Department of Communications and Networking

Energy Efficiency of Wireless Access - Impact of Power Amplifiers and Load Variations

M. M. Aftab Hossain

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 117/2016

Energy Efficiency of Wireless Access - Impact of Power Amplifiers and Load Variations

M. M. Aftab Hossain

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall S3 of the school on 22 June 2016 at 12.

Aalto University
School of Electrical Engineering
Department of Communications and Networking

Supervising professor
Prof. Riku Jäntti, Aalto University, Finland

Thesis advisor
Dr. Konstantinos Koufus, University of Bristol, England

Preliminary examiners
Adjunct Prof. Mehdi Bennis, University of Oulu, Finland
Lecturer Fabien Héliot, University of Surrey, England

Opponent
Prof. Jukka Lempiäinen, University of Tampere, Finland

Aalto University publication series
DOCTORAL DISSERTATIONS 117/2016

ISBN 978-952-60-6863-3 (printed)
ISBN 978-952-60-6864-0 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-6864-0

Unigrafia Oy
Helsinki 2016

Finland

441 697
Printed matter

Author
M. M. Aftab Hossain

Name of the doctoral dissertation
Energy Efficiency of Wireless Access - Impact of Power Amplifiers and Load Variations

Publisher School of Electrical Engineering

Unit Department of Communications and Networking

Series Aalto University publication series DOCTORAL DISSERTATIONS 117/2016

Field of research Communications Engineering

Manuscript submitted 18 February 2016 **Date of the defence** 22 June 2016

Permission to publish granted (date) 19 May 2016 **Language** English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

In order to ensure seamless coverage and sustainable exponential growth of capacity, there is a keen interest in the development and deployment of highly energy efficient wireless systems and solutions. To design an energy efficient network, it is important to consider the facts that the variations in traffic demand in both temporal and spatial domains are significant, and the power consumption of cellular networks is mostly dominated by macro base stations where the power amplifiers (PAs) consume around 55-70 percent of total energy. One of the main challenges lies in coping with this load variations considering that the PAs attain high efficiency only at around the maximum output power level. In this thesis, we propose energy efficient system level solutions for wireless access network that consider the non-ideal efficiency characteristics of the PA and the load variations.

We model and incorporate the PA efficiency in the energy-delay trade-off present in Shannon's channel capacity model in order to investigate the energy saving potential in a wireless access network at the cost of additional flow-level delay. We propose a best response iteration based distributed power control algorithm where the cells identify the power levels for different user locations to minimize energy consumption under delay constraints. We observe that energy saving potential strongly depends on the network load and PA efficiency characteristics. We also investigate the impact of additional delay in the downlink on the energy consumption of the mobile terminal.

Heterogeneous network is the leading technology for the next-generation cellular networks. We investigate the energy-efficient densification and load sharing between the layers of a heterogeneous network while taking into consideration the PA efficiency and temporal load variations (TLV). We also study the impact of PA efficiency and energy-delay trade-off on the energy efficient network densification.

Massive MIMO (MM) is another leading candidate technology to cater for very high capacity demand. We consider a multi-cell MM system and provide the guidelines to dimension the PA for the antennas. We also develop energy efficient antenna adaptation schemes that allow the cells to dynamically adapt the number of antennas to the TLV in order to maintain high energy efficiency (EE) throughout the day. Our results indicate that these proposed antenna adaptation schemes can improve the EE significantly.

Keywords Energy efficiency, power amplifier efficiency, energy-delay trade-off, load variations, heterogeneous network, massive MIMO

ISBN (printed) 978-952-60-6863-3 **ISBN (pdf)** 978-952-60-6864-0

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2016

Pages 167 **urn** http://urn.fi/URN:ISBN:978-952-60-6864-0

Preface

The research work for this doctoral thesis has been carried out at the Department of Communications and Networking of Aalto University, Finland. The main body of this work has been done in the scope of TEKES funded ECEWA and EWINE-S projects and EIT ICT lab funded 5GrEEn project. This thesis could have never been completed without the help and guidance of numerous individuals.

First and foremost, I would like to express my utmost gratitude and appreciation to my supervisor Prof. Riku Jäntti for his continuous guidance, support and encouragement. It has been a pleasure and a great experience to work with him on this research as he was always available with excellent answers to my questions.

I would like to extend my gratitude to Dr. Konstantinos Koufos for his constructive suggestions and insightful feedback. Countless enlightening discussions with him kept me motivated.

I would like to thank and acknowledge my other co-authors Dr. Cicek Cavdar and Prof. Emil Björnson. Their insightful comments and guidance helped improving the quality of a few of the publications.

I would also like to thank the thesis pre-examiners, Prof. Mehdi Bennis and Dr. Fabien Heliot for reviewing this manuscript. Their constructive comments and invaluable suggestions helped in improving the clarity and presentation of the thesis. I would also like to thank Prof. Jukka Lempiäinen for accepting to be the opponent in the public defense of this thesis.

I extend my thanks to Viktor Nässi to make my stay in the department pleasant and smooth. Thanks goes to support personnel particularly Sari Kiveliö and Sanna Patana for travel, study and other arrangements.

My warmest thanks to my colleagues and friends Byungjin Cho, Mirza Nazrul Alam, Dr. Aamir Mahmud, Beneyam Haile, Furqan Ahmed, Shahrukh Bin Ali, Dr. Udesh Sanjika Oruthota and Bikramjit Singh and others for their company and support during my time at the department.

Last but not least, I would like to thank my wife Laila Haque for her understating

and encouragement during the past few years. My deepest gratitude goes to my parents, Mohammad Abdul Bari and Asia Bari for their dedication and unconditional support. Patience and encouragement from my family was absolutely necessary to make this dissertation possible.

Helsinki, June 2, 2016,

M. M. Aftab Hossain

Contents

Preface i

Contents iii

List of Publications v

Author's Contribution vii

List of Abbreviations ix

List of Symbols xi

1. Introduction 1

1.1 Background and Motivation 3

1.1.1 Power Amplifier Efficiency 3

1.1.2 Energy-Delay Trade-off 4

1.1.3 Massive MIMO 5

1.2 Thesis Focus and High Level Questions 6

1.3 Summary of the Publications 7

2. Methodology and Theoretical Models 9

2.1 PA Efficiency Model 9

2.2 Base Station Power Consumption Model 10

2.2.1 Baseband Power Consumption 10

2.2.2 Total Power Consumption for Fixed Power per BS 11

2.2.3 Total Power Consumption for Fixed Power per Antenna . . 11

2.3 Traffic Model 12

2.4 Rate and Interference Model 14

2.4.1 Rate and Interference Model for SISO 14

2.4.2 Rate Model for Massive MIMO 16

2.5 S-modular Games 17

3. Energy-Delay Trade-off **19**

3.1 PA Efficiency Aware Energy-Delay Trade-off 19

3.1.1 Problem Definition 20

3.1.2 Energy Saving Game 22

3.2 Combined Uplink and Downlink 23

3.2.1 Problem Formulation 24

3.2.2 Optimization Algorithm 24

3.3 Impact of PA Efficiency and Delay on Energy Efficient HetNets . . 25

3.3.1 System Model 25

3.3.2 Problem Formulation 26

3.3.3 Optimization Algorithm 27

3.4 Numerical Evaluation 28

3.5 Discussion 34

4. Energy Efficient Antenna Adaptation in Massive MIMO **37**

4.1 Energy Efficiency of Massive MIMO 37

4.2 Problem Definition 38

4.3 Fixed Power per BS 39

4.3.1 Dimensioning of PA 40

4.3.2 Reference Network 40

4.3.3 Numerical Evaluation 40

4.4 Average Fixed Power per Antenna 42

4.4.1 EE Maximization Game 43

4.4.2 Performance Evaluation 44

4.5 Discussion 47

5. Discussion and Future Work **49**

References **53**

Errata **63**

Publications **65**

List of Publications

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

I Hossain M. M. A. and Jäntti R. Impact of efficient power amplifiers in wireless access. In *IEEE Online Conference on Green Communications (GreenCom)*, New York, pp. 36 - 40, Sep 2011.

II Hossain M. M. A., Koufos K. and Jäntti R. Minimum-energy power and rate control for fair scheduling in the cellular downlink under flow level delay constraint. *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3253 - 3263, Jul 2013.

III Hossain M. M. A., Koufos K. and Jäntti R. How much energy can be saved by energy-delay tradeoff in Radio access network?. In *IEEE Online Conference on Green Communications (GreenCom)*, Piscataway, NJ, pp. 121 - 126, Oct 2013.

IV Hossain M. M. A., Koufos K. and Jäntti R. Energy efficient deployment of HetNets: Impact of power amplifier and delay. In *IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, pp. 778 - 782, Apr 2013.

V Hossain M. M. A., Cavdar C. and Jäntti R. Dimensioning of PA for massive MIMO system with load adaptive number of antennas. In *IEEE Globecom Workshops*, Austin, TX, pp. 1102 - 1108, Dec 2014.

VI Hossain M. M. A., Cavdar C., Björnson E. and Jäntti R. Energy-efficint load-adaptive massive MIMO. In *IEEE Globecom Workshops*, SanDiego, CA pp. 1 - 6,

Dec 2015.

VII Hossain M. M. A., Cavdar C., Björnson E. and Jäntti R. Energy saving game for Massive MIMO: Coping with daily load variations. *submitted to IEEE Transactions on Vehicular Technology*, May 2016.

Author's Contribution

Publication I: "Impact of efficient power amplifiers in wireless access"

The problem was motivated by Prof. Riku Jäntti. The author carried out the analysis, generated results and wrote the paper under the supervision of Prof. Riku Jäntti.

Publication II: "Minimum-energy power and rate control for fair scheduling in the cellular downlink under flow level delay constraint"

Prof. Riku Jäntti proposed the power allocation scheme. The author generated the results and wrote the paper together with Dr. Konstantinos Koufos under the supervision of Prof. Riku Jäntti.

Publication III: "How much energy can be saved by energy-delay tradeoff in Radio access network?"

The author formulated the problem, carried out the analysis and generated the results. The author wrote the initial draft and the other co-authors were involved in editing the draft.

Publication IV: "Energy efficient deployment of HetNets: Impact of power amplifier and delay"

The author formulated the problem, carried out the analysis and generated the results together with Dr. Konstantinos Koufos under the supervision of Prof. Riku Jäntti.

Publication V: "Dimensioning of PA for massive MIMO system with load adaptive number of antennas"

The author formulated the problem together with Prof. Riku Jäntti and generated the results. The author wrote the initial draft and the other co-authors of this paper were involved in editing the draft.

Publication VI: "Energy-efficint load-adaptive massive MIMO"

The author formulated the problem together with Prof. Riku Jäntti, and generated the results. The author wrote the initial draft and the other co-authors of this paper were involved in editing the draft.

Publication VII: "Energy saving game for Massive MIMO: Coping with daily load variations"

The author formulated the problem together with Prof. Riku Jäntti and Dr. Cicek Cavdar. The author proposed the best response algorithm and proved its convergence, carried out the numerical analysis and wrote the initial draft. The other co-authors of this paper were involved in editing the draft.

List of Abbreviations

BRI	Best Response Iteration
BRIA	Best Response Iteration Algorithm
BS	Base Station
CDMA	Code Division Multiple Access
CSI	Channel State Information
DC	Direct Current
DLP	Daily Load Profile
EER	Envelope Elimination and Restoration
ET-PA	Envelope Tracking Power Amplifier
HetNet	Heterogeneous Network
MIMO	Multiple Input Multiple Output
MM	Massive MIMO
MT	Mobile Terminal
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PAE	Power-Added Efficiency
PAPR	Peak-to-Average Power Ratio
PASTA	Poisson Arrival See Time Averages
RF	Radio Frequency

RV	Random Variable
SE	Spectral Efficiency
SINR	Signal to Interference plus Noise Ratio
SISO	Single Input Single Output
TDD	Time Division Duplex
TLV	Temporal Load Variation
TPA	Traditional Power Amplifier
TSLV	Temporal and Spatial Load Variation
WCDMA	Wideband Code Division Multiple Access

List of Symbols

$\lvert \cdot \rvert$	cardinality of the set

Greek Symbols

α	pilot reuse factor
γ_u	SINR achieved by the user at location u
$\Delta_{\max}^{BS}$	maximum throughput reduction factor for BS
$\Delta_{\max}^{MT}$	maximum throughput reduction factor for MTs
ϵ	PA dependent parameter for ET-PA.
$\mathcal{E}$	utility function of the players
η	maximum efficiency of the PA
λ	arrival rate (packet per second)
λ_u	arrival rate (packet per second) at location u
λ_T	user arrival rate
$\pi_c(n)$	probability distribution of random number of users n in cell c
ρ_u	flow level throughput for user at location u
σ	standard deviation

Latin Symbols

$f(.)$	factorial
g	power loss mainly due to the path loss
g_{cck}	channel variance from the serving BS
g_{dck}	channel variance from cell d to users in cell c
g_u	distance based pathloss at location u

G_{cc}	average of inverse of the channel variances from the serving BS to the users
G_{cd}	average of ratio between the channel variance from the serving BS and the interference BSs
I_u	interference received by the user at location u
K_c	number of active users in cell c
$K_{\max}$	number of maximum users served at a time
$\mathcal{L}_c$	set of locations in any cell c
M_c	number of active antennas in cell c
$\mathbb{M}_c$	matrix of antennas for all users states and time intervals by cell c
$\mathcal{M}_c^{(h)}$	vector of antennas for different user states in cell c at time interval h
$M_{\max}$	number of maximum anteannas by a BS
N_0	one sided power spectral density of noise
p	BS average transmit power
P_c	total transmit power by the BS c
P_{idle}^{BS}	power consumed at idle state by the BS
P_{idle}^{MT}	power consumed at idle state by the MT
P_{idle}^{M}	power consumed at idle state by the macro BS
P_{idle}^{m}	power consumed at idle state by the micro BS
p_u	BS transmit power for the user at location u
$\mathbf{p}_c$	vector of transmit powers by BS for the users of cell c
$\mathbf{p}_{-c}$	vector of transmit powers for users by any cell other than c
$\mathbf{p}_c^{MT}$	vector of transmit powers by the MTs of cell c
$\mathbf{p}_{-c}^{MT}$	vector of transmit powers by MTs of any cell other than c
$\mathbf{p}_c^{BS}$	vector of BS transmit powers for the MTs of cell c
$\mathbf{p}_{min,c}^{BS}$	vector of minimum BS transmit powers for the MTs of cell c
$\mathbf{p}_{min,c}^{MT}$	vector of minimum transmit powers by the MTs of cell c
$\mathbf{p}^{M}$	vector of transmit powers by macro BS
$\mathbf{p}^{m}$	vector of transmit powers by micro BS
$P_{\text{max,PA}}$	designated maximum output power of the PA
$P_{ET-PA}(p)$	input power to the ET-PA for BS output power p
P_c^{tot}	total power consumption in cell c
P_{idle}	power consumed by at idle state
P_{stat}	static power consumption in the BS
$P_{TPA}(p)$	input power to the TPA for BS output power p
r_u	data transmission rate for the user at location u
R_c	average rate achieved by any user in cell c
s	data traffic contribution of a single user

S_u	packet size for the user at location u
$\mathcal{S}_c$	strategy space of cell c
T_c	length of channel coherence interval (in symbols)
$\mathcal{U}_c$	set of user states served by cell c
W	operating bandwidth
$\mathcal{W}_0$	Lambert function

1. Introduction

"We do not inherit the Earth from our ancestors we borrow it from our children." *Ancient proverb*

In order to meet the exponential growth of mobile data demand and ensure seamless coverage, wireless access networks are undergoing widespread proliferation. The consequent increase in energy consumption by wireless networks is not sustainable as wireless access networks already consume 0.5 percent of global energy supply and there is a general drive to reduce CO_2 emission, e.g., a target to lower CO_2 emissions 20 percent by 2020 was set by the European Commissions in 2008 [1]. Moreover, the share of annual energy cost is constantly increasing and currently it contributes to around 50 percent of the total operating expenditure (OPEX) to run a network [2, 3]. Unfortunately, the revenue from catering for the exponential increase in capacity demand may not increase at all. Business viability as well as environmental awareness have sparked keen interest in developing systems and solutions that can achieve high capacity with a significant increase in energy efficiency (EE) [3–13].

To design an energy efficient network, it is important to consider the fact that there is a significant variation in traffic demand throughout the day. Recent data shows that the daily maximum loads are even two to ten times higher than the daily minimum loads, see Fig. 1.1. Unfortunately, the dimensioning of contemporary wireless networks is dominated by the capacity and coverage requirements during busy hour without assigning much importance to EE. As a result, a base station (BS) running without any actual load consumes around half the energy it requires to run at full load [14]. In a recent study [15], it is shown that the total energy consumed by a network for actual user data transmissions is about 10 percent and the rest is mainly consumed by the BSs to ensure coverage. It is also shown that only 20 percent of the BSs carry 80 percent of the traffic, whereof only five to ten percent of the BSs are highly loaded [16].

Moreover, the power consumption of cellular networks is mostly dominated by

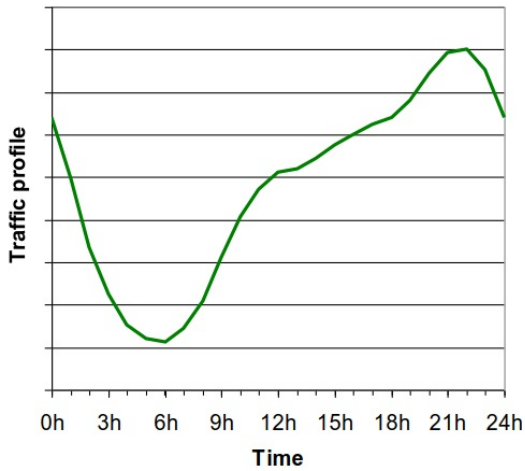


Figure 1.1. Daily data traffic variation in Europe [14]

macro BSs where the power amplifier (PA) consumes around $55-70$ percent of total energy [14]. Unfortunately, the efficiency of PAs are high only near the maximum output power and keeps diminishing when transmit power reduces. One of the main challenges of designing energy efficient wireless networks lies in coping with this temporal and spatial load variations (TSLV) given the non-ideal efficiency characteristics of the PAs.

In this thesis, we aim to provide guidelines for the design and operation of an energy efficient wireless network. We consider the non-ideal efficiency characteristics of the PA while proposing the energy efficient solution for a wireless access network that copes with the TSLV. We model and incorporate the PA efficiency characteristics in the energy-delay trade-off present in Shannon's capacity model in order to investigate the energy saving potential by allowing concedable excess delay. Our best response based distributed algorithm identifies the transmit powers for different users' location in the presence of inter-cell interference and slow fading in order to minimize total energy consumption under delay constraint. We also investigate the impact of this additional delay in the downlink on the energy consumption of the mobile terminal (MT).

In order to cater for very high capacity, the two leading candidate technologies are i) densification of wireless network with small cells, i.e., increasing the number of BSs in a given area and ii) mounting a huge number of antennas at the BS, i.e., deployment of massive MIMO (MM) [17]. In this thesis, we devise schemes to improve the EE of both the technologies taking into consideration the non-ideal efficiency characteristics of the PA and the temporal load variation (TLV). We provide the guidelines for the PA efficiency and TLV aware energy efficient densification of a wireless network with small cells. For MM systems, we provide guidelines to dimension the PA for

the antennas and develop energy efficient antenna adaptation schemes to cope with the TLV in order to maintain high EE throughout twenty-four hour operation. Note that we use the daily load profile (DLP) proposed in [14, 18] to represent TLV. Also note that both the maximum energy saving and EE reported in this thesis refer to the ones achieved through the convergence of the BRI at the Nash equilibrium.

The structure of the thesis is as follows: A general background to the research problem is introduced in Section 1.1. The main research questions addressed in this thesis are defined in Section 1.2. A summary of the main contributions of this thesis is presented in Section 1.3. Chapter 2 contains the general assumptions and models used in this thesis. In Chapter 3, first we discuss the energy saving potential of energy-delay trade-off in the downlink of a wireless system and its impact on the energy consumption of a mobile terminal (MT). Then we study the energy efficient densification of a wireless network with small cells and investigate the impact of DLP and PA efficiency on the EE-aware switching on/off of small cells. Chapter 4 discusses the dimensioning of PAs for MM systems and the schemes developed to improve EE of MM systems by adapting the number of antennas to the DLP. The thesis is concluded in Chapter 5.

1.1 Background and Motivation

1.1.1 Power Amplifier Efficiency

Different classes of traditional power amplifiers (TPAs) are bounded by maximum achievable efficiencies which are attained at around the maximum designated output power, i.e., near the compression region [19]. Unfortunately, recent technologies such as code division multiple access (CDMA) and orthogonal frequency-division multiplexing (OFDM) employ non-constant envelope modulation schemes where the transmitters rarely operate at the maximum transmit power. This requires higher efficiency at output power levels which are lower than the designated maximum. Fortunately, there are several techniques to improve the efficiency of a PA throughout the operating range, e.g., [20]:

- The Envelope Tracking Power Amplifier (ET-PA): ET-PA utilizes a linear PA and a supply modulation circuit where the supply voltage tracks the input envelope [21]. With the help of a dc-dc converter the supply voltage V_{DD} is adjusted according to the RF envelope that needs to be amplified.

- Doherty Amplifier: The Doherty amplifier consists of two or more PAs which operate in parallel. The main device typically saturates at a certain power; for higher power other devices join and altogether reach the maximum output power resulting in multiple efficiency peaks [22, 23].

- Envelope Elimination and Restoration (EER): EER uses a combination of a high-efficiency switch mode PA with an envelope re-modulation circuit. The carrier is amplified by a non-linear amplifier and the envelope of the signal is amplified by a lower frequency but highly linear amplifier. Finally, both signals are combined at the output to resurrect the amplified version of the original signal [20].

Considering $7 - 8$ dB peak-to-average power ratio (PAPR), ET-PA is superior to other techniques that ensure improved efficiency in the required dynamic range of output power levels. This is due to the fact that ET-PA offers higher overall average power-added efficiency (PAE), increased available peak power, very low correction gain and wider tunable bandwidth compared to Doherty amplifiers [24]. On the other hand, though the EER is comparable to ET-PA in terms of efficiency improvement, it is comparatively more sensitive to the path mismatch effect and suffers from the slowness of the envelope restoration feedback loop [25, 26]. Due to its superiority, the ET-PA has already been widely accepted by the wireless industry as the preferred PA and about eight billion ET-PA are expected to be shipped by 2020 [27].

1.1.2 Energy-Delay Trade-off

Shannon's capacity under an Additive White Gaussian Noise (AWGN) channel can be expressed as

$$C = \frac{1}{2}\log_2\left(1 + \frac{gp}{N}\right) \tag{1.1}$$

where g represents the power loss mainly due to the path loss and N is the noise power. Now, if the time necessary to transmit one bit of information is given by $t = \frac{1}{C}$, the energy per bit E_b can be given by [28]

$$E_b(t) = tP_{PA} = \frac{t}{\eta(p)}\frac{N}{g}\left(2^{\frac{2}{t}} - 1\right) \tag{1.2}$$

where $P_{PA} = \frac{p}{\eta(p)}$ is the actual power consumed by the PA when transmitting the average power p as $\eta(p)$ gives the efficiency of the PA when transmitting power p.

It is evident from (1.2) that, if the PA efficiency $\eta(p)$ is a constant number, the energy consumption for transmitting a bit of information can be reduced if additional transmission time is conceded (e.g., see plot for ideal-PA in Fig. 1.2). This trade-off

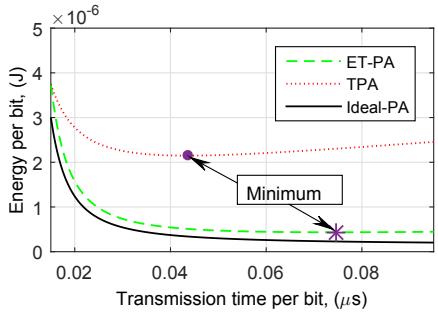


Figure 1.2. Energy per bit versus transmission time per bit corresponding to SNR between 0.15 dB and 20 dB which is typical for a wireless link.

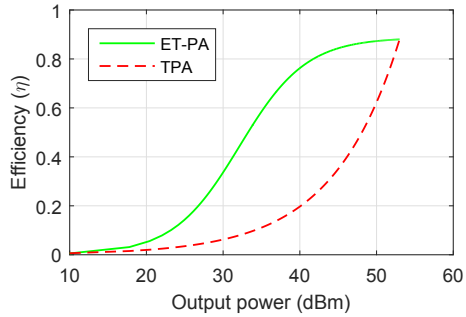


Figure 1.3. Efficiency versus output power characteristics for both TPA and ET-PA based on the schematic proposed in [29].

provides the opportunity to reduce transmit power in the PA at the cost of additional delay and save energy [30–32]. However, Fig. 1.3 shows that the efficiency of a non-ideal PA reduces when it operates away from the maximum output power. In reality, the PAs rarely operate close to its maximum output power due to the fact that modern technologies, e.g., WCDMA and OFDM, employ a waveform whose PAPR is around $7-8$ dB and requires good linearity [33]. If the non-ideal efficiency characteristics of the PAs are taken into account, the energy consumption keeps decreasing until the delay goes beyond a certain limit, see the plots for TPA and ET-PA in Fig. 1.2.

1.1.3 Massive MIMO

The MIMO (Multiple-input and multiple-output) technique is a means to increase throughput, diversity and mitigate interference using multiple transmit and receive antennas to exploit multipath propagation. For single-user MIMO, it has been established that with M transmit antennas and N reception antennas, capacity scales approximately $\min(M, N)$ times compared to a single antenna system [34]. Similar capacity scaling is achievable in multi-user MIMO when a BS with M antennas communicates with N single antenna MTs [35].

Recently, Marzetta T. L. has introduced the concept of massive MIMO showing

the potential SE and EE gains from the deployment of an excess of antennas in TDD mode of operation [36]. In MM systems, each BS uses hundreds of antennas to be able to simultaneously serve tens of MTs on the same time-frequency resource, by downlink multi-user precoding and uplink multi-user detection [37]. MM systems are expected to play a pivotal role in delivering very high capacity in next generation networks (NGNs) as the channel vectors between users and the BS becomes nearly orthogonal due to the massive number of antennas for most propagation environments [37]. The resulting EE gain is also significant as it is possible to get $O(M)$ array gains and $O(K)$ multiplexing gains under such *favorable propagation* [38]. Here, M stands for the number of antennas and K stands for the number of users served at a time. The EE is the benefit-cost ratio of a network and is defined as the number of bits transferred per Joule of energy and hence can be computed as the ratio of average sum rate (in bit/second) and the average total power consumption (in Joule/second) [39]. In [40], the authors show that MM replacing a SISO system improves the spectral efficiency (SE) and radiated EE simultaneously by more than a hundred times. However, in [41], the authors show that if the circuit power is also considered, increasing the number of antennas is likely to decrease the EE given that each antenna has usually its own RF chain. Further improvement in the EE of an MM system is possible by applying different techniques, e.g., adaptation of number of antennas to the network load, transceiver redesign and operation in mm-wave [38].

1.2 Thesis Focus and High Level Questions

The target of several orders of magnitude increase in EE necessitates a paradigm shift in the way the wireless network, infrastructure and its components are designed, dimensioned, managed and operated. The key areas of interest can be broadly summarized as i) hardware components with capability of selective turning on/off ii) link and network level resource management iii) planning and deployment of small cells and/or MM systems, and iv) network operation [3, 6, 7, 10, 14].

In this thesis, we focus on the EE of the wireless access network and devise energy efficient solutions at the network level. The main goal of the thesis is:

*to reduce the **energy consumption** of contemporary wireless access networks and devise a way forward to meet exponential capacity growth in an **energy efficient** manner while considering the impediments imposed by the **load variations** and the **non-ideal efficiency characteristics** of PAs.*

The high level questions addressed in this thesis are

- HQ1: How to reduce energy consumption of the traditional wireless access networks considering the TSLV and the non-ideal efficiency characteristics of PA?
- HQ2: How to design future wireless access networks that can cater for very high capacity and maintain very high EE while coping with the TLV and non-ideal efficiency characteristics of PA?

1.3 Summary of the Publications

This thesis is an overview of seven publications. In the first three publications, we study and model the non-ideal energy efficiency characteristics of the PA, devise mechanisms to save energy in the PA at the cost of small additional delay in a multi-cell environment and then study the impact of this additional delay on the energy consumption of the MTs. We take the DLP into consideration in order to quantify the energy saving potential by the energy-delay trade-off. The next four publications deal with the PA efficiency and DLP aware energy efficient heterogeneous network (HetNet) and load-adaptive MM systems.

In Publication I, we study the impact of PA efficiency characteristics on energy consumption by a cellular network. We model the efficiency characteristics of ET-PA and investigate the energy saving potential by replacing the TPA with the ET-PA in a WCDMA network.

In Publication II, we study the energy saving potential of energy-delay trade-off in the downlink of a cellular system in the presence of inter-cell interference and fading while considering the non-ideal PA efficiency characteristics. We investigate the potential for energy saving by lowering the transmit power level at different user locations in a cell at the cost of additional flow level delay. We model the traffic at flow level and use the concept of flow throughput in order to account for and control additional delay. The BS is modelled as an $M/G/1-PS$ queue (as per Kendall's notation) which mimics the proportional fair scheduling [42]. We develop a best response iteration (BRI) based distributed power control algorithm where each BS independently decides the transmit power for different location based users in the cell in order to minimize energy consumption. The convergence of the algorithm has been proven under a game theoretic approach.

In publication III, we investigate the impact of additional delay in the downlink on the energy consumption of MTs due to prolonged reception time and quantify the energy saving potential in the downlink without any significant increase in power consumption in the MT. We also study how much energy can be saved in both the

uplink and downlink of a wireless access network by energy-delay trade-off under delay constraints.

The HetNet and MM systems are two leading technologies to cater for very high capacity in order to meet the exponential growth of capacity demand. In Publication IV, we study the DLP and PA aware energy efficient HetNet. We consider a HetNet consisting of macro BSs and out-band micro BSs. We show that sequential switching on/off of the micro BSs along with always on macro BS is the way forward to cater for increasing capacity demand. We also investigate the impact of PA efficiency and additional delay on the EE-aware switching on/off of the micro BSs. Our results indicate that inefficient PAs not only consume more energy but also require turning on the micro BSs at lower loads compared to efficient PAs. This leads to the requirements of a higher number of BSs to serve same areas and loads for certain load conditions when using inefficient PAs.

In Publications I-IV, we use an improved version of the BS power consumption model proposed by the EARTH (Energy Aware Radio and neTwork tecHnologies) consortium [14]; more specifically, the dynamic power consumption in the PA incorporates the non-ideal efficiency characteristics of the PA. This model maps the transmit power to the actual total input power consumption of the BS more accurately.

In Publications V-VII, we devise mechanisms to dynamically adapt the number of antennas of BSs with the MM to the DLP in order to maximize the EE throughout the day. In Publication V, we take into consideration the non-ideal efficiency characteristics of the non-ideal PAs along with the DLP in order to investigate the impact of PA dimensioning on EE in a load-adaptive MM system. We consider the total transmit power per BS to be fixed for this analysis. Our results indicate that there is a clear choice for the optimum dimension of the PA and deviating from that choice reduces the EE of a network. In Publication VI and VII, we relax the fixed total power constraint for the BS and dimension an MM system that copes with the TLV and maintains high EE throughout the day while considering the non-ideal efficiency characteristics of PAs. As the BS serves a number of users simultaneously and the user rate depends on the number of active users and active antennas, we model the BS as a state-dependent $M/G/m/m$ queue [43, 44]. We formulate the joint EE maximization problem for the cells where each cell identifies the required number of active antennas at each user state that maximizes EE through BRI. We prove the convergence of the BRI based algorithm under a game theoretic approach. Our results indicate that the EE of the network can be significantly improved, especially, when the network load is low to medium at the cost of user data rate. Also, we improve the power consumption model of MM by considering the non-ideal PA efficiency.

2. Methodology and Theoretical Models

This chapter provides the details about the models and methodology followed in this thesis.

2.1 PA Efficiency Model

We consider both TPA and ET-PA in this analysis. In the TPA case, the maximum efficiency is achieved only when it operates near the compression region, i.e., at the maximum output power. On the other hand, ET-PA achieves better efficiency throughout the operating range by tracking the radio frequency (RF) signal and regulating the supply voltage accordingly. In order to transmit an average output power p, the total input power needed by a BS with TPA can be approximated as [45]

$$P_{TPA}(p) \approx \frac{1}{\eta}\sqrt{p \cdot P_{\text{max,PA}}} \tag{2.1}$$

where η denotes the maximum PA efficiency when transmitting the designated maximum output power $P_{\text{max,PA}}$ of the PA. Note that p reflects the average transmit power of the BS and the PA must be able to handle the peak power, i.e., p must be at least 8 dB less than $P_{\text{max,PA}}$ as the typical PAPR of recent technologies, e.g., OFDM, is 8 dB [46].

In Publication I, we model the efficiency characteristics of ET-PA based on those suggested in [29] where the efficiency characteristics have been derived by taking the ET trajectory (i.e., by modulating the supply voltage according to the RF input power) of a representative sequence of modeled results of efficiency vs. output power at different fixed power supply voltage. By curve fitting, the operating efficiency of the ET-PA $\eta(p)$, at any output power level p can be modeled with the following equation

$$\frac{\eta(p)}{\eta} = \frac{p(1+\epsilon)}{p+\epsilon P_{max,PA}}. \tag{2.2}$$

The total power consumption is approximately a linear function of the actual transmit

power, i.e.,

$$P_{ET-PA}(p) \approx \frac{1}{(1+\epsilon)\eta}\left(p + \epsilon P_{\max,\mathrm{PA}}\right) \tag{2.3}$$

where $\epsilon \approx 0.0082$ is a PA dependent parameter. Note that we do not consider the power consumed by the voltage supply modulation circuit.

2.2 Base Station Power Consumption Model

For the Publications II-IV, we consider single input single output (SISO) systems, i.e., each of the BSs and MTs is equipped with a single antenna. we consider the widely used BS power consumption model proposed by the EARTH project for this SISO system [14]. However, instead of using fixed efficiency for the PAs, we use the dynamic PA efficiency models which are given in Section 2.1. For the Publications V-VII, we analyse the energy efficiency of MM systems. We assume that each antenna has its own PA. The total power consumed in a BS c with M_c active antennas and K_c users is modeled as [39, 47]

$$P_c^{tot}(K_c, M_c) = P_{\mathrm{BB}}(K_c, M_c) + M_c P_{PA}(p) + P_{\mathrm{Oth}} \tag{2.4}$$

where $P_{\mathrm{BB}}(K_c, M_c)$ is the base band signal processing power when the BS serves K_c number of users simultaneously with M_c number of antennas and $P_{PA}(p)$ gives the DC input power in the PA when the average output power is p. P_{Oth} includes the power consumption due to site cooling, control signal, DC-DC conversion loss, etc. For MM systems, base band power consumption is significant as some parts of this consumption scale with the number of antennas and users as discussed in Section 2.2.1.

2.2.1 Baseband Power Consumption

For baseband and fixed power consumption we use the model proposed in [47]. The total circuit power is given by

$$P_{\mathrm{BB}} = P_{\mathrm{TC}} + P_{\mathrm{CE}} + P_{\mathrm{C/D}} + P_{\mathrm{LP}}. \tag{2.5}$$

The power consumed in the transceiver is given by $P_{\mathrm{TC}} = MP_{\mathrm{BS}} + P_{\mathrm{SYN}}$ where M is the number of antennas, P_{BS} is the power required to run the circuit components, e.g., converters, mixers and filters attached to each antenna at the BS, and P_{SYN} is the power consumed by the local oscillator. P_{CE} is the power required for channel estimation. $P_{\mathrm{C/D}} = P_{\mathrm{COD}} + P_{\mathrm{DEC}}$, i.e., the total power required for channel encoding and decoding. P_{LP} is the power consumed by linear processing. According to [47],

the total baseband power can be expressed as

$$P_{\mathrm{BB}}(M_c, K_c) = \underbrace{\mathcal{A}K_c R_c + \sum_{i=0}^{3} C_{0,i} K_c^i}_{C_0^{BB}} + M_c \underbrace{\sum_{i=0}^{2} C_{1,i} K_c^i}_{C_1^{BB}} \tag{2.6}$$

where $C_{0,0} = P_{\mathrm{SYN}}, C_{0,1} = 0, C_{0,2} = 0, C_{0,3} = \frac{B}{3T_c L_{\mathrm{BS}}}, C_{1,0} = P_{\mathrm{BS}}, C_{1,1} = \frac{B}{L_{\mathrm{BS}}}(2 + \frac{1}{T_c}), C_{1,2} = \frac{3B}{L_{\mathrm{BS}}}, \mathcal{A} = P_{\mathrm{COD}} + P_{\mathrm{DEC}}$, K_c is the number of users served at a time by cell c, R_c is the average user rate as discussed in Section 2.4.2 and B is the bandwidth, L_{BS} is computational efficiency in the BS (Gflops/Watt) and T_c is the coherence interval. The power consumption C_1^{BB} gets multiplied with M_c whereas C_0^{BB} does not.

2.2.2 Total Power Consumption for Fixed Power per BS

In Publication V, we assume the total BS transmit power to be fixed. As a result, with M_c antennas and a total transmit power P_c, the power per antenna of a BS c becomes P_c/M_c. Using the power model for ET-PA as in (2.3) the corresponding total input power for the PAs is

$$\begin{aligned} P_{PA}(P_c, M_c) &= M_c \frac{1}{(1+\epsilon)\eta}(P_c/M_c + \epsilon P_{\mathrm{max,PA}}) \\ &= \underbrace{\frac{P_c}{(1+\epsilon)\eta}}_{C_0^{PA}} + \underbrace{\frac{\epsilon P_{\mathrm{max,PA}}}{(1+\epsilon)\eta}}_{C_1^{PA}} M_c. \end{aligned}$$

So, for both the baseband processing and PA, a part of the power consumption is constant and the other part of the consumption linearly increases with the number of antennas. Combining those parts accordingly the total power consumed by cell c can be written as

$$P_c^{tot}(K_c, M_c) \approx C_0 + C_1 M_c \tag{2.7}$$

where $C_0 = C_0^{PA} + C_0^{BB} + P_{Oth}$ and $C_1 = C_1^{PA} + C_1^{BB}$. Note that for TPA case, $C_0^{PA} = 0$ and $C_1^{PA} = \frac{1}{\eta}\sqrt{P_c P_{\mathrm{max,PA}}/M_c}$.

2.2.3 Total Power Consumption for Fixed Power per Antenna

In Publications VI and VII, we do not consider the total BS transmit power to be fixed. Instead we set the average transmit power per antenna to p. The total input

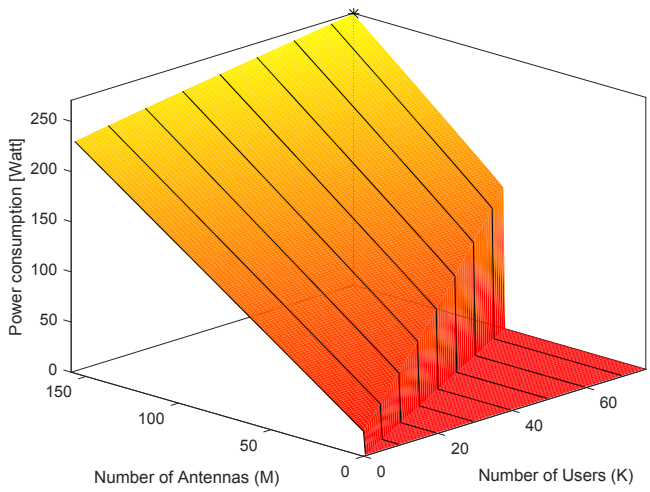


Figure 2.1. Power consumption by BS with TPA according to the model described in Section 2.2.3 with the parameters used in Section 4.4.

power required by M_c antennas is $C_1^{PA} M_c$ where

$$C_1^{PA} = \begin{cases} \frac{1}{\eta}\sqrt{pP_{\max,\mathrm{PA}}}, & \text{if TPA} \\ \frac{p}{(1+\epsilon)\eta} + \frac{\epsilon P_{\max,\mathrm{PA}}}{(1+\epsilon)\eta}, & \text{if ET-PA.} \end{cases} \tag{2.8}$$

Combining with baseband and other power consumptions, the total power consumption can be written as

$$P_c^{tot}(K_c, M_c) \approx C_0 + C_1 M_c \tag{2.9}$$

where $C_0 = C_0^{BB} + P_{Oth}$ and $C_1 = C_1^{PA} + C_1^{BB}$. Fig. 2.1 gives the power consumptions for different combinations of simultaneously served users and active antennas for the parameters used in Section 4.4.

2.3 Traffic Model

In Publications II-IV, we study the energy-delay trade-off under different contexts. In these studies, we model traffic at flow level, i.e., we consider the long-term resource allocation ignoring the short-term packet level complex interactions [48]. In order to quantify additional delay, we utilize the concept of flow throughput which is defined as the ratio of mean flow size to the mean flow transfer delay [48–51]. In order to provide fairness among the users, we consider the most commonly used type of fairness, i.e., proportional fairness [42]. A very good if not accurate throughput performance of proportional fair allocation can be estimated by the so called *Balanced*

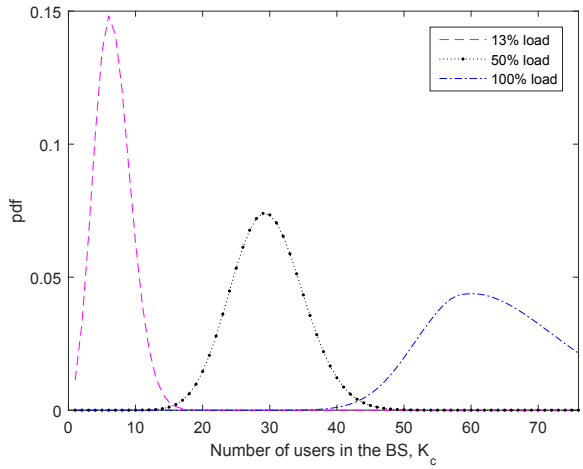


Figure 2.2. Distribution of active users in a single cell at 13%, 50% and 100% of the load with the parameters provided in Section 4.4.2.

fair allocation which can analytically track the steady-state distribution of the flows and is insensitive to any traffic characteristics except the traffic intensity [48, 49, 51]. Balanced fairness requires homothetic allocation which dictates that any change in allocated resource for a given class results in proportional change in data rate of that class [49]. When a BS serves users one at a time and allocates equal time for each user, i.e., mimics a processor sharing queue, it follows proportional fair scheduling. According to [48, 49], this qualifies for homotheticity property.

we model the BS as an $M/G/1-PS$ queue that serves the users in a one-by-one fashion. M indicates that the arrival process is Markovian or memory-less, G indicates that the distribution of service time can be general, 1 means single server as per Kendall's notation [52] and PS stands for processor sharing. The assumption of balanced fairness allows a simple expression for the flow level throughput which has been used in Section 3.1.1.

In Publications V-VII, we study the potential energy saving by adapting the number of antennas of BSs with MM to the DLP. In this case, at any particular time, a BS serves a certain number of users simultaneously (hence termed as user state, i.e, user state n means the BS serves n user simultaneously). The data rate achieved by a user depends on the user state and the number of antennas that are active. In this case, the analytical tractability of the PS queue does not work anymore. For these reasons, we model the BS as a state-dependent $M/G/m/m$ queue (as per Kendall's notation, i.e., number of servers = m, waiting place = 0) [52]. We do not allow any queuing (waiting place = 0) in order to keep the distribution analytically tractable as the $M/G/m$ queue is notorious for its difficulty [53,54]. As any further user requests is not entertained when serving K_{max} users, K_{max} determines the number of servers,

m, i.e, $m = K_{max}$.

The network is assumed to be dimensioned in the way that the data carried by a cell while serving $K_{\max}$ users, corresponds to the peak load of the DLP. In a queuing system with no buffer space, the blocking probability is equal to the probability of having the system 100% loaded, i.e., the probability of having $K_{\max}$ number of users served by the BS simultaneously, according to the Poisson arrivals see time average (PASTA) property [55]. The steady-state probabilities for the random number of users in the BS c, $\pi_c(n) \equiv Pr[K_c{=}n]$, are as follows [43,44]

$$\pi_c(n){=}\left[\frac{\left[\lambda_T \frac{s}{R_c(1)}\right]^n}{n!f(n)f(n-1)...f(2)f(1)}\right]\pi_c(0), n \in \mathcal{U}_c, \tag{2.10}$$

where $\pi_c(0)$ is the probability that there is no user in cell c, and is given by

$$\pi_c^{-1}(0) = 1 + \sum_{i=1}^{K_{\max}}\left(\frac{\left[\lambda_T \frac{s}{R_c(1)}\right]^i}{i!f(i)f(i-1)...f(2)f(1)}\right) \tag{2.11}$$

where $\frac{s}{R_c(1)}$ is the expected service time when BS c serves a single user, $f(n) = R_c(n)/R_c(1)$ where $R_c(n)$ is the data rate per user while serving n number of users, λ_T is the user arrival rate, and s is the total data traffic contribution by a single user. Note that we use (2.13) to find the data rates at different user states. At maximum load the values for λ_T and s are chosen in a way that $K_{\max}$ users in cell c suffer any specific blocking probability (e.g., we use 2%) when the BSs run with $M_{\max}$ antennas. User distributions for other loads are derived using the DLP and λ_T considering s to be fixed. Fig. 2.2 gives three example plots of the user distributions for 13%, 50% and 100% loads with the parameters provided in Section 4.4.2.

2.4 Rate and Interference Model

2.4.1 Rate and Interference Model for SISO

In Publications II-IV, we consider single input single output (SISO) systems, i.e., both the BS and MT are equipped with a single antenna. Assuming that full CSI is available at the BS, the information theoretical capacity of the k-th BS and the u-th user channel is

$$r_u(p_u; I_u(\mathbf{p}_{-k})) = W \log_2\left(\Gamma_u(p_u, I_u(\mathbf{p}_{-k}))\right) \tag{2.12}$$

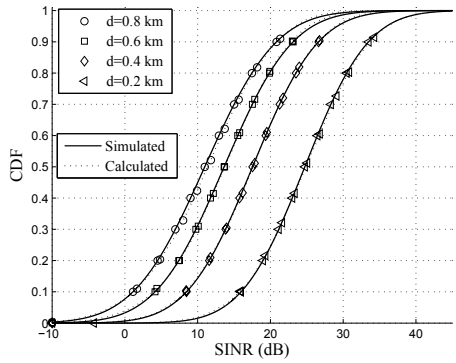


Figure 2.3. SINR distribution at different locations with load 80% and delay increase factor, $1/\Delta = 2$. The parameter d denotes the distance in km from the BS generating the useful signal. The rest of the parameters can be found in Table 3.1.

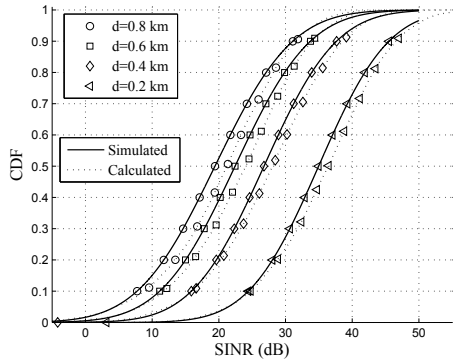


Figure 2.4. SINR distribution at different locations with load 50% and delay increase factor, $1/\Delta = 2$. The parameter d denotes the distance in km from the BS generating the useful signal.The rest of the parameters can be found in Table 3.1.

where $\Gamma_u(p, I) = 1+\gamma_u(p, I)$ and γ_u denotes the SINR at location u, i.e., $\gamma_u(p, I) = \frac{p_u G_u}{I_u+N_0 W}$, p_u denotes the power allocated for the user at location u, G_u incorporates distance-based pathloss g_u and slow fading, I_u is the interference received by the user at location u, N_0 is the one-sided power spectral density of the noise and W is the positive signal bandwidth. The vector, $\mathbf{p}_{-k}$, contains the transmit power values used by the interfering cells.

As our rate function depends on the SINR and we consider outage capacity due to slow fading, we need to derive the distribution of SINR , γ_u. In order to do that, we need to identify the distribution of the useful signal power $p_u G_u$ and the distribution of the aggregate interference plus noise $(N_0 W + I_u(\mathbf{p}_{-k}))$. Hence, we need a model for the propagation pathloss, a fading model and a model for the aggregate interference.

We consider power law based attenuation with slow fading. The slow fading is modeled by a log-normal random variable (RV) with standard deviation σ. The use-

ful signal power at the location u follows the log-normal distribution with mean $m_{S,u}$ and standard deviation σ both expressed in dB. The mean $m_{S,u}$ can be written in terms of the transmission power level p_u and the distance-based path loss g_u: $m_{S,u} = 10\log_{10}(p_u \cdot g_u)$. The aggregate interference plus noise $(I_u(\mathbf{p}_{-k}) + N_0W)$ is a sum of log-normal RVs shifted with the noise power level. In order to model the distribution of the RV $(I_u(\mathbf{p}_{-k}) + N_0W)$, we use the Wilkinson approximation method [56, 57]. According to that method, the sum of log-normal RVs is approximated by another log-normal RV, $(I_u(\mathbf{p}_{-k}) + N_0W) \rightarrow 10^{y/10}$. The first two moments of the RV $10^{y/10}$ and the RV $(I_u(\mathbf{p}_{-k}) + N_0W)$ are forced to be equal to each other. The accuracy of the Wilkinson approach for the sum of log-normal RVs has been examined in [56]. Also, in Publication II, we validate the approximation by taking resort to a Monte-carlo simulation for equally loaded cells. From Fig. 2.3 and Fig. 2.4, it is evident that the calculated and simulated distributions match quite well with each other, particularly in the lower tail which is of interest.

2.4.2 Rate Model for Massive MIMO

In Publications V-VII, we consider the downlink of a multi-cell MM system consisting of cells with indices in the set $\mathcal{C} = \{1, 2, ..., C\}$ where BS $c \in \mathcal{C}$ uses M_c antennas to serve K_c users from the set $\mathcal{U}_c = \{1, 2, ..., K_{\max}\}$. Note that each cell has its own BS and hence we use cell and BS interchangeably. Let us denote the average transmit power of cell c by P_c and of any other cell d, $d \in \mathcal{C}$, $d \neq c$, by P_d. We assume that the BSs and MTs employ a time-division duplex (TDD) protocol, and are perfectly synchronized. The BS obtains perfect channel state information (CSI) from the uplink pilots, which is a reasonable assumption for low-mobility scenarios. We also assume that the aggregate interference suffered by the MTs are obtained by the BS through uplink control channel. Each BS employs zero forcing precoding, and equal power allocation is applied to utilize the pathloss differences to achieve high average data rates. Note that R_c, the average data rate of the users in cell c, is a function of K_c, P_c and P_d. R_c is limited by the capacity, which is generally unknown in multi-cell scenarios, but a tractable and achievable lower bound in cell c under the above assumptions is given by [58]

$$R_c = W\left(1-\frac{\alpha K_{\max}}{T_c}\right)\log_2\left(1+\frac{\frac{P_c}{K_c}(M_c-K_c)}{N_0WG_{cc}+\sum_{d\neq c}P_dG_{cd}}\right) \tag{2.13}$$

where α is the pilot reuse factor, $K_{\max}$ is the maximum number of users assumed to be the same for all cells (i.e., the number of pilot sequences = $\alpha K_{\max}$), T_c is the length of the channel coherence interval (in symbols), $\left(1-\frac{\alpha K_{\max}}{T_c}\right)$ accounts for the necessary overhead for channel estimation, $G_{cc} = \mathbb{E}\{\frac{1}{g_{cck}}\}$ and $G_{cd} = \mathbb{E}\{\frac{g_{dck}}{g_{cck}}\}$,

where the random variable g_{cck} is the channel variance from the serving BS and g_{dck} is the channel variance from cell d to users in cell c, $N_0WG_{cc} + \sum_{d \neq c} G_{cd}P_d$ is the average degradation from inter-cell interference and noise. Note that R_c is achieved by averaging over the locations of the users of cell c, thus it can either be viewed as an average rate among the users in the cell or a rate that every user will achieve when it is moving around in the cell. As a result, the slow fading also gets averaged out.

2.5 S-modular Games

In Publication II and Publication VII, we propose two distributed power control algorithms. In Publication II, we optimize the BS transmit power for users at different locations under inter-cell interference in order to minimize the total energy consumption. Due to inter-cell interference, this power allocation for each user depends on the transmit power of other interfering BSs. In Publication II, we show that such a joint optimization problem is neither convex in terms of the power allocation vectors of the cells nor can be written as the ratio of convex functions for both TPA and ET-PA. However, the utility function has a nice property, which is known as *decreasing differences*, in terms of the power allocation vector of the cells in the sense of Topkis [59]. In Publication VII, the number of antennas at the BSs that maximize EE for different user states is coupled due to inter-cell interference and the joint utility function has *increasing differences* in the number of active antennas used by the cells. Centralized optimization of a network consisting of BSs with MM necessitates exchange of huge information. Therefore, from practical point of view, distributed algorithm is more suitable for multi-cell MM systems. We propose distributed best response iteration algorithm (BRIA) and prove the convergence of this algorithm by relying on a game theoretic approach. Game theory is widely used in order to solve optimization problems in wireless communication [60–62]. We use the framework of s-modular games such that distributed algorithms can be utilized where each BS makes its own decision, i.e., best response, that converges to a Nash equilibrium [59, 63–65].

Supermodular (submodular) games are those characterized by strategic complementarities (substitutabilities) which means that the marginal utility or incremental return of a player's strategy increases (decreases) with increases (decreases) in the other players' strategies [59]. This game is interesting as it possesses important properties [64,65] : 1) there exists a pure strategy Nash equilibrium without necessitating the quasi-convexity of the utility functions, 2) Nash equilibrium can be attained by using BRIA.

In order to study the convergence properties of the best response strategy, we require the following definitions available in [63–66].

Definition 1. *Let a function $\psi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be continuous and twice differentiable. Then, the following two statements are equivalent*

- $\frac{\partial^2 \psi(x,y)}{\partial x \partial y} \geq (\leq) 0, \forall y \in \mathcal{Y}, \forall x \in \mathcal{X}$.

- *The function ψ has increasing (decreasing) differences in (x, y), i.e.*

$$\psi(x', y') - \psi(x, y') \geq (\leq) \psi(x', y) - \psi(x, y), \forall (x', y')$$

where $x' \geq x, y' \geq y$.

The function $\psi(x, y)$ does not have to be nicely behaving, nor does $\mathcal{X}$ and $\mathcal{Y}$ have to be intervals.

Definition 2. *A function ψ is supermodular (submodular) in $x = (x_1, ..., x_m)$ if it has increasing (decreasing) differences in (x_m, x_j) for all $m \neq j$.*

Definition 3. • *If* X *is a convex subset of $\mathbb{R}^n$ and ψ : X $\to$ R is a twice continuously differentiable, then ψ has increasing differences in (x_i, x_j) if and only if $\frac{\partial^2 \psi(x_i,x_j)}{\partial x_i \partial x_j} \geq 0$, and*

- *if $\frac{\partial^2 \psi(x_i,x_j)}{\partial x_i \partial x_j} \geq 0$ for all x and $i \neq j$, then ψ is supermodular.*

Definition 4. *A game is $\mathcal{G}(\mathcal{C}, \mathcal{S}, \mathcal{E})$, is said to be supermodular (submodular) if for all*

- *the strategy space $\mathcal{S}$ is a compact subset of $\mathbb{R}$;*

- *the utility function $\mathcal{E}$ is continuous in all players' strategies ;*

- *the utility function $\mathcal{E}$, has increasing (decreasing) differences in between any component of player k's strategy and any component of any other players strategy.*

Proposition 1. *In a supermodular (submodular) game, a Nash equilibrium exists. Also, if we start with a feasible policy (from the strategy space $\mathcal{S}$), then the sequence of BRIs monotonically increases (decreases) in all components and converges to a Nash equilibrium.*

Proof. These properties of the supermodular (submodular) game follow from [64, Theorem 1]. □

3. Energy-Delay Trade-off

In this chapter, we study the energy-delay trade-off in the downlink of a proportional-fair cellular system with inter-cell interference and fading, while incorporating the efficiency characteristics of non-ideal PAs. In Section 3.1, We develop a game theory based optimization framework that utilizes the energy-delay trade-off in order to minimize energy consumption in the wireless downlink at the cost of additional flow level delay. We utilize this optimization framework in order to study the impact of additional delay in the downlink on the energy consumption of the MT in Section 3.2. We also quantify the potential energy saving in both uplink and downlink under delay constraints. In Section 3.3, we examine the impact of non-ideal efficiency characteristics and energy-delay trade-off on the energy efficient densification of HetNets and load sharing between the layers.

3.1 PA Efficiency Aware Energy-Delay Trade-off

It is evident from (1.2) that the energy required to transmit a 'bit' of information decreases with increasing transmission time monotonically and hence energy consumption to transmit a packet can be reduced by transmitting it over an increased time period with lower transmit power. Based on this observation, Prabhakar B. *et al.* devise the so-called *lazy packet scheduling scheme* to transmit packetized information over a wireless link subject to a delay constraint [30–32]. Based on the packet arrival rate, the *lazy packet scheduling scheme* varies the transmission time of different packets so that the energy consumption is minimized and the delay constraint is met. Berry R. studies the trade-off between average delay and average transmit power for a point-to-point link in a time varying channel with perfect channel state information at the transmitter [67]. Various aspects of the energy-delay trade-off in multi-user scenarios are covered in the literature, see, for instance, different studies in [68–71].

Energy saving potential by discontinuous transmission i.e., letting the BS to trans-

mit at the maximum power as long as needed and then go into micro-sleep mode is studied extensively in [3, 72]. More recent studies in [73–75] under a single BS context show that the joint optimization of BS sleeping control and power matching (i.e., transmit power is adapted to match the target bit rate) have substantial energy saving potential at low traffic load.

The studies in [3, 30–32, 67–75] do not consider the non-ideal efficiency characteristics of the PA while studying the energy-delay trade-off. In reality, the operation of cellular networks is mostly served by macro BSs where the TPA consumes around $55 - 70$ percent of total energy [14, 76]. This is because TPAs attain the maximum efficiency near their compression region, where they rarely operate due to the high PAPR of non-constant envelope signals, e.g., CDMA, OFDM signals.

Han K. *et al.* show that if the PA efficiency characteristics are taken into account, the energy-per-bit starts increasing after the transmission time becomes longer than a certain value [28]. Chen Y. *et al.* show that the non-ideal efficiency of the PA has an impact on beam-forming, modulation level, continuous and discontinuous transmission of a BS, when considered from the perspective of energy efficiency [77]. Wu J. *et al.* summarize the proposed techniques that involve BS or BS component sleeping in a very recent survey paper [7].

A recent study summarizes important open issues regarding energy-delay trade-off which include the investigation of energy-delay trade-off in multi-cell/multi-user interference scenarios [78]. Such analysis should also include the efficiency characteristics of the PA to make the findings concrete. In our work, we aim to fill this gap. We study the energy-delay trade-off in a multi-cell/multi-user cellular downlink scenario and consider slow fading too.

3.1.1 Problem Definition

We consider the downlink of a cellular system consisting of a set $\mathcal{C} = \{1, 2, \cdots C\}$ of cells. A BS is located at the center of each cell. We model the BS as a single server queue that serves the users in a one-by-one fashion. $\mathcal{L}_c$ denotes the finite set of locations from the discretized serving zone (coverage area) of the cell $c \in \mathcal{C}$ and $u \in \mathcal{L}_c$ denotes the location of any user in that cell. The mean energy consumption at the BS for serving a user is obtained by averaging over all the locations $u \in \mathcal{L}_c$

$$\bar{e}_{tot,c}(\mathbf{p}_c, \mathbf{p}_{-c}) = \frac{1}{|\mathcal{L}_c|} \cdot \Big(\sum_{u \in \mathcal{L}_c} \lambda_u S_u \frac{P_{PA}(p_u) + P_{stat} - P_{idle}}{r_u(p_u, I_u(\mathbf{p}_{-c}))} + P_{idle} \Big) \qquad (3.1)$$

where $\mathbf{p}_c$ is the power allocation vector for the c-th BS with elements, $p_u : u \in \mathcal{L}_c$, and $|\cdot|$ denotes the cardinality of the set. λ_u is the flow rate (packets per second) for the location u and S_u is the packet size. $r_u(p_u, I_u(\mathbf{p}_{-c}))$ is the downlink transmission

rate at the location u where $I_u(\mathbf{p}_{-c})$ is the received interference power for location u. The vector, $\mathbf{p}_{-c}$, contains the transmit power values used by the interfering cells. $P_{PA}(p)$ denote the total power consumed by the BS for radiated power p, P_{stat} is the static power consumed during transmission other than in the PA and P_{idle} is the power consumed by the BS during idle state. The details of (3.1) along with the models used for the propagation pathloss, slow fading, aggregate interference and user rate can be found in Publication II.

In order to incorporate delay, we model the BS as a single server queue that serves the users in a one-by-one fashion and utilize the concept of flow throughput which is defined as the ratio of mean flow size to the mean flow transfer delay [48–51]. The flow level throughput at location u, ρ_u, can be written as [48, 49]

$$\rho_u = \left(1 - \sum_{u \in \mathcal{L}_c} \frac{\lambda_u S_u}{r_u}\right) r_u. \tag{3.2}$$

The queue is stable if the following condition holds true: $\sum_{u \in \mathcal{L}_c} \frac{\lambda_u S_u}{r_u} < 1$ [48]. Equation (3.2) is applicable only for balanced fair and proportional fair scheduling as these two qualify for homotheticity property [48]. We consider the achievable flow throughput when all the BSs transmit with the maximum average power $p_{\max}$ as the reference throughput and the related delay as the reference transmission delay. As the flow level delay is inversely proportional to the flow throughput for a given flow size, we can vary delay by regulating the flow throughput. Let us take a fraction Δ of the reference throughput as the target to model a particular delay. Let us denote the target throughput at location u as $\rho_{min,u}$ where

$$\rho_{min,u} = \Delta \left(1 - \sum_{u \in \mathcal{L}_c} \frac{\lambda_u S_u}{r_u}\right) r_u(p_{\max}, I_u(p_{\max})). \tag{3.3}$$

The parameter Δ takes its values in the interval $0 \leq \Delta \leq 1$ and $\frac{1}{\Delta}$ is the delay increase factor. If we increase the delay by a factor of two, $\frac{1}{\Delta} = 2$, then $\Delta = 0.5$ would be introduced in equation (3.3). By varying the factor Δ appropriately, a desired level of additional delay can be introduced. The details for the computations of the minimum transmit power constraint $p_{\min,u}(\mathbf{p}_{-c})$ to ensure $\rho_{min,u}$ (i.e., the delay does not exceed the target) with a 10% outage probability can be found in Publication II.

The energy minimization problem under delay constraint can be written as follows

$$\underset{\mathbf{p}_c,\forall c}{\text{Minimize}}: \sum_{c\in\mathcal{C}} \bar{e}_{tot,c}(\mathbf{p}_c, \mathbf{p}_{-c}). \tag{3.4i}$$

$$\text{Subject to: } p_{\min,u}(\mathbf{p}_{-c}) \leq p_u \leq p_{\max}, \forall u \in \mathcal{L}_c, \forall c \in \mathcal{C}. \tag{3.4ii}$$

Note that the objective function in equation (3.4i) involves summation over all the locations and over all the cells. However, the transmit power level, $p_u, u \in \mathcal{L}_c$, is independent of the power levels used in the other locations of the c-th cell. Therefore, it is sufficient to solve the optimization problem (3.4) separately for each location, i.e.,

$$\underset{p_u}{\text{Minimize}}: \ \lambda_u S_u \frac{P_{PA}(p_u) + P_{stat} - P_{idle}}{r_u(p_u, I_u(\mathbf{p}_{-k}))} + P_{idle} \tag{3.5i}$$

$$\text{Subject to: } \ p_{\min,u} \leq p_u \leq p_{\max}. \tag{3.5ii}$$

3.1.2 Energy Saving Game

The optimization problem (3.5i) is not jointly convex in the power allocation vectors of the cells for neither TPA nor ET-PA. However, for a logarithmic rate function, the optimization function has a nice property of decreasing differences independent of the PA efficiency characteristics [59]. Because of that, we resort to a game theoretic approach.

We define the *energy saving game*, $\mathcal{G}(\mathcal{C}, \mathcal{S}, \mathcal{E})$ where the players are the cells, $S = S_1 \times S_2 \times \cdots S_C$ is the strategy space, i.e., space of power allocations, and $\mathcal{E} = \{\bar{e}_{tot,c}(\mathbf{p}_c, \mathbf{p}_{-c}), c \in \mathcal{C}\}$ is the cost space of the players, i.e., energy consumption of the cells. The allocated transmit power levels at different cells are coupled due to the inter-cell interference. The set $\mathcal{S}_c$ of feasible powers for the c-th cell is a function of the transmit power levels $\mathbf{p}_{-c}$ at the interfering BSs.

$$\mathcal{S}_c(\mathbf{p}_{-c}) = \{p_u : p_{\min,u}(\mathbf{p}_{-c}) \leq p_u \leq p_{\max} \quad \forall u \in \mathcal{L}_c\}. \tag{3.6}$$

In game theory, the *best response* is the strategy (or strategies) that produces the most favorable outcome for a player given other players' strategies. The use of the best response strategy gives rise to a dynamic system of the form

$$\mathbf{p}_c = \arg\min_{\mathbf{p}\in\mathcal{S}_c(\mathbf{p}_{-c})} \bar{e}_{tot,c}(\mathbf{p}, \mathbf{p}_{-c}), \forall c. \tag{3.7}$$

In order to prove the convergence of such a BRI-based dynamic power control scheme, we show that the optimization problem can be modeled as a submodular

game. The convergence properties of such a sub-modular game have been discussed in Section 2.5 and in Publication II. Note that in order to compute the interference level $I_u(\mathbf{p}_{-c})$, it is assumed that the d-th interfering BS $d \neq c$ transmits at the mean power level $\mathrm{E}_u(\mathbf{p}_d) \forall d : d \neq c$. This assumption is valid for proportional scheduling.

For ET-PA and a logarithmic rate model the optimal power vector at each iteration can also be calculated by using the following closed-form expression in terms of the main branch of the Lambert function $\mathcal{W}_0$ [79].

$$p_u = \left(\exp\left(1+\mathcal{W}_0\left(e^{-1}\left(\frac{\epsilon P_{\max,\mathrm{PA}} \cdot g_u}{\delta_u \mathrm{E}\{I_u(\mathbf{p}_{-c}) + N_0 W\}} - 1\right)\right)\right) - 1\right) \frac{\delta_u \mathrm{E}\{I_u(\mathbf{p}_{-c}) + N_0 W\}}{g_u}.$$

where $\delta_u = \exp\left(\frac{-\sqrt{\sigma^2+\sigma^2_{I,u}}Q^{-1}(1-O)}{\xi} - \frac{\sigma^2_{I,u}}{2\xi^2}\right)$. The details can be found in Appendix D in Publication II.

3.2 Combined Uplink and Downlink

The total energy consumption at the radio access part of a wireless system can be written as

$$E_{total} = \underbrace{E^{BS}_{tx} + E^{MT}_{rx}}_{E_{dl}} + \underbrace{E^{MT}_{tx} + E^{BS}_{rx}}_{E_{ul}} + E_{Oth} \tag{3.8}$$

where the downlink energy consumption E_{dl} consists of energy required for the BS to transmit data E^{BS}_{tx} and the MT to receive that data E^{MT}_{rx}. Similarly, for the uplink, E^{MT}_{tx} is consumed at the MT for transmission and E^{BS}_{rx} is consumed at the BS for reception. E_{Oth} refers to cooling, dc-dc conversion loss in the BS, etc. Note that the potential for saving energy mostly lies in the downlink as the power consumed for downlink transmission is much greater than the power consumed in the uplink during reception [80].

However, energy saving in the BS by introducing delay in the downlink increases energy consumption in the MT due to prolonged reception time. Though the energy consumption in the BS dominates the total consumption in the radio access, energy consumption in the MT is also very important from the user's perspective as it determines the frequency of battery charging. In order to compensate for this excess energy consumption, additional delay can be introduced in the uplink likewise. We assume that the total energy consumption (downlink and uplink) in the MT should not increase beyond $\upsilon\%$ with respect to the reference scenario, i.e., downlink transmit power equal to $p_{\max}$ and fractional power control in the uplink.

3.2.1 Problem Formulation

Taking into consideration all the constraints, the energy minimization problem can be formulated as

$$\underset{\Delta^y, \mathbf{p}_c^y \forall c}{\text{Minimize}} : \sum_{c\in\mathcal{C}} \bar{e}_{dl,c}(\mathbf{p}_c^{BS}, \mathbf{p}_{-c}^{BS}) + \sum_{c\in\mathcal{C}} \bar{e}_{ul,c}(\mathbf{p}_c^{MT}, \mathbf{p}_{-c}^{MT}) \tag{3.9i}$$

$$\text{Subject to: } \Delta^y > \Delta^y_{\max} \forall y, \tag{3.9ii}$$

$$\frac{\triangle E_c^{MT}}{E_c^{MT}} \leq \upsilon\% \ \forall c \tag{3.9iii}$$

where $y \in \{BS, MT\}$, $\Delta^y_{\max}$ denotes the maximum reduction in throughput corresponding to the maximum allowable delays, E_c^{MT} refers to the energy consumption of the MTs in the reference scenario and $\triangle E_c^{MT}$ is the difference between energy consumption using energy-delay trade-off and the reference scenario where the total energy consumption in the uplink and downlink can be written following (3.1). For example, the energy consumption for the uplink can be written as

$$\bar{e}_{ul,c}(\mathbf{p}_c, \mathbf{p}_{-c}) = \frac{1}{|\mathcal{L}_c|}\Big(\sum_{u\in\mathcal{L}_c} \lambda_u S_u \frac{P_{PA}^{MT}(p_u) + P_{stat}^{MT} - P_{idle}^{MT} + P_{rx}^{BS}}{r_u(p_u, I_u(\mathbf{p}_{-c}))} + P_{idle}^{MT}\Big). \tag{3.10}$$

Note that the uplink interference condition is different from the downlink. The details of the interference model can be found in Publication II and III.

3.2.2 Optimization Algorithm

In order to solve the optimization problem in (3.9), we discretize the delays in uplink and downlink: $\Delta^{MT} \in [\Delta^{MT}_{max}, 1]$ and $\Delta^{BS} \in [\Delta^{BS}_{max}, 1]$ respectively, and construct the Cartesian product of the delay vectors. For a given delay pair, the minimum required power levels for uplink and downlink, $p^{BS}_{\min,u}, p^{MT}_{\min,u}, \forall u$, can be evaluated following Publication II and the optimization problem can be read as

$$\underset{\mathbf{p}_c^y, \forall c}{\text{Minimize}} : \sum_{c\in\mathcal{C}} \bar{e}_{dl,c}(\mathbf{p}_c^{BS}, \mathbf{p}_{-c}^{BS}) + \sum_{c\in\mathcal{C}} \bar{e}_{ul,c}(\mathbf{p}_c^{MT}, \mathbf{p}_{-c}^{MT}) \tag{3.11i}$$

$$\text{Subject to: } \mathbf{p}^y_{min,c} \prec \mathbf{p}^y_c \prec \mathbf{p}^y_{max,c} \ \forall c, y \in \{BS, MT\} \tag{3.11ii}$$

$$\frac{\triangle E_c^{MT}}{E_c^{MT}} \leq \upsilon\% \ \forall c. \tag{3.11iii}$$

In order to solve the optimization problem (3.11), we decompose it into two independent problems, one for uplink and one for downlink, without taking into account the

constraint (3.11iii). For instance, the optimization problem for the uplink is

$$\underset{\mathbf{p}_c^{MT},\forall c}{\text{Minimize}}: \sum_{c\in\mathcal{C}} \bar{e}_{ul,c}(\mathbf{p}_c^{MT}, \mathbf{p}_{-c}^{MT}) \tag{3.12i}$$

$$\text{Subject to: } \mathbf{p}_{min,c}^{MT} \prec \mathbf{p}_c^{MT} \prec \mathbf{1}.p_{\max,c}^{MT}\ \forall c. \tag{3.12ii}$$

The power allocation vector, $\mathbf{p}_c^{MT}$, is identified using the BRIA. The optimality of the BRIA is proved in Publication II. The optimization problem for the downlink is also solved in a similar way. Note that the pair of power allocation vectors $\mathbf{p}_c^{BS}, \mathbf{p}_c^{MT}$ must be discarded if it does not satisfy the optimization constraint (3.11iii). The optimization process is repeated over all delay pairs and the pair minimizing the objective function in equation (3.11i) is finally selected.

3.3 Impact of PA Efficiency and Delay on Energy Efficient HetNets

There are some proposals exploiting the TLV in order to make the network more energy efficient. For instance, one can shut down a fraction of the PAs or even BSs during off-peak hours [81–83]. The energy consumption of networks with macro BS and low power BS has been compared with the combined operation of both in [84–87] showing the latter as more energy efficient. However, most of these analyses consider the energy and throughput performance during the busy hour, i.e., under full load condition only. While the capacity and coverage requirements during the busy hour dominate the dimensioning of a network, the total energy consumption over the day needs to be considered while evaluating the energy efficiency of a network.

In this section, we consider a HetNet consisting of macro BSs and micro BSs. In order to save energy throughout the day, we vary the density of micro BS based on the offered load and divide the load between the two layers in an energy efficient manner. We also investigate the impacts of the non-ideal efficiency characteristics of the PA and additional delay on the energy efficient densification with small cells. Note that we consider outband micro deployment as it is suggested that in the evolution of cellular networks, outband micro deployment would be necessary to cater for expected high throughput [88–90].

3.3.1 System Model

We consider an outdoor wide area cellular network scenario. Both macro and micro network layers follow classical hexagonal deployment where a BS is placed at the center of each cell. The macro BSs are considered to operate throughout the day as an umbrella cell in order to ensure seamless coverage as suggested in [91, 92] while

the number of micro BSs operating within a macro cell changes according to the TLV.

The maximum number of micro BSs has been fixed based on the busy hour capacity demand. As we consider outband micro, the macro and micro layers can be considered independently once the load is divided between the two layers. We consider that at any location $u \in \mathcal{L}_c$ the offered load λ^{tot} is divided between macro BS λ^M and micro BS λ^m so that $\lambda^{tot} = \lambda^m + \lambda^M$. We consider uniform service demand inside the service area meaning that the same load λ^{tot} is offered at all the locations. The overall energy consumption in a BS can be obtained by summing the energy consumption over all the locations served by that BS

$$E_y = \lambda^y S_u \sum_{u \in \mathcal{L}_c^y} \frac{P_{PA}^y(p_u^y) + P_{stat}^y - P_{idle}^y}{r_u^y(p_u^y, I_u^y(\mathbf{p}_{-c}))} + P_{idle}^y \tag{3.13}$$

where $y \in \{M, m\}$; M represents the macro BS and m the micro BS. Note that during the idle state the BS operates but it does not serve any load. In the macro BS the PA dominates the energy consumption while in micro BS the static part dominates [14].

3.3.2 Problem Formulation

For a given offered load, we identify the number of micro BSs, $|m|$ within a macro cell and the load sharing between the macro BS and micro BS λ^M, λ^m so that the overall energy consumption $E_M + |m|E_m$ is minimized. While identifying the optimal parameter values λ^M, λ^m and $|m|$, we also maintain some quality of service for the end users. We describe the user performance in terms of flow throughput, which is defined as the ratio of mean flow size to mean flow duration. For proportional fairness the transmission time is shared equally among the users and following 3.2, the flow level throughput $\rho_u^y, y \in \{m, M\}$ at location u has the following expression

$$\rho_u^y = \left(1 - \sum_{u \in \mathcal{L}_c^y} \frac{\lambda^y S_u}{r_u^y}\right) r_u^y. \tag{3.14}$$

where the index $y \in \{m, M\}$ is used to distinguish between users served by macro BS and micro BS. The term $\sum_{u \in \mathcal{L}_c} \lambda^y S_u / r_u^y$ describes the total activity in the server queue. The queue is stable i.e. the flow transfer delay is finite if $\sum_{u \in \mathcal{L}_c} \lambda^y S_u / r_u^y < 1$.

Due to the slow fading, the flow throughput at a particular location u is probabilistic. At each location we maintain a flow throughput target ρ_t with probability $1 - O_t$ where O_t denotes the outage probability due to the slow fading. The constraint on the flow throughput at location u can be read as $\Pr(\rho_u^y \leq \rho_t) \leq O_t$.

In order to investigate the impact of energy-delay trade-off in the energy efficient densification of small cells, we assume that all users can compromise their flow throughput by a factor of $\Delta, 0 < \Delta < 1$. The flow throughput at location u with energy-delay trade-off becomes $\rho_u'^y \geq \Delta\rho_u^y$. The optimization problem can be written as

$$\underset{|m|,\lambda^m,\lambda^M,\mathbf{p}^y}{\text{Minimize :}} \quad E_M(\lambda^M) + |m| \cdot E_m(\lambda^m). \tag{3.15i}$$

$$\text{Subject to:} \quad \lambda^M + \lambda^m = \lambda^{tot} \tag{3.15ii}$$

$$\Pr(\rho_u^y \leq \Delta\rho_t) \leq O_t \, \forall u, y \in \{m, M\}\,. \tag{3.15iii}$$

$$\mathbf{p}^y_{\min} \prec \mathbf{p}^y \prec \mathbf{1}.p^y_{\max}, y \in \{m, M\}. \tag{3.15iv}$$

where $\mathbf{p}^y$ is the vector of transmit power levels p_u^y for serving the different locations in a cell and $\mathbf{p}^y_{\min}$ is the vector of transmit power levels $p^y_{u,\min}$. The transmit power level $p^y_{u,\min}$ is used to guarantee that the flow throughput does not drop under $\Delta\rho_u^y$. The details for the computation of the $p^y_{u,\min}$ can be found in Publication II. Note that the optimization problem in (3.15) can be solved with and without the delay constraints i.e., (3.15iv).

3.3.3 Optimization Algorithm

First we solve the optimization problem without considering the delay constraints. From equation (3.14) one can compute the maximum load that can be served by micro BS and macro BS with a finite flow transfer delay. Let us denote the maximum load by λ^y_∞. For any load higher than that, the flow transfer delay grows up to infinity and the flow throughput becomes zero. Given the offered load $\lambda^y \leq \lambda^y_\infty$, one can compute the flow throughput ρ_u^y at each location $u \in \mathcal{L}_c^y$ with outage probability O_t. The details for the computation can be found in Publication II. If the flow throughput at any location is smaller than the target ρ_t, the constraints (3.15iii) are not satisfied. The maximum offered load $\lambda^y_{\max}$ that can be served when constraint (3.15iii) is not violated can be computed by using an exhaustive search in the interval $[0, \lambda^y_\infty]$ with a discretization step $\Delta\lambda$. Finally, the constraints (3.15iii) are turned into linear constraints $\lambda^y \leq \lambda^y_{\max}, y \in \{m, M\}$ which are easier to handle.

After bounding the maximum offered load, we discretize the load λ^y in the interval $[0, \lambda^y_{\max}]$. Given the number of micro cells $|m|$ within a macro cell, the energy consumption (3.15i) is evaluated for all the feasible pairs $(\lambda^m, \lambda^M = \lambda^{tot} - \lambda^m) : \lambda^y \leq \lambda^y_{\max}$. The pair minimizing the cost function (3.15i) should be stored. Since the possible number of micro cells can take few integer values, the pair (λ^m, λ^M) minimizing (3.15i) can be found for each possible value of $|m|$. The 3-tuple $(\lambda^m, \lambda^M, |m|)$

Table 3.1. simulation Parameters

Reference parameters	
Parameter	*Value*
Number of cells	120
Grid size inside each cell	64 points
Cell radius	1 km
PA maximum output power, $P_{\max,\text{PA}}$	53 dBm
Maximum BS transmit power, $p_{\max}$	46 dBm
Maximum PA efficiency at $P_{\max,\text{PA}}$	80%
Path loss exponent	3.6
Shadow fading standard deviation	5.5 dB
Bandwidth	20 MHz
Noise level	-106 dBm
Target outage	10%

minimizing (3.15i) is finally selected.

With energy-delay trade-off, the optimal transmit power levels $\mathbf{p}^y$ for a given feasible 3-tuple $(\lambda^m, \lambda^M, |m|)$ is found by using the best response optimization algorithm and should be stored. The delay increase factor $1/\Delta$ is input to the best response optimization. The algorithm details along with a proof of its optimality can be found in Publication II. The 4-tuple $(\mathbf{p}^y, \lambda^m, \lambda^M, |m|)$ minimizing (3.15i) is finally selected.

3.4 Numerical Evaluation

In order to illustrate that there is a potential to save energy at BSs by lowering the transmission power levels at the cost of additional flow level delay, i.e. by solving optimization problem (3.4), we use the parameters given in Table 3.1, mostly taken from [14]. The reference energy consumption and flow level delay are obtained when the transmit power level is set to p_{max} for all the locations, i.e., no additional delay is allowed. The illustrations are presented for two different types of PAs; TPA and ET-PA.

Fig. 3.1 depicts the results for ET-PA as an example. We observe that for low to medium load, a significant amount of energy can be saved by introducing a small additional transmission delay. When the network load is low, the activity factors of the interfering BSs are also low. This allows having a high transmission rate using the maximum transmit power level p_{max}. The high rate means that the operation takes place at the leftmost side of the energy-delay trade-off curve, see Fig. 1.2. However, when the load is high, the activity factors of the cells are also high. Due to the high interference, the achievable rate using p_{max} decreases in comparison with the case of low to medium network load. The potential for energy saving decreases as the PA operation takes place on the flat part of the trade-off curve. If the transmission

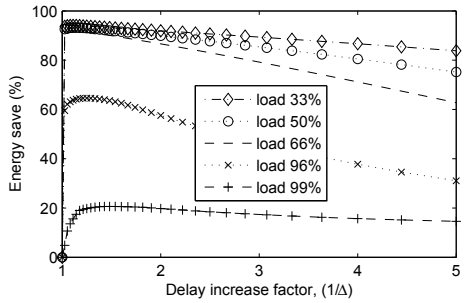


Figure 3.1. Percentage of energy saving potential for different network load with ET-PA.

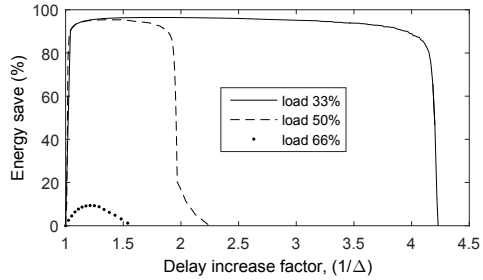


Figure 3.2. Percentage of energy saving potential for different network load with TPA.

delay keeps increasing, the operating point at the energy-delay trade-off curve moves rightwards. At the same time, the increasing transmission delay results in a higher activity factor of interfering BSs and consequently higher interference. At extreme values of transmission delay, the energy saving can become negative.

On the other hand, the energy saving with TPA is significant only for low load, as shown in Fig. 3.2. The percentage save drastically reduces if the network load exceeds 60%.

In order to show that our energy-delay trade-off based optimization scheme also take advantage of spatial load variations, we consider a scenario with non-uniform load where highly loaded urban macro cells are surrounded by lightly-loaded cells covering suburban area. For this purpose, we simulate a cellular network with 19 regular hexagonal cells where wrap around technique is employed to emulate a large network keeping all other parameters the same. We compute the energy saving potential for uniform loads at 96% and 66% along with non-uniform load where the first tier is 96% and the second tier is 66% loaded. The corresponding maximum savings are 68%, 92% and 88% respectively, see Fig. 5, Publication II.

Though the network is normally dimensioned considering the worst case scenario i.e. the peak traffic load, the offered load in a real-life network keeps changing throughout the day. We consider the daily traffic model for both residential and commercial areas as proposed in [18]. We modified the proposed model to introduce a

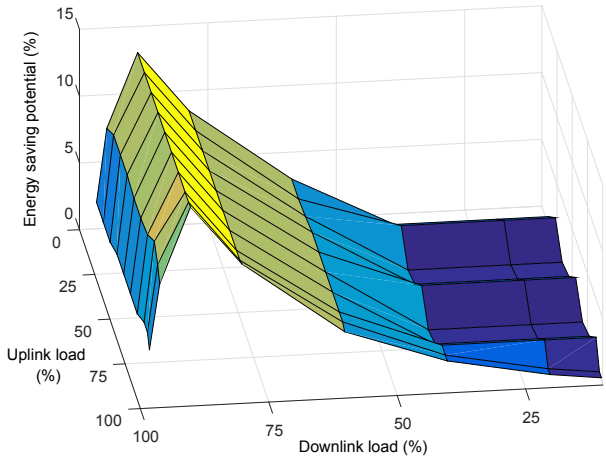


Figure 3.3. Maximum energy saving percentage at the BS when maximum 2% loss at the MT is allowed.

10% minimum load due to control signaling. We found that daily energy saving potential for ET-PA is more than 60% whereas for TPA it is around 20%, see Fig. 6 in Publication II. Note that we ignore the P_{idle} when generating the numerical results for the Section 3.1 in order to focus on the energy saving in the PA only.

Next, in order to investigate the impact of additional delay on the energy consumption of MTs as discussed in Section 3.2, we consider a network with 19 sites. There are three cells per site and each cell has its own BS. Also, the wrap around technique is used to eliminate boundary effects. In this case, we also consider the P_{idle}^{BS} which was ignored while finding the numerical results for Section 3.1. However, for simplicity, we ignore the power needed for packet processing at the BS, i.e., we consider that $P_{idle}^{BS} = P_{c}^{BS}$. Also, P_{idle}^{MT} is ignored in this analysis. In Fig. 3.3 and Fig. 3.4, we show the relation between energy saving in the downlink and its impact on the MT. We restrict the excess energy consumption in the MT to $\upsilon\% = 2\%$. One can see that there is room to save energy both for the BS and the MT for a range of combinations of uplink and downlink loads. However, when the load in the downlink is very high and the uplink load is low, the saving in the uplink is not enough to compensate for the excess energy consumed at the MT due to delay in the downlink.

In these analyses, P_{idle}^{BS}, for the BSs employing TPA and ET-PA has been calculated to be 53.6 W and 45.4 W using the equation suggested in Earth project [14].

For MT, we consider both the static power during transmission and the processing power during reception to be equal, i.e., 1200 mW. However, these power consumptions are different for different technologies, e.g., LTE, 3G [93, 94]. Also, for any technology, the processing power during reception and transmission is neither the same nor remains constant at different network loads. We do not consider the load

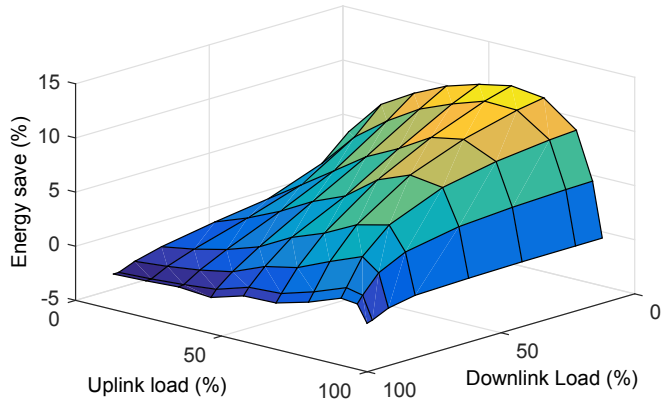
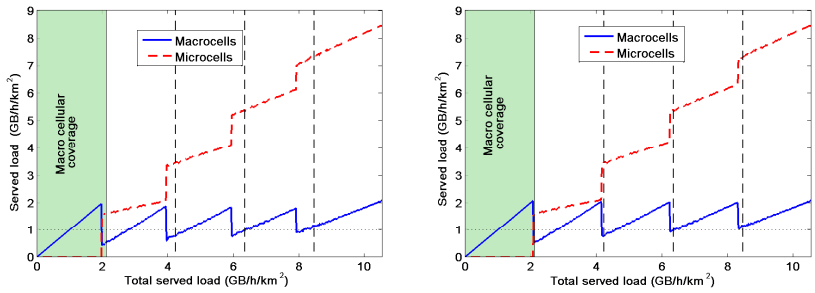


Figure 3.4. Energy saving at the MT during the maximum energy saving at the BS.

dependency of the processing power and also ignore the difference between processing power consumptions during transmission and reception. However, we repeat the calculations for different processing powers, e.g., $\{600, 900, 1200, 1500\}$ mW and found the corresponding best case save in the MT $\{21.59, 16.03, 12.75, 10.60\}\%$ respectively. Similarly, for the worst case, i.e., very high downlink load and very light uplink load, the excess energy consumption in the MT for both 600 mW and 900 mW cases is less than 1%. For the 1500 mW case, we relax the 2% constraint in order to keep the same downlink saving and find that the maximum excess energy consumption of the MT becomes 2.5%.

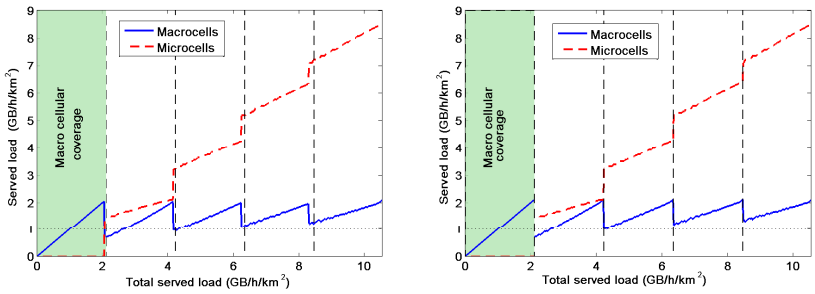
In order to generate numerical results for the impact of PA efficiency and energy-delay trade-off on the energy efficient densification of a wireless network as discussed in Section 3.3, we dimension a network with a maximum capacity demand of 10.2 GB/h/km^2 and a coverage requirement of 200 kbps which is reduced to 180 kbps in the case that additional flow level delay is allowed. The macro cell radius is modified to 600 m and the micro cell radius varies among $\{600, 424, 346, 300\}$ m based on the number of active micro BSs. Note that we use the parameters given in Table 3.1 and consider maximum average power for micro BS to be 38 dBm [14].

Fig. 3.5 illustrates the impact of the PA efficiency on the energy efficient density of micro cells. With increasing load, the activity of the macro BS also increases which pushes up the power consumption at the PA. It becomes energy efficient to offload traffic to micro layer before the load increases to the maximum that can be handled by the macro BS alone. One can see from Fig. 3.5a that for TPA a micro BS is turned on at a load lower than the maximum load, 2.04 GB/h/km^2 shown by the vertical dashed lines. So, switching on an additional micro BS inside the macro cell is an energy efficient strategy each time the traffic in a macro BS approaches this maximum load.



(a) TPA **(b)** ET-PA

Figure 3.5. Splitting the load between macro cells and micro cells for minimizing the energy consumption



(a) TPA **(b)** ET-PA

Figure 3.6. Splitting the load between macro cells and micro cells for minimizing the energy consumption with delay, $\Delta = 0.9$.

Due to the good efficiency characteristics of ET-PA, the macro BSs with ET-PA are able to serve higher load in an energy efficient manner compared to a macro BS with TPA, see Fig. 3.5b. As a result, we can conclude that the efficiency of the PA impacts the network densification.

By comparing Fig. 3.5a with Fig. 3.6a one can see that energy-delay trade-off impacts the load sharing between the macro and micro layers and also the density of micro sites. With delay, a macro BS with any number of micro BSs can handle higher load than the case of no additional delay. Even small additional delay ($\Delta = 0.9$ in Fig. 3.6) reduces the energy consumption in the PA and makes the introduction of an additional micro BS unnecessary until the service demand can be met. We can conclude that if the PA is not efficient, the density of micro BS minimizing the energy consumption can be selected simply based on the offered load at the cost of small additional delay. For efficient PAs, the impact of energy-delay trade-off on load sharing and network densification can be insignificant, compare for instance Fig. 3.5b with Fig. 3.6b. However, introducing delay will still reduce the energy consumption.

In Fig. 3.7, we depict the energy consumption for different loads with different options, e.g., with and without delay. As expected energy is saved if TPA is replaced by

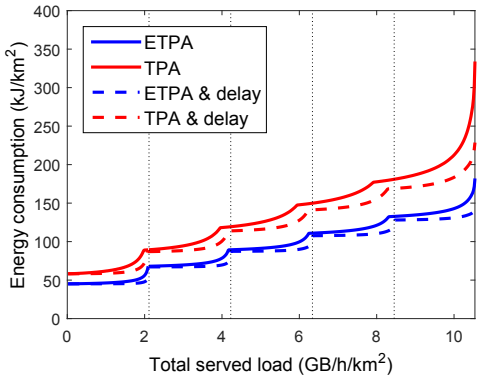


Figure 3.7. Energy consumption for different schemes at different capacity demand.

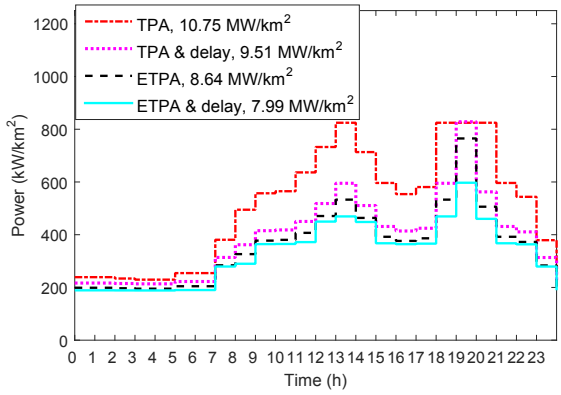


Figure 3.8. Hourly and total energy consumption throughout a day for different schemes for a residential area.

ET-PA. Moreover, introducing delay saves energy for both PAs. Again, energy saving potential by conceding delay is higher when the BSs are running with moderate to high load. It is worth to mention that, the saving percentage of ET-PA with delay is much higher than the case of TPA. However, actual energy saving potential is higher in case of TPA as it consumes higher power compared to ET-PA. In that sense, if the share of idle power consumption for ET-PA can be reduced, energy-delay trade-off will result very efficient network while employing ET-PA.

In Fig. 3.8, we depict the temporal variation of power consumption by taking a daily load variation model for a residential area [18]. It is observed that the energy saving potential through the deployment of efficient PA is high when the network is running close to full load. For low load, the PA activity is low and the impact of its efficiency on the energy consumption is not that significant. Overall, it is possible to save around 30% of energy by replacing the TPA with ET-PA. Furthermore, the overall energy saving by using the energy-delay trade-off scheme is around 8% for ET-PA and 11% for TPA.

Finally, we compute the energy consumption when all the micro BSs operate throughout the day. In that case the daily energy requirement for ET-PA is 3.14 kWh/km^2 and for TPA 4.21 kWh/km^2. This means that if TPA is used, we can save around 30% of the total energy by adopting load dependent switching and energy efficient sharing between the macro and micro layers. The corresponding saving percentage with ET-PA is 34%.

3.5 Discussion

In this chapter first we develop a framework to minimize the energy consumption of a cellular system under flow level delay constraint. The identification of power levels minimizing the total energy consumption in a multi-cell interference scenario is formulated as an energy saving game. The BRI allowed a distributed power allocation implementation which is proven to converge and used to study different aspects of the energy-delay trade-off in a multi-cell environment.

The energy saving potential is found to depend on the loading condition of the network and the PA efficiency characteristics. For low to medium load, a significant percentage of the total energy can be saved even by introducing small additional delay for both TPA and ET-PA. The energy saving drastically decreases with TPA for network load higher than 60%. On the other hand, due to the improved energy-delay trade-off relation of ET-PA, even for a load as high as 90%, the saving in energy is still more than 60%.

Though a considerable amount of energy at a macro BS can be saved by introducing transmission delay in the downlink, the prolonged transmission time at the BS increases the reception time at the MT and subsequently its energy consumption. Fortunately, there is an opportunity to compensate the excess energy consumption at the MT by introducing delay in the uplink too. The uplink and downlink delays that minimize the overall energy consumption depend on the network load and the PA efficiency characteristics. For moderate to high network loads, small delays are sufficient to reduce the total energy consumption at the BS by as much as 15%. At the same time, MT consumption increases slightly for very high download and very low uplink load. For other load combinations, MT also saves energy. This saving potential becomes less significant in the case of a smaller BS (e.g., micro, femto BS) as their energy consumption breakdown is less dominated by the energy share of the PA.

Note that we ignore P_{idle} in Section 3.1 in order to show the energy saving potential in the PA only. However, P_{idle} has been considered in Section 3.2 in order to show the overall energy saving potential in the wireless access network by energy-delay trade-

off. Also note that in these analyses, we do not devise any mechanism to reduce idle power consumption of the BS which is around half of the total power consumed at full load [14]. Putting BS into sleep mode, i.e., employing DTX is shown to reduce this energy waste [3, 72]. Incorporating DTX or partial switching off of the network with this energy-delay trade-off can be a good candidate for future work.

For improving the energy efficiency of HetNet, the density of a BS should vary throughout the day according to the variable service demands. Also, it is important to divide the load between the different layers (e.g., macro and micro layers) in an energy efficient manner. The energy efficient BS density and load sharing depends significantly on the PA efficiency. We observe that as load increases, new micro sites with TPA need to be introduced earlier than the load limit of the system in order to be energy efficient. Whereas in case of ET-PA, traditional load-based dimensioning of the network is also energy efficient. The energy-delay trade-off can be an additional means of saving energy, especially, during the off-peak hours. The significant amount of energy saving potential in our analysis implies that the configuration of a heterogeneous network needs to be dynamic in order to cope with the temporal variation of capacity demand.

We consider the radius of the micro BSs to be variable based on the number of active micro BSs and the load distribution among the layers to be the same in all the locations in the cell. In a practical network the load is usually not homogeneous and the micro BSs do not need to cover the total service area. As a result, more energy can be saved by optimizing the location and coverage areas of the micro BSs based on the spatial load distribution. Note that we do not consider the backhaul energy consumption for the small cells which might impact the energy saving percentage if microwave links are used for the backhauling solution [95–98].

Finally, the simplification of the models in these analyses allows analytical treatment in terms of proving the convergence. However, to the best of our knowledge, our work is the first attempt to utilize energy-delay trade-off under the multi-cell context. We not only consider inter-cell interference and slow fading but also provide a scheme that dynamically adjusts BS power levels to the DLP and save energy throughout the day.

4. Energy Efficient Antenna Adaptation in Massive MIMO

In this chapter, we discuss the dimensioning of MM systems that can cope with TLV to maintain high EE throughout the day while taking into consideration the non-ideal efficiency characteristics of the PAs. We consider two different scenarios, both of which are multi-cell networks. In the first scenario, as in Publication V, the BSs transmit at fixed power irrespective of the number of antennas that are turned on and the number of users being served. In the second scenario, as in Publications VI and VII, we relax this fixed transmit power constraint, i.e., the average transmit power of the BS varies with the number of active antennas. We also provide guidelines regarding how to dimension the PA in the context of the former scenario.

4.1 Energy Efficiency of Massive MIMO

Recently, both MM and EE optimization of wireless systems have gained significant attention from both academia and the industry [39, 40, 45, 47, 99–104]. In [45], it has been shown that the capacity achieving power allocation scheme necessitates allocating as many antennas as possible with the maximum power under the simultaneous per-antenna radiated power constraints and total consumed power constraints. In [40], it has been shown that the transmit power can be reduced proportionally with the number of antennas if the BS has perfect CSI without any fundamental loss in performance. In [47, 99, 100], the impact of the circuit power on the EE of MM has been analyzed. Specifically, in [47], it has been shown that without accounting for circuit power consumption, an infinite EE can be achieved as the number of antennas, $M \to \infty$, which is misleading. In [47] and [39], the authors show that the EE optimal strategy requires increasing the transmit power with the number of antennas if the circuit power consumption is taken into account. They also show that the EE is a quasi-concave function of the three main design parameters; namely, number of BS antennas, number of users and transmit power. In [101], an adaptive antenna selection scheme was proposed where both the number of active RF (radio frequency)

chains and the antenna indices are selected depending on the channel conditions. In [103], the authors show that during low traffic demand, e.g., late night, it is optimal to turn off a fraction of the antennas while minimizing total power consumption if the power consumption in the transceiver circuits is taken into account along with the RF amplifiers.

EE improvement by antenna selection has been studied quite extensively [101,105–108], some of the studies also suggest adapting the number of antennas to improve EE. However, most of these studies along with the existing literature involving MM systems (e.g., [36, 37, 40]) mainly focus on the potential to improve SE and EE when serving peak load. Also, very few of these studies consider multi-cell systems. In [39], the relation among the EE, the number of antennas and simultaneously served users are derived analytically under a single cell context and then a multi-cell scenario is analyzed by simulation to show that a similar relation persists in the presence of inter-cell interference. In [109], the authors suggest that a BS should use as many antennas as are enough to provide the required average sum rate in order to be energy efficient in the presence of inter-cell interference. However, none of these works provide an antenna adaptation scheme that can dynamically cope with the TLV in the presence of inter-cell interference. In our study we not only consider inter-cell interference, but also devise a mechanism to dynamically adapt the number of antennas in order to maximize EE and cope with the TLV. In Publications V, VI and VII, we show that there is a potential to improve EE and save a significant fraction of total energy consumption by adapting the number of antennas to TLV. Note that in Publication V, the BS transmit power is assumed to be fixed for any number of active antennas and users.

4.2 Problem Definition

In this study, we aim at designing a multi-cell MM system that adapts the number of active antennas to the network load and interference condition in order to maximize average EE. The EE is defined as the number of bits transferred per Joule of energy and hence can be computed as the ratio of average sum rate (in bit/second) and the average total power consumption (in Joule/second) [39]. If $P_c^{tot}(K_c, M_c)$ denotes the total power consumed by an arbitrary BS c when serving K_c number of users simultaneously using M_c antennas and $R_c(K_c, M_c, \{M_d\}_{d\neq c})$ denotes the resulting average data rate per user, the corresponding EE will be

$$\text{EE} \quad = \quad \frac{\text{Average sum rate}}{\text{Power consumption}} = \frac{K_c R_c(K_c, M_c, \{M_d\}_{d\neq c})}{P_c^{tot}(K_c, M_c)} \qquad (4.1)$$

where $\{M_d\}_{d\neq c}$ represents the number of antennas used by any other cell d, $d \neq c$.

Note that the network loads vary throughout the day. In order to capture the daily load variation and maximize EE throughout the day, we model the load at each BS as an $M/G/m/m$ state-dependent queue. We divide the day into an arbitrary number of time intervals H and find the steady-state distribution of the users in each time interval. We maximize the average EE by adapting the number of active antennas to the number of active users at each user state while taking the received interference into account. Let us consider that during the time interval, $h, h \in H$, the steady-state probability of the BS c serving n number of users, i.e., $Pr[K_c = n]$ is denoted by $\pi_c(h, n)$. The EE maximization problem for BS c can be written as

$$\underset{\mathbb{M}_c}{\text{maximize}} \quad \sum_{h=1}^{H}\sum_{n=1}^{m}\pi_c(h,n)\frac{nR_c(n, M_c^{(h)}, \{M_d^{(h)}\}_{d\neq c})}{P_c^{tot}(n, M_c^{(h)})} \tag{4.2i}$$

$$\text{subject to:} \quad \mathcal{M}_c^{(h)} \in \mathcal{S}_c \tag{4.2ii}$$

where $R_c(n, M_c^{(h)}, \{M_d^{(h)}\}_{d\neq c})$ is the average rate per user when there are n users in the cell during time interval h. $\mathbb{M}_c = [\mathcal{M}_c^{(1)}\mathcal{M}_c^{(2)}....\mathcal{M}_c^{(H)}]$ where $\mathcal{M}_c^{(h)}$ is the vector that gives the number of antennas that maximizes the EE at different user states in cell c during the time interval h.

It is important to note that EE optimization for any time interval h can be carried out separately following the same procedure. Also, the number of antennas maximizing EE for any user state n, $M_c(n), n \in \mathcal{U}_c$, is independent of the antennas used for the other user states in the same cell. Therefore, it is sufficient to solve the optimization problem (4.2) separately for each user state.

4.3 Fixed Power per BS

In Publication V, we assume the transmit power of a BS to be fixed. Due to this assumption, the inter-cell interference becomes fixed. Under this context, the problem in (4.2i) reduces to finding the optimum number of active antennas for a given number of users in a cell that maximizes the EE. Using (2.7) and (2.13), the optimum number of antennas M_c for a given number of users K_c in (4.2i) becomes a quasi-concave problem and the closed form expression for M_c can be derived as follows

$$M_c = \left\lceil \frac{\exp\left[\mathcal{W}_0\left[\frac{\gamma_c C_0 - (1-K_c\gamma_c)C_1)}{C_1 e}\right]+1\right] - (1-K_c\gamma_c)}{\gamma_c} \right\rceil \tag{4.3}$$

where $\gamma_c = \frac{(P_c/K_c)}{(G_{cc}N_0W + \sum_{d\neq c} G_{cd}P_d)}$ and $\mathcal{W}_0$ is the main branch of the Lambert function [47, Equation (37)]. For TPA, we do not have such closed form expression for

the M_c, however, as M_c and K_c are integer values, it is not computationally demanding to find the optimum number of antennas by exhaustive search for a given number of users.

4.3.1 Dimensioning of PA

The maximum output power of the PA, i.e., $P_{\text{max,PA}}$ is a very important design parameter, especially in the context of energy efficient operation of the network that experiences a highly varying load throughout the day. In order to achieve the highest EE considering TLV as well as dynamic efficiency characteristics of the PA, we need to solve the following problem

$$\underset{P_{\text{max,PA}}}{\arg\max:} \sum_{h=1}^{24}\sum_{n=1}^{m}\pi(h,n)\frac{nR_c(n)}{C_0+(C_1^{BB}+C_1^{*}P_{\text{max,PA}})M_c(n)} \tag{4.4i}$$

$$\text{Subject to: } M_c(n) \geq n+1, P_{\text{max,PA}} \geq \frac{P_c}{M_c(n)}10^{0.8} \tag{4.4ii}$$

where $R_c(n)$ is the rate achieved by any user with the optimum number of antennas, $M_c(n)$ at the state $K_c = n$ and $C_1^{*} = \frac{\epsilon}{(1+\epsilon)\eta}$. The first constraint comes from the prerequisite of zero forcing precoding, i.e., the number of antennas is larger than the number of users and the second constraint ensures that the minimum number of antennas required to transmit at least P_c are active while considering 8 dB PAPR. We solve this problem using an exhaustive search. Given the total transmit power P_c, we can compute the value of $P_{\text{max,PA}}$ for any number of active antennas, $M_c = 1,\ldots,M_{\max}$. We end up with $M_{\max}$ number of different values for $P_{\text{max,PA}}$ and select the one maximizing (4.4i).

4.3.2 Reference Network

Since EE is a quasi-concave function of the number of users and number of antennas [47], there is a global maximizer of $EE(M_{\max}, K_{\max})$ corresponding to the optimum number of antennas $M_{\max}$ and users $K_{\max}$ which is also a stationary point in the plane. In order to compare the performance of this load-adaptive antenna system, we dimension the reference network in a way that it always runs with this $M_{\max}$ antennas and $K_{\max}$ corresponds to the highest load according to DLP, i.e., the reference network attains the highest EE when serving maximum cell load.

4.3.3 Numerical Evaluation

We consider the downlink of a cellular network with 19 regular hexagonal cells where the wrap around technique is applied in order to get rid of the boundary effect. The

Table 4.1. Simulation parameters [39]

Reference parameters	
Parameter	*Value*
Number of cells	19
Grid size inside each cell	15000 points
Cell radius: $d_{\max}$	500 m
Minimum distance: $d_{\min}$	35 m
Maximum PA efficiency at $P_{\max,\text{PA}}$	80%
Path loss at distance d	$\frac{10^{-3.53}}{\|\|d\|\|^{3.76}}$
Local oscillator power: P_{SYN}	2 W
BS circuit power: P_{BS}	1 W
Other power: P_{Oth}	18 W
Power for data coding: P_{COD}	0.1 W/(Gbit/s)
Power for data decoding: P_{DEC}	0.8 W/(Gbit/s)
Computational efficiency at BSs:L_{BS}	12.8 Gflops/W
Bandwidth	20 MHz
Total noise power	-96 dBm
Channel coherence interval: T_c	1800 symbols

parameters used are provided in the Table 4.1. We illustrate the impact of different maximum output power of the PA, i.e., $P_{\max,\text{PA}}$ on EE in Fig. 4.1 and the corresponding number of antennas required to achieve the maximum EE for a given number of users in Fig. 4.2. As the total average transmit power P_c is fixed, the transmit power per antenna P_c/M_c varies with the number of antennas. For the 'variable' curve, $P_{\max,\text{PA}}$ is variable, i.e., just 8 dB higher than the average output power of the PA and hence gives an upper bound on EE. Note that in reality a PA is designed to deliver a fixed $P_{\max,\text{PA}}$ where it attains highest efficiency. Accordingly, the other plots consider fixed $P_{\max,\text{PA}}$ disregarding the actual transmit power transmitted by each antenna. Note that a minimum number of antennas need to be always active in order to guarantee the fixed transmit power. For example, an '8 dB' plot represents the case when at least 20 antennas are required to deliver total transmit power $P_c = 20$ W, see Fig. 4.2. Note that we set pilot reuse factor $\alpha = 1$ for this analysis.

It can be seen from these figures that very high $P_{\max,\text{PA}}$, e.g., 21 dB, is comparatively efficient for small K_c as the number of antennas can be very small. However, for large K_c, M_c is also large as it needs to be larger than K_c as a requirement for zero forcing precoding. This results in low PA efficiency as low power per antenna leads the PA operating point away from the compression region. On the other hand, when the $P_{\max,\text{PA}}$ is low, e.g., '1 dB', it becomes very efficient at high load as contribution of an extra antenna as array gain exceeds its small incremental power contribution. However, when serving only a few users, M_c cannot be reduced as much as required to maximize EE. Due to low maximum transmit power few antennas cannot deliver fixed total transmit power. In Fig. 4.3, we show the benefit of using PA with better efficiency characteristics. In the TPA case, the efficiency reduces drastically if the

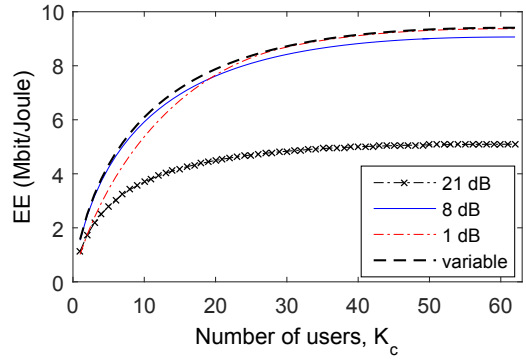


Figure 4.1. Impact of the dimensioning of $P_{\max,\mathrm{PA}}$ on EE at different cell loading (ET-PA).

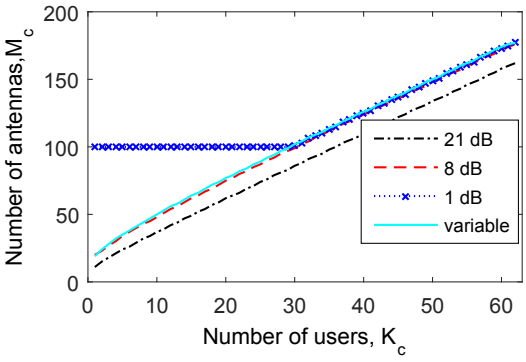


Figure 4.2. Required number of antennas, M_c for a given number of users under different $P_{\max,\mathrm{PA}}$ maximizing EE.

$P_{\max,\mathrm{PA}}$ drifts away from the optimum value compared to the ET-PA. In Fig. 4.4, we illustrate the performance of the load-adaptive antenna system compared to the reference system. It is evident from Fig. 4.4 that there is a significant increase in EE with the adaptive antenna for almost all of the user states. We also observe that over 24 hour operation, overall gain in EE is 12% with an optimally dimensioned PA.

4.4 Average Fixed Power per Antenna

In Publications VI and VII, we design a further energy efficient multi-cell MM system where the BSs are not constrained by fixed total transmit power. Under the assumption of independent fading and considering the fact that power gets averaged over many sub-carriers, each antenna uses the same average power. We denote this average transmit power per antenna by p, hence, the total transmit power of cell c, $P_c = pM_c$. The transmit power of any other cell d: $d \neq c$ with M_d active antennas, $P_d = pM_d$. Note that the average transmit power of a BS is not fixed as it varies with the number of active antennas.

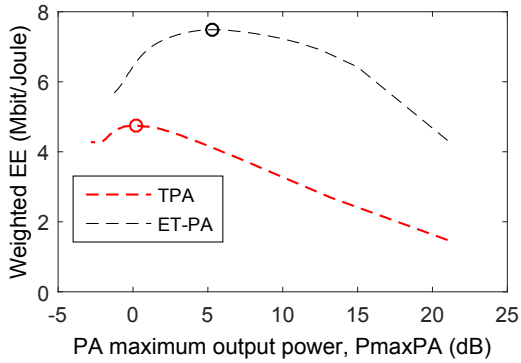


Figure 4.3. EE performance comparison between TPA and ET-PA.

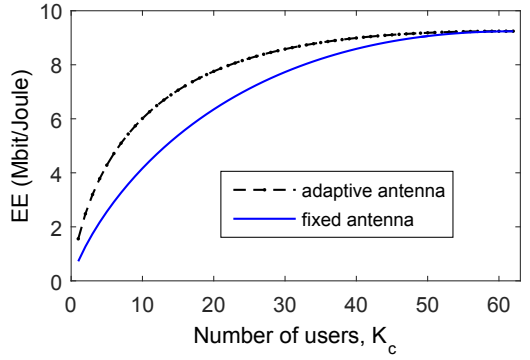


Figure 4.4. Comparison of EE between proposed load-adaptive system and fixed base line system for all user states.

4.4.1 EE Maximization Game

When the BS is serving a particular number of users n following (2.9), (2.13) and (4.2), the objective function for this user state can be broadly written as

$$E_c \approx \frac{\beta}{C_0 + C_1 M_c} n \log\left(1 + \frac{p\frac{M_c}{n}(M_c - n)}{N_0 W G_{cc} + \sum_{d \neq c} p M_d G_{cd}}\right) \tag{4.5}$$

where $\beta = (1 - \frac{\alpha K_{\max}}{T_c})\frac{B}{\ln 2}$. As the transmit power of a BS is proportional to the number of active antennas and depends on the transmit power of interfering BSs due to inter-cell interference, the number of active antennas for different cells are coupled. As a result, the objective function in (4.2) for cell c is dependent on the number of antennas used by other cells. The optimization problem (4.5) is not jointly concave in M_c and M_d for either TPA or ET-PA. However, this objective function has a nice property of increasing differences in M_c and M_d, $\forall d \in \mathcal{C}, d \neq c$, in the sense of Topkis [59]. Also, centralized optimization of a network consisting of BSs with MM necessitates exchange of huge information. From practical point of view, distributed algorithm is more suitable in this context. Because of that we use

a BRI-based distributed algorithm that allows each cell to determine the number of active antennas. We resort to a game theoretic approach to show the convergence of the proposed algorithm to a Nash equilibrium.

We define the *EE maximization game*, $\mathcal{G}(\mathcal{C}, \mathcal{S}, \mathcal{E})$ where the players are the BSs, $S = S_1 \times S_2 \times \cdots \times S_C$ is the strategy space (i.e., space of number of active antennas used at different user states), and $\mathcal{E} = E_c(\mathcal{M}_c, \mathbf{M}_{-c}), c \in \mathcal{C}$ is the utility of the players (i.e., EE of the different cells). Note that the set $\mathcal{S}_c$ is a function of $\mathbf{M}_{-c}$ i.e., the number of antennas used by all other cells that interfere with cell c and can be written as

$$\mathcal{S}_c(\mathbf{M}_{-c}) = \{\mathcal{M}_c(n) : n+1 \leq \mathcal{M}_c(n) \leq M_{\max}, \forall n \in \mathcal{U}_c\} \tag{4.6}$$

where the lower bound comes from the constraint for zero forcing. The *best response* is the strategy (or strategies) that produces the most favorable outcome for a player given other players' strategies. The use of the best response strategy gives rise to a dynamic system of the form

$$\mathcal{M}_c = \arg\max_{\mathcal{M}_c \in \mathcal{S}_c(\mathbf{M}_{-c})} E_c(\mathcal{M}_c, \mathbf{M}_{-c}), \forall c. \tag{4.7}$$

The BRI starts from any cell c by initializing the number of antennas of all the cells as the maximum $M_{\max}$. We iterate over all the cells and optimize the antenna vector for each cell. The iterations are carried out until the antenna vector for each cell converges, i.e., there is one iteration where none of the numbers of antennas changes. The convergence properties of the supermodular game has been discussed in Section 2.5 and in Publication VII.

4.4.2 Performance Evaluation

We consider the similar network scenario and parameters as provided in 4.3.3. The reference scenario described in Section 4.3.2 has been improved by considering the fact that the fraction of time when the BS serves no user is longer under an adaptive antenna scheme compared to the reference case. Note that in the reference case all the antennas are kept on as long required. This results in a high user rate because of high array gain and reduces the active time of the BSs as interferers. This has been taken into account while finding the total energy saving potential for a fair comparison. Also, we use a pilot reuse factor of 7 and a channel coherence interval, $T_c = 5000$ symbols in Publication VII.

Fig. 4.5 gives the interplay between K_c and M_c. When the EE is optimized over K_c, M_c adapts to the user profile resulting in an approximately linear relation between M_c

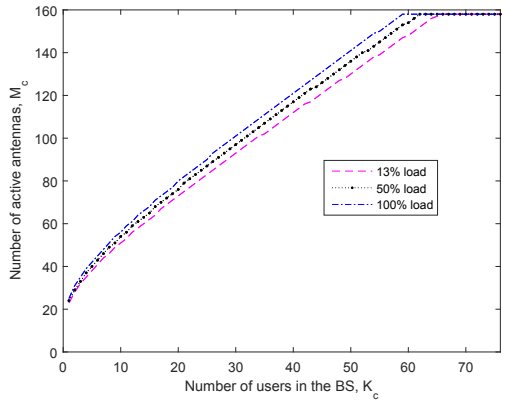


Figure 4.5. Number of antennas as a function of the number of users in the cell at 13%, 50% and 100% load.

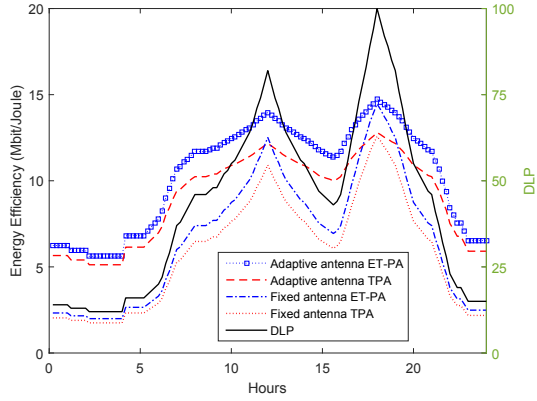


Figure 4.6. Average EE over 24 hours load variation for both fixed and adaptive antenna while using both TPA and ET-PA.

and K_c. However, the ratio between M_c and K_c is quite high when the BS serves only a few users and ends up around two at high load. When K_c is very small, the energy consumed by activation of an additional antenna is not significant compared to the fixed energy consumption but contributes significantly to increase the EE due to higher array gain, i.e., $(M_c - K_c)$. We also observe that even though the number of antennas used for a particular user state at different loads do not vary much (see Fig. 4.5), the average number of antennas used by a BS at different loads varies significantly due to the probability distribution of the users at different loads, see Fig. 2.2.

In Fig. 4.6, we show the hourly average EE throughout the day following the DLP for both the TPA and ET-PA. The plots for fixed antenna cases are generated from the reference case whereas the plots for adaptive antenna cases are obtained by solving the optimization problem in 4.2. In general, the EE increases with the increase in

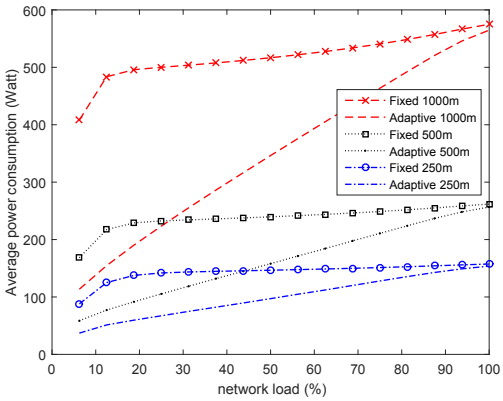


Figure 4.7. Power consumption vs. network load at different cell sizes for both fixed and adaptive antenna schemes.

load for both the reference case and our scheme for both TPA and ET-PA. Our scheme achieves a significantly higher EE compared to the reference case at low load. However, this EE gain keeps decreasing with the increase in load. At the peak load, the gain is insignificant as the probability of having a small number of users which allows EE improvement by reducing the number of antennas is very low, see Fig. 2.2.

In Fig. 4.7 we present the actual power consumption for different network loads in both the adaptive antenna scheme and the fixed antennas, i.e., the reference case. It is observed that power consumption is drastically reduced at low load as the number of active antennas are allowed to be low following the load under the adaptive scheme. We also observe that over 24 hours of operation, the energy saving potential by adopting an adaptive antenna scheme over the reference case is around 40%.

Fig. 4.8 shows the trade-off between the EE and the average user rate at different loads for TPA. At very low load the EE has been increased by around 250% at the cost of an around a 50% reduction of the average user data rate. However, whenever the load of the system increases, both the gain in EE and loss of user rate get reduced. We also observe that over the 24 hours of operation, the EE has been found to be improved around 24% at the cost of around a 12% reduction in user rate. Note that the gain in EE and loss in percentage user rate is quite similar for both TPA and ET-PA.

Another observation indicates that the EE increases with the decrease in cell size, and $M_{\max}$, $K_{\max}$ and optimum p decrease with the decrease in cell size.

It is evident from Fig. 4.6 and Fig. 4.9 that the efficiency of the PA has a significant impact on the energy efficient design of the network. In the TPA case, the efficiency falls drastically when the operating point moves away from $P_{\max,\text{PA}}$, whereas ET-PA achieves higher efficiency over a wider operating region. Therefore, each BS uses more antennas with TPA compared to ET-PA with less power per antenna as

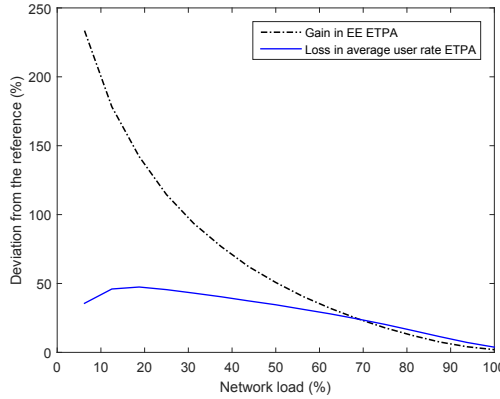


Figure 4.8. Trade-off between EE and user rate at different network loads.

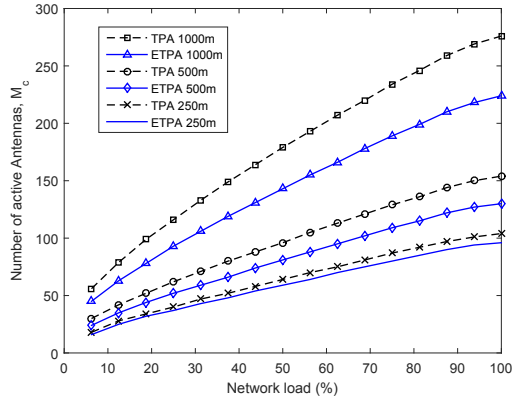


Figure 4.9. Comparison between TPA and ET-PA in terms of average number of active antennas vs. network load for different cell size.

a means of pushing a higher number of antennas to operate near the compression region to reap the benefit of higher PA efficiency. Also note that the TPA requires a comparatively higher number of users to be served simultaneously in order to achieve the maximum EE. Nevertheless, ET-PA yields higher EE for both the reference case and our scheme compared to TPA for any given cell radius. We have also observed that even though the change in cell radius and PA dimensioning impacts the EE, the improvement in EE by our scheme remains quite similar.

4.5 Discussion

In this work, we first design a multi-cell network with load-adaptive MM system and investigate the impact of DLP on the design of the PA and antenna number adaptation while maximizing EE considering the total BS transmit power to be fixed. Modeling a BS with MM as a $M/G/m/m$ queue, we generate the user distribution following

the DLP and show that TLV has a significant impact on the design of the PAs. We observe that there is a clear choice for the optimum value of $P_{\max,\text{PA}}$. From the comparison between the performance of TPA and ET-PA, it has been found that the ET-PA having better efficiency over wider operating range than TPA not only yields higher EE but also provides more flexibility to design the PA. Numerical analysis also suggests that this load-adaptive MM system along with optimally dimensioned PA increases network EE by about 12% even when compared to a base line network where the BSs run with a fixed number of antennas that yield highest EE at 100% cell loading.

Next we consider a system where the BS transmit power varies with the number of antennas and investigate the potential to improve EE by dynamically adapting MM systems to the TLV. We consider the average transmit power per antenna to be fixed. Note that we assume the transmit power to be limited by the PAs and not by any regulatory aspects.

We develop a game theory based distributed algorithm that yields significant gain in EE at the cost of reducing the average user data rate at low user load. However, the rate degradation while increasing the EE comes from the fact that we consider a very tight reference case. In our reference case, the system considers the complete shut-down of all the antennas when the BS is not serving any user. This reduces the interference significantly resulting in high data rates for some users which in turn allows the BS to reduce its activity time further. Over the 24 hours of operation, EE has been found to be improved by around 24% along with an energy saving potential of around 40% while the corresponding reduction in user rates is found to be around 12%.

Note that in these analyses, the algorithms are developed for a simple rate formula based on perfect CSI and does not consider pilot contamination. Also, for the queuing model, we do not consider any waiting place, i.e., we do not allow any users to wait if the system already has the maximum number of users, K_{max}. These models are selected for analytical simplicity but we believe that they nevertheless provide valuable insight to the EE behavior of load-adaptive MM systems.

5. Discussion and Future Work

In this thesis, we provide PA efficiency and DLP aware system level energy efficient schemes for a wireless access network. First we model the efficiency characteristics of the leading PA (i.e., ET-PA) to be used for the NGN wireless network. With the help of the flow level throughput model provided by Bonald *et al.* in [48, 49, 51], we formulate the energy minimization problem in the downlink of a multi-cell wireless network under delay constraint. Our best response based distributed algorithm identifies the transmit powers for different users' location that minimize total energy consumption in the presence of inter-cell interference and slow fading. The results from numerical analysis indicate that the energy saving potential depends on the network load and PA type. At low to medium load, the energy saving potential from the PA consumption part is substantial at the cost of very small additional delay. For ET-PA, energy saving is significant even at network load as high as 90%. However, the TPA does not save any energy beyond a network loading of 60%.

Note that these results have been generated by ignoring the processing power consumption in the BS. Also, the idle power consumption when the BS is not transmitting is also ignored in order to focus on the energy saving potential in the PA. The contemporary work by Tombaz S. in [3] and Hiltunen K. in [72] shows that energy can be saved in the BS by transmitting at the maximum transmit power as long it is needed and then going into micro sleep mode, i.e. employing discontinuous transmission (DTX). Also, energy consumption during DTX is considered to be less than the energy consumed by the BS during the idle period (i.e., few components can be turned off when idle). The total energy saving potential will be larger if DTX can be employed along with energy-delay trade-off as it has been shown to perform better than either of the two mechanisms working alone [73, 74] in a single cell context. However, studies in [73, 74] do not consider the non-ideal PA efficiency characteristics. Similar studies considering non-ideal PA efficiency characteristics under multi-cell context can be a future direction of this work.

Second, we investigate the impact of additional delay in the downlink on the energy

consumption of MT. Though the power consumed in the PA is much less than the processing power of the MT, our results indicate that the excess energy consumed in the MT due to the additional delay in the downlink can be compensated for by applying a small delay in the uplink for the most load combinations except those combinations where the downlink load is very high and uplink load is very low. Note that so called *tail energy* which is consumed due to the PA staying at the high energy state after each transaction dominates the energy consumption of the MT. Using proxy at the BS to accumulate and bundle data before transferring it to the MT has been suggested as a means to reduce the number of tails and reduce energy wastage [110]. Energy-delay trade-off involving proxy can be a future direction of this work.

Next we investigate the impact of delay and PA efficiency on the densification of network with small cells. Varying the number of small cells is found to be energy efficient way to cope with the TLV as also suggested in [3, 111, 112]. It is also found that PA with improved efficiency characteristics not only saves energy but also impacts the energy efficient load sharing between the tiers.

When increasing the number of micro BSs to meet higher capacity demand, the micro BS can be switched on later in case the PA is more energy efficient. When an extra micro BS is required to handle the excess capacity with increasing network load for a TPA case, the ET-PA still stays energy efficient with the same number of micro BSs to some extent with increase in load. This essentially means that a network with inefficient PAs may require more BSs to handle the certain load ranges for energy efficient densification.

Note that the model for flow level throughput we used is valid only for proportional fair allocation. However, to the best of our knowledge, existing literature does not study the utilization of energy-delay trade-off in the presence of inter-cell interference and fading. Our scheme does not only take those into account but also provides a distributed algorithm that ensures energy efficient dynamic operation of the network that is able to cope with the TSLV.

Although MM systems can improve the SE and EE simultaneously by more than 100 times compared to a SISO system [40], it is important that MM maintain high EE disregarding the TLV. In this study first we consider a multi-cell MM system with fixed transmit power per BS in order to investigate the impact of PA dimensioning on the EE. We formulate the problem that takes twenty-four hour load variation into consideration while dimensioning the PA for improved EE. Our results indicate that PA with improved efficiency characteristics not only yields higher EE but also provides close to the maximum EE for a wider range of designated maximum power of the PA whereas for TPA the EE falls sharply if the maximum power drifts away from the optimum value. With this fixed transmit power per BS constraint (i.e., fixed inter-cell

interference), each BS can find its number of antennas maximizing EE analytically and such antenna adaptation can improve EE over the day by 12%.

When we relax the fixed transmit power per BS constraint and we assume that the PA per antenna transmits the same average power, the EE maximization problem for the BSs becomes coupled due to the inter-cell interference. We develop a best response based optimization algorithm where each BS decides the optimum number of antennas iteratively and we use a game theoretic approach to prove the convergence to a Nash equilibrium. Our results indicate that both the increase in EE and reduction in rate is significant during low load and keeps becoming inconsequential with the increase in network load. Overall, EE can be improved by around 24% along with an energy saving potential of around 40% at the cost of around 12% reduction in the user rate over a day.

Note that for both cases, we model the BS as a state dependent M/G/m/m queue that does not allow any queuing as the steady-state distribution of the M/G/m with queuing is not analytically tractable. However, this queuing model is used to find the steady-state distribution of the users and our scheme can be used for any other queuing model that can give the steady-state distribution of the users.

Note that we do not consider the impact of pilot contamination and also utilize a simplified rate formula with the assumption of perfect CSI. Furthermore, in this analysis, we do not consider any residual power consumption when an antenna is turned off, especially with the probability of it being turned on immediately (which can be a item for further study). However, existing literature hardly consider the multi-cell scenario to analyze the EE of MM. In [39], a multi-cell scenario is analyzed by simulation to show that EE is still concave with number of antennas similar to the single cell scenario and in [109], a seven-cell system with MM has been analyzed to conclude that using the minimum number of antennas to provide the required average sum rate is the most energy efficient. However, to the best of our knowledge, no study considers the antenna adaptation of MM systems to improve EE in multi-cell scenario. In our study, we not only consider inter-cell interference, but also devise a mechanism to dynamically adapt the number of antennas in order to maximize EE and cope with the TLV.

The EE performance of MM and small cells have been compared in [113, 114] and the combined operation has shown to be more energy efficient in [115, 116]. Similar analysis involving inter-cell interference can be a future direction of this work as well. Also, the antenna adaptation scheme developed under this study can be compared with the extensive use of micro sleep, i.e., DTX and can be combined with their joint operation with or without minimum rate constraints for the users.

In this analysis, we do not consider any selection policy when reducing the number

of active antennas. However, all the antennas do not contribute equally, especially, when serving a low number of users [117]. Selecting those subset of antennas that contribute most when adapting the number of antennas may result in significant improvement in EE and can be a good candidate as future work.

References

[1] Böhringer C. *et al.*, "The EU 20/20/2020 targets: An overview of the EMF22 assessment," *Elsevier Energy Economics*, vol. 31, pp. 268273, 2009.

[2] Zeller D. *et al.*, "Sustainable Wireless Broadband Access to the Future Internet - The EARTH Project," in *The Future Internet*. Springer, 2013, vol. 7858, pp. 249271.

[3] Tombaz S., "On the Design of Energy Efficient Wireless Access Networks," Ph.D. dissertation, KTH Royal Institute of Technology, 2014.

[4] Blume O. *et al.*, "Approaches to energy efficient wireless access networks," in *Internation Symposium on Communications, Control and Signal Processing (ISCCSP)*, pp.1-5, 2010.

[5] Han C., "Green radio: radio techniques to enable energy-efficient wireless networks," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 46-54, 2011.

[6] Li G.Y. *et al.*, "Energy-efficient wireless communications: tutorial, survey, and open issues," *IEEE Wireless Commun.*, vol. 18, no. 6, pp. 28-35, 2011.

[7] Wu J. *et al.*, "Energy-Efficient Base-Stations Sleep-Mode Techniques in Green Cellular Networks: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 803-826, 2015.

[8] Hossain E. *et al.*, *Green Radio Communication Networks*, Cambridge: Cambridge University Press, 2012.

[9] Wu J., "Green wireless communications: from concept to reality," *IEEE Wireless Commun.*, vol. 19 , no. 4, Aug. 2012.

[10] Feng D. *et al*., "A survey of energy-efficient wireless communications," *IEEE Commun. Surveys & Tutorials*, vol. 15, no. 1, pp. 167-178, 2013.

[11] Vetter P., “Towards energy efficient wireline networks, an update from GreenTouch,” in *OptoElectronics and Communications (OECC)*, Jun. 2013.

[12] Wu J. *et al.*, *Green communications: theoretical fundamentals, algorithms, and applications*. CRC Press, USA, 2012.

[13] Auer G. *et al.*, “How much energy is needed to run a wireless network?,” *IEEE Wireless Commun.*, vol. 18, no. 5, October 2011, pp. 40-49

[14] Auer G. *et al.*, “D2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown,” Energy Aware Radio and Network Technologies (EARTH) INFSO-ICT-247733, ver. 2.0, 2012 [Online]. Available: `http://www.ict-earth.eu/`

[15] Frenger P. and Ericson M., “Assessment of Alternatives for Reducing Energy Consumption in Multi-RAT Scenarios,” in *VTC Spring*, 2014.

[16] Blume O. *et al.*, “Energy savings in mobile networks based on adaptation to traffic statistics,” *Bell Labs Technical Journal*, vol. 15, no. 2, pp. 7794, 2010.

[17] Jungnickel V. *et al.*, “The Role of Small Cells, Coordinated Multipoint and Massive MIMO in 5G,” *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44-51, May 2014.

[18] Leila Z. R., “Traffic dimensioning for multimedia wireless network,” Ph.D. dissertation, Virginia Polytechnic Institute and State University, 2003.

[19] Persson *et al.*, “Amplifier-Aware Multiple-Input Multiple-Output Power Allocation,” *IEEE Commun. Letters*, vol. 17, no. 6, pp. 112-115, Jun. 2013.

[20] Rippke I. A., “Design of Integrated, Efficient Power Amplifier for next-generation Wireless Communications,” Ph.D. dissertation, Cornell University, 2005.

[21] Wang F. *et al.*, “Design of wide-bandwidth envelope-tracking power amplifiers for OFDM applications,” in *IEEE Trans. Microwave Theory Tech.* vol. 53, no. 4, pp. 1244- 1255, 2005.

[22] Doherty W. H., “A new high efficiency power amplifier for modulated waves,” in *IRE*, vol. 24, pp. 1163-1183, 1936.

[23] Srirattana N. *et al.*, “A high efficiency multistage Doherty power amplifier for WCDMA,” in *IEEE Radio and Wireless Conf.*, pp. 397-400, 2003.

[24] Steinbeiser C. *et al.*, “HVHBT Doherty and Envelope Tracking PAs for High Efficiency WCDMA and WiMAX Basestation Applications,” in *IEEE Power Amplifier Symposium*, 2009.

[25] Raab, F.H.; “Efficiency of Doherty RF Power-Amplifier Systems,” in *IEEE Trans. on Broadcasting*, vol. BC-33, no. 3, pp. 77-83, 1987.

[26] Stephen Bruss, *Linearization methods*. 2003 [online]. Available: `http://www.ece.ucdavis.edu/\~spbruss/research/linearizer\_methods\_overview.pdf`.

[27] Hendy J.,“Envelope Tracking Excels in RF Power-Amplifier Efficiency,” *Electronic Design* [online]. Available: `http://electronicdesign.com/communications/envelope-tracking-excels-rf-power-amplifier-efficiency`

[28] K. Han *et al.*, “Power amplifier characteristic-aware energy-efficient transmission strategy,” in *International IFIP-TC6 Networking Conference*, 2007.

[29] Jeong J. *et al.*, “Modeling and Design of RF Amplifiers for Envelope Tracking WCDMA Base -Station Applications,” in *IEEE Trans. Microwave Theory Tech.*, vol. 57, no. 9, pp. 2148-2159, Sept. 2009.

[30] Prabhakar B. *et al.*, “Energy-efficient transmission over a wireless link via lazy packet scheduling,” in *IEEE INFOCOM*, pp. 386-394, 2001.

[31] Gamal A. *et al.*, “Energy-efficient scheduling of packet transmissions over wireless networks,” in *IEEE INFOCOM*, pp. 1773-1782, 2002.

[32] Uysal-Biyikoglu E. *et al.*, “Energy-efficient packet transmission over a wireless link,” *IEEE/ACM Trans. on Netw.*, pp. 487-499, 2002.

[33] Gouba O. A. and Louët Y., “Theoretical analysis of the trade-off between efficiency and linearity of the high power amplifier in OFDM context,” in *European Wireless*, pp.1-7, 2012.

[34] Goldsmith A. *et al.*,“Capacity limits of MIMO channels,” *IEEE J. Sel. Areas in Comm.* vol. 21, no. 5, pp. 684-702, 2003.

[35] Weingarten H. *et al.*, “The Capacity Region of the Gaussian MIMO Broadcast Channel,” in *Conf. Info. Sciences and Systems (CISS)*, 2004.

[36] Marzetta T. L., “Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas,” *IEEE Trans. on Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, 2010.

[37] Larsson E. G. *et al.*, “Massive MIMO for Next Generation Wireless Systems,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186-195, Feb. 2014.

[38] Prasad K. N. R. *et al.*, “Energy Efficiency in Massive MIMO-Based 5G Networks: Opportunities and Challenges,” [Onilne]. Available: `http://arxiv.org/abs/1511.08689`

[39] Björnson E. *et al.*,“Optimal Design of Energy-Efficient Multi-User MIMO Systems: Is Massive MIMO the Answer?,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059-3075, June 2015.

[40] Hien Q. N. *et al.*, “Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems,” *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436-1449, April 2013.

[41] Héliot F. *et al.*, “On the Energy Efficiency-Spectral Efficiency Trade-off over the MIMO Rayleigh Fading Channel,” *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1345-1356, May 2012.

[42] Kelly F.P. *et al.*, “Rate control for communication networks: Shadow prices, proportional fairness and stability,” *Journal of the Operat. Res. Society* vol. 49, pp. 237-252, 1998.

[43] Jen Y. C. and Smith J. M., “Generalized M/G/C/C state dependent queueing models and pedestrian traffic flows,” *Queueing Syst.*, Vol. 15, Issue 1-4, pp. 365-386, 1994.

[44] Cruz F. R. B and Smith J. M., “Approximate analysis of M/G/c/c state Dependent Queeing networks,” *Computers and Operation research*, vol. 34, Issue 8, pp. 2332-2344, 2007.

[45] Persson D. *et al.*, “Amplifier-Aware Multiple-Input Single-Output Capacity,” *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 913-919, March 2014.

[46] Ryu H.G. *et al.*, “Dummy sequence insertion (DSI) for PAPR reduction in the OFDM communication system,” in *IEEE Trans. on Consumer Electronics*, vol. 50, no. 1, pp. 89-94, Feb 2004.

[47] Björnson E. *et al.*, “Designing Multi-User MIMO for Energy Efficiency: When is Massive MIMO the Answer?,” in *IEEE WCNC*, 2014.

[48] Bonald T. *et al.*, “A queuing analysis of max-min fairness, proportional fairness and balanced fairness,” *Queue. Syst.*, vol. 53, pp. 65-84, 2006.

[49] Bonald T., “Throughput Performance in Networks with Linear Capacity Constraints,” in *Annu. Conf. on Inform. Sci. and Syst.*, pp. 644-649, 2006.

[50] Roberts J. W. and Massoulie L., “Bandwidth sharing and admission control for elastic traffic,” *Telecommun. Syst.*, vol. 15, pp. 185-201, 2000.

[51] Bonald T. and Proutiere A., “Insensitive bandwidth sharing in data networks,” *Queueing Syst.*, vol. 44, pp. 69-100, 2003.

[52] Kendall, D. G., “Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain,” *Ann. Math. Statist.* Vol 24, No. 3, pp. 338-354, 1953.

[53] Liu L. and Kulkarni V. G., “Balking and reneging in m/g/s systems exact analysis and approximations,” *Probab. Eng. Inf. Sci.* vol. 22 , pp. 355-371, 2008.

[54] Smith J. M., “M/G/c/K blocking probability models and system performance,” *Performance Evaluation*, vol. 52, Issue 4, pp. 237267, 2003.

[55] Adan I. and Resing J., *Queueing Systems*. 2015 [online]. Available: `http://www.win.tue.nl/~iadan/queueing.pdf`

[56] Cardieri P. and Rappaport T. S., “Statistics of the sum of log-normal variables in wireless communications,” in *IEEE VTC*, vol. 3, pp. 1823-1827, 2000.

[57] Ruttik K. *et al.*, “Modeling of the Secondary System’s Generated Interference and Studying of its Impact on the Secondary System Design,” *J. Radioengineering*, Dec. 2010.

[58] Björnson E. *et al*., “Multi-Objective Signal Processing Optimization: The Way to Balance Conflicting Metrics in 5G Systems,” *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 14-23, 2014.

[59] Topkis D. M., *Supermodularity and Complementarity*. Princeton, NJ, Princeton Univ. Press, 1998.

[60] Han Z. *et al.*, *Game Theory in Wireless and Communication Networks Theory, Models, and Applications*. Cambridge University Press, New York, 2012.

[61] DaSilva L. A. *et al.*, “Game theory in wireless networks,” in *IEEE Commun. Magazine*, vol. 49, no. 8, pp. 110-111, August 2011.

[62] Tembine, H. *et al.*, “Evolutionary Games in Wireless Networks,” in *IEEE Trans. Systems, Man, and Cybernetics*, Part B: vol. 40, no. 3, pp. 634-646, 2010.

[63] Levin J., Supermodular games. *Lecture notes on Game Theory and Economic Applications*, Department of Economics, Stanford University, 2006.

[64] Altman E. and Altman Z., "S-modular games and power control in wireless networks," *IEEE Trans. on Automatic Control*, vol. 48, no. 5, pp. 839-842, May 2003.

[65] Ozdaglar A., "Network Games: Learning and Dynamics," in *Conf. on Decision and Control (CDC)*, 2008 [online]. Available: `https://asu.mit.edu/sites/default/files/presentations/CDC08-tutorial.pdf`

[66] Vives X., "Supermodularity and supermodular games", *Occasional Paper*, OP no 07/18, May 2007. [online]. Available: `http://www.iese.edu/research/pdfs/op-07-18-e.pdf`

[67] Berry R. A. and Gallager R.G., "Communication over fading channels with delay constraints," *IEEE Trans. on Inf. Theory*, vol. 48, no. 5, pp. 1135-1149, May 2002.

[68] Neely M. J., "Optimal Energy and Delay Tradeoffs for Multiuser Wireless Downlinks," *IEEE Trans. on Inf. Theory*, vol. 53, no. 9, pp. 3095-3113, Sept. 2007.

[69] Chen W. *et al.*, "Energy-Efficient Scheduling with Individual Delay Constraints over a Fading Channel," in *Modeling and Optimization in Mobile, WiOpt*, pp. 1-10, 2007.

[70] Gamal A. E. *et al.*, "Throughput-delay trade-off in energy constrained wireless networks," in Symp. on *Inform. Theory*, pp. 439, 2004.

[71] Coleman T. P. and Medard M., "A distributed scheme for achieving energy-delay tradeoffs with multiple service classes over a dynamically varying network," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 5, pp. 929-941, Jun. 2004.

[72] Hiltunen K., "The Performance of Dense and Heterogeneous LTE Network Deployments within an Urban Environment," Ph. D. disseration, Aalto University, Helsinki, Finland, 2014.

[73] Holtkamp H. *et al.*, "Minimizing Base Station Power Consumption," in *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 297-306, 2014.

[74] Wu J. *et al.*, "Traffic-Aware Base Station Sleeping Control and Power Matching for Energy-Delay Tradeoffs in Green Cellular Networks," in *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4196-4209, 2013.

[75] Wu J. *et al.*, "Traffic-aware power adaptation and base station sleep control for energy-delay tradeoffs in green cellular networks," in *IEEE Globecom*, pp. 3171-3176, 2012.

[76] Karl H. (ed.), "An Overview of Energy-Efficiency Techniques for Mobile Communication Systems," TKN Technical Report from university of Berlin, 2003.

[77] Chen Y. *et al.*, "Impact of Non-Ideal Efficiency on Bits Per Joule Performance of Base Station Transmissions," in *IEEE VTC*, pp. 1-5, 2011.

[78] Chen Y. *et al.*, "Fundamental trade-offs on green wireless networks," in *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30-37, Jun. 2011.

[79] Corless R. M. *et al*, "On the Lambert W function," *Advances in Computational Mathematics* vol 5, pp. 329-359, 1996.

[80] Desset C. *et al.*, "Flexible power modeling of LTE base stations," in *IEEE WCNC*, pp. 2858-2862, 2012.

[81] "Alcatel-Lucent demonstrates up to 27 percent power consumption reduction on base stations deployed by China Mobile," Press release, Mobile World Congress, Barcelona, Feb. 2009.

[82] Oh E., Krishnamachari B., "Energy Savings through Dynamic Base Station Switching in Cellular Wireless Access Networks," in *IEEE GLOBECOM 2010*, pp.1-5, 2010.

[83] Gong J. *et al.*, "Traffc-aware base station sleeping in dense cellular networks," in *IEEE IWQoS*, 1-2, 2010.

[84] Fehske A. J. *et al.*, "Energy efficiency improvements through micro sites in cellular mobile radio networks," in *IEEE Globecom workshop*, 2009.

[85] Richter F. and Fettweis G., "Cellular Mobile Network Densification Utilizing Micro Base Stations," in *ICC 2010*, May 2010.

[86] Tombaz S. *et al.*, "Energy Efficiency Improvements Through Heterogeneous Networks in Diverse Traffic Distribution Scenarios," in 6th *Int. ICST Conference on Communications and Networking in China (Chinacom11)*, pp. 708-713, 2011.

[87] Peng J. *et al.*, "Energy-Aware Cellular Deployment Strategy Under Coverage Performance Constraints," in *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 69-80, 2015.

[88] "Deployment strategies for Heterogeneous Networks," whitepaper by NSN, May 2012 [online]. Available : http://networks.nokia.com/sites/default/files/document/nokia_deployment_strategies_for_heterogeneous_networks_white_paper_0.pdf

[89] ZhuYan Z. *et al.*, "TD-LTE Network Evolution with In-Band and Out-Band Micro Cells Deployment," in *IEEE VTC Spring*, pp.1-5, 2012.

[90] Zhang X. *et al.*, "Energy-Efficiency Study for Two-tier Heterogeneous Networks (HetNet) Under Coverage Performance Constraints," *Mobile Networks and Applications*, Volume 18, Issue 4, pp. 567-577, August 2013.

[91] "Designing, Operating and Optimizing Unified Heterogeneous Networks" whitepaper by NSN, 2011.

[92] Oh E. *et al.*, " Towards dynamic energy-efficient operation of cellular network infrastructure," *IEEE Commun. Mag.,* vol. 49, no. 6, pp. 56-61, June 2011.

[93] Balasubramanian N. *et al.*, "Energy Consumption in Mobile Phones: A Measurement Study and Implications for Network Applications," in 9th *ACM SIGCOMM IMC*, pp. 280-293 2009.

[94] Huang J. *et al.*, "A close examination of performance and power characteristics of 4G LTE networks," in 10th *ACM MobiSys*, pp. 225-238, 2012.

[95] Tombaz S. *et al.*, "Impact of Backhauling Power Consumption on the Deployment of Heterogeneous Mobile Networks," in *IEEE GLOBECOM*, 2011.

[96] Tombaz S. *et al.*, "Mobile Backhaul in Heterogeneous Network Deployments: Technology Options and Power Consumption," in *IEEE ICTON*, 2012.

[97] Tombaz S. *et al.*, "Is Backhaul Becoming a Bottleneck for Green Wireless Access Networks?," in *IEEE ICC*, 2014.

[98] Tombaz S. *et al.*, "Green Backhauling for Rural Areas," In *IEEE International Conference on Optical Network Design and Modeling (ONDM)*, 2014.

[99] Yang H. and Marzetta T. L., "Total energy efficiency of cellular large scale antenna system multiple access mobile networks," in *IEEE Online GreenComm*, 2013.

[100] Ha D. *et al.*, "Energy efficiency analysis with circuit power consumption in massive MIMO systems," in *IEEE PIMRC*, pp. 938-942, 2013.

[101] Le N. P. *et al.*, "Adaptive antenna selection for energy-efficient MIMO-OFDM wireless systems," in *International Symp. WPMC*, pp. 60-64, Sept. 2014.

[102] Frenger, P. *et al.*, "Radio network energy performance of massive MIMO beamforming systems," in *IEEE PIMRC*, pp. 1289-1293, 2014.

[103] Cheng H. V. *et al.*, "Massive MIMO at Night: On the Operation of Massive MIMO in Low Traffic Scenarios," in *IEEE ICC*, June 2015.

[104] Lu L. *et al.*, "An Overview of Massive MIMO: Benefits and Challenges," *IEEE J. Sel. Topics in Signal Processing*, vol. 8, no. 5, pp. 742-758, Oct. 2014.

[105] Jiang C. and Cimini L. J., "Antenna Selection for Energy-Efficient MIMO Transmission," in *IEEE Wireless Commun. Letters*, vol. 1, no. 6, pp. 577-580, 2012.

[106] Brante G. *et al.*, "Outage probability and energy efficiency of cooperative MIMO with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5896-5907, 2013.

[107] Zhou X. *et al.*, "Invited Paper: Antenna selection in energy efficient MIMO systems: A survey," in *Communications*, China , vol. 12, no. 9, pp. 162-173, 2015.

[108] Hui Li *et al.*, "Energy Efficiency of Large-Scale Multiple Antenna Systems with Transmit Antenna Selection," in *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 638-647, 2014.

[109] Liu W. *et al.*, "Is Massive MIMO Energy Efficient?" *CoRR*, abs/1505.07187. Online. Available: `http://arxiv.org/abs/1505.07187`

[110] Wang L. *et al.*, "Proxies for energy-efficient web access revisited," in *Int. Conf. on Energy-Efficient Computing and networking* pp. 55-58, 2011.

[111] Hiltunen K., "Improving the energy-efficiency of dense LTE networks by adaptive activation of cells," in *IEEE ICC Workshops*, pp. 1150-1154, 2013.

[112] Huang K. and Lau V., "Enabling wireless power transfer in cellular networks: Architecture, modeling and deployment," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 902912, 2014.

[113] Nguyen H. D. and Sun S., "Massive MIMO versus Small-Cell Systems: Spectral and Energy Efficiency Comparison," submitted to *IEEE Trans. Wireless Commun.*, Sept. 2015.

[114] Liu W. *et al.*, “Energy efficiency comparison of massive MIMO and small cell network,” in *IEEE GlobalSIP*, pp. 617-621, 2014.

[115] Björnson E. *et al.*, “Energy-efficient future wireless networks: A marriage between massive MIMO and small cells,” in *IEEE SPAWC*, pp. 211-215, 2015.

[116] Björnson E. *et al.*, “Deploying Dense Networks for Maximal Energy Efficiency: Small Cells Meet Massive MIMO,” [Online]. Available: `http://arxiv.org/abs/1505.01181`

[117] Gao X. *et al.*, “Massive MIMO in Real Propagation Environments: Do All Antennas Contribute Equally?,” *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 3917-3928, Nov. 2015.

Errata

Publication V

Fig. 2, Fig. 3, Fig. 5, and Fig. 8 should be replaced by the Fig. 4.1, Fig. 4.2, Fig. 4.3 and Fig. 4.4 respectively.

The y-axis labels for Fig. 4, Fig. 6 and Fig. 7 should be scaled down according to Fig. 4.3 where corrected labels for TPA and ET-PA for 20 W case is shown.

The energy savings over a day should be 12% instead of 30%.

Aalto-DD 117/2016

ISBN 978-952-60-6863-3 (printed)
ISBN 978-952-60-6864-0 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University

Department of Communications and Networking
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL DISSERTATIONS