. Fig. 3.12 provides a normalized comparison of the shares of power consumption at the two edge points of Fig. 3.11, namely at zero and at full bandwidth.

In both cases, for $D = 1$ and $D = 2$, respectively, the largest contributors to power consumption, especially at full bandwidth transmission, are found to be the PAs with 56.6% and 53.7%. At zero bandwidth, with 25.8% and 28.5%, the PAs still



(a) One antenna.



(b) Two antennas.

Figure 3.11: BS power consumption per component as a function of bandwidth used.

consume a larger share of the power than other components, but significantly less than for higher transmission bandwidths. Therefore, reducing the PA power consumption significantly affects the power consumption of the entire device.

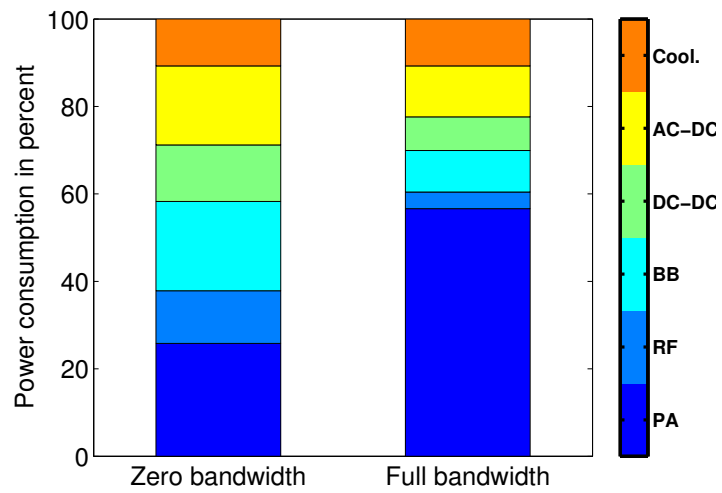
With regard to sleep modes, for $D = 1$ and $D = 2$, respectively, they are estimated to lower consumption to 322 W and 583 W compared to an overall consumption of 437 W and 771 W at zero bandwidth, and 1025 W and 1419 W at full bandwidth usage. That corresponds to a reduction of power consumption by 26% and 24% at zero bandwidth and by 69% and 59% at full bandwidth usage. Thus, sleep modes can reduce the supply power consumption significantly, especially in high load operation.

Overall, it is important to recognize that the power consumption of a BS is far from constant and can be varied through operation parameters by as much as 69%. To reduce its power consumption, it should be operated with low power, low bandwidth or in sleep mode. In addition, the slope of the variation is nearly constant over f for both one and dual antenna operation. This allows for a greatly simplified power consumption model.

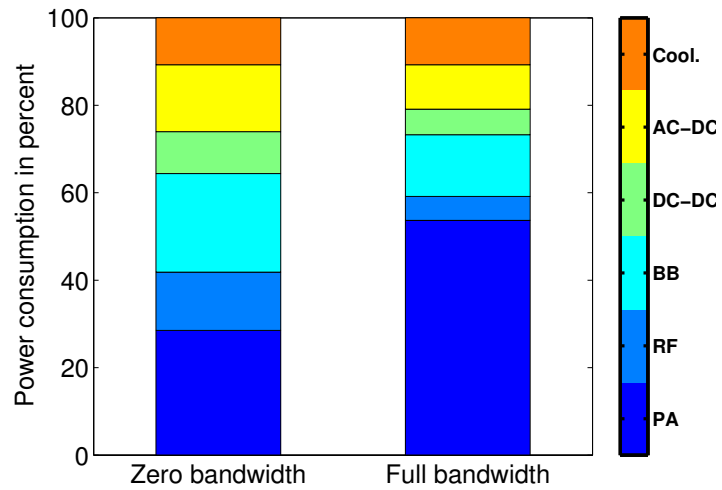
The power model described in this section can be generated for other than macro BS types such as micro, pico, femto or Remote Radio Head (RRH) in similar fashion. However, as this thesis is only concerned with the power consumption of macro type BSs, these types are not discussed here.

3.5 The Parameterized Power Model

The power model described above is very useful for purposes of derivation, understanding, calibration and exploration of changes in architecture. The effect of changes to individual components on the power consumption of the entire BS



(a) One antenna



(b) Two antennas

Figure 3.12: BS power consumption per component at zero and full bandwidth.

can be studied. However, it is often unnecessary to derive the power consumption of a BS from the component level up. When exploring power saving techniques, only few parameters are varied regularly while many remain constant for an entire study. To address this need, this section proposes a simpler, parameterized power model. It abstracts architectural details for simplicity and only keeps those parameters which are regularly altered. It is based on the identification of power saving techniques.

In the literature, specific power saving techniques are suggested: namely, the effects of the number of antennas on power consumption [26], the reduction of transmission power [d, 89, 90], the deactivation of unneeded antennas [49, 52], the adaptation of the transmission bandwidth [42, 57] and the use of low power consumption sleep modes of varied durations [43, 45, 58, 91].

The parameterized model encompasses all of these approaches to allow a direct comparison while abstracting parameters which can either be assumed to be constant or have little effect in the studied scenarios, such as GOPS, lookup tables, equipment manufacturing details or leakage powers. To this extent, only the following is covered in the proposed model:

- The different BS types of a heterogeneous network are modelled by applying different parameter sets to the same model equations.
- The number of transmission antennas and radio chains affects consumption during design and operation.
- The same holds true for transmission power, which affects the design indirectly by choice of a suitable power amplifier as well as the operation directly.
- Also, transmission bandwidth and sleep modes are modelled to allow the investigation of their impact on BS power consumption.

For parameterization, the maximum supply power consumption, P_1 in W, is first found establishing how power consumption scales with bandwidth, W , the number of BS radio chains/antennas, D , and the maximum transmission power, P_{\max} . This requires the consideration the main units of a BS as presented earlier: PA, RF small-signal transceiver, BB unit, DC-DC converter, mains supply and active cooling. Summarizing Section 3.4, the dependence of the BS units on W , D and P_{\max} can be approximated as follows:

- Both BB and RF power consumption, $P_{\text{BB}}^{\text{ppm}}$ in W and $P_{\text{RF}}^{\text{ppm}}$ in W, respectively, are assumed to scale nearly linearly with bandwidth, W , and the number of BS antennas, D .
- The PA power consumption, $P_{\text{PA}}^{\text{ppm}}$, depends on the maximum transmission power per antenna, P_{\max}/D , and the PA efficiency, η_{PA} . Also, possible feeder cable losses, σ_{feed} , have to be accounted for, such that

$$P_{\text{PA}}^{\text{ppm}} = \frac{P_{\max}}{D\eta_{\text{PA}}(1-\sigma_{\text{feed}})}. \quad (3.17)$$

- Losses incurred by DC-DC conversion, AC-DC conversion and active cooling scale linearly with the power consumption of other components and may be approximated by the constant loss factors σ_{DC} , σ_{AC} , and σ_{COOL} , respectively.

These assumptions allow calculating the maximum power consumption of a BS sector,

$$P_1 = \frac{\left[D \frac{W}{10 \text{ MHz}} (P_{\text{BB}}^{\text{ppm}} + P_{\text{RF}}^{\text{ppm}}) + P_{\text{PA}}^{\text{ppm}} \right]}{(1-\sigma_{\text{DC}})(1-\sigma_{\text{AC}})(1-\sigma_{\text{COOL}})}. \quad (3.18)$$

An important characteristic of a PA is that operation at lower transmit powers reduces the efficiency of the PA and that, consequently, power consumption is not a linear function of the PA output power. This is resolved by taking into account

the ratio of maximum transmission power of a PA from its data sheet, $P_{\text{PA,limit}}$ in W, to the maximum transmission power of the PA during operation, $\frac{P_{\text{max}}}{D}$. In typical PAs, the current transmission power can be adjusted by adapting the DC supply voltage, which impacts the offset power of the PA. The efficiency is assumed to decrease by a factor of θ for each halving of the transmission power. The efficiency is thus maximal when $P_{\text{max}} = P_{\text{PA,limit}}$ in single antenna transmission and is heuristically well-described by

$$\eta_{\text{PA}} = \eta_{\text{PA,max}} \left[1 - \theta \log_2 \left(\frac{P_{\text{PA,limit}}}{P_{\text{max}}/D} \right) \right], \quad (3.19)$$

where $\eta_{\text{PA,max}}$ is the maximum PA efficiency.

The reduction of power consumption in sleep mode is achieved by powering off PAs and reduced computations necessary in the BB engine. For simplicity, only the dependence on D is modelled as each PA is powered off. Thus, P_{S} , is approximated as

$$P_{\text{S}} = DP_{\text{S},0}, \quad (3.20)$$

where $P_{\text{S},0}$ is a reference value for the single antenna BS chosen such that P_{S} matches the complex model power consumption for two antennas.

It can be approximated from Fig. 3.11 to treat the supply power consumption as an affine function of the bandwidth and, thus,—given constant power spectral density—of the transmission power. In other words, the consumption can be represented by a static (load-independent) share, P_0 in W, with an added load-dependent share that increases linearly by a power gradient, Δ_{p} , see Figure 3.13. The maximum supply power consumption, P_1 in W, is reached when transmitting at maximum total transmission power, P_{max} . Furthermore, a BS may enter a sleep mode with lowered consumption, P_{S} , when it is not transmitting. Total power

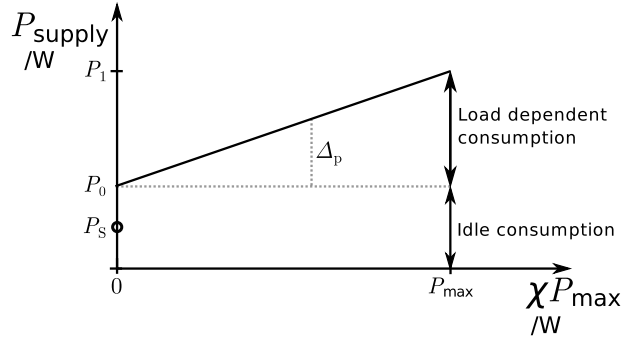


Figure 3.13: Load-dependent power model for an LTE BS.

consumption considering the number of sectors, M_{sec} , is then formulated as

$$P_{\text{supply}}(\chi) = \begin{cases} M_{\text{sec}} (P_1 + \Delta_p P_{\text{max}} (\chi - 1)) & \text{if } 0 < \chi \leq 1 \\ M_{\text{sec}} P_S & \text{if } \chi = 0, \end{cases} \quad (3.21)$$

where $P_1 = P_0 + \Delta_p P_{\text{max}}$. The scaling parameter, χ , is the load share, where $\chi = 1$ indicates a fully loaded system, *e.g.* transmitting at full power and full bandwidth, and $\chi = 0$ indicates an idle system.

The parameterized power model is applied to approximate the consumption of the BS which was described in the component model. Parameters are chosen where possible according to [21], such as losses, efficiencies and power limits. The remaining parameters are adapted such that a closer match to the component model could be achieved. The resulting parameter breakdown is provided in Table 3.4. The two power models are compared over f for one and two antennas in Figure 3.14. Although two parameters, the bandwidth and the number of BS antennas, are varied, the parameterized model can be seen to closely approximate the complex model for all BS types. The largest deviation of the parameterized model from the complex model occurs when modeling four transmit antennas. This is caused by the fact that the parameterized model considers a constant slope, Δ_p , which is independent of D . In contrast, the PA efficiency in the complex

$P_{\text{PA,limit}}$ /W	$\eta_{\text{PA,max}}$	θ	P_{BB} /W	P_{RF} /W	σ_{feed}	σ_{DC}	σ_{COOL}	σ_{AC}	M_{Sec}
80.00	0.36	0.15	29.4	12.9	0.5	0.075	0.1	0.09	3
			P_{max} /W	P_1 /W	Δ_p *10 MHz	$P_{\text{S,0}}$ /W			
			40.00	460.4	4.2	324.0			

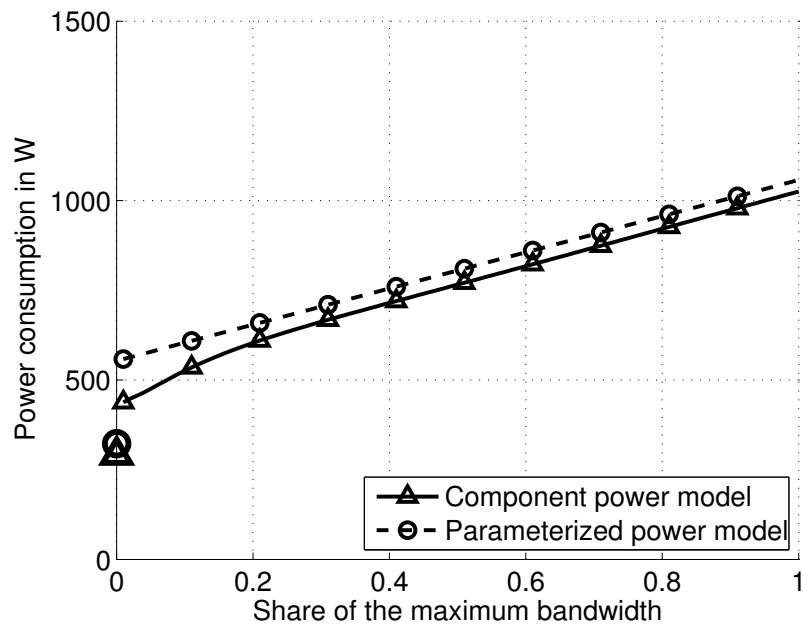
Table 3.4: Parameter breakdown.

model decreases with rising D , leading to an increasing slope which can not be matched by a constant slope. This deviation is a trade-off between simplicity and model accuracy.

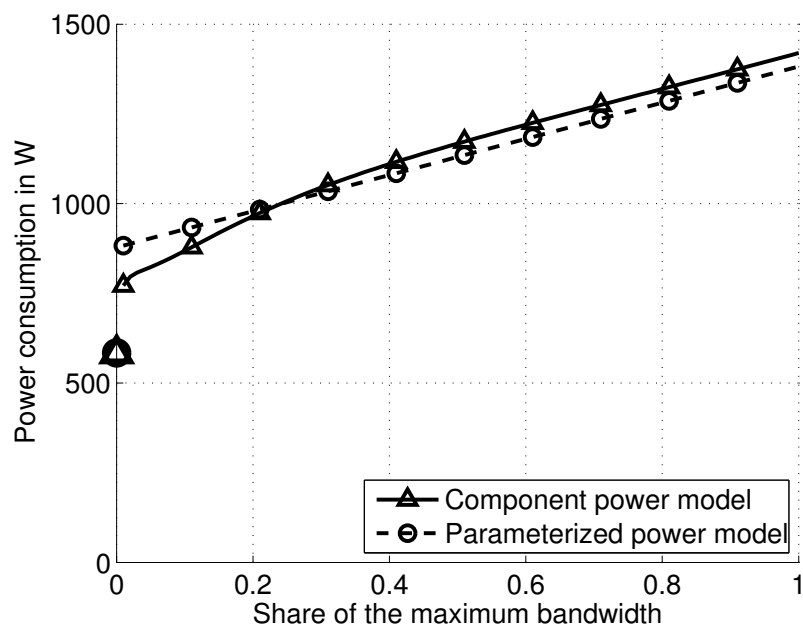
In addition to providing a solid reference, the model and the parameters can provide a basis for exploration. Individual parameters can be changed to observe the resulting variation in power consumption. With regard to the number of antennas, the parameterized model can only be verified up to four antennas, which is the extent of the complex model. Extending the system bandwidth, for example to 20 MHz, is expected to increase the BB and RF power consumption. The other parameters such as the transmission power and losses are expected to remain unaffected by different system bandwidths. Adapting the design maximum transmission power, P_{max} , affects the PA efficiencies, which decrease with P_{max} .

3.6 The Affine Power Model

In preparation for the following chapters of this thesis, one more simplification to power modeling is made. When the input parameters for (3.18) and (3.20) are not changed, the power model can be reduced to (3.21) as a function of the load χ . When considering Antenna Adaptation (AA), a separation has to



(a) One antenna



(b) Two antennas

Figure 3.14: Comparison of the parameterized with the complex model [e] power models for the macro BS type with 40 W transmission power.

be made depending on D . This simplification carries the benefit of turning the power model into an affine function. As will be shown in the next chapter, the treatment of power consumption as an affine function maintains convexity.

Let the general affine model be

$$P_{\text{supply}}(\chi) = \begin{cases} 3(P_{1,D} + \Delta_p P_{\text{max}}(\chi - 1)) & \text{if } 0 < \chi \leq 1, \\ 3 \cdot P_{S,D} & \text{if } \chi = 0, \end{cases} \quad (3.22)$$

with $P_{1,D} = P_{0,D} + \Delta_p P_{\text{max}}$.

For the remainder of this thesis, power consumption will therefore be modelled according to (3.22). This is a combination of the parameterized affine function and the component model sleep mode value in Fig. 3.14.

For later reference, the affine power model for a single antenna BS is explicitly written as

$$P_{\text{supply}}(\chi) = \begin{cases} 3(354 + 4.2P_{\text{max}}(\chi - 1)) & \text{if } 0 < \chi \leq 1, \\ 321 & \text{if } \chi = 0. \end{cases} \quad (3.23)$$

The affine power model for a two-antenna BS is

$$P_{\text{supply}}(\chi) = \begin{cases} 3(460 + 4.2P_{\text{max}}(\chi - 1)) & \text{if } 0 < \chi \leq 1, \\ 648 & \text{if } \chi = 0. \end{cases} \quad (3.24)$$

A tabular summary is provided in Table 3.5.

D	$P_{1,D}$ /W	$P_{0,D}$ /W	$P_{S,D}$ /W	Δ_p
1	354	186	107	4.2
2	460	292	216	4.2

Table 3.5: Summary of affine power model parameters.

3.7 Summary

In this chapter, three different power models have been derived, namely the component model, the parameterized model, and the affine model. The component model was derived by first defining the architecture of a typical LTE BS and then inspecting each component with regard to its power consumption. The PA was found to have a power consumption which strongly depends on the transmission power with some capability for sleep mode. The RF transceiver in its current implementation has limited adaptability and, therefore, constant power consumption. The BB unit's power consumption is determined by the internal computing operations, some of which can be reduced for power saving. The power consumption components such as power conversions and cooling, strongly depends on the power consumption of other components. Altogether, the power consumption of the individual components comprises the component BS power model. While very detailed and necessary for derivation, the component model is too elaborate for many uses. To address this complexity, the parameterized power model was derived from it, which only considers typical operating parameters while abstracting architectural and technical details. It was shown to closely approximate the component model for a range of parameters. Instead of using measured data ranges containing many values in the model, all input parameters are generalized to scalars (single values). When employed in an analytical context, a further fixation of input parameters is advisable, yielding the affine model. Overall, the number of input parameters including experimental data was reduced

from 15 parameters (out of which 8 are data ranges) in the component model to 14 scalar parameters in the parameterized model and to four scalar parameters in the affine model. See Table 3.6 on page 69 for the detailed comparison.

Parameter	Component	Parameterized	Affine
	model	model	model
Feeder loss	S	S	
PA efficiency	V	S	
PA sleep consumption	S		
RF consumption	V	S	
BB consumption	V	S	
Scaling exponent for D	V		
Scaling exponent for f	V		
Sleep switching BB	V		
DC-DC loss	V	S	
Maximum DC-DC output	S		
AC-DC loss	V	S	
Maximum AC-DC output	S		
Cooling efficiency	S	S	
Number of sectors	S	S	
Number of radio chains	S	S	S
Maximum transmission power		S	
PA efficiency decrease		S	
PA limit		S	
Sleep consumption reference		S	S
Power consumption load factor		S	S
Maximum power consumption			S
Number of data range vectors (V)	8	0	0
Number of scalars (S)	7	14	4

Table 3.6: Comparison of required input parameters for different power models.

Chapter 4

Power Saving on the Cell Level (Single-cell)

4.1 Overview

As introduced in Chapter 2, current cellular systems are capacity optimized with little consideration of the power consumed. In contrast, for power saving, networks need to be efficiency optimized. The difference between a power efficient and a capacity maximizing system is that a power efficient system only consumes the resources that are *needed* while a capacity maximizing system consumes those that are *available*. As described in Section 2.2, cellular systems are rarely fully loaded and at times completely unloaded [92]. This implies that not all of the bandwidth, time or transmission power are needed. The power-saving Radio Resource Management (RRM) mechanisms proposed in this chapter address this issue and exploit spare capacity to increase the system efficiency.

In this chapter the previously established power models are applied to study

power-saving Base Station (BS) operation. Section 4.2 discusses the State-Of-The-Art (SotA) of power-saving RRM in the literature. It is then identified in Section 4.3 how a transmission consumes power and which parameters should be optimized to minimize consumption on the basis of general communication theory. Trade-offs are identified between Time Division Multiple Access (TDMA) and Power Control (PC). In Section 4.4, the solution to multi-user BS PC is proposed and studied. Once the system is optimized with regard to PC, Discontinuous Transmission (DTX) is added to further drive down power consumption via joint optimization. The general real-valued TDMA solution is derived. In Section 4.5, the convex optimization problem is then extended into a single-cell multi-user Multiple-Input Multiple-Output (MIMO) Orthogonal Frequency Division Multiple Access (OFDMA) power-saving algorithm. The chapter is summarized in Section 4.6.

The material presented in this chapter has previously been published or submitted for publication in [b, c, d, j, l].

4.2 Power-saving RRM in Literature

In the literature, PC is the most prominent power-saving RRM mechanism with numerous proposals for a range of different problems, for example the related works in [93, 94, 95, 96, 97]. This prominence is due to the fact that, in addition to reducing power consumption, PC is also beneficial to link adaptation and interference reduction. Sinanović *et al.* [93] explore the optimal power points in a network with two interfering links and provide a communication theoretic basis for this work. Wong *et al.* [94] minimize transmit power for multi-user Orthogonal Frequency Division Multiplexing (OFDM) under rate constraints. Although their work is pioneering, it has the drawbacks that the PC algorithm is computationally

complex and does not consider a transmit power constraint or a power model. Miao *et al.* [95] presented an early work on efficiency in which they derive the uplink data rate which maximizes the transmitted information per unit energy (bit per joule). Kivanc *et al.* [96] allocate subcarriers such that transmission powers are reduced. Their algorithm is modified and applied in this chapter. Al-Shatri *et al.* [97] approach sum rate maximization by using margin-adaptive power allocation, a method which is also used towards the end of this chapter.

More recently, acknowledging the significance of transmission-power-independent power consumption for energy efficiency, DTX and Antenna Adaptation (AA) have been identified as promising energy saving RRM techniques. The term *DTX* refers to an interruption of transmission which can be used to enter a short sleep mode. This interruption can be short enough to be unnoticeable to a receiver and fit into an OFDMA frame. This allows to schedule DTX flexibly without affecting the User Equipment (UE). Thus, DTX carries the benefit of not increasing delay or inhibiting discovery compared to longer sleep modes. DTX is first proposed in [43]. To the best of the author's knowledge, there are no prior works which consider DTX in combination with other power saving RRM mechanism. With regard to AA, Cui *et al.* analyze the energy efficiency of MIMO transmissions, being the first to consider the supply power consumption in energy efficient operation [26]. They show that in terms of energy efficiency, the optimal number of transmit antennas used for MIMO transmission is not the highest by default, but rather depends on the Signal-to-Noise-Ratio (SNR). Kim *et al.* [51] establish AA as a MIMO resource allocation problem and adapt the number of transmit antennas on a single link. However, solutions provided for AA on the link level cannot be directly applied to the BS, since for the latter multiple transmissions are active simultaneously on the same set of antennas. Xu *et al.* [52] propose a convex optimization scheme to reduce the number of transmission antennas during transmission. A related approach is applied in this

chapter to the optimization of sleep modes. Hedayati *et al.* [49] propose to perform AA without informing the receiver, thus mimicking deep fades on some antennas. This is predicted to reduce power consumption by up to 50%.

Note that—as mentioned in Chapter 2—there is also a number of works concerned with *long sleep* modes, such as [41, 42, 98, 99]. In comparison with DTX, such long sleep modes are applied for minutes, hours or days. This increases the achievable power saving, as devices are completely turned off. However, long sleep modes cause problems with coverage and discovery, as two communication devices may be completely unaware of one another. Also, such sleep modes clearly affect the Quality of Service (QoS), unless redundant network elements exist. Thus, they are not considered here. The work presented in this chapter can be combined with long sleep modes as part of future work as the two are not mutually exclusive.

4.3 PC and TDMA

When the number of bits transmitted on a link with fixed bandwidth and optimal modulation and coding is to be increased, there are two options: either to increase the link rate by raising the transmission power or to transmit for a longer time. From a power efficiency perspective, providing a fixed rate is thus a trade-off between transmit power and transmission duration. A higher power for smaller duration can provide the same rate as a lower power with higher transmission duration. Therefore, for power minimization on a single link, transmission should always be as long as possible to allow for the lowest transmit power. However, in a shared multi-user channel with orthogonal access, such as Time Division Multiplexing (TDM), Frequency Division Multiplexing (FDM), or OFDM, all links have to be considered which each have individual rate requirements that have to be fulfilled in a given time. This problem is illustrated in Fig. 4.1 for

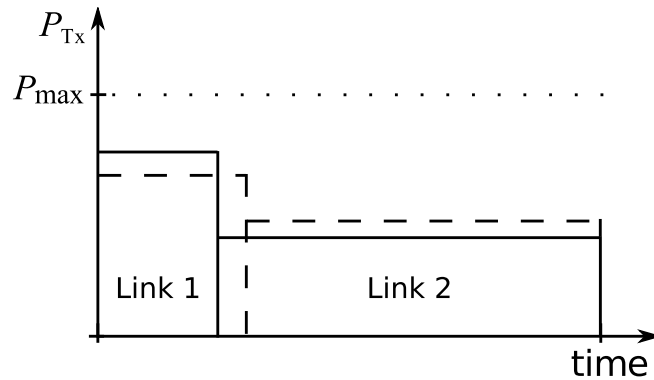


Figure 4.1: Illustration of two possible power/time trade-offs that provide equal rates on both links.

two links. Which combination of transmission duration and transmission power achieves the lowest overall power consumption is the problem discussed in this section.

Power allocation on the link level

It follows a derivation of the power optimal allocation strategy of transmission power and time for a single link. The Shannon bound [100]—as one of the most fundamental laws in communications—is employed to map transmission power to achievable rate. It is assumed that in a cellular system a set of mobiles with known channel gains has a set of rate requirements that needs to be fulfilled.

A cell consists of one BS and K links (mobiles). The capacity in bps per link k is upper bounded by the Shannon bound as

$$R_k = W \log_2(1 + \gamma_k), \quad (4.1)$$

where W is the channel bandwidth in Hz and $\gamma_k = \frac{G_k P_k}{P_N}$ is the SNR with G_k the link channel gain, P_k the transmit power on link k in W and P_N thermal

noise power in W. The noise power is defined as $P_N = Wq_B\vartheta$ with Boltzmann constant q_B and operating temperature ϑ in Kelvin. While capacity and bandwidth are linearly related, capacity and transmit power have a logarithmic relationship. As a consequence it is much more expensive in terms of power to increase the channel capacity than in terms of bandwidth. In other words, if there is a choice between

- a) leaving idle spectrum and transmitting at higher power, and
- b) using all available spectrum and transmitting at the lowest required power,

then the latter will always consume less overall transmit power.

Power allocation on the cell level

Next, an optimization problem is proposed which describes this notion for K users on a shared channel. Since all links in the system share the available resources, the RRM of one link affects all others. Thus, power-optimization can not be performed on the link level only, but has to be approached from the cell level.

Let the normalized transmission time per link be given by

$$\mu_k = \frac{\tau_k}{\tau_{\text{frame}}}, \quad (4.2)$$

where τ_k is the transmission time on link k in seconds, τ_{frame} is the frame duration in seconds. Optimization of normalized time rather than absolute time is chosen here as it results in average power minimization and is more illustrative than energy minimization (which is only meaningful for a known τ_{frame}).

The average capacity, \bar{R}_k , on link k depends on μ_k and R_k , with

$$\bar{R}_k = \mu_k R_k, \quad (4.3)$$

where \bar{R}_k indicates the average of R_k .

Solving (4.1) for P_k and combining with (4.3), the transmission power on link k in W as a function of the required average rate is found to be

$$P_k(\bar{R}_k) = \frac{P_N}{G_k} \left(2^{\frac{\bar{R}_k}{W\mu_k}} - 1 \right), \quad (4.4)$$

where $0 < P_k(\bar{R}_k) < P_{\max}$ for some P_{\max} .

To account for the fact that all links are served by the BS orthogonally within some time τ_{frame} , the system average transmission power at the BS for all links, \bar{P}_{Tx} , is the sum of individual transmit powers weighted with the transmit duration, with

$$\bar{P}_{\text{Tx}} = \sum_{k=1}^K \mu_k P_{\text{Tx},k}(\bar{R}_k) \quad (4.5a)$$

$$= \sum_{k=1}^K \mu_k \frac{P_N}{G_k} \left(2^{\frac{\bar{R}_k}{W\mu_k}} - 1 \right). \quad (4.5b)$$

The combined duration of all transmissions must be less or equal to τ_{frame} . But since it is most efficient to use the entire available τ_{frame} as derived from (4.1), it holds that

$$\sum_{k=1}^K \mu_k = 1. \quad (4.6)$$

Furthermore, time has positive values and a power constraints must be obeyed

with

$$\mu_k > 0 \quad \forall k, \quad (4.7a)$$

$$0 < P_k(\bar{R}_k) < P_{\max} \quad \forall k. \quad (4.7b)$$

The allocation vector $\boldsymbol{\mu}^* = (\mu_1, \dots, \mu_K)$ of transmission durations which minimizes (4.5b) is power optimal.

Evaluation

Next, the affine power model is taken into account, as it is important to consider transmission power in relationship with the consumption of the entire BS. Instead of the transmission power, the supply power consumption is minimized. With the power model from (3.22), the PC cost function (4.5b) and $\chi P_{\max} = P_k(\bar{R}_k)$, the supply power consumption in W is found to be

$$\begin{aligned} P_{\text{supply}}(\bar{R}_k) &= \sum_{k=1}^K \mu_k (P_0 + \Delta_p P_{\text{Tx},k}(\bar{R}_k)) \\ &= \sum_{k=1}^K \mu_k \left(P_0 + \Delta_p \frac{P_N}{G_k} \left(2^{\frac{\bar{R}_k}{W}} - 1 \right) \right). \end{aligned} \quad (4.8)$$

Note that, against intuition, the power model has no effect on the solution of (4.8), *i.e.* $\boldsymbol{\mu}^*$ is equal for (4.5b) and (4.8). As the power model affects all links equally, the PC solution is the same for any power model and, thus, any BS. However, although the *solution* is independent of the hardware, the *benefits* of PC do strongly depend on hardware. As a metric for characterizing hardware by the share of load-dependent consumption of the overall consumption, the load

dependence of a BS is defined as

$$\eta_{\text{ld}} = \frac{\Delta_{\text{p}} P_{\text{max}}}{P_0 + \Delta_{\text{p}} P_{\text{max}}}. \quad (4.9)$$

Using this load dependence, the benefits of PC can be evaluated. Figure 4.2 shows the supply power in a cell with ten users for a range of load dependence values after the application of the PC strategy in (4.5b). The slope of the power model Δ_{p} was chosen for a representative set of load dependences with $\eta_{\text{ld}} = \{0.1, 0.3, 0.5, 0.8\}$ for $P_1 = 460 \text{ W}$ and $P_{\text{max}} = 40 \text{ W}$, from (3.22). Channel gains were calculated based on distance by the 3rd Generation Partnership Project (3GPP) path loss model [82] after dropping users uniformly on a disc with 250 m radius around the BS in the center. The problem (4.8) was solved iteratively using the interior-point algorithm contained in the MATLAB Optimization Toolbox. The remainder of the simulation parameters was set as described in Table 4.1.

As shown in Fig. 4.2, the effectiveness of PC strongly depends on the underlying hardware. A higher load dependence factor yields a lower supply power consumption for PC. The effectiveness of PC in future BSs therefore strongly depends on hardware developments.

Although PC is thus shown to provide significant saving potential in BSs with high η_{ld} , like macro BSs, the power consumption of a device employing PC is always lower bound by its idle power consumption, P_0 . Even when transmission is stopped, a device employing only PC still consumes a supply power consumption of P_0 . Thus, to further reduce power consumption, a method is required which allows power consumption to reach lower values than P_0 . This can be achieved by employing sleep modes or—in the case of OFDMA scheduling—DTX.

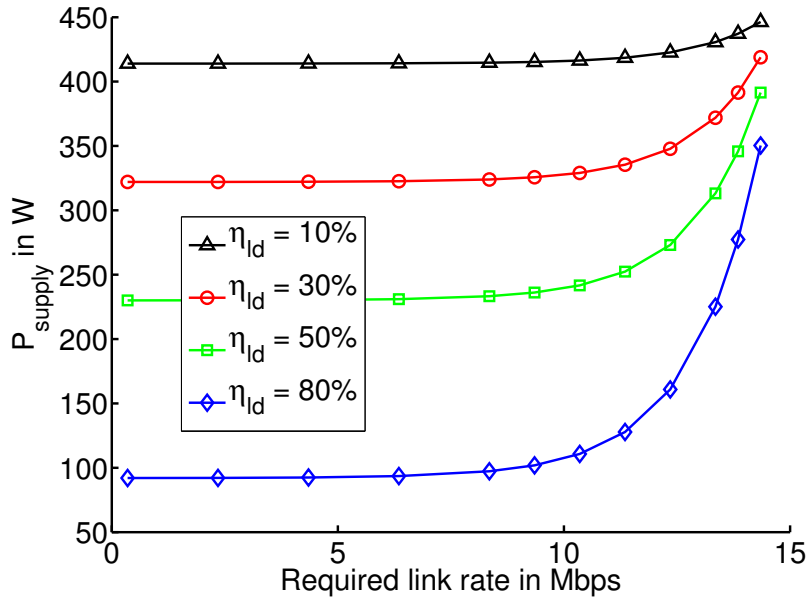


Figure 4.2: Comparison of the effect of load dependence on achievable power savings by PC.

Parameter	Value
Carrier frequency	2 GHz
Cell radius	250 m
Path loss model	3GPP UMa, NLOS, shadowing [82]
Shadowing standard deviation	8 dB
Iterations	1,000
Bandwidth W	10 MHz
Maximum transmission power P_{\max}	46 dBm
Operating temperature T	290 K
System/subcarrier bandwidth	10 MHz/200 kHz
Thermal noise power per subcarrier	$4w \times 10^{-21}$ W

Table 4.1: Simulation parameters.

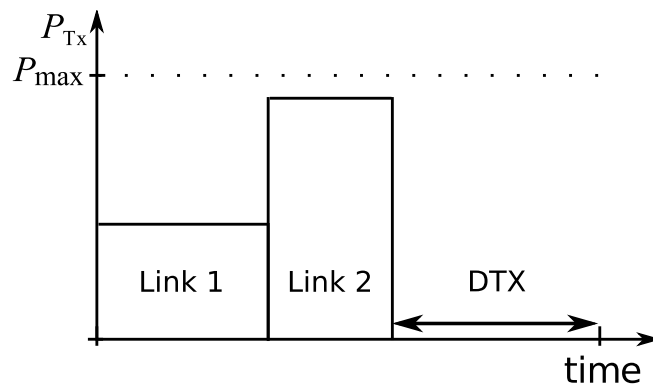


Figure 4.3: Illustration of PRAIS for two links: transmit power, resource share and DTX are allocated optimally in order to serve the requested rate at minimal supply power consumption.

4.4 Power and Resource Allocation Including Sleep (PRAIS)

This section introduces the PRAIS scheme, which combines three techniques: PC, TDMA and DTX. See Figure 4.3 for an illustration of PRAIS on two links.

The parameterized power model in Section 3.5 derived three general mechanisms which affect the power consumption of BSs:

1. The overall power emitted at the antenna(s), P_{Tx} , affinely relates to supply power consumption, see (3.22).
2. If the same BS is operated with fewer Radio Frequency (RF) chains, it consumes less power, see Fig. 3.14.
3. If the BS is put into sleep mode, it consumes less power than in the active (idle) state with zero transmission power, see Fig. 3.14.

These observations lead to the following intuitive saving strategies:

PC : Reduce transmission power.

AA : Reduce the number of RF chains.

DTX : Increase the time the BS spends in DTX.

The first and the last strategy are clearly opposing each other as lower transmission powers lead to lower link rates and thus longer transmission duration (for transmitting a certain number of bits), whereas it would be beneficial for long DTX to have short transmissions. The second and third strategy are related, as AA can be considered a weak form of DTX which still allows transmission on a subset of antenna elements. In this section, PC and DTX are jointly addressed. AA is treated in the following section, Section 4.5.

Joint PC and DTX

When employing DTX individually to minimize power consumption, the optimal strategy is to serve all links at the maximum available transmission power until target rates are fulfilled, and then set the BS to DTX. Alternatively, when employing PC with TDMA individually for maximum efficiency, the transmission duration is stretched over the available time frame with the lowest power necessary to serve the required rates. When PC and DTX are combined, there is a trade-off. Between the two extremes of maximum transmit power with longest DTX and lowest transmit power with no DTX, there are configurations with medium transmit power and medium DTX duration that consume less supply power.

This can be illustrated graphically for a single link example, as shown in Fig. 4.4. Plotted is the supply power consumption caused by transmitting a fixed number of bits. Let $\Phi \in [0, 1]$ denote the share of time spent transmitting. Three operation modes are defined:

- PC: Only PC but no DTX is available. To this extent, the transmission power is adjusted depending on the transmission time normalized to the time slot duration, Φ , such that the target rate is achieved. The BS consumes idle power P_0 when there is no data transmission. Clearly, the point of lowest power consumption in this case is at $\Phi = 1$ where the supply power consumption approaches P_0 . Lower values of Φ increase the required transmission power, until at $\Phi = 0.18$, $P_{\text{Tx}} = P_{\text{max}}$.
- DTX: The BS transmits with full power $P_{\text{Tx}} = P_{\text{max}}$. Thus, the supply power is $P_0 + \Delta_p P_{\text{max}}$ when transmitting or P_S when in DTX mode, yielding an affine function of Φ . At $\Phi = 1$, a higher rate than the target rate is achieved. Reducing Φ from $\Phi = 1$ decreases the supply power consumption linearly up to the point at which the target rate is met with equality. In this mode of operation, the best strategy is to minimize the time transmitting (small Φ), in order to maximize the time in DTX, $(1 - \Phi)$. In Fig. 4.4, the point of lowest power consumption for this mode is at $\Phi = 0.19$ with 154 W.
- Joint application of PC and DTX: In this mode of operation, the DTX time $(1 - \Phi)$ is gradually reduced. The transmission power is adjusted to meet the target rate within Φ . Here, the point of lowest power consumption is at $\Phi = 0.25$ with 135 W.

Inspection of Fig. 4.4 reveals that the joint operation of PC and DTX always consumes less power than either individual mode of operation with an optimal point at $\Phi = 0.25$. This is the power-saving benefit provided by the joint optimization of PC and DTX.

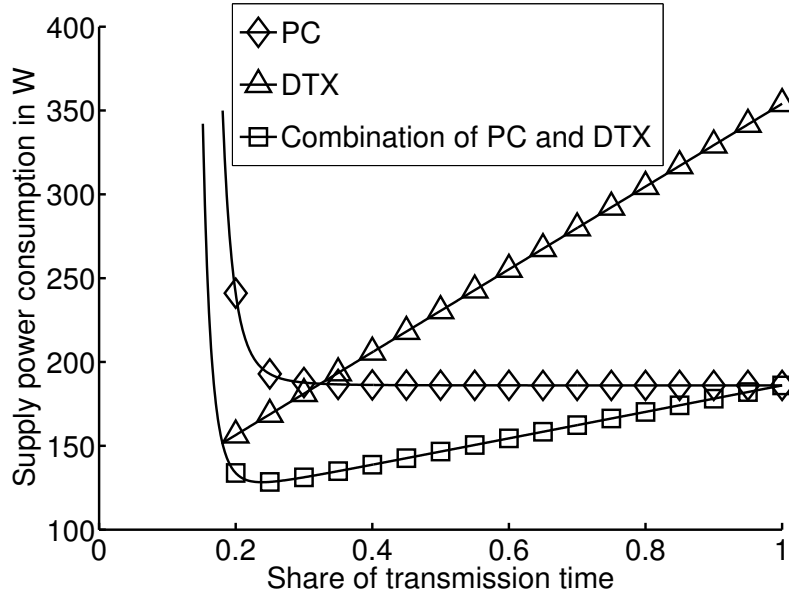


Figure 4.4: Supply power consumption for transmission of a target rate as a function of time spent transmitting, Φ . The combination of DTX and PC achieves lower power consumption than exclusive operation of either DTX or PC.

Problem formulation

This section proceeds to derive the power consumption optimization problem for a multi-user single-cell allocation of transmit powers, DTX times and transmit durations contained within the PRAIS scheme. The normalized time spent transmitting, μ_k , is defined as in (4.2) and the normalized duration spent in DTX is

$$\nu = \frac{\tau_S}{\tau_{\text{frame}}}, \quad (4.10)$$

where τ_S is the time spent in DTX and

$$\sum_{k=1}^K \mu_k + \nu = 1. \quad (4.11)$$

Combining this with the consumption during DTX from (3.22) and (4.5b) yields

the following PRAIS optimization problem.

$$\underset{\mu, \nu}{\text{minimize}} \quad P_{\text{supply}}(\bar{R}_k) = \left[\sum_{k=1}^K \mu_k \left(P_0 + \Delta_p \frac{P_N}{G_k} \left(2^{\frac{\bar{R}_k}{W \mu_k}} - 1 \right) \right) \right] + \nu P_S \quad (4.12a)$$

$$\text{subject to} \quad \sum_{k=1}^K \mu_k + \nu = 1, \quad (4.12b)$$

$$\nu \geq 0, \quad (4.12c)$$

$$\mu_k \geq 0 \quad \forall k, \quad (4.12d)$$

$$0 \leq P_k = \frac{P_N}{G_k} \left(2^{\frac{\bar{R}_k}{W \mu_k}} - 1 \right) \leq P_{\max} \quad \forall k. \quad (4.12e)$$

The constraints (4.12b), (4.12c), (4.12d) reflect the facts that normalized time has to be positive and its sum unity. Transmission powers are positive and bounded by a maximum transmit power in (4.12e). The cost function is a non-negative sum of functions that are convex within the constraint domain and is therefore still convex. See Appendix A.1 for a proof. It can be solved efficiently with appropriate software like the MATLAB optimization toolbox. The solution of (4.12) determines the vector $(\mu_1, \dots, \mu_K, \nu)$, which minimizes the overall power consumption under target rates. Note that selecting $P_S \geq P_0$ is equivalent to disabling DTX, *i.e.* employing PC individually similar to (4.8), whereas fixating $P_{\text{Tx}} = P_{\max}$ is equivalent to disabling PC.

The resource allocation problem (4.12) can be solved for any number of users. Without loss of generality, a ten user scenario for numerical evaluation is selected.

Evaluation

For the numerical analysis, the PRAIS scheme is evaluated in a Monte Carlo simulation using the parameters shown in Table 4.1. Users are dropped uniformly

onto a disk and the associated channel gains G_k are generated by applying the 3GPP urban macro path-loss model including shadowing with a standard deviation of 8 dB [82]. In addition to the individual DTX and PC allocation schemes and the PRAIS scheme, two references that serve as upper limits are presented. First, the maximum transmission power as defined by the Long Term Evolution (LTE) standard provides the theoretical reference. Second, the reference against which gains are measured is the power behavior of a BS as defined in the power models. This is a Bandwidth Adaptation (BA) scheme where each user receives a share of the frequency band as well as the entire considered time slot. This can also be interpreted as ‘DTX in the frequency domain’ where $P_S = P_0$.

For power modeling, representative of current and future developments, the affine power model from Chapter 3, a model from the literature and a purely theoretical model are selected; the parameters are normalized for a single sector, single sector and listed in Table 4.2.

For an assessment of the relevance of DTX, an assumption is that DTX will greatly improve, while standby component consumption cannot be reduced substantially in the coming years [43]. This model is labelled the *Frenger model*. This model particularly emphasizes DTX effects.

As a best-case example, the theoretical power model assumes idealized components which scale perfectly with load. Power consumption is set to scale almost linearly with load with near-zero stand-by consumption. This model is not a prediction of technology advances, but provides theoretical limits. Parameters are selected such that the power consumption at full load matches the affine model.

The simulation results are shown in Figure 4.5 where the per-user link rate

Model	P_0/W	Δ_P	P_S/W
Affine model in (3.23)	186	4.2	107
Frenger model [43]	170	3.4	10
Linear model	1	8.8	1

Table 4.2: Single-sector power model parameters used in Section 4.4.

is plotted against the average supply power consumption under the different schemes. Figures 4.5a, 4.5b, 4.5c reflect the three chosen power models.

It is found in Figure 4.5a that a Single-Input Single-Output (SISO) LTE BS consumes 186 W up to 340 W employing BA, which is considered the operation of the SotA. An important albeit foreseen result is that the consumption curves of BA and PC as well as DTX and PRAIS originate in the same values of P_{supply} at low load. The higher one is P_0 at 186 W and the lower one P_S at 107 W. This is true for all power models and confirms that in an empty cell, power consumption is determined by hardware and resource allocation has no effect.

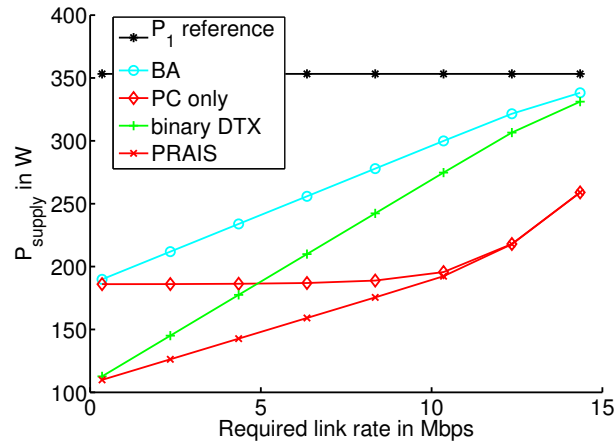
As shown in Fig. 4.5a, the application of PC only allows keeping the overall consumption constant for a large set of low target rates. In this rate region, transmit powers are very low compared to standby consumption. Only when rates above 10 Mbps are targeted is the transmission power high enough to make a noticeable difference compared to the standby consumption. This is reflected in the rising PC only curve at target rates above 10 Mbps. In contrast, the binary DTX scheme (which transmits either at full power or not at all) has much lower power consumption (up to -45%) than PC only or BA at low rates. However, it rises much quicker than the PC only curve. There is a crossover point at 5.6 Mbps between DTX and PC. The PRAIS scheme, which joins PC and DTX, benefits from both individual schemes and has minimal consumption over all rates. At higher target rates, the PRAIS consumption curve joins the PC only curve. This

reflects that it is not feasible to put the BS to sleep at high rates. Employing the affine power model, the PRAIS scheme achieves savings between 42% and 23% over BA.

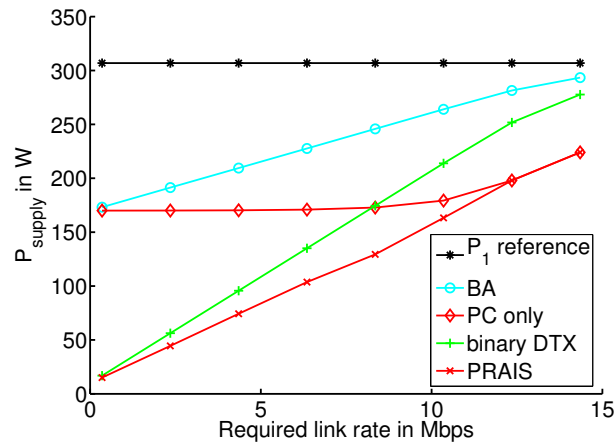
A surprising result is found in Figure 4.5b. Although application of this model has a strong bias towards DTX effects, there remains a cross-over point between DTX and PC after which the use of PC can still save 19% of the supply power consumption in addition to DTX. In the Frenger model, the PRAIS scheme offers between 91% savings at near-zero rates and 23% savings at 15 Mbps.

In the future linear model shown in Figure 4.5c, the behavior of DTX and BA, as well as PC and PRAIS, are identical, due to the fact that there is no gain of sleep modes over standby consumption. BA consumption is significantly lower than for all other power models. All benefits are owed to PC which is strongly amplified by the linear model behavior, delivering significant savings over all target rates.

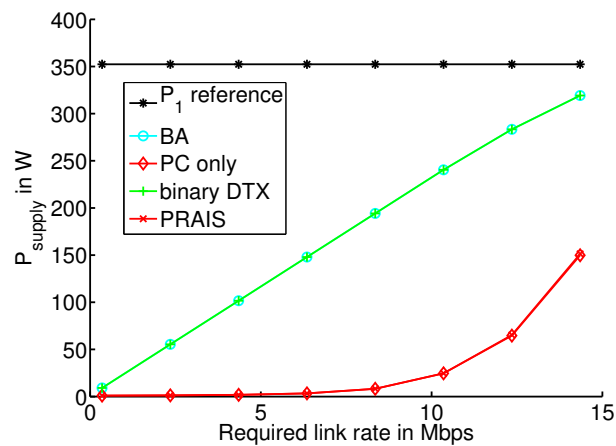
Another finding evident from all figures is that there is no point of ‘full load’ in a cell which causes maximum power consumption. In theory, when a cell is fully loaded—regardless of the resource allocation scheme—the BS is expected to consume the maximum supply power, P_1 . However, the chance that the set of target rates matches the given channel conditions to generate a situation of ‘full load’ is extremely low. It is much more probable that either the channels conditions are good enough to have ‘almost full load’ or that the channel conditions are too bad to fulfill the target rates, generating an overload or outage situation. As target rates increase, the probability of the latter increases. To cover only a representative set of target rates, only those data points are plotted which contain less than 10% overload/outage. Due to this outage, the simulated power consumption never reaches the theoretical maximum. Thus, the selection of the resource allocation scheme does in fact matter, even at near-full loads.



(a) Supply power consumption under the affine model.



(b) Supply power consumption under the Frenger model.



(c) Supply power consumption under the linear model.

Figure 4.5: Fundamental limits for power consumption in BSs.

4.5 Resource allocation using Antenna adaptation, Power control and Sleep modes (RAPS)

To extend the previous analytical work into a comprehensive, practical mechanism for current cellular systems like 3GPP LTE, the RAPS algorithm is proposed in this section, which reduces the BS supply power consumption of multi-user MIMO-OFDM. Given the channel states and target rates per user, RAPS finds the number of transmit antennas, the number of DTX time slots and the resource and power allocation per user. The RAPS solution is found in two steps: First, PC, DTX and resource allocation are joined into a convex optimization problem which can be efficiently solved. Second, subcarrier and power allocation for a frequency-selective time-variant channel pose a combinatorial problem, which is solved by means of a heuristic solution.

One transmission frame is considered in the downlink of a point-to-multipoint wireless communication system, comprising one serving BS and multiple mobile receivers. The BS transmitter is equipped with D antennas and all antennas share the transmit power budget. Mobile receivers have M_R antennas and system resources are shared via OFDMA between K users on N subcarriers and T time slots. In total, there are NT resource blocks as introduced in Section 2.5.2. A frequency-selective time-variant channel is assumed, with each resource unit characterized by the channel state matrix, $\mathbf{H}_{n,t,k} \in \mathbb{C}^{M_R \times D}$, with subcarrier index $n = 1, \dots, N$, time slot index $t = 1, \dots, T$, user index $k = 1, \dots, K$. The vector of spatial channel eigenvalues per resource unit a and user k is $\mathcal{E}_{a,k}$ and its cardinality is $\min\{D, M_R\}$. The system operates orthogonally such that individual resource units cannot be shared among users. MIMO transmission with a variable number

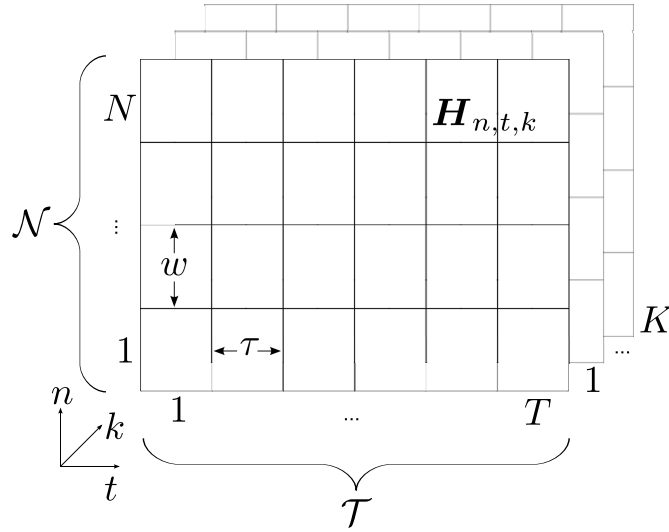


Figure 4.6: OFDM frame structure.

of spatial streams is assumed over the set of resources assigned to each user, \mathcal{U}_k . This frame structure is illustrated in Fig. 4.6.

The PC and DTX trade-off is extended to MIMO-OFDM serving multiple users over frequency-selective channels. The selection of the number of transmit antennas for AA is made once for the entire frame.

For power modeling of single antenna and dual antenna transmission, the affine power models from Chapter 3, (3.23) and (3.24) are employed. It is assumed that the BS is equipped with two antennas out of which one can be deactivated. Since the switch-off process is idealized to be instantaneous, a DTX power consumption of $P_S = 107$ W is assumed for both single and dual antenna transmission.

4.5.1 Problem Formulation

The global problem statement is formulated as follows. Given the channel state matrices $\mathbf{H}_{n,t,k}$ on each channel and the vector of target rates per user $\mathbf{r} = (R_1, \dots, R_K)$, sought are:

- the set of resources allocated to each user \mathcal{U}_k ,

- the power level $P_{a,e}$ per resource block $a = 1, \dots, |\mathcal{U}_k|$ and the spatial channel eigenvalue ϵ_e with $e = 1, \dots, |\mathcal{E}_{a,k}|$,

- and the number of active transmit antennas D ,

such that the supply power consumption is minimized while fulfilling the transmission power constraint, P_{\max} .

Combining (4.1), (4.3) and (2.3a), the sum capacity of user k over one transmission frame of duration τ_{frame} is given by

$$R_k = \frac{w\tau}{\tau_{\text{frame}}} \sum_{a=1}^{|\mathcal{U}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} \log_2 \left(1 + \frac{P_{a,e} \epsilon_{a,k}(e)}{P_N} \right), \quad (4.13)$$

with subcarrier bandwidth w in Hz.

The BS RF transmission power in time slot t is

$$P_{\text{Tx}} = \sum_{a=1}^{|\mathcal{A}_t|} \sum_{e=1}^{|\mathcal{E}_a|} P_{a,e}, \quad (4.14)$$

where \mathcal{A}_t is the set of N resources in time slot t and \mathcal{E}_a is the vector of channel eigenvalues on resource a .

Given (3.22), (4.13) and (4.14), the optimization problem that minimizes the supply power consumption is of the form

$$\underset{\mathcal{U}_k, \forall k, P_{a,e}, \forall (a,e), D}{\text{minimize}} \quad P_{\text{supply}}(\mathbf{r}) = \frac{1}{T} \left(\sum_{t=1}^{T_{\text{Active}}} (P_{0,D} + \Delta_P P_{\text{Tx}}) + \sum_{t=1}^{T_{\text{Sleep}}} P_S \right) \quad (4.15a)$$

$$\text{subject to} \quad P_{\text{Tx}} \leq P_{\text{max}}, \quad (4.15b)$$

$$T_{\text{Active}} + T_{\text{Sleep}} = T, \quad (4.15c)$$

$$R_k \leq \frac{w\tau}{\tau_{\text{frame}}} \sum_{a=1}^{|\mathcal{U}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} \log_2 \left(1 + \frac{P_{a,e} \mathcal{E}_{a,k}(e)}{P_N} \right) \quad (4.15d)$$

with the number of active transmission time slots T_{Active} , and DTX time slots T_{Sleep} . This is a set selection problem over the sets \mathcal{U}_k , and $\mathcal{E}_{a,k}$, as well as a minimization problem in $P_{a,e}$.

Complexity

Dynamic subcarrier allocation is known to be a complex problem for a single time slot in frequency-selective fading channels that can only be solved by suboptimal or computationally expensive algorithms [94, 96, 101]. In this study, two additional degrees of freedom are added by considering AA and DTX, increasing the complexity further. Consequently the problem is considered intractable and divided into two steps:

1. Real-valued estimates of the resource share per user, DTX duration, and number of active RF chains D , are derived based on simplified system assumptions.
2. The power-minimizing resource allocation over the integer set \mathcal{U}_k and power allocation are performed.

4.5.2 Step 1: AA, DTX and Resource Allocation

The first step to solving the global problem is based on a simplification of the system assumptions. This allows defining a convex subproblem with an optimal solution, which is later (sub-optimally) mapped to the solution of the global problem in step 2, described in Section 4.5.3.

Instead of time and frequency-selective fading it is assumed that channel gains on all resources are equal to the center resource block,

$$\mathbf{H}_k = \mathbf{H}_{n^c,t^c,k}, \quad (4.16)$$

where the superscript c signifies the center-most subcarrier and time slot. Step 1 thus assumes a block fading channel per user over $W = Nw$ and τ_{frame} .

The center resource block is selected due to the highest correlation with all other resources. Alternative methods to construct a representative channel state matrix are to take the mean or median of $\mathbf{H}_{n,t,k}$ over the OFDMA frame. However, application of the mean or median over a set of MIMO channels were found to result in a channel with lower capacity. See Section 4.5.4 for a comparison plot between different channel selection methods.

The link capacity is calculated using equal-power precoding and assuming uncorrelated antennas. In contrast to water-filling precoding, equal-power precoding provides a direct relationship between total transmit power and target rate. On block fading channels with real-valued resource sharing, OFDMA and TDMA are equivalent. Without loss of generality, resource allocation via TDMA is selected. These simplifications allow a convex optimization problem to be established that can be efficiently solved.

In a block fading multi-user downlink with equal power precoding, the average

rate for user k is given by

$$\overline{R}_k = \mu_k W \sum_{e=1}^{|\mathcal{E}_k|} \log_2 \left(1 + \frac{P_k}{D} \frac{\mathcal{E}_k(e)}{P_N} \right), \quad (4.17)$$

with the vector of channel eigenvalues per user \mathcal{E}_k , and transmission power on this link, P_k .

The transmission power is a function of the target rate, depending on the number of transmit and receive antennas. For the following configurations, (4.17) reduces to:

1x2 Single-Input Multiple-Output (SIMO),

$$P_k(\overline{R}_k) = \frac{P_N}{\epsilon_1} \left(2^{\left(\frac{\overline{R}_k}{W \mu_k} \right)} - 1 \right), \quad (4.18)$$

2x2 MIMO,

$$\overline{R}_k = \mu_k W \log_2 \left(1 + \frac{P_k}{2P_N} (\epsilon_1 + \epsilon_2) + \left(\frac{P_k}{2P_N} \right)^2 \epsilon_1 \epsilon_2 \right), \quad (4.19)$$

$$P_k(\overline{R}_k) = P_N \frac{-(\epsilon_1 + \epsilon_2) + \sqrt{(\epsilon_1 + \epsilon_2)^2 + 4\epsilon_1 \epsilon_2 (2^{\frac{\overline{R}_k}{W \mu_k}} - 1)}}{\epsilon_1 \epsilon_2}, \quad (4.20)$$

where $\mathcal{E}_k = (\epsilon_1, \epsilon_2)$. These equations can be extended in similar fashion to combinations with up to four transmit or receive antennas. Note that a higher number of antennas would require the general algebraic solution of polynomial equations with degree five or higher, which cannot be found in line with the Abel-Ruffini theorem [102].

Like on the BS side, the number of RF chains used for reception at the mobile could be adapted for power saving. However, the power-saving benefit of receive

AA is much smaller than in transmit AA, where a Power Amplifier (PA) is present in each RF chain. Moreover, multiple receive antennas boost the useful signal power and provide a diversity gain. Therefore, M_{Rx} is assumed to be fixed and set to $M_{\text{R}} = 2$ in the following.

Given the transmission power (4.18), (4.20), and the power model (3.21), the supply power consumption for TDMA is derived, similar to (4.12),

$$P_{\text{supply}}(\mathbf{r}) = \sum_{k=1}^K \mu_k (P_{0,D} + \Delta_{\text{P}} P_k(R_k)) + \nu P_{\text{S}}. \quad (4.21)$$

Consequently, the optimization problem is defined as:

$$\underset{(\mu_1, \dots, \mu_{K+1})}{\text{minimize}} \quad P_{\text{supply}}(\mathbf{r}) = \sum_{k=1}^K \mu_k (P_{0,D} + \Delta_{\text{P}} P_k(R_k)) + \mu_{K+1} P_{\text{S}} \quad (4.22a)$$

$$\text{subject to} \quad \sum_{k=1}^K \mu_k + \nu = 1 \quad (4.22b)$$

$$\nu \geq 0 \quad (4.22c)$$

$$\mu_k \geq 0 \quad \forall k \quad (4.22d)$$

$$0 \leq P_k(R_k) \leq P_{\text{max}} \quad \forall k. \quad (4.22e)$$

The first constraint ensures that all resources are accounted for and upper bounds μ_k . The second and third constraint guarantee positive durations. The fourth encompasses the transmit power budget of the BS and acts as a lower bound on μ_k . Note that due to the block fading assumption and TDMA, the transmission power per user, P_k , is equivalent to P_{Tx} in (3.22) at that point in time.

This problem is convex in its cost function and constraints (as proven in Appendix A.2). It can therefore be solved with available tools like the interior point method [80]. As part of the RAPS algorithm, (4.22) is solved once for each

possible number of transmit antennas. The number of transmit antennas is then selected according to which solution results in lowest supply power consumption.

The solution of this first step yields an estimate for the supply power consumption, the number of transmit antennas, D , the DTX time share, ν , and the resource share per user μ_k . If a solution for step 1 cannot be found, step 2 is not performed and outage occurs. Outage handling is left to a higher system layer mechanism which could, *e.g.* prioritize users and reiterate with reduced system load. If a solution to step 1 can be found, step 2 is performed as described in the following section.

4.5.3 Step 2: Subcarrier and Power Allocation

This section describes the second step of the RAPS algorithm. In step 2, the results of step 1 are mapped back to the global problem to find the resource allocation for each user, \mathcal{U}_k , the power level per resource and spatial channel, $P_{a,e}$, and the number of DTX slots, T_{Sleep} .

First, the real-valued resource share, μ_k , is mapped to the OFDMA resource count per user, $m_k \in \mathbb{N}$, such that

$$m_k = \lceil \mu_k NT \rceil \quad \forall k. \quad (4.23)$$

Possible rounding effects through the ceiling operation in (4.23) are compensated for by adjusting the number of DTX time slots, with

$$T_{\text{Sleep}} = \left\lfloor \frac{TN\mu_{K+1} - K}{N} \right\rfloor = \lfloor T\mu_{K+1} - K/N \rfloor. \quad (4.24)$$

The remaining time slots are available for transmission,

$$T_{\text{Active}} = T - T_{\text{Sleep}}. \quad (4.25)$$

The remaining unassigned resources,

$$m_{\text{rem}} = NT - \sum_{k=1}^K m_k - NT_{\text{Sleep}}, \quad (4.26)$$

are assigned to m_k in a round-robin fashion. After this allocation, it holds that $m_k = |\mathcal{U}_k|$.

Next, the number of assigned resource blocks per user, m_k , is equally subdivided into the number of resources per user and time slot, $m_{k,t}$, with

$$m_{k,t} = \left\lfloor \frac{m_k}{\sum_{l=1}^K m_l} N \right\rfloor. \quad (4.27)$$

The remaining unassigned resources per time slot,

$$m_{t,\text{rem}} = N - \sum_{k=1}^K m_{k,t}, \quad (4.28)$$

are allocated to different $m_{k,t}$ in a round-robin fashion.

Time slots considered for DTX are assigned statically, starting from the back of the frame. Note that the dynamic selection of DTX slots creates additional opportunities for capacity gains or power savings as they could be assigned to time slots with poor channel states, *e.g.* time slots experiencing deep fades. This opportunity is revisited in Chapter 5.

A corner case exists when $TN\mu_{K+1} < K$ and (4.24) becomes negative. This occurs when the DTX time share, ν , is very small and thus traffic load is high. Due to the high traffic load, it is possible that the target rates cannot be fulfilled

within the transmit power constraint, leading to outage for at least one user. Accordingly, if $TN\mu_{K+1} < K$, $T_{\text{Active}} = T$ and the resource mapping strategy in (4.23) is adapted such that

$$m_k = \lfloor \mu_k NT \rfloor, \quad (4.29)$$

which guarantees that $m_k < NT$. The remaining resources are allocated to users as outlined above.

A subcarrier allocation algorithm from Kivanc *et al.* [96] is adopted, which has been shown to work effectively with low complexity. The general idea of the algorithm is as follows: First assign each subcarrier to the user with the best channel. Then start trading subcarriers from users with too many subcarriers to users with too few subcarriers based on a nearest-neighbour evaluation of the channel state. The algorithm is outlined in Algorithm 1 and applied to each time slot t consecutively. The original algorithm from the literature is adapted as indicated in the caption.

At this stage, the MIMO configuration, the number of DTX time slots, and the subcarrier assignment are determined. Transmit powers are assigned in both spatial and time-frequency domains via an algorithm termed margin-adaptive power allocation [103]. The notion of this algorithm is as follows: First, channels are sorted by quality and a ‘water-level’ is initialized on the best channel, such that the rate target is fulfilled. Then, in each iteration of the algorithm, the next best channel is added to the set of used channels, thus reducing the water-level in each step. The search is finished once the water-level is lower than the next channel metric to be added. Refer to Appendix A.3 for the derivation.

Let the number of bits to be transmitted to each user be

$$B_{\text{target},k} = R_k \tau_{\text{frame}}. \quad (4.30)$$

Algorithm 1 Adapted RCG algorithm performed on each time slot t . As compared to [96] the cost parameter $h_{n,t,k}$ has been adapted and the absolute value has been added to the search of the nearest neighbour. $\mathcal{A}_{k,t}$ holds the set of subcarriers assigned to user k .

Ensure: $m_{k,t}$ is the target number of subcarriers allocated to each user k ,

$$h_{n,t,k} = |\overline{\mathbf{H}}_{n,t,k}|^2 \text{ and } \mathcal{A}_{k,t} \leftarrow \{\} \text{ for } k = 1, \dots, K.$$

- 1: **for all** subcarriers n **do**
 - 2: $k^* \leftarrow \arg \max_{1 \leq k \leq K} h_{n,t,k}$
 - 3: $\mathcal{A}_{k^*,t} \leftarrow \mathcal{A}_{k^*,t} \cup \{n\}$
 - 4: **end for**
 - 5: **for all** users k such that $|\mathcal{A}_{k,t}| > m_{k,t}$ **do**
 - 6: **while** $|\mathcal{A}_{k,t}| > m_{k,t}$ **do**
 - 7: $l^* \leftarrow \arg \min_{\{l: |\mathcal{A}_{l,t}| < m_{l,t}\}} \min_{1 \leq n \leq N} | -h_{n,t,k} + h_{n,t,l} |$
 - 8: $n^* \leftarrow \arg \min_{1 \leq n \leq N} | -h_{n,t,k} + h_{n,t,l^*} |$
 - 9: $\mathcal{A}_{k,t} \leftarrow \mathcal{A}_{k,t} / \{n^*\}, \mathcal{A}_{l^*,t} \leftarrow \mathcal{A}_{l^*,t} \cup \{n^*\}$
 - 10: **end while**
 - 11: **end for**
-

To fulfill $B_{\text{target},k}$, the following constraint must be met

$$B_{\text{target},k} - \sum_{a=1}^{|\mathcal{U}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} w\tau \log_2 \left(1 + \frac{P_{a,e} \mathcal{E}_{a,k}(e)}{P_N} \right) = 0, \quad (4.31)$$

which assesses the sum capacity according to (4.13).

The water-level v can be found via an iterative search over the set of channels that contribute a positive power, Ω_k , with

$$\log_2(v) = \frac{1}{|\Omega_k|} \left(\frac{B_{\text{target},k}}{w\tau} - \sum_{e=1}^{|\Omega_k|} \log_2 \left(\frac{\tau \mathcal{E}_{a,k}(e)w}{P_N \log(2)} \right) \right). \quad (4.32)$$

The water-level is largest on the first iteration and decreases on each iteration until it can no longer be decreased.

Using the Lagrangian method detailed in Appendix A.3, one arrives at a power-level per spatial channel of

$$P_{a,e} = \frac{vw\tau}{\log(2)} - \frac{P_N}{\mathcal{E}_{a,k}(e)}. \quad (4.33)$$

The margin-adaptive algorithm is illustrated for four channels in Fig. 4.7. The outcome of margin-adaptive power allocation as part of RAPS are the transmission power levels $P_{a,e}$ for each resource block. The supply power consumption after application of RAPS can be found by summation of transmission powers in each time slot and application of the affine power model. The entire RAPS algorithm is outlined in Fig. 4.8.

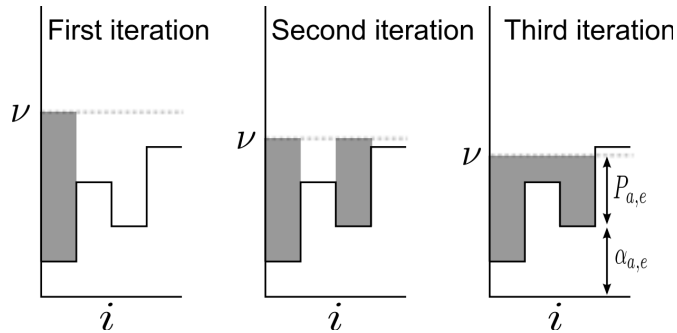


Figure 4.7: Illustration of margin-adaptive power allocation with three steps. The height of each patch i is given by $\alpha_{a,e} = P_N/\mathcal{E}_{a,k}(e)$. The water-level is denoted by ν . The height of the water above each patch is $P_{a,e}$. The first step sets the water-level such that the rate target is fulfilled on the best patch. The second and third step add a patch, thus reducing the water-level. After the third step, the water-level is below the fourth patch level, thus terminating the algorithm.

4.5.4 Results

In order to assess the performance of the RAPS algorithm Monte Carlo simulations are conducted. The simulations are configured as follows: mobiles are uniformly distributed around the BS on a circle with radius 250 m and a minimum distance of 40 m to the BS to avoid peak SNRs. Fading is computed according to the NLOS model described in 3GPP TR 25.814 [82] with 8 dB shadowing standard deviation, and the frequency-selective channel model B5 described in [104] with 3 m/s mobile velocity. All transmit and receive antennas are assumed to be mutually uncorrelated. Further system parameters are listed in Table 4.3.

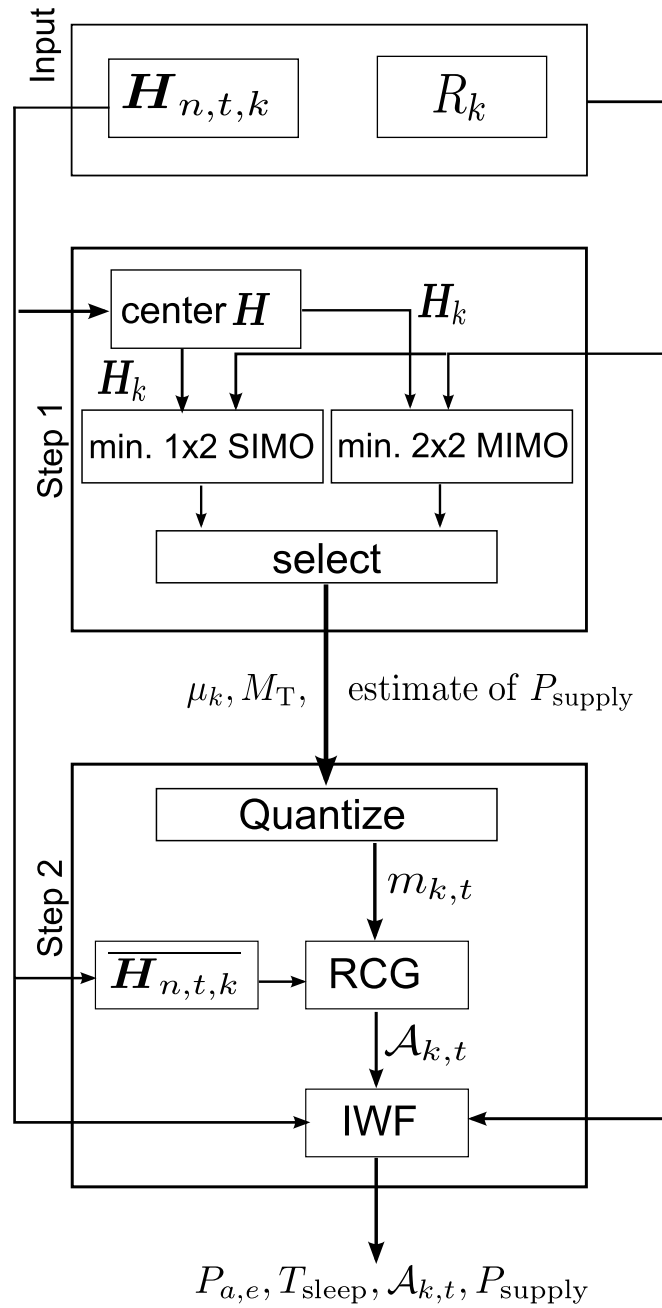


Figure 4.8: Outline of the RAPS algorithm.

Variable		Value
K	Number of users	10
N	Number of subcarriers	50
T	Number of time slots in an OFDMA frame	10
D	Number of transmit antennas	[1,2]
M_R	Number of receive antennas	2
P_{\max}	Maximum transmission power	46 dBm
τ_{frame}/τ	Duration of frame/time slot	10 ms/1 ms
W/w	System/subcarrier bandwidth	10 MHz/200 kHz
P_N	Thermal noise power per subcarrier	$4w \times 10^{-21}$ W

Table 4.3: System parameters

Benchmarks

The following transmission strategies are evaluated for comparison purposes:

- The theoretical limit of the BS power consumption is obtained by constant transmission at maximum power, P_1 .
- BA, which finds the minimum number of subcarriers that achieves the rate target. No sleep modes are utilized and all scheduled subcarriers transmit with a transmission power spectral density of P_{\max}/N . This benchmark represents the power consumption of SotA BSs which are neither capable of DTX nor AA.
- DTX, where the BS transmits with full power, P_1 , and switches to DTX once the rate requirements are fulfilled. This benchmark assesses the attainable savings when only DTX is applied.

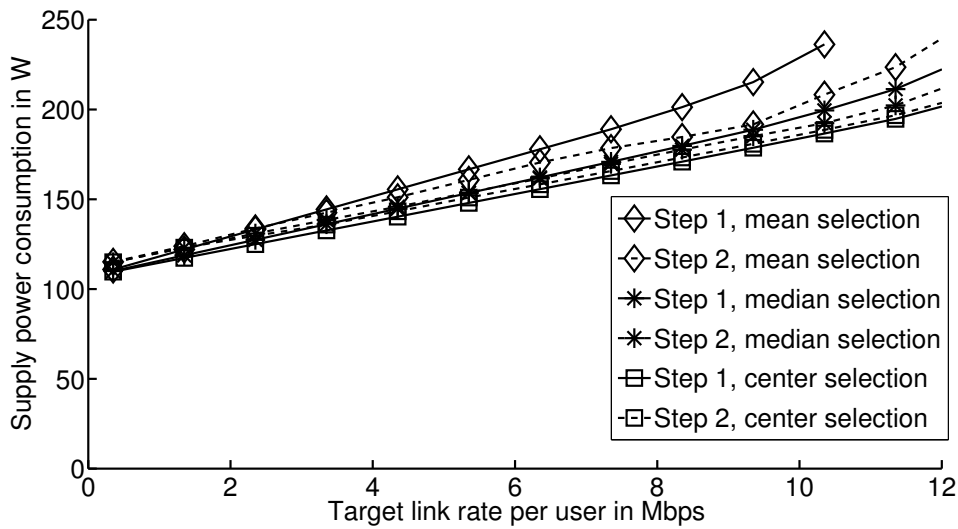


Figure 4.9: Performance comparison of different channel state selection alternatives.

Performance Analysis

The channel value selection in (4.16) serves as the channel gain for the block fading assumption of step 1. How the channel value selection of three possible alternatives (mean, median center) affects the supply power consumption is examined in Fig. 4.9. It can be seen that the selection of the mean channel gain results in the highest supply power consumption estimate after step 1. While the estimate can be improved in step 2, it is still inferior to the other alternatives. Use of the mean channel states causes step 1 to underestimate the channel quality. Consequently, too few time slots are selected for DTX in step 2. Choosing the median provides a better estimate than the mean and after step 2 the solution matches ‘step 2, center selection’. However, in center channel state selection the step 1 estimate and the step 2 solution have both the best match and the lowest supply power consumption. Therefore, as mentioned in Section 4.5.2, center channel state selection is chosen in the RAPS algorithm and all further analyses.

Fig. 4.10 compares the outage probabilities of step 1 with that of the BA

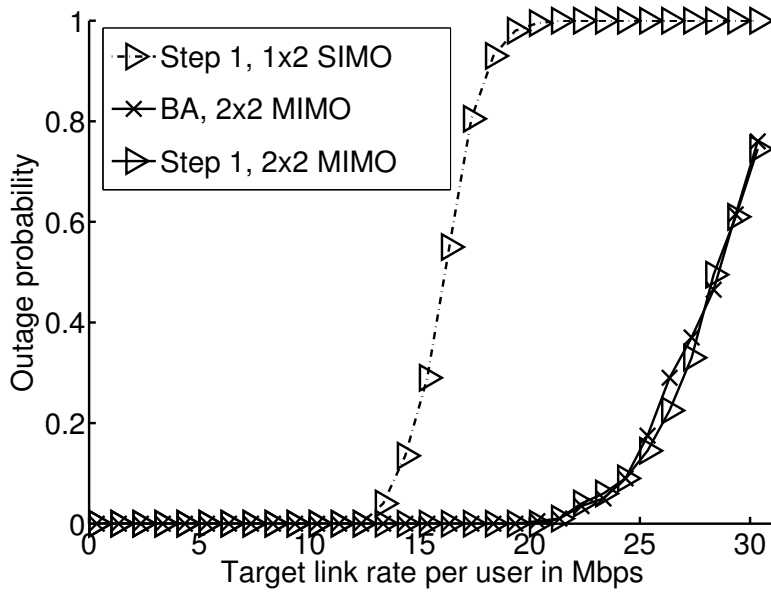


Figure 4.10: Outage probability in step 1 and the BA benchmark.

benchmark. Outage refers to the lack of a solution to step 1; when the user target data rates are too high for the given channel conditions, the convex subproblem has no solution, which causes the algorithm to fail. A reduction of user target data rates based on, *e.g.* priorities, latency or fairness, is specifically not covered by RAPS. One suggested alternative is to introduce admission control, in the way that some users are denied access, so that the remaining users achieve their target rates. Note that step 2 has no effect on the outage; if a solution exists after step 1, then step 2 can be completed. If a solution does not exist after step 1, then step 2 is not performed. Fig. 4.10 illustrates that RAPS selects two transmit antennas for target link rates above 14 Mbps with high probability, as target link rates cannot be achieved with a single transmit antenna. The BA benchmark and the step 1 MIMO solution have similar outage behavior.

Fig. 4.11 depicts the average number of DTX time slots over increasing target link rates. The effect of AA (*i.e.* switching between SIMO and MIMO transmission) can clearly be seen. For low target link rates, a large proportion of all time slots is

selected for DTX. For higher target link rates, the number of DTX time slots must be reduced when operating in SIMO mode. For target link rates above 12 Mbps, which approach the SIMO capacity (as described in the previous paragraph), the system switches to MIMO transmission. The added MIMO capacity allows the RAPS scheduler to allocate more time slots to DTX. In the medium load region (around 15 Mbps) the standard deviation is highest, indicating that here the RAPS algorithm strongly varies the number of DTX time slots depending on channel conditions and whether SIMO or MIMO transmission is selected. These variations contribute strongly to the additional savings provided by RAPS over the benchmarks. Another observation from Fig. 4.11 is that it is unlikely that more than two out of ten DTX time slots are scheduled at target link rates above 8 Mbps. This means that for a large range of target rates, no more than two DTX time slots are required to minimize power consumption. This is an important finding for applications of RAPS in established systems like LTE, where the number of DTX time slots may be limited due to constraints imposed by the standard.

A comparison of the supply power consumption estimates of step 1 and step 2 in Fig. 4.12 verifies that the estimate taken in step 1 as input for step 2 are sufficiently accurate. Although step 1 is greatly simplified with the assumption of block fading and its output parameters cannot be applied readily to an OFDMA system, it precisely estimates power consumption which is the optimization cost function. The slight difference between the step 1 estimate and power consumption after step 2 is caused by quantization loss and resource scheduling. Note that while step 1 supplies a good estimate of the power consumption, it is not a solution to the original OFDMA scheduling problem, since it does not consider the frequency-selectivity of the channel and does not yield the resource and power allocation.

The performance of RAPS in comparison to each benchmark is separately analyzed in Fig. 4.13. Here, a first observation is that even the supply power

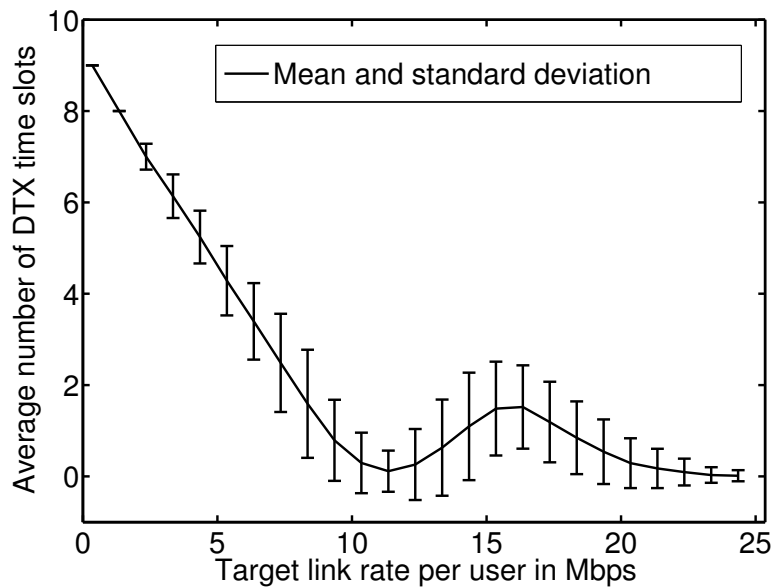


Figure 4.11: Average number of DTX time slots over increasing target link rates. Error bars portray the standard deviation. Total number of time slots $T = 10$.

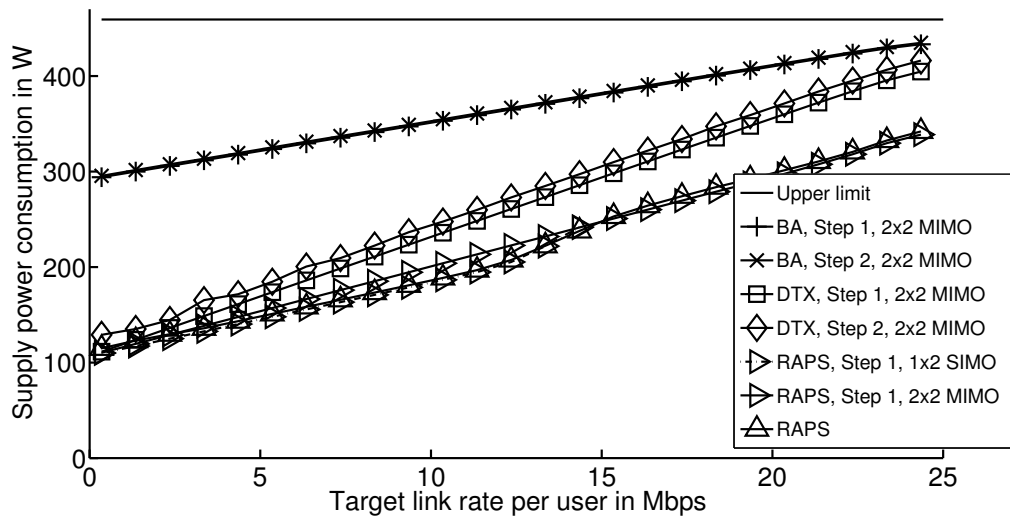


Figure 4.12: Supply power consumption for different RRM schemes on block fading and frequency-selective fading channels (comparison of step 1 and step 2) for ten users. Overlaps indicate the match between the step 1 estimate and the step 2 solution.

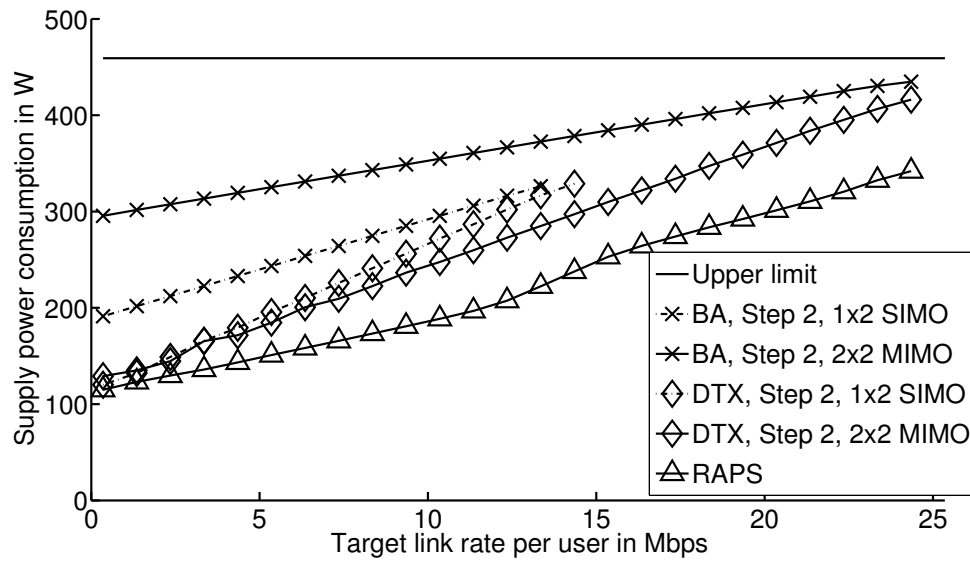


Figure 4.13: Supply power consumption on frequency-selective fading channels for different RRM schemes and RAPS for ten users. For a bandwidth-adapting BS, power consumption is always reduced by AA, while AA is never beneficial for a sleep mode capable BS. For RAPS energy consumption is reduced by AA at low rates. In general, RAPS achieves substantial power savings at all BS loads.

consumption of BA is significantly lower than the theoretical maximum for most target link rates. BA with a single transmit antenna always consumes less power than with two antennas, as long as the rate targets with one antenna can be met. AA is thus a valid power-saving mechanism for BA.

Second, the DTX power consumption is significantly lower than for BA, especially at low target rates, because lower target rates allow the BS to enter DTX for longer periods of time. As the BS load increases, the opportunities of the DTX benchmark to enter sleep mode are reduced, so that DTX power consumption approaches that of BA. Unlike in BA, it is never beneficial for a BS capable of DTX to switch operation to a single antenna, because transmitting for a longer time with a single antenna always consumes more power than a short two-antenna

transmission, which allows for a longer DTX duration. (Note that this finding may not apply to other power models.)

Third, RAPS reduces power consumption further than DTX by employing AA at low rates and PC at high rates. Through PC, the slope of the supply power is kept low between 15 and 20 Mbps. At higher rates an upward trend becomes apparent, since link rates only grow logarithmically with the transmission power. In theory, when the BS is at full load, the supply power of all energy saving mechanism will approach maximum power consumption. However, when the BS load is very high, not all users may achieve their target rate and outage occurs (see Fig. 4.10). In other words, operating the BS with load margins controls outage and allows for large transmission power savings. Power savings of RAPS compared to the SotA BA range from 102.7 W (24.5%) to 136.9 W (41.4%) depending on the target link rate per user.

In addition to absolute consumption, the energy efficiency of RAPS is inspected in Fig. 4.14. Energy efficiency is defined in bit/Joule as

$$E = P_{\text{supply}}^{-1} \sum_{k=1}^K R_k. \quad (4.34)$$

Observe that all data series are monotonically increasing. Therefore, RAPS does not change the paradigm that a BS is most efficiently operated at peak rates. However, RAPS will always offer the most efficient operation at any requested rate.

4.6 Summary

Starting with the Shannon limit, this chapter derived supply power optimal resource allocation in an LTE cell.

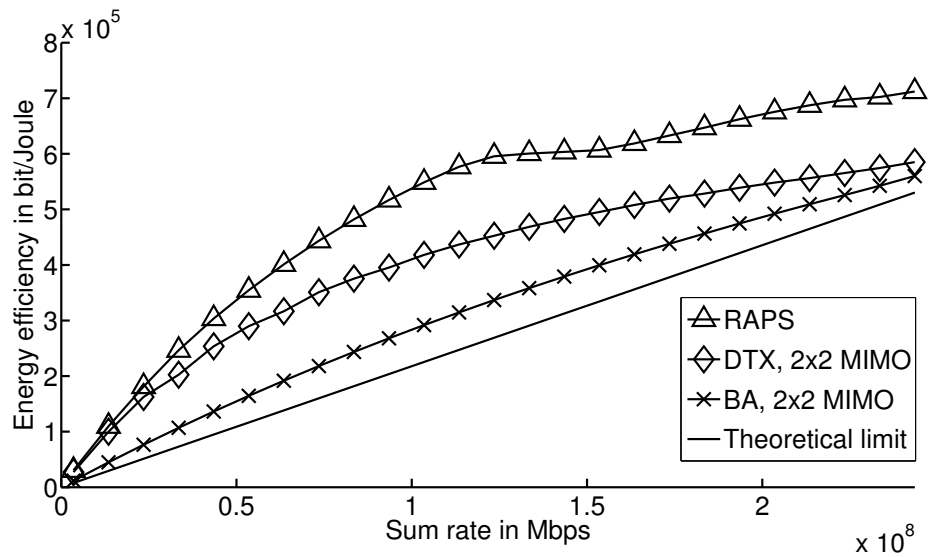


Figure 4.14: Energy efficiency as a function of sum rate. Energy efficiency can be increased by the RAPS algorithm. The BS remains most efficient at peak rate.

First, a literature review described the SotA of energy efficient RRM for cellular networks.

Next, the general interrelationships between bandwidth, power and time in transmissions were studied. As the PA consumption can be reduced by PC, a multi-user downlink PC allocation mechanism was proposed. Consideration of single links was found to be insufficient for power-saving, as a BS schedules multiple users on the same resources. It was shown that the power model, due to its affine mapping, does not affect the PC solution. Thus, it is also the solution to supply power minimization, when sleep modes are not considered.

As the power consumption reduction through PC is lower bounded by the idle power consumption of a transmitter, short sleep modes, *i.e.* DTX, have also been taken into consideration. It was shown that the joint application of PC and DTX brings additional savings over the application of each individual scheme. Using convex optimization, a resource allocation mechanism called PRAIS was proposed. Using PRAIS, fundamental limits of power-saving resource allocation

were identified on several power models. Power savings estimated on the basis of the affine model presented in Chapter 3 are in the range of 50%.

When adding the integer value constraint, which is present in practical systems where resources are only available in chunks and each has an individual channel state, the resource allocation problem becomes a mixed integer program. As this problem is highly complex, it was split into two steps. A convex subproblem was derived and solved in similar fashion as in the PRAIS algorithm. The solution to the subproblem yields an approximation of quantization, providing a configuration for the mixed integer problem. The overall algorithm is labelled RAPS. Application of the RAPS algorithm on this stricter constraint set achieved savings in the order of 25% to 40%.

The RAPS algorithm could be readily applied in current LTE systems. It adds complexity to scheduling, but offers significant power savings over all loads. Limitations lie in the fact that switching is not instantaneous in practice and PAs have limits on the PC range. If switching times are slower than the fraction of a time slot, power savings will not be as high as predicted. Due to linearity requirements, some PAs have a lower limit on transmission powers. Thus, PC may not be as effective as predicted in this chapter. Overall, these limitations only affect the achievable savings, not the functionality of the power-saving RRM algorithms presented in the chapter.

Chapter 5

Power Saving on the Network Level (Multi-cell)

5.1 Overview

Chapter 4 proposed how the power consumption of a Base Station (BS) could be minimized regardless of interference. This chapter expands the power-saving perspective to the network level, thus taking inter-cell interference into account. In particular, it studies how the Discontinuous Transmission (DTX) time slots scheduled by each BS in a network can be aligned constructively to avoid mutual interference. See Fig. 5.1 for an illustration of multiple cells, each choosing to transmit or schedule DTX in the available time slots.

In this chapter, the findings on channel allocation from literature are combined and adapted for power-saving. Four DTX alignment strategies for the cellular downlink are derived, one of which is an original contribution in which each BS

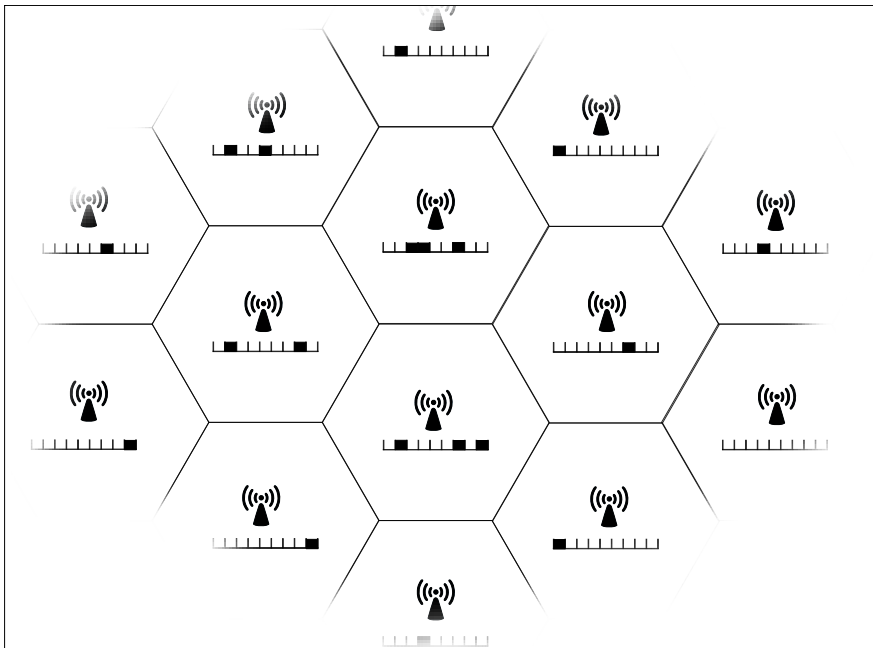


Figure 5.1: Each BS in a network select time slots for transmission (black) or DTX (blank).

independently and without coordination prioritizes time slots for transmission while maintaining system stability.

The remainder of the chapter is structured as follows. Section 5.3 formulates the system model and the problem at hand. In Section 5.4, three State-Of-The-Art (SotA) channel alignment solutions are described and the novel alignment method is introduced. Findings obtained from simulation are presented in Section 5.5. The chapter is summarized in Section 5.6.

The contents of this chapter have been previously accepted for publication in [i] and submitted for publication in [k].

5.2 Channel Allocation in Literature

Generally, in channel allocation, the challenge is to assign frequency channels to different cells or links of a network such that mutual interference is avoided. Several centralized approaches to interference avoidance have been proposed [105, 106, 107, 108]. However, to be applicable for future small-cell cellular networks, such channel allocation needs to be distributed, uncoordinated and dynamic. Distributed operation prevents the delay and backhaul requirements imposed by a central controller. If the allocation is also uncoordinated, it does not require message exchange through a backhaul of limited capacity. Furthermore, as networks, channels and traffic change rapidly, the allocation must be highly dynamic and updated often. Problems of this type are known to be NP-hard [109]. Therefore, different suboptimal approaches have been proposed. The simplest approach to distributing channels over a network is a fixed assignment in which it is predefined which links will use which channels, *e.g.* see [110]. However, this is inflexible to changing or asymmetric traffic loads. For flexibility, dynamic channel allocation methods are required. A simple dynamic channel allocation method, Sequential Channel Search [111], assigns channels with sufficient quality in a predefined order. This technique allows adjusting the number of channels flexibly, but is suboptimal due to its strong channel overlap between neighbors. Taking the channel quality into consideration by measurement is proposed in the Minimum Signal-to-Interference-and-Noise-Ratio (SINR) method [111]. This technique offers increased spectral efficiency, but can lead to instabilities when allocations happen synchronously. Ellenbeck *et al.* [112] introduce methods of game theory and respond to the instabilities by adding p -persistence to avoid simultaneous bad player decisions. However, the proposed method only applies to single user systems and does not address target rates. Dynamic Channel Segregation [113] introduces the notion of memory of channel availability, allowing a transmitter to track which channels tend to be favorable for transmission.

However, the algorithm only applies to sequential channel decisions based on an idle or busy state in circuit switched networks and cannot be applied to the concurrent alignment of channels as is required in Orthogonal Frequency Division Multiple Access (OFDMA) networks.

5.3 System Model and Problem Formulation

The network is considered as follows. BSs in a reuse-one OFDMA cellular network schedule a target rate, B_k , per OFDMA frame to be transmitted to each mobile $k = \{1, \dots, K\}$. All time slots are available for scheduling in all cells. Mobiles report perceived SINR from the previous OFDMA frame, $s_{n,t,k}$, on subcarrier $n = \{1, \dots, N\}$, time slot $t = \{1, \dots, T\}$ to their associated BS. BSs have a DTX mode available for each time slot during which transmission and reception are disabled and power consumption is significantly reduced to P_S compared to power consumption during transmission, which is a function of the power allocated to each resource block, ρ_{TX} . DTX is available fast enough to enable it within individual time slots of an OFDMA transmission such that transmission time slots are not required to be consecutive. A BS can schedule no transmission in one or more time slots and go to DTX mode instead. Scheduling a DTX time slot in one BS reduces the interference on that time slot to all other BSs. The OFDMA frames of all BSs are assumed to be aligned such that the interference over one time slot and subcarrier is flat and that all BSs can perform alignment operations in synchrony. The OFDMA frame consists of NT resource blocks. The channel is subject to block fading as described in Section 2.5.2. Interference is treated as noise.

To describe the resource allocation problem formally, the function

$$\begin{aligned} \Pi : \{1, \dots, T\} \times \{1, \dots, N\} &\rightarrow \{0, 1, \dots, K\} \\ (n, t) &\mapsto k \end{aligned} \quad (5.1)$$

is defined, which maps to each resource block a user k or 0, where 0 indicates that the resource block is not scheduled.

With $r_{n,t,k}$ the capacity of resource (n, t) if it is scheduled to k , the optimization problem for the total BS power consumption, P_{total} , is as follows.

$$\underset{\Pi}{\text{minimize}} \quad P_{\text{total}} = P_{\text{S}} \frac{T_{\text{S}}}{T} + \rho_{\text{Tx}} N_{\text{Tx}} \left(\frac{T - T_{\text{S}}}{T} \right) \quad (5.2a)$$

subject to

$$N_{\text{Tx}} = \quad (5.2b)$$

$$|\{n \in \{1, \dots, N\} \mid (\Pi(n, t) \neq 0 \quad \forall t \in \{1, \dots, T\})\}|$$

$$T_{\text{S}} = \quad (5.2c)$$

$$|\{t \in \{1, \dots, T\} \mid (\Pi(n, t) = 0 \quad \forall n \in \{1, \dots, N\})\}|$$

$$B_k \leq \sum_{\{(n,t) \mid \Pi(n,t)=k\}} r_{n,t,k} \quad \forall k. \quad (5.2d)$$

With $|\cdot|$ the cardinality operator, T_{S} is the number of time slots in which for all n , no user is allocated and the BS can enable DTX and N_{Tx} is the number of resource blocks that are scheduled for transmission. The constraint (5.2d) provides the rate guarantee for each user.

The difficulty lies in finding the mapping Π . Under a brute force approach, there exist $(K + 1)^{NT}$ possible combinations. Due to the large number resource blocks present in typical OFDMA systems like Long Term Evolution (LTE), the computation of the solution is infeasible. Consequently, the next section resorts to heuristic methods.

Some of the methods compared in the following make use of a time slot ranking. In other works, *e.g.* [113], ranking is proposed to be based on SINR. However, in a multi-user OFDMA system each subcarrier and mobile terminal have a different SINR, thus generating a problem of comparability between time slots. Consequently, it is proposed to compare time slots by their hypothetical sum capacity, B_t , over all mobiles and subcarriers with

$$R_t = \sum_{k=1}^K \sum_{n=1}^N \log_2(1 + s_{n,t,k}). \quad (5.3)$$

5.4 DTX Alignment Strategies

In this section, four strategies to tackle the problem at hand are identified which differ in performance and complexity.

5.4.1 Sequential Alignment

In *sequential alignment*, the strategy is to always allocate as many time slots for transmission as required in sequential order as illustrated in Fig. 5.2 and set the remainder to DTX. This leads to strongest overlap and consequently highest interference on the first time slot and lowest overlap and possibly no transmissions at all on the last time slot. This strategy does not make use of the available channel quality information per subcarrier and provides valuable insights into questions of stability, reliability and convergence.

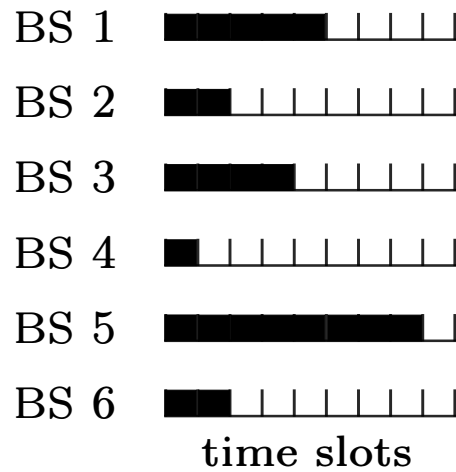


Figure 5.2: Illustration of sequential alignment of time slots for transmission (black) or DTX (blank) in six BSs.

5.4.2 Random Alignment

Random alignment refers to a random selection of transmission time slots after every OFDMA frame and setting the remainder to DTX. This results in chaotic interference patterns. This strategy provides a reference for achievable gains, allows to assess the worst case effects of randomness and represents the SotA in today's unsynchronized, unaligned networks.

5.4.3 P-persistent Ranking

The synchronous alignment of uncoordinated BSs can lead to instabilities, when neighboring BSs—perceiving similar interference information—schedule the same time slots for transmission. This leads to oscillating scheduling which never reaches the desired system state [112]. To address this problem, p-persistence is introduced to break the unwanted synchrony by only changing established DTX schedules with a probability $p = 0.3$. In initial tests, the value of p did not have a

significant effect on the power consumption; exact tuning is presented in [112]. P-persistent ranking first ranks time slots by their sum capacity as in (5.3), schedules time slots in that order, but only applies this new selection with probability p . Otherwise, the schedule from the last iteration remains active. After ranking, time slots are selected for transmission in order of B_t . The remainder of time slots is set to DTX.

5.4.4 Distributed DTX Alignment with Memory

To counter oscillation and achieve a convergent network state memory of past schedules is introduced into the alignment process. The notion of the *distributed DTX alignment with memory* algorithm is as follows. Taking into account current time slot capacities, R_t , past scores and slot allocations, each BS first updates the internal score of each time slot and then returns the priority of time slots by score. In case of equal scores, time slots are further sub-sorted by R_t . The score is updated in integers. All time slots which were used for transmission in the previous OFDMA frame receive a score increment of one. Furthermore, the time slot with highest R_t , Q_0 , receives an increment of one. Time slots which were not used for transmission in the previous OFDMA frame receive a decrement of one, except for Q_0 . Scores have an upper limit, ψ_{ul} , beyond which there is no increment and a lower limit, ψ_{ll} , below which there is no decrement. The difference between ψ_{ul} and ψ_{ll} can be interpreted as the depth of the memory buffer.

The algorithm is shown in Algorithm 2 with scoring map ψ , ranking tuple Q , and priority tuple V . The set of used time slots, Υ_u , and unused time slots, Υ_{uu} , make up the set of time slots, Υ , the cardinality of which is T . The function ‘sort-desc-by-capacity(Υ)’ returns a list of time slots ordered descending by R_t . The function ‘sort-desc-by-score(Q, ψ, Υ)’ returns a list of time slots ordered primarily

Algorithm 2 Distributed DTX alignment with memory

Ensure: ψ, Υ_u

- 1: $Q \leftarrow \text{sort-desc-by-capacity}(\Upsilon)$
 - 2: **for all** v in Υ_u **do**
 - 3: **if** $\psi(v) < \psi_{ul}$ **then**
 - 4: $\psi(v) \leftarrow \psi(v) + 1$
 - 5: **end if**
 - 6: **end for**
 - 7: **for all** v in $\Upsilon_{uu} \setminus \{Q_0\}$ **do**
 - 8: **if** $\psi(v) > \psi_{ll}$ **then**
 - 9: $\psi(v) \leftarrow \psi(v) - 1$
 - 10: **end if**
 - 11: **end for**
 - 12: $\psi(Q_0) \leftarrow \psi(Q_0) + 1$
 - 13: $V \leftarrow \text{sort-desc-by-score}(Q, \psi, \Upsilon)$
 - 14: **return** ψ, V, Υ_u
-

by descending score and secondarily by descending R_t . After the execution of Algorithm 2, the first T_{Tx} time slots in V are scheduled for transmission.

In the following, the iterations over three OFDMA frames of Algorithm 2 for a system with three time slots are illustrated. In the example, the algorithm delays the change of Υ_u from c to b to buffer scheduling changes. The iteration begins with arbitrarily chosen $\Upsilon = \{a, b, c\}$, $\psi = \{a : 0, b : 2, c : 5\}$:

1. $\Upsilon_u = \{c\}$, $Q = (b, c, a)$
 $\rightarrow \psi = \{a : 0, b : 3, c : 5\}$, $V = (c, b, a)$
2. $\Upsilon_u = \{b, c\}$, $Q = (b, c, a)$
 $\rightarrow \psi = \{a : 0, b : 5, c : 5\}$, $V = (b, c, a)$
3. $\Upsilon_u = \{b\}$, $Q = (b, a, c)$
 $\rightarrow \psi = \{a : 0, b : 5, c : 4\}$, $V = (b, c, a)$

When applied, Algorithm 2 strongly benefits time slots which were used for transmission in the past (b, c in step 1 of the example). These time slots tend to repeatedly receive score increments until they all have maximum score ψ_{ul} (b, c in step 2 of the example). When the score is equal for some time slots (b, c in step 2 of the example), the ranking is based on R_t . A time slot which was used for transmission and has highest R_t , Q_0 , receives the highest increment (slot b in step 2 of the example). When a time slot has highest R_t , but was not selected for transmission in the previous OFDMA frame, it receives a score increment, but is not guaranteed to be used for transmission (slot b in step 1 of the example). When a time slot repeatedly has highest R_t , it reaches ψ_{ul} (b in the example). The algorithm thus buffers short term changes in the channel quality setting in favor of long term time slot selection.

5.5 Results

This section analyzes the four alignment strategies in simulation with regard to power consumption, convergence, reliability of delivered rates and algorithmic complexity after the introduction of the simulation environment and the resource block scheduling scheme.

Simulation Environment and Resource Block Scheduling

The four strategies were tested in a network simulation with 19-cell hexagonal arrangement with uniformly distributed mobiles and fixed target rates per mobile. Data was collected only from the center cell which is thus surrounded by two tiers of interfering cells. Power consumption of a cell is modelled by the affine model in (3.23). Table 5.1 lists additional parameters used which approximate an LTE system. The simulation is started with the assumption of full transmission power on all resources with a power consumption of 355 W per cell as a worst-case initial configuration.

In order to assess the performance of the four strategies, it is necessary to make assumptions about how the individual resource blocks are scheduled within a time slot. Resource block scheduling is itself challenging and can have a strong effect on power consumption. In order to avoid masking effects of the time slot scheduling algorithms under test, only a very simple sequential resource block allocation is applied rather than sophisticated solutions from the literature. Sequential resource allocation is performed after the DTX alignment step is completed and allocates as many bits to an OFDMA resource as possible according to the Shannon capacity, followed in order by the next resource in the same time slot (sequentially), until the target rate has been scheduled to each mobile. Time slots are scheduled in the order provided by each of the four strategies. This sequential

Parameter	Value
Carrier frequency	2 GHz
Intersite distance	500 m
Pathloss model	3GPP UMa, NLOS, shadowing [82]
Shadowing standard deviation	8 dB
Bandwidth	10 MHz
Transmission power per resource block	0.8 W
Thermal noise temperature	290 K
Interference tiers	2 (19 cells)
Mobile target rate	1 Mbps to 3 Mbps
OFDMA subframes (time slots)	10
Subcarriers	50
Mobile terminals	10
ψ_{ul}	5
ψ_{ll}	0
Power model parameters, (3.23) (idle; load factor; DTX)	186 W; 4.2; 107 W
Subcarrier bandwidth	200 kHz

Table 5.1: Simulation parameters used.

resource scheduler deliberately omits the benefits of multi-user diversity. This leads to underestimating achievable rates in simulation compared to a system which exploits multi-user diversity, in favor of allowing a fair comparison of the quality of different DTX alignment algorithms.

Power Consumption

To assess achievable power savings and the dynamic adaptivity over a large range of cell loads, the cell total power consumption is shown in Fig. 5.3. At low load very few resource blocks are required to deliver the target rate and more time slots can be scheduled for DTX than at high traffic loads, leading to monotonously rising power consumption over increasing target rates for all alignment methods.

Sequential alignment causes the highest power consumption over any target rate with an almost linear relationship between user target rates and power consumption. Sequential alignment consumes high power, as it schedules many resource blocks to achieve the target rate due to the high interference level present.

This power consumption is significantly lower for the SotA, random alignment. The randomness of time slot alignment creates a much lower average interference than with sequential alignment allowing more data to be transmitted in each resource block. As fewer resource blocks are required, less power is consumed.

P-persistent ranking and DTX with memory both achieve similarly low power consumption of up to 40% less than random alignment. The relationship between power consumption and target rate is noticeably non-linear, as it is flat at low target rates and grows more steeply at high target rates. This behavior is caused by the low interference level these strategies manage to create. Only at high rates, when the number of sleep time slots has to be significantly decreased, does the interference increase, leading to higher power consumption.

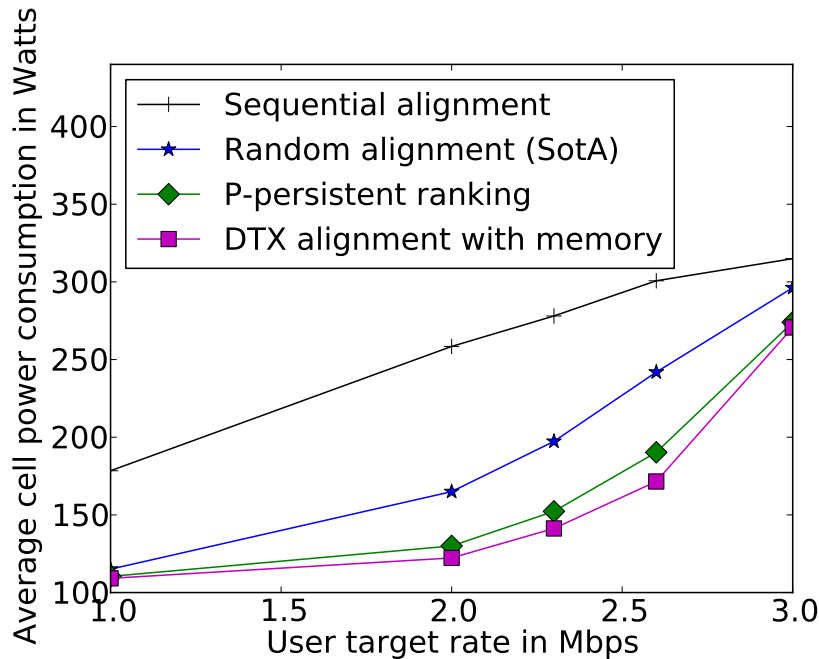


Figure 5.3: BS power consumption over different cell sum rates.

Also noteworthy is the fact that at 1 Mbps and 3 Mbps, random alignment performs nearly as good as p-persistent ranking. At these extreme points the network is almost unloaded and almost fully loaded, respectively. Consequently, either most time slots are scheduled for DTX or none, leaving very little room for optimization compared to randomness. The largest potential for time slot alignment for power saving is in networks which are medium loaded. Under medium load, the number of transmission and DTX time slots is similar, causing the effects of alignment to be most pronounced.

Convergence

Another relevant aspect is the convergence of the network to a stable state. As each BS makes iterative adjustments to its selection of time slots for transmission, the speed of convergence as well as the convergence to a stable point of operation are relevant criteria. The effect of the iterative execution of the four strategies on

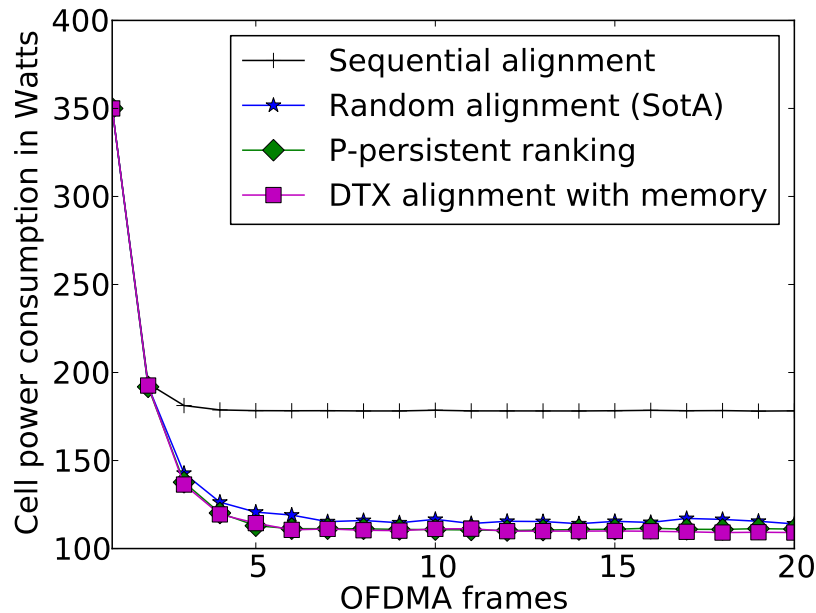


Figure 5.4: BS power consumption over OFDMA frames at 1 Mbps per mobile.

the cell power consumption is illustrated in Fig. 5.4. Power consumption is found to converge to a stable value within within six OFDMA frames (alignment iterations). All strategies converge to the average power consumption values shown in Fig. 5.3. The simulation starts from a worst-case schedule of transmission on all resource blocks and then iteratively schedules time slots for transmission and DTX. In the case of 1 Mbps per user, one transmission time slot is sufficient to schedule the target rate. With transmissions only taking place during one time slot, there is very little difference between a random alignment and p-persistent ranking or DTX with memory.

At 2 Mbps per user, see Fig. 5.5, random alignment occasionally causes higher interference than p-persistent ranking or DTX with memory, leading to higher power consumption. Also, p-persistent ranking converges more slowly than DTX alignment with memory.

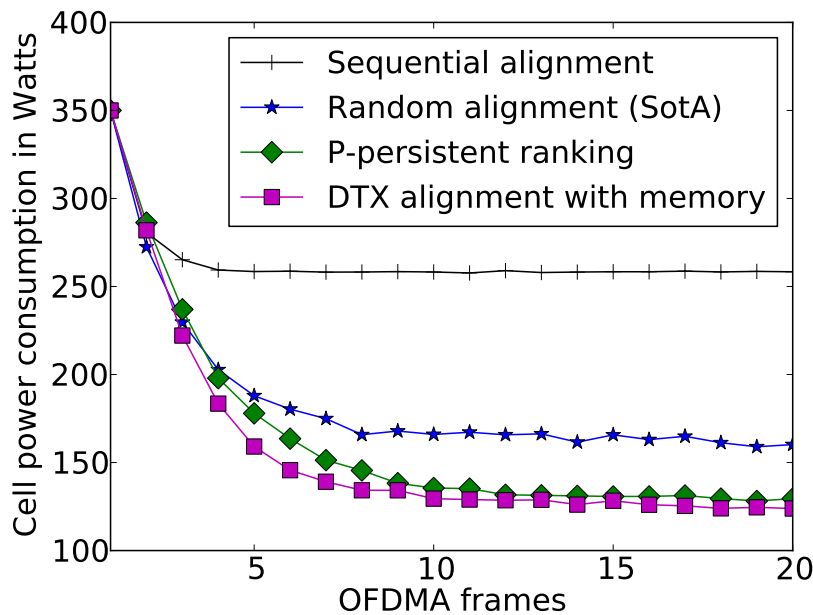


Figure 5.5: BS power consumption over OFDMA frames at 2 Mbps per mobile.

Reliability

An important aspect in dynamical systems with target rates is that scheduled target rates cannot always be fulfilled. As resource block scheduling is based on channel quality information which was collected in the previous OFDMA frame, the actual channel quality during transmission may differ, leading to lower than expected rates. Thus, although certain target rates are scheduled and although the system is not fully loaded, a BS may fail to deliver the targeted rate and require retransmission of some resource blocks. This metric is assessed by considering the retransmission probability for each strategy. The results are shown in Fig. 5.6 against a range of user target rates.

Easiest to interpret is sequential alignment which does not require retransmission for target rates up to 2.3 Mbps due to its determinism. The increase in the retransmission probability at high rates is not caused by a failure of the alignment, but by system overload. When high rates are combined with bad channel

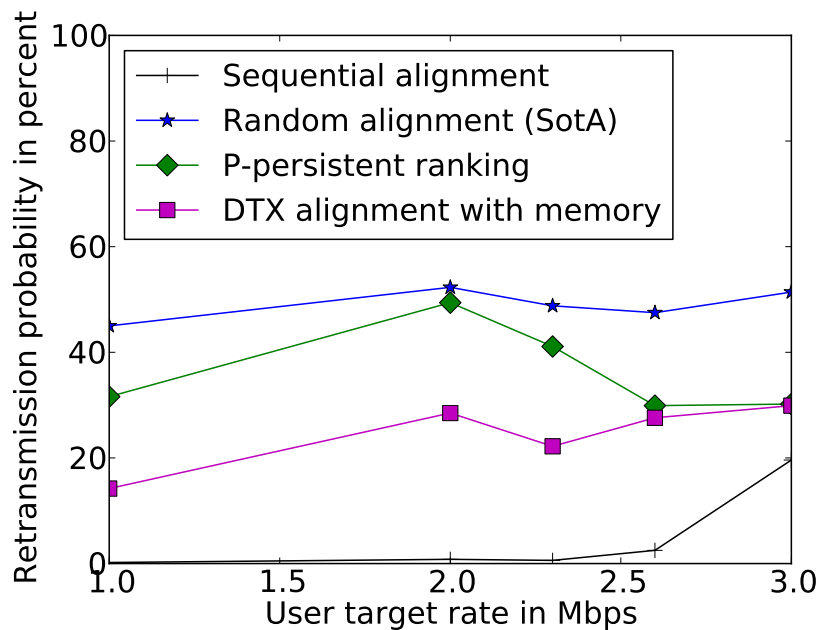


Figure 5.6: Retransmission probability over targeted rate.

conditions, the system may be unable to deliver the target data rate, independent of the alignment strategy. This increase of the retransmission probability at high rates is present for all alignment strategies and constitutes outage.

The retransmission probability is highest for random alignment. This is caused by the strong fluctuation of interference under randomized scheduling. Channel quality measurements used for resource block scheduling are of very little reliability, as the interference changes quickly due to random time slot alignment.

P-persistent ranking performs slightly better than random alignment at 1 Mbps target rate. At 2 Mbps per user, where the alignment potential is highest, oscillation causes the highest retransmission probability.

DTX alignment with memory achieves a much lower retransmission probability in the range of 15% to 20%, due to the reduction in interference fluctuation introduced by the memory score.

Complexity

With regard to complexity, sequential and random alignment are of minimal complexity. These strategies involve no algorithmic decision-making on the set of transmission time slots. P-persistent ranking requires one ranking and time slot selection with probability p per iteration. The highest complexity is present in DTX with memory, which requires two executions of the time slot sort, one for the generation of the score and one for the output of the ranking. Although DTX with memory comprises the highest complexity in comparison, these operations pose a small burden on modern hardware as the number of time slots is typically small. For example, typical LTE systems are designed with 10 subframes (time slots).

Interpretation

DTX with memory was found to provide the best results. Under the present assumptions, it provides both a lower power consumption and lower retransmission probability than the SotA and p-persistent ranking. Future DTX capable networks can and should exploit this alignment potential.

5.6 Summary

In this chapter, the constructive uncoordinated and distributed alignment of DTX time slots between interfering OFDMA BSs for power-saving under rate constraints was studied. The open problem was first formally established. Due to its complexity, it was approached with four alternative heuristic strategies: Sequential alignment, random alignment, p-persistent ranking and distributed DTX with memory. Distributed DTX with memory is an original contribution

of this thesis. It addresses shortcomings of the other strategies by introducing memory to overcome short-term fluctuations and networks oscillations. The performance of the four strategies was tested in simulation. It was found that power consumption can be reduced significantly, especially at medium cell loads. All strategies converge within six OFDMA frames. DTX with memory reduces power consumption up to 40% compared to the SotA, combined with around 20% reduction in retransmission probability.

Chapter 6

Conclusions, Limitations and Future Work

6.1 Summary and Conclusions

In this thesis, the power consumption of Long Term Evolution (LTE) Base Station (BS) devices was studied and modelled. The findings from the model were applied to reduce power consumption through energy efficient Radio Resource Management (RRM) mechanisms while upholding user rate constraints. RRM mechanisms were applied to control links on the cell level and interference on the multi-cell level.

In Chapter 1, the importance of energy efficiency for current mobile networks was emphasized. As mobile traffic, the number of devices and the size of the infrastructure are constantly increasing, the CO₂ emission of mobile networks is expected to double between 2007 and 2020. Green Radio research efforts, such as this thesis, are directed to revert that trend.

Chapter 2 first inspected the power consumption life cycle of a mobile network. While the production of mobile equipment causes significant emissions of CO₂, the operation is power efficient. The situation was found to be almost the opposite for cellular BSs which cause the largest share of emissions during operation. This is caused by the significant number of stations in operation combined with their long life times. This situation motivated the reduction of operation power consumption of BSs as the topic of this thesis. Next, the State-Of-The-Art (SotA) of Green Radio was identified through a detailed literature survey. Finally, the chapter also discussed fundamental technical concepts in preparation for the original contributions in Chapters 3, 4, and 5.

Chapter 3 explored the sources of power consumption in an LTE BS. First, the difficulty of accurate power modeling and other existing power models were described. Next, the typical architecture of a BS was introduced. Each of its components was discussed and individually modelled. Components were found to be either load adaptive, like the Power Amplifier (PA) and Baseband (BB) unit, have constant power consumption, like the Radio Frequency (RF) transceiver, or have a power consumption which scales with other components' consumption, like power conversion and cooling. The power consumptions of the subcomponents were combined to find the power consumption of the entire BS. Since the component model is too elaborate for many applications, a parameterized model was proposed which abstracts many parameters. The parameterized model only considers those parameters which are typically modified during operation such as bandwidth, transmission power, the number of transmission antennas, and sleep mode. Parameters were provided. The parameterized model was once more simplified into an affine model which can be applied in optimization. Overall, it was found that an increase of the energy efficiency in BSs must be based on reducing the PA power consumption through Power Control (PC) and setting the entire BS to a sleep mode. Hardware research should focus on enabling and

amplifying these two mechanisms. Other manipulation such as the adaptation of coding, modulation or bandwidth are less promising. Simple models are found to capture the most relevant aspects sufficiently.

With the power model and saving techniques identified in the prior chapter, Chapter 4 explored three RRM mechanisms to reduce BS power consumption. The first technique is PC, through which power was adapted to match only the requested capacity per User Equipment (UE) instead of over-achieving. The second technique is Discontinuous Transmission (DTX), a very short sleep mode. Of the two, PC was found to provide strong savings in high load when transmission power is generally high, whereas DTX was most effective at low load when there is ample opportunity for sleep mode. It was shown that the joint application of PC and DTX achieves lower power consumption than either individual technique over all cell traffic loads in the Power and Resource Allocation Including Sleep (PRAIS) scheme. This joint problem was shown to be convex and thus efficiently solvable. This finding was combined with the third technique, Antenna Adaptation (AA), and applied by using it as the core of a power-saving Multiple-Input Multiple-Output (MIMO) Orthogonal Frequency Division Multiple Access (OFDMA) resource scheduler, called Resource allocation using Antenna adaptation, Power control and Sleep modes (RAPS). RAPS first solves a simplified convex real-valued subproblem. The solution of the subproblem is then quantized to the integer-valued OFDMA scheduling problem and joined with subcarrier allocation and power allocation. AA was found to provide little added value on top of the other two techniques. Application of RAPS was shown to reduce power consumption of LTE BSs by 25-40% in single-cell simulation. RAPS could be readily built into current LTE BSs.

Going beyond the cell level, Chapter 5 studied the exploration of intercell interference for power saving. It is beneficial to align the DTX mode of BSs constructively, to the effect that interference is reduced. Four alternative

alignment techniques have been tested in simulation: sequential alignment, random alignment, p-persistent ranking, and distributed DTX with memory. The latter is an original contribution by the author. Sequential alignment is a conservative technique which maximizes interference for the benefit of predictability. Random alignment is assumed was the SotA. P-persistent ranking is a popular technique for the suppression of oscillation. Distributed DTX with memory goes further and keeps track of past allocations and channel quality to enable convergence to a stable network state. The important finding was that—unlike for the single cell in Chapter 4—the power consumption could not be reduced through these mechanisms without affecting the Quality of Service (QoS). The interference dynamics cause occasional violations of the rate constraints, thus necessitating retransmissions. Random alignment caused regular rate constraint violations at medium power consumption. P-persistent ranking was slightly more reliable with lower retransmission probability at low power consumption. However, distributed DTX with memory provided a much lower retransmission probability than p-persistent ranking at lower power consumption. It thus provided the best best retransmission probability and power saving. The power saving of distributed DTX with memory over the SotA was found to be up to 40%, depending on the cell loads.

In conclusion, this thesis provides three readily applicable supply power models for LTE BSs, including the model derivations. From these models, DTX and PC were identified as most effective and applied to reduce BS operating power consumption. PC and DTX were optimized jointly and integrated into a comprehensive MIMO OFDMA scheduler. The inherent DTX time slots of such a scheduler were then aligned constructively between neighbouring BSs. Reducing BS power consumption by 50% through software is thus well within reach.

6.2 Limitations and Future Work

The most important limitation to the techniques proposed in this thesis lies in the restrictions imposed on the scheduler by a transmission standard like LTE. In this research, it was assumed that a scheduler can select any resource block for transmission or sleep. In practice, however, this selection will be limited as the standard reserves some resources for signaling, channel sounding, or pilot transmission. Since the scheduler will be less potent due to this limited selection, power consumption is expected to be higher under these constraints.

Next, the power model in Chapter 3 as well as all later chapters assume that a BS can be switched on and off in an instant. Yet, although switching times are assumed to be in the order of tens of microseconds [43], they are not fully instantaneous. This has to be considered in DTX and wake-up scheduling. If sleep modes are deeper, switching times will become longer. Following this line of argument, there may actually be more than one sleep mode available; for example, an OFF state with zero power consumption, a deep sleep with very low consumption, and a light sleep for DTX. In that case, a scheduler would have to consider the wake-up times in its selection of sleep modes.

Finally, the validity of the results presented over the coming years is strongly dependent on the assumptions made in the power model. Drastic changes in the architecture of BS, for example through the proliferation of massive MIMO [114], may change the way BSs consume power. In such a case, the effectiveness of different saving techniques will have to be re-evaluated.

For future research, the most immediate goal is to combine the RAPS scheduler with distributed DTX with memory and assess the joint power savings. As the RAPS scheduler makes no assumptions about the time slots allocated for DTX, it can provide instructions for the DTX alignment scheme. A further planned

extension to the DTX alignment scheme is an decrease of the retransmission probability. It is expected that a rate margin imposed on the rate constraint, *i.e.* requesting a higher rate target, will be beneficial.

Another important addition is the consideration of standards limitations. The strength of the impact of standards limitations has to be assessed. If the standard in its current form severely caps power savings, future standards could be designed to be more open to power saving.

The power-saving RRM mechanisms established herein can also be applied to BSs which are smaller than macro. In smaller BS types, the load dependence is smaller. Thus, PC is less effective compared to DTX. Consideration of smaller BS types is also related to the field of HetNets [115]. Since coverage is generally assumed to overlap in HetNets, new options for longer sleep modes arise as UEs will have guaranteed coverage even when some BSs are asleep.

All analyses in this work assumed a uniform distribution of BSs. However, with the advance of HetNets, it may also be relevant to explore distributions such as Poisson Point Processes [116]. This will have an effect particularly on the multi-cell DTX alignment work.

Appendix A

Appendix

A.1 Proof of Convexity for Problem (4.12)

This proof shows that (4.12) is convex.

Proof. For $P_S = 0$, the first derivative of the cost function for link k is

$$f'_{0,k} = \frac{\partial P_{\text{supply}}(\bar{R}_k)}{\partial \mu_k} = P_0 + \Delta_p \frac{P_N}{G_k} \left[-1 + 2^{\frac{\bar{R}_k}{W\mu_k}} \left(1 - \frac{\bar{R}_k}{W\mu_k} \ln(2) \right) \right]. \quad (\text{A.1})$$

The second derivative of $f_{0,k}$ is given by

$$f''_{0,k} = \frac{\partial^2 P_{\text{supply}}(\bar{R}_k)}{\partial^2 \mu_k} = \Delta_p \frac{P_N}{G_k} \left(\frac{\bar{R}_k}{W} \right)^2 \ln^2(2) \frac{2^{\frac{\bar{R}_k}{W\mu_k}}}{\mu_k^3}. \quad (\text{A.2})$$

All variables in the second derivative are positive, thus $f''_{0,k} \geq 0$ within the parameter bounds. Therefore, each $f_{0,k}$ is convex within the bounds.

The non-negative sum preserves convexity. Thus, f_0 is convex. \square

A.2 Proof of Convexity for Problem (4.15)

This proof shows that (4.15) is convex.

Proof. The partial second derivative of the MIMO cost function in (4.15) with respect to μ_k is

$$\frac{\partial^2 P_{\text{supply},k}(\mathbf{r})}{\partial^2 \mu_k} = \Delta_p \log^2(2) \frac{R_k}{W} \frac{2^{\frac{R_k}{W\mu_k}} \left((\epsilon_1 + \epsilon_2)^2 + 2\epsilon_1\epsilon_2 \left(2^{\frac{R_k}{W\mu_k}} - 2 \right) \right)}{\mu_k^3 \left((\epsilon_1 + \epsilon_2)^2 + 4\epsilon_1\epsilon_2 \left(2^{\frac{R_k}{W\mu_k}} - 1 \right) \right)^{\frac{3}{2}}}, \quad (\text{A.3})$$

which is non-negative since

$$2\epsilon_1\epsilon_2 2^{\frac{R_k}{W\mu_k}} \geq 2\epsilon_1\epsilon_2 \quad (\text{A.4})$$

within the parameter bounds. Thus the cost function is convex in μ_k within the bounds. \square

Convexity of the Single-Input Multiple-Output (SIMO) cost function and the constraint function can be shown similarly and is omitted here for brevity.

A.3 Margin-adaptive Resource Allocation

Margin-adaptive power allocation over a set of resource blocks, \mathcal{U}_k , and a vector of channel eigenvalues, $\mathcal{E}_{a,k}$, minimizes power consumption while fulfilling a target rate [103].

Assuming block-diagonalization precoding, the problem is to

$$\underset{P_{a,e}}{\text{minimize}} \quad \sum_{a=1}^{|\mathcal{U}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} P_{a,e} \quad (\text{A.5a})$$

$$\text{subject to} \quad B_{\text{target},k} = \sum_{a=1}^{|\mathcal{U}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} w\tau \log_2 \left(1 + \frac{P_{a,e} \mathcal{E}_{a,k}(e)}{P_N} \right), \quad (\text{A.5b})$$

$$P_{a,e} \geq 0 \quad \forall a, e, \quad (\text{A.5c})$$

$$\sum_{a=1}^{|\mathcal{U}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} P_{a,e} \leq P_{\max}. \quad (\text{A.5d})$$

With Lagrange multipliers λ, β, ν , the Karush-Kuhn-Tucker optimality conditions for (A.5) are

$$-P_{a,e} \leq 0 \quad \forall a, e \quad (\text{A.6a})$$

$$\sum_{a=1}^{|\mathcal{U}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} P_{a,e} - P_{\max} \leq 0 \quad (\text{A.6b})$$

$$B_{\text{target},k} - \sum_{a=1}^{|\mathcal{U}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} w\tau \log_2 \left(1 + \frac{P_{a,e} \mathcal{E}_{a,k}(e)}{P_N} \right) = 0 \quad (\text{A.6c})$$

$$\lambda_{a,e}^* \geq 0 \quad \forall a, e, \quad \beta \geq 0 \quad (\text{A.6d})$$

$$\lambda_{a,e}^* (-P_{a,e}^*) = 0 \quad \forall a, e \quad (\text{A.6e})$$

$$\beta \left(\sum_{a=1}^{|\mathcal{U}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} P_{a,e} - P_{\max} \right) = 0 \quad (\text{A.6f})$$

$$\frac{\partial \mathcal{L}}{\partial P_{a,e}} = 0, \quad (\text{A.6g})$$

where (A.6a), (A.6b), (A.6c) ensure the primal feasibility, (A.6d) the dual feasibility and (A.6e), (A.6f) the complementary slackness.

The derivative of the Lagrangian at the minimum is

$$\frac{\partial \mathcal{L}}{\partial P_{a,e}} = 1 - \lambda + \beta - \frac{\nu w\tau \mathcal{E}_{a,k}(e)}{\log(2) (P_N + \mathcal{E}_{a,k}(e) P_{a,e})} = 0, \quad (\text{A.7})$$

with λ the multiplier for the first inequality constraint, β for the second (power-limit) constraint and ν the equality constraint multiplier. When reducing the conditions to $P_{a,e} > 0$, $\lambda = 0$. When considering that the power constraint is always loosely bound with $\left(\sum_{a=1}^{|\mathcal{U}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} P_{a,e} - P_{\max} \right) < 0$, then by (A.6f), $\beta = 0$.

Solving the remainder for $P_{a,e}$,

$$P_{a,e} = \frac{\nu w \tau}{\log(2)} - \frac{P_N}{\mathcal{E}_{a,k}(e)}, \quad (\text{A.8})$$

which can be inserted into the equality constraint of (A.5) to yield the water-level, ν , by

$$\log_2(\nu) = \frac{1}{|\Omega_k|} \left(\frac{B_{\text{target},k}}{w\tau} - \sum_{e=1}^{|\Omega_k|} \log_2 \left(\frac{\tau \mathcal{E}_{a,k}(e) w}{P_N \log(2)} \right) \right). \quad (\text{A.9})$$

The water-level can be found via an iterative search over the vector Ω_k of channels that contribute a positive power. Since the sum power is reduced on each iteration, the power constraint (accounted for by the multiplier β) only needs to be tested after the search is finished.

Appendix B

List of Publications

This chapter contains a list of publications related to this thesis ordered by academic publication status (published, accepted, submitted) and other type (project reports and book contributions).

B.1 Published

- Gunther Auer and István Gódor and László Hévizi and Muhammad Ali Imran and Jens Malmudin and Péter FASEKAS and Gergely Biczók and **Hauke Holtkamp** and Dietrich Zeller and Oliver Blume and Rahim Tafazolli, “Enablers for Energy Efficient Wireless Networks”, *Proceedings of the IEEE Vehicular Technology Conference (VTC) Fall-2010*, 2010 [a]
- **Hauke Holtkamp** and Gunther Auer, “Fundamental Limits of Energy-Efficient Resource Sharing, Power Control and Discontinuous Transmission”, *Proceedings of the Future Network & Mobile Summit 2011*, 2011 [b]
- **Hauke Holtkamp** and Gunther Auer and Harald Haas, “Minimal Average

Consumption Downlink Base Station Power Control Strategy”, *Proceedings of the 2011 IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, 2011 [c]

- **Hauke Holtkamp** and Gunther Auer and Harald Haas, “On Minimizing Base Station Power Consumption”, *Proceedings of the Vehicular Technology Conference (VTC) Fall-2011*, 2011 [d]
- Claude Desset and Björn Debaillie and Vito Giannini and Albrecht Fehske and Gunther Auer and **Hauke Holtkamp** and Wieslawa Wajda and Dario Sabella and Fred Richter and Manuel Gonzalez and Henrik Klessig and István Gódor and Per Skillermark and Magnus Olsson and Muhammad Ali Imran and Anton Ambrosy and Oliver Blume, “Flexible Power Modeling of LTE Base Stations”, *Proceedings of the 2012 IEEE Wireless Communications and Networking Conference*., 2012 [e]
- **Hauke Holtkamp** and Gunther Auer and Samer Bazzi and Harald Haas, “Minimizing Base Station Power Consumption”, *IEEE Journal on Selected Areas in Communications*, 2014 [j]

B.2 Accepted

- **Hauke Holtkamp** and Gunther Auer and Vito Giannini and Harald Haas, “A Parameterized Base Station Power Model”, *IEEE Communications Letters*, 2013 [h]
- **Hauke Holtkamp** and Harald Haas, “OFDMA Base Station Power-saving Via Joint Power Control and DTX in Cellular Systems”, *Proceedings of the Vehicular Technology Conference (VTC) Fall-2013*, 2013 [i]

B.3 Submitted

- Hauke Holtkamp and Guido Dietl and Harald Haas, “Distributed DTX Alignment with Memory”, *Proceedings of the 2014 IEEE International Conference on Communications (ICC)*, 2014 [k]

B.4 Project Reports

- EARTH Project Work Package 2, “Deliverable D2.2: Reference Systems and Scenarios”, 2012 [f]
- EARTH Project Work Package 3, “Deliverable D3.3: Green Network Technologies”, 2012 [g]

B.5 Contributions

- Jinsong Wu and Sundeep Rangan and Honggang Zhang, “Green Communications: Theoretical Fundamentals, Algorithms and Applications”, *Taylor & Francis Group*, 2012 [l]

Appendix C

Attached Publications

This chapter contains all work either published or submitted for publication to academic conferences or journals. For brevity, Energy Aware Radio and neTwork tecHnologies (EARTH) project deliverables [g, f] and a book chapter [l] are not listed here.

Enablers for Energy Efficient Wireless Networks

(Invited Paper)

Gunther Auer*, István Gódor[†], László Hévízi[†], Muhammad Ali Imran[‡], Jens Malmodin[§], Péter Fazekas[¶], Gergely Biczók[¶], Hauke Holtkamp*, Dietrich Zeller^{||}, Oliver Blume^{||}, Rahim Tafazolli[‡]

**DOCOMO Euro-Labs, 80687 Munich, Germany, Email: lastname@docomolab-euro.com*

[†]*Ericsson Research, Budapest H-1117, Hungary*

[‡]*CCSR University of Surrey, Guildford GU2 7XH, UK*

[§]*Ericsson Radio Systems, Stockholm, Sweden*

[¶]*Budapest University of Technology and Economics, Budapest H-1117, Hungary*

^{||}*Alcatel-Lucent, Bell Labs Germany, 70435 Stuttgart, Germany*

Abstract—Mobile communications are increasingly contributing to global energy consumption. The EARTH (Energy Aware Radio and neTworking tecHnologies) project tackles the important issue of reducing CO₂ emissions by enhancing the energy efficiency of cellular mobile networks. EARTH is a holistic approach to develop a new generation of energy efficient products, components, deployment strategies and energy-aware network management solutions. In this paper the holistic EARTH approach to energy efficient mobile communication systems is introduced. Performance metrics are studied to assess the theoretical bounds of energy efficiency as well as the practical achievable limits. A vast potential for energy savings lies in the operation of radio base stations. In particular, base stations consume a considerable amount of the available power budget even when operating at low load. Energy efficient radio resource management (RRM) strategies need to take into account slowly changing daily load patterns, as well as highly dynamic traffic fluctuations. Moreover, various deployment strategies are examined focusing on their potential to reduce energy consumption, whilst providing uncompromised coverage and user experience. This includes heterogeneous networks with a sophisticated mix of different cell sizes, which may be further enhanced by energy efficient relaying and base station cooperation technologies. Finally, scenarios leveraging the capability of advanced terminals to operate on multiple radio access technologies (RAT) are discussed with respect to their energy savings potential.

I. INTRODUCTION

The global mobile communications industry is growing rapidly. Today there are more than four billion mobile phone subscribers worldwide, more than half the entire population of the planet. Obviously, this growth is accompanied by increased energy consumption of mobile networks. The objective of the EARTH project is to address the global environmental challenge by identifying effective mechanisms to substantially reduce energy wastage and to improve energy efficiency of mobile communication systems, without compromising network coverage and users perceived quality of service.

In order to quantify the merits of the envisaged energy savings, a holistic system view will be adopted, ensuring that any radio energy efficiency improvement in one area is not wiped out by an increased energy consumption in a different area, e.g. gains at the base station are not neutralized in the core network or at the user terminal. While the performance of

radio access technologies is typically characterized by metrics like spectral efficiency, throughput or coverage, these metrics are unable to quantify the energy efficiency of a network. A thorough analysis of key radio and network technologies in terms of their energy efficiency is not yet available. This is partly due to the fact that there are no suitable methodologies and no sophisticated metrics available that enable a fair and objective evaluation of the entire system, ranging from high power efficiency at the base station site — including the necessary air conditioning — to alternative deployment strategies, such as relaying concepts, heterogeneous networks and base station cooperation. Thus, an important prerequisite for efforts to enhance the energy efficiency of cellular networks is to provide a framework to quantify the attainable energy savings.

Today's mobile networks have a vast potential for energy savings. While the energy efficiency of mobile terminals is highly optimized due to the stringent constraints on the available power supply, until recently power consumption of base stations has been largely ignored. This has led to a situation where terminal energy consumption is only a fraction of the energy consumption of the mobile network, while base stations are the major source of CO₂ emissions [1, 2]. Consequently, the EARTH project has put its focus on enhancing the energy efficiency of radio base stations. In particular, even at low load base stations consume a considerable amount of the available power budget.

Cellular networks exhibit slowly changing daily load patterns as well as highly dynamic traffic fluctuations. However, current networks are configured rather statically. Since most of the energy wastage occurs during low load situations, these load variations can be effectively exploited to reduce network energy consumption. Vendors and operators have started to exploit these load variations by bundling traffic in certain transceivers and switching others off. However, such algorithms are at an early stage and with the emergence of self optimizing networks (SON), there is an emerging potential to react on slowly changing traffic variations. Similarly, highly dynamic load fluctuations may be addressed by energy efficient radio resource management (RRM) strategies. EARTH

will contribute with concepts and algorithms to significantly reduce the energy consumption in low load situations.

One further source of energy wastage is the layout of today's cellular networks. Obviously, a network layout optimized for coverage does not necessarily maximize energy efficiency. EARTH aims at devising fundamental principles of energy efficient network layouts. In particular we will investigate the use of technologies beyond the state of the art, such as advanced repeaters, relaying solutions and heterogeneous deployments that deviate from the conventional cellular deployment paradigm. This includes heterogeneous networks with a sophisticated mix of different cell sizes, optimized for minimal energy consumption. Ultimately, this leads to novel energy efficient network architectures.

In this paper the challenges and key enablers for energy efficient wireless networks are described. A discussion of the key levers for energy efficient radio components is provided in a companion paper [3]. The impact of mobile communication systems on the global CO₂ emission is discussed in Section II. Moreover, the major sources of energy wastage are identified. Section III discusses appropriate energy efficiency metrics that enable a holistic system level comparison of competing technologies. Section IV and V outline the potential of RRM strategies and alternative network deployments that aim to complement the cellular network layout paradigm. Section VI elaborates on energy savings envisaged by terminals that are able to operate on multiple radio access technologies.

II. SOCIO-ECONOMIC IMPACT AND LIFE CYCLE ANALYSIS

As of today, CO₂ equivalents (CO₂-eq) are generally used to measure the potential global warming contributions from organizations and products, e.g. in environmental research, societal push (e.g. Kyoto protocol), and in the media. The largest source of CO₂-eq emissions are fossil fuels. It has become the most common and understandable metric, allowing to directly compare the contribution (the *carbon footprint*) of a given technology sector or organization with other contributions and reduction targets. When calculating carbon footprints, a holistic approach such as life cycle analysis (LCA) is recommended, rather than simple methods that only look at the use phase.

A. Global carbon footprint of wireless communications

Today, Information and Communication Technology (ICT) systems including data networks, fixed and wireless networks, as well as all end-user equipment such as PCs, home network equipment and mobile devices, are responsible for about 2% of global CO₂ emissions and about 1.5% of global CO₂-eq emissions, as shown in Fig. 1. In order to tackle climate change the EU Commission recently issued a call on ICT industry for intensified efforts to reduce its carbon footprint by 20% as early as 2015 for [4]. The mobile part of ICT is responsible for about 0.2% of global CO₂-eq emissions but the number of radio base station (RBS) sites and mobile subscribers are steadily increasing (4 billion subscribers right now) and data-intensive applications are proliferating. Obviously, this

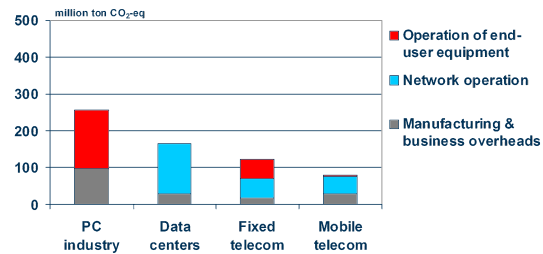


Fig. 1. CO₂-eq emissions for the different ICT subsectors in 2007. Data centers also include data networks and data transport.

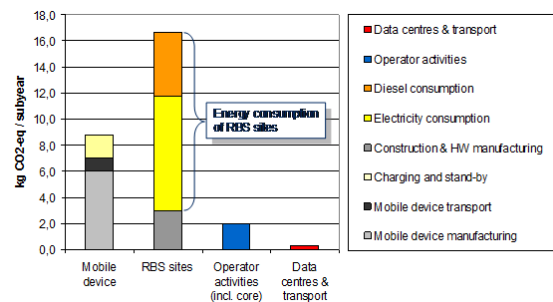


Fig. 2. The carbon footprint (CO₂-eq emissions) for an average mobile subscriber in 2007.

growth is accompanied by an increased energy consumption of mobile networks, with a corresponding increase in the carbon footprint [5].

On the other hand, rising energy costs, especially for networks where a large share of the RBS sites runs on diesel generators, have led to a situation where such sites contribute significantly to the network operation costs. In fact, operator's OPEX figures indicate that their energy costs are nowadays comparable to their personnel costs for network operations.

B. Life cycle analysis

Fig. 2 shows the carbon footprint for an average mobile subscriber, based on data for the full year 2007. The carbon footprint is based on extensive LCA studies covering the whole life cycle of the various components that are part of a mobile network.

It can be clearly seen that the energy consumption of RBS sites is responsible for the largest part of the total life cycle carbon footprint, followed by the mobile phone manufacturing. As a majority of sites has low traffic most of the time, this "stand-by" behavior of the network has a large potential for energy savings in the future.

An important part in an LCA of mobile networks is to assess the emissions from the electricity production sources. A world average electricity production model is used in the ICT and mobile network study presented here. The resulting emission

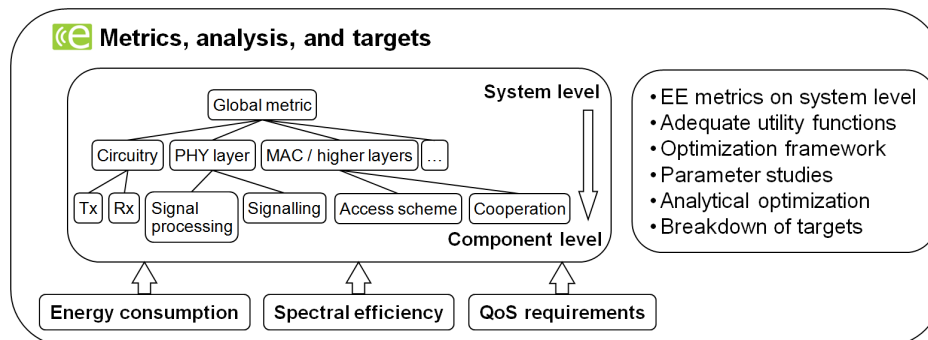


Fig. 3. EARTH approach towards a global energy efficiency metric.

factor is 0.6 kg CO₂-eq/kWh, where 0.5 kg CO₂ comes from the direct production at the power plants.

The carbon footprint for an average mobile subscriber has decreased over time from more than 100 kg CO₂ per subscriber per year in the early 90's to approx. 25 kg in the mid 00's. However, the increased number of RBS sites operated on diesel, "growing" mobile devices, and the rapid growth of new mobile networks has reversed this positive trend.

How much CO₂ is 25 kg? It equals to the amount emitted when 10.5 liter petrol is incinerated in a petrol engine, which takes you about 1 hour or 120 km on the motorway.

III. ENERGY EFFICIENCY METRICS

The classical optimization criterion for wireless network is the areal spectral efficiency, measured in [bit/s/Hz/m²]. This implies maximizing the data rate per unit area for a given transmit power budget. Conventional energy efficiency metrics typically capture only a small fraction of the overall power budget of wireless networks, and may therefore lead to incomplete and potentially misleading conclusions. To this end, an appropriate energy efficiency metric should take into account the following components:

- The input power required to generate a specific output power at the antennas. This input power accounts for the desired power in analogue front-end, the power amplifier losses, air conditioning power consumption within base stations and the power loss of the feeder cables.
- The energy that is consumed by running the digital signal processing equipment (e.g. powerful channel coding with the associated complex decoders).
- The impact of the RRM scheme on the energy efficiency; such as subcarrier and transmit power allocation and base station cooperation (energy required for the exchange of message signals as well as additional control signaling).
- The energy to deliver data to the base station (backhaul power consumption).

A system level view requires a global energy efficiency metric, which is composed of specific metrics on the component level, as illustrated in Fig. 3. The challenge in devising a suitable

energy efficiency metric lies not only in the dimension of the problem space, but also in the fact that the power consumption of the relevant components is interdependent.

Standardization of energy efficiency metrics has already commenced. For instance the ETSI Technical Committee on Environmental Engineering (ETSI-TC-EE) [6] has initiated work on a specification of energy efficiency metrics for mobile cellular networks. However, these specifications only consider energy efficiency metrics on component level (e.g. power efficiency for high power amplifiers) rather than the whole system (end-to-end connectivity including backhaul). Furthermore, criteria related to the quality-of-service are neglected.

For measuring the efficiency of the communication link, several metrics are used in the literature. The most commonly used metric for the energy efficiency of a communication link is *bit per joule* [7], which is an information theoretic measure of the energy efficiency of delay insensitive communication. To account for the data rate as well as the communication distance, a modified metric of *bit meter per joule* may be used. This metric describes the efficiency of reliably transporting the bits over a distance towards the destination per unit of energy consumed [8]. For a cellular area of coverage, this metric is to be modified to *bits per joule per unit area*, so to capture the extent of the coverage area.

The system level energy efficiency of a single cell was studied in [9], where the energy efficiency of communicating over a frequency selective cellular uplink channel was optimized. The authors conclude that both the circuit energy (energy required to keep nodes active) and transmission energy need to be modeled, so to measure the overall energy efficiency in terms of bit per joule.

There exist a multitude of individual figures of merit for optimizing the different aspects and components of wireless systems (e.g. spectral efficiency, power consumption for data transmission and processing, quality of service parameters such as throughput, error rate, delay (jitter), coverage). It is indispensable that an energy efficiency metric takes all of the aforementioned aspects into account; otherwise the optimization might just improve towards one direction, at

the expense of other important parameters. In this regard the EARTH project is developing more sophisticated energy efficiency metrics in the form of desired utility per unit of energy consumption covering a unit area of cellular coverage. The desired utility is to be calculated using a utility function capturing an appropriate mix of the above mentioned figures of merit.

IV. RADIO RESOURCE MANAGEMENT

An essential component of energy efficient operation is that the network management adapts the network configuration and topology to the daily, spatial variation of traffic demand. In GSM/WCDMA networks, the state-of-the-art is to turn on and off some selected transceivers, cells and the associated backhaul links, e.g., during low-traffic night hours [10]. The EARTH project will provide new network management concepts for dynamic reconfiguration of radio networks (e.g. self-optimization, self-configuration and standby-operation). This will allow operators to utilize the different network deployment concepts (e.g. optimal mix of cell sizes, multi-RAT deployments) and to dynamically exploit the potentials in reduction of energy consumption.

In order to save energy, the management procedures should concentrate on the fact that during low traffic hours it is the coverage requirement, in peak hours it is the capacity requirement, which determines the topology. As a consequence of the slow daily variation of capacity requirements, the (self) adaptation is assumed to operate on a longer time scale (minutes, hours) compared to radio resource management (RRM) (below seconds).

It is also important to identify situations where the re-configuration of a system component or the coordinated re-configuration of a group of components may save energy. Such situations arise when the antenna tilt is adapted to modify cell coverage, or by switching between various MIMO functionalities, etc. In addition, the network should be enabled to point the capacity into the place of need, e.g. by directing adaptive antenna arrays towards a traffic hotspot.

Cooperation mechanisms between base stations will be studied from an energy efficiency perspective. The transmission power of individual base stations may be reduced due to increased coverage, increased spectral efficiency, and adaptation to load variations between cells. On the other hand, energy consumption may increase due to more complex signal processing and/or additional backhaul traffic. Fundamental trade-offs between these factors will be studied providing the basis for designing energy efficient site cooperation protocols. As cooperating base stations resemble a distributed MIMO system, studies for energy saving potentials of MIMO systems provide a good starting point [11]. User cooperation for energy efficiency is studied in a simplistic two user scenario in [12].

V. ENERGY EFFICIENT NETWORK DEPLOYMENT

Both environmental concerns and operational cost efficiency recently diverted the attention of wireless broadband research activities towards energy efficient deployment. Recent studies

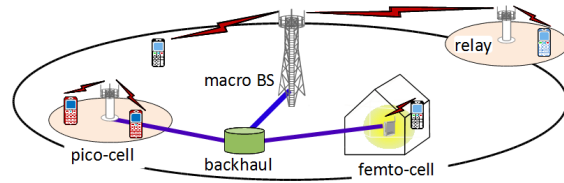


Fig. 4. Heterogeneous network deployment composed of macro-cells that provide wide area coverage, micro and pico-cells serving hot spots, femto-cells that are deployed indoors, and relays.

define specific scenarios that have potential to improve the system energy efficiency. An overview of some promising approaches for improved energy efficient wireless access is provided in [1, 2, 13, 14].

In addition to energy efficient base station and core network design, emerging deployment strategies are to be considered that incorporate heterogeneous networks with cells of various size, ranging from macro-cells, over micro and pico-cells serving hot spots, to femto-cells that are deployed indoors, as shown in Fig. 4. The short distance between transmitter and receiver offered by small cell sizes suggest an enormous potential for energy savings; in turn, however, at low traffic loads the increased number of base stations at standby mode add to the overall power consumption, which may cancel parts of these envisaged energy savings. Hence an optimal balance must be found leading to the most energy efficient deployment scenario. In the EARTH project, the energy consumption to maintain network operation will be evaluated as a function of cell size and deployment scenario for given system loads and environments, ranging from dense urban to rural areas. Several deployment options like sectorization, antenna tilting and distributed antennas also need to be considered. Some initial work in this area is reported in [15–18].

Heterogeneous networks are complemented by relays, repeaters and cooperating base stations for the most energy efficient mix of different deployment strategies. Relay and repeater nodes are often promoted for infrastructure cost saving reasons, since expensive backhaul links are avoided. Another potential advantage is energy saving, since the propagation distance per hop (and thereby transmission power) is decreased [19]. These and other potential energy efficiency saving features [20] of relays and repeaters will be further explored within the EARTH project.

VI. MANAGEMENT OF MULTI-RAT NETWORKS

Currently base stations of multiple radio access technologies (RAT) are usually colocated and several RATs may cover the same area, but they provide different levels of services (cf. GSM, 3G, WLAN). With the projected roll-out of the 3GPP long term evolution (LTE) network, the number of available radio interface technologies further increases.

Multi-RAT terminals may take advantage of coexistence of different RATs. Unfortunately, the present practice is that each RAT has its own network planning, operations & maintenance

and resource management methodologies. Recent efforts from operators and vendors, e.g. carried out within the EU project Ambient Networks [21], aim at joint management of multi-RAT technologies, focusing on capacity and coverage optimization.

In order to harvest energy savings from multi-RAT technologies new network functions are needed, to facilitate the dynamic distribution of load among interfaces (without any user intervention), to support the parallel use of multiple interfaces for the same data transfer (e.g. large file download), and to enable smooth switching of customer traffic between interfaces. Specifically, radio resource management (RRM) is to be enabled to utilize multiple air interfaces as a common resource. EARTH will develop deployment and operating strategies on how the network of different RATs can be deployed and operated in a coordinated fashion.

In order to reduce the energy consumption in multi-RAT deployments, as rule of thumb, the RAT with the lowest carrier frequency should be selected to maintain coverage at minimal power dissipation, due to the favorable propagation conditions at lower frequencies. Along this line, very low power and low coverage RATs (e.g. LTE femto cells) that typically operate in higher frequency bands should cover higher traffic demand areas. In this latter case, the short distance between transmitter and receiver overcompensates the energy efficiency loss of the high carrier frequency.

Naturally, the goals of EARTH go beyond the above examples and try to find the best network management solutions, which provide a joint optimization of multi-RAT networks.

VII. CONCLUSIONS

Beyond today's focus of network element design, the European research project EARTH develops a holistic approach for the design of energy efficient mobile communication networks of the future. Energy efficiency metrics, deployment strategies and radio resource management are identified as key research areas beyond the radio node level research [3].

The project will address both the operational impacts energy efficient mobile radio and the impacts on climate change. We believe that cost saving can make ICT services sustainable both energetically and economically, thus, allowing further reduction of the global CO₂ footprint, e.g. by teleconferencing, telecommuting and dematerialization. The ambitious goal of EARTH is to cut the energy consumption of mobile broadband networks by at least 50%.

The results of EARTH will be introduced to standardization bodies and industry product development in order to ensure impact of the project to the technical community.

VIII. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contribution of the EARTH consortium to the technical definition of the project.

The work leading to this paper has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement n° 247733 — project EARTH.

REFERENCES

- [1] M. Gruber, O. Blume, D. Ferling, D. Zeller, M. A. Imran, and E. Calvanese-Strinati, "EARTH - Energy Aware Radio and Network Technologies," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Cannes, France, 2008.
- [2] J. T. Louhi and H.-O. Scheck, "Energy efficiency of cellular networks," in *Proc. Int. Symp. Wireless Personal Multimedia Communications (WPMC)*, Lapland, Finland, 2008.
- [3] D. Ferling, T. Bohn, D. Zeller, P. Frenger, I. Godor, Y. Jading, and W. Tomaselli, "Energy Efficiency Approaches for Radio Nodes (submitted)," in *Future Network & Mobile Summit*, Florence, Italy, 2010.
- [4] "EU Commissioner calls on ICT industry to reduce its carbon footprint by 20% as early as 2015," 2009. [Online]. Available: <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/09/140>
- [5] Ericsson, "Long Term Evolution (LTE): an introduction," 2007. [Online]. Available: http://www.ericsson.com/technology/whitepapers/lte_overview.pdf
- [6] B. Gorini, "ETSI work programme on energy saving, presented at Telecommunications Energy Conference," 2007.
- [7] V. Rodoplu and T. Meng, "Bits-per-Joule Capacity of Energy-Limited Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 3, pp. 857–865, 2007.
- [8] J. L. Gao, "Analysis of Energy Consumption for Ad Hoc Wireless Sensor Networks Using a Bit-Meter-per-Joule Metric," 2002.
- [9] G. Miao, N. Himayat, and Y. Li, "Energy-efficient transmission in frequency-selective channels," in *Proc. IEEE Globecom*, 2008.
- [10] Ericsson, "Sustainable energy use in mobile communications," 2007. [Online]. Available: http://www.ericsson.com/technology/whitepapers/sustainable_energy.pdf
- [11] S. Cui, A. Goldsmith, and A. Bahai, "Energy-Efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1089–1098, 2004.
- [12] T. Himsoon, W. Siriwongpairat, and K. Liu, *Energy-efficient cooperative transmission over multiuser OFDM networks: who helps whom and how to cooperate?*. IEEE, 2005.
- [13] L. Hérault, "Sustainability of mobile network solutions," Santander, Spain, 2009.
- [14] L. Hérault, E. Calvanese Strinati, D. Zeller, O. Blume, M. A. Imran, R. Tafazolli, J. Lundsjö, Y. Jading, and M. Meyer, "Green Communications: A Global Environmental Challenge," in *Proc. Int. Symp. Wireless Personal Multimedia Communications (WPMC)*, Sendai, Japan, 2009.
- [15] H. Claussen, L. T. W. Ho, and F. Pivit, "Leveraging advances in mobile broadband technology to improve environmental sustainability," *Telecommunications Journal of Australia*, vol. vol59, 2009.
- [16] —, *Effects of joint macrocell and residential picocell deployment on the network energy efficiency*. IEEE, 2008.
- [17] E. Bogenfeld and I. Gaspard, "A White Paper by the FP7 project End-to-End Efficiency (E3) Self-x in Radio Access Networks," *Framework*, pp. 1–24, 2008. [Online]. Available: https://www.ict-e3.eu/project/white_papers/Self-x_WhitePaper_Final_v1.0.pdf
- [18] H. Karl, "An overview of energy efficiency techniques for mobile communication systems," Berlin, 2003.
- [19] P. Herhold, W. Rave, and G. Fettweis, "Relaying in CDMA networks: Pathloss reduction and transmit power savings," in *Proc. IEEE Vehicular Technology Conference (VTC)*, 2003.
- [20] A. Radwan and H. S. Hassanein, *NXG04-3: Does Multi-hop Communication Extend the Battery Life of Mobile Terminals?* IEEE, 2006.
- [21] EU FP6 IP, "Ambient Networks," 2007. [Online]. Available: <http://www.ambient-networks.org>

Minimal average consumption downlink base station power control strategy

Hauke Holtkamp, Gunther Auer

DOCOMO Euro-Labs
D-80687 Munich, Germany
Email: {holtkamp, auer}@docomolab-euro.com

Harald Haas

Institute for Digital Communications
Joint Research Institute for Signal and Image Processing
The University of Edinburgh, EH9 3JL, Edinburgh, UK
E-mail: h.haas@ed.ac.uk

Abstract—We consider single cell multi-user OFDMA downlink resource allocation on a flat-fading channel such that average supply power is minimized while fulfilling a set of target rates. Available degrees of freedom are transmission power and duration. This paper extends our previous work on power optimal resource allocation in the mobile downlink by detailing the optimal power control strategy investigation and extracting fundamental characteristics of power optimal operation in cellular downlink. We find that only a system wide allocation of transmit powers is optimal rather than on link level. The allocation strategy that minimizes overall power consumption requires the transmission power on all links to be increased if only one link degrades. Furthermore, we show that for mobile stations with equal channels but different rate requirements, it is power optimal to assign equal transmit powers with proportional transmit durations. To relate the effectiveness of power control to live operation, we take the power model into consideration which maps transmit power to supply power. We show that due to the affine mapping, the solution is independent of the power model. However, the effectiveness of power control measures is completely dependent on the underlying hardware and the load dependence factor of a base station (instead of absolute consumption values). Finally, we conclude that power control measures in base stations are most relevant in macro stations which have load dependence factor of more than 50%.

I. INTRODUCTION

Mobile traffic volume is growing at a rapid pace. This requires more Base Stations (BSs) which are more powerful and more densely deployed. It is widely agreed that the operation of today's mobile networks causes 0.3% of global CO₂ emissions and that 80% of the operating energy is spent at the radio BS sites causing significant electricity and diesel bills for network operators [1], [2]. These combined environmental concerns and rising energy costs provide strong incentives to reduce the power consumption of future BS generations.

Traffic statistics reveal that network load varies significantly over the course of a day and different deployment patterns such as rural, suburban or urban. While BSs operate at maximum efficiency during peak (full load) hours, their load adaptability is limited resulting in low energy efficiency in low load situations. We seek to actively decrease the power consumption via power control during such low load times when the spectral and computational resources far outweigh the traffic demand.

Recent studies show that the overall power consumption of a radio transceiver is dominated by the consumption of the Power Amplifier (PA) [3]. Reduction of transmit power causes a significant decrease of the consumed power in the PA. This general notion is analytically included via the EARTH¹ power model, which determines the overall BS power consumption on the basis of transmit power. In this fashion, it is possible to determine absolute savings of power control strategies.

In our previous work, we have considered power control as an added improvement to systems which are capable of sleep modes, also labeled Discontinuous Transmission (DTX). We evaluated the combined gains on several generations of base stations. In this work, we provide a detailed investigation of the stand-alone power control strategy.

Power control has been extensively studied as a tool for rate maximization [4], [5] when spectral resources are sparse. But these strategies need to be adjusted when spectral efficiency is not the cost function. Wong et al. [6] provide an algorithmic strategy for power allocation in multiuser Orthogonal Frequency Division Multiple Access (OFDMA), but only for static modulation. By employing the Shannon limit directly, we presume optimal modulation. Al-Shatri et al. [7] optimize the operating efficiency, but requires a predetermined static number of resources per user. Cui et al. [8] optimize the modulation scheme on the link level, which we extend to the BS level.

The remainder of this paper is structured as follows. The system model is presented in Section II which results in the optimal allocation strategy. In Section III, first the optimal allocation of transmission power and time are discussed. It is then derived what transmit powers users with equal channels should receive. The power model is added in Section IV to change the optimization variable from transmit power to supply power. The paper is concluded in Section V.

II. SYSTEM MODEL

When a number of bits transmitted on a link with fixed bandwidth and optimal modulation is to be increased, there are two options. Either increase the link rate by raising the

¹EU funded research project EARTH (Energy Aware Radio and neTwork technologies), FP7-ICT-2009-4-247733-EARTH, Jan. 2010 to June 2012. <https://www.ict-earth.eu>

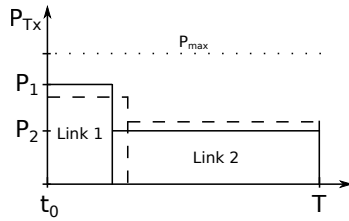


Fig. 1. Illustration of two possible power/time trade-offs that provide equal rates on both links.

transmission power or transmit for a longer time. From a power efficiency perspective providing a fixed rate is thus a trade-off between transmit power and transmission duration. As illustrated in Figure 1, a higher power for smaller duration will provide the same rate as a lower power with higher transmission duration. For a single link, transmission should always be the longest to allow for the lowest transmit power. However, in a shared multi-user channel with orthogonal access, all links have to be considered which each have individual rate requirements that have to be fulfilled in a set time.

We proceed with a derivation of the optimal allocation strategy. We employ the Shannon bound to map transmission power to achievable rate as the most fundamental law of energy consumption in communications. We assume that in a system a set of mobiles with known Channel State Information (CSI) has a set of rate requirements that needs to be fulfilled under minimum average system-wide power consumption.

A cell consists of a BS and N_L links (mobiles). The rate per link i is

$$R_i = W \log_2(1 + \gamma_i), \quad (1)$$

where W is the channel bandwidth in Hz and γ_i is the signal-to-noise ratio and is a function of the transmit power. With G_i the link channel gain, P_i the transmit power on link i in W and N_0 thermal noise in W, $\gamma_i = \frac{G_i P_i}{N_0}$. The noise power is defined by $N_0 = Wk\vartheta$ with Boltzmann constant k and operating temperature ϑ in Kelvin. While rate and bandwidth are linearly related, rate and transmit power have a logarithmic relationship. As a consequence it is much more expensive in terms of power to increase channel rate than in terms of bandwidth. In other words, if there is a choice between leaving idle bands and transmitting at higher power and using all available bands and transmitting at the lowest required power, then the latter will always consume less overall transmit power.

The average target rate \bar{R}_i per link in bps has to be fulfilled within the total time frame, T . Total energy consumed by the system is defined as

$$E_{\text{SYS}} = \bar{P}_{\text{SYS}} T \quad (2)$$

in J, where \bar{P}_{SYS} is the system average power consumption.

Normalized transmission time per link in s is given by

$$\mu_i = \frac{t_i}{T} \quad (3)$$

where t_i is the transmission time on link i and $\mu_i > 0$. Optimization of normalized results in average power minimization which is more illustrative than energy minimization (which is only meaningful for a known T).

The average target rate on link i over the time slot T depends on the transmission time and the rate during transmission, R_i :

$$\bar{R}_i = \mu_i R_i. \quad (4)$$

We are now able to find the transmission power per link as a function of the required average rate:

$$\bar{P}_{\text{Tx},i}(\bar{R}_i) = \frac{N_0}{G_i} \left(2^{\frac{\bar{R}_i}{W\mu_i}} - 1 \right), \quad (5)$$

where $0 < \bar{P}_{\text{Tx},i}(\bar{R}_i) < P_{\text{max}}$ for some P_{max} .

To account for the fact that all links are served by the BS orthogonally on the shared resource, the system average transmission power at the base station for all links over T is the sum of individual transmit powers weighted with the transmit duration [9]

$$\begin{aligned} \bar{P}_{\text{SYS}}(\bar{R}_i) &= \sum_{i=1}^{N_L} \mu_i \cdot P_{\text{Tx},i}(\bar{R}_i) \\ &= \sum_{i=1}^{N_L} \mu_i \frac{N_0}{G_i} \left(2^{\frac{\bar{R}_i}{W\mu_i}} - 1 \right) \end{aligned} \quad (6)$$

where all links have to be served in the available time T . The combined duration of all transmissions must be less or equal to T . But since it is clearly most efficient to use the entire available T , it holds that

$$\sum_{i=1}^{N_L} \mu_i = 1. \quad (7)$$

It is that allocation of transmission durations is power optimal which minimizes (6).

III. METHODOLOGY AND RESULTS

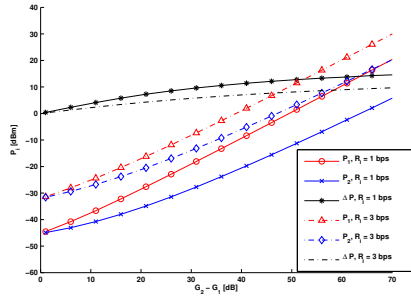
A. On transmission power and transmission time

For a particular target rate on a link, a lower transmit time results in a higher transmission power and vice versa. Since transmission power is capped by P_{max} , this in turn lower bounds μ_i . If the target rates cannot be fulfilled without violation of P_{max} on all links (or if $\sum_i \mu_i > 1$), then the system is overloaded. Thus, the optimal allocation is a trade-off of transmission times between links which depend on the individual channel gains and target rates. First, we inspect the behavior of transmit powers and times as they depend on the channel gains.

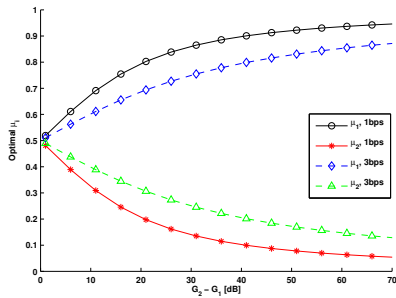
For illustration, the individual optimal transmission powers and times of a system with two links and equal target rates are plotted in Figure 2. In this setting, the channel gain on link 2 is constant, while the it is swept over a range of 100 dB for link 1. Plotted in Figure 2(a) are the system optimal transmit powers. The x-axis contains the ratio of the two channel gains (which equals the difference in dB) to emphasize

the relative gap in channel gains. When channel gains are equal the power optimal allocation is clearly equal powers. As link 2 degrades, more power should be allocated to it. The important finding is here that for optimal power allocation, transmit powers have to be increased on both links as channel gain difference increases. It is not possible to minimize BS power consumption by considering individual links. Rather, all links have to be considered simultaneously.

Figure 2(b) shows the corresponding transmission times. It can be seen that transmission times should compensate for the change in transmit powers on each link. The target rates are fulfilled by appropriate selection of transmission times μ_i . For example, in a cellular system, this occurs in the downlink case between a cell edge user with low channel gain on link 1 and a center user with high channel gain on link 2. The difference in transmission power in Figure 2(a) shows that at higher target rates there is less flexibility in the selection of μ_i resulting in a smaller difference of transmission powers.



(a) Optimal transmission powers in a 2-link system as a function of channel gain difference. The difference in powers ΔP in dB is equivalent to the ratio in linear terms.



(b) Optimal transmission times in a 2-link system as a function of channel gain difference.

Fig. 2. Illustration of optimal transmission powers and times.

For more than two users, the same holds true. If one link in the system degrades, all links should increase their power. To uphold the rate requirements, the unchanged links will reduce transmission time and the bad link receives more transmission time.

B. Optimal strategy on equal channels is equal powers

When considering the system which depends on the given channel gains and target rates, finding the optimal transmission times μ_i is an analytically difficult problem to which to our knowledge no closed form solution exists [9]. However, it turns out that there exists a subproblem, which can be analytically solved. If users are grouped such that we only consider links with equal channel gains at a time, the problem changes in nature. For $G_i = G$,

$$\bar{P}_{\text{SYS}}(\bar{R}_i) = \frac{N_0}{G} \sum_{i=1}^{N_L} \mu_i \left(2^{\frac{\bar{R}_i}{W\mu_i}} - 1 \right). \quad (8)$$

This new problem can be solved analytically using a Lagrange multiplier.

Theorem 1.

$$\mu_i = \frac{R_i}{\sum_{i=1}^{N_L} R_i} \quad (9)$$

is the minimizer of

$$\bar{P}_{\text{SYS}}(\bar{R}_i) = \sum_{i=1}^{N_L} \mu_i \left(2^{\frac{\bar{R}_i}{W\mu_i}} - 1 \right). \quad (10)$$

Proof: The partial derivatives of (10) and (7) are

$$\frac{\partial \bar{P}_{\text{SYS}}(\bar{R}_i)}{\partial \mu_i} = 2^{\frac{R_i}{W\mu_i}} - 1 - \left(\frac{R_i}{W\mu_i} \right) \log(2) 2^{\frac{R_i}{W\mu_i}}$$

and unity, respectively; log refers to the natural logarithm.

Therefore, the gradients of the Lagrange function for any i are

$$\nabla_{\mu_i, \lambda} \Lambda(\mu_i, \lambda) = \frac{\partial \bar{P}_{\text{SYS}}(\bar{R}_i)}{\partial \mu_i} + \lambda. \quad (11)$$

The partial derivatives are independent of each other and only depend on λ . Thus, they have to be equal.

Functions of the type $x \mapsto 2^x (1 - \log(2)x)$ are monotone on $x \geq 0$, hence there cannot be two different arguments for one i yielding the same function value. That is, one must have $\mu_i = cR_i$ for a given c . Since $\sum \mu_i = 1$, $1/c$ is the sum of R_i over i and the optimum is given by

$$\mu_i = \frac{R_i}{R_1 + \dots + R_N}. \quad (12)$$

This means that transmission times are directly proportional to the link target rates. The optimal transmission power on each link can be directly found from the sum of target rates using (5) and (12). Optimization of mobile user subsets can be achieved by assigning mobiles in groups according to their channel quality (or distance from the BS) or by appropriate scheduling in time of users with similar link quality. ■

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Bandwidth W	10 MHz
Thermal noise	-103 dBm
Static channel gain	-100 dB

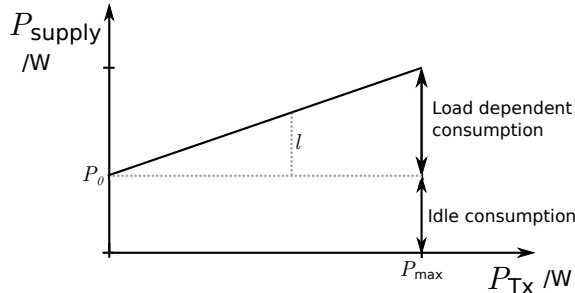


Fig. 3. Illustration of the power model with load dependent and idle consumption.

IV. POWER CONTROL STRATEGY IN THE POWER MODEL

The power model [10] is a detailed representative model of how such BS components as radio transceiver, baseband interface, power amplifier, AC-DC-converter, DC-DC-converter and cooling fans act together forming an overall power consumption behavior. It was found that although highly complex in detail, the combined consumption of today's BSs can be represented by an affine function. There is a static consumption that applies independent of load, P_0 , which we refer to as the idle power consumption. In addition, there is consumption which depends on the power delivered to the antenna. The rate of increase is represented by a load factor l . The model is analytically represented by

$$P_{\text{supply}} = P_0 + lP_{\text{Tx}}, \quad (13)$$

with $P_{\text{Tx}} \leq P_{\text{max}}$ and is illustrated in Figure 3.

As a metric for the share of load-dependent consumption of the overall consumption, we define the load dependence of a BS as

$$\eta_{\text{ld}} = \frac{lP_{\text{max}}}{P_0 + lP_{\text{max}}}. \quad (14)$$

It is important to consider power control and transmission power in relationship to the consumption of the entire BS. The power model provides this by mapping transmission power to supply power. Instead of the transmission power, we are now able to minimize the supply power consumption. Addition of the power model to the power control strategy changes (6) in the following fashion:

$$\begin{aligned} \bar{P}_{\text{supply}}(\bar{R}_i) &= \sum_{i=1}^{N_L} \mu_i (P_0 + lP_{\text{Tx},i}(\bar{R}_i)) \\ &= \sum_{i=1}^{N_L} \mu_i \left(P_0 + l \frac{N_0}{G_i} \left(2^{\frac{\bar{R}_i}{W\mu_i}} - 1 \right) \right) \end{aligned} \quad (15)$$

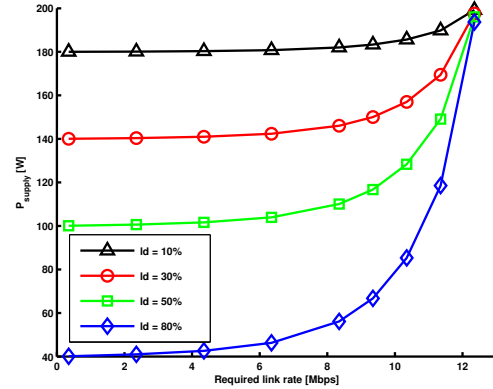


Fig. 4. Comparison of power models on the effectiveness of power control.

TABLE II
TYPICAL LOAD DEPENDENCE FACTORS BY BS TYPE [10]

BS type	P_{max}	Load dependence factor η_{ld}
Macro	46 dBm	50%
Micro	38 dBm	30%
Pico	21 dBm	14%
Femto	17 dBm	10%

Although not intuitive, it turns out that addition of the power model does not affect the power control strategy. Extending the Lagrange multiplier analysis to the power model inclusive problem, the partial derivative of the cost function now becomes

$$\frac{\partial \bar{P}_{\text{SYS}}(\bar{R}_i)}{\partial \mu_i} = P_0 + l \frac{N_0}{G_i} \left(2^{\frac{\bar{R}_i}{W\mu_i}} \left(1 - \left(\frac{\bar{R}_i}{W\mu_i} \right) \log(2) \right) - 1 \right). \quad (16)$$

In line with the proof for Theorem 1, it holds now that still all partial derivatives have to be equal. Hence, P_0 and l have no influence on the solution. This bears the important consequence that the power control strategy is independent of the underlying power model, and thus the underlying BS. It is therefore optimal for all BSs.

However, although the strategy is independent of the hardware, the benefits of power control strongly depend on hardware. See Figure 4 for an illustration in a cell with ten users. Only when the load dependence factor η_{ld} of a BS is high enough, can power control provide relevant savings. The effectiveness of power control in future BSs therefore strongly depends on hardware developments. For comparison, we consider typical load dependence factors of different BS types in Table II. Due to the low range and therefore low transmit powers of smaller BS types, they have load dependence factors much smaller than 50%. Therefore, power control has limited applicability in smaller BS types than in macro stations.

V. CONCLUSION

In this paper, supply power optimal transmission power control is studied in the multi-user downlink setting where the Quality of Service (QoS) criterion is a target link rate. The minimization variable is the average supply power of a BS as a function of the required rates and channel gains. The system can be analytically represented. The first important finding is that in power optimal allocation, the system has to be considered as a whole. If a channel in the system degrades, the transmission powers for all links have to be increased while transmission durations for all links except the degrading link are decreased. While this convex problem does not have a closed-form solution, it can be simplified if only users with equal channel gains are considered at one time. It has been shown that in that case the optimal transmission time is proportional to the required link rate over the sum of link rates and the optimal transmission powers are equal for all links. To account for realistic hardware consumption, a power model is considered which represents static and load dependent consumption of a BS. The addition of the power model allows estimation of absolute consumption values rather than transmission power. Inclusion of the power model into the power allocation problem reveals that the optimal strategy holds independent of the chosen power models. However, only if the power consumption of a BS is largely load-dependent, does power control result in significant savings. Otherwise, the idle consumption is several magnitudes larger than the effect of transmission power. Since BS types other than macro stations have much lower transmission powers and thus lower load dependence factors, power control is most effective in macro stations.

ACKNOWLEDGEMENTS

This work has received funding from the European Community's 7th Framework Programme [FP7/2007-2013. EARTH,

Energy Aware Radio and neTwork tecHnologies] under grant agreement n° 247733.

The authors gratefully acknowledge the invaluable insights and visions received from partners of the EARTH consortium.

REFERENCES

- [1] A. Fehske, J. Malmudin, G. Biczók, and G. Fettweis, "The Global Carbon Footprint of Mobile Communications - The Ecological and Economic Perspective," *IEEE Communications Magazine*, 2010. IEEE Communications Magazine.
- [2] G. Auer, I. Gódor, L. Hévízi, M. A. Imran, J. Malmudin, P. Fasekas, G. Biczók, H. Holtkamp, D. Zeller, O. Blume, and R. Tafazolli, "Enablers for Energy Efficient Wireless Networks," in *Proc. of the Vehicular Technology Conference (VTC)*, 2010.
- [3] D. Ferling, T. Bohn, D. Zeller, P. Frenger, I. Gódor, Y. Jading, and W. Tomaselli, "Energy Efficiency Approaches for Radio Nodes," in *Future Network & Mobile Summit 2010*, 2010.
- [4] A. J. Goldsmith, "The Capacity of Downlink Fading Channels with Variable Rate and Power," *IEEE Transactions on Vehicular Technology*, vol. 46, pp. 569–580, 1997.
- [5] S. Sinanović, N. Serafimovski, H. Haas, and G. Auer, "Maximising the System Spectral Efficiency in a Decentralised 2-link Wireless Network," *Eurasip Journal on Wireless Communications and Networking*, vol. 2008, p. 13, 2008. doi:10.1155/2008/867959.
- [6] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1747–1758, Oct. 1999.
- [7] H. Al-Shatri and T. Weber, "Fair Power Allocation for Sum-Rate Maximization in Multiuser OFDMA," in *Proceedings of the International ITG Workshop on Smart Antennas 2010*, 2010.
- [8] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-Constrained Modulation Optimization," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 2349–2360, 2005.
- [9] H. Holtkamp, G. Auer, and H. Haas, "On Minimizing Base Station Power Consumption," in *Proceedings of the IEEE 74th Vehicular Technology Conference*, 2011.
- [10] L. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. Van Der Perre, "Challenges and Enabling Technologies for Energy Aware Mobile Radio Networks," *Communications Magazine*, *IEEE*, vol. 48, no. 11, pp. 66–72, 2010.



Future Network and Mobile Summit 2011 Conference Proceedings
 Paul Cunningham and Miriam Cunningham (Eds)
 IIMC International Information Management Corporation, 2011
 ISBN: 978-1-905824-25-0

Fundamental Limits of Energy-Efficient Resource Sharing, Power Control and Discontinuous Transmission

Hauke HOLTkamp¹, Gunther AUER¹

¹DOCOMO Euro-Labs, Landsbergerstr. 312, Munich 80687, Germany

Email: {lastname}@docomolab-euro.com

Abstract: The achievable gains via power-optimal scheduling are investigated. Under the QoS constraint of a guaranteed link rate, the overall power consumed by a cellular Base Station (BS) is minimized. Available alternatives for the minimization of transmit power consumption are presented. The transmit power is derived for the two-user downlink situation. The analysis is extended to incorporate a BS power model (which maps transmit power to supply power consumption) and the use of Discontinuous Transmission (DTX) in a BS. Overall potential gains are evaluated by comparison of a conventional State-Of-The-Art (SOTA) BS with one that employs DTX exclusively, a power control scheme and an optimal combined DTX and power control scheme. Fundamental limits of the achievable savings are found to be at 5.5 dB under low load and 2 dB under high load when comparing the SOTA consumption with optimal allocation under the chosen power model.

Keywords: Green Radio, Energy Efficiency, Wireless Networks, Cellular Networks, LTE, EARTH project, Scheduling, RRM, power control

1. Introduction

Today, the battle against global warming – pushed by recent initiatives – requires energy optimization also on the BS side – in addition to the battery-limited User Equipment (UE) side. A decrease in total CO₂ emission by 20% by 2020 is requested [1]. In a thriving sector like the Information and Communication Technologies (ICT), which is predicted to grow at 24% percent *p.a.*, this is a challenging goal. Globally, the ICT sector causes 2% of CO₂ in the year 2007. Mobile communication accounts for a fraction of 0.3% of global CO₂. Achieving a 20% decrease in emission by 2020 while growing in volume by 200% necessitates a decrease in total energy consumption of today's mobile networks by 74%. In addition to the reduction of CO₂ emission, a lowering of Operating Expenses (OPEX) provides a strong incentive for network operators to reduce BS energy consumption.

Over the evolution of cellular networks, the focus of research and development has been on maximizing throughput, spectral efficiency, reliability and Quality of Service (QoS). It has always been of utmost importance to provide the mobile user with a seamless, fast and dependable connection. While these aspects have not lost relevance, a new consideration is coming into play – the energy efficiency of network operation. Due to the battery-powered nature of the mobile device, care has always been taken to minimize processing load on the UE by passing it asymmetrically to the BS which is usually mains powered. Examples for this design in Long Term Evolution (LTE) include the use of OFDMA with a high Peak-to-Average Power Ratio (PAPR) for downlink transmission (requiring more complex amplifiers) and Single-carrier FDMA

(SC-FDMA) on the uplink (placing complexity on the receiver side) and the BS always-on protocol design. In fact, operation of the mobile device over its life-time is so efficient today that it causes negligible power consumption compared to its manufacturing [2]. While this asymmetry has achieved long battery lives on the mobile equipment side, little attention has been paid to the the overall network power consumption. Studies have shown that the power consumption of the infrastructure network per subscriber is between 10 and 100 times higher than the power consumption of the mobile device [3, 4]. In addition to this power consumption asymmetry, high spectrum efficiency is only required during peak traffic hours while during off-peak hours large parts of the available resources are unused. Fully exploiting these idle resources (*i.e.*, transmit power, time and/or frequency) can greatly increase system efficiency [5].

While the routine refinement of components has regularly been able to reduce energy consumption over multiple generations of devices [1], little work has been done on the energy efficiency of radio transmission. Often, these are just side mentions in throughput oriented research [6, 7]. Furthermore, minimization of transmit power does not always lead to a minimization of supply (mains) power. The Energy Aware Radio and neTwork tecHnologies (EARTH) project is based upon the above and has established the following mission: Reduce power consumption of the world-wide mobile networks by 50% by improving the radio transceiver BS, hereby focusing on LTE and LTE-Advanced as the upcoming mobile network. This work is concerned with effects of DTX and power control on power savings while upholding a required link rate. Taking into account the supply power in addition to the transmit power, it is studied which approach uses the least energy.

In this paper, we present the results of a study summarizing the fundamental limits in energy-efficient scheduling. Section 2 provides the necessary background for this analysis. In Section 3 the problem of minimizing transmit power is formally described. This model is extended in Section 4 to incorporate overall power consumption. We present the results and conclude in Section 5.

2. Considerations for Energy-efficient Radio Resource Management

When inspecting the internal power dissipation of a typical macro BS, it turns out that around 65% of the power is lost in the power amplifier [8]. Power consumption of the power amplifier can be reduced in two ways. On the one hand, the efficiency of the component can be improved and thus the consumption reduced. On the other hand, the required transmit power can be reduced via appropriate scheduling. This study focuses on the latter. A more efficient and linear power amplifier will scale its consumption more directly with transmit power, thus both approaches support each other. BS transmit power is reduced via downlink Power Control (PC). By adjusting transmit power on each transmitted resource element individually, overall consumption is minimized.

In general, transmit power can be reduced whenever the channel conditions, in terms of the Signal-to-Interference-and-Noise-Ratio (SINR), are better than required. In addition, the existence of idle resources plays an important role. Current cellular systems and BS are designed to provide maximum throughput and deployed such that the user demand can be fulfilled during peak hours. However, for large parts of a day, this capacity is overachieving. As a result, resource elements – and thus parts of the

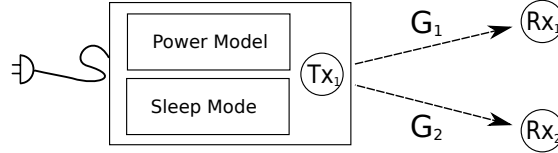


Figure 1: The employed network model, where G_i is the path gain on link i

available frequency band – are unused. So instead of transmitting at high power on few resource elements, it is advantageous from an energy perspective to transmit on all available resource elements with lower transmit power. When there is more than one user requesting data, the resource elements need to be scheduled.

3. Minimizing transmit power

In order to describe the transmit power consumption of a BS under varying loads, two user downlink is selected as the appropriate model, as shown in Fig. 1. We proceed to derive and minimize the system transmit power consumption. According to Shannon [9] the channel capacity of an ideal bandlimited additive white Gaussian noise channel is described by

$$C = W \log_2(1 + \gamma), \quad (1)$$

where C is the channel capacity in bps, W is the channel bandwidth and γ is the SINR $\gamma = \frac{GP}{N}$ with G the channel gain, P the transmit power and N the thermal noise.

While the bandwidth is proportional to the capacity C in (1), the transmit power has a logarithmic relationship to C . As a consequence, to increase data rate is much more expensive in terms of power than in terms of bandwidth. In other words, if there is a choice between a) leaving idle bands and transmitting at higher power and b) using all available bands and transmitting at the lowest required power, then option b) will always consume less overall transmit power.

The idea of a trade-off of resources against transmit power receives an additional dimension in the two-user case. There is now the option to allocate resources efficiently over two users, such that overall power consumption is minimal. To account for the spread use of the resource, we introduce the weighting factor μ , representing share of the resource. In analytical terms, this is represented as follows:

$$C_{\text{SYS}} = W [\mu \log_2(1 + \gamma_1) + (1 - \mu) \log_2(1 + \gamma_2)], \quad (2)$$

where $\gamma_1 = \frac{G_1 P_1}{N}$, $\gamma_2 = \frac{G_2 P_2}{N}$ on link i . Weighting factor $\mu = 1$ signifies the assignment of all resources to link 1, none to link 2 and vice versa for $\mu = 0$; $0 \leq \mu \leq 1$. Allocation (weighting) of resources is referred to as Resource Sharing (RS). For a more general analysis, we consider the spectral efficiency ς instead of the link capacity C , where $\varsigma = \frac{C}{W}$.

It is our goal to identify the set of parameters that causes least transmit power consumption for a guaranteed link spectral efficiency ς_{min} that is assumed to be equal on all links as the QoS constraint:

$$\min (P_{\text{SYS}}) : \varsigma_n \geq \varsigma_{\text{min}}. \quad (3)$$

Parameter	Minimum	Maximum	Unit
ς	0.1	6	bps/Hz
N	-100	-100	dBm
G_i	-150	-50	dB
P_i	-50	50	dBm

Table 1: Evaluated parameter ranges

The overall transmit power consumed by the system is

$$P_{Tx} = \mu P_1 + (1 - \mu) P_2. \quad (4)$$

The individual required link spectral efficiencies of the system on link i are

$$\varsigma_i = \mu \log_2 (1 + \gamma_i). \quad (5)$$

Thus

$$P_1 = \left(2^{\frac{\varsigma_1}{\mu}} - 1\right) \frac{N}{G_1} \quad \text{and} \quad P_2 = \left(2^{\frac{\varsigma_2}{1-\mu}} - 1\right) \frac{N}{G_2}. \quad (6)$$

Due to the monotonically increasing shape of the log-function, the minimum power consumption on a static link occurs for the lowest possible spectral efficiency ς_{\min} . Therefore, the system operates optimally in terms of power consumption when

$$\varsigma_1 = \varsigma_2 = \varsigma_{\min}. \quad (7)$$

The combination of (4) and (6) yields

$$P_{Tx} = \mu \frac{N}{G_1} \left(2^{\frac{\varsigma_{\min}}{\mu}} - 1\right) + (1 - \mu) \frac{N}{G_2} \left(2^{\frac{\varsigma_{\min}}{1-\mu}} - 1\right), \quad (8)$$

which describes the system transmit power consumption for a short enough time during which G_1 and G_2 remain unchanged, *i.e.*, less than the coherence time of the channel.

For an initial analysis, Table 1 lists the parameter range for the spectral efficiency ς , thermal noise N , channel gains G_i and transmit powers P_i , $i=\{1, 2\}$, which corresponds to today's terrestrial cellular networks, such as LTE or WiMAX. The noise power $N = -100$ dBm represents thermal noise for $B=10$ MHz bandwidth with $N = BN_0 = BkT$, $k=1.38e-23$ dBm/Hz, $T=290$ K.

For each combination of parameters, there is one optimal weighting factor μ that minimizes power consumption. Figure 2 shows the optimal weighting factor for selected spectral efficiencies. For high required spectral efficiencies, μ approaches linearity.

4. From Transmit Power to Supply Power

While the findings from the previous section establish a necessary foundation for energy-efficient transmission, there is one more step necessary to generate a complete picture. So far, only transmit power has been considered. By means of a power model, which translates transmit power into supply power ("power at the 230V plug") of a BS, it is possible to account for component behavior. While the model has recently been described to behave linearly [5], it still adds one important dimension to the problem: A BS consumes significant power even when it is not transmitting. For transmit power

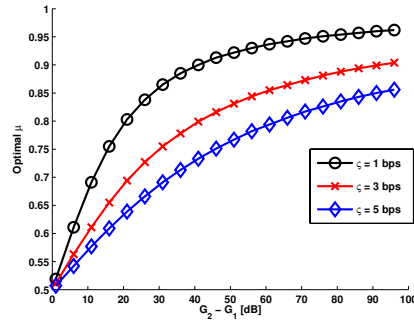


Figure 2: Optimal weighting factor μ under selected spectral efficiencies.

optimization this bears the significant consequence that savings at low transmit powers have no significant effect on the supply power side. In low power/low load operation, the standby consumption causes the bulk of overall consumption. Instead an intermediate step can be taken via sleep mode. This means that the BS enters a micro sleep as soon as it is not needed. There exist several levels or “depths“ of sleep which take different amounts of time to enter and leave. For this study, we assume that there exists exactly one sleep mode that is available instantly. In the following, the problem presented in the previous section is extended by two additional steps. First, transmit power is translated into supply power via the linear power model. Second, there exists the alternative for the BS to be in sleep mode during which it consumes a lower-than-standby amount of power. Depending on the share of time it spends in sleep mode, the standby consumption is reduced. However, as a tradeoff, the rate has to be increased during awake-time. This in turn causes higher necessary transmit power.

Formally, the power model effect is added by

$$P_{PM} = P_0 + mP_{Tx} \quad (9)$$

where P_0 is the standby power consumption and m is the slope of the load dependent power consumption.

To account for the effect of DTX, the activity factor t is introduced which indicates the share of time spent in active mode (rather than sleep mode). DTX power is modeled as

$$P_{supply} = (1 - t)P_s + tP_{PM}, \quad (10)$$

where t is the share of time spent in active (awake) mode and P_s is the BS power consumption in sleep mode.

The trade-off of DTX is an increased rate during active time. This is included in the model via $\zeta_{DTX} = \frac{\zeta}{t}$.

As has been shown so far, minimum transmit power leads to minimum power consumption under the linear power model. However, additional savings are possible by employing DTX, which completely changes the transmit-power-only analysis for low data rates.

Parameter	Value
Carrier frequency	2 GHz
Cell radius	250 m
Pathloss model	3GPP UMa, NLOS, shadowing
Shadowing standard deviation	8 dB
Iterations	10,000
Bandwidth	10 MHz
Noise power	-100 dBm
P_0 , BS power consumption at zero load	233 W
m , Increase per Watt PTx (model slope)	5
P_s , BS power consumption in sleep mode	50 W
P_{max}	46 dBm

Table 2: Simulation Parameters

5. Results and Discussion

The described two-user downlink system has been evaluated in simulation using the system parameters listed in Table 2. Two users are dropped uniformly on a disk, generating a distribution of channel gains. Results of the achievable data rate are plotted over the bandwidth. Power consumption is plotted for the power-optimal selection of weighting factor μ and awake time share t in Figure 3.

While the RS-PC-DTX scheme provides power-optimal resource allocation, it is necessary to define alternative allocation schemes for reference and estimation of achievable gains. For comparison, the following schemes are included in the analysis: A constant transmit power of 46 dBm which represents the upper limit of an LTE macro BS; a weakly load dependent SOTA BS as the reference; RS-PC without DTX; a scheme without power control, but with DTX; and the optimal RS-PC-DTX scheme. The 46 dBm transmit power reference provides an absolute upper limit, which represents the supply power consumption of a BS that is always-on and transmitting at maximum power. However, this is not a realistic assumption since even BS today (SOTA) that are not manufactured for energy efficiency have some load dependent behavior. As an estimate, we assume that such a SOTA BS consumes only half of its maximum consumption (*i.e.*, -3 dB) when it is not loaded with a linear raise in consumption up to full load [5]. The first allocation scheme is one where allocation of resource blocks and PC are enabled, but the BS does not have the capability of sleep modes. This is informative to estimate gains in addition to DTX. In contrast, it is also instructive to consider a scheme which does not have PC available, but can go into sleep mode, which is labelled 'ON/OFF DTX'. Ultimately, the optimal RS-PC-DTX allocation scheme utilizes all three available options.

Figure 3 summarizes our findings. The overall system power consumption P_{supply} is plotted vs. the required link rate. It is important to note that only link rates up to around 6×10^7 bps can be provided without outage. Thus, the plot values for link rates higher than 6×10^7 bps are provided as a trend. For the same reason, the SOTA BS consumption is only plotted over this range.

Most relevant are the achievable gains of the optimal scheme (RS-PC-DTX) compared to the SOTA behavior. Here, it can be seen that the largest gains are expected in low load (5.5 dB) and slowly decrease (to 2 dB) as the required link rate approaches 6×10^7 bps. In percentages, these fundamental limits translate into a 73% savings po-

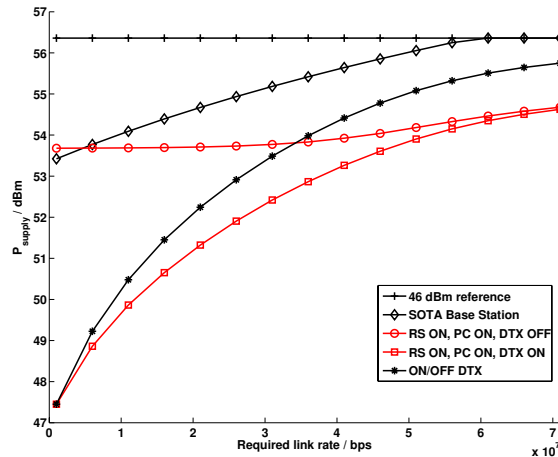


Figure 3: Comparison of supply power consumption as a function of link rate under varying schemes

tential in low load going towards a 27% saving in high load situations. In addition to these cornerstone numbers, further relevant behavior can be identified. Comparing 'ON/OFF DTX' with the optimal scheme shows that in low load PC has hardly an effect. This is due to standby power consumption outmatching the transmit power consumption by orders of magnitude in the power model. In other words, saving transmit power in a low load situation bears insignificant benefit. Only as load increases (higher required rates) to the curves part, yielding a 1 dB benefit of the optimal scheme over 'ON/OFF DTX'. This benefit comes at the cost of allowing downlink power control which stretches today's standards specification and technical capabilities. The merging of the two RS-PC curves illustrates that — as rates grow — DTX becomes less feasible. In this operating region, situations occur in which it is more efficient to keep the BS active (*i.e.*, no sleep mode) and decrease transmit power than to go into sleep mode.

6. Conclusions

This study reveals that DTX is needed in future energy-efficient BS, if significant gains are expected. Power control alone can only reduce power consumption by around 20% in high load situations. When optimizing for low load, DTX by itself delivers similar gains as a scheme which employs RS, PC and DTX. However, only the combination of all three techniques is overall power-optimal.

When considering today's hardware and standards definitions, DTX is the appropriate energy saving scheme in range. However, consideration of the control signalling in LTE is part of future work in this field and is expected to reduce the applicability of DTX, because during necessary control signalling the BS is not allowed to enter sleep mode.

Acknowledgements

The authors gratefully acknowledge the contribution of the EARTH consortium to the technical definition of the project.

The work leading to this paper has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013. EARTH, Energy Aware Radio and neTwork tecHnologies] under grant agreement n° 247733.

References

- [1] O. Blume, D. Zeller, and U. Barth, "Approaches to Energy Efficient Wireless Access Networks," in *4th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pp. 1–5, 3-5 2010.
- [2] G. Auer, I. Gódor, L. Hévízi, M. A. Imran, J. Malmudin, P. FASEKAS, G. Biczók, H. Holtkamp, D. Zeller, O. Blume, and R. Tafazolli, "Enablers for Energy Efficient Wireless Networks," in *Proc. of the Vehicular Technology Conference (VTC)*, 2010.
- [3] M. Etoh, T. Ohya, and Y. Nakayama, "Energy Consumption Issues on Mobile Network Systems," in *Proc. of the International Symposium on Applications and the Internet (SAINT)*, (Turku, Finland), pp. 365–368, IEEE, July/Aug. 2008.
- [4] A. Fehske, J. Malmudin, G. Biczok, and G. Fettweis, "The Global Carbon Footprint of Mobile Communications - The Ecological and Economic Perspective," 2010. IEEE Communications Magazine.
- [5] L. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. Van Der Perre, "Challenges and Enabling Technologies for Energy Aware Mobile Radio Networks," *Communications Magazine, IEEE*, vol. 48, no. 11, pp. 66–72, 2010.
- [6] S. Sinanović, N. Serafimovski, H. Haas, and G. Auer, "Maximising the System Spectral Efficiency in a Decentralised 2-link Wireless Network," *Eurasip Journal on Wireless Communications and Networking*, vol. 2008, p. 13, 2008. doi:10.1155/2008/867959.
- [7] A. Gjendemsjø, D. Gesbert, G. E. Øien, and S. G. Kiani, "Optimal Power Allocation and Scheduling for Two-Cell Capacity Maximization," in *Proc. of the 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, (Boston, USA), pp. 1–6, Apr. 3–7 2006.
- [8] D. Ferling, T. Bohn, D. Zeller, P. Frenger, I. Gódor, Y. Jading, and W. Tomaselli, "Energy Efficiency Approaches for Radio Nodes," in *Future Network & Mobile Summit 2010*, 2010.
- [9] C. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 & 623–656, July & Oct. 1948.

On Minimizing Base Station Power Consumption

Hauke Holtkamp, Gunther Auer

DOCOMO Euro-Labs
D-80687 Munich, Germany
Email: {holtkamp, auer}@docomolab-euro.com

Harald Haas

Institute for Digital Communications
Joint Research Institute for Signal and Image Processing
The University of Edinburgh, EH9 3JL, Edinburgh, UK
E-mail: h.haas@ed.ac.uk

Abstract—We consider resource allocation over a wireless downlink where Base Station (BS) power consumption is minimized while upholding a set of required link rates. A Power and Resource Allocation Including Sleep (PRAIS) method is proposed that combines resource sharing, Power Control (PC), and Discontinuous Transmission (DTX), such that downlink power consumption is minimized. Unlike conventional approaches that aim at minimizing transmit power, in this work the BS mains supply power is chosen as the relevant metric. Based on a linear power model, which maps a certain transmit power to the necessary mains supply power, we quantify the fundamental limits of PRAIS in terms of achievable BS power savings. The fundamental limits are numerically evaluated on link level for four sets of BS power model parameters representative of envisaged future hardware developments. For BSs installed around 2014, PRAIS provides 63% to 34% energy savings over conventional resource allocation schemes, depending on the rate target per link.

I. INTRODUCTION

Recent surveys on the energy consumption of cellular network components throughout the whole life cycle, including Base Station (BS), mobile terminals and core network, reveal that around 80% of the electricity bill of a mobile network operator are generated at BS sites [1], [2]. This highlights the need to improve the energy efficiency at the BS.

Moreover, analyzing the traffic data of mobile operators suggests that only during few hours per day BSs are running under full load — for which they were designed — while operating in low load for the rest of the time [2], [3]. This creates potential for BS energy savings by tailoring the resource and power allocation for low load situations. However, energy savings should be accomplished without affecting the perceived Quality of Service (QoS) of the user.

In the past, energy efficiency research on wireless transmitters has focused on the minimization of transmit power. The EARTH¹ power model establishes a relation between a given transmit power and the necessary input supply power of a BS [4]. Two fundamental characteristics of a BS are reported: first, the power model is closely approximated by a linear function; and second, the power consumption in idle mode (when no data is transmitted) may be significantly reduced with respect to the active transmission mode, since some hardware components and circuits may be switched-off. This

¹EU funded research project EARTH (Energy Aware Radio and neTwork technologies), FP7-ICT-2009-4-247733-EARTH, Jan. 2010 to June 2012. <https://www.ict-earth.eu>

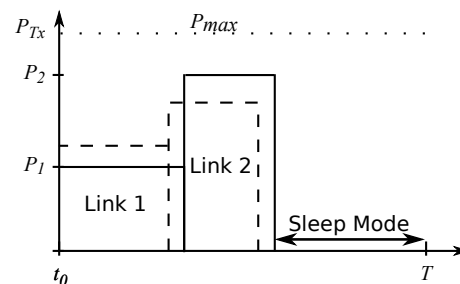


Fig. 1. Illustration of PRAIS for two links: transmit power, resource share and sleep mode are allocated optimally in order to serve the requested rate at minimal overall power consumption.

linear power model allows to minimize the *overall* BS supply power consumption, rather than transmit power only.

In this study, the EARTH power model is utilized to extract fundamental limits on efficient energy resource allocation. The proposed Power and Resource Allocation Including Sleep (PRAIS) scheme combines three techniques: Power Control (PC), appropriate resource sharing between multiple users by means of Time Division Multiple Access (TDMA) and Discontinuous Transmission (DTX). See Figure 1 for an illustration on two links. DTX refers to the ability to put the BS into a sleep mode, which has lower consumption than the active or idle state. To date, DTX or PC have only been considered individually [5], [6], thus missing the combined gains. The contribution of this work is to optimize PC, TDMA and DTX jointly, such that the BS supply power is minimized. For some representative power models that reflect anticipated developments in BS hardware, numerical results for the expected power consumption of future BSs are provided.

The remainder of this paper is structured as follows. Section II introduces the EARTH power consumption model. Section III derives the PRAIS scheme and the resulting optimization problem and states the scenario parameters. Section IV presents the quantitative results. The paper is concluded in Section V.

II. POWER MODEL

The basis of our analysis is the BS power model. A typical Long Term Evolution (LTE) BS consists of such components

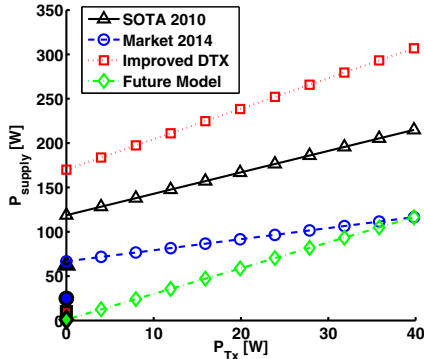


Fig. 2. Load behavior of the selected power models. Sleep mode consumption P_s is marked on the y-axis by the respective marker.

TABLE I
POWER MODEL PARAMETERS

Power Model	P_0 /W	m	P_s /W
SOTA 2010	119	2.4	63
Market 2014	67	1.25	25
Improved DTX	170	3.4	25
Future Model	1	2.9	1

as radio transceiver, baseband interface, power amplifier, AC-DC-converter, DC-DC-converter and cooling fans, each possibly manifold depending on the number of sectors and antennas employed. Each of these components has an individual power consumption that may either be constant or load-dependent. It was found in [4] that the power consumption of such a State-Of-The-Art (SotA) BS can be approximated by a linear function. This reflects the fact that some components have constant consumption independent of the load and the sum of load-dependent components creates an affine function, resulting in a straight line plot with a y-axis offset. Analytically, the linear power model for the instantaneous power consumption as a function of the transmit power P_{Tx} can be written as

$$P_{\text{supply}} = \begin{cases} P_0 + mP_{Tx} & \text{if } 0 < P_{Tx} \leq P_{\text{max}} \\ P_s & \text{if } P_{Tx} = 0. \end{cases} \quad (1)$$

where m denotes the slope of the trajectory that quantifies the load dependence, P_{max} is the maximum transmit power, P_0 and P_s account for the stand-by and sleep mode consumption of the BS, respectively. The power models used in this study are depicted in Figure 2 as functions of the transmit power P_{Tx} . Also shown in Figure 2 is the consumption in sleep mode P_s as markers on the y-axis for $P_{Tx}=0$.

Note that in this study, 'load' refers to the power density in an Orthogonal Frequency Division Multiple Access (OFDMA) system like LTE. High load signifies that many data Resource Elements (REs) in a particular time slot are used for transmission, each adding to the overall transmit power. Full load refers to all REs in use, leading to highest allowed transmit power.

Representative of current and future developments, we select four different power models; three from literature and a best-case assumption with the parameters listed in Table I. The first set of parameters represents the SotA as of 2010. Power consumption is reported to be $P_{\text{supply}} = 1292$ W and 712 W at full and zero load. Since the BS studied in [7] covers three sectors with two transmit antennas each, and the consumption of a BS scales approximately linearly with the number of transmit antennas and sectors, we scale down P_{supply} by a factor of six to represent a single antenna, one-sector BS.

The second model is a prediction of developments that will be available on the market in 2014 [4]. It is based on anticipated component improvements over four years. We label this the 2014 Market model.

Third, for the benefit of sleep modes, we follow an assumption where sleep modes will greatly improve, while standby component consumption cannot be reduced substantially in the coming years [5]. As this model assumes BS from some years before 2010, the load-dependent power consumption is higher than in the SotA 2010 model, thus particularly emphasizing DTX effects.

As a best-case example, the fourth power model assumes idealized components that scale perfectly with load. Power consumption is set to scale linearly with load with near-zero stand-by consumption. This model is not a prediction of technology advances, but provides theoretical limits. Assuming that operating efficiency at full load will not change after 2014, the future model has consumption equal to the 2014 Market model at full load/maximum transmit power.

Knowledge of the dependence of supply power on transmit power allows the targeted optimization via the PRAIS scheme.

III. PRAIS

When employing DTX individually to minimize power consumption, all required links are served at the maximum allowed transmission power until target rates are fulfilled, then the BS may go to sleep mode. Alternatively, when employing PC with TDMA individually for maximum efficiency, the transmission duration is stretched over the available time frame with the lowest power necessary to serve the required rates. When PC and DTX are combined, there is a trade-off. Between the two extremes of maximum transmit power with longest sleep mode and lowest transmit power with no sleep mode, there is a configuration with medium transmit power and medium sleep mode duration that consumes less overall power. This is exploited in the PRAIS scheme.

We proceed to derive the power consumption optimization problem for a multi-user single-cell allocation of transmit powers, sleep times and transmit durations contained within the PRAIS scheme. We define the normalized duration

$$\mu_i = \frac{t_i}{T} \quad (2)$$

where t_i is the time allocated to link i and T is the duration of the considered time frame in seconds.

$$\begin{aligned}
& \underset{\mu, \nu}{\text{minimize}} && \bar{P}_{\text{supply}}(\bar{R}_i) = \left[\sum_{i=1}^{N_L} \mu_i \left(P_0 + m \frac{N_i}{G_i} \left(2^{\frac{\bar{R}_i}{W_i \mu_i}} - 1 \right) \right) \right] + \nu P_s \\
& \text{subject to} && \sum_i \mu_i + \nu = 1, \quad \nu \geq 0, \quad \mu_i \geq 0 \quad \forall i \\
& && 0 \leq P_i = \frac{N_i}{G_i} \left(2^{\frac{\bar{R}_i}{W_i \mu_i}} - 1 \right) \leq P_{\max}
\end{aligned} \tag{9}$$

The normalized duration spent in sleep mode is

$$\nu = \frac{t_s}{T} \tag{3}$$

where t_s is the time spent in sleep mode and

$$\sum_{i=1}^{N_L} \mu_i + \nu = 1 \tag{4}$$

with N_L the number of links served.

Since we are interested in the fundamental limits we employ the Shannon bound to associate transmit powers with link rates. The rate on each link is

$$R_i(P_i) = W_i \log_2(1 + \gamma_i(P_i)), \tag{5}$$

where $\gamma_i(P_i) = \frac{G_i P_i}{N_i}$ denotes the Signal-to-Noise-Ratio (SNR) and W_i is the transmission bandwidth in Hz, with channel gain G_i , transmit power $P_i \leq P_{\max}$ in W. The noise power is defined by $N_i = W_i k \vartheta$ with Boltzmann constant k and operating temperature ϑ in Kelvin.

The average rate on link i over the time slot T is

$$\bar{R}_i = \mu_i R_i. \tag{6}$$

Thus, the average transmit power required to fulfill a target average rate \bar{R}_i is

$$\bar{P}_{\text{Tx},i}(\bar{R}_i) = \frac{N_i}{G_i} \left(2^{\frac{\bar{R}_i}{W_i \mu_i}} - 1 \right). \tag{7}$$

The overall power consumption caused when transmitting is the weighted average of the link supply power consumptions found by summation of (1) over all links:

$$\bar{P}_{\text{active}}(\bar{R}_i) = \sum_{i=1}^{N_L} \mu_i \left(P_0 + m \bar{P}_{\text{Tx},i}(\bar{R}_i) \right). \tag{8}$$

Combining with the consumption during sleep mode from (1) and (7) we generate the cost function for P_{supply} yielding the optimization problem (9), displayed at the top of this page. The constraints reflect the facts that normalized time has to be positive and their sum unity. Transmission powers are positive and bounded by a maximum transmit power. The cost function is a non-negative sum of functions that is convex within the constraint domain and is therefore still convex. See the appendix for a proof. It can be solved with appropriate software like the MATLAB optimization toolbox. The solution of (9) determines the vectors μ and ν which minimize the overall power consumption under target rates. Note that fixing $P_0 = P_s$ is equivalent to disabling DTX,

TABLE II
SIMULATION PARAMETERS

Parameter	Value
Carrier frequency	2 GHz
Cell radius	250 m
Pathloss model	3GPP UMa, NLOS, shadowing [8]
Shadowing standard deviation	8 dB
Iterations	10,000
Bandwidth W	10 MHz
Maximum transmission power P_{\max}	46 dBm
Operating temperature T	290 K

i.e. employing PC individually, whereas fixing $P_{\text{Tx}} = P_{\max}$ is equivalent to disabling PC.

The resource allocation problem (9) can be solved for any number of users. Without loss of generality, we select a ten user scenario for numerical evaluation.

IV. RESULTS

For the numerical analysis, we evaluate the PRAIS scheme in a Monte Carlo simulation under the parameters shown in Table II. Users are dropped uniformly onto a disk and the associated channel gains G_i are generated applying the 3GPP urban macro path-loss model including shadowing with a standard deviation of 8 dB. In addition to the individual DTX and PC allocation schemes and the PRAIS scheme, we present two references that serve as upper limits. First, the maximum transmission power as defined by the LTE standard provides the theoretical reference. Second, we set the reference against which gains are measured to be the power behavior of a BS as defined in the power models. Practically, this is a Frequency Division Multiple Access (FDMA) scheme where each user receives a share of the frequency band as well as the entire considered time slot. This can be interpreted as 'DTX in the frequency domain' where $P_s = P_0$. In essence, this is bandwidth adaptation and is our reference against which we assess the achieved savings.

The simulation results are shown in Figure IV where the per-user link rate is plotted against the average supply power consumption under the different schemes. Figures 3(a), 3(b), 3(c), 3(d) reflect the four chosen power models.

In Figure 3(a) we see that a Single-Input and Single-Output (transmission) (SISO) link in an LTE BS of 2010 consumes 120 W up to 220 W in bandwidth adaptation, which we consider the operation of the SotA. The first result is that the consumption curves of bandwidth adaptation and PC as well as DTX and PRAIS originate in the same values of

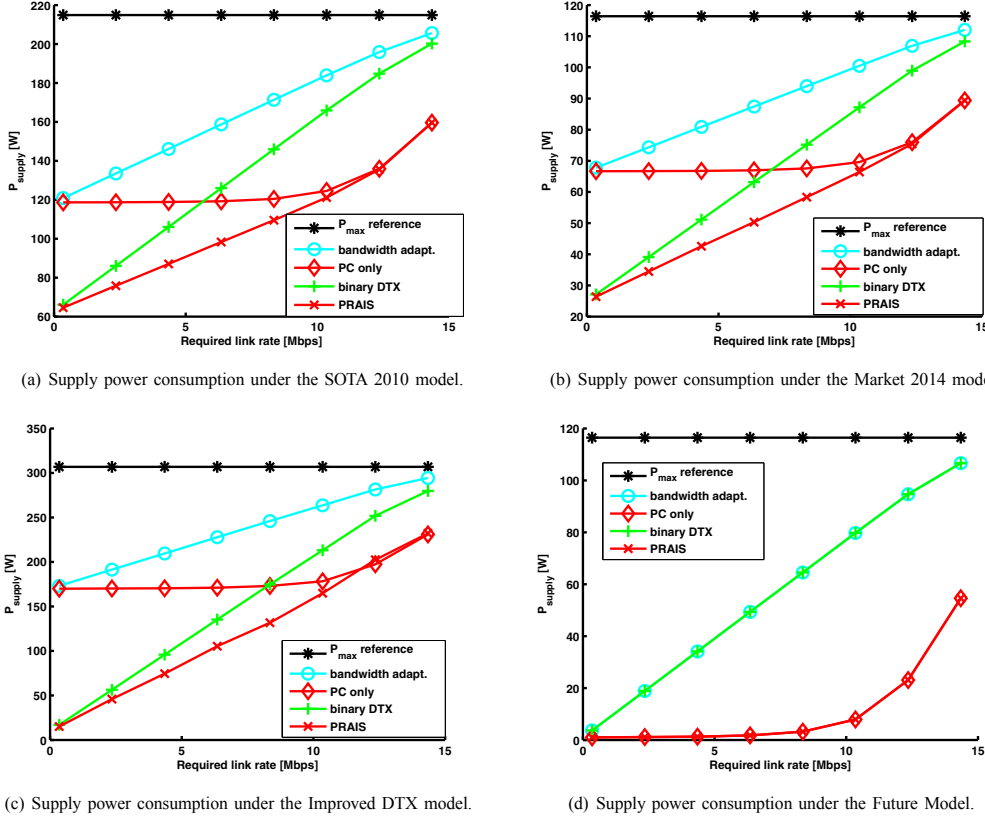


Fig. 3. Fundamental limits for power consumption in BSs.

P_{supply} . The higher one is P_0 at 119 W and the lower one P_s at 67 W.

The use of PC only allows to keep the overall consumption constant for a large set of low target rates. In this rate region, transmit powers are very low compared to standby consumption. Only when rates above 10 Mbps are requested is the transmission power high enough to make a noticeable difference compared to the standby consumption. This is reflected in the rising PC only curve above 10 Mbps. In contrast, the binary DTX scheme has much lower power consumption (up to -45%) at low rates. However, it rises much quicker than the PC only curve. There is a crossover point at 5.6 Mbps between DTX and PC. The combined PRAIS scheme has minimal consumption over all rates. At higher target rates, the PRAIS consumption curve joins the PC only curve. This reflects the fact, that it is not feasible to put the BS to sleep at high rates. In the 2010 model, the PRAIS scheme achieves savings between 45% and 31% over bandwidth adaptation.

Similar behavior can be found in the 2014 Market model in Figure 3(b). The model predicts that sleep mode consumption

will be less than half of the standby consumption at 25 W. This favors the application of DTX which is lower bounded by this value. The cross-over point between the individual DTX and PC schemes shifts to the right to around 7 Mbps. This means that DTX is more efficient than PC for a larger set of rates. We see this as a trend, especially since sleep modes in BSs are a new trend which will receive strong research efforts on the hardware side. In the 2014 model, PRAIS delivers savings between 61% and 34% over bandwidth adaptation.

A surprising result is found in Figure 3(c). Although this model has a strong bias towards DTX effects, there remains a cross-over point between DTX and PC after which the use of PC can still add an extra 15% savings on top of DTX. In the Improved DTX model, the PRAIS scheme offers between 90% savings at near-zero rates and 38% savings at 10 Mbps.

In the future linear model in Figure 3(d) the behavior of DTX and bandwidth adaptation, as well as PC and PRAIS, are identical, due to the fact that there is no gain of sleep modes over standby consumption. Bandwidth adaptation consumption is significantly lower than for all other power models. All gains are realized by PC which is strongly amplified by the linear

behavior, delivering gains of up to 75% over the bandwidth allocation scheme.

With regard to fundamental limits we find that highest consumption and thus the upper limit of the downlink power consumption is found to be a 2010 BS without energy efficient allocation at 119–220 W. When upscaling by three sectors and two transmit radio chains, we arrive at a consumption of 712 W to 1121 W. For the year 2014, we assume the default configuration of BSs to be equipped with four transmit antennas and full application of PRAIS. Thus, the consumption of 25–71 W must be upscaled by a factor of 12, resulting in an expected consumption of 300 W to 852 W per BS in 2014. This is an important finding, because it states that although BSs will contain more radio chains in the future, power consumption can still be expected to decrease (if energy efficiency measures are applied).

The attainable power savings of PC, DTX and PRAIS hardly depend on the number of users; power savings decrease slightly as the number of users increases (not shown in Fig. IV).

V. CONCLUSION

In this paper we have presented a comparison of TDMA, PC, DTX and bandwidth adaptation individually in the cellular downlink on the power supply metric. This leads to the optimal PRAIS scheme which exploits TDMA, PC and DTX in combination. It was found that the PRAIS scheme is convex which allows to find a global minimum of the supply power cost function. It can be seen in simulation that DTX contributes the highest savings compared to the other schemes. The PRAIS scheme can lower the power consumption of the 2014 Market BS by between 61% and 34%.

We find that from the two investigated individual schemes, DTX provides higher individual gains under the assumptions made in this study. But even under extreme assumptions that strongly benefit the use of DTX, the combined use of DTX and PC within the PRAIS scheme still delivers significant gains over the individual scheme. The two power saving techniques of PC and DTX are shown to have different applicability with DTX in the low rates and PC in high rates.

As a result of the overall predicted consumption development, we consider the adaptation of the number of active radio chains to the load an important future research topic.

Since practical limitations may reduce the applicability of PRAIS our future work will take more limitations into account, e.g. like LTE control signalling.

Ultimately, we establish 27 W at near-zero load and 68 W at 10 Mbps per 10 MHz bandwidth and 10 users as the lower fundamental limit of a BS in 2014 when employing resource and power allocation via PRAIS. This represents a saving over bandwidth allocation of 45% to 31% in 2010 and 61% to 34% in 2014, respectively.

APPENDIX

We show that (9) is monotonically decreasing and convex without sleep modes and that it is still convex (but no longer

monotonically decreasing) when sleep modes are considered. The first derivative of the cost function for link i is

$$f'_{0,i} = \frac{N_i}{G_i} \left[-1 + 2^{\frac{R_i}{W_i \mu_i}} \left(1 - \frac{1}{\mu_i} \frac{R_i}{W_i} \ln(2) \right) \right]. \quad (10)$$

It is seen that for $\mu_i \rightarrow 0$ the negative term goes towards infinity, thus $f'_{0,i} \rightarrow -\infty$. For $\mu_i \rightarrow 1$, the outcome depends on $\frac{R_i}{W_i}$. For $\frac{R_i}{W_i} \rightarrow +0$, $f'_{0,i} \rightarrow -0$. For $\frac{R_i}{W_i} \rightarrow +\infty$, $f'_{0,i} \rightarrow -\infty$. In all cases, the first derivative is negative, thus the cost function is monotonically decreasing in μ_i . The second derivative of $f_{0,i}$ is given by

$$f''_{0,i} = \frac{N_i}{G_i} \left(\frac{R_i}{W_i} \right)^2 \ln^2(2) \frac{2^{\frac{R_i}{W_i \mu_i}}}{\mu_i^3} \quad (11)$$

All variables in the second derivative are positive, thus $f''_{0,i} \geq 0$ within the parameter bounds. Therefore, each $f_{0,i}$ is convex within the bounds.

The non-negative sum preserves convexity. Thus, f_0 is convex.

When $P_s > 0$, the first derivative is no longer negative for all μ , thus the cost function is no longer monotonically decreasing in μ_i . However, the linear terms drops out in the second derivative, such that only the second derivative (11) remains, which has been shown to be convex.

ACKNOWLEDGEMENTS

This work has received funding from the European Community's 7th Framework Programme [FP7/2007-2013. EARTH, Energy Aware Radio and neTwork tecHnologies] under grant agreement n° 247733.

The authors gratefully acknowledge the invaluable insights and visions received from partners of the EARTH consortium.

REFERENCES

- [1] A. Fehske, J. Malmudin, G. Biczók, and G. Fettweis, "The Global Carbon Footprint of Mobile Communications - The Ecological and Economic Perspective," *IEEE Communications Magazine*, 2010. IEEE Communications Magazine.
- [2] G. Auer, I. Gódor, L. Hévízi, M. A. Imran, J. Malmudin, P. Fasekas, G. Biczók, H. Holtkamp, D. Zeller, O. Blume, and R. Tafazolli, "Enablers for Energy Efficient Wireless Networks," in *Proc. of the Vehicular Technology Conference (VTC)*, 2010.
- [3] Sandvine, "Mobile Internet Phenomena Report." http://www.sandvine.com/downloads/documents/2010_Global_Internet_Phenomena_Report.pdf, 2010.
- [4] L. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. Van Der Perre, "Challenges and Enabling Technologies for Energy Aware Mobile Radio Networks," *Communications Magazine, IEEE*, vol. 48, no. 11, pp. 66–72, 2010.
- [5] P. Frenger, P. Moberg, J. Malmudin, Y. Jading, and I. Gódor, "Reducing Energy Consumption in LTE with Cell DTX," in *Vehicular Technology Conference Proceedings, 2011. VTC 2011-Spring Budapest, 2011 IEEE 73rd*, 2011.
- [6] S. Sinanović, N. Serafimovski, H. Haas, and G. Auer, "Maximising the System Spectral Efficiency in a Decentralised 2-link Wireless Network," *Eurasip Journal on Wireless Communications and Networking*, vol. 2008, p. 13, 2008. doi:10.1155/2008/867959.
- [7] G. Auer, V. Giannini, I. Gódor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. J. Gonzalez, and C. Desset, "Cellular Energy Efficiency Evaluation Framework," in *Proceedings of the 2011 IEEE 73rd Vehicular Technology Conference (VTC)*, 2011.
- [8] 3GPP, "Further Advancements for E-UTRA Physical Layer Aspects (Release 9)." 3GPP TR 36.814 V0.4.1 (2009-02), Sept. 2009. Retrieved June 2, 2009 from www.3gpp.org/ftp/Specs/.

Flexible power modeling of LTE base stations

Claude Dessel¹, Björn Debaillie¹, Vito Giannini¹, Albrecht Fehske², Gunther Auer³, Hauke Holtkamp³, Wieslawa Wajda⁴, Dario Sabella⁵, Fred Richter², Manuel J. Gonzalez⁶, Henrik Klessig², István Gódor⁷, Magnus Olsson⁸, Muhammad Ali Imran⁹, Anton Ambrosy⁴, Oliver Blume⁴

¹Imec, Kapeldreef, 75, B-3001 Leuven, Belgium

²Technische Universität Dresden, 01062 Dresden, Germany

³DOCOMO Euro-labs, Landsberger Str. 312, 80687 Munich, Germany

⁴Alcatel-Lucent Deutschland AG, Lorenzstrasse 10, ZFZ/WA1, 70435 Stuttgart, Germany

⁵Telecom Italia S.p.A. via G.Reiss Romoli 274, 10148 Turin, Italy

⁶TTI Norte, S.L., Albert Einstein 14 PCTCAN, 39011 Santander, Cantabria

⁷Ericsson Magyarország Kft. H-1037 Budapest, Laborc u. 1. Hungary

⁸Ericsson AB, SE-164 80 Stockholm, Sweden

⁹Centre for Comm. Systems Research (CCSR), BA Building, University of Surrey, Guildford, GU2 7XH, UK
contact author: dessel@imec.be — +32.16.28.11.63

Abstract—With the explosion of wireless communications in number of users and data rates, the reduction of network power consumption becomes more and more critical. This is especially true for base stations which represent a dominant share of the total power in cellular networks. In order to study power reduction techniques, a convenient power model is required, providing estimates of the power consumption in different scenarios. This paper proposes such a model, accurate but simple to use. It evaluates the base station power consumption for different types of cells supporting the 3GPP LTE standard. It is flexible enough to enable comparisons between state-of-the-art and advanced configurations, and an easy adaptation to various scenarios. The model is based on a combination of base station components and sub-components as well as power scaling rules as functions of the main system parameters.

Index Terms—LTE, base station, power consumption, power model, energy efficiency, green radio

I. INTRODUCTION

In a world of exploding wireless communications, improving the power efficiency of radio networks is an important research topic [1]. To evaluate the energy efficiency of today's mobile communication systems and to identify improvement areas for next generation systems, a high level energy efficiency evaluation framework (E³F) has been developed within the Energy Aware Radio and neTwork technologies (EARTH) project [2]. This framework [3] covers the complete system, including network and radios. It enables a quantitative evaluation in terms of efficiency under different traffic and load scenarios. The E³F builds on state-of-the-art radio network evaluation methodology from system level simulations extended with power models, traffic models, and deployment models.

This paper¹ describes the detailed power models of the base station components and sub-components and focuses on how the power is scaled over different scenarios. Four different types of base stations are considered: macro, micro, pico and femto cells. To increase the application area, power is used instead of energy because it is a more natural metric to use

¹Acknowledgment: The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement n 247733 project EARTH.

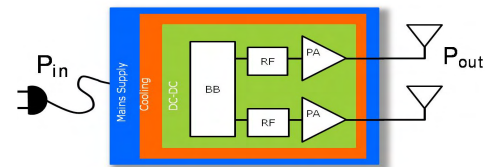


Fig. 1. Overview of base station components included in the power model.

as a function of required throughput or system load, while energy is depending on a given amount of data to transmit in a batch but does not directly relate to system load and efficiency. The model has been designed for good accuracy, simplicity, and flexibility. It enables relative comparisons over many scenarios, even if fully accurate absolute power numbers cannot be guaranteed. This would require extensive physical measurements, and also prevent the model from being used to explore non-measured scenario.

Section II describes the general model structure and its parameters as visible to the model user. Section III details the model construction for all the main components of a typical base station. Section IV illustrates the way the model is used and shows the kind of results that can be obtained. Finally, Section V summarizes the model and its goals.

II. MODEL STRUCTURE

The power model is built around the split of a base station into a number of components and sub-components, as shown in Figure 1. This section introduces those components, distinguishes between different base station types, and presents the main parameters used to compute the power consumption in a specific scenario.

A. Types of base stations and general parameters

Four types of base stations are included in the model: macro, micro, pico, and femto. They are different in the sub-components they contain as well as in the power figures associated with those sub-components, based on different constraints

TABLE I
PARAMETERS AFFECTING SCALING OF BASEBAND AND RF POWER CONSUMPTION. DEFAULT VALUES CORRESPOND TO THE BASELINE CONFIGURATION OF THE BASE STATION. MODULATION IS EXPRESSED IN BITS PER SYMBOL, I.E., 1, 2, 4 OR 6 FOR BPSK, QPSK, 16-QAM OR 64-QAM, RESPECTIVELY.

Notation	Description	Range	Default
BW	Bandwidth [MHz]	1.4 – 20	10
Ant	Number of antennas	1, 2, 4	2
M	Modulation	1, 2, 4, 6	6
R	Coding rate	$1/3 - 1$	$5/6$
dt	Time-domain duty-cycling	0 – 1	1
df	Frequency-domain duty-cycling	0 – 1	1

(output power, maximum load, signal accuracy...). They translate into different basic architectures, for example small base stations (pico and femto) may use more power-efficient dedicated components while large base stations (macro and micro) require more reconfigurability, for example using more FPGAs and less dedicated hardware. This translates into different intrinsic power efficiencies.

Macro base stations are also characterized in the model by having a variable number of sectors while other base station types only have one. The four base station types also differ in maximum output power.

Next to the base station type, the silicon technology is a general parameter affecting the power consumption obtained from the model. It can be specified as either CMOS feature size in nm or year of deployment. The default value for technology is 65 nm or 2010 deployment.

The remaining general parameters define the losses of different sub-components related to the power systems of the base station. This includes losses from cooling, AC/DC conversion, DC/DC conversion and antenna feeder. The three first terms are described in Subsection III-D. Antenna feeder losses are only present between the PA and the antenna in macro base stations. They are included together with the PA model. Other base station types do not have feeder losses due to their more compact design.

B. Base station sub-components and scaling parameters

The sub-component split is based on both real hardware and architecture splits, e.g., analog RF part versus digital baseband part, as well as functional splits, e.g., time-domain processing versus frequency-domain processing. Architecture splits include digital baseband, RF (analog), power amplifier, and overhead (power systems and cooling):

$$P_{\text{Total}} = P_{\text{BB}} + P_{\text{RF}} + P_{\text{PA}} + P_{\text{Overhead}} \quad (1)$$

The reason for those splits is twofold: making clear links to the power consumption of specific components in the system in order to get reference numbers but also identifying how each power figure scales with parameters (time-domain computations may not scale the same way as frequency-domain computations when changing the modulation or system load). As a consequence, the model is centered around two types of tables. Power tables contain reference power figures that were obtained from measurements for the different sub-components of a base station and reasoning on the architecture and specifications of the different types of base stations (see Tables II, III, V and VI). Scaling tables contain specific

scaling factors telling how each power figure evolves with each specific parameter (see Table IV and Section III-B).

This approach is used for baseband as well as RF power consumption. Defining I_{BB} the set of baseband sub-components, I_{RF} the set of analog sub-components, and $X = \{BW, Ant, M, R, dt, df\}$ the list of parameters as described in Table I, we get the following expression²:

$$P_{\text{Total}} = \sum_{i \in I_{\text{BB}}} P_{i,\text{ref}} \prod_{x \in X} \left(\frac{x_{\text{act}}}{x_{\text{ref}}} \right)^{s_{i,x}} + \sum_{i \in I_{\text{RF}}} P_{i,\text{ref}} \prod_{x \in X} \left(\frac{x_{\text{act}}}{x_{\text{ref}}} \right)^{s_{i,x}} + P_{\text{PA}} + P_{\text{Overhead}} \quad (2)$$

The number of antennas is assumed to be the same in transmission and in reception, and also the same as the number of spatial streams. Time-domain duty-cycling represents the fraction of time during which the base station is operating, assuming it fully sleeps during the rest of the time. It is indeed an interesting technique in order to reduce the average power consumption [4]. However, depending on the hardware design and on the speed at which transitions between active and sleeping states should be performed, the full sleeping assumption may be too optimistic, as further discussed in Section IV. Frequency-domain duty-cycling represents the fractional load of the system in frequency resources, i.e., PRBs or physical resource blocks.

Scaling factors are exponents relating each power contributor in the model to each parameter of the model. For example, if the power of a mixer is not impacted by the number of resource blocks currently in use, the exponent will be zero, meaning that a change in resource block allocation (frequency-domain duty-cycling) has no influence on the power of that mixer. Similarly, the baseband demapping complexity will scale linearly with the number of loaded resource blocks, so the corresponding exponent will be one. Other values than zero and one are occasionally used, for example in order to model the quadratic or cubic MIMO processing complexity as function of the number of antennas.

For digital computations, we use complexity figures (GOPS or Giga Operations Per Second) that can be translated into power figures depending on the intrinsic efficiency of the selected technology (GOPS/W). Although this is a crude approximation, this is the only one compatible with the simplicity and flexibility of the model. The power numbers obtained from this estimation have been benchmarked to actual values in existing base stations in order to validate the approach and ensure meaningful results.

III. MODEL COMPONENTS

This section details the power models of the different components of a base station, i.e., the digital baseband, the analog RF, the power amplifier and the power system (power conversion and cooling). The corresponding power figures have been derived from existing designs.

²Baseband sub-components are listed in Table II; RF sub-components are listed in Tables V and VI.

TABLE II
COMPLEXITY OF BASEBAND OPERATIONS IN DOWNLINK (GOPS IN REFERENCE SCENARIO).

<i>GOPS per operation type</i>	Macro	Micro	Pico	Femto
DPD	160	160	0	0
Filter	200	160	120	100
CRPI/SERDES	360	300	0	0
OFDM	80	80	70	60
FD (linear)	30	30	20	20
FD (non-linear)	10	10	5	5
FEC	20	20	20	20
CPU	200	200	30	20

A. Digital baseband processing

The digital processing is modeled based on estimated complexity in GOPS, multiplied by a technology-dependent factor expressing the number of operations that can be performed per second and per Watt. This factor is 40 GOPS/W for large base stations and default technology, i.e., 65 nm General Purpose CMOS. It is three times larger for pico and femto cells given the more dedicated hardware used, as discussed in Subsection II-A. The digital complexity is split into a number of sub-components:

- DPD: Digital Pre-Distortion
- Filter: up/down-sampling and filtering
- CRPI/SERDES: serial link to backbone network
- OFDM: FFT and OFDM-specific processing
- FD: Frequency-Domain processing (mapping/demapping, MIMO equalization); it is split into two parts, scaling linearly and non-linearly with the number of antennas
- FEC: Forward Error Correction
- CPU: platform control processor

The non-linear part of frequency-domain processing accounts for the MIMO operations with quadratic or cubic scaling. The relatively small contributions for frequency-domain processing in Table II (single-antenna reference case) can hence make up a more significant share of the total power with 2 or 4 antennas.

Next to those terms, the leakage power is also taken into account, leading the following total baseband power consumption:

$$P_{BB} = P_{\text{Dynamic}} + P_{\text{Leak}}, \text{ where} \quad (3)$$

$$P_{\text{Dynamic}} = P_{\text{DPD}} + P_{\text{Filter}} + P_{\text{OFDM}} + P_{\text{FD,lin}} + P_{\text{FD,nl}} + P_{\text{CRPI}} + P_{\text{FEC}} + P_{\text{CPU}} \quad (4)$$

The reference leakage power is defined as function of the reference dynamic power, where η_{Leak} is 0.1 in 65 nm technology.

$$P_{\text{Leak,Ref}} = \eta_{\text{Leak}} P_{\text{Dynamic,Ref}} \quad (5)$$

Both dynamic and leakage power have their own scaling rules with technology. We assume for each CMOS generation, e.g., moving from 65 nm to 45 nm, a two-fold reduction of the dynamic power but a three-fold worsening of leakage power, i.e., η_{Leak} is multiplied by 3 [5]. This can be used to assess the power for 45 nm (2012 deployment) instead of the default 65 nm in 2010.

The complexity values given in Table II are used as reference case for the baseband in downlink. This reference case assumes 20 MHz bandwidth, single-antenna, 64-QAM, rate-

TABLE III
COMPLEXITY OF BASEBAND OPERATIONS IN UPLINK (GOPS).

<i>GOPS per operation type</i>	Macro	Micro	Pico	Femto
Filter	200	160	160	150
CRPI/SERDES	360	300	0	0
OFDM	80	80	80	60
FD (linear)	60	60	40	30
FD (non-linear)	20	20	10	10
FEC	120	120	120	110
CPU	200	200	30	20

TABLE IV
SCALING EXPONENTS FOR THE BASEBAND SUB-COMPONENTS IN DOWNLINK (PARAMETERS ARE DEFINED IN SECTION II-B).

<i>Digital scaling exponents</i>	BW	M	R	Ant.	dt	df
DPD, Filter and OFDM	1	0	0	1	1	0
CRPI/SERDES	1	1	1	1	1	1
FD (linear)	1	0	0	1	1	1
FD (non-linear)	1	0	0	2	1	1
FEC	1	1	1	1	1	1
CPU	0	0	0	1	0	0
Leakage	1	0	0	1	0	0

1 encoding³ and a load⁴ of 100%. This reference scenario is used within the model and its tables; it should not be confused with the baseline scenario representing the default operation of a base station as presented in Table I and illustrated in Section IV. The reference CMOS technology is 65 nm (2010 reference year). Values are given in GOPS for the different types of base stations. The figures come from a mixture of system knowledge — estimating the relative complexity of different sub-components — and base station design expertise — matching the total power for selected technology options to the known order of magnitude in existing designs.

Some of the differences between larger and smaller base stations can be explained as follows:

- Predistortion is not applied to small base stations given the different type of PA and power levels.
- Filtering and signal accuracy constraints are more stringent for large base stations, with impact on both up/down-sampling filters and MIMO-OFDM processing.
- The network link is done differently for small base stations (no specific backbone network, general IP connection instead).
- The platform control overhead is larger for a large base station due to the larger platform size including specific additional components.

The corresponding uplink numbers are provided in Table III. Most functional blocks are similar to the downlink case (Table II) but doing reverse operations (the filter is performing down-conversion in uplink, FEC is decoding...). Comparing to the downlink cases, we can note the following differences:

- Predistortion is not used in uplink.
- MIMO processing (equalization and detection) is more complex than the corresponding transmitter steps.
- FEC decoding is significantly more complex (for turbo or convolutional) than encoding.

³In reality the channel coding is deactivated when the rate is equal to 1, but the corresponding value is used as reference for the largest possible throughput.

⁴The load is defined as fractional use of time and frequency resources, i.e., $load = dt \times df$.

TABLE V
RF ANALOG COMPONENT POWER (mW), TRANSMITTER CASE
(DOWNLINK).

Power per analog component [mW]	Macro	Micro	Pico	Femto
IQ modulator	1000	1000	1000	1000
Variable attenuator	10	10	0	0
Buffer	300	300	0	0
Forward Voltage-Contr. Osc. (VCO1)	170	170	170	170
Feedback Voltage-Contr. Osc. (VCO2)	170	170	0	0
Feedback mixer	1000	1000	0	0
Clock generation and buffering	990	990	990	990
Digital-to-Analog Converter (DAC)	1370	1370	200	200
Analog-to-Digital Converter (ADC)	730	730	140	140
Predriver (not incl. in total)	2250	2250	0	0
TOTAL (w/o predriver)	5740	5740	2500	2500
Downscaling factor	1	2	7	12
Total after downscaling [W]	5.7	2.9	.4	.2

Table IV gives the scaling exponents in downlink for each baseband sub-component and each scaling parameter. This table has been derived by analyzing whether the amount of computations to perform in each sub-component was depending on each of the scaling parameters or not. DPD, filtering and OFDM are put together as time domain processing category, with the same scaling properties. Leakage power is also a category of its own. Clock gating is assumed when duty-cycling components in time domain, but not power gating (leakage remains). In uplink, all exponents are the same, except for the non-linear frequency-domain processing (cubic dependency — exponent 3 — in uplink instead of the quadratic dependency — exponent 2 — in downlink).

Example of baseband power scaling: As illustration, let us assume we want to compute the power of frequency-domain, linear processing in case of 10-MHz, 2x2 MIMO, 16-QAM, coding rate 3/4, 100% of time-domain duty-cycling and 30% of frequency occupation, we start from the reference scenario (20-MHz, single-antenna, 64-QAM, coding rate 1, 100% time-domain and frequency-domain duty-cycling), leading for a macrocell in uplink to 60 GOPS (Table III) or $P_{FD,lin,ref} = 1.5$ W (with 40 GOPS/W in 65 nm). The scaling vector with respect to the 6 parameters is $[s_1, s_2 \dots s_6] = [1, 0, 0, 1, 1, 1]$ (Table IV)⁵. It is applied to the ratio of all parameters between actual and reference scenarios, leading the actual power consumption of linearly-scaling frequency-domain processing:

$$\begin{aligned}
 P_{FD,lin} &= P_{FD,lin,ref} \left(\frac{BW_{act}}{BW_{ref}} \right)^{s_1} \left(\frac{M_{act}}{M_{ref}} \right)^{s_2} \left(\frac{R_{act}}{R_{ref}} \right)^{s_3} \\
 &\quad \left(\frac{Ant_{act}}{Ant_{ref}} \right)^{s_4} \left(\frac{dt_{act}}{dt_{ref}} \right)^{s_5} \left(\frac{df_{act}}{df_{ref}} \right)^{s_6} \\
 &= 1.5 \text{ W} \times \left(\frac{10}{20} \right)^1 \left(\frac{4}{6} \right)^0 \left(\frac{3/4}{1} \right)^0 \left(\frac{2}{1} \right)^1 \\
 &\quad \left(\frac{100}{100} \right)^1 \left(\frac{30}{100} \right)^1 \\
 &= 0.45 \text{ W} \tag{6}
 \end{aligned}$$

B. RF sub-components

The RF architecture contains the different elements of a low-IF/zero-IF architecture, in particular clock/carrier gener-

⁵Formally the scaling exponents are named $[s_{FD,lin,BW} \dots s_{FD,lin,df}]$ according to (2). The shorter notation $[s_1 \dots s_6]$ has been selected in order to have a more readable equation (6).

TABLE VI
RF ANALOG COMPONENT POWER (mW), RECEIVER CASE (UPLINK).

Power per analog component [mW]	Macro	Micro	Pico	Femto
First Low-Noise Ampl. (LNA1)	300	300	300	300
Main variable attenuator	10	10	10	10
Second Low-Noise Ampl. (LNA2)	1000	1000	0	0
Dual mixer	1000	1000	1000	1000
Dual IF Variable Gain Ampl. (VGA)	650	650	0	0
Clock generation and buffering	990	990	990	990
Analog-to-Digital Conv. (ADC)	1190	1190	290	290
TOTAL	5140	5140	2590	2590
Downscaling factor	1	2	7	12
Total after downscaling [W]	5.1	2.6	.4	.2

ation and distribution, modulator, mixers, low-noise amplifier (LNA), variable-gain amplifier (VGA), analog/digital converters (DAC and ADC), filters, buffer and predriver, and feedback chain. Only the PA is considered separately (Section III-C) because it cannot be captured in the same modeling approach. All the RF elements consume a specific amount of power, as given in Table V for downlink in reference technology of 65 nm CMOS.

The predriver is shown for information but included into the PA power. Some of the sub-components are not present in smaller base stations. Moreover, an overall downscaling factor reduces the power on smaller base stations due to less constraining specs and different hardware implementation of smaller cells. The reason is two-fold. Firstly, the amount of blockers a small base station has to face is smaller, leading to more relaxed linearity specs and hence less power is needed. Secondly, smaller base stations can work from a lower supply voltage, further reducing their consumption.

Scaling of RF power with input parameters is done as follows. All RF sub-components have a scaling exponent 1 with respect to number of antennas and time-domain duty-cycling. For carrier and clock generation sub-components, all the other scaling exponents have the value zero. For the other RF sub-components, the scaling exponent is also 1 with respect to frequency-domain duty-cycling, assuming a scalable implementation, and for small base stations with respect to bandwidth as well.

Technology scaling is used for analog, too, based on a scaling factor as function of the selected CMOS technology compared to the reference $tech = 65$ nm case. This is an empirical rule from designers' experience, not a physical law:

$$\text{Power}(tech) = \text{Power}(65 \text{ nm}) \left(1 + \frac{tech/65 - 1}{2} \right) \tag{7}$$

Table VI provides the numbers for uplink direction, when the analog front-end is receiving. The same downscaling factors are used as in the downlink case.

C. Power amplifier

The power amplifier behavior cannot be captured by a single reference power number and scaling rules. Hence, the PA model is represented by a table containing measurements of output power versus consumed power. Measured points differ in requested output power and in tuning of the 1 dB compression point. The power model picks up the point with minimal power consumption that is satisfying the output power and linearity constraints.

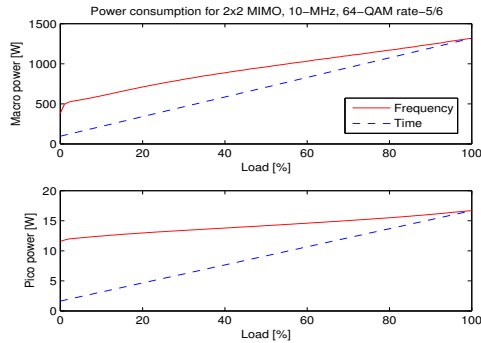


Fig. 2. Power consumption as function of load for a 3-sector macro cell and a pico cell. Load dependency is based on either frequency-domain duty-cycling or time-domain duty-cycling. FDD is expected (uplink plus downlink).

The maximum total output power is 46 dBm for macro, 41 dBm for micro, 24 dBm for pico and 20 dBm for femto cells. Additionally, feeder cable losses amount to 3 dB in macro base stations only. When using multiple antennas, this output power is divided between the antennas, keeping the same total value (within one sector). Large base station only have to support a PAPR of 8 dB thanks to the pre-distortion; small base stations have to support a PAPR of 12 dB.

The PA selected for macro and micro base stations can output up to 54 dBm at 1 dB compression for a consumption of 390 W. Output power reduces linearly with frequency-domain load or explicit power control. The reduction in consumed power is not proportional given that the PA efficiency degrades at reduced load. The PA minimal power consumption at low output power is 11 W. Similarly, small cells (pico and femto) use a PA design capable of up to 36 dBm for a consumption of 8 W. Its minimum power consumption is 650 mW.

D. Overhead

In the overhead or power system category, we place all the sub-components that are related to the system powering, including AC/DC and DC/DC conversion as well as cooling. In order to keep a simple model, the power is computed as a fixed overhead linearly depending on the total power of the rest of the base station. The corresponding factors can be given as input parameters (otherwise default values are used, as specified in Section II-A):

$$P_{\text{overhead}} = (P_{\text{BB}} + P_{\text{RF}} + P_{\text{PA}}) \times ((1 + \eta_{\text{cool}})(1 + \eta_{\text{dc/dc}})(1 + \eta_{\text{ac/dc}}) - 1) \quad (8)$$

Default values for the linear loss model are 10% for cooling but in macro base stations only, 5% for DC/DC conversion, and 10% for AC/DC conversion. There is no cooling for smaller base stations.

IV. MODEL USAGE AND RESULTS

The model has been implemented in Matlab. This enables a flexible generation of output power for different scenarios. All input parameters described in Section II are implemented including their default values for the baseline scenario. The

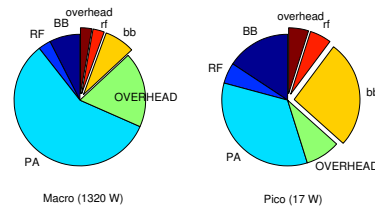


Fig. 3. Relative share of the different components for two types of base stations (configuration of Figure 2). UPPERCASE labels denote DOWNLINK; lowercase labels and shares shifted out of the pie center denote uplink.

output separates power consumption in uplink and downlink, for each of the main components described in Section III.

Figure 2 illustrates the power obtained model as function of the system load. Two of the four base station types have been selected as example. Time-domain duty-cycling outperforms frequency-domain duty-cycling due to the possibility to let specific components sleep during inactive periods. However, the way time-domain duty-cycling it is implemented in the model and illustrated in this Figure is optimistic in the sense that duty-cycled components are assumed to be switched on and off without any overhead. This suits scenarios where sufficiently long sleeping time is present between transitions. When considering micro-sleeping for which this assumption does not hold, more realistic numbers are obtained by using the sleep mode approach of [3].

In the case of a macro-cell, 87% of the power is used for downlink and 13% for uplink. The relative share of the different components is shown on Figure 3. This illustrates that while large base stations are strongly dominated by PA and to a lesser extent by the power systems (overhead), smaller base stations are less PA-dominated and also contain a significant baseband share.

V. CONCLUSIONS

In this paper, we propose a high-level power consumption model for LTE base stations. It covers various types of base stations and is flexible due to a large number of parameters. It provides power figures for the different sub-components present in a typical base station. This model is mostly designed not to give very accurate absolute figures but to enable exploring different architectures as proposed within exploration projects (more antennas, trade-offs between macro and small cells, load impact, role of duty-cycling...).

REFERENCES

[1] A. Fehske, G. Fettweis, J. Malmolin, and G. Biczok, "The global footprint of mobile communications: The ecological and economic perspective," *IEEE Communications Magazine*, vol. 49, no. 8, pp. 55–62, Aug. 2011.
 [2] "EARTH project under european community's seventh framework programme FP7/2007–2013, grant agreement n 247733."
 [3] Gunther Auer, Vito Giannini, István Gódor, Per Skillermark, Magnus Olsson, Muhammad Ali Imran, Dario Sabella, Manuel J. Gonzalez, Claude Desset, Oliver Blume, and Albrecht Fehske, "How much energy is needed to run a wireless network?," *IEEE Wireless Communications Magazine, special issue on Technologies for Green Radio Communication Networks*, vol. 18, no. 4, Oct. 2011.
 [4] P. Frenger et al., "Reducing energy consumption in LTE with cell DTX," in *IEEE VTC Spring*, Budapest, Hungary, May 2011.
 [5] Y. Taur, "CMOS design near the limit of scaling," *IBM Journal of Research and Development*, vol. 46, no. 2.3, pp. 213–222, march 2002.

Minimizing Base Station Power Consumption

Hauke Holtkamp, *Member, IEEE*, Gunther Auer, *Member, IEEE*, Samer Bazzi, *Member, IEEE*, and Harald Haas, *Member, IEEE*

Abstract—We propose a new radio resource management algorithm which aims at minimizing the base station supply power consumption for multi-user MIMO-OFDM. Given a base station power model that establishes a relation between the RF transmit power and the supply power consumption, the algorithm optimizes the trade-off between three basic power-saving mechanisms: antenna adaptation, power control and discontinuous transmission. The algorithm comprises two steps: a) the first step estimates sleep mode duration, resource shares and antenna configuration based on average channel conditions and b) the second step exploits instantaneous channel knowledge at the transmitter for frequency selective time-variant channels. The proposed algorithm finds the number of transmit antennas, the RF transmission power per resource unit and spatial channel, the number of discontinuous transmission time slots, and the multi-user resource allocation, such that supply power consumption is minimized. Simulation results indicate that the proposed algorithm is capable of reducing the supply power consumption by between 25% and 40%, dependent on the system load.

Index Terms—Energy efficiency, green radio, resource allocation, power control, sleep mode, antenna adaptation, discontinuous transmission (DTX), inverse water-filling

I. INTRODUCTION

THE POWER consumption of cellular radio networks is constantly increasing due to the growing number of mobile terminals and higher traffic demands, which require a densification of the network topology. Environmental concerns, rising energy costs and a growing need for self-sufficient (power-grid independent) Base Stations (BSs) reinforce efforts to reduce the cellular networks' power consumption [1]. Shortcomings of the state-of-the-art lie in the fact that current BSs are designed to serve peak demands without considering energy efficient off-peak operation. Depending on the time of day or geographic location, a BS may be idle—and thus over-provisioning—for a significant portion of its operation time. In such low-load situations, spectral efficiency can be traded for energy efficiency without reducing the user experience. The importance here is to operate BSs such that they can flexibly adjust to traffic demand. Furthermore, any power-saving operation on the BS side must not negatively affect

the mobile terminal, *e.g.* by increasing its power consumption or reducing the perceived quality of service.

With regard to different power-saving Radio Resource Management (RRM) mechanisms, Power Control (PC) is the most prominent in literature. This is due to the fact that, in addition to reducing power consumption, PC is also beneficial to link adaptation and interference reduction [2], [3]. In [2] Wong *et al.* minimize transmit power for multi-user Orthogonal Frequency Division Multiplexing (OFDM) under rate constraints. Their PC algorithm is computationally complex and does not consider a transmit power constraint or power model. Cui *et al.* analyze the energy efficiency of Multiple-Input Multiple-Output (MIMO) transmissions, being the first to consider the supply power consumption in energy efficiency studies [3]. They show that in terms of energy efficiency, the optimal number of transmit antennas used for MIMO transmission depends on the Signal-to-Noise-Ratio (SNR). Miao *et al.* [4] derive the data rate that maximizes the transmitted information per unit energy (bit per joule). More recently, acknowledging the significance of circuit power consumption, Discontinuous Transmission (DTX) and Antenna Adaptation (AA) have been identified as energy saving RRM techniques. Kim *et al.* [5] establish AA as a MIMO resource allocation problem and adapt the number of transmit antennas on a single link. However, solutions provided for AA on the mobile side cannot be directly applied to the BS, since for the latter multiple links have to be considered. Some works have assessed the energy saving potentials of sleep modes and AA in Long Term Evolution (LTE) BSs, but have applied very crude decision mechanisms [6]–[8], *e.g.* a BS is only allowed to sleep when a cell is empty. To the best of our knowledge, there are no prior works which consider DTX in combination with other power saving RRM mechanism.

The closest related works to Inverse Water-filling, which is introduced in Section IV, have studied the problem of allocating transmit power, user rates and resources per user with suboptimal solutions [3], [9], [10]. Bit capacities, modulation as well as transmission power consumption are investigated in general MIMO point-to-point transmissions [3], but without considering power-allocation to orthogonal channels. A later work assigns power sub-optimally in favor of a fairness constraint [9]. Al-Shatri *et al.* [10] have used Lagrange multipliers to invert water-filling in a rate-maximizing fairness application. We extend their work by relaxing the equal rate assumption among users and present a complete power-minimizing algorithm.

In this paper, we address the issue of overly generic assumptions about underlying hardware by identifying the relevant energy saving schemes for cellular networks on the basis of a

Manuscript received April 13, 2012; revised October 31, 2012. This work has received funding from the European Community's 7th Framework Programme [FP7/2007-2013. EARTH, Energy Aware Radio and neTwork technologies] under grant agreement n° 247733. The material in this paper was in part presented at IEEE Vehicular Technology Conference (VTC), San Francisco, USA, September 2011.

H. Holtkamp and S. Bazzi are with DOCOMO Euro-Labs, Munich, Germany (e-mail: lastname@docomolab-euro.com).

G. Auer is with Ericsson AB, Stockholm, Sweden (e-mail: gunther.auer@ericsson.com).

H. Haas is with the University of Edinburgh (e-mail: h.haas@ed.ac.uk).
Digital Object Identifier 10.1109/JSAC.2014.141210.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

2

IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 32, NO. 12, DECEMBER 2014

realistic and detailed power model [11]. As a consequence, and unlike previous energy saving studies [2], this work minimizes the *supply power* (or 'AC-plug' power) of the cellular BS rather than the Radio Frequency (RF) transmission power. To provide an applicable mechanism for current cellular systems like 3rd Generation Partnership Project (3GPP) LTE, we present the Resource allocation using Antenna adaptation, Power control and Sleep modes (RAPS) algorithm which reduces the BS supply power consumption of multi-user MIMO-OFDM. Given the channel states and target rates per user, RAPS finds the number of transmit antennas, the number of DTX time slots and the resource and power allocation per user. DTX refers to a micro sleep on the transmitter side, which is sufficiently short so that the regular operation of the mobile is not affected. The RAPS solution is found in two steps: First, PC, DTX and resource allocation are joined into a convex optimization problem which can be efficiently solved. Second, subcarrier and power allocation for a frequency-selective time-variant channel pose a combinatorial problem, which is solved by means of a heuristic solution. To allocate minimal power levels to resources and spatial channels, we derive the Inverse Water-filling algorithm.

The paper is organized as follows: The system and power model is presented in Section II. The global energy efficiency problem is defined in Section III. Section IV presents Step 1 of the two-step RAPS algorithm which determines DTX duration, antenna number, and resource shares on the basis of block fading channels. Step 2 extends these estimates by subcarrier and power allocation in frequency-selective time-varying channels in Section V. Simulation results are presented in Section VI. The paper is concluded in Section VII.

II. SYSTEM AND POWER MODEL

A. System Model

We consider one transmission frame in the downlink of a point-to-multipoint wireless communication system, comprising one serving BS and multiple mobile receivers. The BS transmitter is equipped with M_T antennas and all antennas share the transmit power budget. Mobile receivers have M_R antennas and system resources are shared via Orthogonal Frequency Division Multiple Access (OFDMA) between K users on N subcarriers and T time slots. In total there are TN resource units. OFDMA is used, for example, in 3GPP LTE. A frequency-selective time-variant channel is assumed, with each resource unit characterized by the channel state matrix $\mathbf{H}_{n,t,k} \in \mathbb{C}^{M_R \times M_T}$, with subcarrier index $n = 1, \dots, N$, time slot index $t = 1, \dots, T$, user index $k = 1, \dots, K$. The vector of spatial channel eigenvalues per resource unit a and user k is $\mathcal{E}_{a,k}$ and its cardinality is $\min\{M_T, M_R\}$. The system operates orthogonally such that individual resource units cannot be shared among users. MIMO transmission with a variable number of spatial streams is assumed over the set of resources assigned to each user \mathcal{A}_k . The frame structure is illustrated in Fig. 1.

Although cellular wireless networks are generally interference limited, co-channel interference between neighbouring cells is not considered in this work. The main contribution of this work lies in the detailed derivation of the BS supply

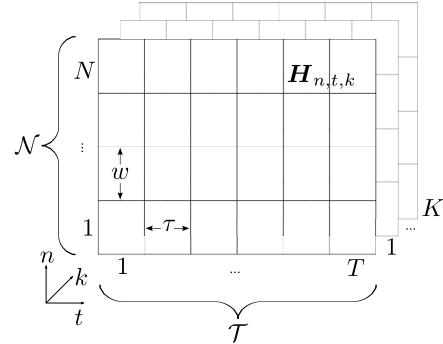


Fig. 1. OFDM frame structure.

power minimization via a power model. The consideration of interference and resulting network dynamics are, therefore, beyond scope of this paper and are left for future work.

B. Power Model

Supply power consumption, *i.e.* the overall device power consumption of the BS, is calculated via a power model. Power consumption models for cellular BSs are established in [11], [12]. The model allows to map RF transmission power to supply power consumption. The power model is constructed from a real BS hardware implementation, and is able to capture the most relevant power consumption effects in commercially available BSs. It provides a simple linear model which is verified by an advanced realistic and detailed hardware model [13]. Therefore, the linear power model provides a sufficient foundation for the BS energy efficiency analysis conducted in this paper. In the power model [11], [12], a RF chain refers to a set of hardware components composed of a small-signal transceiver section, a Power Amplifier (PA) and the antenna interface. The PA consumes a large share of the overall power due to its low efficiency [11]. Each transmit antenna is connected to an RF chain, thus the term AA implies the adaptation of the respective RF chain. Unlike, *e.g.* a shared baseband component, it is assumed that RF chains can be switched off or put to sleep individually when there is no transmission on the respective antenna. The BS consumes less power when fewer RF chains are active. Setting the entire BS to sleep interrupts transmission but further reduces the power consumption. For simplicity, we assume instantaneous and cost-free on-off-switching of hardware, *i.e.* there is no delay or additional power cost incurred by putting a BS into DTX mode or enabling/disabling an RF chain. This assumption is supported by [7], in which a switching time of $30 \mu\text{s}$ is assumed, which is much smaller than the duration of an OFDMA frame. We, therefore, presume that the additional power penalty due to the settling time of the RF components can be neglected.

Power model parameters are provided in [11] only for a BS with two RF chains. We derive the supply power consumption value when operating with a single RF chain as follows: In [11] power consumption characteristics for a macro BS with three sectors are provided. The power consumption is rated at a

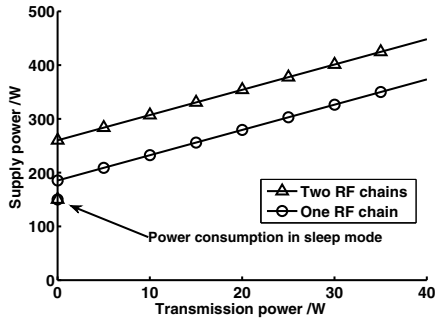


Fig. 2. Supply power model [11].

minimum active consumption P_{0,M_T} , which depends on M_T , a linear transmission power dependence factor Δ_{PM} (the slope) and a power consumption in DTX mode P_S . Transmission power consumption in any time slot t is P_t . Under the assumption of unchanged efficiencies of all components, we use the tables provided in [11] to find the maximum power consumption of a BS operated with a single RF chain and assume Δ_{PM} to be unchanged. Since we presume instantaneous switching capabilities for all components, the sleep mode consumption is the same in both RF chain settings. Lastly, we normalize the power consumption values to a single sector, leading to the power model depicted in Fig. 2. Formally, the supply power consumption is

$$P_{\text{supply}}(P_t) = \begin{cases} P_{0,M_T} + \Delta_{PM}P_t, & \text{if } P_t > 0, \\ P_S, & \text{if } P_t = 0. \end{cases} \quad (1)$$

III. GLOBAL PROBLEM STATEMENT

In this section, we deduce power-saving mechanisms from the power model and derive the global OFDMA supply power minimization problem.

A. Energy Saving Strategies

The power model in Fig. 2 reflects three general mechanisms that affect the power consumption of BSs: First, the overall power emitted at the PA P_t affinely relates to supply power consumption (1). Second, if the same BS is operated with fewer RF chains, it consumes less power. Third, if the BS is put into sleep mode, it consumes less power than in the active state with zero transmission power.

These observations lead to the following saving strategies:

- PC : Reduce transmission power on each resource unit.
- AA : Reduce the number of RF chains.
- DTX : Increase the time the BS spends in DTX.

The first and the last strategy are clearly opposing each other as lower transmission powers lead to lower link rates and thus longer transmission duration (for an equal bit-load), whereas it would be beneficial for long DTX to have short transmissions. The second and third strategy are related, as AA can be considered a weak form of DTX which still allows transmission on a subset of antenna elements.

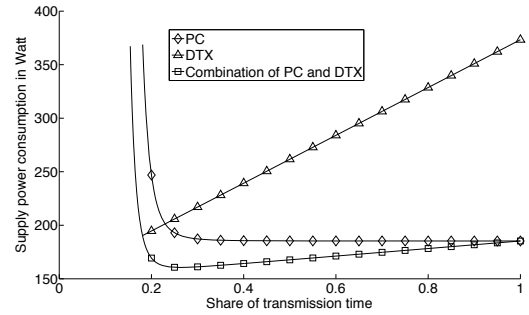


Fig. 3. Supply power consumption for transmission of a target spectral efficiency as a function of time spent transmitting, Φ . The combination of DTX and power control achieves lower power consumption than exclusive operation of either DTX or power control.

B. Joint Strategies

For a single link, the energy efficiency trade-off between PC and DTX is illustrated in Fig. 3. Shown is the supply power consumption caused by transmitting a fixed amount of data. Three operation modes are depicted in Fig. 3:

- PC: Only PC but no DTX is available. The transmission power is adjusted depending on the transmission time normalized to the time slot duration, Φ , such that the target bit load is transmitted. The BS consumes idle power P_0 when there is no data transmission. Clearly, the strategy for lowest power consumption in this case is to set $\Phi=1$. Reducing Φ increases the required transmission power, until at $\Phi = 0.18$, $P_{TX} = P_{\max}$.
- DTX: The BS always transmits with full power $P_{TX} = P_{\max}$. The supply power is $P_0 + mP_{\max}$ when transmitting or P_S when in DTX mode, yielding an affine function of Φ . At $\Phi = 1$, more data than the target bit load is transmitted. Reducing Φ from $\Phi = 1$ decreases the supply power consumption linearly up to the point where the target bit load is met. Here, the best strategy clearly is to minimize the time transmitting (small Φ), so to maximize the time in DTX ($1-\Phi$).
- Joint application of PC and DTX: In this mode of operation, the DTX time ($1-\Phi$) is gradually reduced. The transmission power is adjusted to meet the target bit load within Φ .

Fig. 3 shows that the joint operation of PC and DTX consumes less power than each individual mode of operation with an optimal point at $\Phi=0.25$.

C. Global Problem

We extend the power control and DTX trade-off to MIMO-OFDM serving multiple users over frequency-selective channels. The selection of the number of transmit antennas for AA is made once for the entire frame.

We formulate the global problem statement: Given the channel state matrices $\mathbf{H}_{n,t,k}$ on each channel and the vector of target rates per user $\mathbf{r} = (R_1, \dots, R_K)$, we seek:

- the set of resources allocated to each user \mathcal{A}_k ,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4

IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 32, NO. 12, DECEMBER 2014

- the power level $P_{a,e}$ per resource unit $a = 1, \dots, |\mathcal{A}_k|$ and spatial channel $e = 1, \dots, |\mathcal{E}_{a,k}|$,
- and the number of active transmit antennas M_T ,

such that the supply power consumption is minimized while fulfilling the transmission power constraint P_{\max} .

The sum capacity of user k over one transmission frame of duration τ_{frame} is given by:

$$R_k = \frac{w\tau}{\tau_{\text{frame}}} \sum_{a=1}^{|\mathcal{A}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} \log_2 \left(1 + \frac{P_{a,e} \mathcal{E}_{a,k}(e)}{N_0 w} \right), \quad (2)$$

with subcarrier bandwidth w in Hz, time slot duration τ in seconds, and noise spectral density N_0 in W/Hz.

The RF transmission power in time slot t is

$$P_t = \sum_{a=1}^{|\mathcal{A}_t|} \sum_{e=1}^{|\mathcal{E}_a|} P_{a,e}, \quad (3)$$

where \mathcal{A}_t is the set of N resources in time slot t and \mathcal{E}_a is the vector of channel eigenvalues on resource a .

Given (1), (2) and (3), the optimization problem that minimizes the supply power consumption is of the form

$$\begin{aligned} & \underset{\mathcal{A}_k \forall k, P_{a,e} \forall a,e, M_T}{\text{minimize}} && P_{\text{supply,frame}}(\mathbf{r}) = \\ & \frac{1}{T} \left(\sum_{t=1}^{T_{\text{Active}}} (P_{0,M_T} + \Delta_{\text{PM}} P_t) + \sum_{t=1}^{T_{\text{Sleep}}} P_S \right) \end{aligned} \quad (4)$$

subject to $P_t \leq P_{\max}$, $T_{\text{Active}} + T_{\text{Sleep}} = T$,

$$R_k \leq \frac{w\tau}{\tau_{\text{frame}}} \sum_{a=1}^{|\mathcal{A}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} \log_2 \left(1 + \frac{P_{a,e} \mathcal{E}_{a,k}(e)}{N_0 w} \right)$$

with the number of active transmission time slots T_{Active} , and DTX slots T_{Sleep} . This is a set selection problem over the sets \mathcal{A}_k , and $\mathcal{E}_{a,k}$, as well as a minimization problem in $P_{a,e}$.

D. Complexity

Dynamic subcarrier allocation is known to be a complex problem for a single time slot in frequency-selective fading channels that can only be solved by suboptimal or computationally expensive algorithms [2], [14], [15]. In this study, we add two additional degrees of freedom by considering AA and DTX, increasing the complexity further. We consequently divide the problem into two steps: First, real-valued estimates of the resource share per user, DTX duration, and number of active RF chains M_T , are derived based on simplified system assumptions. Second, the power-minimizing resource allocation over the integer set \mathcal{A}_k and consecutive Inverse Water-filling power allocation are performed.

IV. STEP 1: ANTENNA ADAPTATION, DTX AND RESOURCE ALLOCATION

The first step to solving the global problem is based on a simplification of the system assumptions. This allows defining a convex subproblem with an optimal solution, which is later (sub-optimally) mapped to the solution of the global problem in Step 2, described in Section V.

Instead of time and frequency-selective fading we assume that channel gains on all resources are equal to the center resource unit:

$$\mathbf{H}_k = \mathbf{H}_{n^c, t^c, k}, \quad (5)$$

where the superscript c signifies the center-most subcarrier and time slot. Step 1 thus assumes a block fading channel per user over $W = Nw$ and τ_{frame} .

We select the center resource unit due to the highest correlation with all other resources. Alternative methods to construct a representative channel state matrix are to take the mean or median of $\mathbf{H}_{n,t,k}$ over the OFDMA frame. However, application of the mean or median over a set of MIMO channels were found to result in a channel with lower capacity. See Section VI for a comparison plot between different channel selection methods.

The link capacity is calculated using equal-power precoding and assuming uncorrelated antennas. In contrast to water-filling precoding, equal-power precoding provides a direct relationship between total transmit power and target rate. On block fading channels with real-valued resource sharing, OFDMA and Time Division Multiple Access (TDMA) are equivalent. Without loss of generality, we select resource allocation via TDMA. These simplifications allow to establish a convex optimization problem that can be efficiently solved.

In a block fading multi-user downlink with target rate R_k and equal power precoding, the spectral efficiency for user k is given by

$$C_k = \frac{R_k}{W \mu_k} = \sum_{e=1}^{|\mathcal{E}_k|} \log_2 \left(1 + \frac{P_k \mathcal{E}_k(e)}{M_T N_0 W} \right), \quad (6)$$

with the vector of channel eigenvalues per user \mathcal{E}_k , and transmission power P_k . The resource share $\mu_k \in (0, 1]$ is the normalized (unit-less) representation of the share of time.

The transmission power is a function of the target rate, depending on the number of transmit and receive antennas. For the following configurations, (6) reduces to:

1x2 SIMO:

$$P_k(R_k) = \frac{1}{\epsilon_1} \left(2^{\left(\frac{R_k}{W \mu_k} \right)} - 1 \right) \quad (7)$$

2x2 MIMO:

$$P_k(R_k) = \frac{-(\epsilon_1 + \epsilon_2) + \sqrt{(\epsilon_1 + \epsilon_2)^2 + 4\epsilon_1\epsilon_2 \left(2^{\frac{R_k}{W \mu_k}} - 1 \right)}}{\epsilon_1\epsilon_2} \quad (8)$$

where ϵ_i is the i -th member of \mathcal{E}_k . These equations can be extended in similar fashion to combinations with up to four transmit or receive antennas. Note that a higher number of antennas would require the general algebraic solution of polynomial equations with degree five or higher, which cannot be found in line with the Abel-Ruffini theorem.

Like on the BS side, the number of RF chains used for reception at the mobile could be adapted for power saving. However, the power-saving benefit of receive AA is much smaller than in transmit AA, where a PA is present in each RF chain. Moreover, multiple receive antennas boost the useful signal power and provide a diversity gain. Therefore, M_{Rx} is assumed to be fixed and set to $M_{\text{Rx}}=2$ in the following.

Given the transmission power (7), (8), and the power model (1), we derive the supply power consumption for TDMA. With DTX a BS may go to sleep when all K users have been served; the overall average consumption is the weighted sum of power consumed during transmission and DTX time share μ_S :

$$P_{\text{supply}}(\mathbf{r}) = \sum_{k=1}^K \mu_k (P_{0,M_T} + \Delta_{\text{PM}} P_k(R_k)) + \mu_S P_S \quad (9)$$

Consequently, we define the optimization problem with $\mu_{K+1} = \mu_S$:

$$\begin{aligned} & \underset{(\mu_1, \dots, \mu_{K+1})}{\text{minimize}} && P_{\text{supply}}(\mathbf{r}) = \\ & && \sum_{k=1}^K \mu_k (P_{0,M_T} + \Delta_{\text{PM}} P_k(R_k)) + \mu_{K+1} P_S \\ & \text{subject to} && \sum_{k=1}^{K+1} \mu_k = 1 \\ & && \mu_k \geq 0 \quad \forall k \\ & && 0 \leq P_k(R_k) \leq P_{\text{max}} \end{aligned} \quad (10)$$

The first constraint ensures that all resources are accounted for and upper bounds μ_k . The second constraint encompasses the transmit power budget of the BS and acts as a lower bound on μ_k . Note that due to the block fading assumption and TDMA, the transmission power per user P_k is constant over time and equivalent to P_t in (1).

This problem is convex in its cost function and constraints (proof in Appendix A). It can therefore be solved with available tools like the interior point method [16]. As part of the RAPS algorithm, (10) is solved once for each possible number of transmit antennas. M_T is then selected according to which solution results in lowest supply power consumption.

The solution of the first step yields an estimate for the supply power consumption, the number of transmit antennas M_T , the DTX time share μ_S and the resource share per user μ_k . If a solution for Step 1 cannot be found, Step 2 is not performed and outage occurs. We leave outage handling to a higher system layer mechanism which could be avoided by, *e.g.*, prioritize users and reiterate with reduced system load.

V. STEP 2: SUBCARRIER AND POWER ALLOCATION

This section describes the second step of the RAPS algorithm. In Step 2, the results of Step 1 are mapped back to the global problem to find the subcarrier allocation for each user \mathcal{A}_k , the power level per resource and spatial channel $P_{a,e}$ and the number of DTX slots T_{Sleep} .

First, the real-valued resource share μ_k is mapped to the OFDMA resource count per user $m_k \in \mathbb{N}$,

$$m_k = \lceil \mu_k N T \rceil \quad \forall k = 1, \dots, K. \quad (11)$$

Possible rounding effects through the ceiling operation in (11) are compensated for by adjusting the number of DTX time slots

$$T_{\text{Sleep}} = \left\lfloor \frac{TN\mu_{K+1} - K}{N} \right\rfloor = \lfloor T\mu_{K+1} - K/N \rfloor. \quad (12)$$

The remaining time slots are available for transmission,

$$T_{\text{Active}} = T - T_{\text{Sleep}}. \quad (13)$$

The remaining unassigned resources

$$m_{\text{rem}} = NT - \sum_{k=1}^K m_k - NT_{\text{Sleep}} \quad (14)$$

are assigned to m_k in a round-robin fashion. After this allocation, it holds that $m_k = \lfloor \mathcal{A}_k \rfloor$.

Next, the number of assigned resource units per user m_k is equally subdivided into the number of resources per user and time slot $m_{k,t}$ with

$$m_{k,t} = \left\lfloor \frac{m_k}{\sum_{l=1}^K m_l} N \right\rfloor. \quad (15)$$

The remaining unassigned resources

$$m_{t,\text{rem}} = N - \sum_{k=1}^K m_{k,t} \quad (16)$$

are allocated to different $m_{k,t}$ in a round-robin fashion.

Time slots considered for DTX are assigned statically, starting from the back of the frame. Note that the dynamic selection of DTX slots would create additional opportunities for capacity gains or power savings as they could be assigned to time slots with poor channel states, *e.g.* time slots experiencing deep fades.

A corner case exists when $TN\mu_{K+1} < K$ and (12) becomes negative. This occurs when the DTX share $\mu_{K+1} = \mu_S$ is very small and thus traffic load is high. Due to the high traffic load, it is possible that the target rates cannot be fulfilled within the transmit power constraint, leading to outage for at least one user. Accordingly, if $TN\mu_{K+1} < K$, we set $T_{\text{Active}} = T$ and the resource mapping strategy in (11) is adapted such that

$$m_k = \lfloor \mu_k N T \rfloor \quad (17)$$

which guarantees that $m_k < NT$. The remaining resources are allocated to users as outlined above.

A subcarrier allocation algorithm from Kivanc *et al.* [14] is adopted, which has been shown to work effectively with low complexity. The general idea of the algorithm is as follows: First assign each subcarrier to the user with the best channel. Then start trading subcarriers from users with too many subcarriers to users with too few subcarriers based on a nearest-neighbor evaluation of the channel state. The algorithm is outlined in Algorithm 1 and applied to each time slot t consecutively.

At this stage, the MIMO configuration, the number of DTX time slots, and the subcarrier assignment are determined. Transmit powers are assigned in both spatial and time-frequency domains via an algorithm termed Inverse Water-filling (IWF). The notion of this algorithm is as follows: First, channels are sorted by quality and the water-level is initialized on the best channel, such that the bit load target is fulfilled. Then, in each iteration of the algorithm, the next best channel is added to the set of used channels, thus reducing the water-level in each step. The search is finished once the water-level is lower than the next channel metric to be added. Refer to Appendix B for the derivation.

Let us define for each user the target bit-load

$$B_{\text{target},k} = R_k \tau_{\text{frame}}. \quad (18)$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6

IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 32, NO. 12, DECEMBER 2014

Algorithm 1 Adapted Rate Craving Greedy (RCG) algorithm performed on each time slot t . As compared to [14] the cost parameter $h_{n,t,k}$ has been adapted and the absolute value has been added to the search of the nearest neighbor. $\mathcal{A}_{k,t}$ holds the set of subcarriers assigned to user k .

Ensure: $m_{k,t}$ is the target number of subcarriers allocated to each user k , $h_{n,t,k} = |\overline{\mathbf{H}}_{n,t,k}|^2$ and $\mathcal{A}_{k,t} \leftarrow \{\}$ for $k = 1, \dots, K$.

- 1: **for all** subcarriers n **do**
- 2: $k^* \leftarrow \arg \max_{1 \leq k \leq K} h_{n,t,k}$
- 3: $\mathcal{A}_{k^*,t} \leftarrow \mathcal{A}_{k^*,t} \cup \{n\}$
- 4: **end for**
- 5: **for all** users k such that $|\mathcal{A}_{k,t}| > m_{k,t}$ **do**
- 6: **while** $|\mathcal{A}_{k,t}| > m_{k,t}$ **do**
- 7: $l^* \leftarrow \arg \min_{\{l: |\mathcal{A}_{l,t}| < m_{l,t}\}, 1 \leq l \leq N} | -h_{n,t,k} + h_{n,t,l} |$
- 8: $n^* \leftarrow \arg \min_{1 \leq n \leq N} | -h_{n,t,k} + h_{n,t,l^*} |$
- 9: $\mathcal{A}_{k,t} \leftarrow \mathcal{A}_{k,t} / \{n^*\}, \mathcal{A}_{l^*,t} \leftarrow \mathcal{A}_{l^*,t} \cup \{n^*\}$
- 10: **end while**
- 11: **end for**

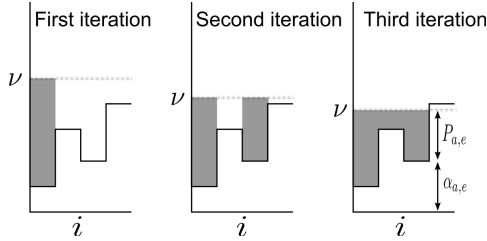


Fig. 4. Illustration of Inverse Water-filling (IWF) with three steps. The height of each patch i is given by $\alpha_{a,e} = N_0 w / \mathcal{E}_{a,k}(e)$. The water-level is denoted by ν . The height of the water above each patch is $P_{a,e}$. The first step sets the water-level such that the bit load target is fulfilled on the best patch. The second and third step add a patch, thus reducing the water-level. After the third step, the water-level is below the fourth patch level, thus terminating the algorithm.

To fulfill the target bit-load, the following constraint must be met

$$B_{\text{target},k} - \sum_{a=1}^{|\mathcal{A}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} w\tau \log_2 \left(1 + \frac{P_{a,e} \mathcal{E}_{a,k}(e)}{N_0 w} \right) = 0, \quad (19)$$

which assesses the sum capacity over a set of resources and spatial channels.

The water-level ν can be found via an iterative search over the set Ω_k channels that contribute a positive power

$$\log_2(\nu) = \frac{1}{|\Omega_k|} \left(\frac{B_{\text{target},k}}{w\tau} - \sum_{e=1}^{|\Omega_k|} \log_2 \left(\frac{\tau \mathcal{E}_{a,k}(e)}{N_0 \log(2)} \right) \right). \quad (20)$$

The water-level is largest on the first iteration and decreases on each iteration until it can no longer be decreased.

Using the Lagrangian method detailed in Appendix B, we arrive at a power-level per spatial channel of

$$P_{a,e} = \frac{\nu w \tau}{\log(2)} - \frac{N_0 w}{\mathcal{E}_{a,k}(e)}. \quad (21)$$

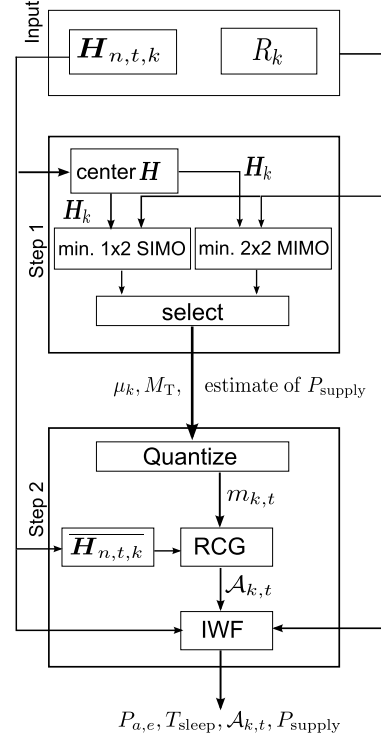


Fig. 5. Outline of the RAPS algorithm.

The inverse water-filling algorithm is illustrated for four channels in Fig. 4. The outcome of Inverse Water-filling as part of RAPS are the transmission power levels $P_{a,e}$ for each resource unit. The supply power consumption after application of RAPS can be found by summation of transmission powers in each time slot and application of the supply power model (1). The entire RAPS algorithm is outlined in Fig. 5.

VI. RESULTS

In order to assess the performance of the RAPS algorithm Monte Carlo simulations are conducted. The simulations are configured as follows: mobiles are uniformly distributed around the BS on a circle with radius 250 m and a minimum distance of 40 m to the BS to avoid peak SNRs. Fading is computed according to the NLOS model described in 3GPP TR25.814 [17] with 8 dB shadowing standard deviation, and the frequency-selective channel model B5 described in [18] with 3 m/s mobile velocity. All transmit and receive antennas are assumed to be mutually uncorrelated. Further system parameters are listed in Table I.

A. Benchmarks

The following transmission strategies are evaluated for comparison purposes:

- The maximum BS power consumption is obtained by constant transmission at maximum power, $P_{\text{supply,max}} = P_{0,M_T} + \Delta_{\text{PM}} P_{\text{max}}$.

TABLE I
SYSTEM PARAMETERS

Variable		Value
K	Number of users	10
N	Number of subcarriers \mathcal{N}	50
T	Number of time slots in set \mathcal{T}	10
M_T	Number of transmit antennas	[1,2]
M_R	Number of receive antennas	2
P_{0,M_T}	Circuit power consumption	185 W/260 W
ΔP_M	Load-dependence factor	4.7
P_S	Power consumption in DTX	150 W
P_{\max}	Maximum transmission power	46 dBm
τ_{frame}/τ	Duration of frame/time slot	10 ms/1 ms
W/w	System/subcarrier bandwidth	10 MHz/200 kHz
N_0	Noise power spectral density	4×10^{-21} W/Hz

- Bandwidth Adaptation (BA), which finds the minimum number of subcarriers that achieves the rate target. No sleep modes are utilized and all scheduled subcarriers transmit with a transmission power spectral density of P_{\max}/N . This benchmark is chosen to represent the power consumption of state-of-the-art BSs which are not capable of DTX or AA.
- DTX where the BS transmits with full power $P_{\text{supply,max}}$ and switches to micro sleep once the rate requirements are fulfilled. This benchmark assesses the attainable savings when only DTX is applied.

B. Performance Analysis

The channel value selection in (5) serves as the channel gain for the block fading assumption of Step 1. How the channel value selection of three possible alternatives (mean, median center) affects the supply power consumption is examined in Fig. 6. It can be seen that the selection of the mean channel gain results in the highest supply power consumption estimate after Step 1. While the estimate can be improved in Step 2, it is still inferior to the other alternatives. Use of the mean channel states causes Step 1 to underestimate the channel quality. Consequently, too few time slots are selected for DTX in Step 2. Choosing the median provides a better estimate than the mean and after Step 2 the solution matches 'Step 2, center selection'. However, in center channel state selection the Step 1 estimate and the Step 2 solution have both the best match and the lowest supply power consumption. Therefore, as mentioned in Section IV, center channel state selection is chosen in the RAPS algorithm and all further analyses.

Fig. 7 compares the outage probabilities of Step 1 with that of the BA benchmark. Outage refers to the lack of a solution to Step 1; when the user target data rates are too high for the given channel conditions, the convex subproblem has no solution, which causes the algorithm to fail. A reduction of user target data rates based on, *e.g.*, priorities, latency or fairness, is specifically not covered by RAPS. One suggested alternative is to introduce admission control, in the way that some users are denied access, so that the remaining users achieve their target rates. Note that Step 2 has no effect on the outage; if a solution exists after Step 1, then Step 2 can be completed. If a solution does not exist after Step 1, then Step 2 is not performed. Fig. 7 illustrates that RAPS selects two transmit antennas for target link rates above 14 Mbps with high probability, as target link rates cannot be achieved with

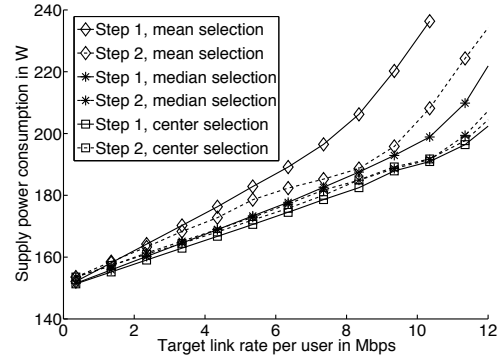


Fig. 6. Performance comparison of different channel state selection alternatives in Step 1.

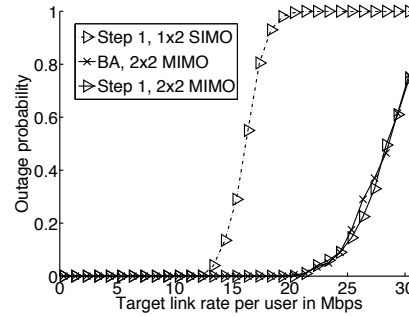


Fig. 7. Outage probability in Step 1 and the BA benchmark.

a single transmit antenna. The BA benchmark and the Step 1 MIMO solution have similar outage behavior.

Fig. 8 depicts the average number of DTX time slots over increasing target link rates. The effect of AA (*i.e.*, switching between Single-Input Multiple-Output (SIMO) and MIMO transmission) can clearly be seen in Fig. 8. For low target link rates, a large proportion of all time slots is selected for DTX. For higher target link rates, the number of DTX time slots must be reduced when operating in SIMO mode. For target link rates above 12 Mbps, which approach the SIMO capacity (as described in the previous paragraph), the system switches to MIMO transmission. The added MIMO capacity allows the RAPS scheduler to allocate more time slots to DTX. In the medium load region (around 15 Mbps) the standard deviation is highest, indicating that here the RAPS algorithm strongly varies the number of DTX time slots depending on channel conditions and whether SIMO or MIMO transmission is selected. These variations contribute strongly to the additional savings provided by RAPS over the benchmarks. Another observation from Fig. 8 is that it is unlikely that more than two out of ten DTX time slots are scheduled at target link rates above 8 Mbps. This means that for a large range of target rates, no more than two DTX time slots are required to minimize power consumption. This is an important finding for applications of RAPS in established

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8

IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 32, NO. 12, DECEMBER 2014

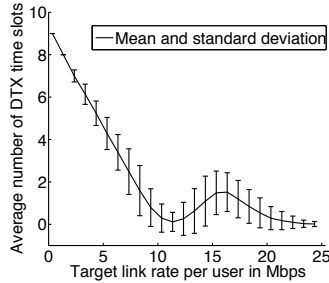


Fig. 8. Average number of DTX time slots over increasing target link rate. Error bars portray the standard deviation. Total number of time slots $T = 1$

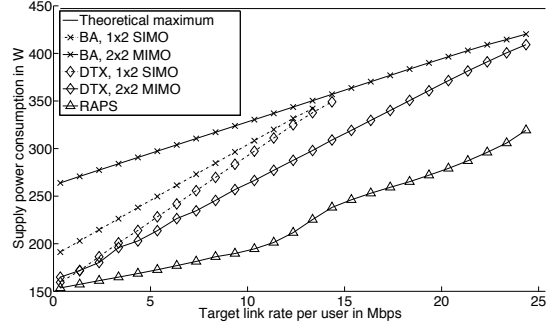


Fig. 10. P_{supply} on frequency-selective fading channels for different RRM schemes and RAPS for ten users. For a bandwidth-adapting BS power consumption is always reduced by AA, while AA is never beneficial for a sleep mode capable BS. For RAPS energy consumption is reduced by AA at low rates. In general, RAPS achieves substantial power savings at all BS loads.

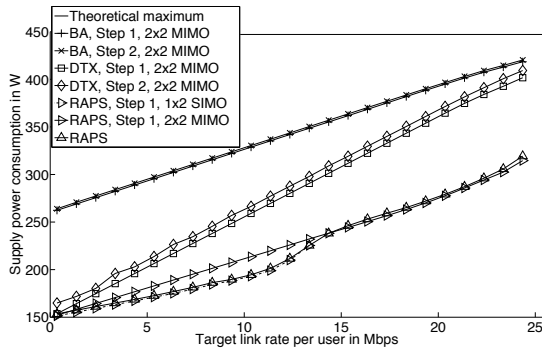


Fig. 9. Supply power consumption for different RRM schemes on block fading and frequency-selective fading channels (comparison of Step 1 and Step 2) for ten users. Overlaps indicate the match between the Step 1 estimate and the Step 2 solution.

systems like LTE, where the number of DTX time slots may be limited due to constraints imposed by the standard.

A comparison of the supply power consumption estimates of Step 1 and Step 2 in Fig. 9 verifies that the estimate taken in Step 1 as input for Step 2 are sufficiently accurate. Although Step 1 is greatly simplified with the assumption of block fading and its output parameters cannot be applied readily to an OFDMA system, it precisely estimates power consumption which is the optimization cost function. The slight difference between the Step 1 estimate and power consumption after Step 2 is caused by quantization loss and resource scheduling. Note that while Step 1 supplies a good estimate of the power consumption, it is not a solution to the original OFDMA scheduling problem, since it does not consider the frequency-selectivity of the channel and does not yield the resource and power allocation.

The performance of RAPS in comparison to the benchmarks is separately analyzed in Fig. 10. An initial observation is that even the supply power consumption of BA is significantly lower than the theoretical maximum for most target link rates. BA with a single transmit antenna always consumes less power than with two antennas, as long as the rate targets with one antenna can be met. AA is thus a valid power-saving mechanism for BA.

The DTX power consumption is significantly lower than for BA, especially at low target rates, because lower target rates allow the BS to enter DTX for longer periods of time. As the BS load increases, the opportunities of the DTX benchmark to enter sleep mode are reduced, so that DTX power consumption approaches that of BA. Unlike in BA, it is never beneficial for a BS capable of DTX to switch operation to a single antenna, because transmitting for a longer time with a single antenna always consumes more power than a short two-antenna transmission, which allows for a longer DTX duration. (Note that this finding may not apply to other power models.)

RAPS reduces power consumption further than DTX by employing AA at low rates and PC at high rates. Through PC, the slope of the supply power is kept low between 15 and 20 Mbps. At higher rates an upward trend becomes apparent, since link rates only grow logarithmically with the transmission power. In theory, when the BS is at full load, the supply power of all energy saving mechanism will approach maximum power consumption. However, when the BS load is very high, not all users may achieve their target rate and outage occurs (see Fig. 7). In other words, operating the BS with load margins controls outage and allows for large transmission power savings. Power savings of RAPS compared to the state-of-the-art BA range from 102.7 W (24.5%) to 136.9 W (41.4%) depending on the target link rate per user.

In addition to absolute consumption we inspect the energy efficiency of RAPS in Fig. 11. We define energy efficiency $E = P_{\text{supply}}^{-1} \sum_{k=1}^K R_k$ in bit/Joule. Observe that all data series are monotonically increasing. Therefore, RAPS does not change the paradigm that a BS is most efficiently operated at peak rates. However, when operating below peak rates, RAPS will always offer the most efficient operation at the requested rate.

VII. SUMMARY AND CONCLUSION

In this paper, we have examined three BS supply power saving mechanisms, namely antenna adaptation, power control and DTX, and evaluated their performance with a base

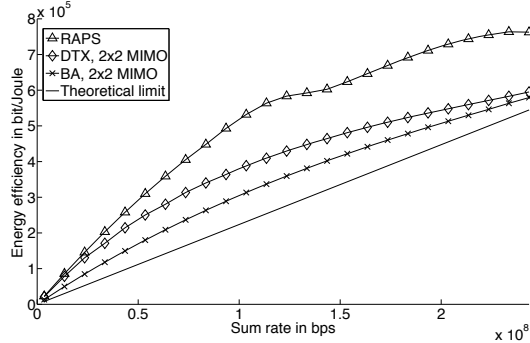


Fig. 11. Energy efficiency as a function of sum rate. Energy efficiency can be increased by the RAPS algorithm. The BS remains most efficient at peak rate.

station power consumption model. The scheduling problem of minimizing BS supply power consumption for the downlink of multiuser MIMO-OFDM was defined. To solve the scheduling problem the RAPS algorithm was proposed, which comprises two steps: first an estimate from a convex subproblem is found assuming constant channel gains on all resource units; then this estimate is applied in the second step to determine the resource and power allocation for each resource unit for a frequency-selective time-variant channel. As part of the second step, we have developed the Inverse Water-filling algorithm which finds the minimal sum power allocation for a target bit-load. Simulation results indicate that RAPS is capable of reducing supply power consumption over all target link rates by 25% to 40%. Savings at low target data rates are mostly attributed to DTX and AA, while at high rates PC is most effective.

While RAPS can be applied to many types of BSs, the reported simulation results clearly depend on the used power model. Future BSs may have different power consumption characteristics. For example, smaller cells can be expected to have lower transmission power compared to circuit consumption. This work is also meant to assist hardware developments in predicting which changes will be most useful for higher-level RRM mechanisms. Further investigations are planned into the effects of hardware switching times and hardware limitations on AA due to the limited operating regions of PAs. Extension of RAPS to the multi-cell interference setting is part of future work.

APPENDIX A

PROOF OF CONVEXITY FOR PROBLEM (10)

The partial second derivative of the MIMO cost function in (10) with respect to μ_k reads as

$$\frac{\partial^2 P_{\text{supply},k}(\mathbf{r})}{\partial^2 \mu_k} = m \log^2(2) \frac{R_k}{W} \frac{2^{\frac{R_k}{W\mu_k}} \left((\epsilon_1 + \epsilon_2)^2 + 2\epsilon_1\epsilon_2 \left(2^{\frac{R_k}{W\mu_k}} - 2 \right) \right)}{\mu_k^3 \left((\epsilon_1 + \epsilon_2)^2 + 4\epsilon_1\epsilon_2 \left(2^{\frac{R_k}{W\mu_k}} - 1 \right) \right)^{\frac{3}{2}}} \quad (22)$$

which is non-negative since

$$2\epsilon_1\epsilon_2 2^{\frac{R_k}{W\mu_k}} \geq 2\epsilon_1\epsilon_2. \quad (23)$$

Thus the cost function is convex in μ_k . Convexity of the SIMO cost function and the constraint function can be shown similarly and is omitted here for brevity.

APPENDIX B

INVERSE WATER-FILLING ALGORITHM

Inverse Water-filling over the set of resources \mathcal{A}_k and the vector of channel eigenvalues $\mathcal{E}_{a,k}$ minimizes power consumption over a set of channels while fulfilling a target bit load.

Assuming block-diagonalization precoding, the problem reads

$$\text{minimize} \quad \sum_{a=1}^{|\mathcal{A}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} P_{a,e} \quad (24)$$

subject to

$$B_{\text{target},k} - \sum_{a=1}^{|\mathcal{A}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} w\tau \log_2 \left(1 + \frac{P_{a,e} \mathcal{E}_{a,k}(e)}{N_0 w} \right) = 0,$$

$$P_{a,e} \geq 0 \quad \forall a, e,$$

$$\sum_{a=1}^{|\mathcal{A}_k|} \sum_{e=1}^{|\mathcal{E}_{a,k}|} P_{a,e} \leq P_{\text{max}}.$$

The derivative of the Lagrangian at the minimum is

$$\frac{\partial \mathcal{L}}{\partial P_{a,e}} = 1 - \lambda + \beta - \frac{\nu w\tau \mathcal{E}_{a,k}(e)}{\log(2) (wN_0 + \mathcal{E}_{a,k}(e) P_{a,e})} = 0, \quad (25)$$

with λ the multiplier for the first inequality constraint, β for the second (power-limit) constraint and ν the equality constraint multiplier. The variable λ can be omitted since it acts as a slack variable.

From (25) we arrive at a power-level per spatial channel of

$$P_{a,e} = \frac{\nu w\tau}{\log(2)} - \frac{N_0 w}{\mathcal{E}_{a,k}(e)}, \quad (26)$$

which can be inserted into the equality constraint of (24) to yield the water-level ν

$$\log_2(\nu) = \frac{1}{|\Omega_k|} \left(\frac{B_{\text{target},k}}{w\tau} - \sum_{e=1}^{|\Omega_k|} \log_2 \left(\frac{\tau \mathcal{E}_{a,k}(e)}{N_0 \log(2)} \right) \right). \quad (27)$$

The water-level can be found via an iterative search over the vector Ω_k of channels that contribute a positive power. Since the sum power is reduced on each iteration, the power constraint (accounted for by the multiplier β) only needs to be tested after the search is finished.

REFERENCES

- [1] A. Fehske, J. Malmolin, G. Biczók, and G. Fettweis, "The Global Carbon Footprint of Mobile Communications - The Ecological and Economic Perspective," *IEEE Commun. Mag.*, 2010.
- [2] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.
- [3] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-Efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks," *IEEE J. Sel. Areas Commun.*, 2004.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10

IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 32, NO. 12, DECEMBER 2014

- [4] G. Miao, N. Himayat, Y. Li, and D. Bormann, "Energy Efficient Design in Wireless OFDMA," in *IEEE International Conference on Communications*, 19-23 2008, pp. 3307–3312.
- [5] H. Kim, C.-B. Chae, G. de Veciana, and R. W. Heath, "A Cross-layer Approach to Energy Efficiency for Adaptive MIMO Systems Exploiting Spare Capacity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4264–4275, Aug. 2009.
- [6] M. Hedayat, M. Amirijoo, P. Frenger, and J. Moe, "Reducing Energy Consumption Through Adaptation of Number of Active Radio Units," in *Proc. IEEE VTC 2012-Spring*, 2012.
- [7] P. Frenger, P. Moberg, J. Malmmodin, Y. Jading, and I. Gódor, "Reducing Energy Consumption in LTE with Cell DTX," in *Proc. IEEE VTC 2011-Spring*, 2011.
- [8] R. Wang, J. Thompson, H. Haas, and P. Grant, "Sleep Mode Design for Green Base Stations," *IET Communications*, vol. 5, no. 18, pp. 2606–2616, 2011.
- [9] Z. Shen, J. Andrews, and B. Evans, "Optimal power allocation in multiuser OFDM systems," in *Global Telecommunications Conference, 2003. GLOBECOM'03. IEEE*, vol. 1. IEEE, 2003, pp. 337–341.
- [10] H. Al-Shatri and T. Weber, "Fair Power Allocation for Sum-Rate Maximization in Multiuser OFDMA," in *Proc. International ITG Workshop on Smart Antennas 2010*, 2010.
- [11] G. Auer, V. Giannini, I. Gódor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. J. Gonzalez, and C. Desset, "Cellular Energy Efficiency Evaluation Framework," in *Proc. VTC 2011-Spring*, 2011.
- [12] G. Auer, V. Giannini, I. Gódor, P. Skillermark, M. Olsson, M. A. Imran, M. J. Gonzalez, C. Desset, O. Blume, and A. Fehske, "How Much Energy is Needed to Run a Wireless Network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct 2011.
- [13] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, W. Wajda, D. Sabella, F. Richter, M. Gonzalez, H. Klessig, I. Gódor, P. Skillermark, M. Olsson, M. A. Imran, A. Ambrosy, and O. Blume, "Flexible power modeling of LTE base stations," in *2012 IEEE Wireless Communications and Networking Conference: Mobile and Wireless Networks (IEEE WCNC 2012 Track 3 Mobile & Wireless)*, 2012.
- [14] D. Kivanc, G. Li, and H. Liu, "Computationally Efficient Bandwidth Allocation and Power Control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1150–1158, Nov. 2003.
- [15] J. Jang and K. B. Lee, "Transmit Power Adaptation for Multiuser OFDM Systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171–178, Feb. 2003.
- [16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Mar. 2004.
- [17] 3GPP, "Physical Layer Aspects for Evolved Universal Terrestrial Radio Access (UTRA)," 3GPP TR 25.814 V7.1.0 (2006-09), Sep. 2006. Retrieved Nov. 2, 2009 from www.3gpp.org/ftp/Specs/.
- [18] IST-2003-507581 WINNER, "D5.4 v1.0 Final Report on Link Level and System Level Channel Models," Retrieved Apr. 15, 2007, from <https://www.ist-winner.org/DeliverableDocuments/>, Nov. 2005.



Hauke Holtkamp received his B.Sc. degree in electrical engineering at Jacobs University Bremen, Germany, in 2005, and the M.Sc. degree in communication networks from Technical University Munich, Germany, in 2008. He is currently working on the energy efficiency of cellular networks at DOCOMO Euro-Labs GmbH in Munich, Germany, while pursuing a Ph.D. degree in communication networks at the University of Edinburgh, UK.



Gunther Auer received the Dipl.-Ing. degree in Electrical Engineering from the Universität Ulm, Germany, in 1996, and the Ph.D. degree from the University of Edinburgh, UK in 2000. From 2000 to 2001 he was a research and teaching assistant with Universität Karlsruhe (TH), Germany. From 2001 to 2012 he is with NTT DOCOMO Euro-Labs, Munich, Germany, where he was team leader and research manager in wireless technologies research. Since February 2012 he is with Ericsson AB, Stockholm, Sweden. His research interests include



green radio, heterogeneous networks and multi-carrier based communication systems, in particular medium access, cross-layer design, channel estimation and synchronization techniques.

Samer Bazi (M10) received a B.E. degree in Computer and Communications Engineering from the American University of Beirut (AUB) in 2008, and a M.S. in Communications Engineering from the Technical University of Munich (TUM) in 2010. He is currently a member of DOCOMO Euro-Labs, located in Munich, Germany and an external Ph.D. candidate at the Associate Institute for Signal Processing at TUM. His research interests include signal processing techniques and interference mitigation for multi-cell multi-user MIMO systems in addition to convex optimization for wireless systems.



Harald Haas (SM'98-AM'00-M'03) holds the Chair of Mobile Communications in the Institute for Digital Communications (IDCOM) at the University of Edinburgh, and he currently is the CTO of a university spin-out company pureVLC Ltd. His main research interests are in interference coordination in wireless networks, spatial modulation and optical wireless communication. Prof. Haas holds more than 25 patents. He has published more than 60 journal papers including a Science Article and more than 160 peer-reviewed conference papers. Nine of his papers are invited papers. Since 2007 Prof. Haas has been a Regular High Level Visiting Scientist supported by the Chinese '111 programme' at Beijing University of Posts and Telecommunications (BUPT). He was an invited speaker at the TED Global conference 2011, and his work on optical wireless communication was listed among the '50 best inventions in 2011' in Time Magazine. He recently has been awarded a prestigious Fellowship from the Engineering and Physical Sciences Research Council (EPSRC) in the UK.

A Parameterized Base Station Power Model

Hauke Holtkamp, *Member, IEEE*, Gunther Auer, *Member, IEEE*, Vito Giannini, *Member, IEEE*, Harald Haas, *Member, IEEE*

Abstract—Power models are needed to assess the power consumption of cellular base stations (BSs) on an abstract level. Currently available models are either too simplified to cover necessary aspects or overly complex. We provide a parameterized linear power model which covers the individual aspects of a BS which are relevant for a power consumption analysis, especially the transmission bandwidth and the number of radio chains. Details reflecting the underlying architecture are abstracted in favor of simplicity and applicability. We identify current power-saving techniques of cellular networks for which this model can be used. Furthermore, the parameter set of typical commercial BSs is provided and compared to the underlying complex model. The complex model is well approximated while only using a fraction of the input parameters.

I. INTRODUCTION

Recently, the power consumption of cellular networks has become a point of interest in research and even been taken into consideration for the standardization of future cellular networks like Long Term Evolution (LTE)-Advanced [1]. It was found in [2] that in cellular networks the element which causes the largest share of overall consumption is the base station (BS). Numerous techniques have consequently been proposed by which the power consumption of BSs can be reduced [3]–[8]. Some of these techniques only consider transmission power while others take into consideration that the generation of the radio signal also consumes power in circuitry by employing power models. Such power models describe abstractly how much power a transmitter consumes and how this consumption depends on operating parameters. In the past, the modelling of BS power consumption often had to be based on intuition until the first power models were published [9]–[13]. Simple models like [9], [10] allow computing the power consumption of a BS for specific configurations. In contrast, the non-linear complex model described by Desset *et al.* [13] is derived from the combination of each of a BS' subcomponents. This allows inspecting the power consumption to such detailed level as the effect of giga operations per second or transistor gate lengths, but is unwieldy to apply. In this paper, we extend the work in [11], [12] by maintaining simplicity while integrating two relevant operating variables into the model, namely the power amplifier (PA)'s output range and the transmission bandwidth. The proposed model allows assessing the power consumption of all techniques that are currently employed to reduce the power consumption of BS while conserving simplicity.

H. Holtkamp is with DOCOMO Euro-Labs, Munich, Germany, e-mail: holtkamp@docomolab-euro.com

G. Auer was with DOCOMO Euro-Labs when this work was carried out. He is now with Ericsson AB, Stockholm, Sweden.

V. Giannini is with IMEC, Leuven, Belgium, e-mail: vito@imec.be

H. Haas is with the University of Edinburgh, UK, e-mail: h.haas@ed.ac.uk

The scope of the model is described in Section II. The model is subsequently presented in Section III. In Section IV, it is discussed and compared to the complex model. The paper is concluded in Section V.

II. POWER MODEL SCOPE

Power saving techniques in literature can be generally divided into design changes and operating approaches. Design changes affect the layout of the network or the hardware architecture. For example, in [14] it is proposed that the use of heterogeneous networks will positively affect the network power consumption under certain conditions. Cui *et al.* [4] show how adjusting the number of antennas affects power consumption. In contrast to design changes, operating approaches manipulate the functionality of a BS during operation. Here, proposed techniques are the reduction of transmission power [5], the deactivation of unneeded antennas [6], the adaptation of the transmission bandwidth [7] and the use of low power consumption sleep modes of varied durations [8].

The model presented in this paper encompasses all of these approaches to allow for a direct comparison while abstracting parameters which can either be assumed to be constant or have been shown to have little effect in the studied scenarios, such as modulation and coding settings, equipment manufacturing details and leakage powers [13]. To this extent, the following is covered in the proposed model:

- The different BS types of a heterogeneous network are modelled by applying different parameter sets to the same model equations.
- The number of transmission antennas and radio chains affects consumption during design and operation.
- The same holds true for transmission power, which affects the design indirectly by choice of a suitable power amplifier as well as the operation directly.
- Also, transmission bandwidth and sleep modes are modelled in their effect on BS power consumption.

III. BASE STATION POWER MODEL

It was found in [11] that the supply power consumption of a BS can be approximated as an affine function of transmission power. In other words, the consumption can be represented by a static (load-independent) share, P_0 , with an added load-dependent share that increases linearly by a power gradient, Δ_p . The maximum supply power consumption, P_1 , is reached when transmitting at maximum total transmission power, P_{\max} . Furthermore, a BS may enter a sleep mode with lowered consumption, P_{sleep} , when it is not transmitting. Fig. 1 shows an illustration. Total power consumption considering the

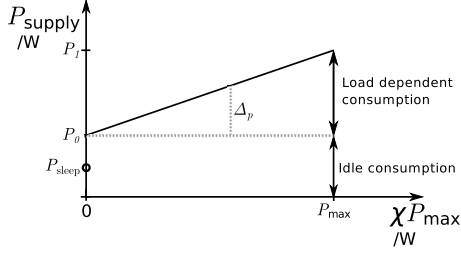


Fig. 1. Load-dependent power model for an LTE BS.

number of sectors, M_{sec} , is then formulated as

$$P_{\text{supply}}(\chi) = \begin{cases} M_{\text{sec}}(P_1 + \Delta_p P_{\text{max}}(\chi - 1)) & \text{if } 0 < \chi \leq 1 \\ M_{\text{sec}} P_{\text{sleep}} & \text{if } \chi = 0, \end{cases} \quad (1)$$

where $P_1 = P_0 + \Delta_p P_{\text{max}}$. The scaling parameter χ is the load share, where $\chi = 1$ indicates a fully loaded system, e.g. transmitting at full power and full bandwidth, and $\chi = 0$ indicates an idle system. To further understand the contribution of different parameters on this basic model, we parameterize the maximum supply power consumption, P_1 . We first establish how power consumption scales with the transmission bandwidth in Hz, W , the number of BS radio chains/antennas, D , and the maximum transmission power in W, P_{max} . This requires to consider the main units of a BS: PA, radio frequency (RF) small-signal transceiver, baseband (BB) engine, direct-current (DC)-DC converter, active cooling and mains supply (MS). The dependence of the BS units on W , D and P_{max} can be approximated as follows [13]:

- Both the power consumptions of BB and RF, P_{BB} and P_{RF} , respectively, scale linearly with bandwidth, W , in and the number of BS antennas D . For some basic consumptions, P'_{BB} and P'_{RF} , we thus define

$$P_{\text{BB}} = D \frac{W}{10 \text{ MHz}} P'_{\text{BB}} \quad (2)$$

and

$$P_{\text{RF}} = D \frac{W}{10 \text{ MHz}} P'_{\text{RF}}. \quad (3)$$

- The PA power consumption P_{PA} depends on the maximum transmission power per antenna P_{max}/D and the PA efficiency η_{PA} . Also, possible feeder cable losses, σ_{feed} , have to be accounted for:

$$P_{\text{PA}} = \frac{P_{\text{max}}}{D \eta_{\text{PA}} (1 - \sigma_{\text{feed}})}. \quad (4)$$

- Losses incurred by DC-DC conversion, MS and active cooling scale linearly with the power consumption of other components and may be approximated by the loss factors σ_{DC} , σ_{MS} , and σ_{cool} , respectively. These losses are included in the model as losses of a total according to [11]. Active cooling is typically only applied in Macro type BSs.

These assumptions are combined to calculate the maximum power consumption of a BS sector,

$$P_1 = \frac{P_{\text{BB}} + P_{\text{RF}} + P_{\text{PA}}}{(1 - \sigma_{\text{DC}})(1 - \sigma_{\text{MS}})(1 - \sigma_{\text{cool}})} \quad (5a)$$

$$= \frac{D \frac{W}{10 \text{ MHz}} (P'_{\text{BB}} + P'_{\text{RF}}) + \frac{P_{\text{max}}}{D \eta_{\text{PA}} (1 - \sigma_{\text{feed}})}}{(1 - \sigma_{\text{DC}})(1 - \sigma_{\text{MS}})(1 - \sigma_{\text{cool}})}. \quad (5b)$$

An important characteristic of a PA is that operation at lower transmit powers reduces the efficiency of the PA and that, consequently, power consumption is not a linear function of the PA output power. This is resolved by taking into account the ratio of maximum transmission power of the PA from the data sheet, $P_{\text{PA,limit}}$ to the maximum transmission power of the PA during operation $\frac{P_{\text{max}}}{D}$. The current transmission power can be adjusted by adapting the DC supply voltage, which impacts the offset power of the PA. The efficiency is assumed to decrease by a factor of γ for each halving of the transmission power. The efficiency is thus maximal when $P_{\text{max}} = P_{\text{PA,limit}}$ in single antenna transmission and was heuristically found to be well-described by

$$\eta_{\text{PA}} = \eta_{\text{PA,max}} \left[1 - \gamma \log_2 \left(\frac{P_{\text{PA,limit}}}{P_{\text{max}}/D} \right) \right], \quad (6)$$

where $\eta_{\text{PA,max}}$ is the maximum PA efficiency.

The reduction of power consumption during sleep modes is achieved by powering off PAs and reduced computations necessary in the BB engine. For simplicity, we only model the dependence on D as each PA is powered off. Thus, P_{sleep} , is approximated as

$$P_{\text{sleep}} = D P_{\text{sleep},0}, \quad (7)$$

where $P_{\text{sleep},0}$ is a reference value for the single antenna BS chosen such that P_{sleep} matches the complex model value for two antennas.

IV. RESULTS AND DISCUSSION

The parameterized power model is applied to approximate the consumption of the Macro, Pico and Femto BSs which are described in the complex model. Parameters are chosen where possible according to [11], such as losses, efficiencies and power limits. The remaining parameters are adapted such that a closer match to the complex model could be achieved. The resulting parameter breakdown is provided in Table I. The proposed and the complex power models are compared for a bandwidth sweep with a varying number of transmit antennas in Fig. 2, Fig. 3, and Fig. 4, for a Macro, a Pico and a Femto station, respectively. Although two parameters, the bandwidth and the number of BS antennas, are varied, the parameterized model can be seen to closely approximate the complex model for all BS types. The largest deviation of the parameterized model from the complex model occurs when modeling four transmit antennas. This is caused by the fact that the parameterized model considers a constant slope, Δ_p , which is independent of D . In contrast, the PA efficiency in the complex model decreases with rising D , leading to an increasing slope which can not be matched by a constant slope. This deviation is a trade-off between simplicity and model accuracy.

BS type	$P_{PA,limit}$ /W	$\eta_{PA,max}$	γ	P_{BB} /W	P_{RF} /W	σ_{feed}	σ_{DC}	σ_{cool}	σ_{MS}	M_{Sec}	P_{max} /W	P_1 /W	Δ_p *10 MHz	$P_{sleep,0}$ /W
Macro	80.00	0.36	0.15	29.4	12.9	0.5	0.075	0.1	0.09	3	40.00	460.4	4.2	324.0
Pico	0.25	0.08	0.20	4.0	1.2	0.0	0.09	0.0	0.11	1	0.25	17.4	4.0	4.9
Femto	0.10	0.05	0.10	2.5	0.6	0.0	0.09	0.0	0.11	1	0.10	12.0	4.0	3.3

TABLE I
PARAMETER BREAKDOWN.

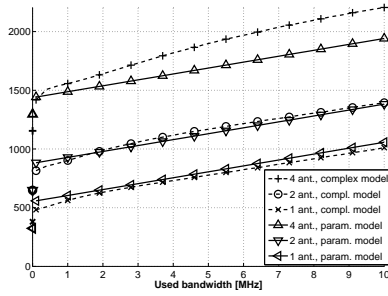


Fig. 2. Comparison of the parameterized with the complex model [13] power models for the Macro BS type with 40 W transmission power.

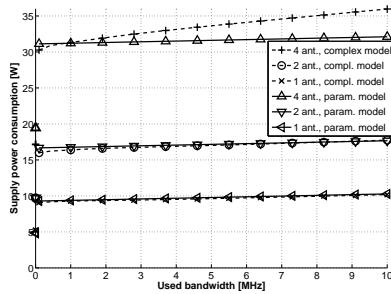


Fig. 3. Comparison of the parameterized with the complex model [13] power models for the Pico BS type with 0.25 W transmission power.

In addition to providing a solid reference, the model and the parameters can provide a basis for exploration. Individual parameters can be changed to observe the resulting variation in power consumption. With regard to the number of antennas, the parameterized model can only be verified up to four antennas, which is the extent of the complex model. Extending the system bandwidth, for example to 20 MHz, is expected to increase the BB and RF power consumption. The other parameters such as the transmission power and losses are expected to remain unaffected by different system bandwidths. Adapting the design maximum transmission power, P_{max} , affects the PA efficiencies, which decrease with P_{max} .

V. CONCLUSION AND REMARKS

In this paper, we have provided a parameterized power model which allows calculating the power consumption of a modern BS based on important design and operation param-

eters. The model is much simpler and more applicable than the

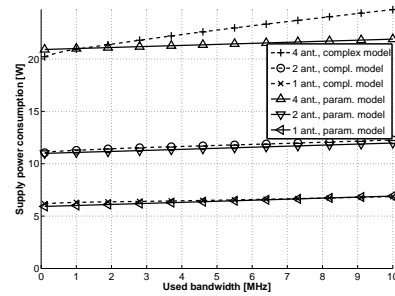


Fig. 4. Comparison of the parameterized with the complex model [13] power models for the Femto BS type with 0.1 W transmission power.

model it was derived from. A comparison of the parameterized model with the source model is provided.

REFERENCES

- [1] 3GPP TR 36.927, "Evolved Universal Terrestrial Radio Access (E-UTRA); Potential solutions for energy saving for E-UTRAN," Sep. 2012.
- [2] A. Fehske, J. Malmolin, G. Biczok, and G. Fettweis, "The Global Carbon Footprint of Mobile Communications - The Ecological and Economic Perspective," *IEEE Communications Magazine*, 2010.
- [3] H. Claussen, L. Ho, and F. Pivit, "Effects of Joint Macrocell and Residential Picocell Deployment on the Network Energy Efficiency," in *PIMRC 2008*, 2008, pp. 1-6.
- [4] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-Efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks," *IEEE Journal on Selected Areas in Communications*, 2004.
- [5] H. Holtkamp, G. Auer, and H. Haas, "On Minimizing Base Station Power Consumption," in *Proceedings of the IEEE VTC 2011-Fall*, 2011.
- [6] Z. Xu, C. Yang, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-Efficient MIMO-OFDMA Systems based on Switching Off RF Chains," in *Proceedings of the VTC Fall 2011*, 2011.
- [7] A. Ambrosy, M. Wilhelm, O. Blume, and W. Wajda, "Dynamic Bandwidth Management for Energy Savings in Wireless Base Stations," in *Proceedings of Globecom 2012*, 2012.
- [8] P. Frenger, P. Moberg, J. Malmolin, Y. Jading, and I. Gódor, "Reducing Energy Consumption in LTE with Cell DTX," in *Proceedings of the IEEE VTC 2011-Spring*, 2011.
- [9] M. Deruyck, W. Joseph, and L. Martens, "Power Consumption Model for Macrocell and Microcell Base Stations," *Transactions on Emerging Telecommunications Technologies*, pp. n/a-n/a, 2012.
- [10] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power Consumption Modeling of Different Base Station Types in Heterogeneous Cellular Networks," in *Future Network and Mobile Summit*, 2010.
- [11] Gunther Auer *et al.*, "How Much Energy is Needed to Run a Wireless Network?" *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40-49, Oct 2011.
- [12] G. Auer, V. Giannini, I. Gódor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. J. Gonzalez, and C. Desset, "Cellular Energy Efficiency Evaluation Framework," in *Proceedings of the VTC 2011-Spring*, 2011.
- [13] Claude Desset *et al.*, "Flexible Power Modeling of LTE Base Stations," in *IEEE WCNC 2012*, 2012.
- [14] A. Fehske, F. Richter, and G. Fettweis, "Energy efficiency improvements through micro sites in cellular mobile radio networks," in *GLOBECOM Workshops, 2009 IEEE*, 30 2009-dec. 4 2009, pp. 1-5.

Distributed DTX Alignment with Memory

Hauke Holtkamp* and Guido Dietl*[‡]

*DOCOMO Euro-Labs

[‡]University of Applied Sciences Landshut

Email: {holtkamp,dietl}@docomolab-euro.com

Harald Haas

Institute for Digital Communications

Joint Research Institute for Signal and Image Processing

The University of Edinburgh, EH9 3JL, Edinburgh, UK

Email: h.haas@ed.ac.uk

Abstract—This paper addresses the assignment of transmission and sleep time slots between interfering transmitters with the objective of minimal power consumption. In particular, we address the constructive alignment of Discontinuous Transmission (DTX) time slots under link rate constraints. Due to the complexity of the combinatorial optimization problem at hand, we resort to heuristic assignment strategies. We derive four time slot alignment solutions (sequential alignment, random alignment, p-persistent ranking and DTX alignment with memory) and identify trade-offs. One solution, DTX alignment with memory, addresses trade-offs of the other three by maintaining memory of past alignment and channel quality to buffer short term changes in channel quality. All strategies are found to exhibit similar convergence behavior, but different power consumption and retransmission probabilities. DTX alignment with memory is shown to achieve up to 40% savings in power consumption and more than 20% lower retransmission probability than the State-Of-The-Art (SotA).

I. INTRODUCTION

To reduce their power consumption, future Base Stations (BSs) will use very short sleep modes—called Discontinuous Transmission (DTX)—which interrupt their transmission [1]. DTX can be employed for energy saving when a communication system has spare capacity, *e.g.* at night in cellular networks. It can be easily integrated into Orthogonal Frequency Division Multiple Access (OFDMA) systems without incurring delays by not scheduling resources for transmission during one or more time slots and then setting the transmitter to DTX mode instead for the duration of these unscheduled time slots. For the transmitter, this is a local power saving measure. When seen from a network perspective, these regular interruptions in transmission can be aligned to reduce interference and power consumption. Such alignment has not been studied in detail, but is related to the field of channel allocation.

Generally, in channel allocation, the challenge is to assign frequency channels to different cells or links of

a network such that mutual interference is minimized. To be applicable for future small-cell cellular networks, such channel allocation needs to be distributed, uncoordinated and dynamic. Distributed operation prevents the delay and backhaul requirements imposed by a central controller. If the allocation is also uncoordinated, it does not require message exchange through a backhaul of limited capacity. Also, as networks, channels and traffic change rapidly, the allocation must be highly dynamic and update often. Problems of this type are known to be NP-hard [2]. Therefore, different suboptimal approaches have been proposed. The simplest approach to distributing channels over a network is a fixed assignment in which it is predefined which links will use which channels, *e.g.* see [3]. However, this is inflexible to changing or asymmetric traffic loads. For flexibility, dynamic channel allocation methods are required. A simple dynamic channel allocation method, Sequential Channel Search [4], assigns channels with sufficient quality in a predefined order. This technique allows adjusting the number of channels flexibly, but is suboptimal due to its strong channel overlap between neighbors. Taking the channel quality into consideration by measurement is proposed in the Minimum Signal-to-Interference-and-Noise-Ratio (SINR) method [4]. This technique offers increased spectral efficiency, but can lead to instabilities when allocations happen synchronously. Ellenbeck *et al.* [5] introduce methods of game theory and respond to the instabilities by adding p-persistence to avoid simultaneous bad player decisions. However, the proposed method only applies to single user systems and does not address target rates. Dynamic Channel Segregation [6] introduces the notion of memory of channel availability, allowing a transmitter to track which channels tend to be favorable for transmission. However, the algorithm only applies to sequential channel decisions based on an idle or busy state in circuit switched networks and can not be applied to the concurrent alignment of channels as is required in OFDMA networks.

In this paper, we combine the findings from past channel allocation research and adapt them for aligning DTX time slots in an OFDMA frame. We derive four distributed uncoordinated dynamic time slot alignment strategies for the cellular downlink, in which BSs independently prioritize time slots for transmission while maintaining system stability. The remainder of the paper is structured as follows. Section II formulates the system model and the problem at hand. The considered four alternative solutions are described in Section III. Findings obtained from simulation are presented in Section IV. The paper is concluded in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We make the following assumptions about the network. BSs in a reuse-one OFDMA cellular network schedule a target rate, B_k , per OFDMA frame to be transmitted to each mobile $k = \{1, \dots, K\}$. All time slots are available for scheduling in all cells. Mobiles report perceived SINR from the previous OFDMA frame, $s_{n,t,k}$, on subcarrier $n = \{1, \dots, N\}$, time slot $t = \{1, \dots, T\}$ to their associated BS. BSs have a DTX mode available for each time slot during which transmission and reception are disabled and power consumption is significantly reduced to P_S compared to power consumption during transmission, which is a function of the power allocated to each Resource Block (RB), ρ_{Tx} . DTX is available fast enough to enable it within individual time slots of an OFDMA transmission such that transmission time slots are not required to be consecutive. A BS can schedule no transmission in one or more time slots and go to DTX mode instead. Scheduling a DTX time slot in one BS reduces the interference on that time slot to all other BSs. The OFDMA frames of all BSs are assumed to be aligned such that the interference over one time slot and subcarrier is flat and that all BSs can perform alignment operations in synchrony. The OFDMA frame consists of NT RBs. The channel is subject to block fading. Interference is treated as noise.

To describe the resource allocation problem formally, we define the function

$$\Pi : \{1, \dots, T\} \times \{1, \dots, N\} \rightarrow \{0, 1, \dots, K\} \quad (1)$$

$$(n, t) \mapsto k,$$

which maps to each RB a user k or 0, where 0 indicates that the RB is not scheduled.

Each cell tries to minimize its power consumption by scheduling time slots for transmission and sleep. With $r_{n,t,k}$ the capacity of resource (n, t) if it is scheduled

to k , the optimization problem for the total BS power consumption, P_{total} , is as follows.

$$\underset{\Pi}{\text{minimize}} \quad P_{\text{total}} = P_S \frac{T_S}{T} + \rho_{Tx} N_{Tx} \left(\frac{T - T_S}{T} \right) \quad (2a)$$

subject to

$$N_{Tx} = \quad (2b)$$

$$|\{n \in \{1, \dots, N\} | (\Pi(n, t) \neq 0 \quad \forall t \in \{1, \dots, T\})\}|$$

$$T_S = \quad (2c)$$

$$|\{t \in \{1, \dots, T\} | (\Pi(n, t) = 0 \quad \forall n \in \{1, \dots, N\})\}|$$

$$B_k \leq \sum_{\{(n,t)|\Pi(n,t)=k\}} r_{n,t,k} \quad \forall k. \quad (2d)$$

With $|\cdot|$ the cardinality operator, T_S is the number of time slots in which for all n , no user is allocated and the BS can enable DTX and N_{Tx} is the number of RBs that are scheduled for transmission. The constraint (2d) provides the rate guarantee for each user.

The difficulty lies in finding the mapping Π . Under a brute force approach, there exist $(K+1)^{NT}$ possible combinations. Due to the large number RBs present in typical OFDMA systems like Long Term Evolution (LTE), the computation of the solution is infeasible. Consequently, we resort to heuristic methods in the next section.

Some of the methods compared in this paper make use of a time slot ranking. In other works, *e.g.* [6], ranking is proposed to be based on SINR. However, in a multi-user OFDMA system each subcarrier and mobile terminal have a different SINR, thus generating a problem of comparability between time slots. Consequently, we propose to compare time slots by their hypothetical sum capacity, B_t , over all mobiles and subcarriers with

$$B_t = \sum_{k=1}^K \sum_{n=1}^N \log_2(1 + s_{n,t,k}). \quad (3)$$

III. DTX ALIGNMENT STRATEGIES

In this section, we derive four strategies to tackle the problem at hand which differ in performance and complexity.

A. Sequential alignment

This strategy is to always allocate as many time slots for transmission as required in sequential order and set the remainder to DTX. This leads to strongest overlap and consequently highest interference on the first time slot and lowest overlap and possibly no transmissions at

all on the last time slot. This strategy does not make use of the available channel quality information and provides valuable insights into questions of stability, reliability and convergence. It serves as a deterministic upper bound.

B. Random alignment

Random alignment refers to randomly selecting transmission time slots for every OFDMA frame and setting the remainder to DTX. This strategy provides a reference for achievable gains, allows to assess the worst cast effects of randomness and represents the state of the art in today's unsynchronized unaligned networks.

C. P-persistent ranking

The synchronous alignment of uncoordinated BSs can lead to instabilities, when neighboring BSs—perceiving similar interference information—schedule the same time slots for transmission. This leads to oscillating scheduling which never reaches the desired system state [5]. To address this problem, we introduce p-persistence to break the unwanted synchrony by only changing established DTX schedules with a probability $p = 0.3$. In initial tests, the value of p did not have a significant effect on the power consumption. The exact tuning of p is out of the scope this paper. P-persistent ranking first ranks time slots by their sum capacity as in (3), schedules time slots in that order, but only applies this new selection with probability p . Otherwise, the schedule from the last iteration remains active. After ranking, time slots are selected for transmission in order of B_t . The remainder of time slots is set to DTX.

D. Distributed DTX alignment with memory

To counter oscillation and achieve a convergent network state we introduce memory of past schedules into the alignment process. The rationale behind the distributed DTX alignment with memory algorithm is as follows. Taking into account current time slot capacities, B_t , past scores and slot allocations, each BS first updates the internal score of each time slot and then returns the priority of time slots by score. In case of equal scores, time slots are further sub-sorted by B_t . The score is updated in integers. All time slots which were used for transmission in the previous OFDMA frame receive a score increment of one. Furthermore, the time slot with highest B_t , R_0 , receives an increment of one. Time slots which were not used for transmission in the previous OFDMA frame receive a decrement of one, except for R_0 . Scores have an upper limit ψ_{ul} beyond which there

is no increment and a lower limit ψ_{ll} below which there is no decrement. The difference between ψ_{ul} and ψ_{ll} can be interpreted as the depth of the memory buffer.

Algorithm 1 Distributed DTX alignment with memory

Input: ψ, Υ_u

- 1: $R \leftarrow \text{sort-desc-by-capacity}(\Upsilon)$
- 2: **for all** v in Υ_u **do**
- 3: **if** $\psi(v) < \psi_{ul}$ **then**
- 4: $\psi(v) \leftarrow \psi(v) + 1$
- 5: **end if**
- 6: **end for**
- 7: **for all** v in $\Upsilon_{uu} \setminus \{R_0\}$ **do**
- 8: **if** $\psi(v) > \psi_{ll}$ **then**
- 9: $\psi(v) \leftarrow \psi(v) - 1$
- 10: **end if**
- 11: **end for**
- 12: $\psi(R_0) \leftarrow \psi(R_0) + 1$
- 13: $V \leftarrow \text{sort-desc-by-score}(R, \psi, \Upsilon)$
- 14: **return** ψ, V, Υ_u

The described steps are summarized in Algorithm 1 with scoring map ψ , ranking tuple R , and priority tuple V . The set of used time slots Υ_u and unused time slots Υ_{uu} make up the set of time slots Υ , the cardinality of which is T . The function 'sort-desc-by-capacity(Υ)' returns a list of time slots ordered descending by B_t . The function 'sort-desc-by-score(R, ψ, Υ)' returns a list of time slots ordered primarily by descending score and secondarily by descending B_t . After the execution of Algorithm 1, the first T_{Tx} time slots in V are scheduled for transmission.

In the following, we illustrate the iterations over three OFDMA frames of Algorithm 1 for a system with three time slots. In the example, the algorithm delays the change of Υ_u from c to b to buffer scheduling changes. The iteration begins with arbitrarily chosen $\Upsilon = \{a, b, c\}$, $\psi = \{a : 0, b : 2, c : 5\}$:

- 1) $\Upsilon_u = \{c\}$, $R = (b, c, a)$
 $\rightarrow \psi = \{a : 0, b : 3, c : 5\}$, $V = (c, b, a)$
- 2) $\Upsilon_u = \{b, c\}$, $R = (b, c, a)$
 $\rightarrow \psi = \{a : 0, b : 5, c : 5\}$, $V = (b, c, a)$
- 3) $\Upsilon_u = \{b\}$, $R = (b, a, c)$
 $\rightarrow \psi = \{a : 0, b : 5, c : 4\}$, $V = (b, c, a)$

When applied, Algorithm 1 strongly benefits time slots which were used for transmission in the past (b, c in step 1 of the example). These time slots tend to repeatedly receive score increments until they all have maximum score ψ_{ul} (b, c in step 2 of the example). When the

score is equal for some time slots (b, c in step 2 of the example), the ranking is based on B_t . A time slot which was used for transmission and has highest B_t, R_0 , receives the highest increment (slot b in step 2 of the example). When a time slot has highest B_t , but was not selected for transmission in the previous OFDMA frame, it receives a score increment, but is not guaranteed to be used for transmission (slot b in step 1 of the example). When a time slot repeatedly has highest B_t , it reaches ψ_{ul} (b in the example). The algorithm thus buffers short term changes in the channel quality setting in favor of long term time slot selection.

IV. RESULTS

We analyze the four alignment strategies using computer simulations with regard to power consumption, convergence, reliability of delivered rates and algorithmic complexity after the introduction of the simulation environment and the RB scheduling scheme.

A. Simulation environment and RB scheduling

The four strategies were tested in a network simulation with 19-cell hexagonal arrangement with uniformly distributed mobiles and fixed target rates per mobile. Data was collected only from the center cell which is thus surrounded by two tiers of interfering cells. Power consumption of a cell is modeled as a function of transmission power as described in [7]. Table IV-A lists additional parameters used which approximate an LTE system. The simulation is started with the assumption of full transmission power on all resources with a power consumption of 350 W per cell as a worst-case initial configuration.

In order to assess the performance of the four strategies, it is necessary to make assumptions about how the individual RBs are scheduled within a time slot. In order to avoid masking effects of the algorithms under test, we have applied a sequential RB allocation. Sequential resource allocation is performed after the DTX alignment step is completed and allocates as many bits to an OFDMA resource block as possible according to the Shannon capacity, followed in order by the next resource in the same time slot (sequentially), until the target rate has been scheduled to each mobile. Time slots are scheduled in the order provided by each of the four strategies. This sequential resource scheduler deliberately omits the benefits of multi-user diversity. This leads to underestimating achievable rates in simulation compared to a system which exploits multi-user diversity,

Parameter	Value
Carrier frequency	2 GHz
Intersite distance	500 m
Pathloss model	3GPP UMa, NLOS, shadowing [8]
Shadowing standard deviation	8 dB
Bandwidth	10 MHz
Transmission power per RB	0.8 W
Thermal noise temperature	290 K
Interference tiers	2 (19 cells)
Mobile target rate	2 Mbps and 3 Mbps
OFDMA subframes (time slots)	10
Subcarriers	50
Mobiles	10
ψ_{ul}	5
ψ_{ll}	0
Power model factors [7] (idle; load factor; DTX)	200 W; 3.75; 90 W

TABLE I
SIMULATION PARAMETERS

but allows a fair comparison of the quality of different DTX alignment algorithms.

B. Power consumption

To assess achievable power savings and the dynamic adaptivity over a large range of cell loads, we inspect the cell total power consumption in Fig. 1. At low load very few RBs are required to deliver the target rate and more time slots can be scheduled for DTX than at high traffic loads, leading to monotonously rising power consumption over increasing target rates for all alignment methods.

Sequential alignment causes the highest power consumption over any target rate with an almost linear relationship between user target rates and power consumption. Sequential alignment consumes high power, as it schedules many RBs to achieve the target rate due to the high interference level present.

This power consumption is significantly lower for the State-Of-The-Art (SotA), random alignment. The randomness of time slot alignment creates a much lower average interference than with sequential alignment allowing more data to be transmitted in each RB. As fewer RBs are required, less power is consumed.

P-persistent ranking and DTX with memory both achieve similarly low power consumption of up to 40% less than random alignment. The relationship between power consumption and target rate is noticeably non-linear, as it is flat at low target rates and grows more steeply at high target rates. This behavior is caused by the low interference level these strategies manage to create. Only at high rates, when the number of sleep time slots

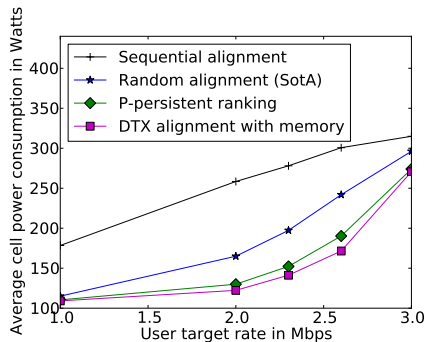


Fig. 1. BS power consumption over different cell sum rates.

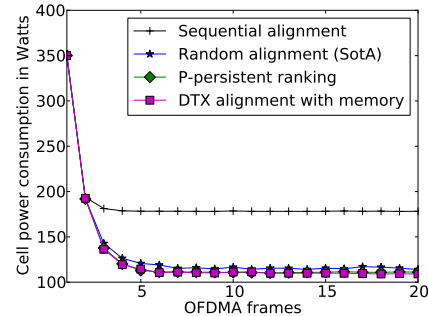


Fig. 2. BS power consumption over OFDMA frames at 1 Mbps per mobile.

has to be significantly decreased, does the interference increase, leading to higher power consumption.

Also noteworthy is the fact that at 1 Mbps and 3 Mbps, random alignment performs nearly as good as p-persistent ranking. At these extreme points the network is almost unloaded and almost fully loaded, respectively. Consequently, either most time slots are scheduled for DTX or none, leaving very little room for optimization compared to randomness. The largest potential for time slot alignment for power saving is in networks which are medium loaded. Under medium load, the number of transmission and DTX time slots is similar, causing the effects of alignment to be most pronounced.

C. Convergence

Another relevant aspect is the convergence of the network to a stable state. As each BS makes iterative adjustments to its selection of time slots for transmission, the speed of convergence as well as the convergence to a stable point of operation are relevant criteria. The effect of the iterative execution of the four strategies on the cell power consumption is illustrated in Fig. 2. Power consumption is found to converge to a stable value within within six OFDMA frames (alignment iterations). All strategies converge to the average power consumption values shown in Fig. 1. The simulation starts from a worst-case schedule of transmission on all RBs and then iteratively schedules time slots for transmission and DTX. In the case of 1 Mbps per user, one transmission time slot is sufficient to schedule the target rate. With transmissions only taking place during one time slot, there is very little difference between a random alignment and p-persistent ranking or DTX with memory. At 2 Mbps per user, see Fig. 3, random alignment occasionally causes higher interference than p-persistent ranking or DTX with memory, leading to higher power

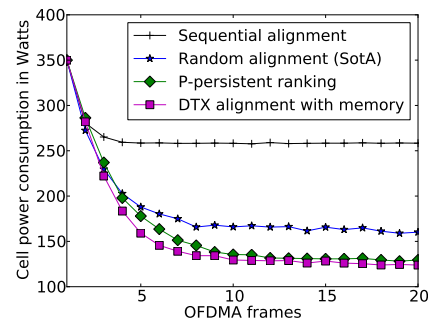


Fig. 3. BS power consumption over OFDMA frames at 2 Mbps per mobile.

consumption. Also, p-persistent ranking converges more slowly than DTX alignment with memory.

D. Reliability

An important aspect in dynamical systems with target rates is that scheduled target rates cannot always be fulfilled. As RB scheduling is based on channel quality information which was collected in the previous OFDMA frame, the actual channel quality during transmission may differ, leading to lower than expected rates. Thus, although certain target rates are scheduled and although the system is not fully loaded, a BS may fail to deliver the targeted rate and require retransmission of some RBs. We assess this metric by considering the retransmission probability for each strategy. The results are shown in Fig. 4 against a range of user target rates.

Easiest to interpret is sequential alignment which does not require retransmission for target rates up to 2.3 Mbps due to its determinism. The increase in the retransmission probability at high rates is not caused by a failure of the alignment, but by system overload. When

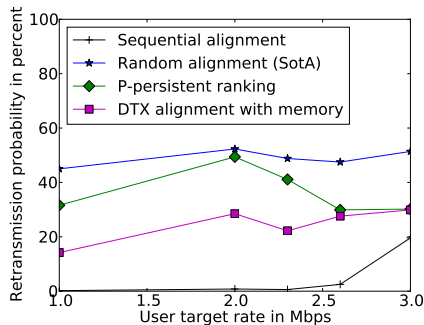


Fig. 4. Retransmission probability over targeted rate.

high rates are combined with bad channel conditions, the system may be unable to deliver the target data rate, independent of the alignment strategy. This increase of the retransmission probability in user satisfaction at high rates is present for all alignment strategies and constitutes outage.

The retransmission probability is highest for random alignment. This is caused by the strong fluctuation of interference under randomized scheduling. Channel quality measurements used for RB scheduling are of very little reliability, as the interference changes quickly due to random time slot alignment, causing low user satisfaction.

P-persistent ranking performs slightly better than random alignment at 1 Mbps target rate. At 2 Mbps per user, where the alignment potential is highest, oscillation causes the highest retransmission probability.

DTX alignment with memory achieves a much lower retransmission probability in the range of 15% to 20%, due to the reduction in interference fluctuation introduced by the memory score.

E. Complexity

With regard to complexity, sequential and random alignment are of minimal complexity. These strategies involve no algorithmic decision-making on the set of transmission time slots. P-persistent ranking requires one ranking and time slot selection with probability p per iteration. The highest complexity is present in DTX with memory, which requires two executions of the time slot sort, one for the generation of the score and one for the output of the ranking. Although DTX with memory comprises the highest complexity in comparison, these operations pose a small burden on modern hardware as the number of time slots is typically small. For example, typical LTE systems are designed with 10 subframes.

F. Interpretation

To conclude our evaluation, we have found that DTX with memory provides the best results. Under the present assumptions, it provides both a lower power consumption and lower retransmission probability than the SotA and p-persistent ranking. Future DTX capable networks can and should exploit this alignment potential.

V. CONCLUSION

In this paper, we have studied the constructive uncoordinated and distributed alignment of DTX time slots between interfering OFDMA BSs for power-saving under rate constraints. We first established the open problem formally. Due to its complexity, we have approached it with four alternative heuristic strategies. One of these strategies, DTX with memory, is an original contribution of this paper and introduces memory to overcome short-term fluctuations and networks oscillations. The performance of the four strategies was tested in simulation. It was found that power consumption can be reduced significantly, especially at medium cell loads. All strategies converge within six OFDMA frames. DTX with memory reduces power consumption up to 40% compared to the SotA, combined with around 20% reduction in resubmission probability.

REFERENCES

- [1] P. Frenger, P. Moberg, J. Malmodin, Y. Jading, and I. Gódor, "Reducing Energy Consumption in LTE with Cell DTX," in *Proceedings of the IEEE VTC 2011-Spring*, 2011.
- [2] I. Katzela and M. Naghshineh, "Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey," *Personal Communications, IEEE*, vol. 3, no. 3, pp. 10–31, Jun 1996.
- [3] J. G. Proakis, *Digital Communications*, 4th ed., ser. McGraw-Hill Series in Electrical and Computer Engineering. McGraw-Hill Higher Education, Dec. 2000.
- [4] M. Serizawa and D. Goodman, "Instability and Deadlock of Distributed Dynamic Channel Allocation," in *Vehicular Technology Conference, 1993., 43rd IEEE*, May 1993, pp. 528–531.
- [5] J. Ellenbeck, C. Hartmann, and L. Berlemann, "Decentralized Inter-Cell Interference Coordination by Autonomous Spectral Reuse Decisions," in *Proc. of the 14th European Wireless Conference (EW)*, Prague, Czech Republic, Jun. 22–25 2008, pp. 1–7.
- [6] Y. Akaiwa and H. Andoh, "Channel Segregation – A Self-organized Dynamic Channel Allocation Method: Application to TDMA/FDMA Microcellular System," *Selected Areas in Communications, IEEE Journal on*, vol. 11, no. 6, pp. 949–954, Aug. 1993.
- [7] C. Desset *et al.*, "Flexible Power Modeling of LTE Base Stations," in *IEEE WCNC 2012*, 2012.
- [8] 3GPP, "Further Advancements for E-UTRA Physical Layer Aspects (Release 9), TR 36.814 V9.0.0," www.3gpp.org/ftp/Specs/, Mar. 2010, retrieved 15 Jan. 2013.

OFDMA Base Station Power-saving Via Joint Power Control and DTX in Cellular Systems

Hauke Holtkamp

DOCOMO Euro-Labs
D-80687 Munich, Germany
E-mail: h.holtkamp@gmail.com

Harald Haas

Institute for Digital Communications
Joint Research Institute for Signal and Image Processing
The University of Edinburgh, EH9 3JL, Edinburgh, UK
E-mail: h.haas@ed.ac.uk

Abstract—In this paper, we extend the previously proposed power-saving Resource Allocation using Power Control and Sleep (RAPS) algorithm to the interference environment. The RAPS algorithm allocates resources, power and Discontinuous Transmission (DTX) time slots jointly for Orthogonal Frequency Division Multiple Access (OFDMA) downlink resource scheduling such that Base Station (BS) supply power consumption is reduced while providing guaranteed link rates. The RAPS scheduler was previously proposed and studied only as a single-cell mechanism. It is here adapted to and tested in the multi-cell interference setting. The resulting dynamical system is analyzed with respect to convergence, power consumption and load dependence. It is found that the Signal-to-Interference-and-Noise-Ratio (SINR) distribution in the network is greatly improved and average power consumption can be reduced by around 50% in comparison to a Proportional-Fair (PF) scheduling benchmark.

I. INTRODUCTION

With the increasing number and density of cellular BSs, power consumption is becoming a growing concern for network operators [1]. Rising energy costs, the desire for better disaster recovery and the reduction of CO₂ emissions all motivate more energy efficient BSs [2]. While energy efficiency is highly relevant, it still ranks behind capacity and reliability in operator priorities. Thus, for an operator, it is not acceptable to increase energy efficiency by degrading user Quality of Service (QoS). However, studies have shown that in many locations and during a significant share of operation time, traffic load is far below cell peak capacity [1]. This opens the opportunity to deliberately reduce cell capacities as long as all user demands can still be fulfilled, thus gaining energy efficiency without affecting the QoS. Exploiting this provisioning gap is the tool for power-saving Radio Resource Management (RRM) mechanisms.

Energy efficient RRM can broadly be divided into two approaches, namely saving through sleep mode and saving through transmission Power Control (PC). Sleep mode approaches can further be separated into long sleep and micro sleep. Long sleep refers to turning devices off completely for durations of minutes and more. This clearly saves most energy as devices' power consumption can possibly be reduced to zero, but creates issues with regard to QoS, coverage and wake-up triggering. In contrast, micro sleep operates on time scales of milliseconds and can be arranged to be transparent

to User Equipments (UEs). Micro sleep is also referred to as DTX because it is more closely related to a short interruption in transmission rather than a deep sleep state. A device in DTX will not turn off entirely and remain available to continue transmission, but has significantly reduced (although not zero) power consumption. The alternative approach, PC, is directly linked to the allocation of subcarriers or Resource Blocks (RBs) when considering OFDMA transmission systems. Most works on OFDMA PC, including this paper, therefore also include resource allocation mechanisms. As we specifically aim at maintaining the QoS, long sleep is not considered in this work.

For a comparison with the state of the art, we proceed to present a literature overview on DTX, PC and resource allocation: DTX in BSs is still a relatively new topic. Frenger et al. [3] propose to use Long Term Evolution (LTE) Multicast-Broadcast Single Frequency Network (MBSFN) subframes to allow DTX in current standards, but do not show how resources should be scheduled to align transmission with the subframes. Abdallah et al. [4] align DTX time slots via a central controller and find that orthogonal time slot assignment to neighboring cells is desirable in terms of rate maximization. In comparison to DTX, PC literature has a longer history as it is also beneficial for link adaptation and interference mitigation. Wong et al. [5] proposed the first, but highly complex transmission power minimizing OFDMA resource and power allocation method. Miao et al. [6] maximize energy efficiency on an individual link while considering circuit consumption. Videv et al. [7] spread signals over unused bandwidth and thus achieve a lower transmission power at unchanged rates. Xu et al. [8] jointly allocate subcarriers, transmission power and antenna chains to minimize device consumption. López-Pérez et al. [9] reduce interference and transmission power via PC and resource allocation, but do not consider DTX. The only consideration of both sleep modes and PC is presented by Domenico [10] who divides traffic into delay tolerant and real-time traffic. Under real-time traffic, transmission power is reduced by bandwidth spreading while under delay-tolerant traffic, the device is put into sleep mode.

In [11], we have recently presented the RAPS algorithm, which allocates OFDMA resources, transmission power and DTX time slots in a multi-user cell under rate guarantees at

reduced device power consumption. This work is extended here to the multi-cell interference system with the following contributions:

- Allocation decisions are based on the RB SINR rather than the Channel State Information (CSI).
- We include an iterative technique which continuously adapts to the dynamical system in contrast to the single iteration previously proposed.
- We derive a power-saving adaptation of a PF scheduler as the state-of-the-art benchmark.
- The effect of power-saving resource allocation is heuristically analyzed in terms of convergence, power consumption, SINR distribution and energy-per-bit and benchmarked against the comparable PF resource allocation mechanism.

The paper is organized as follows. The system model, the multi-cell extensions and the benchmarks are presented in Section II. Simulation results are presented in Section III. The paper is concluded in Section IV.

II. SYSTEM MODEL

The RAPS algorithm [11] reduces power consumption via resource, power and DTX time slot allocation in a cell with multiple UEs while providing a guaranteed rate, R_k , to each UE k over each OFDMA frame based on the reported RB CSI. The reduction in supply power consumption is achieved by first computing the joint PC and DTX solution to a real-valued and convex sub-problem and then quantizing the solution to find the integer-valued RB allocation. The allocation is based on CSI values on each RB as reported from the UEs. See Fig. 1 for a block diagram of the RAPS algorithm.

We make the following assumptions about the network. A transmission power limit applies in each time slot, which is determined by the BS hardware. Each RB can only be allocated to one user, *i.e.* there is no RB sharing. The weighting of the decision on the number of DTX time slots is made on the basis of a linear power model [12]. The linear power model approximates the supply power consumption of LTE BSs as an affine function of the transmission power and provides a reduced power consumption value during DTX, P_S .

Assuming flat-fading channels for each user in a Time Division Multiple Access (TDMA) transmission system, a convex optimization problem is used to approximate real-valued resource shares, μ_k , as part of the RAPS algorithm:

$$\begin{aligned}
 & \underset{(\mu_1, \dots, \mu_{K+1})}{\text{minimize}} && P_{\text{supply}}(\mathbf{r}) = \sum_{k=1}^K \mu_k (P_0 + \Delta_{\text{PM}} P_k(R_k)) \\
 & && + \mu_{K+1} P_S \\
 & \text{subject to} && \sum_{k=1}^{K+1} \mu_k = 1 \\
 & && \mu_k \geq 0 \quad \forall k \\
 & && 0 \leq P_k(R_k) \leq P_{\text{max}},
 \end{aligned} \tag{1}$$

where $P_{\text{supply}}(\mathbf{r})$ is the average cell supply power consumption when providing a rate vector \mathbf{r} to K UEs. P_0 and Δ_{PM} are

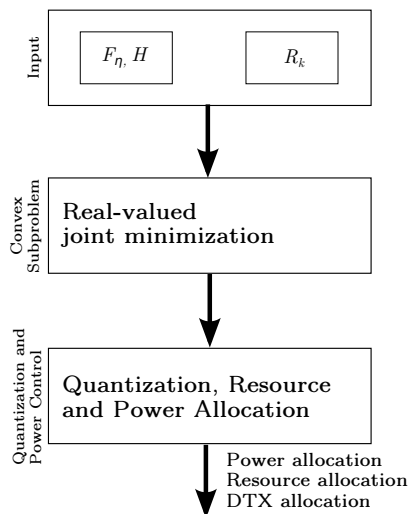


Fig. 1. Outline of the RAPS algorithm [11].

the power consumption while idle and the transmission power load factor, respectively. $P_k(R_k)$, the transmission power required to transmit rate R_k to user k , can be determined for up to 4×4 Multiple-Input Multiple-Output (MIMO) in closed form via inverting the Shannon capacity. Note that in a flat-fading TDMA system, the transmission power per user, P_k , is equal to the system transmission power at that time instant. The resource share allocated to DTX is represented by μ_{K+1} .

The solution vector $\mu^* \in (\mu_1, \dots, \mu_{K+1})$ provides the supply-power-minimal resource shares that each UE should receive as well as the time share of DTX. Due to the integer nature of OFDMA resource allocation, this result needs to be quantized. The quantized resource shares are used as input to Rate Craving Greedy (RCG) scheduling and power allocation via Margin Adaptation, concluding the RAPS algorithm by returning RB power levels, DTX time slots and UE RB allocation.

In the multicell setting, which is inherently interference limited, applying the CSI without consideration of the interference as the basis for scheduling results in an overestimation of achievable rates. In order to apply the RAPS algorithm effectively in the multicell setting, we propose the following two extensions.

First, instead of calculating achievable rates, R_k , based on CSI alone, we take interference into account for scheduling by considering the interference and noise covariance, \mathbf{F}_η , in addition to the CSI matrix, \mathbf{H} . We assume that as described in [13] on each RB there can exist a flat-fading 2x2 point-to-point MIMO link with capacity

$$\begin{aligned}
 R_{\text{RB}} &= \log_2 \left| \mathbf{I} + \mathbf{H} \mathbf{F}_\eta^{-1} \mathbf{H}^H \mathbf{F}_y \right| \\
 &= \sum_{i \in \text{spatial streams}} \log_2 \left(1 + (\text{eig}(\mathbf{H} \mathbf{F}_\eta^{-1} \mathbf{H}^H))_i P_i \right), \tag{2}
 \end{aligned}$$

where

$$\mathbf{F}_\eta = \mathbf{I}\sigma_n^2 + \sum_{j \in \text{interferers}} \mathbf{H}_j \mathbf{P}_j \mathbf{H}_j^H \quad (3)$$

and

$$\mathbf{H} \in \mathbb{C}^{N_{\text{Rx}} \times N_{\text{Tx}}}. \quad (4)$$

The determinant operator is $|\cdot|$, \mathbf{I} is the identity matrix, $\text{eig}(\cdot)$ is the eigenvalue operation and σ_n^2 is the thermal noise power. Each entry of \mathbf{H} is composed of path gain, shadowing gain, antenna gain and a Rayleigh distributed fast fading component. The achievable rate per UE, R_k , is the sum of all scheduled R_{RB} .

Second, each BS performs the RAPS algorithm on each consecutive OFDMA frame instead of a single execution. As a consequence, estimation errors can be corrected over time and the scheduling decisions of neighboring BSs are incorporated into the current allocation via the change in \mathbf{F}_η .

These extensions transform the single-cell mechanism into a dynamical system consisting of multiple cells, each of which performs the RAPS algorithm independently. Each user has a rate requirement and reports its SINR to the BS it is connected to. The perceived SINR depends on the transmission powers of all BSs. Thus, the scheduling decisions which the RAPS algorithm makes in one cell affect all other cells by means of the changing interference setting. We consider a system where, first, all UEs report their SINR per RB for an entire OFDMA frame and, then, all BSs simultaneously perform the RAPS allocation over this frame. Note that the BSs do not exchange information. Consequently, the SINR profile reported by each UE is always outdated by one iteration. As transmission power in each time slot is reduced by the RAPS algorithm due to PC or is even set to zero in a DTX slot, interference between BSs is inherently reduced.

In order to assess the performance of the RAPS algorithm in the interference scenario, we have identified Proportional-Fair (PF) scheduling as the closest comparable scheduler. By default, PF scheduling will allocate all resources regardless of requested rates. While it prevents service starvation of users with bad channels, PF is still a rate maximizing scheduler (not a power minimizing scheduler). Therefore, we make the adaptation that PF will stop scheduling more resources to a UE once a rate target has been fulfilled, thus providing guaranteed rates like RAPS while limiting power consumption. In this fashion, many resources will potentially remain unscheduled, reducing the sum bandwidth in each time slot. This is, hence, a bandwidth adapting opportunistic scheduler. Power-saving is achieved by the reduction in transmission power as a consequence of reduced transmission bandwidth at constant power spectral density.

Sequential bandwidth adaptation is taken as a second benchmark. This technique allocates RBs sequentially (not opportunistically) in the time domain UE-by-UE until all rate targets are fulfilled. When the system is under-loaded, some RBs will remain unused, thus reducing the used bandwidth. Due to its negligence of multiuser diversity, this benchmark provides a lower bound on scheduling performance.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Carrier frequency	2 GHz
Intersite distance	500 m
Pathloss model	3GPP UMa, NLOS, shadowing [14]
Shadowing standard deviation	8 dB
Bandwidth	10 MHz
Maximum transmission power	46 dBm
Thermal noise temperature	290 K
Interference tiers	2 (19 cells)
User target rate	2 Mbps
OFDMA subframes (time slots)	10
Frequency chunks	50
Power model [15] (idle; load factor; DTX)	200 W; 3.75; 90 W

To the best of our knowledge, the multi-cell system (like the single-cell system [11]) does not have a closed form solution for the resource, power and DTX allocation as functions of RB SINR and target rates. We therefore inspect the performance of the RAPS algorithm heuristically via Monte Carlo simulation.

III. RESULTS

We tested a hexagonally arranged macro BS network with a homogeneous UE distribution. The network was tested with the set of parameters shown in Table I. Three tiers of cells were placed around the center cell to generate a calibrated SINR distribution over all resources within the center cell. All cells perform RAPS resource allocation, but only data from the center cell was collected. The simulation comprises sectorization and frequency-selective fading. Unless stated otherwise, the initial power allocation state is an equal distribution of the maximum allowed power over all resources. This creates a high interference initial setting, which the RAPS algorithm improves, as well as a fairer comparison to the benchmarks which also allocate maximum power per RB. Signalling overhead is not considered in this paper. Consideration of signalling RBs prevents some time slots from being scheduled for DTX and is a topic of future study. DTX time slots are allocated sequentially from last-to-first within the OFDMA frame. Supply power consumption is averaged over an entire OFDMA frame.

We begin inspection of the network behavior by means of the SINR distribution over those RBs which were scheduled for transmission by RAPS, see Fig. 2. The SINR distributions are shown for the first ten iterations of the network after startup. As interference is reduced in each iteration, the scheduler assigns lower transmission powers. As a result, the overall SINR distribution of the scheduled RBs remains largely unchanged.

The reduction of interference can be more clearly observed when inspecting the SINR assuming unit power on all RBs in Fig. 3. It can be seen that before the first application of RAPS (Iteration 1), the unit power SINRs are in the region of -20 to +20 dB. With each iteration, the unit power SINR distribution improves, yielding higher unit power SINRs over all RBs. This is caused by the decreased interference through reduced

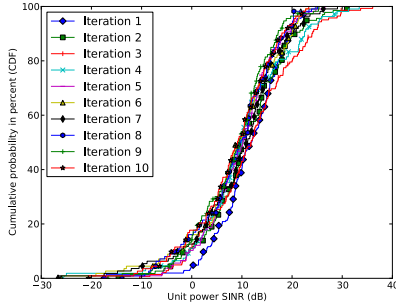


Fig. 2. RAPS SINR distribution of scheduled RBs over 10 iterations. Cell sum rate of 10 Mbps.

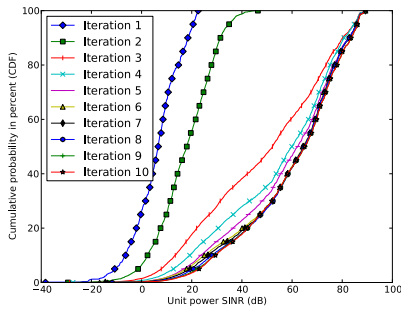


Fig. 3. RAPS unit power SINR distribution of all RBs over 10 iterations. Cell sum rate of 10 Mbps.

transmission powers in interfering cells as a consequence of zero-power DTX resources or reduced-power PC resources. Although there is no communication between the BSs, the RAPS scheduler only selects resources for transmission which have good SINR, thus reducing interference on resources with a low SINR. We conclude that RAPS acts as an interference reduction mechanism. The SINR distribution remains constant after the fifth iteration, reaching a stable network state. SINRs are improved between 20 dB for low SINRs to 50 dB for high SINRs within six iterations.

To assess the dynamics and convergence of the system and the reduction in power consumption, we next inspect the power consumption of the center cell. In Fig. 4, the center cell average power consumption is shown for 40 independent trials for different initial power allocations. As computing an initial worst-case power allocation over all RBs and cells by brute force is infeasible, we start each trial with one of three power configurations, namely zero power on all RBs, maximum equal power on all RBs and a random allocation over all RBs within the transmission power constraint. In addition, the PF benchmark is applied with a maximum equal power starting condition. It is found that for all of the

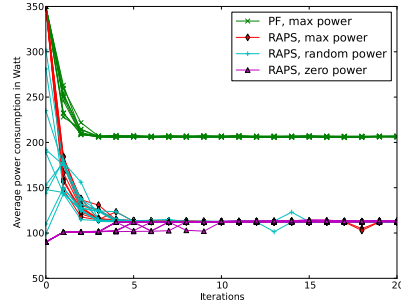


Fig. 4. Average center cell power consumption over 40 independent trials with varying initial power levels. Cell sum rate of 10 Mbps.

starting configurations, most RAPS trials converge within four iterations and all RAPS trials converge within ten iterations. Convergence is reached from all three power configurations (max, random and zero power), reinforcing the prediction that the RAPS scheduler converges for any starting configuration in large networks. Although each RAPS trial is performed in an individual simulation with different channel realizations, all trials converge to the same power consumption level of 112 W. This is caused by the following effect: As transmission powers are reduced in each iteration, P_0 and P_S begin to dominate the overall consumption. As for the same target rate per UE, two DTX time slots turn out to be the solution, and a total average power consumption of $8P_S + 2P_0 = 112$ W is eventually reached. This consumption level is significantly lower than the upper limit at 350 Watt. This saving is achieved by the exploitation of the provisioning gap between target capacity and available capacity. By only using resources with good channel conditions for transmission, optimal power allocation and DTX, no transmission or device power consumption is wasted. In comparison, the benchmark PF is much less adaptable to the low load and converges to a consumption of 207 W. Overall, RAPS operation allows to reduce power consumption by almost 50% compared to the PF benchmark.

A noticeable effect is the convergence to distinct levels of power consumption. These are owed to the coarse quantization nature of the DTX time slots. A degradation in SINR profiles reported by UEs triggers the RAPS algorithm to allocate one more time slot for transmission, thus increasing power consumption to a higher level. In turn, improved SINRs in a cell lead to one less time slot used for transmission yielding lower power consumption. In this region of operation transmission powers are low compared to other device consumption, emphasizing the effect of DTX time slot selection and creating the stepped pattern (most prominent in 'RAPS, zero power').

Lastly, we assess the performance of the RAPS algorithm over different cell sum data rates. In Fig. 5, it is found that PF scheduling only achieves marginal gains compared to sequential bandwidth adaptation, a very weak benchmark. The power savings of PF over this weak benchmark are in

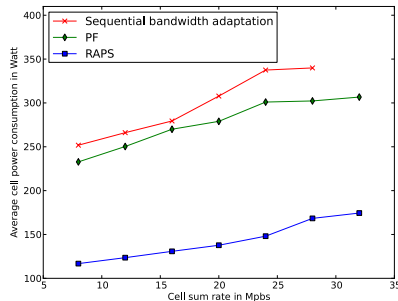


Fig. 5. Average supply power consumption over cell sum rate.

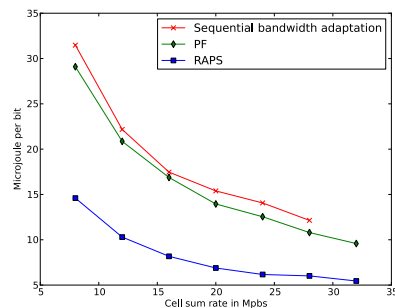


Fig. 6. Energy per bit over cell sum rate.

the order of 20 W to 40 W. Savings are larger at higher cell sum rates (*i.e.* higher traffic load) than at lower cell sum rates. In contrast, RAPS scheduling achieves a reduction in power consumption of up to 50% or 150 W compared to PF scheduling. The savings are almost of constant magnitudes over all cell sum rates tested. Note the missing data point at 32 Mbps for sequential bandwidth adaptation. This is due to the fact that a cell sum rate of 32 Mbps could not be achieved in trials for sequential bandwidth adaptation due to its lack of multi-user diversity exploitation. PF and RAPS can provide these rates due to opportunistic resource allocation.

However, although the PF and RAPS schedulers can provide similar rates, they can do so at significantly differing average power consumption. This contribution is further analyzed by considering the invested energy-per-bit. The same data set as in Fig. 5 is plotted in Fig. 6 on the energy-per-bit metric to emphasize the non-linear relationship between data transmitted and cell power consumption. Generally, at high cell sum rates the energy-per-bit metric is lower than at low cell sum rates for all schedulers. This is not surprising as BSs operate more efficiently at high capacities. However, application of the RAPS algorithm nearly halves the energy-per-bit cost over all cell sum rates, thus doubling efficiency compared to PF scheduling.

IV. CONCLUSION

In this paper, we have presented the first adaptation of the RAPS algorithm for a cellular system that is subject to co-channel interference. OFDMA SINR is added as an input parameter to provide the basis for RAPS power, DTX and resource allocation. In contrast to the single cell, application of RAPS in the multi-cell setting creates a dynamical system in which the RAPS algorithm has to be executed repeatedly. We find the system to be converging after six iterations. Simulation results indicate that SINR distributions can be improved by more than 20 dB by employing RAPS in OFDMA BSs. A RAPS system is shown to converge independent of the initial power allocation configuration. Without exchanging information between the BSs, cell power consumption is reduced by around 50% compared to the PF scheduler benchmark. The RAPS scheduler is, thus, able to double the energy efficiency of interference limited cellular networks.

REFERENCES

- [1] G. Auer, I. Gódor, L. Hévízi, M. A. Imran, J. Malmodin, P. FASEKAS, G. Biczók, H. Holtkamp, D. Zeller, O. Blume, and R. Tafazolli, "Enablers for Energy Efficient Wireless Networks," in *Proc. of the Vehicular Technology Conference (VTC)*, 2010.
- [2] A. Fehske, J. Malmodin, G. Biczók, and G. Fettweis, "The Global Carbon Footprint of Mobile Communications - The Ecological and Economic Perspective," *IEEE Communications Magazine*, 2010.
- [3] P. Frenger, P. Moberg, J. Malmodin, Y. Jading, and I. Gódor, "Reducing Energy Consumption in LTE with Cell DTX," in *Proceedings of the IEEE VTC 2011-Spring*, 2011.
- [4] K. Abdallah, I. Cerutti, and P. Castoldi, "Energy-Efficient Coordinated Sleep of LTE Cells," in *Communications (ICC), 2012 IEEE International Conference on*, pp. 5238–5242, June 2012.
- [5] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1747–1758, Oct. 1999.
- [6] G. Miao, N. Himayat, and G. Li, "Energy-Efficient Transmission in Frequency-Selective Channels," *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008*, 2008.
- [7] S. Videv and H. Haas, "Energy-Efficient Scheduling and Bandwidth-Energy Efficiency Trade-Off with Low Load," in *ICC*, pp. 1–5, 2011.
- [8] Z. Xu, C. Yang, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-Efficient MIMO-OFDMA Systems based on Switching Off RF Chains," in *Proceedings of the VTC Fall 2011*, 2011.
- [9] D. López-Pérez, X. Chu, A. V. Vasilakos, and H. Claussen, "Minimising Cell Transmit Power: Towards Self-Organized Resource Allocation in OFDMA Femtocells," in *Proceedings of the ACM SIGCOMM 2011 Conference, SIGCOMM '11*, (New York, NY, USA), pp. 410–411, ACM, 2011.
- [10] A. de Domenico, *Energy Efficient Mechanisms for Heterogeneous Cellular Networks*. PhD thesis, Université de Grenoble, 2012.
- [11] H. Holtkamp, G. Auer, S. Bazzi, and H. Haas, "Minimizing Base Station Power Consumption," *Selected Areas in Communications, IEEE Journal on*, vol. PP, Dec. 2013.
- [12] H. Holtkamp, G. Auer, and H. Haas, "On Minimizing Base Station Power Consumption," in *Proceedings of the IEEE VTC 2011-Fall*, 2011.
- [13] E. Biglieri, A. R. Calderbank, A. G. Constantinides, A. Goldsmith, and A. Paulraj, *MIMO Wireless Communications*. Cambridge University Press, 2007.
- [14] 3GPP, "Further Advancements for E-UTRA Physical Layer Aspects (Release 9)," 3GPP TR 36.814 V9.0.0, Mar. 2010. Retrieved Jan. 15, 2013 from www.3gpp.org/ftp/Specs/.
- [15] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, W. Wajda, D. Sabella, F. Richter, M. Gonzalez, H. Klessig, I. Gódor, P. Skillermark, M. Olsson, M. A. Imran, A. Ambrosy, and O. Blume, "Flexible Power Modeling of LTE Base Stations," in *2012 IEEE Wireless Communications and Networking Conference: Mobile and Wireless Networks (IEEE WCNC 2012 Track 3 Mobile & Wireless)*, 2012.

Publications

- [a] Gunther Auer, István Gódor, László Hévízi, Muhammad Ali Imran, Jens Malmudin, Péter FASEKAS, Gergely Biczók, **Hauke Holtkamp**, Dietrich Zeller, Oliver Blume, and Rahim Tafazolli. Enablers for Energy Efficient Wireless Networks. In *Proceedings of the IEEE Vehicular Technology Conference (VTC) Fall-2010*, 2010.
- [b] **Hauke Holtkamp** and Gunther Auer. Fundamental Limits of Energy-Efficient Resource Sharing, Power Control and Discontinuous Transmission. In *Proceedings of the Future Network & Mobile Summit 2011*, 2011.
- [c] **Hauke Holtkamp**, Gunther Auer, and Harald Haas. Minimal Average Consumption Downlink Base Station Power Control Strategy. In *Proceedings of the 2011 IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 2430–2434, Sept 2011.
- [d] **Hauke Holtkamp**, Gunther Auer, and Harald Haas. On Minimizing Base Station Power Consumption. In *Proceedings of the Vehicular Technology Conference (VTC) Fall-2011*, 2011.
- [e] Claude Desset, Björn Debaillie, Vito Giannini, Albrecht Fehske, Gunther Auer, **Hauke Holtkamp**, Wiesława Wajda, Dario Sabella, Fred Richter, Manuel Gonzalez, Henrik Klessig, István Gódor, Per Skillermark, Magnus Olsson, Muhammad Ali Imran, Anton Ambrosy, and Oliver Blume. Flexible Power Modeling of LTE Base Stations. In *Proceedings of the 2012 IEEE Wireless Communications and Networking Conference*., 2012.
- [f] EARTH Project Work Package 2. Deliverable D2.2: Reference Systems and Scenarios. Retrieved April 18, 2013, from <https://www.ict-earth.eu/publications/deliverables/deliverables.html>, June 2012. **Hauke Holtkamp** is an editor.
- [g] EARTH Project Work Package 3. Deliverable D3.3: Green Network Technologies. Retrieved April 18, 2013, from <https://www.ict-earth.eu/publications/deliverables/deliverables.html>, June 2012. **Hauke Holtkamp** is an author.

-
- [h] **Hauke Holtkamp**, Gunther Auer, Vito Giannini, and Harald Haas. A Parameterized Base Station Power Model. *IEEE Communications Letters*, 2013. Accepted for publication on 16 August 2013.
 - [i] **Hauke Holtkamp** and Harald Haas. OFDMA Base Station Power-saving Via Joint Power Control and DTX in Cellular Systems. In *Proceedings of the Vehicular Technology Conference (VTC) Fall-2013*, 2013.
 - [j] **Hauke Holtkamp**, Gunther Auer, Samer Bazzi, and Harald Haas. Minimizing Base Station Power Consumption. *IEEE Journal on Selected Areas in Communications*, PP(99), December 2014.
 - [k] **Hauke Holtkamp**, Guido Dietl, and Harald Haas. Distributed DTX Alignment with Memory. In *Proceedings of the 2014 IEEE International Conference on Communications (ICC)*, Sydney, Australia, June 2014. Submitted on Aug 15, 2013.
 - [l] Jinsong Wu, Sundeep Rangan, and Honggang Zhang. *Green Communications: Theoretical Fundamentals, Algorithms and Applications*. Taylor & Francis Group, September 2012. **Hauke Holtkamp** is a coauthor of Chapter 17.

Literature References

- [1] Q. Bi, G. Zysman, and H. Menkes, “Wireless Mobile Communications at the Start of the 21st Century,” *Communications Magazine, IEEE*, vol. 39, no. 1, pp. 110–116, 2001.
- [2] G. Wu, S. Talwar, K. Johnsson, N. Himayat, and K. Johnson, “M2M: From Mobile to Embedded Internet,” *Communications Magazine, IEEE*, vol. 49, no. 4, pp. 36–43, 2011.
- [3] J. Brazell, L. Donoho, J. Dexheimer, P. Robert Hannenman, and G. Langdon, *M2M: The Wireless Revolution*. Texas State Technical College Publishing, 2005.
- [4] Wireless Week, “Cisco: 1 trillion connected devices by 2013.” <http://www.wirelessweek.com/news/2010/03/cisco-1-trillion-connected-devices-2013>, Mar. 2010. Retrieved 14 Apr. 2013.
- [5] Cisco, “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012–2017.” <http://www.cisco.com/>, 2013. Retrieved 18 Apr. 2013.
- [6] A. Fehske, J. Malmodin, G. Biczók, and G. Fettweis, “The Global Carbon Footprint of Mobile Communications - The Ecological and Economic Perspective,” *IEEE Communications Magazine*, vol. 49, pp. 55–62, 2010.
- [7] G. Auer, V. Giannini, I. Gódor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. J. Gonzalez, and C. Desset, “Cellular Energy Efficiency Evaluation Framework,” in *Proceedings of the VTC 2011-Spring*, 2011.
- [8] EARTH Project Work Package 2, “Deliverable D2.3: Energy Efficiency Analysis of the Reference Systems, Areas of Improvements and Target Breakdown.” <https://www.ict-earth.eu/publications/deliverables/deliverables.html>, Nov. 2010. Retrieved Apr. 18, 2013.
- [9] M. Gonzalez, D. Ferling, W. Wajda, A. Erdem, and P. Maugars, “Concepts for Energy Efficient LTE Transceiver Systems in Macro Base Stations,” in *Future Network Mobile Summit (FutureNetw)*, 2011, pp. 1–8, 2011.

- [10] EARTH Project Work Package 4, “Deliverable D4.3: Final Report on Green Radio Technologies.” <https://www.ict-earth.eu/publications/deliverables/deliverables.html>, June 2012. Retrieved Apr. 18, 2013.
- [11] “EARTH - Energy Aware Radio and neTwork TecHnologies.” <http://www.ict-earth.eu>. Retrieved 17 June 2013.
- [12] J. Malmodin, A. Moberg, D. Lundén, G. Finnveden, and N. Lövehagen, “Greenhouse Gas Emissions and Operational Electricity Use in the ICT and Entertainment & Media Sectors,” *Journal of Industrial Ecology*, vol. 14, no. 5, pp. 770–790, 2010.
- [13] J. Louhi and H. Scheck, “Energy Efficiency of Cellular Networks,” in *Proc. Int. Symp. Wireless Personal Multimedia Communications (WPMC), Lapland, Finland, 2008*.
- [14] K. Aleklett, M. Höök, K. Jakobsson, M. Lardelli, S. Snowden, and B. Söderbergh, “The Peak of the Oil Age Analyzing the World Oil Production Reference Scenario in World Energy Outlook 2008,” *Energy Policy*, vol. 38, no. 3, pp. 1398 – 1414, 2010.
- [15] G. Fettweis and E. Zimmermann, “ICT Energy Consumption – Trends and Challenges,” in *Proceedings of the 11th International Symposium on Wireless Personal Multimedia Communications (WPMC 2008)*, 2008.
- [16] P. Nema, R. Nema, and S. Rangnekar, “Minimization of Green House Gases Emission by using Hybrid Energy System for Telephony Base Station Site Application,” *Renewable and Sustainable Energy Reviews*, vol. 14, no. 6, pp. 1635–1639, 2010.
- [17] C. Han, T. Harrold, S. Armour, I. Krikidis, S. Videv, P. M. Grant, H. Haas, J. S. Thompson, I. Ku, C.-X. Wang, *et al.*, “Green Radio: Radio Techniques to Enable Energy-efficient Wireless Networks,” *Communications Magazine, IEEE*, vol. 49, no. 6, pp. 46–54, 2011.
- [18] Y. Sugiyama, “Green ICT toward Low Carbon Society,” in *Design for Innovative Value Towards a Sustainable Society* (M. Matsumoto, Y. Umeda, K. Masui, and S. Fukushige, eds.), pp. 739–742, Springer Netherlands, 2012.
- [19] NTT DOCOMO Technical Journal Editorial Office, “Measures for Recovery from the Great East Japan Earthquake using NTT DOCOMO R&D Technology,” *NTT DOCOMO Technical Journal*, vol. 13, no. 4, pp. 96–106, 2012.
- [20] Y. Ran, “Considerations and Suggestions on Improvement of Communication Network Disaster Countermeasures after the Wenchuan Earthquake,” *Communications Magazine, IEEE*, vol. 49, no. 1, pp. 44–47, 2011.

- [21] G. Auer, V. Giannini, I. Gódor, P. Skillermark, M. Olsson, M. A. Imran, M. J. Gonzalez, C. Desset, O. Blume, and A. Fehske, “How Much Energy is Needed to Run a Wireless Network?,” *IEEE Wireless Communications*, vol. 18, pp. 40–49, Oct 2011.
- [22] O. Arnold, F. Richter, G. Fettweis, and O. Blume, “Power Consumption Modeling of Different Base Station Types in Heterogeneous Cellular Networks,” in *Future Network and Mobile Summit, 2010*, pp. 1–8, June 2010.
- [23] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, “Providing Quality of Service over a Shared Wireless Link,” *Communications Magazine, IEEE*, vol. 39, no. 2, pp. 150–154, 2001.
- [24] O. Blume, D. Zeller, and U. Barth, “Approaches to Energy Efficient Wireless Access Networks,” in *4th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pp. 1–5, 3-5 2010.
- [25] W. Ye, J. Heidemann, and D. Estrin, “An Energy-Efficient MAC Protocol for Wireless Sensor Networks,” *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, pp. 1567–1576, 2002.
- [26] S. Cui, A. J. Goldsmith, and A. Bahai, “Energy-Efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 22, pp. 1089–1098, 2004.
- [27] S. Cui, A. J. Goldsmith, and A. Bahai, “Energy-Constrained Modulation Optimization,” *IEEE Transactions on Wireless Communications*, vol. 4, pp. 2349–2360, 2005.
- [28] R. Bhatia and M. Kodialam, “On Power Efficient Communication over Multi-hop Wireless Networks: Joint Routing, Scheduling and Power Control,” in *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, pp. 1457–1466, IEEE, 2004.
- [29] J. Malmudin, L. Oliv, and P. Bergmark, “Life Cycle Assessment of Third Generation (3G) Wireless Telecommunication Systems at Ericsson,” in *Environmentally Conscious Design and Inverse Manufacturing, 2001. Proceedings EcoDesign 2001: Second International Symposium on*, pp. 328–334, 2001.
- [30] E. Weidman and S. Lundberg, “Life Cycle Assessment of Ericsson Third Generation Systems,” in *Electronics and the Environment, 2000. ISEE 2000. Proceedings of the 2000 IEEE International Symposium on*, pp. 136–142, 2000.

- [31] Mobile VCE, "Website of the Green Radio Project." <http://www.mobilevce.com/green-radio>. Retrieved 12 June, 2013.
- [32] M. Gruber, O. Blume, D. Ferling, D. Zeller, M. Imran, and E. Strinati, "EARTH: Energy Aware Radio and Network Technologies," in *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, pp. 1–5, 2009.
- [33] D. Ferling, T. Bohn, D. Zeller, P. Frenger, I. Gódor, Y. Jading, and W. Tomaselli, "Energy Efficiency Approaches for Radio Nodes," in *Future Network & Mobile Summit 2010*, pp. 1–9, 2010.
- [34] S. Bories, L. Dussopt, A. Giry, and C. Delaveaud, "Duplexer-less RF Front-end for LTE Pico-Cell using a Dual Polarization Antenna," in *Wireless Conference 2011 - Sustainable Wireless Technologies (European Wireless), 11th European*, pp. 1–3, 2011.
- [35] O. Hammi, A. Kwan, M. Helaoui, and F. M. Ghannouchi, "Green Power Amplification Systems for 3G+ Wireless Communication Infrastructure," in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, pp. 1–5, IEEE, 2010.
- [36] H. Klessig, A. Fehske, and G. Fettweis, "Energy Efficiency Gains in Interference-limited Heterogeneous Cellular Mobile Radio Networks with Random Micro Site Deployment," in *Sarnoff Symposium, 2011 34th IEEE*, pp. 1–6, 2011.
- [37] F. Richter, A. Fehske, and G. Fettweis, "Energy Efficiency Aspects of Base Station Deployment Strategies for Cellular Networks," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, pp. 1–5, 2009.
- [38] B. Badic, T. O'Farrell, P. Loskot, and J. He, "Energy Efficient Radio Access Architectures for Green Radio: Large versus Small Cell Size Deployment," in *Proc. of the 70th Vehicular Technology Conference (VTC)*, (Anchorage, USA), pp. 1–5, Sept. 2009.
- [39] C. Khirallah and J. S. Thompson, "Energy Efficiency of Heterogeneous Networks in LTE-Advanced," *Journal of Signal Processing Systems*, vol. 69, no. 1, pp. 105–113, 2012.
- [40] J. Hoydis, M. Kobayashi, and M. Debbah, "Green Small-Cell Networks," *Vehicular Technology Magazine, IEEE*, vol. 6, no. 1, pp. 37–43, 2011.
- [41] E. Oh and B. Krishnamachari, "Energy Savings through Dynamic Base Station Switching in Cellular Wireless Access Networks," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, pp. 1–5, IEEE, 2010.

-
- [42] I. Ashraf, F. Boccardi, and L. Ho, "Sleep Mode Techniques for Small Cell Deployments," *Communications Magazine, IEEE*, vol. 49, pp. 72–79, Aug. 2011.
- [43] P. Frenger, P. Moberg, J. Malmudin, Y. Jading, and I. Gódor, "Reducing Energy Consumption in LTE with Cell DTX," in *Proceedings of the IEEE VTC 2011-Spring*, pp. 1–5, 2011.
- [44] L. Saker, S.-E. Elayoubi, and T. Chahed, "Minimizing Energy Consumption via Sleep Mode in Green Base Station," in *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, pp. 1–6, 2010.
- [45] K. Abdallah, I. Cerutti, and P. Castoldi, "Energy-Efficient Coordinated Sleep of LTE Cells," in *Communications (ICC), 2012 IEEE International Conference on*, pp. 5238–5242, June 2012.
- [46] T. Han and N. Ansari, "ICE: Intelligent Cell BrEathing to Optimize the Utilization of Green Energy," *Communications Letters, IEEE*, vol. 16, no. 6, pp. 866–869, 2012.
- [47] C. Abgrall, E. C. Strinati, and J.-C. Belfiore, "Distributed Power Allocation for Interference Limited Networks," in *Proc. of the 21st IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), (Istanbul, Turkey)*, pp. 1118–1122, Sept. 26–29 2010.
- [48] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell Zooming for Cost-efficient Green Cellular Networks," *Communications Magazine, IEEE*, vol. 48, no. 11, pp. 74–79, 2010.
- [49] M. Hedayati, M. Amirijoo, P. Frenger, and J. Moe, "Reducing Energy Consumption Through Adaptation of Number of Active Radio Units," in *Proceedings of the IEEE VTC 2012-Spring*, pp. 1–5, 2012.
- [50] P. Skillermark and P. Frenger, "Enhancing Energy Efficiency in LTE with Antenna Muting," in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pp. 1–5, 2012.
- [51] H. Kim, C.-B. Chae, G. de Veciana, and R. W. Heath, "A Cross-layer Approach to Energy Efficiency for Adaptive MIMO Systems Exploiting Spare Capacity," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 4264–4275, Aug. 2009.
- [52] Z. Xu, C. Yang, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-Efficient MIMO-OFDMA Systems based on Switching Off RF Chains," in *Proceedings of the VTC Fall 2011*, pp. 1–5, 2011.

- [53] X. Wu, S. Sinanovic, M. Di Renzo, and H. Haas, "Base Station Energy Consumption for Transmission Optimised Spatial Modulation (TOSM) in Correlated Channels," in *Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2012 IEEE 17th International Workshop on*, pp. 261–265, 2012.
- [54] A. Stavridis, S. Sinanovic, M. Di Renzo, H. Haas, and P. Grant, "An Energy Saving Base Station Employing Spatial Modulation," in *Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2012 IEEE 17th International Workshop on*, pp. 231–235, IEEE, 2012.
- [55] R. Gupta and E. Strinati, "Green Scheduling to Minimize Base Station Transmit Power and UE Circuit Power Consumption," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, pp. 2424–2429, 2011.
- [56] Z. Chong and E. Jorswieck, "Analytical Foundation for Energy Efficiency Optimisation in Cellular Networks with Elastic Traffic," in *Mobile Lightweight Wireless Systems*, pp. 18–29, Springer, 2012.
- [57] A. Ambrosy, M. Wilhelm, O. Blume, and W. Wajda, "Dynamic Bandwidth Management for Energy Savings in Wireless Base Stations," in *Proceedings of Globecom 2012*, pp. 3502–3507, 2012.
- [58] S. Videv and H. Haas, "Energy-Efficient Scheduling and Bandwidth-Energy Efficiency Trade-Off with Low Load," in *International Conference on Communications*, pp. 1–5, 2011.
- [59] F. Meshkati, V. Poor, and S. Schwartz, "Energy-Efficient Resource Allocation in Wireless Networks: An Overview of Game Theoretic Approaches," *IEEE Signal Processing Magazine: Special Issue on Resource-Constrained Signal Processing, Communications and Networking*, May 2007.
- [60] EARTH Project Work Package 3, "Deliverable D3.3: Green Network Technologies." <https://www.ict-earth.eu/publications/deliverables/deliverables.html>, June 2012. Retrieved 18 Apr. 2013.
- [61] L. Scalia, T. Biermann, C. Choi, K. Kozu, and W. Kellerer, "Power-efficient Mobile Backhaul Design for CoMP Support in Future Wireless Access Systems," in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, pp. 253–258, 2011.
- [62] S. Han, C. Yang, G. Wang, and M. Lei, "On the Energy Efficiency of Base Station Sleeping with Multicell Cooperative Transmission," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, pp. 1536–1540, IEEE, 2011.

- [63] L. G. Hevizi and I. Gódor, “Power Savings in Mobile Networks by Dynamic Base Station Sectorization,” in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, pp. 2415–2417, IEEE, 2011.
- [64] E. Strinati, A. De Domenico, and A. Duda, “Ghost Femtocells: A Novel Radio Resource Management Scheme for OFDMA Based Networks,” in *Wireless Communications and Networking Conference (WCNC), 2011 IEEE*, pp. 108–113, 2011.
- [65] E. Ternon, Z. Bharucha, and H. Taoka, “A Feasibility Study for the Detached Cell Concept,” in *International ITG Conference on Systems, Communications and Coding*, Jan. 2013.
- [66] X. Wang, A. V. Vasilakos, M. Chen, Y. Liu, and T. T. Kwon, “A Survey of Green Mobile Networks: Opportunities and Challenges,” *Mobile Networks and Applications*, vol. 17, no. 1, pp. 4–20, 2012.
- [67] A. Ephremides, “Energy Concerns in Wireless Networks,” *Wireless Communications, IEEE*, vol. 9, no. 4, pp. 48–59, 2002.
- [68] F. Capozzi, G. Piro, L. Grieco, G. Boggia, and P. Camarda, “Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey,” *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 1–23, 2012.
- [69] E. Dahlman, S. Parkvall, and J. Sköld, *4G LTE/LTE-Advanced for Mobile Broadband*. Academic Press, 1 ed., 2011.
- [70] GSA Secretariat, “GSA Evolution to LTE report,” tech. rep., Global mobile Suppliers Association, Oct. 2013.
- [71] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. Wiley, 1 ed., 2009.
- [72] Ericsson AB, “Ericsson Mobility Report,” tech. rep., Ericsson AB, Nov. 2012.
- [73] I. A. Glover and P. M. Grant, *Digital Communications*. Prentice Hall, 1998.
- [74] V. Garg, *Wireless Communications & Networking*. The Morgan Kaufmann Series in Networking, Elsevier Science, 2010.
- [75] S. Pike, “The Radio Interface for UMTS,” in *Personal Communications in the 21st Century (I) (Ref. No. 1998/214)*, *IEE Colloquium on*, (London), pp. 1–725, Feb. 5, 1998.

- [76] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) Protocol Specification (Release 9).” 3GPP TS 36.321 V 9.2.0 (2010-03), Mar. 2010.
- [77] V. Kühn, *Wireless Communications over MIMO Channels*. John Wiley & Sons Ltd., 2006.
- [78] E. Biglieri, A. R. Calderbank, A. G. Constantinides, A. Goldsmith, and A. Paulraj, *MIMO Wireless Communications*. Cambridge University Press, 2007.
- [79] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, “Capacity Limits of MIMO Channels,” *IEEE Journal on Selected Areas in Communication*, vol. 21, pp. 684–702, June 2003.
- [80] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Mar. 2004.
- [81] D. P. Palomar and Y. C. Eldar, *Convex Optimization in Signal Processing and Communications*. Cambridge University Press New York, 2010.
- [82] 3GPP, “Further Advancements for E-UTRA Physical Layer Aspects (Release 9), TR 36.814 V9.0.0.” www.3gpp.org/ftp/Specs/, Mar. 2010. Retrieved 15 Jan. 2013.
- [83] IST-4-027756 WINNER II, “D1.1.2 v1.2 WINNER II Channel Models.” <https://www.ist-winner.org/WINNER2-Deliverables/>, 2006. Retrieved 5 Feb. 2008.
- [84] A. Wächter and K. Laird, “IPOPT - Interior Point OPTimizer.” retrieved May 2, 2013.
- [85] Condor Team, “Condor Manual.” <http://research.cs.wisc.edu/htcondor/>, May 2013. Retrieved 17 June 2013.
- [86] “MATLAB - The Language of Technical Computing.” <http://www.mathworks.com/products/matlab>. Retrieved 17 June 2013.
- [87] “Python Programming Language Official Website.” <http://www.python.org>. Retrieved 17 June 2013.
- [88] M. Deruyck, W. Joseph, and L. Martens, “Power Consumption Model for Macrocell and Microcell Base Stations,” *Transactions on Emerging Telecommunications Technologies*, 2012.
- [89] D. López-Pérez, X. Chu, A. V. Vasilakos, and H. Claussen, “Minimising Cell Transmit Power: Towards Self-Organized Resource Allocation in OFDMA Femtocells,” in *Proceedings of the ACM SIGCOMM 2011 Conference*, SIGCOMM ’11, (New York, NY, USA), pp. 410–411, ACM, 2011.

- [90] G. Miao, N. Himayat, and G. Li, "Energy-Efficient Transmission in Frequency-Selective Channels," *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008*, vol. 58, pp. 545–554, 2008.
- [91] A. de Domenico, *Energy Efficient Mechanisms for Heterogeneous Cellular Networks*. PhD thesis, Université de Grenoble, 2012.
- [92] Sandvine, "Mobile Internet Phenomena Report." <http://www.sandvine.com/downloads/documents/2010%20Global%20Internet%20Phenomena%20Report.pdf>, 2010. Retrieved 17 June 2013.
- [93] S. Sinanović, N. Serafimovski, H. Haas, and G. Auer, "System Spectral Efficiency Analysis of a 2-Link Ad Hoc Network," in *Proc. of the 50th IEEE Global Telecommunications Conference (GLOBECOM)*, (Washington, USA), pp. 3684–3688, IEEE, Nov. 26–30 2007.
- [94] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1747–1758, Oct. 1999.
- [95] G. Miao, N. Himayat, Y. Li, and D. Bormann, "Energy Efficient Design in Wireless OFDMA," in *IEEE International Conference on Communications*, pp. 3307–3312, 19–23 2008.
- [96] D. Kivanc, G. Li, and H. Liu, "Computationally Efficient Bandwidth Allocation and Power Control for OFDMA," *IEEE Transactions on Wireless Communications*, vol. 2, pp. 1150–1158, Nov. 2003.
- [97] H. Al-Shatri and T. Weber, "Fair Power Allocation for Sum-Rate Maximization in Multiuser OFDMA," in *Proceedings of the International ITG Workshop on Smart Antennas 2010*, pp. 350–354, 2010.
- [98] W. Vereecken, M. Deruyck, D. Colle, W. Joseph, M. Pickavet, L. Martens, and P. Demeester, "Evaluation of the Potential for Energy Saving in Macrocell and Femtocell Networks using a Heuristic Introducing Sleep modes in Base Stations," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, pp. 1–14, 2012.
- [99] L. Xiang, F. Pantisano, R. Verdone, X. Ge, and M. Chen, "Adaptive Traffic Load-balancing for Green Cellular Networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, pp. 41–45, IEEE, 2011.
- [100] C. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 & 623–656, July & Oct. 1948.

- [101] J. Jang and K. B. Lee, "Transmit Power Adaptation for Multiuser OFDM Systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 171–178, Feb. 2003.
- [102] M. Rosen, "Niels Hendrik Abel and Equations of the Fifth Degree," *The American Mathematical Monthly*, vol. 102, no. 6, pp. 495–505, 1995.
- [103] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed Multiuser Power Control for Digital Subscriber Lines," *Selected Areas in Communications, IEEE Journal on*, vol. 20, no. 5, pp. 1105–1115, 2002.
- [104] IST-2003-507581 WINNER, "D5.4 v1.0 Final Report on Link Level and System Level Channel Models." <https://www.ist-winner.org/DeliverableDocuments/>, Nov. 2005. Retrieved 15 Apr. 2007.
- [105] G. Li and H. Liu, "Downlink Radio Resource Allocation for Multi-Cell OFDMA System," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 3451–3459, Dec. 2006.
- [106] S. Uygungelen, G. Auer, and Z. Bharucha, "Graph-Based Dynamic Frequency Reuse in Femtocell Networks," in *Proc. of the 73rd IEEE Vehicular Technology Conference (VTC)*, (Budapest, Hungary), May 15–18 2011.
- [107] M. Rahman and H. Yanikomeroglu, "Enhancing Cell-edge Performance: a Downlink Dynamic Interference Avoidance Scheme with Inter-cell Coordination," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 4, pp. 1414–1425, 2010.
- [108] Y.-S. Liang, W.-H. Chung, C.-M. Yu, H. Zhang, C.-H. Chung, C.-H. Ho, and S.-Y. Kuo, "Resource Block Assignment for Interference Avoidance in Femtocell Networks," in *Vehicular Technology Conference (VTC Fall), 2012 IEEE*, pp. 1–5, 2012.
- [109] I. Katzela and M. Naghshineh, "Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey," *Personal Communications, IEEE*, vol. 3, no. 3, pp. 10–31, Jun 1996.
- [110] J. G. Proakis, *Digital Communications*. McGraw-Hill Series in Electrical and Computer Engineering, McGraw-Hill Higher Education, 4 ed., Dec. 2000.
- [111] M. Serizawa and D. Goodman, "Instability and Deadlock of Distributed Dynamic Channel Allocation," in *Vehicular Technology Conference, 1993., 43rd IEEE*, pp. 528–531, May 1993.
- [112] J. Ellenbeck, C. Hartmann, and L. Berlemann, "Decentralized Inter-Cell Interference Coordination by Autonomous Spectral Reuse Decisions," in

- Proc. of the 14th European Wireless Conference (EW)*, (Prague, Czech Republic), pp. 1–7, June 22–25 2008.
- [113] Y. Akaiwa and H. Andoh, “Channel Segregation – A Self-organized Dynamic Channel Allocation Method: Application to TDMA/FDMA Microcellular System,” *Selected Areas in Communications, IEEE Journal on*, vol. 11, pp. 949–954, Aug. 1993.
- [114] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, “Massive MIMO for Next Generation Wireless Systems,” *arXiv preprint arXiv:1304.6690*, 2013.
- [115] R. Q. Hu, Y. Qian, S. Kota, and G. Giambene, “HetNets – a new Paradigm for Increasing Cellular Capacity and Coverage,” *Wireless Communications, IEEE*, vol. 18, no. 3, pp. 8–9, 2011.
- [116] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, “Stochastic Geometry and Random Graphs for the Analysis and Design of Wireless Networks,” *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 7, pp. 1029–1046, 2009.
-