Maynooth University

National University
of Ireland Maynooth

Resource Allocation for
Next Generation Radio Access Networks

In Partial fulfilment of the requirements for the degree of
PhD

by:

Bahar Partov Poor

Supervisor:
Professor Douglas J. Leith

Hamilton Institute
Maynooth University

February, 2016

Contents

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy from the Hamilton Institute is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____

Date: 1st Feb, 2016.

Acknowledgement

I feel truly fortunate to be given the opportunity to pursue a research path in stimulating environments of the Hamilton Institute, and Bell Laboratories and alongside a number of very bright people. There are many people that I would like to thank for their feedback and support in the past three years.

First and foremost, I'm forever grateful for the insight, and generous guidance of my supervisor Prof. Douglas Leith, whose endless patience is beyond description. I'm most thankful for his help to making my time working on the material of this thesis, a thoroughly constructive experience.

I also like to thank Dr. Holger Claussen and Dr. Rouzbeh Razavi of Bell Labs, for providing me the opportunity to have a taste of research in industry and for the helpful discussions. Thanks are also due to Dr. David Lopez-Perez, Dr. Lester Ho, Dr. Vijay Venkateswaran, and Dr. Afra Mashhadi of Bell Labs for their friendship and the technical conversations that we had in the past few years. I have also benefited from discussions with Dr. David Malone, Dr. Oliver Mason, Prof. Ken Duffy, Dr. Buket Benek Gursoy, Dr. Andreas Garcia Saavedra, Dr. Cristina Cano and Dr. Mohammad Karzand of Hamilton Institute, as well as many other fellow students. Also thanks go to Rosemary Hunt and Kate Moriarty for their help with the various administrative issues that needed to be dealt with, during the course of my research at the Institute.

On a more personal level, I'm forever indebted to my parents, to my dearest friend, Bijan and to my good humoured brother, for not only tolerating me but also being a constant supporting force over the years.

Abstract

Driven by data hungry applications, the architecture of mobile networks is moving towards that of densely deployed cells where each cell may use a different access technology as well as a different frequency band. Next generation networks (NGNs) are essentially identified by their dramatically increased data rates and their sustainable deployment. Motivated by these requirements, in this thesis we focus on (i) capacity maximisation, (ii) energy efficient configuration of different classes of radio access networks (RANs). To fairly allocate the available resources, we consider proportional fair rate allocations. We first consider capacity maximisation in co-channel 4G (LTE) networks, then we proceed to capacity maximisation in mixed LTE (including licensed LTE small cells) and 802.11 (WiFi) networks. And finally we study energy efficient capacity maximisation of dense 3G/4G co-channel small cell networks. In each chapter we provide a network model and a scalable resource allocation approach which may be implemented in a centralised or distributed manner depending on the objective and network constraints.

Notations and Abbreviations

Notations:

Symbol	Meaning
$\mathbb{R}$	The set of real numbers
$\mathbb{R}^n$	The set of all n-tuples of real numbers
$\mathbb{N}$	The set of natural numbers
$\mathbb{E}[.]$	The expected value of a random variable
$\mathbf{I}$	The indicator matrix
$[.]^+$	The projection on the positive orthant
$\mathbb{1}_{(x)}$	The indicator function
$\mathcal{B}$	The set of base stations
$\mathcal{B}_m$	The set of macro base stations
$\mathcal{B}_{sc}$	The set of small cell base stations
$\mathcal{I}$	The set of available LTE sub-channels
$\mathcal{A}$	The set of access points
$\mathcal{A}_u$	The set of access points in the geographic area of user u
$\mathcal{U}$	The set of users
$\mathcal{U}_a$	The set of users which can connect to AP a
$\mathcal{U}_{a,t}$	The set of users which can connect to AP a at time slot t
$\mathcal{U}_b$	The set of users in the geographic area of base station b
$\mathcal{M}_{a,t}$	The set of MAC slots of access point a that are fully contained within time slot t
$\bar{\mathcal{M}}_{a,t}$	The set of MAC slots which are only partially contained in time slot t
$\mathcal{E}^i$	The set of user-BS pairs for which transmissions on sub-channel i interfere
Λ	The set of Lagrangian multipliers
B	A compact set in $\mathbb{R}^n$
C	A compact set of the feasible solutions to a constrained optimisation problem
$\mu(\Gamma \cap \Sigma)$	The volume of the intersection of two sets Γ and Σ
$\mathcal{T}$	The training data set of RF fingerprint vectors

Symbol	**Meaning**
$\mathcal{N}_i$	The set of neighbouring cells associated with the RF fingerprint vector X_i
$\mathcal{S}$	The set of snapshots
$\mathbf{S}$	A zero mean Gaussian Random Variable
$\mathbf{A}_{u,t}$	A random variable denoting the access point which user u selects at time slot t
$\mathbf{N}_{a,t}$	A random variable denoting the number of users associated with access point a at time slot t
$\boldsymbol{\mathcal{X}}_{i,u}$	A random variable indicating transmissions for the 802.11 user u at MAC slot i
$\boldsymbol{\mathcal{Y}}_{i,u}$	A random variable indicating successful transmission for 802.11 user u at MAC slot i
$\boldsymbol{\mathcal{Z}}_i$	A random variable indicating absence of transmissions at MAC slot i of the 802.11 channel
$U(.)$	The utility function
$L(.,\Lambda)$	The Lagrangian function
$P(.)$	The penalty function associated a constrained optimisation problem
r_u	The throughput of user u
$\bar{r}$	The maximum achievable throughput
$\underline{r}$	The minimum acceptable throughput
γ_u	The average signal to interference plus noise ratio for user u
γ_u^i	The signal to interference plus noise ratio for user u on sub-channel i
$\gamma_{b,u}^i$	The signal to noise ratio for base station b- user u pair on sub-channel i
$\zeta_{b,u}^i$	The fraction of time sub-channel i of BS b is used by user u
ω	Total available bandwidth
ω^i	Bandwidth of sub-channel i
p_b	The transmit power of base station b
σ_n^2	The noise power at the receiver
p_u	transmit power of user u
p_b^p	transmit power of pilot channels for base station b
p_b^d	transmit power of data channels for base station b
θ_b	The tilt angle of base station b

Symbol	Meaning
$\theta_{b,u}$	Horizontal angle between reference plane of user u and its direction towards base station b
$\hbar$	The height difference between a base station and a user
$d_{b,u}$	The distance between base station b and user u
$\ell_{b,u}$	The path loss between base station b and user u
ℓ_0	The fixed path loss factor
ϱ	The path Loss exponent
$\tilde{G}_{b,u}(\theta_b)$	The base station antenna gain
$\tilde{G}_0$	The maximum gain of the antenna
$\tilde{G}_v(\theta_b, d_{b,u})$	The base station antenna vertical gain
$G_v(\theta_b, d_{b,u})$	The base station antenna vertical gain exponent
$\hat{G}_v(\theta_b, d_{b,u})$	The linear approximation to base station antenna vertical gain exponent
$P_{R,u}(\theta_{b(u)})$	The received power from base station $b(u)$ at user u as a function of the tilt angle, $\theta_{b(u)}$
$\hat{P}_{R,u}(\theta_{b(u)})$	The approximate received power from base station $b(u)$ at user u
$h_{b,u}$	Channel gain between base station b and user u
$h_{b,u}^i$	sub-channel i gain between base station b and user u
$\mathbf{h}_u$	SIMO link gain vector for user u
$\boldsymbol{\Upsilon}_{b,u}$	The Rayleigh flat fading vector in the SIMO link between base station b and user u
$\mathbf{v}_u$	The inter-cell interference vector in the SIMO link of user u
$\mathbf{n}_u$	The noise vector in the SIMO link of user u
$\boldsymbol{\kappa}_u$	The MMSE combining vector in the SIMO link of user u
$\boldsymbol{\Phi}_u$	The autocorrelation of the interference vector in the SIMO link of user u
$c_{a/b,u}$	The physical rate of access point a/ base station b
s_u	The aggregate throughput of user u in the 802.11 network
$\rho_{a,u}$	The normalised throughput of the user u associated with access point a
T	The duration of 802.11 scheduling time slot t
$z_{a,u}$	The fraction of time that user u is associated with access point a
$\tau_{a,u}$	The probability of transmission for user u connected to access point a

Symbol	Meaning
τ	The probability of transmission in the 802.11 channel, (Also) the probability of LTE/3G user activity
$p_{a,u,n}$	The probability having n users connected to access point a, conditioned on that user u is associated to access point a
$L_{a,u}$	The payload in bits of a successful 802.11 transmission when user u is associated to base station b
$T_{succ,a,u}$	The mean duration of a successful transmission time of user u to access point a
T_{coll}	The mean duration of a collision
σ	The mean duration of an idle MAC slot
X_u	The RF fingerprint vector of user u (a vector of received powers)
Y_u	The query RF fingerprint vector reported by user u
$\mathcal{F}$	The RF fingerprint classifier
$\hat{\mathcal{F}}(X)$	The predicated cell association vector w.r.t RF fingerprint vector X
$\mathrm{P}(\boldsymbol{b}, \hat{\mathcal{F}}(X_u))$	The loss function identifying mismatch between the classifier's prediction, $\hat{\mathcal{F}}(X_u)$ and the target vector, $\boldsymbol{b}$
$Err_{\mathcal{T}}$	The misclassification error of a classifier trained on data set $\mathcal{T}$
Err	The expected misclassification error
$\Omega(s)$	The normalised traffic load factor at snapshot s
$N_{au}(s)$	The number of active users at snapshot s
$e_b(p_b^d, p_b^p)$	The energy consumption of base station b
$\Xi(p^d, p^p)$	The cost function to minimise using simulated annealing
N_b	The number of users served by base station b
N_{max}	The maximum allowable number of users for each small cell base station
N	The number of users per km^2
F_{sc}	The hotspot factor
N_o	The operator market share
A_h	The area of a hotspot
P_0	The power consumption of a small cell base station in the idle state
Δ	The rate of change in electrical power usage with varying RF power
P_{max}	The maximum allowable transmit power for each small cell base station
α	The step size that is used in primal-dual subgradient updates
c	The penalty parameter
ϑ	The simulated annealing temperature parameter
ς	The cooling schedule in the simulated annealing algorithm

Abbreviations:

Symbol	Meaning
3GPP	3rd Generation Partnership Project
ABS	Almost Blank Subframes
ACK	Acknowledgement
Ant.	Antenna
AP	Access Point
AMPDU	Aggregated MAC Protocol Data Unit
BS	Base Station
CDF	Cumulative Distribution Function
CRE	Cell Range Expansion
CDMA	Code Division Multiple Access
CoMP	Coordinated Multi Point
CSMA	Collision Sense Multiple Access
CW	Contention Window
E911	Enhanced 911
eICIC	enhance Inter-cell Interference Coordination
FCC	Federal Communication Commission
FDD	Frequency Division Duplexing
FCS	Frame Check Sequence
GPS	Global Positioning System
GSM	Groupe Spécial Mobile
IPN	Interference Plus Noise
KKT	Karush-Kuhn-Tucker
KNN	K- Nearest Neighbours
LBS	Location Based Services
LMMSE	Linear Minimum Mean Square Error
LTE	Long Term Evolution
MAC	Medium Access Control
MBS	Macro Base Station
MDP	Markov Decision Process
MIMO	Multiple Input Multiple Output
FDM	Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PHY	Physical

Symbol	Meaning
PLCP	Physical Layer Convergence Protocol
PSDU	PLCP Service Data Unit
RAN	Radio Access Networks
RAT	Radio Access Technology
SCBS	Small Cell Base Station
SC-OFDMA	Single Carrier Orthogonal Frequency Multiple Access
SD	Standard Deviation
SDMA	Space Division Multiple Access
SIMO	Single Input Multiple Output
SISO	Single Input Single Output
SINR	Signal to Interference Noise Ratio
SNR	Signal to Noise Ratio
SO	Self Organisation
RF	Radio Frequency
RIPN	Residual Interference Plus Noise
RSCP	Received Signal Code Power
RSRP	Reference Signal Received Power
Rx	Receiver
TCP	Transmission Control Protocol
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
Tx	Transmitter
QoS	Quality of Service
UE	User Equipment
UMTS	Universal Mobile Telecommunication Systems
WLAN	Wireless Local Area Network

List of Figures

List of Tables

Introduction

In this chapter we briefly discuss the characteristics and challenges of the next generation of radio access networks (RANs). We then summarise the approach that we have taken in the thesis. After outlining the structure of the thesis, we conclude by summarising the principal contributions of the work.

1.1 Next Generation Radio Access Networks

The next generation of mobile networks (long term evolution- LTE) is steadily expanding across the globe, providing users with similar data rates to a wired broadband network [1],[2]. This expansion is fuelled by the popularity of smart devices/objects with small form factors [3]. The popularity of such devices has led to a number of data hungry applications in the area of mobile health, social media, and video streaming. Worldwide LTE-related subscriptions reached 200 million in 2013, and the number of LTE subscription is estimated to exceed 2 billion by the end of 2019 [4].

As observed by Martin Cooper, increase in cellular network capacity is due to exploitation of spatial diversity, better use of the spectrum and improvement in link efficiency in descending order [5]. The dependency of the capacity gain on the above factors, can also be explained by the Shannon theorem for the capacity of a noisy channel [6].

For example, consider LTE systems established based on orthogonal frequency division multiple access (OFDMA) technology where each cell divides the available spectrum into orthogonal sub-channels. The sub-channels are then assigned to the users of a cell by the cell scheduler. To maximise spectral efficiency, LTE was originally designed to operate with a frequency reuse factor of one [7] i.e. all cells exploit the resources of the same frequency band. In a system with a frequency reuse factor of one, each cell scheduler indepen-

dently distributes the bandwidth resources among its users. As a result users within a cell operate in orthogonal channels and do not interfere with each other. However lack of coordination among cells results in interference among the users of neighbouring cells. From the Shannon theorem on the capacity of a noisy channel and by treating interference between co-channel cells as Gaussian noise, an upper bound of a user's data rate is given by

$$r_u \leq \frac{\omega}{N_b} \log(1 + \gamma_u) \tag{1.1}$$

where r_u and γ_u, denote the data rate, and signal to interference noise ratio (SINR) of user u respectively. We let ω denote the available bandwidth and N_b be the number of users that equally share the available bandwidth of a base station (BS) b (note that for simplicity, in this chapter the terms BS and cell are used interchangeably). Equal allocation of bandwidth is achieved for example by a LTE cell that assigns bandwidth to users in a round robin manner. It can be seen from Eq (1.1), that an increase in the available spectrum, ω will result in a linear increase in the capacity. An increase in the link efficiency on the other hand will result in the bound in inequality (1.1) becoming tighter. By increasing the number of cells, the number N_b of users per cell b, decreases resulting in an increase in the capacity (assuming that SINR, γ_u remains unchanged).

Moreover additional gains in channel capacity can be achieved by spatial multiplexing i.e. introducing independent streams of data through multiple input multiple output (MIMO) antennas at the transmitter and the receiver [8], [9]. Suppose that a number n_T of transmit antennas and n_R of received antennas are used in the system. The seminal work of E. Telatar [9] shows that the MIMO capacity with channel knowledge at the receiver scales as $min(n_T, n_R)\ log(\gamma_u) + \mathcal{O}(1)$, where γ_u is the signal to noise ratio (SNR) of a Gaussian channel. By treating interference as Gaussian noise, the results will also hold in a system with co-channel interference.

Here we refer to the effect of decreasing cell sizes and use of multiple data streams as spatial diversity. Network densification refers to the combined effect of spatial diversity and spectrum aggregation. We start this thesis with consideration of capacity optimisation in current LTE networks and then proceed to consider network densification. With the industry estimating that the majority (nearly 80%) of mobile traffic will be generated in indoor locations [10], we will focus on the cell size reduction element when concerned with spatial diversity.

1.1.1 Network Densification

1.1.1.1 Spatial Diversity: Small Cells

Small cell networks have emerged as a result of the need for network densification. Small cells are low cost, low powered BSs which may co-habit on the
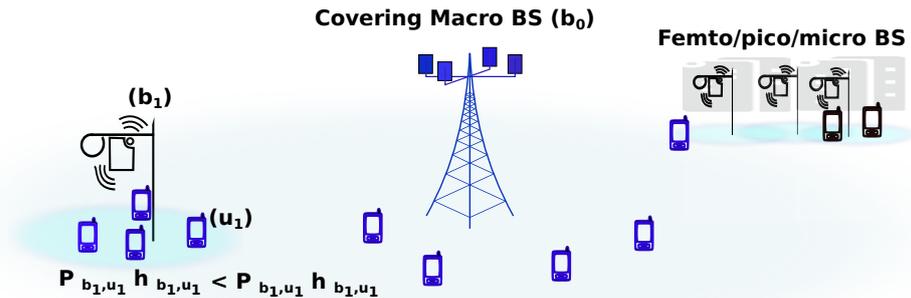
Figure 1.1: UE to BS assignment defined by Eq (1.2). UEs which are served by the macro BS are identified by blue color.

same frequency band as the macro BSs (co-channel), or might be deployed in a different part of the spectrum (e.g. unlicensed femto-cells). In addition to reduced load per cell, increasing the number of cells (reducing the cell sizes) has another favourable effect i.e. reducing the path loss and increasing the line of sight component of the signal. This means that the normally fast-fading wireless channel contains strong deterministic non-fading components [11] which in turn may make the radio propagation predictions difficult in comparison to that of a macro BSs network. In a conventional cellular system a user connects to the cell which provides that user with the maximum received signal level [7, Section 22.3.1] i.e.

$$b_u \in \arg\max_{b \in \mathcal{B}} p_b h_{b,u} \tag{1.2}$$

where $\mathcal{B}$ is the set of BSs, p_b is the transmit power of BS b, and $h_{b,u}$ is the average channel gain between BS b and user u. For example consider the scenario in Fig 1.1. Although the user equipment (UE), u_1 is closer to the small cell BS (SCBS) b_1, its received power from the covering macro BS, b_0 is higher. This is due to the higher transmit power of the macro BSs compared to that of a small cell. Therefore the association rule in Eq (1.2) will result in assignment of u_1 to b_0. This assignment prevents efficient offloading through small cell BSs in heavy traffic load conditions or in hotspot locations where spatial distribution of traffic is non-uniform and it is concentrated in busy areas. An efficient UE to BS assignment based on the traffic demand is similar to the task of assigning jobs to servers i.e. load balancing in computer networks. In this thesis, we may use the two terms interchangeably.

In addition to the effect of asymmetric transmit powers on user association, co-channel small cells may suffer from increased levels of interference caused by the covering macro BSs or nearby small cell BSs. Release 11 of the 3GPP (LTE) standard [12] addresses these issues under the enhanced inter cell interference coordination (eICIC) guidelines. The proposed eICIC framework

includes cell range expansion (CRE) and the use of almost blank subframes (ABS). CRE enhances interference management on the downlink by assigning a bias value to the transmit power of small cell BSs, therefore users receiving a weaker signal from a small cell will not become associated with a covering macro BS with a stronger signal. Although CRE may solve the load balancing problem specifically for non-uniform load distributions, it does not mitigate the downlink interference caused by the covering macro BSs and neighbouring small cells. To counteract the interference caused by macro cells, macro BSs can use time sharing by applying the ABS method. In ABS mode, macro BSs only transmit control data on the common control channel and set a side the rest of the (ABS) subframe resources for the transmitting small cell BSs. Although both CRE and ABS provide general guidelines, an efficient use of the resources requires optimisation of the bias values and time sharing periods depending on the user load distribution.

1.1.1.2 Spectrum Aggregation

With the consistent increase in mobile data demand, spectrum aggregation methods have attracted significant interest in the literature and in the future unlicensed spectrum is expected to become an essential part of mobile broadband services. This includes the use of 2.4GHz, 5GHz and 60GHz bands. Although the existing technologies in these bands don't support full mobility, they may be used to improve capacity and quality of service (QoS). 802.11 networks are a popular class of technologies in the unlicensed band, and in Chapter 3 we will study principled approaches towards their integration with licensed LTE networks.

1.1.2 Self Organisation

Suppose a macro cell network has a static spatio-temporal traffic distribution. In this case fixed design parameters can be used for long periods of time. Currently, tuning of these parameters is often carried out by the operators manually, using network planning tools or drive tests [13]. From the operator's perspective manual optimisation of a large scale and dynamic network incurs an additional operational cost. As already discussed, network densification and reduced cell sizes are becoming key elements of the next generation of cellular networks. However confining cells to a smaller geographic area, implies an increased probability of arrival and departures of users per cell. Hence the spatial and temporal distribution of the traffic load per small cell is also expected to become more dynamic compared to that of a macro cell. Autonomous optimisation of networks parameters which uses a minimum amount of overhead is therefore attractive for both the operator and the users. Such autonomous techniques are often referred to as self organisation (SO) methods by the network standardisation bodies [14]. SO covers three classes: self-configuration, self-optimisation and self-healing. These classes allow for auto-configuration of basic system parameters, resource allocation

and recovery from node failures. For a more detailed description of different self-organisation modes refer to [15]. In this thesis we mainly consider the self-optimisation aspects of SO. However in some instances the proposed self-optimised solutions can also be applied for self-healing purposes as we show in Chapter 2 or self-configuration as we show in Chapters 4 and 5.

Self organisation techniques have been extensively investigated for other communication networks, for example wireless sensor and ad hoc networks. Although in the context of cellular networks they are relatively new, an extensive body of research has targeted the challenges in this area in the last few years. To name a few see [16], [17],[18], [19] and references therein. In this thesis we develop a number of self organisation methods specifically designed for the next generation of cellular networks where dense deployment and heterogeneity are prevalent.

1.2   Introductory Remarks

In this thesis we deal with optimal allocation or network configuration for next generation RANs. We will present detailed system models for the access technologies considered in the individual chapters. In this Section we briefly outline the multiple access strategies that are historically used in RANs. We describe notions of optimality and informally discuss some aspects of optimisation that are used in this thesis.

1.2.1   Multiple Access Strategies

Being concerned with multi-user resource allocation, we are primarily interested in modelling RANs based on their medium access strategies. These are often implemented in the highest level of the data link layer and governed by the medium access control (MAC) interface. Below is a summary of the multiple access strategies that are frequently applied in wireless systems:

- FDMA: Frequency division multiple access refers to sharing the spectrum between users by a one to one mapping of frequency channels to users. This allocation can be made permanently or temporarily. For more details see [20, Chapter 2].

- TDMA: Time division multiple access (North America cellular standard IS-136) refers to a one to one mapping of time intervals to users so that all users are allowed to use the entire available spectrum. For more details see [20, Chapter 2].

  CSMA: Carrier sense multiple access is a randomised variation of TDMA. To identify idle time intervals, CSMA applies a probabilistic method of accessing the spectrum by sensing the channel. In its simplest form if the channel is sensed as idle, the user attempts to transmit, if the the channel is sensed as busy, the transmission is

scheduled after a random back off time. For more details see [20, Chapter 4].

- CDMA: Code division multiple access (North America cellular standard IS-95) is when all users use the available spectrum. However to avoid interference, signals must be encrypted with distinct codes [8, Chapter 4].

- SDMA: Space division multiple access is to share the spectrum by exploiting the spatial distribution of the users and use of beam-forming techniques [8, Chapter 10].

Next generation mobile networks aim to increase data rates, cost efficiency and utilisation of the network resources. To achieve these goals LTE technology has emerged as a set of radio and core network specifications and as a technological evolution of its predecessors (GSM, UMTS, and wide band CDMA networks). In reference to the multiple access mode, LTE adopts a variation of FDMA referred to as OFDMA. Similar to FDMA, OFDMA is based on frequency sharing. But it also allows overlapping sub-channels [1].

As discussed in Section 1.1, it is common that the system frequency reuse factor is one and cells independently assign their users on the orthogonal sub-channels of a given frequency spectrum. This autonomous OFDMA cell scheduling may result in interference in the uplink or the downlink signals of nearby cells. However in order to optimise frequency allocations, most recent releases of the LTE standard [21] allow for coordinated scheduling between neighbouring cells. In Chapters 2, 4 and 5 of this thesis, we consider independent OFDMA cell scheduling in LTE systems with frequency reuse factor of one. While in Chapter 3 we consider coordinated scheduling among LTE BSs. We also show in Chapters 4 and 5 that the introduced dynamic power allocation method can be applied to both OFDMA and CDMA systems.

In Chapter 3 where we consider LTE and 802.11 integration, we obtain 802.11 network models based on the CSMA mechanism. We additionally introduce time sharing procedures among LTE and 802.11 BSs/Access Points (APs) in order to utilise the network performance.

---

[1]In OFDMA, each data symbol is individually modulated by an OFDMA subcarrier, resulting in high variations in the amplitude of the OFDM modulated signal [7]. This becomes problematic due to the non-linearity of the power amplifiers outside their dynamic range. To avoid additional cost and power consumption at the UE, a variation of OFDMA i.e Single-Carrier OFDMA (SC-OFDMA) is selected for the uplink transmissions. Despite what the name suggests, SC-OFDMA also modulates the signal on multiple subcarriers. However to avoid high amplitude variations, the signal modulated onto a given subcarrier is a linear combination of all the data symbols transmitted at the same time instant.

1.2.2   Fairness and Utility Functions

By optimising the network parameters we aim to better utilise the limited network capacity while ensuring each user experiences satisfactory performance. Network resources however can be allocated amongst users according to various policies. To perform the task of resource allocation, we aim to provide users with efficiency as well as fairness. Although there is no notion of fairness that is universally accepted, max-min and proportional fair allocations [22] are commonly referred to as fair allocation schemes.

1.2.2.1   Max-min Fairness

Let $\mathcal{U} = \{1, 2, \ldots, n\}$ denote the set of users. A throughput allocation $[r_1^*, r_2^*, \ldots, r_n^*]$ to the set of users $\mathcal{U}$, is max-min fair if it is feasible and if r_u^* for any $u \in \mathcal{U}$ cannot be increased without decreasing $r_{u'}^*$ for some $u' \in \mathcal{U}$ for which $r_{u'}^* \leq r_u^*$. i.e. a decrease in $r_{u'}^*$ that leads to increase to r_u^* is forbidden [23]. Max-min fairness therefore ensures that the needs of users with minimum amount of resources is met. As shown in [24] max-min fairness may result in performance inefficiencies at the expense of favouring users with smaller rates.

1.2.2.2   Proportional Fairness

An alternative notion of fairness is proportional fairness which is introduced in the context of the game theory [25]. A feasible throughput allocation $[r_1^*, r_2^*, \ldots, r_n^*]$ is proportionally fair if and only if for any other feasible allocation $[r_1, r_2, \ldots, r_n]$, we have

$$\sum_{u \in \mathcal{U}} \frac{r_u - r_u^*}{r_u^*} \leq 0 \tag{1.3}$$

In a proportional fair rate allocation scheme, if the proportion by which the rate of a user changes is positive, there exists at least another user for which the proportional change in rate is negative.

1.2.2.3   Utility Functions

The utility of a user can be interpreted both as the level of satisfaction that the user experiences and as a notion of fairness that is defined by the network [26]. We let $U(r_u)$ denote the utility function of user u transmitting/receiving at rate r_u. The goal of resource allocation is to maximise the sum of the user utilities i.e. max. $\sum_{u \in \mathcal{U}} U(r_u)$ subject to network constraints. It can be shown that different notions of fairness can be unified by an α-optimal utility function that is described in [26, p18] and that the log utility function, $U(r_u) = \log r_u$, yields a proportional fair rate allocation [24, Theorem 1.2.3.].

Note that the mean delay experienced by a user is inversely proportional to the user's rate. For example the mean delay for a large file transfer is equal to

the file size divided by the mean data rate. A delay optimal objective, which is also referred to as delay potential fairness [27], is described by min $\sum_{u \in \mathcal{U}} \frac{1}{r_u}$. This objective will penalise very low rates (throughputs).

In this thesis when we are concerned with optimising capacity, we often adopt the proportional definition of fairness. Moreover with network densification, improving energy efficiency is becoming an essential part of the next generation mobile networks. To address energy efficiency in large small cell networks, in Chapter 4 we introduce methods for energy efficient configuration of the network. While in Chapter 5 we consider energy efficiency together with the proportional fair rate allocation.

1.2.3 Optimisation Tools

To solve a network utility optimisation problem, a range of optimisation tools maybe used. Effective methods exists to solve convex optimisation problems of the following form

$$\min_{\mathbf{x}} \quad f(\mathbf{x})$$
$$s.t. \quad g(\mathbf{x}) \leq 0 \tag{1.4}$$

with $\mathbf{x} \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ being convex functions. The interested reader may refer to [28] for a detailed description of convex optimisation. If a feasible solution to a convex optimisation problem exists, the solution is a global optimum of the optimisation problem i.e. no better solution than the solution found exists. However the optimisation problem may contain a non-convex objective or constraints. Convex optimisation methods are not applicable to such non convex problems. However in some instances it is possible to reduce non convex problems to convex problems e.g. by appropriate change of variables.

Note that in a large scale network it is desirable to develop algorithms that solve a network utility optimisation problem in a distributed manner i.e. each network node (a node can be interpreted as a user or a cell) updates its solution locally and with some limited information exchange from the neighbours. In Chapters 2 and 3, we develop methods that assist solving optimisation problems in a distributed fashion. However in Chapter 5 constrained by the nature of the problem at hand, we propose centralised optimisation schemes for network optimisation and configuration.

1.3 Outline of the Thesis

The overview of the thesis is as follows.

- In Chapter 2 we consider capacity maximisation in an LTE network. We formulate adaptation of macro BS antenna tilt angle as a utility fair

optimisation task. We show that under reasonable conditions this problem can be reformulated as a convex optimisation. Using this insight, we develop a lightweight method for finding the optimal antenna tilt angles, making use of measurements which are already available at BSs, and suited to distributed implementation.

- In Chapter 3 we consider proportional fair rate allocation in a heterogeneous network with a mix of LTE and 802.11 cells which supports multipath and multihomed operation (simultaneous connection of a user device to multiple LTE BSs and 802.11 APs). We show that the utility fair optimisation problem is non-convex but that a global optimum can be found by solving a sequence of convex optimisations in a distributed fashion. The result is a principled approach to offload from LTE to 802.11 and for exploiting LTE/802.11 path diversity to meet user traffic demands.

- In Chapters 4 and 5 we consider dense small cell deployments, where energy efficiency and maintaining QoS are important. We emphasis the importance of the network traffic load dynamics for making informed decisions on small cells transmit power allocations.

  . In Chapter 4 we propose a method for dynamic idle mode selection in small cells which is based on classification of historical user measurements.

  . Using the periodic nature of the traffic load, we then develop a semi-static power adaptation approach in Chapter 5. That is, jointly scheduling of deep sleep and allocating pilot and data power in small cells with the aim of minimising energy consumption while maximising user utility within a mixed macro/small cell network. While the proposed dynamic approach quickly adapts to variations in traffic demand, the semi-static approach provides longer term (in order of hours) scheduling decisions. This semi-static power adaptation differs from the dynamic approach by exploiting the predictable nature of the traffic load. Hence, this approach enables deep sleep modes i.e. additional energy saving gains as well as avoiding additional signalling to control the idle mode of the SCBSs.

1.4 Contributions

This thesis contributes to enhancing resource allocation in different classes of next generation RANs. These contributions cover optimisation techniques in LTE networks, mixed LTE/small cells and 802.11 networks, and mixed LTE/small cell networks. The following publications were prepared during the PhD study:

1. Bahar Partov, Douglas Leith, "Utility Fair RAT Selection in Multi-Homed LTE/802.11 Networks". To appear on *IEEE 53rd Annual Allerton Conference on Communication, Control and Computing, Sep 2015.*

2. Bahar Partov, Douglas Leith, "Utility Fair Rate Allocation in LTE/802.11 Networks". Submitted to *IEEE Transactions/ACM on Networking, June 2015: http://arxiv.org/abs/1506.01058.*

3. Bahar Partov, Douglas Leith, Rouzbeh Razavi, and Holger Claussen, "Dynamic Idle Mode Control in Small Cell Networks". *IEEE International Conference on Communications, ICC, June 2015.*

4. Bahar Partov, Douglas Leith, and Rouzbeh Razavi, "Energy-Aware Configuration of Small Cell Networks". *IEEE 25th annual symposium on personal, indoor, and mobile radio communications, PIMRC, Sep 2014.*

5. Bahar Partov, Douglas Leith, and Rouzbeh Razavi, "Tilt Angle Adaptation in LTE Networks with Advanced Interference Mitigation". *IEEE 25th annual symposium on personal, indoor, and mobile radio communications, PIMRC, Sep 2014.*

6. Bahar Partov, Douglas Leith, and Rouzbeh Razavi, " Utility Fair Optimisation of Antenna Tilt Angles in LTE Networks". *IEEE Trans/ACM on Networking , 23(1), pp 175-185, DOI: 10.1109/TNET.2013.2294965.*

The problems that we have addressed in the context of this thesis are of notable importance. As a result there exist an extensive body of relevant work in regards with some of these problems. These related works are separately summarised in the individual chapters. Our motivations however remain to address each problem in a different perspective. Where we aim to provide methods that are easy to implement while they maintain efficiency.

For example the effect of antenna tilt angles in network performance has been studied in multiple works. However most of them address the problem by using heuristic methods or by using rather simplistic utility functions. While in Chapter 2 we aim to find a unifying solution that can be used in scale. We prove that the global optimums can be found. Moreover we introduce a distributed algorithm with low complexity that makes use of measurements that are already available at BSs.

Similar to adaptation of antenna tilt angles, radio access technology (RAT) selection in multihomed networks is studied in a number of related works. However the choices of the utility functions or the throughput models are rather simplistic. In Chapter 3 we consider a detailed model of LTE and 802.11 networks and a proportional fair objective. Moreover we show that a global optimum can be found in a distributed fashion and by applying a number of light-weight updates.

However in Chapters 4 and 5 we take different approaches compared to those of Chapters 2 and 3. This is where cost of reducing the complex problems into a simpler framework is high. Therefore we aim to apply non-convex solutions that still maintain users QoS and are easy to implement.

CHAPTER 2

Antenna Tilt Angle Adaptation in LTE Networks

In this Chapter we formulate adaptation of antenna tilt angle as a utility fair optimisation task. This optimisation problem is non-convex, but we show that under reasonable conditions it can be reformulated as a convex optimisation. Using this insight, we develop a lightweight method for finding the optimal antenna tilt angles, making use of measurements which are already available at base stations, and suited to distributed implementation.

2.1   Introduction

The antenna tilt angle of wireless base-stations is known to be a key factor in determining cell coverage and to play a significant role in interference management [29], [30]. While traditionally adjustment of tilt angle has largely been carried out manually, modern base stations increasingly allow automated adjustment. This creates the potential for more dynamic adaptation of tilt angle, for example to better match cell coverage to the distribution of user equipments and traffic, to reduce coverage holes created by failures in neighbouring stations, to better manage interference from the user deployment of femtocells, *etc.* The benefits of self configuration and self optimisation are already recognised in LTE release 9 [14], and automated adaptation of tilt angle in particular has been the subject of recent interest.

In this chapter we formulate adaptation of antenna tilt angle as a utility fair optimisation task. Namely, the objective is to jointly adjust antenna tilt angles within the cellular network so as to maximise network utility, subject to network constraints. Adjustments at base stations must be carried out jointly in a coordinated manner in order to manage interference. This optimisation

problem is non-convex, but in this chapter we show that under certain conditions it can be reformulated as a convex optimisation. Specifically, we show that (i) in the high SINR operating regime and with an appropriate choice of decision variables, the optimisation is convex for any concave utility function, and (ii) in any SINR regime the optimisation can be formulated in a convex manner when the objective is a proportional fair rate allocation. Since the optimisation is not well-suited to solution using standard dual methods, we develop a primal-dual method for finding the optimal antenna tilt angles. This approach is lightweight, making use of measurements which are already available at base stations, and suited to distributed implementation.

The rest of the chapter is organised as follows. In Section 2.2 we summarize the existing work in the area. In Section 2.3 we introduce our network model, which is based on 3GPP standard , and in Section 2.4 we analyse its convexity properties in the high SINR regime. In Section 2.5 we extend the analysis to general SINR regimes. In Section 2.6 we carry out a performance evaluation of a realistic setup and finally, in Section 2.8, we summarise our conclusions.

2.2 Related Work

The analysis and modelling of the impact of the antenna tilt angle on cell performance has been well studied, see for example [31, 32] and references therein. Recently, self-optimisation of tilt angle has started to attract attention, but most of this work makes use of heuristic approaches. In [33] a heuristic method is proposed for adjusting tilt to maximise average spectral efficiency within the network, while [34] proposes a combination of fuzzy and reinforcement learning. In [35] simulated annealing is considered for joint self-configuration of antenna tilt angle and power and in [36] a non-cooperative game approach between neighbouring base stations is studied. Offline planning of tilt angle is considered, for example, in [37], using a heuristic search method combined with a mixed integer local search. In the present chapter, we take a more formal, rigorous approach and show that tilt angle optimisation can, in fact, be formulated as a convex problem. Building on this result, we then introduce a lightweight distributed algorithm based on primal-dual subgradient updates and show that this algorithm is guaranteed to converge arbitrarily closely to the network optimum.

2.3 Network Model

2.3.1 Network Architecture

The network consists of a set $\mathcal{B}$ of base stations and a set $\mathcal{U}$ of User Equipment (UE), with UE $u \in \mathcal{U}$ receiving downlink traffic transmitted from base station $b(u) \in \mathcal{B}$. For base stations with sectoral antennas, we define a separate element in $\mathcal{B}$ for each antenna. We denote the (x, y) geographical co-ordinates of base station b by (x_b, y_b) and of user equipment u by (x_u, y_u). The distance between user u and base station b is therefore given by

$$d_{b,u} = \sqrt{(x_u - x_b)^2 + (y_u - y_b)^2} \tag{2.1}$$

2.3.2   Antenna Gain and Path Loss

The received power on sub-channel i from base station $b \in \mathcal{B}$ at user $u \in \mathcal{U}$ is given by $\tilde{G}_{b,u}(\theta_b)\ell_{b,u}p_b$, where $\tilde{G}_{b,u}(\theta_b)$ is the base station antenna gain, $\ell_{b,u}$ the path loss between b and u. Without loss of generality, we assume that transmit power of BS b for all sub-channels $i \in \mathcal{I}$ are equal and are denoted by p_b. For simplicity, shadowing and fast fading are not considered in the equations. We model path loss, as recommended in [38], by

$$\ell_{b,u} = \ell_0 d_{b,u}^{-\varrho} \tag{2.2}$$

with fixed path loss factor ℓ_0, path loss exponent ϱ and distance d in kilometres. For a given antenna type, the antenna gain $\tilde{G}_{b,u}(\theta_b)$ can be determined given the relative positions of b and u, the antenna tilt angle θ_b and the azimuth angle. With regard to the latter, changing the tilt and/or azimuth angles changes the direction of the antenna's main lobe, see Fig 2.1. We will assume that the azimuth angle is held fixed but allow the antenna tilt angle to be adjusted within the interval $[\underline{\theta}, \bar{\theta}]$. Following [38], the antenna gain can then be modelled by:

$$\tilde{G}_{b,u}(\theta_b) = \tilde{G}_0 \tilde{G}_v(\theta_b, d_{b,u}) \tag{2.3}$$

where $\tilde{G}_0$ is the maximum gain of the antenna,

$$\tilde{G}_v(\theta_b, d_{b,u}) = 10^{-1.2\left(\frac{\theta_{b,u} - \theta_b}{\theta_{3dB}}\right)^2} \tag{2.4}$$

is the antenna vertical attenuation, $\theta_{b,u} = \tan^{-1}(\hbar/d_{b,u})$, $\hbar$ is the height difference between the base station and UE (which, for simplicity, we assume is the same for all base stations and users) and θ_{3dB} the vertical half power beam width of the antenna. Figure 2.2 illustrates the ability of Eq (2.4) to accurately model the main lobe of an antenna which is popular in cellular networks.

It will prove useful to use the quantity $G_v(\theta_b, d_{b,u}) := \log \tilde{G}_v(\theta_b, d_{b,u})$. It will also prove useful to consider the following linear approximation $\hat{G}_v(\theta_b, d_{b,u})$ to antenna gain exponent $G_v(\theta_b, d_{b,u})$ about tilt angle θ_0,

$$\hat{G}_v(\theta_b, d_{b,u}) = \frac{-1.2 \log 10}{\theta_{3dB}^2}\left((\theta_{b,u} - \theta_0)^2 + 2(\theta_{b,u} - \theta_0)\theta_b\right) \tag{2.5}$$

This linear approximation is illustrated by the solid line in Fig 2.3. It is reasonably accurate provided adaptation of the antenna angle θ_b does not
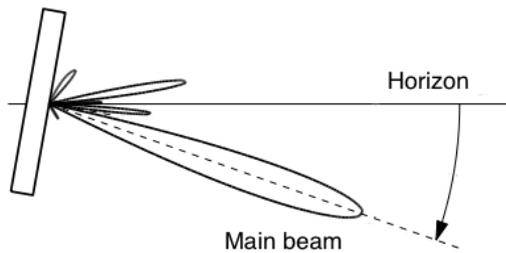
Figure 2.1:   Schematic illustrating relationship between antenna main lobe and tilt angle.
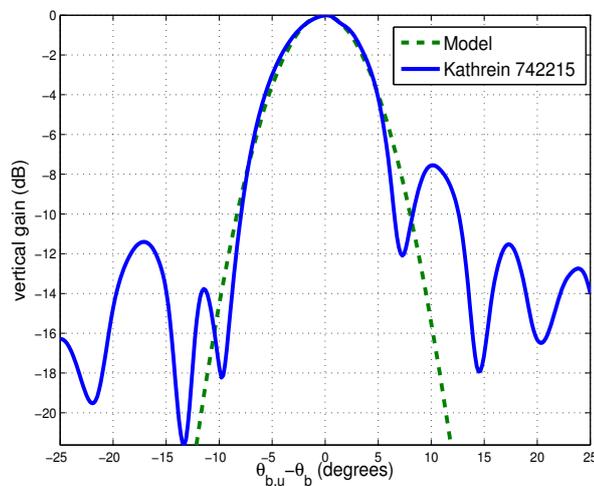
Figure 2.2:   Comparison of antenna main lobe vertical gain model (2.4) (dashed line) and measured antenna gain (solid line) for a Kathrein 742215 antenna, $\theta_{3dB} = 9°$.

cause $\theta_{b,u} - \theta_b$ to change sign (in which case the side of the main antenna lobe facing the user changes and so the slope of the linear approximation changes sign). This is assumed to be the case for the antennas of base stations other than that to which the UE is associated, which is only a mild assumption since otherwise interference from these base stations can be expected to be excessive.

2.3.3   User Throughput

The downlink throughput of the user equipment $u \in \mathcal{U}$ associated with base station $b(u)$ is given by:

$$R_u(\Theta) = \min\{\bar{r}, r_u(\Theta)\}, \quad u \in \mathcal{U} \tag{2.6}$$

15
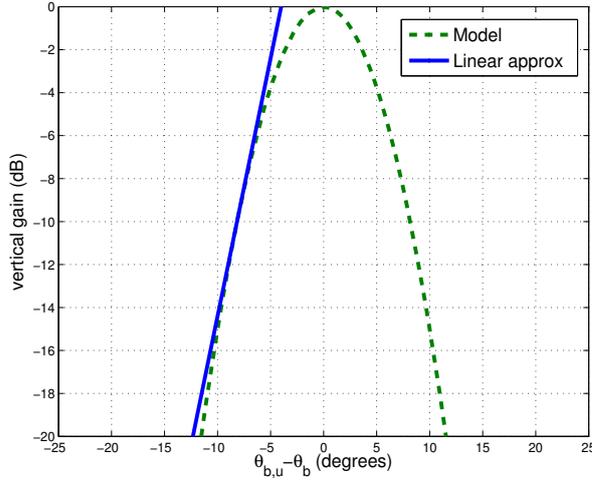
Figure 2.3:   Illustrating linear approximation $\hat{G}_v(\theta_b, d_{b,u})$ to $G_v(\theta_b, d_{b,u})$, $\theta_{3dB} = 9°$.

where Θ is the vector $[\theta_b]$, $b \in \mathcal{B}$ of tilt angles, $\bar{r}$ is the maximum achievable throughput (limited by the available modulation and coding schemes), and

$$r_u(\Theta) = \frac{\beta_1 \omega}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \log(1 + \gamma_u^i(\Theta)/\beta_2) \tag{2.7}$$

Here $\mathcal{I}$ denotes the set of available sub-channels, ω total available bandwidth. Note that the available bandwidth, ω is equally shared among all sub-channels i.e. $\omega^i = \frac{\omega}{|\mathcal{I}|}$ with ω^i denoting bandwidth of sub-channel $i \in \mathcal{I}$. Here parameters β_1 and β_2 reflect the LTE bandwidth and SINR implementation efficiencies [39]. Finally, $\gamma_u^i(\Theta)$ denotes the SINR on sub-channel i for user u,

$$\gamma_u^i(\Theta) = \frac{P_{R,u}(\theta_{b(u)})}{\sum_{c \in \mathcal{B} \setminus \{b(u)\}} \hat{P}_{R,u}(c) + \sigma_n^2} \tag{2.8}$$

where $P_{R,u}(\theta_b) := e^{G_{b,u}(\theta_b)} \ell_{b,u} p_b$ is the received power from base station $b(u)$ by user u, $\hat{P}_{R,u}(\theta_c) := e^{\hat{G}_{c,u}(\theta_c)} l_{c,u} p_c$ is the received power from base station $c \neq b(u)$ by user u and σ_n^2 noise power at the receiver. Observe that in $\hat{P}_{R,u}(\theta_c)$ we make use of linear approximation $\hat{G}_v$.

2.4   High SINR Regime

In the high SINR regime, the downlink throughput (2.6) can be accurately approximated by

$$\hat{R}_u(\Theta) = \min\{\bar{r}, \hat{r}_u(\Theta)\}, \quad u \in \mathcal{U}. \tag{2.9}$$

where

$$\hat{r}_u(\Theta) = \frac{\beta_1 \omega}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \log(\gamma_u^i(\Theta)/\beta_2) \tag{2.10}$$

2.4.1 Utility Fair Optimisation of Tilt Angle

Under the assumption of high SINR operation, we can formulate the selection of antenna tilt angles as the following optimisation problem $(P1)$:

$$\max_{\Theta} \sum_{u \in \mathcal{U}} U(\hat{R}_u(\Theta)) \tag{2.11}$$

$$s.t. \quad \underline{\theta} \le \theta_b \le \bar{\theta}, \quad b \in \mathcal{B} \tag{2.12}$$

$$\underline{r} \le \hat{R}_u(\Theta), \quad u \in \mathcal{U} \tag{2.13}$$

where $U(\cdot)$ is a concave increasing utility function and $\underline{r} \ge 0$. Constraint (2.12) captures restrictions on the range of feasible antenna tilt angles, while (2.13) ensures that each user receives a specified minimum throughput (which is expected to mainly be important for users at the edge of a cell who might otherwise be assigned too low a throughput).

2.4.2 Convexity Properties

Lemma 2.4.1. $\log P_{R,u}(\theta_b)$ *is strictly concave in* θ_b.

Proof: We have

$$\log P_{R,u}(\theta_b) = G_{b,u}(\theta_b) + \log \ell_{b,u} + \log p_b$$
$$= -1.2 \frac{\log 10}{\theta_{3dB}^2} (\theta_{b,u} - \theta_b)^2 + \log \ell_{b,u} + \log p_b$$

Now $\log \ell_{b,u}$, $\log p_b$, θ_{3dB} and $\theta_{b,u}$ are constants, so we only need to consider concavity with respect to θ_b. It can be verified that $\partial_{\theta_b}^2 \log P_{R,u}(\theta_b) = -2.4 \frac{\log 10}{\theta_{3dB}^2} < 0$. $\square$

Lemma 2.4.2. $\log \left(\sum_{c \in \mathcal{B} \backslash \{b(u)\}} \hat{P}_{R,u}(\theta_c) + \sigma_n^2 \right)$ *is convex.*

Proof: Rewrite as $\log \left(\sum_{c \in \mathcal{B} \backslash \{b(u)\}} e^{\hat{G}_{c,u}(\theta_c) + \log l_{c,u}p_c} + \sigma_n^2 \right)$. This can be expressed as $\log \left(\sum_{c \in \mathcal{B} \backslash \{b(u)\}} e^{h_{c,u}(\theta_c)} + \sigma_n^2 \right)$ with $h_{c,u}(\theta_c) := \hat{G}_{c,u}(\theta_c) + \log l_{c,u}p_c$. By [28, p74], the log of a sum of exponentials is convex. The additive term σ_n^2 acts as a translation and by [28, p79] convexity is preserved under translation. By approximation (2.5), the function $h_{c,u}(\theta_c)$ is affine in θ_c, and by [28, p79] when composed with the log of a sum of exponentials the resulting function remains convex. $\square$

It follows from Lemmas 2.4.1 and 2.4.2 that

Theorem 2.4.3. $\hat{R}_u(\Theta)$ *is concave in* Θ.

Proof: Recalling from (2.8) and (2.10):

$$\hat{r}_u(\Theta) = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \left(-\log \beta_2 + \log P_{R,u}(\theta_{b(u)}) - \log \Big(\sum_{c \in \mathcal{B} \backslash \{b(u)\}} \hat{P}_{R,u}(c) + \sigma_n^2 \Big) \right)$$

(2.14)

By Lemmas 2.4.1 and 2.4.2, $\hat{r}_u(\Theta)$ is concave in Θ. Since the min function is concave and non-decreasing, it follows that $\hat{R}_u(\Theta) := \min\{\hat{r}_u(\Theta), \bar{r}\}$ is also concave in Θ. $\square$

Note that $\hat{R}_u(\Theta)$ is not strictly concave in Θ since $\log P_{R,u}(\theta_{b(u)})$ is only strictly concave in $\theta_b(u)$ but not in the other elements of Θ. Nevertheless, under mild conditions $\sum_{u \in \mathcal{U}} \hat{R}_u(\Theta)$ is strictly concave in Θ:

Theorem 2.4.4. *Suppose $\mathcal{B} = \cup_{u \in \mathcal{U}} b(u)$, i.e. every base station $b \in \mathcal{B}$ has at least one associated UE $u \in \mathcal{U}$. Then $\sum_{u \in \mathcal{U}} \hat{r}_u(\Theta)$ is strictly concave in Θ (and so the solution to problem $(P1)$ is unique).*

Proof: We have

$$\sum_{u \in \mathcal{U}} \hat{r}_u(\Theta) = \frac{\beta_1 \omega}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \left(-\sum_{u \in \mathcal{U}} \log \beta_2 + \sum_{u \in \mathcal{U}} \log P_{R,u}(\theta_{b(u)}) \right.$$
$$\left. -\sum_{u \in \mathcal{U}} \log\Big(\sum_{c \in \mathcal{B} \backslash \{b(u)\}} \hat{P}_{R,u}(c) + \sigma_n^2 \Big) \right)$$

Recall $P_{R,u}(\theta_{b(u)})$ is strictly concave in $\theta_{b(u)}$ (by Lemma 2.4.1). The sum $\sum_{u \in \mathcal{U}} \log P_{R,u}(\theta_{b(u)})$ is therefore strictly concave in every θ_b, $b \in \cup_{u \in \mathcal{U}} b(u) = \mathcal{B}$. It is therefore strictly concave in Θ. In more detail, for any Θ^1, Θ^2 and $\alpha \in [0, 1]$ we have

$$\sum_{u \in \mathcal{U}} \log P_{R,u}(\alpha \theta_{b(u)}^1 + (1-\alpha)\theta_{b(u)}^2) > \sum_{u \in \mathcal{U}} \Big(\alpha \log P_{R,u}(\theta_{b(u)}^1) + (1-\alpha) \log P_{R,u}(\theta_{b(u)}^2) \Big)$$
$$= \alpha \sum_{u \in \mathcal{U}} \log P_{R,u}(\theta_{b(u)}^1) + (1-\alpha) \sum_{u \in \mathcal{U}} \log P_{R,u}(\theta_{b(u)}^2)$$

The result then follows from the fact that the sum of a strictly concave function and a concave function is strictly concave. $\square$

And we have the following corollary,

Corollary 2.4.5. *When each base station has at least one user with throughput less that $\bar{r}$, then $\sum_{u \in \mathcal{U}} \hat{R}_u(\Theta)$ is strictly concave in Θ.*

Proof: Let $\bar{\mathcal{U}} \subset \mathcal{U}$ denote the set of users with throughput less than $\bar{r}$. When each base station has at least one user with throughput less that $\bar{r}$ then $\mathcal{B} = \cup_{u \in \bar{\mathcal{U}}} b(u)$. Now $\sum_{u \in \bar{\mathcal{U}}} \hat{R}_u(\Theta) = \sum_{u \in \bar{\mathcal{U}}} \hat{r}_u(\Theta)$ is strictly concave in Θ by Theorem 2.4.4. It then follows immediately that $\sum_{u \in \mathcal{U}} \hat{R}_u(\Theta) = \sum_{u \in \bar{\mathcal{U}}} \hat{R}_u(\Theta) + \sum_{u \in \mathcal{U} \backslash \bar{\mathcal{U}}} \bar{r}$ is strictly concave in Θ. $\square$

2.4.3 Convex Optimisation

The objective function in optimisation problem (P1) is concave in Θ (since $U(\cdot)$ is concave increasing and $\hat{R}_u(\Theta)$ is concave by Theorem 2.4.3, then $U(\hat{R}_u(\Theta))$ is concave) and constraints (2.12)-(2.13) are linear (and so convex). Hence, optimisation problem (P1) is convex. It follows immediately that a solution exists. The Slater condition is satisfied and so strong duality holds.

2.4.4 Difficulty of Using Conventional Dual Algorithms

The Lagrangian is

$$L(\Theta, \Lambda) = -\sum_{u \in \mathcal{U}} U(\hat{R}_u(\Theta)) + \sum_{u \in \mathcal{U}} \lambda_u^1 (\underline{r} - \hat{R}_u(\Theta)) + \sum_{b \in \mathcal{B}} \lambda_b^2 (\underline{\theta} - \theta_b)$$
$$+ \sum_{b \in \mathcal{B}} \lambda_b^3 (\theta_b - \bar{\theta}) \tag{2.15}$$

where Λ denotes the set of multipliers λ_u^1, λ_b^2, λ_b^3, $u \in \mathcal{U}$, $b \in \mathcal{B}$. The dual function is $g(\Lambda) := L(\Theta^*(\Lambda), \Lambda)$, where $\Theta^*(\Lambda) = \arg\max_\Theta L(\Theta, \Lambda)$. The main KKT conditions are $\partial_{\theta_b} L(\Theta, \Lambda) = 0$, $b \in \mathcal{B}$. That is,

$$\sum_{u \in \mathcal{U}} \left(1 + \lambda_u^1\right) \partial_{\theta_b} U(\hat{R}_u(\Theta)) = \lambda_b^3 - \lambda_b^2, \quad b \in \mathcal{B} \tag{2.16}$$

Given Λ, we can use Eq (2.16) to find $\Theta^*(\Lambda)$. The optimal vector of multipliers is $\Lambda^* = \arg\max_{\Lambda \geq 0} g(\Lambda)$. Since $g(\Lambda)$ is concave, a standard dual function approach is to find Λ^* using subgradient ascent techniques, and then find the optimal tilt angle $\Theta^*(\Lambda^*)$. However, solving Eq (2.16) to obtain the primal variables is tricky in general since it imposes complex, implicit dual constraints for a solution to exist. Consequently, the dual subgradient approach is unattractive for solving problem (P1).

2.4.5 Distributed Algorithm for Finding Optimal Solution

We consider the following primal-dual algorithm:

where in Algorithm 2.1 projection $[z]^+$ equals z when $z \geq 0$ and 0 otherwise,

$$\partial_{\theta_b} L(\Theta, \Lambda) = -\sum_{u \in \mathcal{U}} \left(1 + \lambda_u^1\right) \partial_{\theta_b} U(\hat{R}_u(\Theta)) - \lambda_b^2 + \lambda_b^3 \tag{2.20}$$

$$\partial_{\lambda_u^1} L(\Theta, \Lambda) = \underline{r} - \hat{R}_u(\Theta) \tag{2.21}$$

$$\partial_{\lambda_b^2} L(\Theta, \Lambda) = \underline{\theta} - \theta_b \tag{2.22}$$

$$\partial_{\lambda_b^3} L(\Theta, \Lambda) = \theta_b - \bar{\theta} \tag{2.23}$$

and $\partial_{\theta_b} U(\hat{R}_u(\Theta))$ denotes any subgradient of $U(\hat{R}_u(\Theta))$ with respect to θ_b.

Algorithm 2.1 High SINR

Initialise: $t = 0$, $\Theta(0)$, $\Lambda(0)$, step size $\alpha > 0$

do

$$\theta_b(t+1) = \theta_b(t) - \alpha \partial_{\theta_b} L(\Theta(t), \Lambda(t)), \qquad \theta_b \in \mathcal{B} \tag{2.17}$$

$$\lambda_u^1(t+1) = \left[\lambda_u^1(t) + \alpha \partial_{\lambda_u^1} L(\Theta(t), \Lambda(t)) \right]^+, \; u \in \mathcal{U} \tag{2.18}$$

$$\lambda_b^i(t+1) = \left[\lambda_b^i(t) + \alpha \partial_{\lambda_b^i} L(\Theta(t), \Lambda(t)) \right]^+, \; b \in \mathcal{B}, i = 2, 3 \tag{2.19}$$

$$t \leftarrow t + 1$$

loop

Observe that each iteration (2.17)-(2.19) of Algorithm 2.1 simultaneously updates both the primal variable Θ and the multipliers λ_u^1, λ_b^2, λ_b^3. It possesses the following convergence property:

Lemma 2.4.6. *For Algorithm 2.1 suppose* $(\Theta(t), \Lambda(t))$ *is bounded for all t. Then there exists constant $M \geq 0$ such that*

$$0 \leq \frac{1}{t} \sum_{\tau=0}^{t} \left(L(\Theta(\tau), \Lambda^*) - L(\Theta^*, \Lambda(\tau)) \right) \leq \frac{\Delta}{2\alpha t} + \frac{\alpha M}{2}$$

where (Θ^*, Λ^*) is a solution to optimisation problem (P1), $\Delta = ||\Theta(0) - \Theta^*)||_2^2 + ||\Lambda(0) - \Lambda^*||_2^2$ and $|| \cdot ||_2$ denotes the usual Euclidean norm.

Proof: Optimisation problem $(P1)$ is convex, the objective and constraint functions are differentiable and the Slater condition is satisfied. The result now follows by direct application of Lemma A.3.2 in the Appendix. $\square$

Since $\Delta/(2\alpha t) \to 0$ as $t \to \infty$, Lemma 2.4.6 tells us that update (2.17)-(2.19) converges to a ball around an optimum (Θ^*, Λ^*), the size of the ball decreasing with step size α. The size of the ball is measured in terms of metric $L(\Theta, \Lambda^*) - L(\Theta^*, \Lambda)$, and recall that by complementary slackness $L(\Theta^*, \Lambda^*) = \sum_{u \in \mathcal{U}} U(\hat{R}_u(\Theta^*))$.

2.4.6   Message Passing and Implementation

Algorithm 2.1 can be implemented in a distributed manner. Namely, each base station $b \in \mathcal{B}$ carries out local tilt angle updates according to Eqs (2.17) and (2.19), and also carries out update (2.18) for each user u associated with base station b. For this, each base station b needs to evaluate Eqs (2.20)-(2.23). Evidently Eqs (2.21)-(2.23) can be evaluated using locally available information (the tilt angle of base station b and the current downlink throughput of user u associated with base station b). In contrast, evaluating (2.20) requires information sharing between base stations. Specifically, it is necessary

to evaluate

$$\sum_{u \in \mathcal{U}} \frac{\partial \hat{r}_u}{\partial \theta_b} = \sum_{u \in \{u \in \mathcal{U}: b(u)=b\}} \frac{\partial \hat{r}_u}{\partial \theta_b} + \sum_{u \in \{u \in \mathcal{U}: b(u) \neq b\}} \frac{\partial \hat{r}_u}{\partial \theta_b} \qquad (2.24)$$

The first term in (2.24) is the sensitivity of the throughput of users associated to base station b to changes in its tilt angle θ_b. This can either be directly measured by base station b (by perturbing the tilt angle), or calculated using

$$\frac{\partial \hat{r}_u}{\partial \theta_b} = \frac{\beta_1 \omega}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \frac{\partial G_v(\theta_b, d_{b,u})}{\partial \theta_b} \qquad (2.25)$$

where

$$\frac{\partial G_v(\theta_b, d_{b,u})}{\partial \theta_b} = \frac{2.4 \log 10}{\theta_{3dB^2}} (\theta_{b(u),u} - \theta_b) \qquad (2.26)$$

This calculation requires knowledge of the pointing angle $\theta_{b(u),u}$ between base station b and user u. This pointing angle can be determined from knowledge of the location of users, information which is usually available to modern base stations since location based services (LBS) are of high importance for mobile network providers. For example, in the US carriers are required by FCC to provide location-based information of the mobile users for E911 services and to within a specified accuracy [40]. Within Release 9 of 3GPP a set of enhanced positioning methods are standardized for LTE[41].

The second term in (2.24) is the sensitivity of the throughput of users associated to base stations other than b to changes in tilt angle θ_b. This can be calculated as

$$-\frac{1}{|\mathcal{I}|} \sum_{u \in \{u \in \mathcal{U}: b(u) \neq b\}} \sum_{i \in \mathcal{I}} \gamma_u^i(\Theta) \frac{\partial \hat{G}_v(\theta_b, d_b)}{\partial \theta_b} \frac{\hat{P}_{R,u}(\theta_b)}{P_{R,u}(\theta_{b(u)})} \qquad (2.27)$$

This requires user received power $\hat{P}_{R,u}(\theta_b)$ from base station b, user received power $P_{R,u}(\theta_{b(u)})$ from the base station to which it is associated, the user SINR γ_u^i and the pointing angle $\theta_{b(u),u}$. All of this information is available to the base station to which the user is associated (via user equipment received power and SINR reports), but not to neighbouring base stations and so must be communicated to them.

We note that antenna tilt angle updates are likely to occur on a relatively long time-scale in practice. Capturing hourly based traffic patterns of the mobile users may therefore also provide relatively reliable traffic distribution information, which might also be used.

2.4.7 Example

We illustrate the application of the foregoing high SINR analysis to the scenario shown in Fig 2.4(a). We use a simple scenario here to help gain insight,

with a more realistic setup considered in detail in Section 2.6. The scenario consists of regularly spaced base stations each with three sector antennas. The base station radio parameters are detailed in Table 3.2 based on 3GPP standard [38]. The users are primarily located in two clusters, as indicated in Fig 2.4(a). One cluster of 16 users is associated with the first sector of base station 1, and the other cluster of 16 users with the third sector of base station 2. Clustering of users creates a challenging tilt angle assignment task since a poor choice of tilt angles will have a strong effect on network performance. Additionally, two users are located close to the mid-point between these base stations. Ensuring adequate coverage at cell edges is commonly an issue for network operators and so we expect a performance tradeoff between serving these edge users and serving users located in the clusters. For concreteness, we select utility function $U(z) = z$, so that optimisation problem (P1) corresponds to maximising the network sum-throughput, subject to every user obtaining a minimum throughput of $64kps$ and to physical constraints that the allowable tilt angles must lie in the interval $[5, 20]$ degrees.

Table 2.1: simulation parameters

Parameter	Value
BS and UE	
Inter-site distance	$500m$
Number of sectors	3
Number of sub-channels $|\mathcal{I}|$	1
BS Antenna max gain $\tilde{G}_0$	$15dBi$
UE Antenna Gain/Type	$0dBi$/Omni
BS Antenna height	$25m$
UE Antenna height	$1.5m$
Vertical half power beamwidth θ_{3dB}	$10°$
Transmit power p_b	$46dBm$
Channel	
Exponential path loss factor ϱ	3.76
Fixed path loss factor ℓ_0	$10^{2.8}$
Bandwidth ω	$10MHz$
Carrier Frequency	$2GHz$
Bandwidth efficiency β_1	1
SINR efficiency β_2	1
UE noise power σ_n^2	$-94.97dBm$
Optimisation	
Min tilt angle $\underline{\theta}$	$5°$
Max tilt angle $\bar{\theta}$	$20°$
Min throughput $\underline{r}$	$64kbps$
Max throughput $\bar{r}$	$10Mbps$
Step size α	0.05
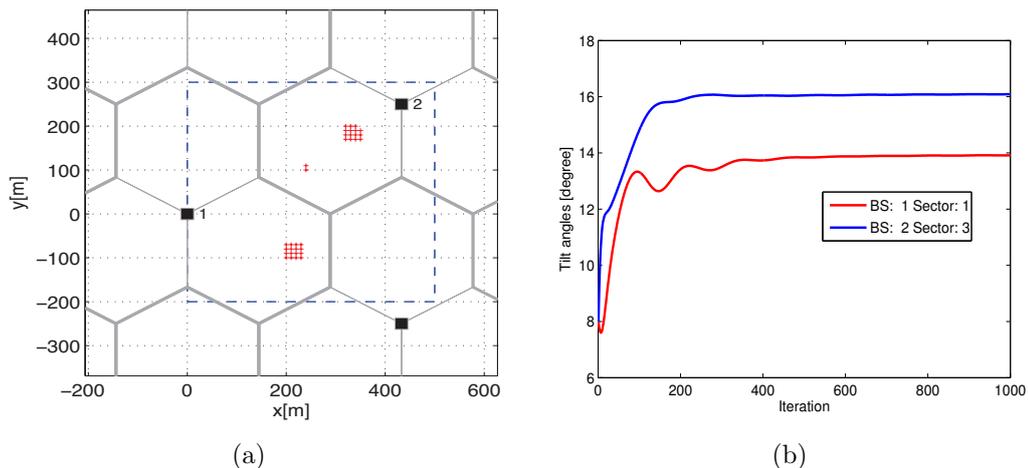
Figure 2.4: (a) Example network topology. Base-stations are indicated by solid squares labelled 1,2 and 3, UEs by dots. (b) Tilt angle vs iteration number with Algorithm 2.1.

Figure 2.4(b) shows tilt angle time histories for the two base stations when using Algorithm 2.1. It can be seen that the tilt angles converge to the optimum in less than 600 iterations. Figure 2.5(a) shows the corresponding network sum-throughput vs time. Also shown in the network sum-throughput for fixed antenna angles of 8°. Optimising the tilt angles increases the network sum throughput by almost factor of 18 compared to the use of fixed angles. As already noted, the improvement is expected to be particularly pronounced in this simple example since the users are grouped into clusters, and so angling the antennas to point towards their respective clusters both greatly increases received power and decreases interference. We note that significant performance gains are, however, also observed in the more realistic scenario studied in Section 2.6 and this reflects the fundamental importance of antenna tilt angle to network performance. Fig 2.5(b) illustrates the impact of the minimum throughput constraint $\underline{r}$ on network sum-throughput. It can be seen that as $\underline{r}$ is increased from zero to 2Mbps, the network sum-throughput decreases but that the impact is minor. Note that as $\underline{r}$ is increased beyond 2Mbps the optimisation becomes infeasible as the stations at the cell edge are unable to support such high rates.

As discussed in Section 2.4.6, UE location information is used when calculating Eq (2.20) in Algorithm 2.1. In practice this location information will be approximate in nature. Figure 2.6 plots the optimised network sum-throughput vs the standard deviation of the location error when zero-mean gaussian noise is added to the true user locations. It can be seen that, as might be expected, the optimised sum-throughput falls as the noise level is increased. However, the decrease is small (less than 5%) even for relatively large location errors.

(a)   (b)
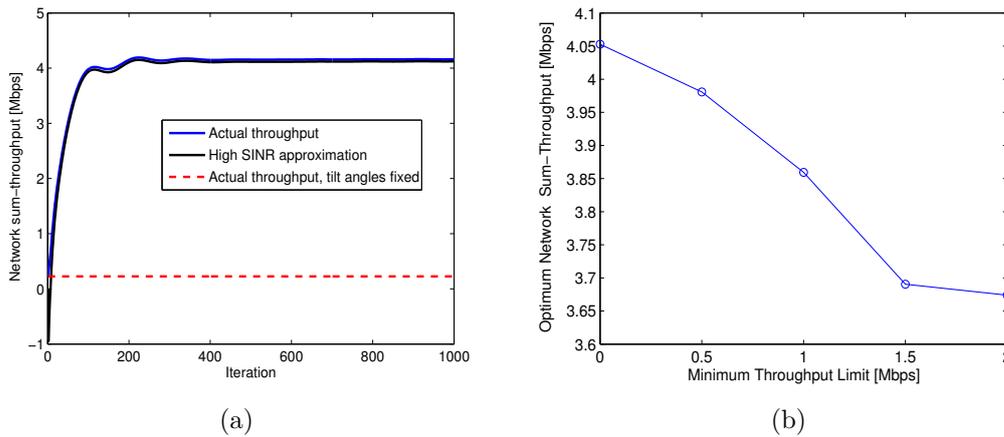
Figure 2.5: (a) Normalised network sum-throughput vs iteration number with Algorithm 2.1. Here, the network sum-throughput is normalised by dividing by the number of users in the network. (b) Normalised network sum-throughput as minimum throughput constraint r is varied. The network sum-throughput is normalised by dividing by the number of users in the network.
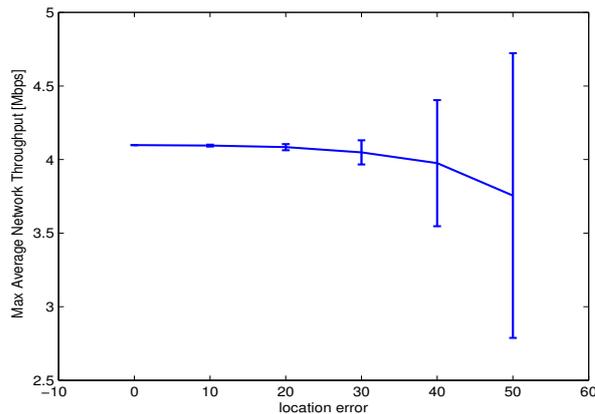
Figure 2.6:   Optimised network sum-throughput vs. magnitude of location error, error bars indicate the standard deviation for 100 runs of simulation for each point. The network sum-throughput is normalised by dividing by the number of users in the network.

2.5   Any SINR: Proportional Fair Rate Allocation

In this section we relax the assumption of operation in the high SINR regime. However, this comes at the cost of restricting attention to proportional fair rate allocations. We consider the following utility fair optimisation problem

(P2):

$$\max_{\Theta} \sum_{u \in \mathcal{U}} \log R_u(\Theta) \tag{2.28}$$

$$s.t. \quad \underline{\theta} \leq \theta_b \leq \bar{\theta}, \quad b \in \mathcal{B} \tag{2.29}$$

$$\log \underline{r} \leq \log R_u(\Theta), \quad u \in \mathcal{U} \tag{2.30}$$

where $R_u(\Theta)$ is given by (2.6).

2.5.1 Convexity Properties

We recall the following,

Lemma 2.5.1 ([42]). $h(x) = \log(\log(1 + e^x))$ *is concave and non-decreasing in* $x \in \mathcal{R}$.

Turning now to $R_u(\Theta)$, we begin by observing that

Lemma 2.5.2. $\log(r_u(\Theta))$ *is concave in* Θ.

Proof: From (2.7) we have

$$\log(r_u(\Theta)) = \frac{\beta_1 \omega}{|\mathcal{I}|} \sum_{i \in |\mathcal{I}|} \log(\log(1 + \gamma_u^i(\Theta)/\beta_2) = \frac{\beta_1 \omega}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \log(\log(1 + e^{\hat{r}_u^i(\Theta)})) \tag{2.31}$$

where $\hat{r}_u^i(\Theta) = \log(\gamma_u^i(\Theta)/\beta_2)$. That is, the mapping from vector Θ to $\log(r(\Theta))$ is the vector composition of $h(x)$ in Lemma 2.5.2 and $\hat{r}_u^i(\Theta)$. By Lemmas 2.4.1 and 2.4.2, $\hat{r}_u^i(\Theta)$ is concave in Θ. By [28, p86], the vector composition of a non-decreasing concave function and a concave function is concave. $\square$

and

Theorem 2.5.3. $\log R_u(\Theta)$ *is concave in* Θ.

Proof: *From (2.6) we have*

$$\log R_u(\Theta) = \log \min\{\bar{r}, r_u(\Theta)\} \stackrel{(a)}{=} \min\{\log \bar{r}, \log r_u(\Theta)\} \tag{2.32}$$

where (a) follows from the fact that the log *function is monotonically increasing. By Lemma 2.5.2,* $\log r_u(\Theta)$ *is concave. Since the* min *function is concave non-decreasing, when composed with* $\log r_u(\Theta)$ *it is concave i.e.* $\log \hat{R}_u(\Theta)$ *is concave in* Θ. $\square$

2.5.2 Convex Optimisation

It follows from Theorem 2.5.3 that the objective of optimisation problem (P2) is concave. Constraints (2.29) are linear (so convex). The RHS of constraint (2.30) is concave, again by Theorem 2.5.3, and so this constraint is convex. It follows that optimisation problem (P2) is convex and a solution exists.

2.5.3 Distributed Algorithm

The Slater condition is satisfied and strong duality holds. We can therefore apply a similar approach as in Section 2.4.5 to develop a distributed algorithm for finding the optimal antenna tilt angles.

The Lagrangian is:

$$L(\Theta, \Lambda) = -\sum_{u \in \mathcal{U}} \log R_u(\Theta) + \sum_{u \in \mathcal{U}} \lambda_u^1 (\log \underline{r} - \log R_u(\Theta)) + \sum_{b \in \mathcal{B}} \lambda_b^2 (\underline{\theta} - \theta_b)$$
$$+ \sum_{b \in \mathcal{B}} \lambda_b^3 (\theta_b - \bar{\theta}) \tag{2.33}$$

We can now apply Algorithm 2.1 to solve (P2) provided we use the appropriate gradients:

$$\partial_{\theta_b} L(\Theta, \Lambda) = -\sum_{u \in \mathcal{U}} \left(1 + \lambda_u^1\right) \partial_{\theta_b} (\log R_u(\Theta)) - \lambda_b^2 + \lambda_b^3 \tag{2.34}$$

$$\partial_{\lambda_u^1} L(\Theta, \Lambda) = \log \underline{r} - \log R_u(\Theta) \tag{2.35}$$

$$\partial_{\lambda_b^2} L(\Theta, \Lambda) = \underline{\theta} - \theta_b \tag{2.36}$$

$$\partial_{\lambda_b^3} L(\Theta, \Lambda) = \theta_b - \bar{\theta} \tag{2.37}$$

with

$$\frac{\partial r_u}{\partial \theta_b} = \begin{cases} \frac{\partial G_v(\theta_b, d_{b,u})}{\partial \theta_b} \frac{P_{R,u}(\theta_{b(u)})}{P_{R,u}(\theta_{b(u)}) + \beta_2 I_{u \in b(u)}} & b = b(u) \\ \frac{\partial \hat{G}_v(\theta_b, d_{b,u})}{\partial \theta_b} \left(\frac{\hat{P}_{R,u}(\theta_{b(u)})}{P_{R,u}(\theta_c)/\beta_2 + I_{u \in \mathcal{B} \setminus \{b(u)\}}} - \frac{\hat{P}_{R,u}(\theta_{b(\acute{u})})}{I_{u \in \mathcal{B} \setminus \{b(u)\}}} \right) & b \neq b(u) \end{cases}$$

and

$$I_{u \in b(u)} = \sum_{c \in \mathcal{B} \setminus \{b(u)\}} \hat{P}_{R,u}(\theta_c) + \sigma_n^2 \tag{2.38}$$

2.5.4 Example

We revisit the example in section 2.4.7. Figure 2.7 compares the results for optimisation problems $(P1)$ and $(P2)$. Figure 2.7(a) shows the sum-throughput from Fig 2.5(a) and when solving proportional fair allocation problem $(P2)$. As expected, the sum throughput is lower for the proportional fair allocation. Fig 2.7(b) compares the sum-log-throughput. As expected, the sum-log-throughput is higher for the proportional fair allocation problem $(P2)$. Fig 2.8 shows detail of the throughputs assigned to individual users to maximise sum-throughput and for proportional fairness. It can be seen that the throughput assignments are broadly similar in both cases, with the primary difference being the throughputs assigned to the two users located at the cell edge (numbered 33 and 34 in Fig 2.8). The proportional fair allocation assigns significantly higher rate to these edge stations than does the max sum-throughput allocation.
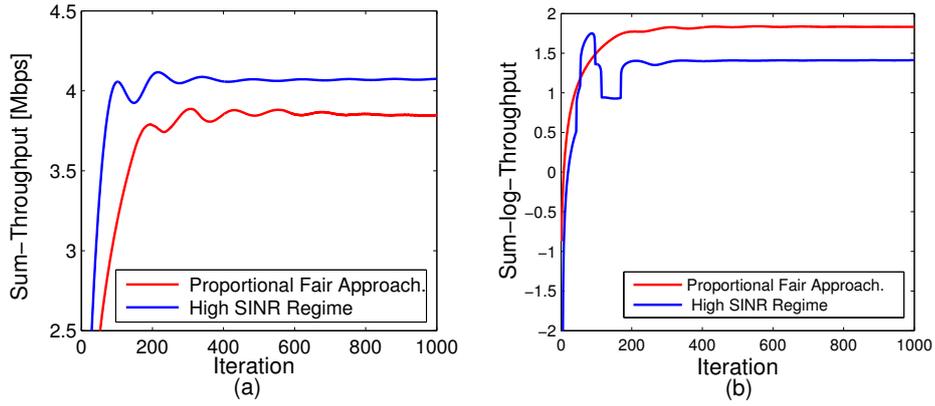
Figure 2.7: (a) Comparing normalised sum-throughput from Fig 2.5(a) and when solving proportional fair allocation problem ($P2$) – as expected, the sum throughput is lower for the proportional fair allocation. (b) Comparing normalised sum-log-throughput – as expected, the sum-log-throughput is higher for the proportional fair allocation. In all cases the network throughput is normalised by dividing by the number of users in the network.
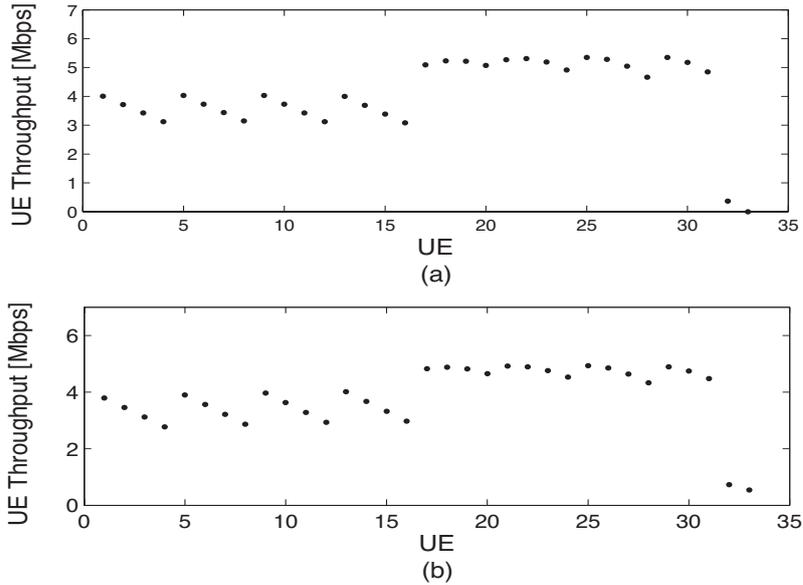
Figure 2.8: User throughput assignments for (a) sum-throughput maximisation in the high SINR regime and (b) proportional fair rate allocation.

2.6 Performance Evaluation

In this section we consider a realistic example based on data from the cellular network covering Grafton Street and Dawson Street in downtown Dublin, Ireland, see Fig 2.9(a). These are major shopping streets close to the centre

27

(a) Street map

(b) User and base-station locations.
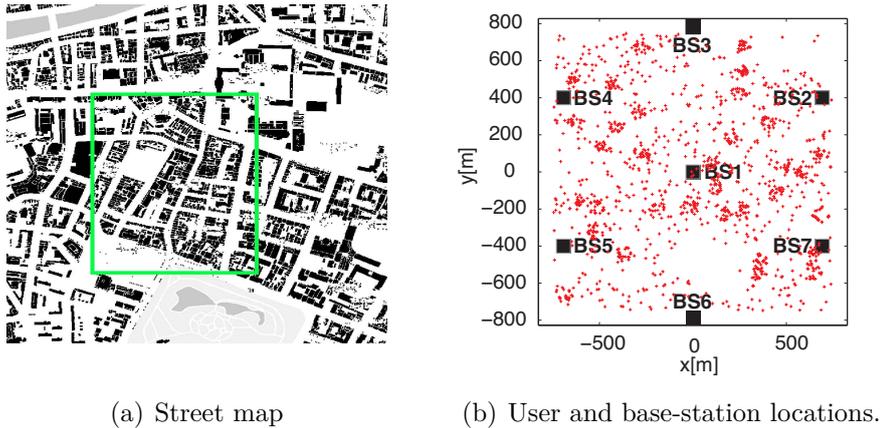
Figure 2.9: Dublin, Ireland example.

of Dublin city, with a large number of cellular users. We consider a section of the network with 21 sectors in a $1500m \times 1500m$ area and with an inter-site distance of $800m$. Environmental characteristics are derived from experimental measurement data with a combination of non-line of sight and line-of sight paths. Path loss and log-normal shadow fading parameters are derived from [38] for macro urban scenarios and detailed Table 2.2. There are 1350 users, with locations as shown in Fig 2.9(b). We focus on the performance experienced by the 388 users associated with the centre base station (indicated by BS1 in Fig 2.9(b)). Figures 2.10 shows the proportional fair rate allocation.

Table 2.2: Dublin scenario simulation parameters

Parameter	Value
Site and Sector	
Inter-site distance	$800m$
Channel	
NLOS exponential path loss factor	3.9
NLOS fixed path loss factor	$10^{2.1}$
LOS exponential path loss factor	2.2
LOS fixed path loss factor	$10^{3.4}$
Shadow fading standard deviation	6
Shadow fading mean	0
Optimisation	
Step size α	0.01

For comparison, results are also shown when a fixed tilt angle of 8° is used. It can be seen from Fig 2.10(a) that the sum-log-throughput objective function is improved by 22% by tilt angle optimisation, and that Algorithm 2.1 converges rapidly to the optimal allocation. From the cumulative distribution

fucntion (CDF) in Fig 2.10(b) it can be seen the user throughputs are also significantly increased, with the median throughput increased by almost a factor of 4 compared to use of fixed angles.

(a) Normalised Sum-log-throughput

(b) CDF of user throughputs.
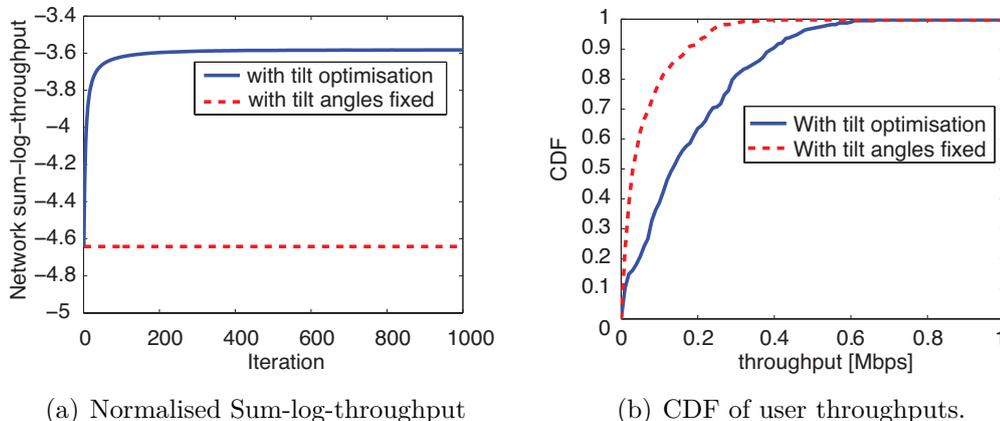
Figure 2.10: Proportional fair rate allocation, Dublin example. For comparison, results are also shown when a fixed tilt angle of 8° is used (indicated by dashed lines).

2.7 LTE SIMO Links and MMSE Post-Processing

In this section we extend the performance evaluation to consider LTE SIMO links with one transmit antenna on the BS and two receive antennas at the UE. The presence of two antennas at the receiver allows the UE to cancel one interferer. Hence, if interference is dominated by a single transmitter then we expect the use of SIMO links will allow inter-cell interference to be significantly reduced. Our interest here is in the impact that this interference cancellation has on the size of throughput gain achievable by tilt angle adjustment.

We consider a SIMO link with linear minimum mean square error (LMMSE) post processing applied to the received signal to mitigate neighbouring cell interference. We denote the channel vector for user u by $\mathbf{h}_{u,b(u)}$. Having user u associated to BS b, allows us to simplify notation by letting $\mathbf{h}_{u,b(u)} = \mathbf{h}_u$. Defining channel vector $\mathbf{h}_u = [h_1 \ h_2]^T$, the channel gain for user u is:

$$\mathbf{h}_u = \sqrt{\frac{P_{R,u}(\theta_b)}{p}} \sqrt{10^{S_{b,u}/10}} \ \boldsymbol{\Upsilon}_{b,u}. \tag{2.39}$$

where $\boldsymbol{S}$ is a zero mean Gaussian random variable representing slow fading effects, $\boldsymbol{\Upsilon}_{b,u}$ is a Rayleigh flat fading vector and p is the power of the transmitted signal assuming all base stations transmit at $p = p_b$. We can consider the elements of $\boldsymbol{\Upsilon}_{b,u}$ to be independent complex random Gaussian processes

corresponding to the channels of base station b and user u, provided that the antenna elements are sufficiently separated (typically on the order of half a wavelength apart). We identify the inter-cell interference vector $\mathbf{v}_u = [v_1 \ v_2]^T$ for user u by the strongest interferer:

$$\mathbf{v}_u = \sqrt{\max_{c \in \mathcal{B}\backslash\{b(u)\}}\{\hat{P}_{R,u}(\theta_c)10^{S_{c,u}/10}\}}\ \boldsymbol{\Upsilon}_{c_{max},u} \tag{2.40}$$

The remaining inter-cell interference is modelled as spatially white Gaussian noise [43], which comprises the noise vector $\boldsymbol{n}_u = \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$ where n_1 and n_2 are independent Gaussian variables:

$$N_0 = \mathbb{E}[n_1 n_1^H] = \mathbb{E}[n_2 n_2^H] = \sum_{\substack{c \in \mathcal{B}\backslash\{b(u)\} \\ c \neq c_{max}}} \hat{P}_{R,u}(\theta_c) + \sigma_n^2 \tag{2.41}$$

Hence the received signal ($\mathbf{y}$) is given by:

$$\mathbf{y}_u = \mathbf{h}_u \mathbf{x} + \mathbf{v}_u + \mathbf{n}_u \tag{2.42}$$

with $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = p$. The linear MMSE combining vector $\boldsymbol{\kappa}_u = [\kappa_1 \ \kappa_2]^T$, is given by:

$$\boldsymbol{\kappa}_u = \mathbf{h}_u^H(\mathbf{h}_u \mathbf{h}_u^H + \frac{\boldsymbol{\Phi}_u + N_0\mathbf{I}}{p})^{-1} \tag{2.43}$$

where Φ_u is the autocorrelation of interference vector $\mathbf{v}_u$:

$$\boldsymbol{\Phi}_u = \mathbb{E}[\mathbf{v}_u \mathbf{v}_u^H] = \begin{bmatrix} |v_1|^2 & v_1 v_2^H \\ v_2 v_1^H & |v_2|^2 \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \tag{2.44}$$

By applying the MMSE weights on the received signal, the post processing SINR is calculated as:

$$\gamma_{MMSE,u} = \frac{p|\kappa_1 h_1 + \kappa_2 h_2|^2}{|\kappa_1|^2\phi_{11} + |\kappa_2|^2\phi_{22} + 2Re\{\kappa_1\kappa_2^H\phi_{12}\} + N_0(|\kappa_1|^2 + |\kappa_2|^2)} \tag{2.45}$$

We average the post processing SINRs over the multipath fading samples. Using the averaged post processing SINRs, user throughputs with and without tilt optimisation can be calculated using Eq (2.6).

Figure 2.11 shows CDF of the user throughputs for SIMO links with MMSE detection, with and without flat fading. As expected, the use of MMSE detection yields significant improvements in the user throughputs. The throughput gains achieved by tilt optimisation can be compared for SISO links and for SIMO links with an optimal LMMSE detector by comparing Figs 2.10(b) and 2.11. The gain in the mean user throughput achieved by tilt optimisation is decreased from 83.07% to 67.42% when MMSE detection is employed.

(a) with fast fading (b) without fast fading
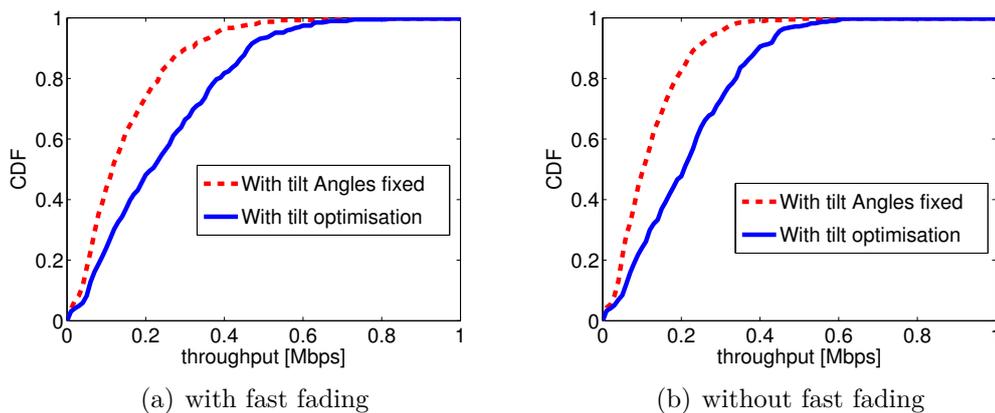
Figure 2.11: User throughput CDFs with SIMO links and MMSE detection, Dublin example. Fast fading is modelled by generating 300 samples using the 3GPP typical urban channel model, where the speed of the mobile user and carrier frequency are $3km/h$ and $2GHz$ respectively.

However, the gain in the log-sum-rate (which is the objective function of optimisation P2) only changes from 22.29% to 22.00%. That is, while MMSE detection enhances inter-cell interference mitigation, tilt optimisation can still yield significant improvements in network capacity.

We can investigate this behaviour in more detail as follows. Let

$$v_u = \frac{\max_{c \in \mathcal{B} \setminus \{b(u)\}} \{\hat{P}_{R,u}(\theta_c) 10^{S_{c,u}/10}\}}{\sum_{c \in \mathcal{B} \setminus \{b(u)\}} \hat{P}_{R,u}(\theta_c) 10^{S_{c,u}/10}} \tag{2.46}$$

be the ratio of the largest interferer to the total interference experienced by a user u. The CDF of v for the Dublin example is shown in Fig 2.12(a). It can be seen that approximately 40% of users have v values less than 0.5 i.e. for 40% of users the the strongest interferer power is less than the sum of the power of the other interferers. Figure 2.12(b) shows the corresponding spatial distribution of v. It can be seen that the strongest interferer is dominant at the edge of antenna sectors and along the nulls of the sector antennas. However, the intensity of the strongest interferer decreases along the edges of the base station coverage area and alongside the antennas. Table 2.3 details the throughput gains achieved by tilt angle optimisation for both SISO and SIMO links and for users with different v ratios. It can be seen that the throughput gain obtained by tilt angle optimisation for users with $v > 0.5$ is reduced when SIMO links are used. However, the gain is similar for both SISO and SIMO links for users with $v \leq 0.5$, once MMSE post processing is applied, and as noted above this consists of approximately 40% of users.

In summary, although the mean user throughput is improved for both fixed and optimal tilt angles for SIMO links with MMSE, tilt optimisation can

still yield considerable performance gains. Note that in a separate work [44], we investigate the role of tilt angle adjustment where SIMO receivers and adaptive OFDM/TDMA transmission scheduling may additionally be used to mitigate interference. And we find that even when SIMO/LMMSE reception and adaptive transmission scheduling are used to mitigate interference, tilt angle adjustment still offers the potential for significant performance gains, namely increases in mean user throughput of more than 65%.
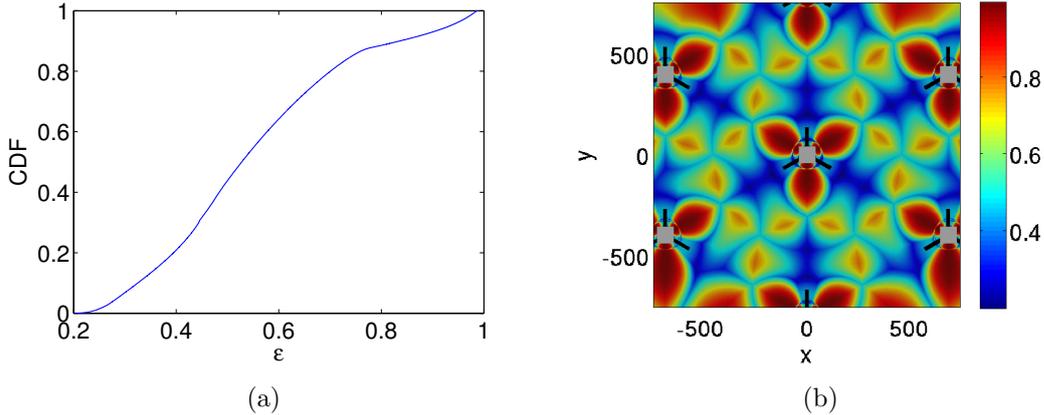
Figure 2.12: Contribution of the strongest interference to the total Interference: (a) cumulative distributed function of υ in the central BS coverage area. (b) Distribution of υ relative to the user positions.

Table 2.3: User throughput gains due to tilt angle optimisation for both SISO and SIMO links and vs υ, Dublin example.

	Mean Throughput Gain [%]	
	SISO	SIMO & MMSE Detection
$0 \leq \upsilon_u \leq 0.5$	54.5	52.3
$0.5 < \upsilon_u \leq 0.8$	128	85.99
$0.8 < \upsilon_u \leq 1$	147	91.99
$0 \leq \upsilon_u \leq 1$	83.07	67.42

2.8   Conclusions

In this chapter we formulate adaptation of antenna tilt angle as a utility fair optimisation task. Namely, the objective is to jointly adjust antenna tilt angles within the cellular network so as to maximise user utility, subject to network constraints. Adjustments at base stations must be carried out jointly in a coordinated manner in order to manage interference. This optimisation problem is non-convex, but we show that under certain conditions it can be

reformulated as a convex optimisation. Specifically, we show that (i) in the high SINR operating regime and with an appropriate choice of variables the optimisation is convex for any concave utility function, and (ii) in any SINR regime the optimisation can be formulated in a convex manner when the objective is a proportional fair rate allocation. Since the optimisation is not well-suited to use of standard dual methods, we develop a primal-dual method for finding the optimal antenna tilt angles. This approach is lightweight, making use of measurements which are already available at base stations, and suited to distributed implementation. The effectiveness of the proposed approach is demonstrated using a number of simulation examples, including a realistic example based on the cellular network in Dublin, Ireland, and is found to yield considerable performance gains.

User Association in LTE/802.11 Networks

In this chapter we consider proportional fair rate allocation in a heterogeneous network with a mix of LTE and 802.11 cells which supports multipath and multihomed operation (simultaneous connection of a user device to multiple LTE BSs and 802.11 APs). We show that the utility fair optimisation problem is non-convex but that a global optimum can be found by solving a sequence of convex optimisations in a distributed fashion. The result is a principled approach to offload from LTE to 802.11 and for exploiting LTE/802.11 path diversity to meet user traffic demands.

3.1 Introduction

In this chapter we consider rate allocation in a heterogeneous network with a mix of LTE and 802.11 cells. Integrated design of LTE and 802.11 is topical in view of the continuing increases in data traffic, the fact that many cellular network operators also operate a large network of 802.11 hotspots and that user handsets are now typically equipped with both LTE and 802.11 interfaces. Rather than an either/or handover-like setting where the question is which network to use our interest is instead in settings where user devices jointly use the LTE and 802.11 networks and may send data across both simultaneously. Further, we consider situations where user devices may connect to multiple LTE BSs and 802.11 APs simultaneously. This allows us to encompass the coordinated multi point (CoMP) features in release 11 of LTE [45] which allow coordinated transmission and reception across multiple BSs and also multihoming to multiple APs and through a single WLAN card as proposed in, for example, [46]. Of course connection to a single BS or AP remains as a special case within this more general framework. Simultaneous transmission

across multiple interfaces might be implemented by routing each connection across one or other network (in a manner akin to load balancing) or by striping connections across both networks (e.g. via use of transport layer protocols such as multi-path TCP [47]).

Our focus in this chapter is on how to allocate the available LTE and 802.11 bandwidth amongst user devices in a heterogeneous network, and in particular how to determine a proportional fair rate allocation. Our main contributions are as follows. Firstly, we develop a throughput model for heterogeneous networks that include both LTE and 802.11 links and which encompasses multipath and multihomed operation (simultaneous connection of a user device to multiple LTE BSs and 802.11 APs). We show that the rate region is non-convex, and is also not log-convex. Our second contribution is therefore a sequential convex optimisation approach, based on determining a sequence of maximal convex subsets, that we show converges to a global optimum. This optimisation approach is suited to distributed implementation. Thirdly, we present a number of application examples demonstrating how this framework can be used to develop principled approaches to offload from LTE to 802.11 and for exploiting LTE/802.11 path diversity to meet user traffic demands.

3.1.1 Motivating Example

Consider the simple example in Fig 3.1, where the network consists of one cellular BS and one 802.11 AP. UE u_2 is located close to the 802.11 AP and so uses a physical rate of 54Mbps. UE u_1 is further from the AP and so uses a lower physical rate of 1Mbps. Both UEs are located a similar distance from the LTE base station and use the same physical rate of 10Mbps for their LTE connection. The physical rates available on the 802.11 and LTE links are summarised in Table 3.1. For simplicity, we assume both UEs are saturated *i.e.* always have a packet to send.
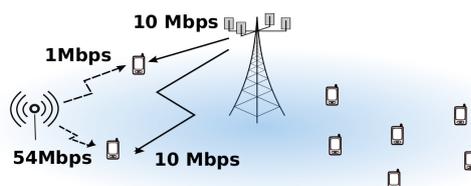
Figure 3.1: Example of a mixed LTE/802.11 network, with one 802.11 AP, one LTE BS and two UEs u_1 and u_2 each equipped with both an 802.11 and an LTE interface.

When the UEs are contending for the available resources, the achievable throughput is determined by the channel access method. For 802.11 the random access approach leads to collisions and reduced throughput when more than one station contends for access. When u_1 and u_2 use only the 802.11

WLAN, their data rates are, respectively 0.46Mbps and 24.9Mbps for the standard 802.11n MAC settings (recall the UE physical rates are 1Mbps and 54Mbps). In the case of LTE, time-frequency slots are allocated to UEs by the cell BS. When the UEs only use the LTE cell, the BS allocates a rate of 5Mbps to each UE.

It can be seen that u_1 enjoys better throughput via the LTE link than over 802.11, while u_2's throughput over the 802.11 link is significantly better than that over the LTE connection. Due to this connection diversity the potential exists for more efficient decision-making to improve performance. For this example we later show that if u_1 sends all traffic via the BS, and u_2 sends all traffic via the AP, rates of 10Mbps and 45Mbps can be achieved by the UEs and that this is the proportional fair rate allocation. This is summarised in Table 3.1. Note that simply splitting the traffic for each UE equally over the 802.11 and LTE networks would yield rates of 5.46Mbps and 29.9Mbps *i.e.* worse for *both* users than the optimised allocation.

Table 3.1: Motivating Example: Achievable data rates

	PHY Rates [Mbps]		Technology Rates [Mbps]		
	BS	AP	LTE Only	802.11 Only	Optimised Multi-RAT
u_1	10	1	5	0.46	10
u_2	10	54	5	24.9	42.4

3.2   Related Work

The problem of resource allocation in heterogeneous wireless networks (Het-Nets) has been the subject of significant interest in the literature in the recent years. Much of this work has focussed on network selection, namely the task of selecting the best network for a UE to use. See [48] for a survey on network selection in HetNets. In network selection two modes of UE operation are typically considered: multi-modal and multi-homed. Although both multi-mode and multi-home operations support multiple radio interfaces, only the latter supports multiple TCP flows across disjoint paths. The task of network selection is usually based on a specified utility function with various methods (combinatorial, fuzzy logic, MDP, game theory, *etc.*) proposed to solve the resulting utility-based network selection problem. For example, [49], [50] consider combinatorial optimisation, [49] considers a linear utility function in a multi-mode UMTS/GSM network and [50] a linear utility function which is the sum of logarithmic utilities of individual BS-UE pairs. In [51] a network-user association problem in a WLAN/UMTS hybrid network for multi-mode operations is solved using Markov Decision Processes (MDPs), although the complexity scales poorly with network size [52]. In [53] a non-cooperative

game is formed among users of two different classes corresponding to 3G/LTE and 802.11.

Our work differs from this previous work in that (i) we consider concurrent association of the users among OFDMA LTE BSs and 802.11 CSMA APs by considering both access network models, (ii) we formulate the problem as a proportional fair resource allocation problem, (iii) we show that although the problem is non-convex it can be solved by solving a sequence of convex optimisation methods and (iv) we use a low-complexity distributed optimisation method to solve each convex optimisation problem.

3.3 Network model

3.3.1 Network topology

We consider a fairly general multi-RAT setup where UEs can potentially connect to multiple LTE base stations and multiple 802.11 APs simultaneously. Networks where UEs are constrained to connect to a single LTE BS and/or 802.11 AP are then special cases within this general setup. We note that the CoMP features in release 11 of LTE already allow transmission and reception across multiple BSs [45],[54]. For 802.11 devices, connection to multiple APs might be implemented using the AP-hopping approach described in [46], where an 802.11 user periodically hops between a set of APs, informing those APs with which it is not currently transmitting/receiving that it is in power-save mode so that these APs buffer packets directed to the user until it next connects. This software implementation is a client-side only change that does not require more than a single 802.11 radio and which uses standard hardware.

Let $\mathcal{A}$ denote the set of 802.11 access points and $\mathcal{U} := \{1, 2, \ldots, U\}$ the set of users. Associated with each access point $a \in \mathcal{A}$ is the set of users $\mathcal{U}_a \subset \mathcal{U}$ which can associate to it (so capturing geographic and access control constraints). We also let $\mathcal{A}_u := \{a \in \mathcal{A} : u \in \mathcal{U}_a\} \cup \{\emptyset\}$ denote the set of access points with which user u can associate, where $\emptyset$ corresponds to the case where the user is not associated with any access point. This also defines a bipartite graph $\mathcal{G}$ with one set of vertices corresponding to users, a second set of vertices corresponding to the access points and edges between each user vertex and the set of access point vertices to which it can connect.

Similarly, let $\mathcal{B}$ denote the set of LTE base stations, $\mathcal{U}_b$ the set of users located in the cell operated by $b \in \mathcal{B}$ and $\mathcal{B}_u$ the set of base stations with which user u can connect.

3.3.2 LTE throughput

Similar to the notations used in Chapter 2, we let $\mathcal{I}$ denote the set of available LTE sub-channels. However note that unlike the system model in Chapter 2 we consider coordinated scheduling between neighbour BSs. We let $\zeta_{b,u}^i$ denote fraction of time sub-channel i of BS b is used by user u. Let $\mathcal{E}^i$ be the set of

user-BS pairs for which transmissions interfere on sub-channel i (this defines a conflict graph). We consider joint sub channel allocation and time sharing so that at a given time interfering BSs do not transmit on the same sub-channel [55]. Hence, the total allocation for each sub-channel must satisfy the following:

$$\sum_{(u,b)\in\mathcal{E}^i} \zeta_{b,u}^i \leq 1 \quad , \forall i \in \mathcal{I} \tag{3.1}$$

When there is no frequency reuse, this constraint simplifies to $\sum_{u\in\mathcal{U}}\sum_{b\in\mathcal{B}}\zeta_{b,u}^i \leq 1$, $\forall i \in \mathcal{I}$. Letting ω^i denote the sub-channel bandwidth, then the achievable rate of user u is given by:

$$r_u = \sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}} \zeta_{b,u}^i \beta_1 \omega^i \log\left(1 + \frac{\gamma_{b,u}^i}{\beta_2}\right) \quad \forall u \in \mathcal{U} \tag{3.2}$$

with $\gamma_{b,u}^i = \frac{p_b h_{b,u}^i}{\sigma_n^2}$ denoting the SNR on sub-channel i of BS b where p_b, $h_{b,u}^i$, σ_n^2 denote the BS power on channel i (without loss of generality, we assumed that all sub-channels experience equal transmit power allocations), channel gain, and noise power at the receiver respectively. Hence the rate region of the LTE network is the following

$$\mathcal{R}_{lte} = \left\{\mathbf{r} : r_u = \sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}} \zeta_{b,u}^i \beta_1 \omega^i \log\left(1 + \frac{\gamma_{b,u}^i}{\beta_2}\right), \underline{r} \leq r_u \leq \bar{r}, \ 0 \leq \zeta_{b,u}^i \leq 1, \right.$$
$$\left. \sum_{u\in\mathcal{U}}\sum_{b\in\mathcal{B}} \zeta_{b,u}^i \leq 1, \ \forall i \in \mathcal{I} \right\} \tag{3.3}$$

where $\mathbf{r} \in \mathbb{R}^U$ is the vector formed by stacking the user throughputs r_u, $u \in \mathcal{U}$ and we have also added the constraint that each user has maximum and minimum rates $\bar{r}$ and $\underline{r}$ respectively. Note that $\underline{r}$ may be 0 and $\bar{r}$ may be ∞. It can be seen that the LTE rate region is convex (a polytope in fact, since the constraints are linear) in the time-frequency sharing factors $\zeta_{b,u}^i$.

We note that this model can be extended to include dynamic power allocation per sub-channel, for example using a similar approach to [55] and will still be convex. Additionally, observe that Eq (3.2) can be used to describe both uplink and downlink throughputs, where by replacing p_b with p_u, $\gamma_{b,u}^i$ will denote the uplink SINR as opposed to the downlink SINR in the discussion above.

3.3.3 802.11 WLAN scheduling

We consider an AP hopping approach where time is partitioned into scheduling slots of duration T indexed by $t = 1, 2, \ldots$ such that in slot t user $u \in \mathcal{U}$ operates its 802.11 MAC in association with at most one access point (reflecting the constraint that user u only has a single 802.11 radio). Extension to allow simultaneous connection to multiple access points is straightforward.

The access point which user u selects at time slot t is a random variable $\mathbf{A}_{u,t}$ which takes value a from $\mathcal{A}_u \setminus \{\emptyset\}$ with probability $z_{a,u}$ (so $0 \le z_{a,u} \le 1$ and $\sum_{a \in \mathcal{A}_u} z_{a,u} \le 1$). Access points are selected independently at each slot and by each user. Note that with this randomised schedule a user is not associated with any access point in a time slot (i.e. $\mathbf{A}_{u,t} = \emptyset$) with probability $p_{\emptyset,u} := 1 - \sum_{a \in A_u} z_{a,u}$.

	$t = 1$	$t = 2$	$t = 3$
$\mathbf{A}_{1,t}$	a_1	a_2	a_2
$\mathbf{A}_{2,t}$	a_1	a_1	a_2
$\mathbf{N}_{1,t}$	2	1	0
$\mathbf{N}_{2,t}$	0	1	2

Figure 3.2: Illustrating the 802.11 random variables $\mathbf{A}_{u,t}$ and $\mathbf{N}_{a,t}$ related to user–AP association, for a two user, two AP example.

Let random variable $\mathbf{N}_{a,t}$ denote the number of users associated with access point a in time slot t. This is illustrated in Fig 3.2 for a setup with two UEs, and two APs a_1, a_2. In slot 1 both UEs are associated with AP a_1. In slot 2 user u_1 is associated with AP a_2, while user u_2 is associated with AP a_1 and in slot 3 both users are associated with AP a_2.

By construction, random variables $\mathbf{N}_{a,t}$, $t = 1, 2, \ldots$ are i.i.d. Let $2^{\mathcal{S}}$ denote the set consisting of the subsets of set $\mathcal{S}$, and let $\mathcal{P}_\kappa(\mathcal{S}) = \{s : s \in 2^{\mathcal{S}}, |s| = \kappa\}$ denote the subsets of set $\mathcal{S}$ which have cardinality κ. Therefore $\mathcal{P}_{n-1}(\mathcal{U}_a \setminus \{u\})$ is all subsets of $\mathcal{U}_a \setminus \{u\}$ with cardinality $n - 1$. Define

$$p_{a,u,n} := Prob(\mathbf{N}_{a,t} = n | \mathbf{A}_{u,t} = a) \tag{3.4}$$

to be the probability of n users being associated with AP a conditioned on user u being associated with a. We have

$$p_{a,u,n} = \begin{cases} \prod\limits_{v \in \mathcal{U}_a \setminus \{u\}} (1 - z_{a,v}) & n = 1 \\ \sum\limits_{\substack{\tilde{\mathcal{U}}_a \in \\ \mathcal{P}_{n-1}(\mathcal{U}_a \setminus \{u\})}} \prod\limits_{v \in \tilde{\mathcal{U}}_a} z_{a,v} \prod\limits_{v \in \mathcal{U}_a \setminus \{\tilde{\mathcal{U}}_a, u\}} (1 - z_{a,v}) & n > 1 \end{cases} \tag{3.5}$$

Letting $w_{u,a} = \frac{z_{a,u}}{1 - z_{a,u}}$, this can be rewritten equivalently as

$$p_{a,u,1} = \frac{1}{\prod\limits_{v \in \mathcal{U}_a \setminus \{u\}} (1 + w_{a,v})} \tag{3.6}$$

and for $n > 1$

$$p_{a,u,n} = \sum_{\substack{\tilde{\mathcal{U}}_a \in \\ \mathcal{P}_{n-1}(\mathcal{U}_a \setminus \{u\})}} \prod_{v \in \tilde{\mathcal{U}}_a} \frac{z_{a,v}}{1 - z_{a,v}} \prod_{v \in \mathcal{U}_a \setminus \{u\}} (1 - z_{a,v}) = \frac{\sum\limits_{\tilde{\mathcal{U}}_a \in \mathcal{P}_{n-1}(\mathcal{U}_a \setminus \{u\})} \prod\limits_{v \in \tilde{\mathcal{U}}_a} w_{a,v}}{\prod\limits_{v \in \mathcal{U}_a \setminus \{u\}} (1 + w_{a,v})}$$

$$(3.7)$$

Defining,

$$q_{a,u,n} := \begin{cases} 1 & n = 1 \\ \sum\limits_{\tilde{\mathcal{U}}_a \in \mathcal{P}_{n-1}(\mathcal{U}_a \setminus \{u\})} \prod\limits_{v \in \tilde{\mathcal{U}}_a} w_{a,v} & n > 1 \end{cases} \tag{3.8}$$

then the expression for probability $p_{a,u,n}$ simplifies to

$$p_{a,u,n} = q_{a,u,n} p_{a,u,1}, \ n = 1, 2, \ldots \tag{3.9}$$

Note that $\sum_{n=1}^{|\mathcal{U}_a|} p_{a,u,n} = p_{a,u,1} \sum_{n=1}^{|\mathcal{U}_a|} q_{a,u,n} = 1$ and so $\sum_{n=1}^{|\mathcal{U}_a|} q_{a,u,n} = \frac{1}{p_{a,u,1}}$.

3.3.4 802.11 MAC slots

The number of users actively using an AP varies from time slot to time slot. We therefore have to take some care to define the MAC slots within each WLAN appropriately. Within each WLAN MAC slots are induced by carrier-sensing, which may be idle slots of duration σ (no user transmits) or busy slots of duration T_b (at least one user transmits, this includes both successful transmission and collision times). In order to distinguish between MAC slots associated with different access points, we index MAC slots by $i \in \mathcal{A} \times \mathbb{N}$ i.e. the MAC slots associated with access point a are $i = (a, 1), (a, 2), \ldots$.

We let $\mathcal{M}_{a,t}$ denote the set of MAC slots of access point a that are fully contained within time slot t and $\bar{\mathcal{M}}_{a,t}$ be the set of MAC slots which are only partially contained in time slot t. Since the MAC slot duration is random, generally there will be boundary MAC slots which are only partially contained in time slot t but there are at most two such slots. We will assume that the time slot duration T is sufficiently large that partial MAC slots $\bar{\mathcal{M}}_{a,t}$ can be neglected when calculating user airtime and throughput. This is illustrated in Figure 3.3.

3.3.5 802.11 throughput

Let $L_{a,u}$, $T_{succ,a,u}$, T_{coll} and σ denote, respectively, the payload in bits of a successful transmission by user u when associated with AP a, the mean duration of a successful transmission, the mean duration of a collision and the idle slot duration. To simplify notation we confine consideration to user uplink transmissions. Extension to include AP downlink transmissions is straightforward, simply requiring appropriate book-keeping of which fraction of AP transmissions is directed to each user. For simplicity we also assume that $T_{succ,a,u}$ is the same for all users transmitting in WLAN a and let $T_{succ,a} = T_{succ,a,u}$.
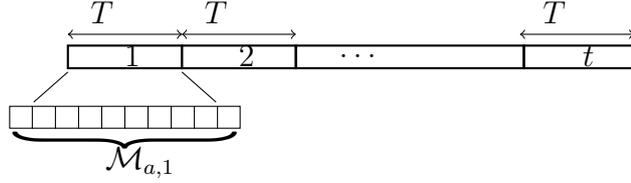
Figure 3.3: Illustrating 802.11 time slotting : each user u can hop between multiple APs, but is associated to the same AP for time slots of duration T. Within a time slot of duration T, 802.11 carrier sense defines MAC slots. $\mathcal{M}_{a,t}$ denotes the set of MAC slots fully contained within time slot t.

In time slot t let $\mathcal{U}_{a,t} \subset \mathcal{U}_a$ denote the set of users associated with access point a and recall that $\mathbf{N}_{a,t} = |\mathcal{U}_{a,t}|$. Let $\boldsymbol{\mathcal{X}}_{i,u}$ be a random variable which is equal to 1 if user u transmits in MAC slot $i \in \mathcal{M}_{a,t}$ and 0 otherwise where $u \in \mathcal{U}_a$. We let $\boldsymbol{\mathcal{X}}_{i,u} = 0$ for users $u \notin \mathcal{U}_{a,t}$ not associated with access point a in slot $i \in \mathcal{M}_{a,t}$. For $u \in \mathcal{U}_{a,t}$ we assume that the $\{\boldsymbol{\mathcal{X}}_{i,u}\}$ are i.i.d. in both u and i i.e. $\boldsymbol{\mathcal{X}}_{i,u} \sim \boldsymbol{\mathcal{X}}$, and we let $\tau := Prob(\boldsymbol{\mathcal{X}} = 1)$. This can be achieved in practice by setting $CW_{max} = CW_{min}$ and using the same value of CW_{min} for all associated users, in which case $\tau = 2/(CW_{min} + 1)$ [56]. We let $\boldsymbol{\mathcal{Y}}_{i,u}$ be a random variable that equals 1 if user u transmits successfully in MAC slot i, and 0 otherwise. We let $\boldsymbol{\mathcal{Z}}_i$ equal 0 when one or more users transmits in MAC slot i, and 1 otherwise. Since transmissions $\{\boldsymbol{\mathcal{X}}_{i,u}\}$ are i.i.d. for $u \in \mathcal{U}_{a,t}$, $i \in \mathcal{M}_{a,t}$, we have that $\boldsymbol{\mathcal{Y}}_{i,u} \sim \boldsymbol{\mathcal{Y}}^{|\mathcal{U}_{a,t}|}$ with $E[\boldsymbol{\mathcal{Y}}^{|\mathcal{U}_{a,t}|}] = \tau(1-\tau)^{|\mathcal{U}_{a,t}|-1}$, and $\boldsymbol{\mathcal{Z}}_i \sim \boldsymbol{\mathcal{Z}}^{|\mathcal{U}_{a,t}|}$ with $E[\boldsymbol{\mathcal{Z}}^{|\mathcal{U}_{a,t}|}] = (1-\tau)^{|\mathcal{U}_{a,t}|}$.

The throughput of user u in WLAN a is given by

$$s_{a,u} = \lim_{k \to \infty} \frac{\sum_{t=1}^{k} \sum_{i \in \mathcal{M}_{a,t}} \boldsymbol{\mathcal{Y}}_{i,u} L_{a,u}}{kT} = \lim_{k \to \infty} \frac{1}{k} \sum_{t \in \{s \in \{1,2,...,k\}: \mathbf{A}_{u,s}=a\}} \sum_{i \in \mathcal{M}_{a,t}} \boldsymbol{\mathcal{Y}}_{i,u} \frac{L_{a,u}}{T}$$

$$= \lim_{k \to \infty} \sum_{n=1}^{|\mathcal{U}_a|} \frac{|\mathcal{T}_{a,n}^k|}{k} \frac{1}{|\mathcal{T}_{a,n}^k|} \sum_{t \in \mathcal{T}_{a,n}^k} \sum_{i \in \mathcal{M}_{a,t}} \boldsymbol{\mathcal{Y}}_{i,u} \frac{L_{a,u}}{T}$$

$$(3.10)$$

where $\mathcal{T}_{a,n}^k := \{s \in \{1, 2, \ldots, k\} : A_{u,s} = a, \mathbf{N}_{a,s} = n\}$ and we have used the fact that $\mathbf{x}_{i,u} = 0$ when $u \notin \mathcal{U}_a$.

Defining

$$P_{succ,n} = \sum_{j=1}^{n} E[\boldsymbol{\mathcal{Y}}^n] = n\tau(1-\tau)^{n-1} \tag{3.11}$$

$$P_{idle,n} = E[\boldsymbol{\mathcal{Z}}^n] = (1-\tau)^n \tag{3.12}$$

$$P_{coll,n} = 1 - P_{idle,n} - P_{succ,n} \tag{3.13}$$

we have

$$\lim_{k \to \infty} \frac{1}{|\mathcal{T}_{a,n}^k|} \sum_{t \in \mathcal{T}_{a,n}^k} \sum_{i \in \mathcal{M}_{a,t}} \mathcal{Y}_{i,u} \frac{L_{a,u}}{T} = E[|\mathcal{M}_{a,t}| | \mathbf{N}_{a,t} = n, \mathbf{A}_{u,t} = a] E[\mathcal{Y^n}] \frac{L_{a,u}}{T}$$

$$= \frac{\tau(1-\tau)^{n-1} L_{a,u}}{P_{idle,n}\sigma + P_{succ,n}T_{succ,a} + P_{coll,n}T_{coll}} \quad (3.14)$$

We also have $\lim_{k \to \infty} \frac{|\mathcal{T}_{a,n}^k|}{k} = Prob(\mathbf{A}_{u,t} = a, \mathbf{N}_{a,t} = n) = Prob(\mathbf{A}_{u,t} = a)Prob(\mathbf{N}_{a,t} = n | \mathbf{A}_{u,t} = a) = z_{a,u}p_{a,u,n}$. Thus, the throughput of user u in the WLAN operated by AP a is given by

$$s_{a,u} = z_{a,u} \sum_{n=1}^{|\mathcal{U}_a|} \frac{p_{a,u,n}\tau(1-\tau)^{n-1} L_{a,u}}{P_{idle,n}\sigma + P_{succ,n}T_{succ,a} + P_{coll,n}T_{coll}} \quad (3.15)$$

3.3.5.1 Each User Associates to a Single WLAN

It is helpful to consider the special case where each user u connects only to a single WLAN a, in which case $z_{a,u} = 1$ and $z_{a',u} = 0$ for $a' \neq a$. Letting n_a denote the number of users connected to WLAN a then $p_{a,u,n_a} = 1$ and $p_{a,u,n} = 0$ for $n \neq n_a$. It follows that (3.15) simplifies to,

$$s_{a,u} = \frac{\tau(1-\tau)^{n_a-1} L_{a,u}}{P_{idle,n_a}\sigma + P_{succ,n_a}T_{succ,a} + P_{coll,n_a}T_{coll}} \quad (3.16)$$

which is identical to classical 802.11 WLAN models *e.g.* that in [57]. However, when the number of users using a WLAN varies from time slot to time slot, (3.15) yields the shared throughput.

3.3.5.2 Change of variables

It will prove useful to work in terms of the quantities $\psi = \frac{\tau}{1-\tau}$ and $w_{a,u} = \frac{z_{a,u}}{1-z_{a,u}}$ rather than τ and $z_{a,u}$s respectively. Also we will use the normalised throughput $\rho_{a,u} = \frac{s_{a,u}}{c_{a,u}}$, where $c_{a,u} = \frac{L_{a,u}}{T_{succ_a}}$. Recalling $p_{a,u,n} = q_{a,u,n}p_{a,u,1}$, we have

$$\rho_{a,u} = \frac{w_{a,u}}{1+w_{a,u}} \sum_{n=1}^{|\mathcal{U}_a|} \frac{q_{a,u,n}}{\prod_{v \in \mathcal{U}_a \setminus \{u\}} (1+w_{a,v})} \frac{\psi}{\Psi_n(\psi)} \frac{T_{succ_a}}{T_{coll}} \quad (3.17)$$

with

$$\Psi_n(\psi) = \frac{\sigma}{T_{coll}} + n\left(\frac{T_{succ_a}}{T_{coll}} - 1\right)\psi + (\psi+1)^n - 1 \quad (3.18)$$

3.3.5.3   802.11 Rate Region

With the above change of variables the rate region of the 802.11 network is,

$$
\mathcal{R}_{wifi} = \Bigg\{ \mathbf{s} : s_u = \sum_{a \in \mathcal{A}_u} \rho_{a,u} c_{a,u},
$$

$$
\rho_{a,u} \le \sum_{a \in \mathcal{A}_u} \frac{w_{a,u}}{\prod\limits_{v \in \mathcal{U}_a} (1 + w_{a,v})} \sum_{n=1}^{|\mathcal{U}_a|} q_{a,u,n} \frac{\psi}{\Psi_n(\psi)} \frac{T_{succ_a}}{T_{coll}},
$$

$$
q_{a,u,n} \le \sum_{\substack{\bar{\mathcal{U}}_a \in \\ \mathcal{P}_{n-1}(\mathcal{U}_a \setminus \{u\})}} \prod_{v \in \tilde{\mathcal{U}}_a} w_{a,v}, \ n > 1, \ \sum_{a \in \mathcal{A}} \frac{w_{a,u}}{1 + w_{a,u}} \le 1,
$$

$$
0 \le \rho_{a,u} \le 1, \ w_{a,u} \ge 0, \ q_{a,u,n} \ge 0, \forall n \in \mathcal{U}, \ \forall a \in \mathcal{A}, \ \forall u \in \mathcal{U} \Bigg\}
$$

$$(3.19)$$

where $\mathbf{s} \in \mathbb{R}^U$ is the vector obtained by stacking the user throughputs s_u, $u \in \mathcal{U}$. Letting $\tilde{\rho}_{a,u} = \log \rho_{a,u}$, $\tilde{w}_{a,u} = \log w_{a,u}$, we also have that the rate region of the 802.11 network can be written as

$$
\hat{\mathcal{R}}_{wifi} = \Bigg\{ \mathbf{s} : s_u = \sum_{a \in \mathcal{A}_u} e^{\tilde{\rho}_{a,u}} c_{a,u},
$$

$$
\tilde{\rho}_{a,u} \le \tilde{w}_{a,u} - \sum_{v \in \mathcal{U}_a} \log(1 + e^{\tilde{w}_{a,v}}) + \log \sum_{n=1}^{|\mathcal{U}_a|} \frac{T_{succ_a}}{T_{coll}} \frac{\psi}{\Psi_n} q_{a,u,n},
$$

$$
q_{a,u,n} \le \sum_{\substack{\bar{\mathcal{U}}_a \in \\ \mathcal{P}_{n-1}(\mathcal{U}_a \setminus \{u\})}} \prod_{v \in \tilde{\mathcal{U}}_a} e^{\tilde{w}_{a,v}}, \ n > 1, \ \sum_{a \in \mathcal{A}} \frac{e^{\tilde{w}_{a,u}}}{1 + e^{\tilde{w}_{a,u}}} \le 1,
$$

$$
q_{a,u,n} \ge 0, \forall n \in \mathcal{U}, \ \forall a \in \mathcal{A}, \ \forall u \in \mathcal{U} \Bigg\}
$$

$$(3.20)$$

3.3.5.4   Example 802.11 Rate Region

Recall the simple example in Fig 3.1 where we have one BS, one AP and two users. The 802.11 rate region is,

$$
\mathcal{R}_{wifi} = \Bigg\{ \mathbf{s} : s_1 = \rho_{1,1} c_{1,1}, \ s_2 = \rho_{1,2} c_{1,2},
$$

$$
\rho_{1,1} \le \frac{w_{1,1}}{(1 + w_{1,1})(1 + w_{1,2})} \Big(\frac{\psi}{\Psi_1} + q_{1,1,2} \frac{\psi}{\Psi_2} \Big) \frac{T_{succ_{1,1}}}{T_{coll}}, \ q_{1,1,2} \le w_{1,2},
$$

$$
\rho_{1,2} \le \frac{w_{1,2}}{(1 + w_{1,1})(1 + w_{1,2})} \Big(\frac{\psi}{\Psi_1} + q_{1,1,2} \frac{\psi}{\Psi_2} \Big) \frac{T_{succ_{1,2}}}{T_{coll}}, \ q_{1,2,2} \le w_{1,1},
$$

$$
0 \le \rho_{1,u} \le 1, \ w_{1,u} \ge 0, \ q_{1,u,n} \ge 0, \ \big\{ u, n | u, n \in \{1, 2\} \big\} \Bigg\}
$$

$$(3.21)$$

(a) 1 AP: $\tau_{a_1} = 0.95$

(b) 2 APs: $\tau_{a_1} = 0.95, \tau_{a_2} = 0.2$
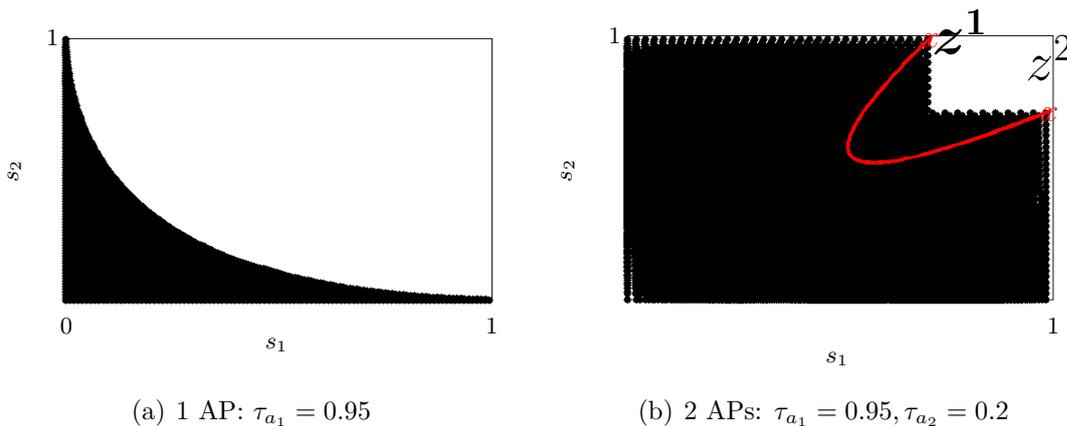
Figure 3.4: Examples of 802.11 rate region $\mathcal{R}_{wifi}$ where hopping between APs is admissible. In (a) two users hop between connecting and not connecting to an AP and in (b) two users hop between two APS for which the transmission probabilities differ.

3.3.6   Non-convexity of Network Rate Region

The network rate region is the joint LTE/802.11 rate region $\mathcal{R}_{lte} \times \mathcal{R}_{wifi}$. The LTE rate region $\mathcal{R}_{lte}$ is convex and so the convexity/non-convexity of the network rate region depends on the convexity/non-convexity of the 802.11 component of the network rate region $\mathcal{R}_{wifi}$.

In the special case (see above) where for each user u we force $z_{a,u} = 1$ for one AP a, then each user associates to a single, fixed AP and the only design freedom lies in the LTE allocation which is a convex problem. However, when we additionally allow the UEs to adjust the fraction of time that they connect to this single AP then the rate region becomes non-convex, as illustrated in Figure 3.4(a), and when users may connect to more than one AP is illustrated in Fig 3.4(b). Fig 3.4(b) plots the rate region $\mathcal{R}_{wifi}$ for networks with two APs and two UEs. The non-convexity of the rate region in Fig 3.4(b) is evident. Note that the two extreme points of the "notch" in the rate region indicated in the figure correspond to parameter values $\mathbf{z}^1 = [1\ 0\ 0\ 1]^T$, and $\mathbf{z}^2 = [0\ 1\ 1\ 0]^T$, where $\mathbf{z} := [z_{1,1}\ z_{1,2}\ z_{2,1}\ z_{2,2}]^T$. That is, point $\mathbf{z}^1$ corresponds to u_1 being connected to AP a_1 100% of the time and u_2 to AP a_2, while $\mathbf{z}^2$ is the reverse. The UE rates along the time-sharing chord $\alpha\mathbf{z}^1 + (1 - \alpha)\mathbf{z}^2$, $\alpha \in [0, 1]$ are indicated by the line marked on Fig 3.4(b). It can be seen that this lies in the *interior* of the rate region. Simple time-sharing therefore does not yield convexity, and this is due to the collisions in 802.11 WLANs shared by more than one UE.

3.4 Proportional Fair Rate Allocation

3.4.1 Utility Fair Optimisation

We are now in a position to consider the main focus of this work, namely finding a proportional fair rate allocation for the joint LTE/802.11 network. This is the solution to the following utility optimisation P,

$$\max_{\mathbf{s},\mathbf{r}} \sum_{u \in \mathcal{U}} \log\left(s_u + r_u\right) \tag{3.22}$$

$$s.t. \quad \mathbf{s} \in \hat{\mathcal{R}}_{wifi}, \ \mathbf{r} \in \mathcal{R}_{lte} \tag{3.23}$$

As already noted, the rate region $\mathcal{R}_{wifi}$ (and so $\hat{\mathcal{R}}_{wifi}$) is generally non-convex and so the utility fair optimisation P is non-convex. However, as we will see the optimisation possesses sufficient convex structure to allow near-optimal solutions to be found in an efficient manner. We proceed by first considering methods for approximating a non-convex set by a maximal convex subset. In this way we can define a convex approximation to P for which solutions can be efficiently found. We then consider a concave-convex procedure that adaptively selects the maximal convex subsets with the aim of maximising the optimisation objective. This procedure is guaranteed to converge to a stationary point of non-convex optimisation P. Although this stationary will in general be sub-optimal, in practice we find that it is usually near-optimal.

3.4.2 Approximate Optimisation Via Maximal Convex Subsets

3.4.2.1 Maximal convex subsets

Sets of the form $E := \{(x,y) : y \leq e^x, (x,y) \in \mathbb{R}^2\}$ will prove important in our analysis. Set E is concave, see Figure 3.5. Nevertheless, large convex subsets of this set can be readily identified. In particular, the set $F := \{(x,y) : y \leq e^{\bar{x}}(1 + x - \bar{x}), (y,x) \in \mathbb{R}^2\}$, with parameter $\bar{x} \in \mathbb{R}$, is convex with $F \subset E$, as illustrated in Figure 3.5. Convexity follows from the fact that the constraint defining set F is linear in x and y. That F is contained in E follows from the observation that the complement $\bar{E} := \{(x,y) : y \geq e^x, (x,y) \in \mathbb{R}^2\}$ of set E is convex and the boundary of F is a supporting hyperplane to $\bar{E}$ at point $(x, e^{\bar{x}})$.
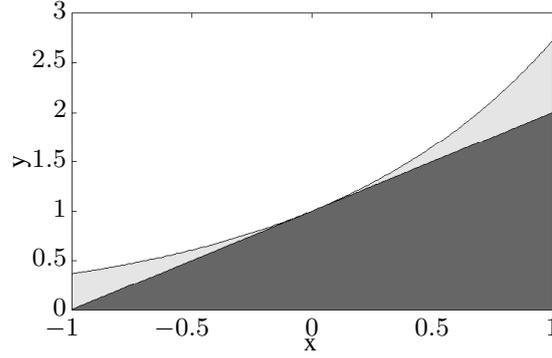
Figure 3.5: Illustrating a maximal convex subset (indicated in darker shade) of concave set $y \leq e^x$.

3.4.2.2   Maximal Convex Subsets of 802.11 Rate Region

The main constraints in the 802.11 rate region $\hat{\mathcal{R}}_{wifi}$ can be written in standard form as

$$\tilde{\rho}_{a,u} - \tilde{w}_{a,u} + \sum_{v \in \mathcal{U}_a} \log(1 + e^{\tilde{w}_{a,v}}) - \log\left(\sum_{n=1}^{|\mathcal{U}_a|} \frac{\psi}{\Psi_n} q_{a,u,n}\right) \leq 0 \tag{3.24}$$

$$q_{a,u,n} - \sum_{\substack{\tilde{\mathcal{U}}_a \in \\ \mathcal{P}_{n-1}(\mathcal{U}_a \backslash \{u\})}} \prod_{v \in \tilde{\mathcal{U}}_a} e^{\tilde{w}_{a,v}} \leq 0 \tag{3.25}$$

The terms on the LHS of the $\tilde{\rho}_{a,u}$ constraint are all convex. However, the second term on the LHS of the $q_{a,u,n}$ constraint is concave. We proceed by finding a maximal convex subset.

Letting

$$E := \{(q_{a,u,n}, \{\tilde{w}_{a,v}\}_{v \in \mathcal{U}_a}) : \tilde{w}_{a,v} \in \mathbb{R}, v \in \mathcal{U}_a, \ q_{a,u,n} \leq \sum_{\substack{\tilde{\mathcal{U}}_a \in \\ \mathcal{P}_{n-1}(\mathcal{U}_a \backslash \{u\})}} \prod_{v \in \tilde{\mathcal{U}}_a} e^{\tilde{w}_{a,v}}\}$$
$$\tag{3.26}$$

then a maximal convex subset is

$$F := \{(q_{a,u,n}, \{\tilde{w}_{a,v}\}_{v \in \mathcal{U}_a}) : \tilde{w}_{a,v} \in \mathbb{R}, v \in \mathcal{U}_a, \ q_{a,u,n} \leq \sum_{\substack{\tilde{\mathcal{U}}_a \in \\ \mathcal{P}_{n-1}(\mathcal{U}_a \backslash \{u\})}} \prod_{v \in \tilde{\mathcal{U}}_a} W_{\tilde{\mathcal{U}}_a,v}\}$$
$$\tag{3.27}$$

where $W_{\tilde{\mathcal{U}}_a,v} := e^{\bar{w}_{a,v}}(1 + \sum_{v \in \tilde{U}_a \backslash \{u\}}(\tilde{w}_{a,v} - \bar{w}_{a,v}))$ and $\bar{w}_{a,u}$ is a design parameter.

The 802.11 rate region also contains the constraint

$$s_u - \sum_{a \in \mathcal{A}_u} e^{\tilde{\rho}_{a,u}} c_{a,u} \leq 0 \tag{3.28}$$

46

It can be seen that the LHS is concave. Once again, we adopt a maximal convex subset approach

$$s_u - \sum_{a \in \mathcal{A}_u} e^{\bar{\rho}_{a,u}}(1 + \tilde{\rho}_{a,u} - \bar{\rho}_{a,u})c_{a,u} \leq 0$$

where $\bar{\rho}_{a,u}$ is a design parameter. The LHS is now convex in $\tilde{\rho}_{a,u}$ as required.

Lastly, the 802.11 rate region contains the constraint

$$\sum_{a \in \mathcal{A}} \frac{e^{\tilde{w}_{a,u}}}{1 + e^{\tilde{w}_{a,u}}} \leq 1 \tag{3.29}$$

the LHS of which is sigmoidal in form, so neither convex not concave. Rewriting $\frac{e^{\tilde{w}_{a,u}}}{e^{\tilde{w}_{a,u}}+1}$ equivalently as $e^{\tilde{w}_{a,u}} - \frac{e^{2\tilde{w}_{a,u}}}{1+e^{\tilde{w}_{a,u}}}$ observe that the first term in the RHS is convex while the second term is concave. Linearising the second term yields the following inequality

$$\sum_{a \in \mathcal{A}} e^{\tilde{w}_{a,u}} - \left(\frac{2e^{2\bar{w}_{a,u}}}{1+e^{\bar{w}_{a,u}}} + \frac{e^{3\bar{w}_{a,u}}}{(1+e^{\bar{w}_{a,u}})^2} \right)(\tilde{w}_{a,u} - \bar{w}_{a,u}) \leq 1 \tag{3.30}$$

which defines a maximal convex subset.

3.4.2.3 Solving the Convex Optimisation

Using the maximal convex subset approach described above we obtain a convex optimisation the solution of which is feasible for non-convex optimisation P but will, in general, be sub-optimal. The Slater condition is satisfied and so strong duality holds. We can therefore solve this convex optimisation in a distributed manner using a primal-dual subgradient approach.

Let $\mathbf{x} := [\boldsymbol{\zeta}^T \ \mathbf{r}^T \ \mathbf{s}^T \ \tilde{\boldsymbol{\rho}}^T \ \tilde{\mathbf{w}}^T \ \mathbf{q}^T]^T \in \mathbb{R}^n$ be the vector obtained by stacking LTE and 802.11 rate region variables with $\boldsymbol{\zeta} \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{B}| \times U}$, $\tilde{\boldsymbol{\rho}}, \tilde{\mathbf{w}} \in \mathbb{R}^{|\mathcal{A}| \times U}$, $\mathbf{q} \in \mathbb{R}^{|\mathcal{A}| \times U \times U}$ denoting, respectively, the vectors with elements $\zeta^i_{b,u}$, $\tilde{\rho}_{a,u}$, $\tilde{w}_{a,u}$, $q_{a,u,n}$ for $b \in \mathcal{B}$, $i \in \mathcal{I}$, $a \in \mathcal{A}$, $n \in \mathcal{U}$, $u \in \mathcal{U}$. We re-write the optimisation problem P in the following form,

$$\min_{\mathbf{x}} \ f(\mathbf{x}) \tag{3.31}$$

$$s.t. \quad h^{(i)}(\mathbf{x}) - g^{(i)}(\mathbf{x}) \leq \mathbf{0}, \ i = 1, 2, \dots, l \tag{3.32}$$

with $f : \mathbb{R}^n \to \mathbb{R}$, $h^{(i)} : \mathbb{R}^n \to \mathbb{R}$, $g^{(i)} : \mathbb{R} \to \mathbb{R}$ being convex continuously differentiable functions, expressions for which are summarised below.

With the obvious abuse of notation we add subscripts to the functions $h^{(i)}$ and $g^{(i)}$ to streamline the presentation since there is no risk of confusion. Details

of the objective and constraints functions are as follows:

$$f(\mathbf{x}) = -\sum_{u\in\mathcal{U}} \log(s_u + r_u)$$

$$h_u^{(1)}(\mathbf{x}) = s_u, \quad g_u^{(1)}(\mathbf{x}) = \sum_{a\in\mathcal{A}_u} e^{\tilde{\rho}_{a,u}} c_{a,u}$$

$$h_{a,u}^{(2)}(\mathbf{x}) = \tilde{\rho}_{a,u} - \tilde{w}_{a,u} + \sum_{v\in\mathcal{U}_a} \log(1 + e^{\tilde{w}_{a,v}}) - \log\left(\sum_{n=1}^{|\mathcal{U}_a|} \frac{T_{succ_a}}{T_{coll}} \frac{\psi}{\Psi_n} q_{a,u,n}\right),$$

$$g_{a,u}^{(2)}(\mathbf{x}) = 0$$

$$h_{a,u,n}^{(3)}(\mathbf{x}) = q_{a,u,n}, \quad g_{a,u,n}^{(3)}(\mathbf{x}) = \sum_{\substack{\tilde{\mathcal{U}}_a\in \\ \mathcal{P}_{n-1}(\mathcal{U}_a\backslash\{u\})}} \prod_{v\in\tilde{\mathcal{U}}_a} e^{\tilde{w}_{a,v}}$$

$$h_u^{(4)}(\mathbf{x}) = \sum_{a\in\mathcal{A}_u} e^{\tilde{w}_{a,u}}, \quad g_u^{(4)}(\mathbf{x}) = \sum_{a\in A_u} \frac{e^{2\tilde{w}_{a,u}}}{1 + e^{\tilde{w}_{a,u}}}$$

$$h_i^{(5)}(\mathbf{x}) = \sum_{u\in\mathcal{U}}\sum_{b\in\mathcal{B}} \zeta_{b,u}^i - 1, \quad g_i^{(5)}(\mathbf{x}) = 0$$

$$h_u^{(6)}(\mathbf{x}) = r_u - \sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}} \zeta_{b,u}^i \beta_1 \omega^i \log\left(1 + \frac{\gamma_{b,u}^i}{\beta_2}\right), \quad g_u^{(6)}(\mathbf{x}) = 0$$

Let $-\hat{g}^{(i)}(\mathbf{x}; \bar{\mathbf{x}})$ be the maximal convex subset expression for non-convex function $-g^{(i)}(\mathbf{x})$.

$$-\hat{g}^{(i)}(\mathbf{x}; \bar{\mathbf{x}}) = -g^{(i)}(\bar{\mathbf{x}}) - \partial g_{\mathbf{x}}^{(i)}(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) \tag{3.33}$$

Then the approximate optimisation problem $P_{\bar{\mathbf{x}}}$ is given by

$$\min_{\mathbf{x}} \ f(\mathbf{x}) \tag{3.34}$$

$$s.t. \quad h^{(i)}(\mathbf{x}) - \hat{g}^{(i)}(\mathbf{x}; \bar{\mathbf{x}}) \leq \mathbf{0}, \ i = 1, 2, \ldots, l \tag{3.35}$$

Letting $\Lambda := [\lambda^{(1)}, \cdots, \lambda^{(l)}]^T$ denote the vector of multipliers associated with the rate region constraints 1 to l, the Lagrangian for optimisation problem $P_{\bar{\mathbf{x}}}$ is

$$L(\mathbf{x}, \Lambda; \bar{\mathbf{x}}) = f(\mathbf{x}) + \sum_{i=1}^{l} \lambda^{(i)} \left(h^{(i)}(\mathbf{x}) - \hat{g}^{(i)}(\mathbf{x}; \bar{\mathbf{x}})\right)$$

The standard primal-dual subgradient approach in Algorithm 3.2 can then be used, for example, to find a solution to optimisation $P_{\bar{\mathbf{x}}}$.

3.4.2.4 Message Passing Required

1. To update the LTE sub-channel airtime $\zeta_{b,u}^i$ each UE needs (i) the SNR to its LTE BSs (which it already knows) and (ii) the multiplier associated with the sub-band constraint (which can be communicated by a BS).

Algorithm 3.2 Distributed primal-dual algortithm

Initialise: $t = 0$, $\mathbf{x}(0)$, $\Lambda(0)$, step size $\alpha > 0$

do

$$\mathbf{x}(t+1) = \mathbf{x}(t) - \alpha \partial_{\mathbf{x}} L\left(\mathbf{x}\left(t\right), \Lambda\left(t\right); \bar{\mathbf{x}}\right)$$
$$\Lambda(t+1) = \left[\Lambda(t) + \alpha \partial_{\Lambda} L\left(\mathbf{x}\left(t\right), \Lambda\left(t\right); \bar{\mathbf{x}}\right)\right]^{+}$$
$$t \leftarrow t + 1$$

loop

2. The 802.11 association probability $\tilde{w}_{a,u}$ and WLAN parameters $q_{a,u,n}$ can be updated using information available locally at AP a (no need for message passing) together with knowledge of the multiplier associated with the constraint that association probabilities for user u sum to one. This requires that $\tilde{w}_{a,u}$ as well as $\bar{w}_{a,u}$ be communicated by AP a to UE u and the multiplier then communicated back from UE u to AP a.

3. The 802.11 normalised throughput $\tilde{\rho}_{a,u}$ can be updated by UE u using local information together with knowledge of the multiplier for the rate constraint on $\tilde{\rho}_{a,u}$ (which can be communicated by AP a).

3.4.3 Adaptation of Maximal Convex Subsets

The convex optimisation in the preceding section yields a solution which is feasible for non-convex optimisation P but which is, in general, sub-optimal for P. The degree of sub-optimality is dependent on the choice of maximal convex subsets used to derive the convex optimisation, *e.g.* when the convex subsets contain at least one point which is an optimum of the non-convex problem P then the solution to the convex optimisation $P_{\bar{\mathbf{x}}}$ will in fact be optimal for P.

When we have some knowledge of the part of the network rate region where the optima of problem P are likely to lie, we can use this to select the maximal convex subsets. Such knowledge might be available from prior experience, *e.g.* previous solutions to similar network configurations, or from structural insight. In the next section we illustrate how such prior information can be used to obtain near-optimal solutions.

However, it is also possible to automate this process and this is the focus of the present section. The basic idea is to iteratively update the choice of maximal convex subsets based on the current solution to the convex optimisation. Having obtained the solution to the convex optimisation for the current choice of maximal convex subsets, one natural approach is to use the components $\tilde{w}_{a,u}$, $\tilde{\rho}_{a,u}$ of this solution as the values for the parameters $\bar{w}_{a,u}$, $\bar{\rho}_{a,u}$ and in this way to define a new maximal convex subset.

3.4.3.1 Convergence

In more detail, let $C := \{\mathbf{x} \in B : h^{(i)}(\mathbf{x}) - g^{(i)}(\mathbf{x}) \leq 0, \ i = 1, \cdots ., l\}$ with $B \subset \mathbb{R}^n$ non-empty, convex and compact (closed and bounded). As we will see, set B is needed for technical reasons, to ensure that C is compact, but B can otherwise be chosen arbitrarily and can be viewed as augmenting the set of convex constraints $h^{(i)}$. Of course, we assume that set C is non-empty. Now consider the iterative update

$$\mathbf{x}_{k+1} \in D_{\mathbf{x}_k} f(\mathbf{x}), \ k = 1, \cdots \tag{3.36}$$

with $\mathbf{x}_1 \in C$,

$$D_{\mathbf{x}_k} f(\mathbf{x}) = \begin{cases} \arg\min_{\mathbf{x} \in C_{\mathbf{x}_k}} f(\mathbf{x}) & \text{if } f(\mathbf{x}_k) = f^*_{\mathbf{x}_k} \\ \{\mathbf{x} \in C_{\mathbf{x}_k} : f(\mathbf{x}) < f(\mathbf{x}_k)\} & \text{otherwise} \end{cases} \tag{3.37}$$

and $f^*_{\mathbf{x}_k} := \min_{\mathbf{x} \in C_{\mathbf{x}_k}} f(\mathbf{x})$, $C_{\mathbf{x}_k} := \{\mathbf{x} \in B : h^{(i)}(\mathbf{x}) - \hat{g}^{(i)}(\mathbf{x}; \mathbf{x}_k) \leq 0\}$.

Observe that the solution to convex optimisation $P_{\mathbf{x}_k}$ lies in set $D_{\mathbf{x}_k} f(\mathbf{x})$ and so can be used in update (3.36). However, set $D_{\mathbf{x}_k} f(\mathbf{x})$ defines a wider class of updates, including suitable approximate solutions to $P_{\mathbf{x}_k}$ that may be easier/faster to find, and generalises the concave-convex approach of [58].

We have the following convergence result.

Theorem 3.4.1 (Local Convergence)**.** *Iterative update (3.36) converges to a stationary point of non-convex optimisation P.*

Proof: See Appendix B. □

By adding a stochastic search component to update (3.36) we can strengthen this to obtain a global convergence result.

Theorem 3.4.2 (Global Optimisation)**.** *Consider iterative update*

$$\mathbf{x}_{k+1} \in \begin{cases} \mathbf{X}_{k+1} & \text{w.p. } \epsilon \\ D_{\mathbf{x}_k} f(\mathbf{x}) & \text{otherwise} \end{cases}$$

$$y_{k+1} = \arg\min\{f(y_k), f(\mathbf{x}_{k+1})\} \tag{3.38}$$

with $\mathbf{x}_1 \in C$ and where $\{\mathbf{X}_k\}$ is a sequence of random variables taking values in C satisfying $Prob(\mathbf{X}_{k+1} \in \Sigma | \mathbf{X}_k \in \Upsilon) \geq \eta\mu(\Gamma \cap \Sigma)$ for any Γ, $\Sigma \subset C$ where $\mu(\Gamma \cap \Sigma)$ denotes the volume of their intersection and parameter $\eta > 0$. Then, y_k converges to the optimum of optimisation P with probability one.

Proof: See Appendix B. □

Update (3.38) is no longer purely a descent update, but rather with probability ϵ an update is made which may lead to the objective f increasing, allowing

escape from unfavourable stationary points (saddle points *etc*) and from local minima. The requirement on random variable $\mathbf{X}_{k+1}$ can be satisfied by, for example, selecting $\mathbf{X}_{k+1}$ uniformly at random within a ball about the current point $\mathbf{x}_k$.

Although the proof of Theorem 3.4.2 could be used to upper bound the convergence, this bound would be very loose and so while Theorem 3.4.2 provides some comfort as to the ability of update (3.38) to find a global optimum, evaluation of the convergence rate really needs to be carried out via numerical experiments.

Note that keeping track of the best x_{k+1} at each step of Update (3.38) doesn't lend itself to distributed implementation and so we drop that step in the proceeding examples (although this also means that we lose the convergence result of theorem 2).

3.4.3.2 Descent Step

Descent step $D_{\mathbf{x}_k} f(\mathbf{x})$ can be realised in many ways. One is by use of a truncated version of the primal-dual subgradient approach:

Algorithm 3.3 $D_{\mathbf{x}_k} f(\mathbf{x})$: Truncated Primal-Dual Updates

do

$$\mathbf{z}(t) = \mathbf{x}(t)$$
$$\mathbf{z}(t+1) = \mathbf{z}(t+1) - \alpha \partial_{\mathbf{z}} L\big(\mathbf{z}(t), \Lambda(t); \mathbf{x}_k\big)$$
$$\Lambda^{(i)}(t+1) = \Big[\Lambda^{(i)}(t) + \alpha \partial_{\Lambda} L\big(\mathbf{z}(t), \Lambda(t); \mathbf{x}_k\big)\Big]^+$$
$$t = t+1$$

until $f(\mathbf{z}(t+1)) \le f(\mathbf{x}_k)$ and $\mathbf{z}(t+1) \in C_{\mathbf{x}_k}$
$\mathbf{x}_{k+1} = \mathbf{z}(t+1)$

While this primal-dual approach might not lead to a feasible point at every step, the requirement that $\mathbf{x}_{k+1}$ be feasible can be relaxed to the requirement that (i) $\mathbf{x}_{k+1}$ is bounded and (ii) $\mathbf{x}_{k+1}$ is feasible for sufficiently large k. Theorem 3.4.1 then continues to hold, with only minor changes to the proof.

3.4.3.3 Example

The following simple example illustrates the convergence of update (3.38) to a global optimum. Consider the optimisation problem

$$\min_{x \in [-4,4]} \quad x^2 - x^4$$

It can be verified that in the interval $[-4, 4]$ this has a global minimum at $x^* = \pm 1/\sqrt{2}$ with $f(x^*) = -0.25$, but it also has a stationary point at $x = 0$.

Fig 3.6(a) compares update (3.38) with update (3.36) when starting from initial condition $x = 0$ (a stationary point). It can be seen that update (3.36) gets stuck at this stationary point whereas update (3.38) is able to escape and find a global minimum. Fig 3.6(b) shows realisations of update (3.38) for a range of initial conditions, illustrating its insensitivity to the choice of initial condition.

(a)   (b)
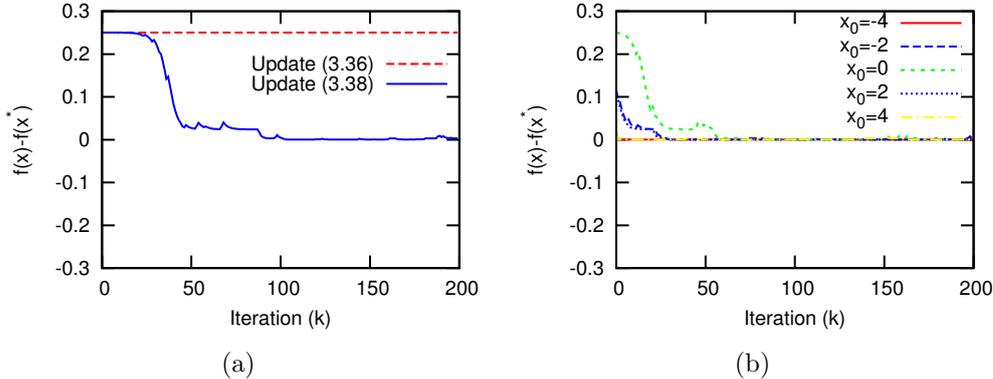
Figure 3.6:   Example illustrating global convergence, $\alpha = 0.1$, $\epsilon = 0.2$.

3.5   Motivating Example Revisited

Armed with our proportional fair problem formulation and non-convex optimisation tools we now revisit the example in Fig 3.1 in more detail.

3.5.1   Network Setup

The network consists of one LTE base-station, one 802.11 AP and two UEs. We assume that the AP uses 802.11n [59] with the settings detailed in Table 3.2 and the aggregate MAC protocol data unit (A-MPDU) frame structure illustrated in Fig 3.7. For simplicity we assume that only two LTE sub-channels are available, denoted sub-band one and sub-band two. We consider LTE FDD systems with 10MHz system bandwidth and uplink transmissions with one data stream from single antenna UEs. The LTE system parameters are also detailed in Table 3.2.

3.5.2   Rate Allocations

Using the maximal convex subset $\bar{\mathbf{x}}_1$ given in Table 3.2, Algorithm 3.2 yields rate allocations of 10 Mbps and 42.4Mbps for UEs u_1 and u_2 respectively. Using update (3.36) to adapt the maximal convex subset the rate allocation improves to be 10 Mbps for UE u_1 and 46.5Mbps for UE u_2. Details of the solutions found using the two approaches are given in Table 3.3.

For comparison, when u_1 and u_2 use only the 802.11 WLAN their data rates are, respectively 0.46Mbps and 24.9Mbps while when the UEs use only the
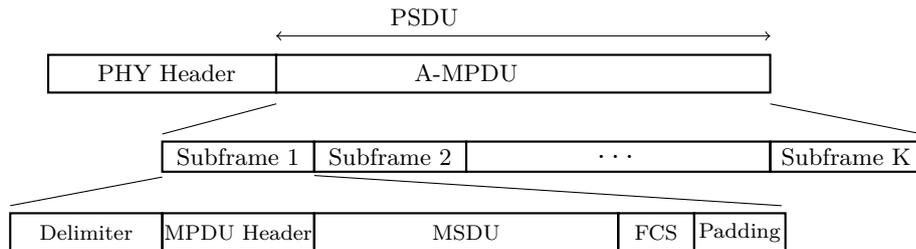
Figure 3.7: 802.11n frame format for Aggregated-Mac Packet Data Units (A-MPDUs).

LTE the BS allocates a rate of 5Mbps to each UE. Splitting the traffic for each UE equally over the 802.11 and LTE networks would yield rates of 5.46Mbps and 29.9Mbps.

3.5.3 Convergence

Figure 3.8 illustrates the convergence of update (3.36) for various choices of initial convex subset, detailed in Table 3.2. The first choice $\bar{\mathbf{x}}_1$ makes use of knowledge of the network to estimate the rate region in which the optimum is likely to lie. The second and third choices $\bar{\mathbf{x}}_2$ and $\bar{\mathbf{x}}_3$ are randomly selected.

For comparison, Figure 3.9 shows the convergence of both updates (3.36) and (3.38). Observe that in this example the extra complexity of update (3.38) does not yield a better optimum, and indeed we also observe this in the other examples presented below (we also carried out further tests, not shown here, that exhibit similar behaviour). This suggests that for the class of optimisation problems considered here update (3.36) tends to converge to a near optimal solution.

3.6 Example Scenarios

3.6.1 802.11 Offload

Our problem formulation can be readily extended to provide a principled approach for offload of data traffic from LTE to 802.11. In an offload setting we would like to preferentially use the 802.11 network where possible. We can capture this requirement by augmenting the proportional fair utility function with a cost related to LTE usage whereby the objective to minimise is given by:

$$-\sum_{u\in\mathcal{U}}\log\left(s_u+r_u\right)+\sum_{u\in\mathcal{U}}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}}\frac{\zeta_{b,u}^i}{\beta_1\omega^i\log\left(1+\frac{\gamma_{b,u}^i}{\beta_2}\right)} \tag{3.39}$$

The added function in the objective associates a cost with the LTE airtime used by each UE (the airtime being inversely proportional to the data rate).

Table 3.2: Network parameters.

Parameter	Value
802.11 Parameters	
Minimum contention window size CW_{MIN}	15
Physical Rate, $(c_{a,1}, c_{a,2})$	$(1, 54)$ Mbps
Physical Rate, $(c_{b,1}, c_{b,2})$	$(5, 1)$ Mbps
Basic Rate R_{basic}	1 Mbps
MAC slot Time σ	$9\mu s$
PLCP Preamble	$16\mu s$
PLCP Header	48 bits @$c_{a,u}$
MAC Header	192 bits @$c_{a,u}$
Payload size $L_{a,u}$, $\forall a \in \mathcal{A}$, $\forall u \in \mathcal{U}$	1500 bytes @$c_{a,u}$
Number of MPDU subframes for u_1, K_1	1
Number of MPDU subframes for u_2, K_2	54
FCS	32 bits @$c_{a,u}$
Delimiter	32 bits @$c_{a,u}$
Padding	32 bits @$c_{a,u}$
ACK	112 bits @R_{basic}
RTS	160 bits @R_{basic}
CTS	112 bits @R_{basic}
LTE Parameters	
Duplex Mode	FDD
System Bandwidth	2×10 MHz
Carrier Frequency	2 GHz
Number of Antennas at BS	1
Number of Antennas at UE	1
Guard Band Overhead	10%
DMRS Overhead	14.3%
Random Access Overhead	0.625%
Cyclic Prefix Overhead	6.66%
Bandwidth Efficiency β_1	71%
SNR Efficiency β_2	1
SNR ,$(\gamma_{1,1}^{(i)}, \gamma_{1,2}^{(i)})$, $i = 1, 2$	$(4.9, 4.9)$ dB
Optimisation	
Maximal Convex Subset 1 $(\bar{\mathbf{x}}_1)$ with $\{(\bar{\rho}_{a,1}, \bar{\rho}_{a,2}), (\bar{w}_{a,1}, \bar{w}_{a,2})\}$	$\{(-1,0),(-2,2)\}$
Maximal Convex Subset 2 $(\bar{\mathbf{x}}_2)$ with $\{(\bar{\rho}_{a,1}, \bar{\rho}_{a,2}),(\bar{w}_{a,1}, \bar{w}_{a,2})\}$	$\{(0,0) ,(0,0)\}$
Maximal Convex Subset 3 $(\bar{\mathbf{x}}_3)$ with $\{(\bar{\rho}_{a,1}, \bar{\rho}_{a,2}),(\bar{w}_{a,1}, \bar{w}_{a,2})\}$	$\{(0,0),(1,1)\}$
α (step size)	0.01
ϵ	0.2

Consider a simple network setup with one BS, one AP and two UEs. Suppose the 802.11 physical rate is 54Mbps for both users and the LTE rates are 4Mpbs (corresponding to an SNR of 0.75dB), see Table 3.4. Other 802.11 LTE parameters are detailed in Table 3.2.

Using update (3.36) to solve the optimisation problem, Table 3.5 summarises the solution found. The UEs share the 802.11 AP and do not send traffic via the LTE network (the LTE rates $r_{u_1} = 0 = r_{u_2}$). The throughput of each UE is 16.5Mbps, which is higher than the data rate of 2Mbps provided by the LTE

Table 3.3: Solutions given by Algorithm 3.2 and update (3.36).

	Algorithm 3.2, Convex subset $\bar{\mathbf{x}}_1$		Update (3.36)	
Parameter	u_1	u_2	u_1	u_2
$\zeta_{b,u}$	1	0	1	0
$\rho_{a,u}$	0.05	0.85	0	0.98
r_u	10	46.5	10	48.8
$p_{a,u,1}$	0.06	0.98	0.01	0.99
$p_{a,u,2}$	0.85	0.01	0.97	0.03
$z_{a,u}$	0	0.93	0.01	0.99

Table 3.4: 802.11 offload example data rates.

	PHY Rates [Mbps]		Technology Rates [Mbps]		
	BS	AP1	LTE only	802.11 only	Optimised Multi-RAT
u_1	4	54	2	16.5	16.5
u_2	4	54	2	16.5	16.5

Table 3.5: Rate allocation for 802.11 offload example.

Parameter	u_1	u_2
r_u	0	0
$\rho_{a,u}$	0.48	0.48
s_u	16.47	16.47
$p_{a,u,1}$	0.046	0.046
$p_{a,u,2}$	0.954	0.954
$z_{a,u}$	0.95	0.95

network. Figure 3.10(a) plots how the fraction $\zeta_{b,u}^i$ of LTE airtime used by each UE changes as the LTE data rate is increased, while Figure 3.10(b) shows the aggregate throughput (LTE plus 802.11) of each UE. It can be seen that for LTE data rates less than 5Mbps the LTE network is not used and data is fully offloaded to the 802.11 network. However, as the LTE data rate increases the LTE network is increasingly used to enhance the user throughputs. Observe also from Figure 3.10(a) that the probability $z_{a,u}$ of associating to the 802.11 network remains almost constant as the use of the 802.11 network always results in an increase in the aggregate throughput. We can readily extend consideration to include situations where UEs have a specified traffic load by adding additional constraints to the optimisation problem. For example, suppose that both UEs have a traffic load of 10Mbps video, the LTE available rate for each UE is 10Mbps (i.e. the total LTE capacity is 20 Mbps, enough to support the traffic load) and the 802.11 physical rate for UE u_1 is fixed at 6Mbps while that of UE u_2 is varied. Using (3.39) to account for the cost of using the LTE connection we expect the traffic to be offloaded to the 802.11

(a) Throughput of UE u_1

(b) Throughput of UE u_2

(c) $z_{a,1}$

(d) $z_{a,2}$

(e) $\zeta_{b,1}^i, i = 1, 2$

(f) $\zeta_{b,2}^i, i = 1, 2$
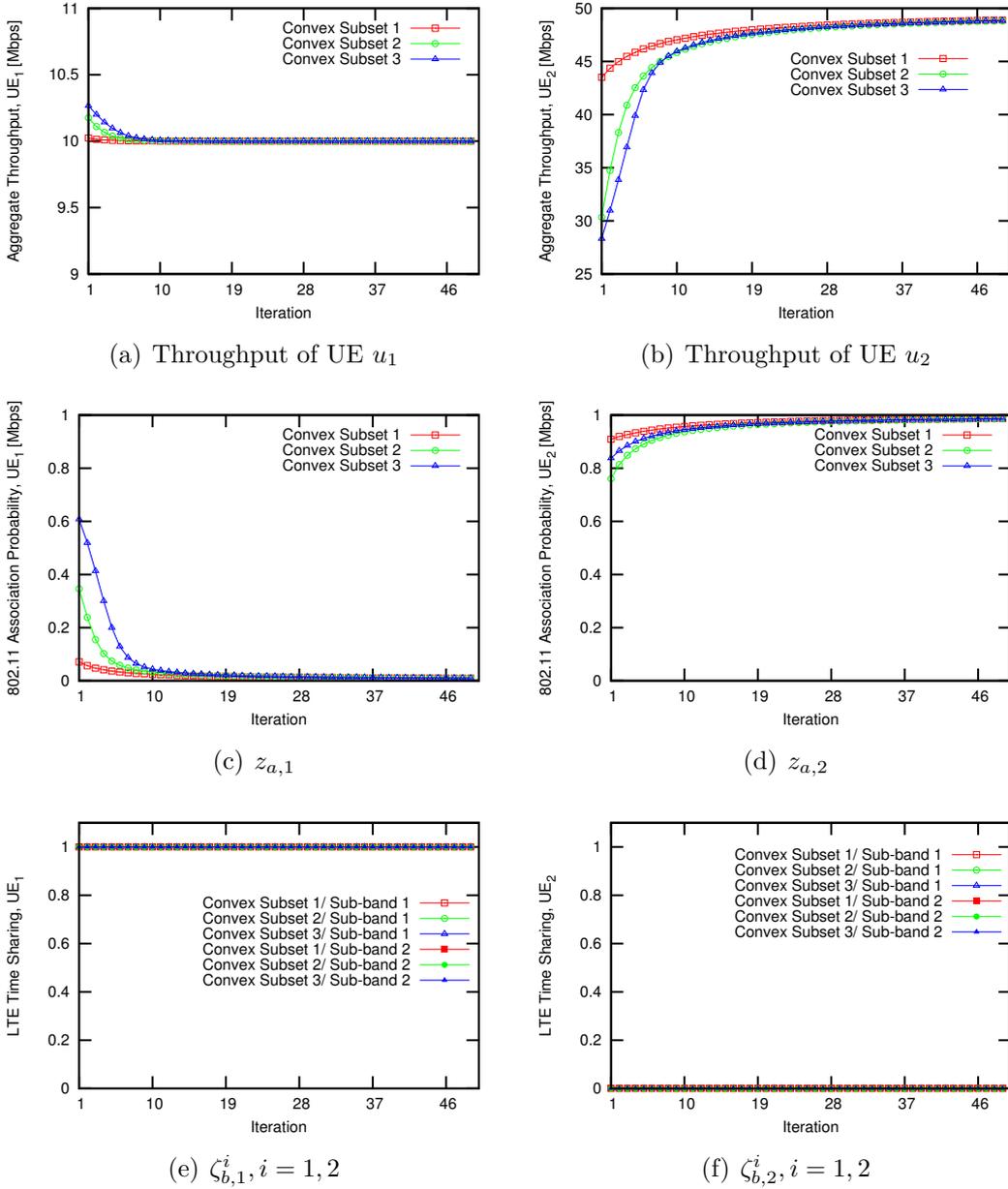
Figure 3.8: Convergence of update (3.36) for various choices of initial convex subset, detailed in Table 3.2.

network as long as its capacity is sufficient to meet the load. Figure 3.11 illustrates the LTE air time for UEs u_1 and u_2 vs the 802.11 rate available to UE u_2. Since the 802.11 rate for u_1 is fixed at 6Mbps, u_1 must use the LTE connection in order to meet its traffc demand and indeed, as expected, it can be seen in Fig 3.11(a) that $\zeta_{b,1}^1$ and $\zeta_{b,1}^2$ (the fractions of LTE airtime used by UE u_1 on each subband) are always non-zero. In contrast, it can be seen from Fig 3.11(b) that the LTE airtime used by UE u_2 falls to zero

(a) Throughput of UE u_1

(b) Throughput of UE u_2

(c) $z_{a,1}$

(d) $z_{a,1}$
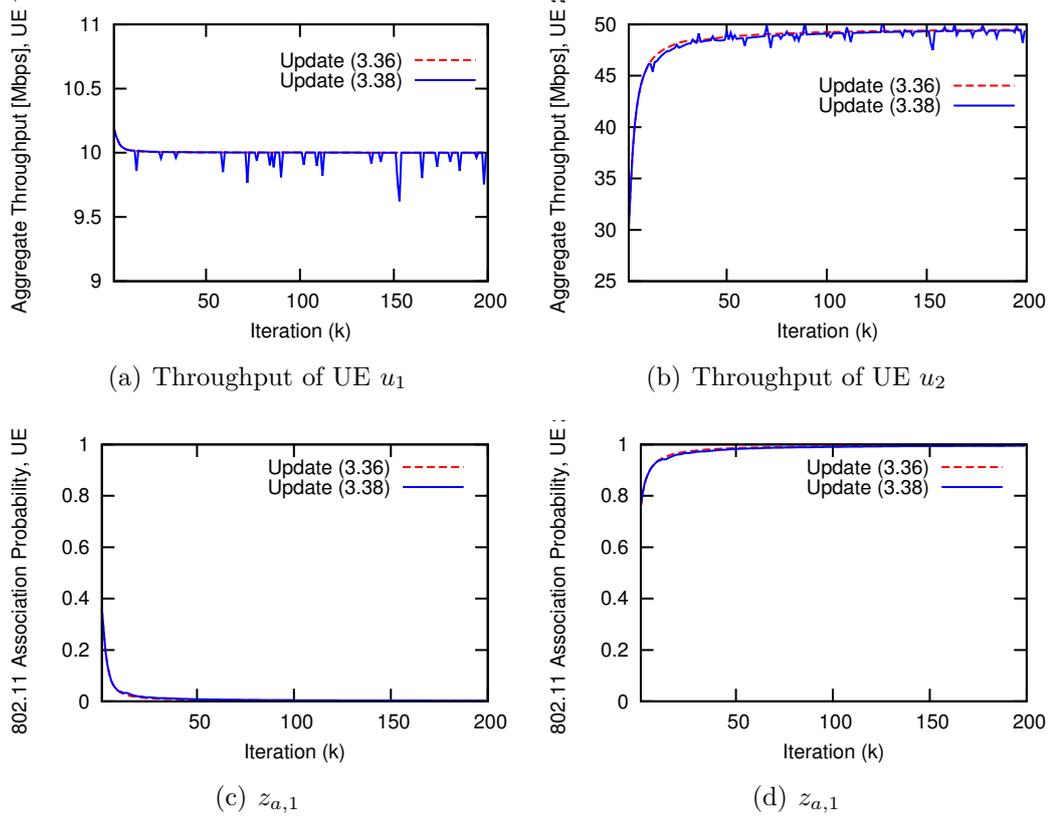
Figure 3.9: Convergence of updates (3.36) and (3.38) when $\mathbf{X}_{k+1}$ is drawn uniformly random within the ball of radius 0.025.

(a)

(b)
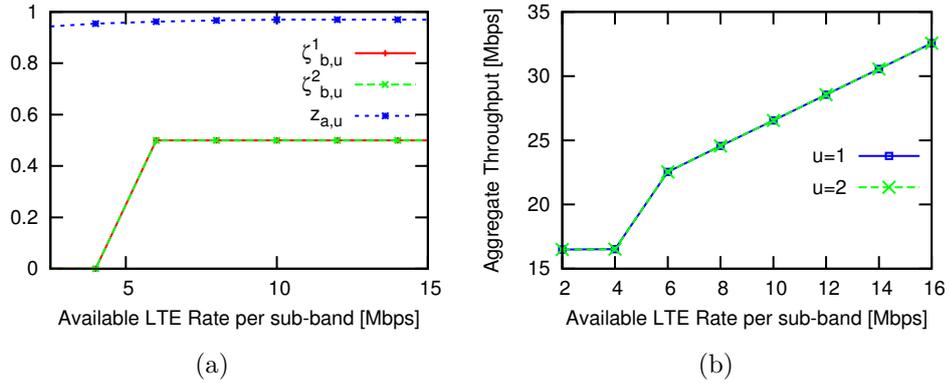
Figure 3.10: Illustrating 802.11 offload as LTE data rate is varied, (a): LTE airtime, (b): Aggregate LTE plus 802.11 UE throughput.

once the available 802.11 rate rises above 10Mbps *i.e.* once sufficient 802.11 bandwidth is available to support the traffic load it is fully offloaded from LTE onto the 802.11 link. Observe also that the 802.11 usage by UE u_1 falls to zero

when the 802.11 rate is 10 Mbps. This allows UE u_2 to make full use of the 802.11 network to meet its traffic demand, without incurring the overhead of collisions. Once the 802.11 rate increases further, the extra capacity is then used by UE u_1.

(a) UE u_1 (b) UE u_2
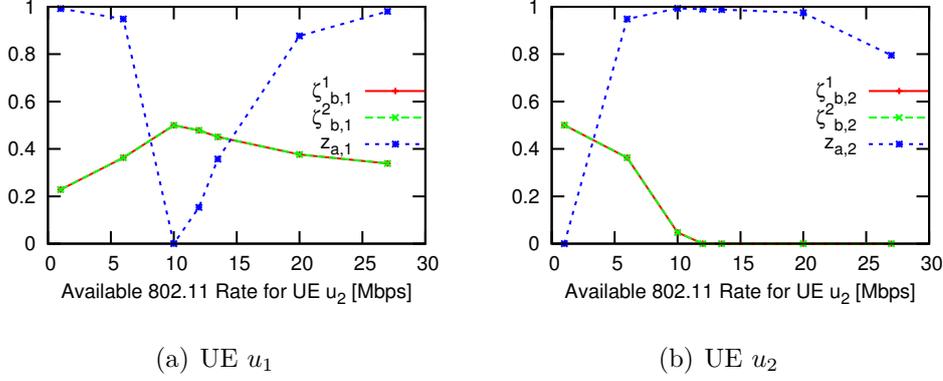
Figure 3.11: Illustrating 802.11 offload as 802.11 rate is varied.

3.6.2 802.11 Multihoming

We next illustrate how our framework can be used to manage 802.11 multihoming. Consider an example with one BS, two APs (a_1 and a_2) and two UEs (u_1 and u_2). The available data rates are summarised in Table 3.6. Each user can individually achieve 10Mbps over the LTE network (and 5Mbps each when sharing the LTE network). The users have a physical rate of 36Mbps and 54Mbps over the first 802.11 WLAN and of 36Mbps and 18Mbps over the second WLAN, so user u_2 has a better connection via AP a_1 and user u_1 will have a better connection via AP a_1 by avoiding contention with user u_2. Note that although 54Mbps link has a higher physical rate compared to that of the 36Mbps link, when the duration of a successful transmission is held constant the achievable throughputs are roughly equal due to the MAC framing and contention overhead.

Table 3.7 details the rate allocation found using update (3.36). It can be seen that UE u_1 is mainly associated with AP a_2 ($z_{a_2,u_1} = 0.99$) and u_2 with AP a_1 ($z_{a_1,u_2} = 0.99$), as might be expected in view of the link characteristics. The LTE network is shared equally by both users ($\zeta_{b,u_1} = 0.5 = \zeta_{b,u_2}$).

Figure 3.12 shows how the allocation changes as the capacity of the 802.11 link between AP a_2 and UE u_2 varies. It can be seen that as the rate on the a_2-u_2 link increases UE u_2 increasingly makes use of this link ($z_{a_2,2}$ increases while $z_{a_1,2}$ falls) and conversely traffic for UE u_1 increasingly makes use of AP a_1 ($z_{a_1,1}$ increases while $z_{a_2,1}$ falls) and of the LTE network ($\zeta_{b,1}^1$ and $\zeta_{b,1}^2$ both rise). When both UEs have equal rates of 36Mbps and 54Mbps through APs a_1 and a_2 respectively, they make use of APs equally

Table 3.6: 802.11 multihoming example data rates.

	PHY Rates [Mbps]			Technology Rates [Mbps]			
	BS	AP a_1	AP a_2	LTE	WLAN a_1	WLAN a_2	Optimised Multi-RAT
u_1	10	36	36	5	5.9	8.9	22.39
u_2	10	54	18	5	8.87	4.3	22.39

Table 3.7: Rate allocation for 802.11 multihoming example.

Parameter	u_1	u_2
$\zeta_{b,u}$	0.5	0.5
$\rho_{a_1,u}$	0.0045	0.92
$\rho_{a_2,u}$	0.94	0.0037
r_u	5	5
s_u	17.39	17.4
$p_{a_1,u,1}$	0.007	1
$p_{a_2,u,1}$	1	0.007
$z_{a_1,u}$	0	0.99
$z_{a_2,u}$	0.99	0

(a) UE u_1 (b) UE u_2
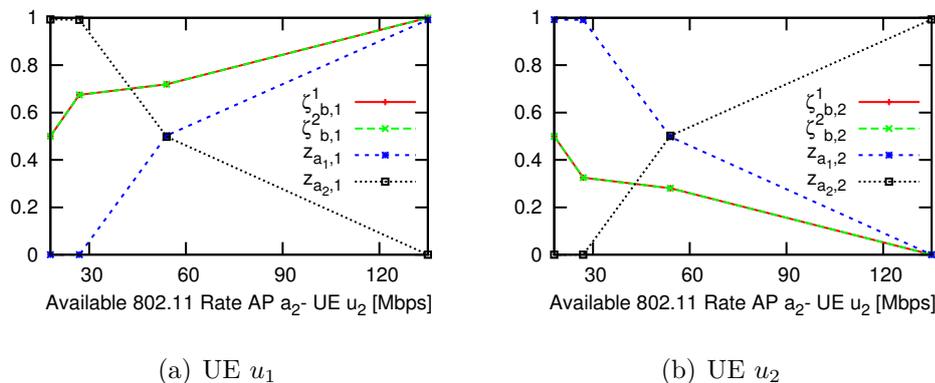
Figure 3.12: Illustrating 802.11 multi-homing as the rate on the UE u_2- AP a_2 802.11 link is varied.

3.6.3 LTE Multihoming

We next consider an example with two BSs (b_1 and b_2), one AP, and 4 UEs (u_1, u_2, u_3, and u_4). The physical rates available on the 802.11 and LTE links are summarised in Table 3.8. We also consider the conventional user assignments to LTE BSs *i.e.* each UE associates with the LTE BS that provides the UE with the maximum received power. The achievable LTE rates when using this maximum received power assignment are also summarised in Table 3.8. Figure 3.13 shows the proportional fair LTE sub-band allocations and 802.11

Table 3.8: LTE multihoming example: data rates.

	PHY Rates [Mbps]			Technology Rates [Mbps]				
	BS b_1	BS b_2	AP	LTE b_1 only	LTE b_2 only	LTE (Maximum Rx Power)	802.11 only	Optimised Multi-RAT
u_1	26	25	54	6.5	6.25	8.27	7.51	15.36
u_2	10	25	27	2.5	6.25	2.4	3.75	12.5
u_3	5	29	54	1.25	7.25	3.56	7.51	14.8
u_4	11	10	13.5	2.75	2.5	2.29	1.88	5.5
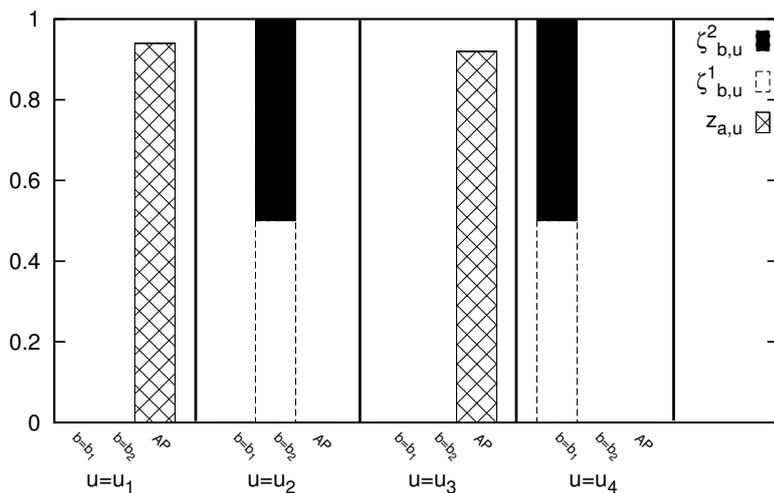
Figure 3.13: LTE multihoming, user association probabilities per sub-band of the LTE BS ($\zeta_{b,u}^i$) and per 802.11 AP ($z_{a,u}$) given by Update (3.36).

association probabilities obtained using update (3.36). From Table 3.8 it can be seen that a relatively large physical rate of 54 Mbps is available to UEs u_1 and u_3 and so it can be seen from Figure 3.13 that in the proportional fair allocation these UEs make use of the 802.11 network but not the LTE network, so freeing up capacity in the LTE network for UEs u_2 and u_4. UEs u_2 and u_4 make use of the LTE rather than the 802.11 network (so reducing collisions and increasing the 802.11 capacity available to UEs u_1 and u_3), and share the LTE network evenly between them.

It can be seen from the right-hand column of Table 3.8 that all users benefit from this use of multi-homing. In addition observe that a simple aggregation of both LTE (Maximum Power association) and 802.11 resources results in 13% reduction in the proportional fair rate objective compared with that of a near-optimal solution.

3.7 Conclusions

We consider proportional fair rate allocation in a heterogeneous network with a mix of LTE and 802.11 cells which supports multipath and multihomed operation (simultaneous connection of a user device to multiple LTE BSs and

802.11 APs). We show that the utility fair optimisation problem is non-convex but that a global optimum can be found by solving a sequence of convex optimisations in a distributed fashion. The result is a principled approach to offload from LTE to 802.11 and for exploiting LTE/802.11 path diversity to meet user traffic demands.

Dynamic Power Adaptation in LTE/3G Small Cell Networks

In this chapter we propose a load aware approach for dynamic idle mode selection. Our approach allows pilot signals and most of the processing power of a SCBS to be completely switched off when no active user is in its vicinity, or when the users may be served by the underlying macro base stations. We evaluate the efficiency of our approach using field measurements in central Dublin as well as with system-level simulations.

4.1 Introduction

It is predicted that global mobile data traffic will increase by 17-fold between 2012 and 2017 [60]. In order to meet this traffic demand, the cellular network architecture is shifting towards the use of a heterogeneous approach comprising a mix of macro and small cells [21]. The number of deployed small cells is to increase from 2.3 million at the end of 2010 to 54 million by 2015 [61]. In contrast to macro BS deployment, SCBSs are deployed with a higher degree of randomness since they can be installed by the end user or in an opportunistic manner. Consequently in a large scale deployment, it is increasingly critical to address the following issues:

- Energy Saving: It is well established that offloading the macro base station load can potentially yield significant energy savings [62]. However the deployment of millions of small cells raises concerns about the aggregate power consumption of those BSs. For example, an average of 12W operation power for each small cell equates to power consumption of 105.12 kWh per year; resulting in a consumption of 5.25×10^9 kWh per year for 50 million small cells. The CO_2 emissions of a dense deployment

can also become significant depending on the type of power source used. Further, provisioning of power and back-haul for dense deployments is in itself a major challenge for network operators [63]. A potential solution is to use wireless back-haul and alternative energy sources, e.g. solar energy [64], in which case it is essential for small cells to efficiently manage their power consumption. While in sleep mode small cells have reduced power usage, depending on the hardware architecture and sleep mode algorithms [65] the reduction is typically limited to no more than 30% to 60% of the total power consumption. To achieve the largest energy savings, it is necessary to place the small cell base stations into deep sleep mode where the energy consumption is essentially zero.

- Interference and Pilot Pollution: SCBSs can cause interference to the users of the macro BSs as well as interference to the users of the other SCBSs, resulting in undesirable impacts on the performance. Moreover denser small cells result in pilot pollution where multiple strong transmitters are detected by a UE. As described in Chapter 1, eICIC methods are developed in releases 10 and 11 of 3GPP [12] to enhance interoperability between the layers. eICIC encompasses CRE and ABS modes by introducing more efficient cell selection [66] and interference mitigation mechanisms receptively. However putting SCBSs into sleep mode when no pilot power is transmitted, potentially provides an additional means of interference management.

- Signalling Overhead: In busy areas significant signalling overhead can be introduced by the handover/hand-off messages of passers-by in the vicinity of a small cell. Further enhancements in release 12 of the LTE specification address these problems. For example dual connectivity modes allow simultaneous connection of the user to both macro and small cell layers, enabling separation of the data and control messages, where the control signalling is provided by the macro layer and high speed data is provided by the small cell layer. Once again, appropriate control of the small cell powers potentially provides an additional mechanism to manage signalling overhead.

Although each of the issues above can and must be addressed separately, it should be noted that the traffic load demand plays a critical role in defining dynamic configurations of the network. It is known that traffic demand fluctuates over time [67] and is not static. Moreover the spatial distribution of the mobile traffic load may be clustered and vary over time.

In this chapter we consider a dynamic approach towards power control of small cells, whereby certain subsets of small cells may be de-activated (i.e. completely switched off) if no user is detected in their vicinity, while other small cells remain active to serve their nearby users.

To determine the activity of small cells, two natural solutions may be envisioned: 1) To keep small cells in idle mode at all times so that they can detect user activity in their vicinity by sniffing the channel [68]. 2) To report the (GPS measured) geographical location of the users to a central network entity and use this information to wake up small cells if needed. A sniffer based wake up procedure may not be desirable since:

- It uses electrical power and circuit functionality for detecting UEs at all times. The idle mode electrical power consumption typically contributes 30% to 60% of the total power consumption of a small cell [65].

- It may result in waking multiple neighbour BSs up due to the lack of coordination among the small cells.

Detecting nearby users by reporting GPS coordinates, is also problematic for the following reasons:

- GPS functionality is not available in indoor locations.

- Switching on the GPS functionality results in additional energy consumption of the mobile UEs.

- Reporting the coordinate values to the BSs leads to increased signalling overhead and also raises privacy concerns.

In this chapter we therefore determine the relative position (not the exact coordinates) of a user to the small cells. This can be achieved by using readily available received power measurements at the UEs. In ideal channel conditions (no fading, no noise and no receiver sensitivity to the angles of arrival), a unique set of measurements will identify an individual measurement location. However in practice, these measurements may not be unique. We therefore adopt a classification approach to map each set of measurements to a serving BS (a macro or a small cell). Data is initially collected with all small cells switched on so that appropriate serving cells can be determined. New measurements then can be compared against this baseline. This concept is illustrated in Fig 4.1.

4.2 Related Work

Although substantial work has been done to reduce energy consumption in the macro base stations, from improving base station hardware design [64], to power control mechanisms [69], and use of alternative energy sources [64], introduction of the small cells has been shown to result in significant power reduction gains [62]. In [11] the authors detail the challenges and energy efficiency concerns in large small cell networks. They further show how a
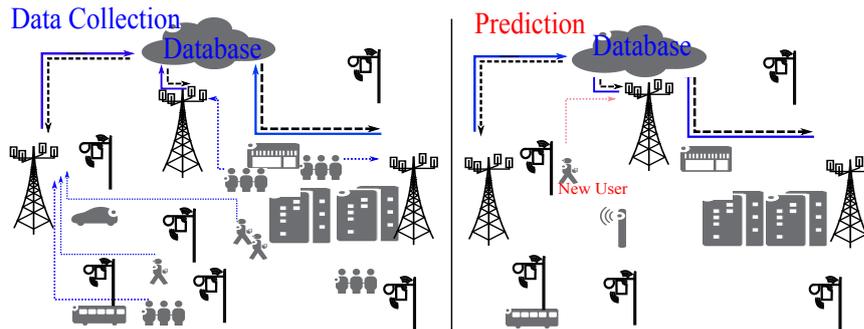
Figure 4.1: Idle mode control using received power measurements.

large system analysis (of small cells) based on random matrix theory is useful for the performance prediction of cooperating groups of small cells. Idle mode control procedures for small cells however, have received relatively little attention in the literature to date. In [65] the authors propose sleep mode algorithms by enabling dynamic adjustment of the small cell transmit powers. In [68], the authors propose distributed procedures for idle mode control, where each small cell base station remains in idle mode unless it detects user activity in its vicinity. A small cell determines user activity by sniffing the uplink channel between a UE and its serving macro BS. Multiple wake ups, pilot pollution and plug-in power consumption are potential disadvantages of these distributed methods. In release 11 of the 3GPP [70], a number of waking up procedures are proposed for hotspot cells. By our proposed dynamic idle mode control procedure, we complement these wakeup based approaches by reducing pilot transmissions and by enabling complete shut-down of the small cells. To estimate the topology of the network load and to control idle modes dynamically, we make use of the received power vectors or the "RF Fingerprints" reported by the UE.

4.3   System Setup

Let $\mathcal{B}$ denote the set of base stations, where $\mathcal{B}_m \subset \mathcal{B}$ is the set of macro BSs, and $\mathcal{B}_{sc} \subset \mathcal{B}$ is the set of SCBSs. We denote the set of users in the geographic area of $\mathcal{B}$ by $\mathcal{U}$. We let $\mathcal{L}$ denote the set of geographical locations that contain the received power measurement records.

4.3.1   RF Fingerprints

We describe the vector of the received pilot powers from nearby macro BSs as the "RF Fingerprints" vector and denote it by X_l (note that each RF fingerprint vector is associated with a set of measurements in a distinct location, l). Here nearby refers to BSs for which received pilot power is above the detection level of the UE. As a routine part of their operation, UEs periodically send measurement reports of the received pilot powers. In addition, a BS can

explicitly request UEs to perform and report measurements for making handover decisions. Hence RF-fingerprints can be collected without additional overhead or operational cost.

We collected field measurements of RF fingerprint vectors in central Dublin outdoor locations in June 2014. We equipped an Android mobile device with a major network provider's 3G sim card and collected the measurements by war-driving. We modified an open source Android App [71] to collect signal strength measurements from neighbouring cells in addition to collecting the geographical coordinates. Note that we will only use the geographical coordinates for validation purposes. Figure 4.2 shows a sample path where the measurements were taken. Measurements were taken with an average walking speed of $3km/h$ and at $1s$ intervals i.e. the measurement points are roughly 1m apart. Table 4.1 illustrates a sub-sample of the central Dublin field mea-
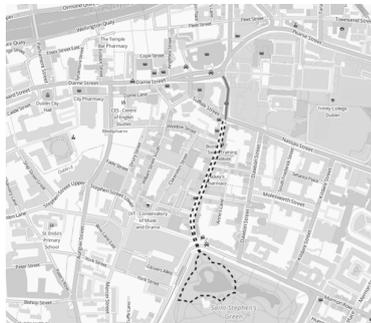
Figure 4.2: RF fingerprint measurement path

surements in the form of RF fingerprint vectors.

Table 4.1: Sub-samples of RF fingerprint pilot powers in dBm where each dimension is marked by a distinct Primary Scrambling Code, PSC and each distinct measurement point is identified by a geographical coordinates in the form of (latitude, longitude).

Location\PSCs	212	204	252	300	120	45	236	292
(53.3400988,-6.2607508)	-	-69	-53	-	-105	-91	-68	-65
(53.3401079,-6.2607396)	-	-	-	-	-	-	-	-63
(53.3401169,-6.2607290)	-	-	-	-	-	-	-	-63
(53.3401227,-6.2607128)	-	-	-51	-	-	-	-	-
(53.3401297,-6.2607026)	-	-	-51	-	-	-	-	-
(53.340137,-6.2606895)	-	-	-51	-	-	-	-60	-59

4.3.2   Classification

In this section we first describe the state of the art cell selection rules that are commonly used in cellular networks, we then proceed to use this approach for the RF fingerprint records, in order to determine the serving cells associated to each RF fingerprint vector.

4.3.2.1 Cell Selection

As described in Section 1.1.1.1, a user to cell assignment is determined by the levels of the received signal powers. In more detail, a limited number of the LTE sub-channels over the available bandwidth and time slots are allocated to what are called reference signals. Here we refer to these signals as pilot signals. The linear average of the received pilot powers within the measurement bandwidth over a defined measurement period (referred to as the reference signal received power- RSRP) is then used to specify the serving cells [7, Section 22.3.1]. Considering power adaptation in this chapter, we use separate notation for the pilot and total transmit powers of a BS, denoting these powers for a given BS b by p_b^p and p_b respectively. Also note that within UMTS systems, the received signal code power (RSCP) that is measured over the physical layer measurement period is equivalently used to identify the serving cells [7, Section 22.3.1].

When all small cell base stations are active, a user is scheduled to a base station according to one of the following rules (See Chapter 1 and [7, Section 3.3.3]):

- Signal Strength: Maximum received pilot power:

$$b_u \in \arg\max_{b \in \mathcal{B}} p_b^p h_{b,u} \qquad (4.1)$$

- Signal Quality: Maximum pilot SINR:

$$b_u \in \arg\max_{b \in \mathcal{B}} \gamma_{b,u} \qquad (4.2)$$

where

$$\gamma_{b,u} = \frac{p_b^p h_{b,u}}{\sigma_n^2 + \sum_{k \in \mathcal{B} \setminus \{b\}} p_k h_{k,u}} \qquad (4.3)$$

Observe that with the obvious abuse of notation, in this chapter we let $h_{b,u}$ and $\gamma_{b,u}$ denote the averaged over bandwidth values of the received signal power and SINR between base station b and user u.

When small cells are dynamically activated, the cell selection rules above will not take into account all the existing BSs (i.e. $\forall b_i \in \mathcal{B}, \ i = 1, .., |\mathcal{B}|$). This is due to the absence of the pilot signals from the deactivated BSs. For example consider a previously deactivated small cell in the vicinity of a recently arrived UE. In this situation directly applying Eq (4.1) or Eq (4.2) results in a selection which may not provide the UE with the highest signal quality and therefore may lead to decreased performance.

In our approach instead we make use of an existing dataset of RF fingerprints (a set of $\{X_l\}$), to estimate the serving cell for each user. Each vector X_l is

associated with a class label, namely the UE's serving BS at location l. We denote this serving cell by b_l. For a given RF fingerprint vector (X_l), the serving cell can be calculated using Eq 4.2. We denote the collection of RF fingerprint vectors X_l together with their corresponding class label b_l by $\mathcal{T}$ and refer to it as the training dataset. Note that the class labels i.e. the serving cells for each RF fingerprint vector are calculated according to Eq (4.2) with the knowledge of SINR from $\forall b_i \in \mathcal{B},\ i = 1, .., |\mathcal{B}|$ [1].

4.3.2.2 Deployment of Small Cells

At the time of the measurements detailed in Section 4.3.1, we did not detect small cells in the measurement area. We therefore introduced simulated small cells and calculated the received pilot powers from these small cells by using standard urban path loss models [72] (we will use this data to determine the serving cells in the original dataset). This allows us to analyse a number of deployment scenarios. We briefly summarise the scenarios that we use in this chapter:

(i) Six small cells are located in the measurement area. Each small cell is co-located with the areas where clusters of users tend to form. Co-locating small cells with these clusters can be expected to maximise the scope for macro cell offload. The location of the small cells i.e. the white circles and the received power map from the SCBSs are illustrated in Fig 4.3. Note that the received powers are in dBm scale.
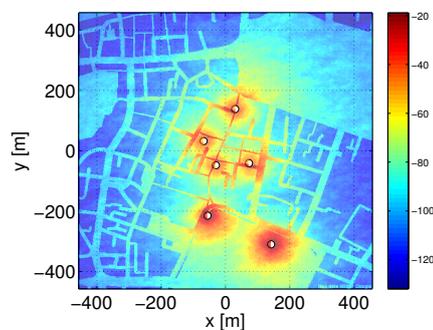
(a)

Figure 4.3: Received power map from SCBSs located at busy areas

---

[1] SINR values from all candidate cells at location l, can be directly calculated from the RF fingerprint vector. In more detail, the useful received signal power of BS b is derived from the bth element of the RF fingerprint vector X_l, while the total interference is the sum of the received signal powers from all other candidate cells $c \in \mathcal{B} \setminus \{b\}$. When calculating the signal quality, we consider use of blank subframes by the macro cells so that interference from the macro to neighbour small cells is minimal. Also note that are simulation have shown similar results when cell selection rules in Eq (4.1) are used.

(ii) Small cells are equally spaced within the measurement path of Fig 4.2. These might be deployed for example on lamp posts in busy urban areas. We consider this scenario, to analyse the effect of changing density of small cells by varying the number of small cells and by keeping the distance between them equal.

(iii) Four Small cells are deployed at the edge of each macro cell i.e. radius R from the macro cells. Placing small cells at the edge of the macro cells requires knowledge of macro cell locations. This information cannot be collected by the mobile UE. But we obtained this data from [73]. We also calculate the received pilot powers from the macro cells using an urban path loss model. This scenario is illustrated in Figs 4.4(a) and 4.4(b). Edge deployment of small cells is favourable in that they may result in improved performance gains for the edge users by offloading from the macro load. To evaluate energy efficiency under a range of topologies, in Section 4.4.2, we vary the radius R.

(a) $R = 40m$ (b) $R = 100m$
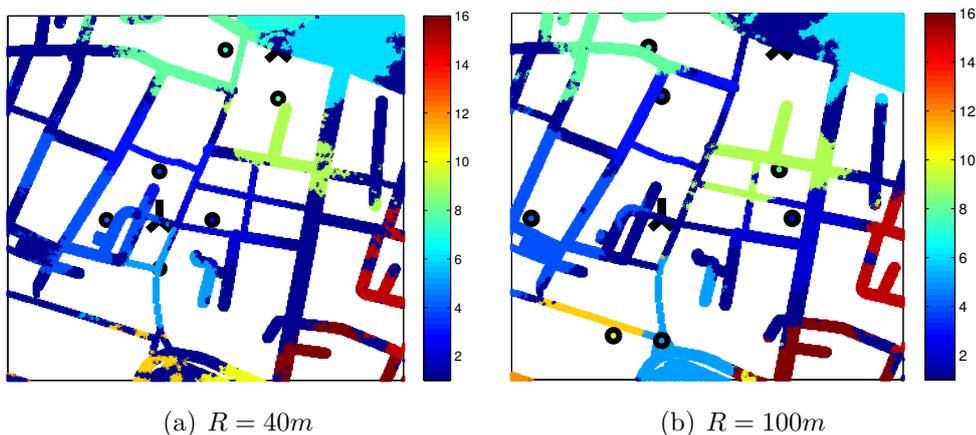
Figure 4.4: Estimated serving cell where SCBSs are positioned in radius R from the centre of each macro BS. Colour map indicates the cell ID, all macro BS are identified by ID 1 and SCBSs are numbers 2 to 17. Black 3-sectors icon and small coloured-circles represent the MBS and SCBS respectively.

4.3.2.3 Predictions

We denote a classifier by $\mathcal{F}$. The task of the classifier $\mathcal{F}$, is to use the training data $\mathcal{T}$, to predict the cell association of a query fingerprint vector, $Y_{l'}$. We use the terms observation and test vectors to refer to the fingerprints within the dataset $\mathcal{T}$, and the query fingerprints respectively. Since the RF fingerprint vectors may contain missing entries, we use the following two stage classification approach.

We define a set of neighbouring BSs, $\mathcal{N}_i$ for the RF fingerprint vector X_i. Recall that each RF fingerprint vector relates to a distinct location with a number

of detectable BSs i.e. $|\mathcal{N}_i|$. We first find the n_{jac} most similar observations to the test data in terms of their neighbouring BS sets. This is achieved by applying the "Jaccard" similarity index. Then we apply a K-Nearest Neighbour (KNN) search on the remaining n_{jac} observations.

Algorithm 4.4 Two stage classifier

---

(1) Find Jaccard similarity index for observation-query pairs:
while $i \leq |\mathcal{T}|$ **do**
  $JS(Y_u, X_i) = \frac{|Y_u \cap X_i|}{|Y_u \cup X_i|}$
end while
 Sort observations based on their Jaccard index
(2) Perform 1 norm K-Nearest Neighbour search on first n_{jac} entries of $JS(Y_u, \mathcal{T})$

The test vector is classified by majority vote of the K neighbours. Finding the K-nearest observations is typically carried on by calculating the pair-wise Euclidean distance between the observations and the test vector. To reduce sensitivity to abrupt changes in the RSRP or the RSCP values, we instead use the 1 norm distance. It is also common to apply weights to the votes to reduce error in KNN classifiers. In this work we consider weights proportional to the inverse of the 1 norm distance for each test-observation pair.

4.4   Performance Evaluation

To evaluate performance, we use field measurements from Dublin city centre in addition to synthetic data for small cells as described in Section 4.3.2.2. We first consider deployment scenario (i), where SCBSs are placed deterministically in busy areas as shown in Fig 4.3. The SCBS parameters can be found in Table 4.2 (the MBS parameters in Table 4.2 are only used later in Section 4.4.2).

Table 4.2: Dynamic power control, simulation parameters

Parameter	Value	Parameter	Value
MBS Tx Power	$46dBm$	SCBS Tx Power	$30dBm$
MBS Ant. Height	$25m$	SCBS Ant. Height	$10m$
MBS Ant. Gain	$15dBi$	SCBS Ant. Gain/Type	$2.15dBi$/Dipole
MBS Ant Tilt	$10°$	UE Ant. Height	$1.5m$
UE Ant. Gain/Type	$0dBi$/Omni	UE Noise Figure	$9dB$
Path Loss model	[72]	Shadow fading SD	$6dB$
P_0	13.6	Δ	4

4.4.1   Cross Validation and Misclassification Error

We consider the following three classifiers:

(a) A classifier which performs a 1 norm K-NN search through the observations in the training set with common neighbouring cells i.e. $\{X_i \in \mathcal{T} \mid \mathcal{N}_u \cap \mathcal{N}_i \neq \emptyset\}$.

(b) A classifier which performs a 1 norm KNN search on the entire training set.

(c) The two stage classifier proposed in Section 4.3.2.3.

We consider performance evaluation for a range of K values. To determine the misclassification error, we apply a 10-fold cross validation to the entire measurement set. The misclassification error is defined as follows. Let $\hat{\mathcal{F}}(X)$ and $\boldsymbol{b}$ denote the predicted and target cell association vectors respectively. A loss function is defined as a mismatch between the classifier's predictions and the target values:

$$\mathrm{P}(\boldsymbol{b}, \hat{\mathcal{F}}(X_u)) = \mathbb{1}_{(\boldsymbol{b} \neq \hat{\mathcal{F}}(X))} \tag{4.4}$$

where $\mathbb{1}_{(x)}$ is the indicator function, taking value 1 when conditional x is true 0 otherwise. Hence the generalisation error can be defined as the prediction error over an independent test sample:

$$Err_{\mathcal{T}} = \mathbb{E}_X[\mathrm{P}(\boldsymbol{b}, \hat{\mathcal{F}}(X)) \mid \mathcal{T}] \tag{4.5}$$

Expected prediction error, on the other hand averages over everything that is random including the randomness in the training set that produced $\hat{\mathcal{F}}$ [74]:

$$Err = \mathbb{E}[\mathrm{P}(\boldsymbol{b}, \hat{\mathcal{F}}(X))] = \mathbb{E}[Err_{\mathcal{T}}] \tag{4.6}$$

Here we referred to $Err_{\mathcal{T}}$ as the misclassification error of classifier trained on $\mathcal{T}$, and Err as the expected misclassification error.

Figure 4.5 shows the misclassification error of the three classifiers when a 1 norm $K = 1$ nearest neighbours search is used. It can be seen that a misclassification error of 1.9% is achieved using classifier (c) and the complete data set. Figure 4.5 also indicates that the performance improves for all classifiers with increasing size of the training data. The proposed two stage classifier slightly enhances the performance compared to the other classifiers. As the size of the training data decreases, all classifiers perform similarly. This is related to the correlation of the fingerprint samples: measurement records suggest that there can be abrupt changes in fingerprint vectors which are a few tens of meters apart. A decrease in the size of the training data implies more dissimilarities between the training set and validation data and hence a larger mean and variance in the prediction error.

Figures 4.6 and 4.7 show the misclassification error versus size of the training data for classifiers (b) and (c) respectively and when number of nearest neighbours (K), is varied from 1 to 9. It can be seen that increasing the number
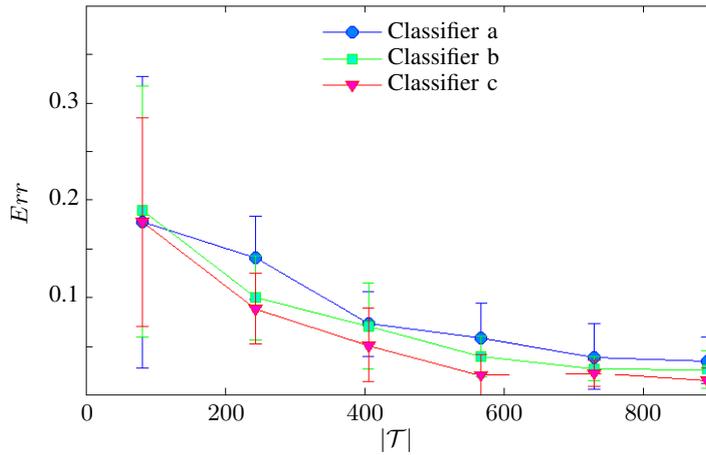
Figure 4.5: Misclassification error when K=1 nearest neighbour is used with classifiers (a), (b), and (c) and versus size of the training data, $|\mathcal{T}|$.
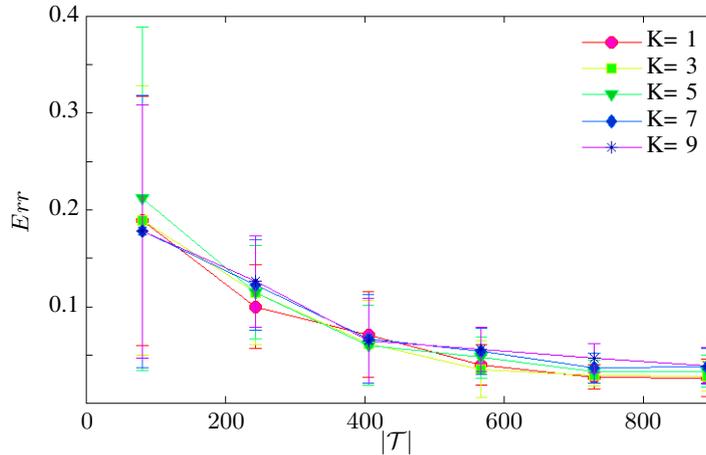
Figure 4.6: Misclassification error of the classifier (b) with KNN versus size of the training data, $|\mathcal{T}|$.

of neighbours does not necessarily improve the performance of the classifier. Moreover with decreasing size of the training data, the 1 norm $K = 1$ nearest neighbours method outperforms the 1 norm K-nearest methods with $K > 1$, for the same reason of decreased similarity among observations and test data.

To evaluate the classification error for other network configurations, SCBSs are are located at equally spaced positions, across the measurement path as detailed in deployment scenario (ii) in Section 4.3.2.2. Figure 4.8 shows the misclassification error for a varying number of SCBSs and with a complete training set. Each box plot represents the prediction error for a number of $|\mathcal{B}_{sc}|$ SCBSs. The boxes have lines at the lower quartile, median, and upper
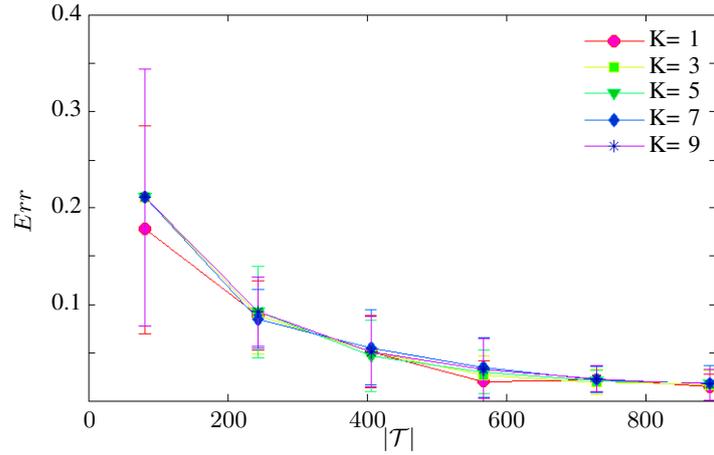
Figure 4.7: Misclassification error of the classifier (c) when KNN used at the second stage, versus size of the training data, $|\mathcal{T}|$.

quartile values. It can be observed that the misclassification error is increasing with increasing density of the SCBSs. By increasing the number of small cells, a larger fraction of the observations lie at the edge of two or more neighbouring cells. At these locations, there may no longer exist a dominant serving BS. In this situation, RF fingerprints at the edge areas and only few meters apart (similar vectors) may be assigned to different class labels. Hence test (validation) points at these locations are more likely to be misclassified.
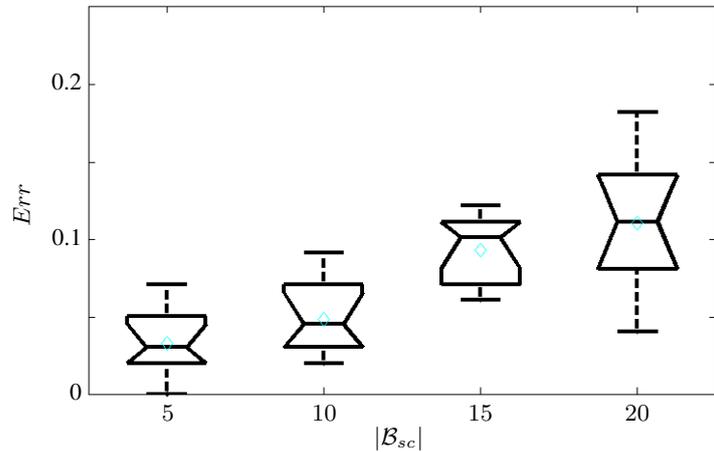
Figure 4.8: Misclassification error with Classifier (c)/$k = 1$ nearest neighbour and complete training set versus number of SCBSS, $|B_{sc}|$.

The observed errors can be split into three categories:

73

- User should be served by MBS but it is classified as being served by a SCBS.

- User should be served by SCBS but it is classified as being served by a MBS.

- User association is misclassified between different SCBSs.

The type of misclassification error depends on the configuration of the network i.e. the density of the small cells and their relative positions to the macro cells. For example when small cells are located closer to the centre of a macro BS, the misclassification probability between the macro and small cell users increases. To assess the different types of misclassification, we consider SCBS deployment scenario (ii) in Section 4.3.2.2, also described above. Figure 4.9 illustrates the split of misclassification errors as the density of SCBSs is varied. It can be seen from the figure that the probability of misclassification among
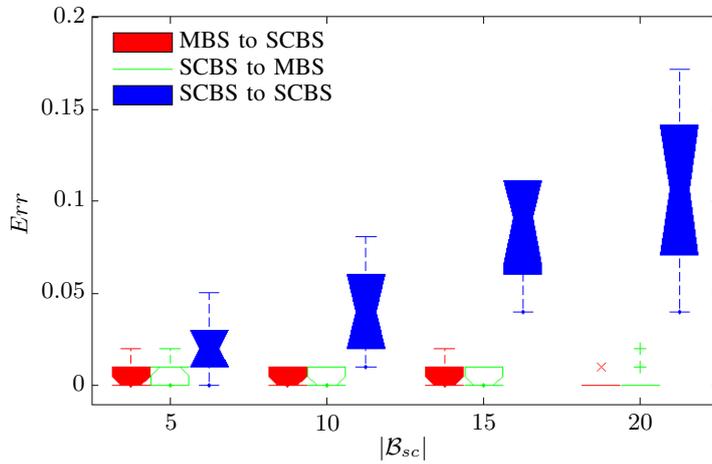
Figure 4.9: Different types of misclassification with Classifier (c)/$k = 1$ nearest neighbour versus number of SCBSs, $|\mathcal{B}_{sc}|$.

SCBSs is relatively higher than other types of misclassification error. This becomes more pronounced with increasing the density of the SCBSs.

4.4.2 Evaluation Using Synthetic Data Records

Recalling that our interest is in dynamic control of small cell idle modes, in this section we assess the potential energy efficiency gains that may be achieved using our proposed classification method. Figure 4.10 illustrates the normalised traffic load within the simulation area and in a 24 hours period. The temporal load data illustrated in Fig 4.10 is obtained from a major provider in Ireland [75]. This is the same provider that we used for the field measurements. However a shortcoming of the field measurements is that they are only collected
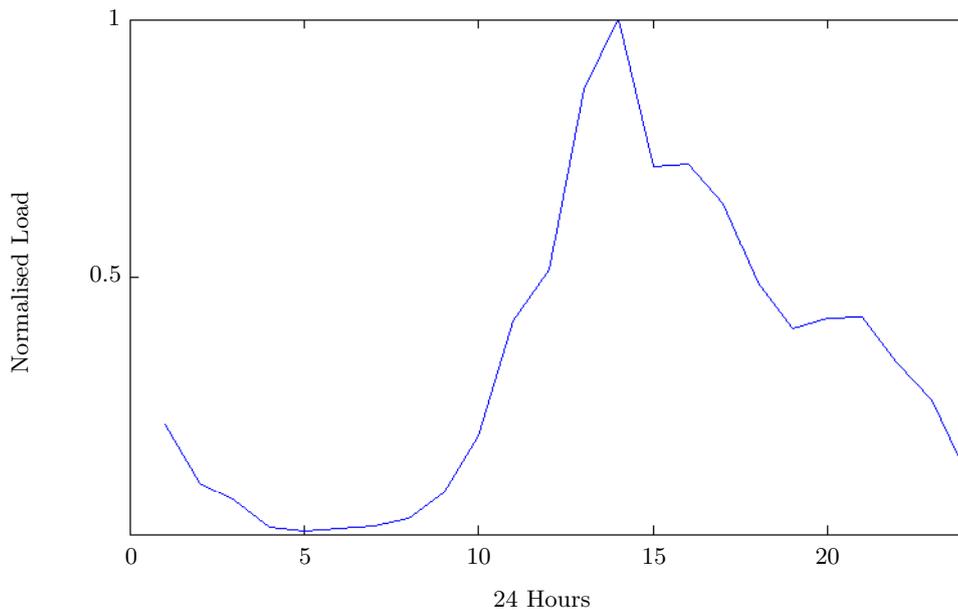
Figure 4.10: Normalised daily traffic demand

by a single user device. Therefore, although we have access to peak temporal variations by [75], we do not have information on the spatial distribution of the RF fingerprint vectors by different UEs and at diffident time instances. As a result in the following sections, we use simulation models to analyse the effects of our approach under varying load conditions.

1) Error and Network Configuration

We consider deployment scenario (iii) summarised in Section 4.3.2.2. Here we refer to the synthetic data as the RF fingerprint vectors generated by simulations and by using urban path loss models. We consider an outdoor simulation area of $300m \times 300m$ with Figs 4.4(a) and 4.4(b) illustrating the locations of SCBSs and macro BSs together with the estimated serving cell maps. Table 4.2 summarises the simulation parameters.

Figure 4.11 illustrates the misclassification error versus network configuration. Unlike the results in Fig 4.9, the percentage of misclassification of an estimated SCBS with another SCBS is relatively low. This is due to the different network topology and the position of the SCBSs.

2) Energy Efficiency Gains

Improved energy efficiency is a natural by-product of the proposed idle-mode control procedure. To evaluate potential energy savings the spatio-temporal characteristics of the user traffic demand must be considered. Here we consider

75
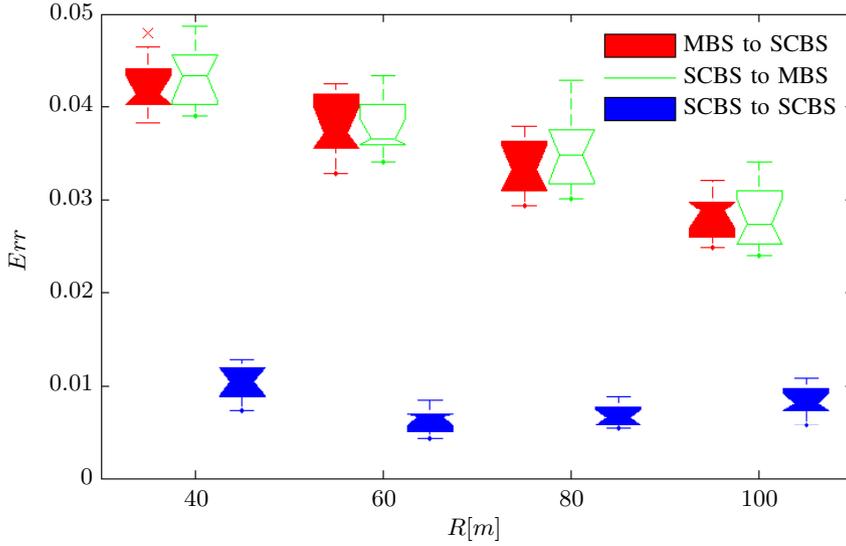
Figure 4.11: Misclassification error of Classifier (b) versus radius (R) of the SCBS from the centre of each Macro BS.

Table 4.3: User traffic characteristics

User Traffic parameter	Value
Number of users per Km^2 (N)	6000
Hotspot factor (F_{sc})	6
Operator market share (N_o)	0.4
Voice activity per month	$512mins$
Data activity per month	$850MB$
Mean throughput activity	$330Kbps$

deployment scenario (iii) which was detailed in the section above. To simulate temporal and spatial variations of the load, we divide 24 hours into 5 minute snapshots. We used the normalised traffic load statistics in Fig 4.10 and user traffic parameters in Table 4.3 to calculate the number of active users during snapshot s. This is given by Eq (4.7).

$$N_{au}(s) = \frac{N F_{sc} \tau \Omega(s)}{A_h N_o} \tag{4.7}$$

where A_h is the area of the hotspot ($300m \times 300m$), F_{sc} a scaling factor, τ the probability of a typical user being active. This probability is calculated by the mean voice and data activity periods per month that are summarised in Table 4.3. These are typical values that are found in different operators' quarterly reports e.g. [76], [4]. $\Omega(s)$ is the normalised load factor that is derived from the data in [75]. The spatial distribution of users is sampled from a correlated

distribution, that tends to yield clusters of users, using a similar approach to [77] which was introduced to generate spatially correlated shadow maps.

For the purpose of this work, we only focus on the performance of the outdoor users affected by the outdoor SCBS. To estimate the power consumption of SCBSs, we adopt the power model in [78], with the energy consumption of base station b given by:

$$e_b(p_b^d, p_b^p) = \begin{cases} P_0 + \Delta(p_b^d + p_b^p) & p_b^d + p_b^p > 0 \\ 0 & p_b^d + p_b^p = 0 \end{cases}$$

where P_0 is idle power usage and Δ the rate of change in electrical power usage with changing RF power where the values of P_0 and Δ are detailed in Table 4.2. Observe also that power consumption is measured over one day of user activity period. Figure 4.12 illustrates the kWh power consumption of three different idle mode selection procedures i.e.

- Network-assisted idle mode selection with perfect estimation (misclassification error is zero.)

- Network-assisted idle mode selection using classifier (b)

- Sniffer-based waking-up of the SCBSs [68] where SCBSs monitor the uplink received power from the users and wake up if the received power is larger than a certain threshold which is set to $-60dBm$ in this work.

Observe that for readability reasons we have not included power consumption values when no idle mode procedure is in place. In fact concurrent operation of all SCBSs in the scenario leads to $2.07kWh$ power consumption. Network-assisted idle mode selection using classifier (b) leads to an average of $0.15kWh$ power consumption. This is a significant energy saving i.e. approximately 90%. It can be observed that erroneous misclassification results in better energy efficiency. This may be due to the fact that a large percentage of misclassification errors involve misclassifying a SCBS user as a MBS user (see Fig 4.11). The cost of this favourable effect is the reduced throughput performance of the users when compared with the perfect detections (see Fig 4.13). As expected up-link power monitoring methods result in lower energy efficiency due to the lack of communication between the cells and the higher idle mode power consumption. Moreover Fig 4.13 suggests improved user throughputs with dynamic selection of the small cells when compared with the simultaneous operation of all the SCBSs. Deactivating SCBSs at low traffic times/areas, results in reduced interference among these base stations. Moreover traffic load of the small cells may be handed over to macro cells in such low traffic activity periods. As a result, the hypothetical users of a now dormant small cell are not effected by the interference signals from the underlying macro BS.
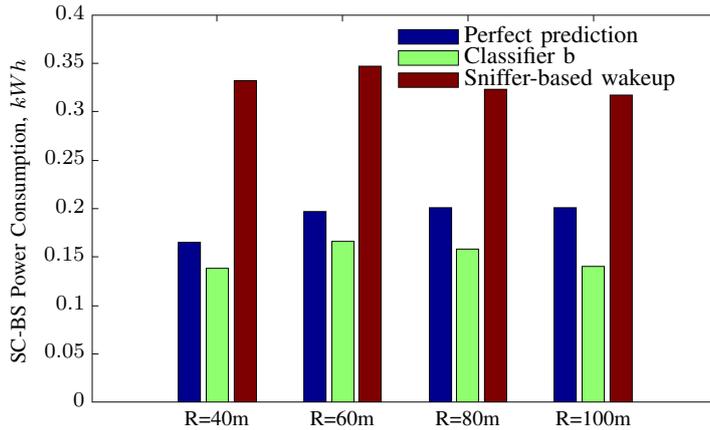
Figure 4.12: Power Consumption (kWh), of the various idle mode procedures, deployment scenario (iii).

Table 4.4 depicts the relationship between size of the training data and median of user throughputs as well as median of 5-percentile user throughputs. From Table 4.4 it can be observed that with decreasing size of the training data, median user throughput tend to decrease. This decrease in user throughputs is due to an increased probability of misclassification with decreasing size of the training data.

Table 4.4: Median user throughput versus size of the training data. 5-percentile user throughput denotes the user throughputs below the 5-percentile point of the CDF of user throughputs.

Size of	Median user throughput [Mbps]				Median 5-percentile user throughput [Mbps]			
the training data	$R = 40m$	$R = 60m$	$R = 80m$	$R = 100m$	$R = 40m$	$R = 60m$	$R = 80m$	$R = 100m$
2585	6.78	8.6	9.05	8.02	0.775	0.85	0.8	0.8
1293	5.36	6.84	7.86	6.04	0.57 5	0.625	0.65	0.6
862	7	7.66	7.87	7.82	0.65	0.65	0.65	0.65
674	5.35	7.26	7.48	7.82	0.55	0.65	0.65	0.65
517	4.66	4.8	4.63	5.49	0.525	0.525	0.525	0.55
259	4.41	4.62	4.34	4.9	0.525	0.55	0.525	0.525

4.5 Conclusions

In this chapter we exploit records of the users' RF fingerprint vectors to estimate their serving cells. We introduce the use of a number of classifiers and we show that an average 1.9% misclassification error rate can be achieved when the proposed two stage classifier is applied on the field measurements. We evaluate our method for various network configurations and various density of the small cells. And we have shown that by using this approach and taking advantage of the traffic load variations, energy efficiency gains as large as 90% are achievable compared to a configuration without idle modes for a given realistic network configuration.

(a) $R = 40m$

(b) $R = 60m$

(c) $R = 80m$

(d) $R = 100m$
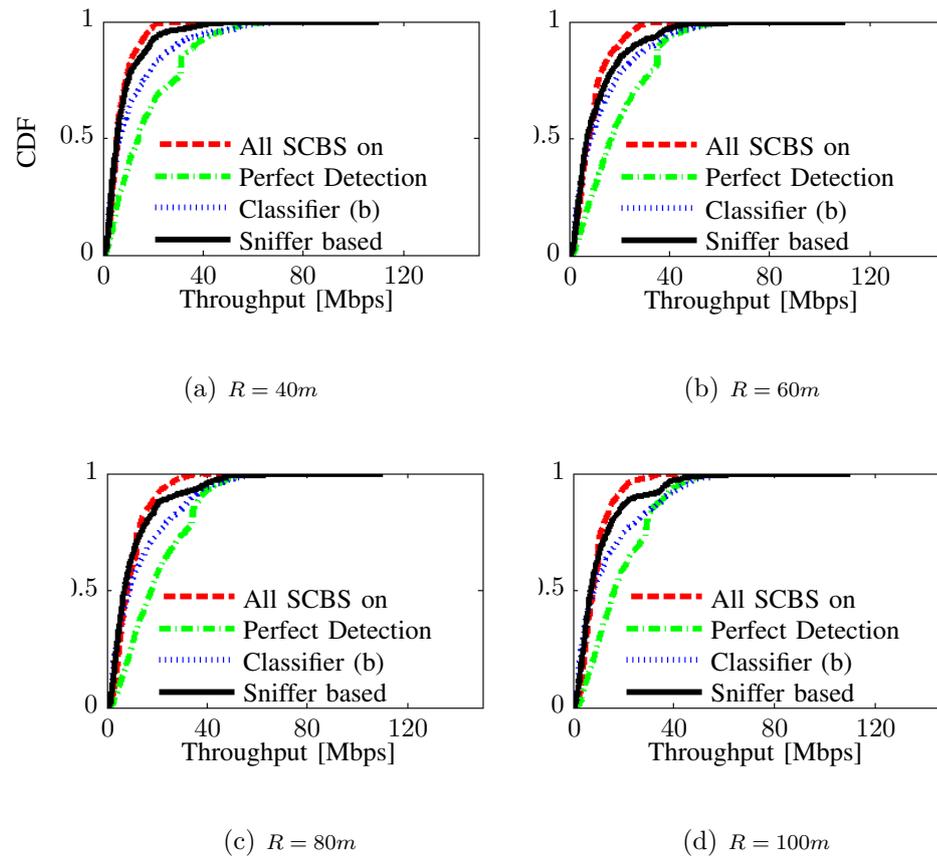
Figure 4.13: CDF of the user throughputs for various idle mode procedures when 4 SCBSs are position on a circle of radius R from the centre of each macro BS.

CHAPTER 5

Semi-Static Power Adaptation in LTE Small Cell Networks

In this chapter we propose a new approach for jointly scheduling deep sleep and allocating pilot and data transmit powers in small cells with the aim of minimising energy consumption while maximising user utility within a mixed macro/small cell network. Our approach exploits the relatively predictable nature of traffic load across each day/week and is demonstrated to achieve significant energy efficiency gains.

5.1   Introduction

Traffic load variations frequently follow predictable patterns during different hours of the day and across days of the week. This predictable behaviour of the traffic load therefore allows long term tuning of the network parameters. Motivated by the reasons described in Section 4.1, we introduce a semi-static approach for adjusting the small cell network parameters. In this approach instead of pure idle mode control, we introduce deep sleep mode scheduling as well as independent adjustment of the pilot and data powers according to long term traffic load statistics and whenever the existing resources satisfy minimum network requirements.

In idle mode, small cells have reduced power usage but the reduction is typically limited to no more than 30% to 60% of the total power consumption, depending on the hardware structure and sleep mode algorithms [65]. To achieve the largest energy savings, it is necessary to place the small cell base stations into deep sleep mode where the energy consumption is essentially zero. However, the increased energy saving comes at the cost of recovery from deep sleep mode being relatively slow. Fortunately, temporal variations in network

traffic load have a broadly predictable component over time scales of a few tens of minutes. This is illustrated by Fig 5.1, which shows measurements of the aggregate traffic load over the course of one week in the centre of Dublin, Ireland. Further, at a given time of day the fluctuations from day to day are relatively small. Hence, the potential exists to predictively determine a small cell deep sleep schedule without significantly compromising user quality of service.
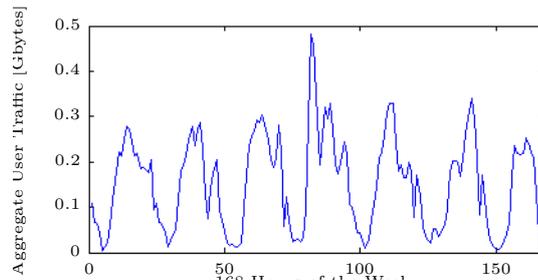
Figure 5.1: Measured aggregate mobile traffic load in Dublin city centre over the course of one week (21st-27th Feb 2011) . The x-axis indicates the week hours [75]

5.2 Related Work

As already discussed in Section 4.2, in addition to our proposed idle mode control method in the previous chapter, a number of sleep mode algorithms which enable dynamic use of the small cell transmit powers have been proposed in the literature e.g. [65] and [68]. Moreover release 11 of the 3GPP standard [70] provisions waking up procedures for the hotspot cells. Long term energy saving procedures have been first noted by the authors in [67]. The authors in [67] highlight that switching-off an entire sector or cell, is of particular interest as it is also feasible in existing radio networks. They further suggest that the simulations in hexagonal scenarios provide a strong indication that the radio coverage can be kept intact by appropriate reconfiguration of the neighbour cells. Following the observations in [67], our work complements wakeup based approaches in that it enables progressively switching off the small cell base stations according to long term traffic load statistics and whenever the existing resources satisfy minimum network requirement in addition of taking account for a proportional fair rate allocation.

5.3 System Model

We consider a macro base station that coordinates a number of linked small cells located within the macro cell coverage area. That is, the network comprises a set of base stations $\mathcal{B}$ where b_0 is the host macro base station and $b_1, .., b_n$ the linked small cell base stations. We let $\mathcal{U}$ denote the set of users.

5.3.1 User Association

The received pilot power from base station $b \in \mathcal{B}$ at user $u \in \mathcal{U}$ is given by $h_{b,u}p_b^p$, where p_b^p is the base station pilot power. As we have previously described in Section 4.3.2.1, each user u is associated with the base station $b(u) \in \mathcal{B}$ which provides the largest received pilot power or the largest signal quality to user u. Hence the base station to which a user is associated may change as the number of active users changes and/or as the pilot power of active stations is adjusted.

Assuming a user to cell association rule based on the received pilot powers, the number of users that are served by base station b is given by:

$$N_b(\boldsymbol{p}^p; \mathcal{U}) = \sum_{u \in \mathcal{U}} \mathbb{1}_{(\arg\max_{b \in \mathcal{B}} h_{b,u}p_b^p = b)} \tag{5.1}$$

where $\boldsymbol{p}^p = (p_b^p)_{b \in \mathcal{B}}$ is the vector of base station pilot powers.

We formally interpret the the allocation of a zero pilot power p_b^p to base station b as meaning that the base station has been placed in deep sleep. The macro station is assumed always to be active, and so no users will be associated with a small cell base station having $p_b^p = 0$ since the received macro pilot power will be non-zero.

5.3.2 User Throughput

We use a throughput model similar to that of Section 2.3. However as the number of users per BS is varying with the change in the power levels, the round robin bandwidth allocation introduced in Eq (2.7) is no longer valid. Observe also that for the reasons described in Section 4.3.2.1, we have treated the pilot and the data signals separately. The downlink throughput of user $u \in \mathcal{U}$ associated with base station $b(u)$ is then given by:

$$r_u(\boldsymbol{p}^d, \boldsymbol{p}^p; \mathcal{U}) = \frac{\omega\beta_1}{N_{b(u)}(\boldsymbol{p}^p; \mathcal{U})} \log(1 + \gamma_u(\boldsymbol{p}^d, \boldsymbol{p}^p)/\beta_2) \tag{5.2}$$

where $\boldsymbol{p}^d = (p_b^d)_{b \in \mathcal{B}}$ is the vector of base station data transmit powers, γ_u is the SINR for user u:

$$\gamma_u(\boldsymbol{p}^d, \boldsymbol{p}^p) = \frac{h_{b(u),u}p_{b(u)}^d}{\sum_{k \in \mathcal{B} \backslash \{b(u)\}} h_{k,u}p_c + \sigma_n^2} \tag{5.3}$$

where $p_k = p_k^p + p_k^d$ is overall transmit power of base station k. Both data and pilot powers contribute to the interference term in the denominator. Observe that in this chapter similar to the definitions in Chapter 4, we let $h_{b,u}$ and $\gamma_{b,u}$ denote the average values of the received signal power and SINR over the available bandwidth.

5.3.3 Base Station Energy Model

To evaluate BS power consumption we use the energy model that we introduced in Section 4.4.2.

5.3.4 Time Snapshots

Time is partitioned into 5 minute periods called *snapshots*. Each snapshot is indexed by a unique integer from $\mathcal{S} = \{1, 2, \ldots, S\}$.

The set $\mathcal{S}$ of snapshots is partitioned into intervals $\mathcal{S}_i := \{\underline{S}_i, \underline{S}_i+1, .., \bar{S}_i\} \subset \mathcal{S}$, with $\underline{S}_i, \bar{S}_i \in \mathcal{S}$, $\underline{S}_i \leq \bar{S}_i$ and so we have a sequence of optimisations, one for each interval. This partitioning can be carried out in a number of ways. In this chapter, we divide the day into four intervals and select the interval boundaries so as to minimise the standard deviation of the normalised traffic load within each interval.

Within each snapshot $s \in \mathcal{S}$, the set of users $\mathcal{U}(s)$ is fixed. However, the set of users may vary between snapshots. For all snapshots $s \in \mathcal{S}_i$ within the i'th interval, the base station transmit powers $\boldsymbol{p}^d(i), \boldsymbol{p}^p(i)$ are held fixed. Hereafter we will often drop the i argument from the base station powers as we will focus on how to select the base station powers within a single interval – selecting the powers within other intervals then simply involves repeating this process.

5.3.5 User Report Data

Our predictive scheduling approach makes use of user measurement reports, which are already commonly available to base stations. Within each snapshot the following historical measurement information is available:

1. Snapshot ID$s \in \mathcal{S}$

2. Number of active users being served by each base station b ($N_b(\boldsymbol{p}^p; \mathcal{U}(s))$)

3. SINR for all active users ($\gamma_u, u \in \mathcal{U}(s)$)

4. Received pilot power $h_{b,u}p_b^p$ from all base stations $b \in \mathcal{B}$ as measured by each active user $u \in \mathcal{U}(s)$.

5. Pilot power for all base stations ($\boldsymbol{p}^p$)

6. Dedicated base station downlink transmit power per active user ($\boldsymbol{p}^d$)

Note that all of this data is already commonly collected by base stations. Using this snapshot data, the path-loss $h_{b,u}$ between each active user u and each base station b can be estimated (from pilot power of each base station and the received pilot power at each user). The useful signal component $h_{b(u),u}p_{b(u)}^d$ can then be estimated.

5.4 Balancing Energy Minimisation and User QoS

5.4.1 Optimisation Problem

We formulate the energy minimisation task over interval $\mathcal{S}_i$ as the following optimisation problem P:

$$\min_{\boldsymbol{p}^d, \boldsymbol{p}^p} f(\boldsymbol{p}^d, \boldsymbol{p}^p) \tag{5.4}$$

$$s.t.\ p_b^d + p_b^p \leq P_{max},\ b \in \mathcal{B} \backslash \{b_0\} \tag{5.5}$$

$$\sum_{s \in \mathcal{S}_i} \sum_{u \in \mathcal{U}(s)} \frac{\mathbb{1}_{(r_u(\boldsymbol{p}^d, \boldsymbol{p}^p; \mathcal{U}(s)) < r_{min})}}{|\mathcal{S}_i||\mathcal{U}(s)|} < 0.02 \tag{5.6}$$

$$N_b(\boldsymbol{p}^p, \mathcal{U}(s)) \leq N_{max},\ b \in \mathcal{B} \backslash \{b_0\}, s \in \mathcal{S}_i \tag{5.7}$$

$$p_b^p \geq 0, p_b^d \geq 0,\ b \in \mathcal{B} \backslash \{b_0\} \tag{5.8}$$

with

$$f(\boldsymbol{p}^d, \boldsymbol{p}^p) = \sum_{b \in \mathcal{B}}^{n} e(p_b^d, p_b^p) - \lambda \sum_{s \in \mathcal{S}} \sum_{u \in \mathcal{U}(s)} \log r_u(\boldsymbol{p}^d, \boldsymbol{p}^p; \mathcal{U}(s))$$

where constraint (5.5) limits the maximum transmit power, constraint (5.6) ensures that more than 98% of the users have throughputs greater than r_{min} (e.g. the throughput needed to support a voice call) and constraint (5.7) restricts the number of users served by each small cell base station.

The first term in the optimisation objective $f(\boldsymbol{p}^d, \boldsymbol{p}^p)$ is the sum of the base station energy usages. The design parameter λ adjusts the balance between minimising this energy efficiency and maximising user utility. The latter is measured by the sum-log of user throughputs, maximisation of which corresponds to a proportional fair rate allocation. When $\lambda = 0$, the optimisation problem minimises energy without regard to user utility i.e. only respecting the basic operational constraint (5.6) that most users achieve a throughput of at least r_{min}. When λ is large, energy costs are essentially ignored and the optimisation problem allocates the base station pilot and data transmit powers to achieve a proportional fair rate allocation. By adjusting λ we can select operating points between these two extremes.

5.4.2 Non-Convexity of Optimisation

Using a similar analysis to [42], it can be shown that the log-rate $\log r_u(\boldsymbol{p}^d, \boldsymbol{p}^p; \mathcal{U})$ is convex with respect to the data signal power $\boldsymbol{p}^d$. However, the log-rate is not convex with respect to the pilot power $\boldsymbol{p}^p$ due to its dependence on $N_b(\boldsymbol{p}^p; \mathcal{U})$, which involves a sum of (non-convex) indicator functions. This non-convexity is unsurprising, as association of users to base stations is inherently discrete in our model (there is no soft handover). Further, switching of base stations into/out of deep sleep is also an inherently discrete process due to the idle power term P_0 in the energy model which introduces a discontinuity into the energy cost.

5.4.3 Predictive Solution of Optimisation

Solving non-convex optimisation P is extremely challenging. For example, if we consider a simple network with 4 small cell base stations and 10 choices of pilot and data transmit power level, the network configuration search space is $\mathbf{20^4 \approx 16e^4}$ in size.

One important consequence of this is that it is unrealistic to try to solve optimisation P dynamically in real-time as network conditions change, even when substantial computing power is available. This general observation has been a major obstacle to the use of dynamic adaptation within cellular networks, combined with understandable concerns about introducing new types of failure mode via the adaptation itself.

One of the contributions of the present chapter is the observation that, while the traffic load on a network is changing, a large component of this change is *predictable* based in daily usage patterns (recall Fig 5.1). Hence, it is possible to find good solutions to optimisation P predictively ahead of time. That is, without being subject to strong realtime deadlines and with scope for error checking to eliminate adaptation-related failures. In this way, a practically viable adaptation approach can be realised.

5.4.4 Simulated Annealing

Many numerical methods are available to solve non-convex optimisation problems, once we relax the requirement of meeting a realtime deadline. Here we adopt a simulated annealing approach [79]. Specifically, following [80], we use a Cauchy distribution to generate samples from the solution space, and then apply a Gibbs sampler on the cost function of each newly generated sample. As a temperature parameter is decreased, the stationary distribution converges to the global minima.

To account for the constraints in optimisation P, we use a penalty method [81] whereby the cost function Ξ minimised by the annealing algorithm is given by:

$$\Xi(\boldsymbol{p}^d, \boldsymbol{p}^p) = f(\boldsymbol{p}^d, \boldsymbol{p}^p) + c \sum_i P(g_i(\boldsymbol{p}^d, \boldsymbol{p}^p)) \tag{5.9}$$

where penalty function $P(x)$ equals x when $x \leq 0$ and equals 0 otherwise, c is a large constant, and the g_i functions enforce the inequality constraints in optimisation P.

The resultant algorithm is detailed in Algorithm 5.5. The maximum number of iterations (iter_{max}) is affected by the dimension of the solution space and the method of generating the sampled solutions. We found that significantly fewer iterations are required when solution samples are drawn from a Cauchy probability function, consistent with the observations in [80]. We use a geometric cooling schedule, decreasing the temperature parameter ϑ at rate

ς. The initial temperature is selected to be significantly larger than typical values of the objective function so that solutions at high temperatures are accepted with high probability. The minimum temperature ϑ_{min} value is selected experimentally by finding the largest minimum temperature for which the annealing converges.

Algorithm 5.5 Simulated Annealing Algorithm

Initialise: ϑ, $\boldsymbol{p}^d$, $\boldsymbol{p}^p$
while $\vartheta > \vartheta_{min}$ **do**
  $\boldsymbol{p}^d(0) \leftarrow \boldsymbol{p}^p$, $\boldsymbol{p}^p(0) \leftarrow \boldsymbol{p}^p$,$C_0 = \Xi(\boldsymbol{p}^d, \boldsymbol{p}^p)$, $i \leftarrow 1$
  while $i < \text{iter}_{max}$ **do**
    Generate solution vector $\boldsymbol{p}^d(i), \boldsymbol{p}^p(i)$
    Calculate cost function $\Xi_i = E(\boldsymbol{p}^d(i), \boldsymbol{p}^p(i))$
    if $\Xi_{i-1} - \Xi_i < 0$ **then**
      $\boldsymbol{p}^d \leftarrow \boldsymbol{p}^d$, $\boldsymbol{p}^p \leftarrow \boldsymbol{p}^p$
    else
      Draw X uniformly at random from $[0, 1]$
      if $X \le e^{\frac{\Xi_{i-1} - \Xi_i}{\vartheta}}$ **then**
        $\boldsymbol{p}^d \leftarrow \boldsymbol{p}^d(i)$, $\boldsymbol{p}^p \leftarrow \boldsymbol{p}^p(i)$
      end if
    end if
    $i \leftarrow i + 1$
  end while
  $\vartheta \leftarrow \varsigma\vartheta$
end while

5.5 Performance Evaluation: Berlin City Centre

To investigate the practical utility of the proposed energy saving approach, we applied it to a realistic network scenario based on the cellular network data from a top tier provider in the south east of Berlin, in the vicinity of Grünauer Straße. We consider a $200m \times 200m$ hotspot area within the coverage area by a single macro base station. This hotspot area is distance d metres from the macro base station (we present results below for various values of d) and within the hotspot we deploy 4 small cell base stations. There are three neighbouring macro base stations, with associated interference. We consider two methods of deployment for the small cells:

1. Scenario A: the small cell base stations are placed uniformly at random within the hotspot area i.e. their positions are independent of the user locations.

2. Scenario B: the small cell base stations are co-located with clusters of users. At peak traffic hours users tend to be located in clusters, e.g. in coffee shops, market squares, *etc*, and co-locating small cells with these clusters can be expected to maximise the scope for macro cell offload within the hotspot.

The simulation parameters are based on the 3GPP standard [82] and are summarised in Table 5.1 alongside the optimisation parameters.

Table 5.1: Semi static power control: simulation parameters

	Parameter	value
Base station parameters	P_{b0}	$46dBm$
	P_{max}	$250mW$
	N_{max}	16
	P_0	$13.6W$
	Δ	4
Channel	ϱ, LOS	2.2
	ℓ_0, LOS	$10^{3.4}$
	ϱ, NLOS, Macro BS	3.9
	ϱ, NLOS, Small BS	3.67
	ℓ_0, NLOS, macro BS	$10^{2.1}$
	ℓ_0, NLOS, Small BS	10^3
	ω	$10MHz$
	Bandwidth Efficiency β_1	1
	SNR Efficiency β_2	1
	UE noise power σ_n^2	$-94.97dBm$
Optimisation	P_{max}	$250mW$
	ς	0.9
	ϑ	10^7
	ϑ_{min}	1
	iter$_{max}$	2000

5.5.1 User traffic load and locations

We used normalised daily load profile shown in Fig 4.10 for evaluation. We divide the 24 hours period into 5 minute snapshots and determine four intervals as previously discussed. Table 4.3 summarises the user traffic parameters. The number of active users in the hotspot area during snapshot s is given by Eq (4.7). The number of the active users in each macro cell is obtained similarly. The spatial distribution of users is sampled from a correlated distribution, that tends to yield clusters of users, using a similar approach to [77] which was introduced to generate spatially correlated shadow maps.

5.5.2 Scenario A

For Scenario A (small cells located uniformly at random), the optimised pilot and data transmit power configurations are calculated for various values of λ and with $d = 484.55m$ (*i.e.* the hotspot is located towards the edge of the macro cell). The resulting aggregate network power consumption is shown in Fig 5.2. The power consumption for each λ optimal configuration is compared with full power operation (where all small cell base stations are operating at their maximum power). These results indicate that when $\lambda = 0.1$ the optimised power schedule achieves an average of 37.8% energy efficiency gain compared with full power operation. Even higher energy efficiency gains of 76.8% are achieved when the traffic load is low (second interval $\mathcal{S}_2$). Figs
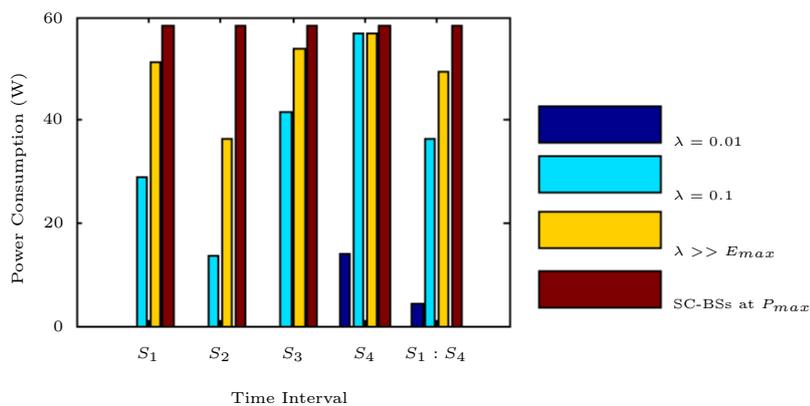

Figure 5.2: Aggregate network power consumption vs λ for Scenario A.

5.3(a) to 5.3(d) show the CDF of user throughputs for each traffic regime. It can be seen that, as expected, as the design parameter λ is increased the energy efficiency is deprioritised relative to the proportional fair rate allocation. It can be seen that when $\lambda = 0.1$ a reasonable balance between energy efficiency and throughput is achieved.

5.5.3 Scenario B

In Scenario B the small cell base stations are positioned where the peak traffic demand is maximum. The resulting energy usage and user throughputs are shown in Figs 5.4-5.5. It can be seen that the energy usage is slightly increased compared to Scenario A, with the average reduction in power consumption decreasing from 37.8% to 29%. However, this is balanced by a significant average increase of 80% in median user throughput compared to the random small cell placements used in Scenario A.

5.5.4 Impact of Hotspot Location

The performance benefit provided by the small cells can be expected to be strongly dependent on the distance d between the hotspot area and the macro

(a) Interval $\mathcal{S}_1$

(b) Interval $\mathcal{S}_2$

(c) Interval $\mathcal{S}_3$

(d) Interval $\mathcal{S}_4$
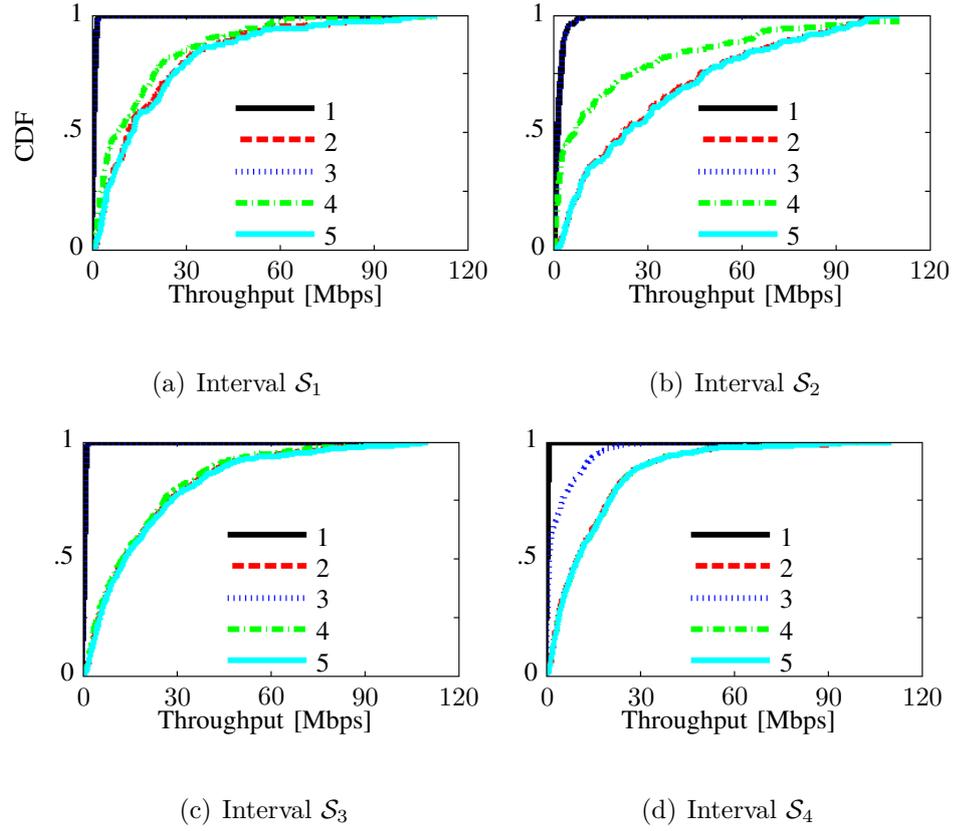
Figure 5.3: CDF of user throughputs in Scenario A. **1**: SCBS off, **2**: SCBS @P_{max}, **3**: $\lambda = 0.01$, **4**: $\lambda = 0.1$, **5**: $\lambda >> E_{max}$.
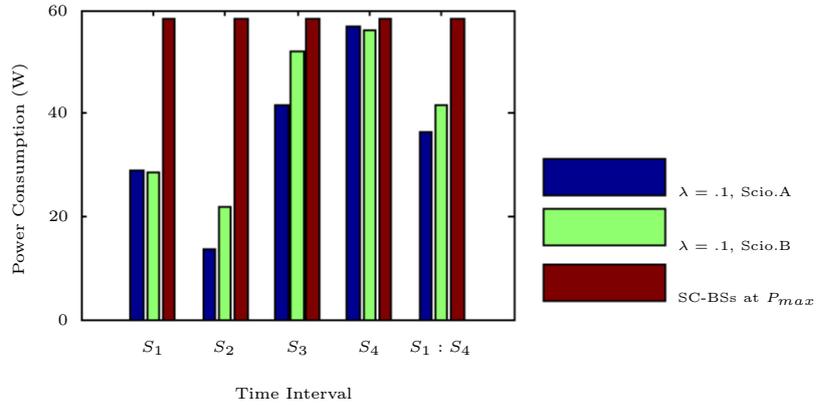
Figure 5.4: Aggregate network power consumption in Scenario B.

base station for the macro cell in which the hotspot is located. As the hotspot area is moved towards the edge of the host macro cell (d is increased), the received power from the macro base station reduces and users can be expected to be less satisfied by the macro base station, and so more small cells may

(a) Interval $\mathcal{S}_1$

(b) Interval $\mathcal{S}_2$

(c) Interval $\mathcal{S}_3$

(d) Interval $\mathcal{S}_4$
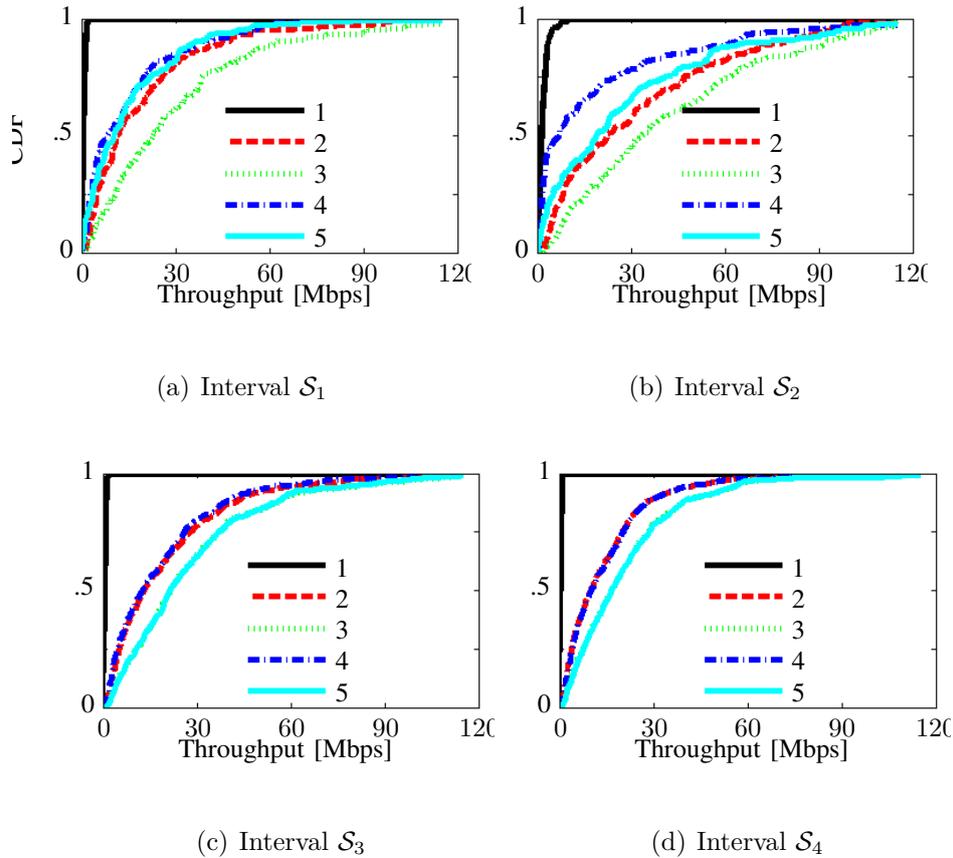
Figure 5.5: CDF of user throughputs in Scenario B compared to that of Scenario A. **1**: SCBS off, **2**: SCBS @P_{max}- Scenario A, **3**: SCBS @P_{max}- Scenario A, **4**: $\lambda = 0.1$- Scenario A, **5**: $\lambda = 0.1$- Scenario B.

needed to meet traffic demand. Moreover, due to the reduced received power from the macro cell, the interference between the host macro and the small cells can also be expected to reduce, resulting in improved small cell SINR and throughput. The energy usage and user throughputs vs distance d is shown in Figs 5.6-5.7. It can be seen that the reduction in average power consumption increases from 37.8% to 73.8% as the hotspot area is moved from the edge of the cell to beside the host macro station. Intuitively, when the traffic load is low, small cells in the vicinity of the host macro base station can be switched off and so larger energy efficiency can be achieved.
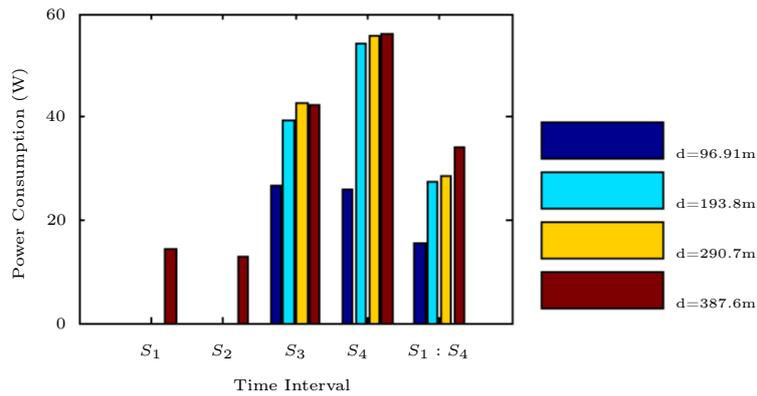
Figure 5.6: Aggregate network power consumption vs hotspot distance d from macro base station, Scenario A.

(a) Interval $\mathcal{S}_1$

(b) Interval $\mathcal{S}_2$

(c) Interval $\mathcal{S}_3$

(d) Interval $\mathcal{S}_4$
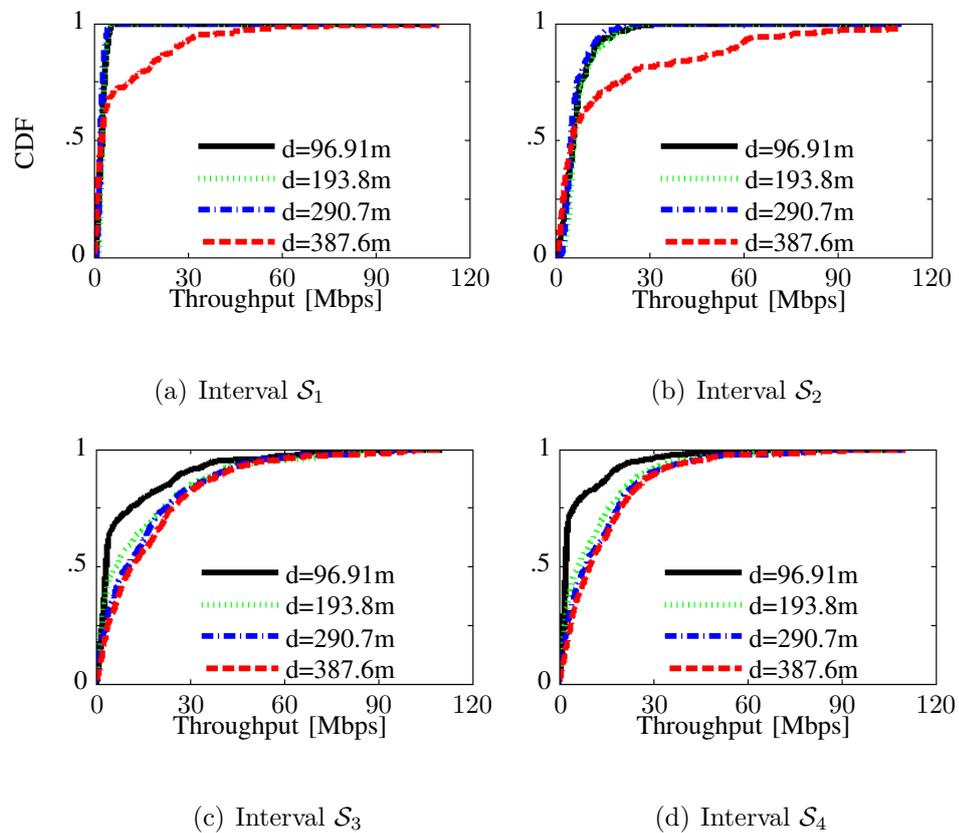
Figure 5.7: CDF of user throughputs vs hotspot distance d from macro base station, Scenario A.

5.6 Conclusions

In this chapter, we make use of the long term behaviour of users' traffic demand to optimise the configuration of small cell base stations. We formulate adjust-

ment of the data and pilot powers independently, where range expansion can be achieved by adjustment of pilot powers, and interference can be controlled by the total transmit power. The aim is to reach a balance between energy efficiency and throughput fairness of the users while network constraints are satisfied. The effectiveness of the proposed approach is demonstrated using number of simulations where it is found that even when small cells are located at uniformly random positions (and so not tailored to traffic demand), an average of more than 37.8% energy efficiency gain is achieved.

Concluding Remarks

In this final chapter, we discuss the results presented in the thesis and suggest directions for future research.

6.1 Overview and Discussions

The results presented in this thesis emerged from an attempt to find practical (yet sustainable) answers to a number of questions that currently concern researchers in the wireless industry. The outcome is a number of self-optimisation and self-configuration techniques that are designed for next generation RANs. Having considered capacity maximisation and energy efficiency as the key objectives, we present algorithms that are relatively easy to implement. We further show that the proposed algorithms result in significant improvements in a number of measured network utility.

As already discussed in Chapter 2, the practical impact of antenna tilt angles on network performance has been well studied in the literature. Although solving this optimisation problem has attracted attention in the literature, most of this work makes use of heuristic approaches. In Chapter 2, we first show that in high SINR regimes with an appropriate change of variables this problem can be described in a convex form. We then proceeded to show that the seemingly non-convex optimisation problem for any SINR regime can be reduced to a proportional fair convex optimisation framework. This is achieved by a linear approximation of the vertical gain pattern of the antennas, which is valid for the antennas of base stations other than those to which the UE is associated (emphasising that this is only a mild assumption). Solving the proportional fair optimisation problem with standard convex optimisation methods imposes computational complexities. We therefore develop a primal-dual approach that is light-weight and suitable for distributed implementation.

93

To evaluate the impact of the proposed approach, we use realistic simulation scenarios in central Dublin where we show that significant performance gains can be achieved (a 22% gain in the proportional fair objective and a factor of 4 improvement in the median of the user throughputs compared with the use of fixed tilt angles). Moreover we study the performance gain of the proposed method in conjunction with other interference mitigation techniques i.e. use of MMSE post processing for interference cancellation. We found that considerable performance gains due to tilt angle optimisation were still evident.

In view of the continuing increase in data traffic, in Chapter 3 we consider a topical class of integrated LTE and 802.11 networks. As discussed in Chapter 3, RAT selection strategies implemented by the UE or the network have also attracted attention in the literature, although only simple linear models are mostly used to describe both the cellular and 802.11 networks. In our approach we considered detailed system models of both LTE and 802.11 networks including support for both multipath and multihomed operation. We established a utility fair optimisation problem in order to maximise network capacity. This optimisation problem is non-convex. We show that with some knowledge of the part of network rate region that the optima of the problem are likely to lie, sub-optimal solutions can be found by solving a convex optimisation, by appropriate change of variables and re-writing the constraints as a difference of convex functions. We also present a concave-convex procedure that converges to a stationary point of the non-convex optimisation problem. We then extend this approach to find global optimums by randomisation. This provides a principled approach for offload from LTE to 802.11 and for exploiting LTE/802.11 path diversity.

In Chapters 4 and 5, we consider load aware idle/sleep mode scheduling of dense small cell networks. As discussed in both chapters, it is well known that traffic load varies over time as well as varying spatially. Exploiting these variations in order to gain energy efficiency and to improve network performance is a natural approach. In both Chapters however, we aimed to find effective solutions which were easy to implement in current networks.

In Chapter 4 we use a dataset of "RF Fingerprints" to estimate relative position of a new user. RF fingerprints are readily available at the UEs in the form of received signal powers. In order to construct a dataset, it is only required to report these information to a central network entity. By collecting data from central Dublin, we showed that real field RF Fingerprints are in fact more sparse than the synthetic data produced by simulation models. Hence we proposed the use of a two stage classifier based on Jaccard similarity distance. Using field measurements as well as system level simulations, we show that our approach results in low prediction errors and energy efficiency gains as large as 90%.

To exploit the predictable nature of the traffic load, in Chapter 5, we apply a more rigorous approach where we formulate adjustment of the data and pilot powers independently, with range expansion achieved by adjustment of pilot powers, and interference controlled by adjusting the total transmit power. The aim is to reach a balance between energy efficiency and throughput fairness for the users. To solve this optimisation problem, we adopt a simulated annealing approach which requires collaboration between neighbouring small cells as well as knowledge of the traffic load statistics. We demonstrate the effectiveness of this approach by using a number of simulations and we show that on average a 37.8% gain in energy efficiency (over a period of 24 hours) is achieved.

6.2 Future Work

In this final section we discuss a number of questions and possible future directions that arise from the work of this thesis.

1. It is possible to further develop the proposed primal-dual updates in Chapter 2 to a more light-weight method, using approaches similar to those of [83]. This may be useful for reducing the communication overhead in practical implementations, and also it would be a theoretically interesting result to obtain.

2. Due to the structure of the optimisation problem described by Eqs (3.31),(3.32) in Chapter 3, it may also be possible to describe the optimisation problem in a convex form [84]. However this requires a more detailed study of the non-convex (concave) functions in the optimisation problem.

3. It may also be of practical interest to compare the complexity of the proposed utility fair user association approach in Chapter 3, with a learning based approach which may require less information exchange e.g. the distributed learning methods introduced in [85]. However it must be noted that these methods are likely to lead to longer convergence rates.

4. The tilt angle adaptation approach in Chapter 2 may be used in conjunction with the dynamic or semi-static transmit power adaptation described in Chapters 4 and 5. In this case, the number of users associated to a macro base station varies with changes in the small cell configuration. Therefore to jointly adapt the small cell transmit powers and macro cell tilt angles, we may need to extend the system model and the optimisation framework.

5. In Chapters 4 and 5, the choice of dynamic or semi-static power control depends on operator requirements. Perhaps as a future study, these methods can be compared in terms of their cost efficiency and appeal for operators.

Bibliography

[1] *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2*, 2012.

[2] *LTE; Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced) (3GPP TR 36.913 version 9.0.0 Release 9)*, 2010.

[3] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5g era," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 90–96, 2014.

[4] "Erricson mobility report," June, 2015. [Online]. Available at www.ericsson.com/res/docs/2015/ericsson-mobility-report-june-2015.pdf.

[5] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *Communications Magazine, IEEE*, vol. 46, no. 9, pp. 59–67, 2008.

[6] C. E. Shannon, "A note on the concept of entropy," *Bell System Tech. J*, vol. 27, pp. 379–423, 1948.

[7] S. Sesia, I. Toufik, and M. Baker, *LTE: the UMTS long term evolution.* Wiley Online Library, 2009.

[8] D. Tse and P. Viswanath, *Fundamentals of wireless communication.* Cambridge university press, 2005.

[9] I. E. Telatar *et al.*, "Capacity of multi-antenna gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[10] "Indoor deployment strategies," June, 2014. [Online]. Available at http://networks.nokia.com/portfolio/solutions/indoor-solutions.

[11] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *Vehicular Technology Magazine, IEEE*, vol. 6, no. 1, pp. 37–43, 2011.

[12] *R1-104968, Summary of the description of candidate eICIC solutions; 3GPP Std., Madrid, Spain*, Aug. 2010.

[13] J. M. Graybeal and K. Sridhar, "The evolution of son to extended son," *Bell Labs Technical Journal*, vol. 15, no. 3, pp. 5–18, 2010.

[14] *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Self-configurating , Self-optimizating Network (SON); use cases and solutions; 3GPPTR 36.902 (Release 9)*, 2010.

[15] S. Feng and E. Seidel, "Self-organizing networks (son) in 3gpp long term evolution," *Nomor Research GmbH, Munich, Germany*, 2008.

[16] J. Huang, R. Berry, M. L. Honig, *et al.*, "Distributed interference compensation for wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 5, pp. 1074–1084, 2006.

[17] A. L. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse in ofdma systems," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, IEEE, 2008.

[18] B. Rengarajan, A. L. Stolyar, and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse on the uplink of ofdma systems," in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, pp. 1–6, IEEE, 2010.

[19] E.-V. Belmega, S. Lasaulce, and M. Debbah, "Power allocation games for mimo multiple access channels with coordination," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 6, pp. 3182–3192, 2009.

[20] R. Rom and M. Sidi, *Multiple access protocols: performance and analysis.* Springer Science & Business Media, 2012.

[21] *ETSI TR 136 912 V9.0.0 (2009-09) LTE; Feasibility study for Further Advancements for E-UTRA (LTE- Advanced) (3GPP TR 36.912 version 9.0.0 Release 9)*, 2009.

[22] F. P. Kelly, A. K. Maulloo, and D. K. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research society*, pp. 237–252, 1998.

[23] D. P. Bertsekas and R. G. Gallager, *Data networks.* Prentice-hall, 1987.

[24] J.-Y. Le Boudec, "Rate adaptation, congestion control and fairness: A tutorial," *Web page, November*, 2005.

[25] R. Mazumdar, L. G. Mason, and C. Douligeris, "Fairness in network optimal flow control: Optimality of product forms," *Communications, IEEE Transactions on*, vol. 39, no. 5, pp. 775–782, 1991.

[26] R. Srikant and L. Ying, *Communication Networks: An Optimization, Control, and Stochastic Networks Perspective.* Cambridge University Press, 2013.

[27] L. Massoulié and J. Roberts, "Bandwidth sharing: objectives and algorithms," in *INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, pp. 1395–1403, IEEE, 1999.

[28] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge university press, 2004.

[29] I. Forkel, A. Kemper, R. Pabst, and R. Hermans, "The effect of electrical and mechanical antenna down-tilting in umts networks," in *3G Mobile Communication Technologies, 2002. Third International Conference on (Conf. Publ. No. 489)*, pp. 86–90, IET, 2002.

[30] F. Athley and M. Johansson, "Impact of electrical and mechanical antenna tilt on lte downlink system performance," in *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, pp. 1–5, IEEE, 2010.

[31] E. Benner and A. Sesay, "Effects of antenna height, antenna gain, and pattern downtilting for cellular mobile radio," *Vehicular Technology, IEEE Transactions on*, vol. 45, no. 2, pp. 217–224, 1996.

[32] O. Yilmaz, S. Hamalainen, and J. Hamalainen, "Analysis of antenna parameter optimization space for 3gpp lte," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, pp. 1–5, IEEE, 2009.

[33] H. Eckhardt, S. Klein, and M. Gruber, "Vertical antenna tilt optimization for lte base stations.," in *VTC Spring*, pp. 1–5, IEEE, 2011.

[34] R. Razavi, S. Klein, and H. Claussen, "A fuzzy reinforcement learning approach for self optimization of coverage in lte networks," *Bell Lab. Tech. J.*, vol. 15, pp. 153–175, Dec. 2010.

[35] A. Temesváry, "Self-configuration of antenna tilt and power for plug & play deployed cellular networks," in *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, pp. 1–6, IEEE, 2009.

[36] G. Calcev and M. Dillon, "Antenna tilt control in cdma networks," in *Proceedings of the 2nd annual international workshop on Wireless internet*, p. 25, ACM, 2006.

[37] A. Eisenblatter and H. Geerdes, "Capacity optimization for umts: Bounds and benchmarks for interference reduction," in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pp. 1–6, IEEE, 2008.

[38] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9)*, 2010.

[39] P. Mogensen, W. Na, I. Z. Kovács, F. Frederiksen, A. Pokhariyal, K. I. Pedersen, T. Kolding, K. Hugl, and M. Kuusela, "Lte capacity compared to the shannon bound," in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pp. 1234–1238, IEEE, 2007.

[40] FCC, "Wireless E911 Location Accuracy Requirements." http://www.fcc.gov/document/wireless-e911-location-accuracy-requirements-1/, 2011.

[41] *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); LTE Positioning Protocol (LPP); 3GPP TS 36.355 (Release 9)*, 2010-02.

[42] J. Papandriopoulos, S. Dey, and J. Evans, "Optimal and distributed protocols for cross-layer design of physical and transport layers in manets," *IEEE/ACM Transactions on Networking (TON)*, vol. 16, no. 6, pp. 1392–1405, 2008.

[43] *Universal Mobile Telecommunications System (UMTS), Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (Release 11)*, 2012.

[44] B. Partov, D. J. Leith, and R. Razavi, "Tilt angle adaptation in lte networks with advanced interference mitigation," in *IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pp. 1959 – 1964, IEEE, 2014.

[45] *3GPP TR 36.819 V11.1.0 (2011-12) Technical Specification Group Radio Access Network; 3rd Generation Partnership Project; Coordinated multipoint operation for LTE physical layer aspects (Release 11)*, 2011.

[46] D. Giustiniano, E. G. Llairo, J. M. Pozo, A. L. Toledo, and P. Rodriguez, "Clubadsl: When your neighbors are your friends," in *Computers and Communications, 2009. ISCC 2009. IEEE Symposium on*, pp. 25–29, IEEE, 2009.

[47] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure, "TCP Extensions for Multipath Operation with Multiple Addresses," Internet Draft draft-ietf-mptcp-rfc6824bis-04, IETF, Mar. 2015.

[48] L. Wang and G.-S. Kuo, "Mathematical modeling for network selection in heterogeneous wireless networks—a tutorial," *Communications Surveys & Tutorials, IEEE*, vol. 15, no. 1, pp. 271–292, 2013.

[49] I. Blau, G. Wunder, I. Karla, and R. Sigle, "Decentralized utility maximization in heterogeneous multicell scenarios with interference limited and orthogonal air interfaces," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, p. 2, 2009.

[50] M. Ismail and W. Zhuang, "A distributed multi-service resource allocation algorithm in heterogeneous wireless access medium," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 2, pp. 425–432, 2012.

[51] D. Kumar, E. Altman, and J. M. Kelif, "User-network association in a wlan-umts hybrid cell: Global & individual optimality," *CoRR*, vol. abs/cs/0608025, 2006.

[52] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, "An overview of load balancing in hetnets: Old myths and open problems," *arXiv preprint arXiv:1307.7779*, 2013.

[53] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "Rat selection games in hetnets," in *INFOCOM, 2013 Proceedings IEEE*, pp. 998–1006, IEEE, 2013.

[54] S. Brueck, L. Zhao, J. Giese, and M. A. Amin, "Centralized scheduling for joint transmission coordinated multi-point in lte-advanced," in *Smart Antennas (WSA), 2010 International ITG Workshop on*, pp. 177–184, IEEE, 2010.

[55] J. Huang, V. Subramanian, R. Berry, and R. Agrawal, "Scheduling and resource allocation in ofdma wireless systems," *book chapter, submitted*, 2009.

[56] G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *Selected Areas in Communications, IEEE Journal on*, vol. 18, no. 3, pp. 535–547, 2000.

[57] D. J. Leith, V. G. Subramanian, and K. R. Duffy, "Log-convexity of rate region in 802.11 e wlans," *arXiv preprint arXiv:1102.4469*, 2011.

[58] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, 2003.

[59] "802.11n-2009 - IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 11: Wireless LAN Medium Access Control (MAC)and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput," *IEEE Std*, 2009.

[60] C. V. N. Index, "Global mobile data traffic forecast update, 2012-2017," *Cisco white paper*, 2013.

[61] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 3, pp. 497–508, 2012.

[62] R. Razavi and H. Claussen, "Urban small cell deployments: Impact on the network energy consumption," in *Wireless Communications and Networking Conference Workshops (WCNCW), 2012 IEEE*, pp. 47–52, IEEE, 2012.

[63] A. Mesodiakaki, F. Adelantado, A. Antonopoulos, E. Kartsakli, L. Alonso, and C. Verikoukis, "Energy impact of outdoor small cell backhaul in green heterogeneous networks," in *IEEE CAMAD 2014*, IEEE, DOI: 10.13140/2.1.3618.0162, 2014.

[64] J. T. Louhi, "Energy efficiency of modern cellular base stations," in *Telecommunications Energy Conference, 2007. INTELEC 2007. 29th International*, pp. 475–476, IEEE, 2007.

[65] I. Ashraf, F. Boccardi, and L. Ho, "Sleep mode techniques for small cell deployments," *Communications Magazine, IEEE*, vol. 49, no. 8, pp. 72–79, 2011.

[66] D. Lopez-Perez, I. Guvenc, G. De La Roche, M. Kountouris, T. Q. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, no. 3, pp. 22–30, 2011.

[67] O. Blume, H. Eckhardt, S. Klein, E. Kuehn, and W. M. Wajda, "Energy savings in mobile networks based on adaptation to traffic statistics," *Bell Labs Technical Journal*, vol. 15, no. 2, pp. 77–94, 2010.

[68] H. Claussen, I. Ashraf, and L. T. Ho, "Dynamic idle mode procedures for femtocells," *Bell Labs Technical Journal*, vol. 15, no. 2, pp. 95–116, 2010.

[69] F. Meshkati, H. V. Poor, S. C. Schwartz, and N. B. Mandayam, "An energy-efficient approach to power control and receiver design in wireless data networks," *Communications, IEEE Transactions on*, vol. 53, no. 11, pp. 1885–1894, 2005.

[70] *ETSI TR 136 927 V11.0.0 (2012-10) LTE ,Evolved Universal Terrestrial Radio Access (E-UTRA), Potential solutions for energy saving for E-UTRAN (3GPP TR 36.927 version 11.0.0 Release 11)*, 2012.

[71] https://github.com/andr3jx/Mobilog.

[72] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9)*, 2010.

[73] http://www.askcomreg.ie/mobile/siteviewer.273.LE.asp.

[74] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani, *The elements of statistical learning*, vol. 2. Springer, 2009.

[75] E. Carolan, S. C. McLoone, S. F. McLoone, and R. Farrell, "Analysing ireland's interurban communication network using call data records.," in *23rd Irish Signals and Systems Conference, June 28-29 2012*.

[76] "Ericson mobility report," Novermber, 2012. [Online]. Available at www.ericsson.com/res/docs/2015/ericsson-mobility-report-november-2012.pdf.

[77] H. Claussen, "Efficient modelling of channel maps with correlated shadow fading in mobile radio systems," in *Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. IEEE 16th International Symposium on*, vol. 1, pp. 512–516, IEEE, 2005.

[78] G. Auer, O. Blume, V. Giannini, I. Godor, M. Imran, Y. Jading, E. Katranaras, M. Olsson, D. Sabella, P. Skillermark, *et al.*, "D2. 3: energy efficiency analysis of the reference systems, areas of improvements and target breakdown," *INFSOICT-247733 EARTH (Energy Aware Radio and NeTwork TecHnologies), Tech. Rep*, 2010.

[79] S. Kirkpatrick, D. G. Jr., and M. P. Vecchi, "Optimization by simmulated annealing," *science*, vol. 220, no. 4598, pp. 671–680, 1983.

[80] H. Szu and R. Hartley, "Fast simulated annealing," *Physics letters A*, vol. 122, no. 3, pp. 157–162, 1987.

[81] B. W. Wah and T. Wang, "Simulated annealing with asymptotic convergence for nonlinear constrained global optimization," in *Principles and Practice of Constraint Programming–CP'99*, pp. 461–475, Springer, 1999.

[82] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); LTE physical layer, general description (Release 11)*, 2012.

[83] K. Kar, S. Sarkar, and L. Tassiulas, "A simple rate control algorithm for max total user utility," in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 1, pp. 133–141, IEEE, 2001.

[84] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.

[85] J. R. Marden, H. P. Young, and L. Y. Pao, "Achieving pareto optimality through distributed learning," *SIAM Journal on Control and Optimization*, vol. 52, no. 5, pp. 2753–2770, 2014.

[86] P. Billingsley, *Probability and measure*. John Wiley & Sons, 2008.

Primal-Dual Updates: Convergence

Consider the optimisation problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \qquad s.t. \qquad g_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, m$$

with $\mathbf{x} \in \mathcal{R}^n$, $f(\mathbf{x}) : \mathcal{R}^n \to \mathcal{R}$ convex, $g_i(\mathbf{x}) : \mathcal{R}^n \to \mathcal{R}$, $i = 1, \cdots, m$ convex. For simplicity we will assume that $f(\mathbf{x})$, $g_i(\mathbf{x})$ are differentiable, but this could be relaxed. The optimisation problem is convex and so at least one solution exists, let X^* denote the set of solutions Assuming the Slater condition is satisfied, then strong duality holds and the KKT conditions are necessary and sufficient for optimality. The Lagrangian is

$$L(\mathbf{x}, \mathbf{u}) = f(\mathbf{x}) + \sum_{i=1}^{m} u_i g_i(\mathbf{x})$$

where u_i is the multiplier associated with constraint $g_i(\mathbf{x}) \geq 0$ and $\mathbf{u} = (u_1, \cdots, u_m)$. At an optimum $\mathbf{x}^* \in X^*$, the multipliers must lie in set

$$U(\mathbf{x}^*) = \{\mathbf{u} : \mathbf{u} = \arg\sup_{\mathbf{u} \geq 0} L(\mathbf{x}^*, \mathbf{u})\}$$

A.1 Gradient Algorithm

We consider the following primal-dual update

$$x_j(t+1) = x_j(t) - \alpha\partial_{x_j}L(\mathbf{x}(t), \mathbf{u}(t)), \quad j = 1, \cdots, n \qquad \text{(A.1)}$$

$$u_i(t+1) = [u_i(t) + \alpha\partial_{u_i}L(\mathbf{x}(t), \mathbf{u}(t))]^+, \quad i = 1, \cdots, m \qquad \text{(A.2)}$$

where step size $\alpha > 0$ and $\partial_{x_j}L(\mathbf{x}, \mathbf{u})$, $\partial_{u_i}L(\mathbf{x}(t), \mathbf{u}(t))$ are subgradients of $L(\mathbf{x}(t), \mathbf{u}(t))$ with respect to x_j and u_i respectively. We have $\partial_{x_j}L(\mathbf{x}, \mathbf{u}) = \partial_{x_j}f(\mathbf{x}) + \sum_{i=1}^{m} u_i\partial_{x_j}g_i(\mathbf{x})$ and $\partial_{u_i}L(\mathbf{x}(t), \mathbf{u}(t)) = g_i(\mathbf{x})$ with $\partial_{x_j}f(\mathbf{x})$ a subgradient of $f(\mathbf{x})$ with respect to x_j, $\partial_{x_j}g_i(\mathbf{x})$ a subgradient of $g_i(\mathbf{x})$ with respect to x_j. Projection $[z]^+ = z$ when $z \geq 0$, 0 otherwise.

A.2 Fixed Points

Lemma A.2.1 (Fixed points). $(\mathbf{x}^*, \mathbf{u}^*)$ *with* $\mathbf{x}^* \in X^*$, $\mathbf{u}^* \in U(\mathbf{x}^*)$ *is a fixed point of the dynamics (A.1)-(A.2).*

Proof: From the KKT conditions, $\partial_{x_j} L(\mathbf{x}^*, \mathbf{u}^*) = 0$ and so $(\mathbf{x}^*, \mathbf{u}^*)$ is a fixed point of (A.1). Since $(\mathbf{x}^*, \mathbf{u}^*)$ is feasible, $\partial_{u_i} L(\mathbf{x}^*, \mathbf{u}^*) = g_i(\mathbf{x}^*) \leq 0$. We need to consider two cases: (i) $\partial_{u_i} L(\mathbf{x}^*, \mathbf{u}^*) = 0$ in which case $(\mathbf{x}^*, \mathbf{u}^*)$ is a fixed point of (A.2) and (ii) $\partial_{u_i} L(\mathbf{x}^*, \mathbf{u}^*) < 0$ in which case by complementary slackness $u_i^* = 0$ and this is also a fixed point of (A.2). Hence, every $(\mathbf{x}^*, \mathbf{u}^*)$ is a fixed point of the dynamics (A.1)-(A.2). $\square$

A.3 Convergence

Let $V(\mathbf{x}, \mathbf{u}) := \min_{\mathbf{x}^* \in X^*, \mathbf{u}^* \in U(\mathbf{x}^*)} \sum_{j=1}^{n}(x_j - x_j^*)^2 + \sum_{i=1}^{m}(u_i - u_i^*)^2$. Observe that (i) $V(\mathbf{x}, \mathbf{u}) \geq 0$ and (ii) $V(\mathbf{x}, \mathbf{u}) = 0$ if and only if $\mathbf{x} \in X^*$ and $\mathbf{u} \in U(\mathbf{x}^*)$.

Lemma A.3.1. *Under update (A.1)-(A.2),*

$$V(\mathbf{x}(t+1), \mathbf{u}(t+1)) \leq \Big[V(\mathbf{x}(t), \mathbf{u}(t)) - 2\alpha \left(L(\mathbf{x}(t), \mathbf{u}^*(t)) - L(\mathbf{x}^*(t), \mathbf{u}(t)) \right)$$

$$+ \alpha^2(t)\epsilon(\mathbf{x}(t), \mathbf{u}(t)) \Big]$$

where $\epsilon(\mathbf{x}, \mathbf{u}) = \sum_{j=1}^{n} \left(\partial_{x_j} L(\mathbf{x}, \mathbf{u}) \right)^2 + \sum_{i=1}^{m} g_i^2(\mathbf{x})$ *and*

$$(\mathbf{x}^*(t), \mathbf{u}^*(t)) = \arg\min_{\mathbf{x}^* \in X^*, \mathbf{u}^* \in U(\mathbf{x}^*)} \sum_{j=1}^{n}(x_j(t) - x_j^*)^2 + \sum_{i=1}^{m}(u_i(t) - u_i^*)^2$$

.

Proof: From (A.1), for any $x^* \in X^*$ we have

$$\sum_{j=1}^{n} \left(x_j(t+1) - x_j^* \right)^2$$

$$= \sum_{j=1}^{n} \left(x_j(t) - x_j^* - \alpha \partial_{x_j} L(\mathbf{x}(t), \mathbf{u}(t)) \right)^2$$

$$= \sum_{j=1}^{n}(x_j(t) - x_j^*)^2 + 2\alpha \sum_{j=1}^{n}(x_j^* - x_j(t)) \partial_{x_j} L(\mathbf{x}(t), \mathbf{u}(t))$$

$$+ \alpha^2(t) \sum_{j=1}^{n} \left(\partial_{x_j} L(\mathbf{x}(t), \mathbf{u}(t)) \right)^2$$

$$\overset{(a)}{\leq} \sum_{j=1}^{n}(x_j(t) - x_j^*)^2 + 2\alpha \left(L(\mathbf{x}^*, \mathbf{u}(t)) - L(\mathbf{x}(t), \mathbf{u}(t)) \right)$$

$$+ \alpha^2(t) \sum_{j=1}^{n} \left(\partial_{x_j} L(\mathbf{x}(t), \mathbf{u}(t)) \right)^2 \tag{A.3}$$

where (a) follows from the fact that

$$L(\mathbf{x}^*, \mathbf{u}) - L(\mathbf{x}, \mathbf{u}) \geq \sum_{j=1}^{n} \left(x_j^* - x_j \right) \partial_{x_j} L(\mathbf{x}, \mathbf{u})$$

(from the definition of a subgradient). From (A.2) we have for any $\mathbf{u}^* \in U(\mathbf{x}^*)$ that

$$\sum_{i=1}^{m} \left(u_i(t+1) - u_i^* \right)^2$$
$$= \sum_{i=1}^{m} \left(u_i(t) + \alpha g_i(\mathbf{x}(t)) - u_i^* \right)^2$$
$$= \sum_{i=1}^{m} \left(u_i(t) - u_i^* \right)^2 + 2\alpha \sum_{i=1}^{m} \left((u_i(t) - u_i^*) g_i(\mathbf{x}(t)) + \alpha^2(t) g_i^2(\mathbf{x}(t)) \right)$$
$$\overset{(a)}{=} \sum_{i=1}^{m} \left(u_i(t) - u_i^* \right)^2 + 2\alpha \left(L(\mathbf{x}(t), \mathbf{u}(t)) - L(\mathbf{x}(t), \mathbf{u}^*) \right) + \alpha^2(t) g_i^2(\mathbf{x}(t)) \quad \text{(A.4)}$$

where (a) follows from the observation that

$$L(\mathbf{x}, \mathbf{u}) - L(\mathbf{x}, \mathbf{u}^*) = f(\mathbf{x}) + \sum_{i=1}^{m} u_i g_i(\mathbf{x}) - f(\mathbf{x}) - \sum_{i=1}^{m} u_i^* g_i(\mathbf{x}) \sum_{i=1}^{m} (u_i - u_i^*) g_i(\mathbf{x})$$

Then, from (A.3) and (A.4),

$$V_{\mathbf{u}^*}(\mathbf{x}(t+1), \mathbf{u}(t+1)) \leq \sum_{j=1}^{n} (x_j(t+1) - x_j^*(t))^2 + \sum_{i=1}^{m} (u_i(t+1) - u_i^*(t))^2$$
$$\leq V_{\mathbf{u}^*}(\mathbf{x}(t), \mathbf{u}(t)) - 2\alpha \left(L(\mathbf{x}(t), \mathbf{u}^*(t)) - L(\mathbf{x}^*(t), \mathbf{u}(t)) \right)$$
$$+ \alpha^2(t) \epsilon(\mathbf{x}(t), \mathbf{u}(t))$$

$\square$

Lemma A.3.2. *Under update (A.1)-(A.2), when $\frac{1}{t} \sum_{\tau=0}^{t} \epsilon(\mathbf{x}(\tau), \mathbf{u}(\tau)) \leq M$ (e.g. this holds when $(\mathbf{x}(\tau), \mathbf{u}(\tau))$ is bounded and $f(\mathbf{x})$, $g(\mathbf{x})$ are continuous) we have*

$$0 \leq \frac{1}{t} \sum_{\tau=0}^{t} \left(L(\mathbf{x}(\tau), \mathbf{u}^*(\tau)) - L(\mathbf{x}^*(\tau), \mathbf{u}(\tau)) \right) \leq \frac{1}{2\alpha t} V(\mathbf{x}(0), \mathbf{u}(0)) + \frac{\alpha M}{2}$$

where $\mathbf{x}^*(\tau) \in X^*$ and $\mathbf{u}^*(\tau) \in U^*(\mathbf{x}^*(\tau))$.

Proof: By Lemma A.3.1,

$$V(\mathbf{x}(t+1), \mathbf{u}(t+1)) - V(\mathbf{x}(0), \mathbf{u}(0)) \leq$$
$$\sum_{\tau=0}^{t} \left(-2\alpha \left(L(\mathbf{x}(\tau), \mathbf{u}^*(\tau)) - L(\mathbf{x}^*(\tau), \mathbf{u}(\tau)) \right) + \alpha^2(t) \epsilon(\mathbf{x}(\tau), \mathbf{u}(\tau)) \right)$$

Hence,

$$\frac{1}{t} \sum_{\tau=0}^{t} \left(L(\mathbf{x}(\tau), \mathbf{u}^*(\tau)) - L(\mathbf{x}^*(\tau), \mathbf{u}(\tau)) \right) \leq$$

$$\frac{1}{2\alpha t} V(\mathbf{x}(0), \mathbf{u}(0)) + \frac{\alpha}{2t} \sum_{\tau=0}^{t} \epsilon(\mathbf{x}(\tau), \mathbf{u}(\tau))$$

For $x^* \in X^*$, $\mathbf{u}^* \in U(\mathbf{x}^*)$ recall $\mathbf{u}^* = \arg\sup_{\mathbf{u} \geq 0} L(\mathbf{x}^*, \mathbf{u})$ and $\mathbf{x}^* = \arg\inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{u}^*)$. Hence, $L(\mathbf{x}^*, \mathbf{u}) \leq L(\mathbf{x}^*, \mathbf{u}^*) \leq L(\mathbf{x}, \mathbf{u}^*)$ and $L(\mathbf{x}, \mathbf{u}^*) - L(\mathbf{x}^*, \mathbf{u}) \geq 0$. Therefore, $L(\mathbf{x}(\tau), \mathbf{u}^*(\tau)) - L(\mathbf{x}^*, \mathbf{u}(\tau)) \geq 0$. Substituting for $\frac{1}{t} \sum_{\tau=0}^{t} \epsilon(\mathbf{x}(\tau), \mathbf{u}(\tau)) \leq M$ then yields the result. $\square$

LTE/802.11 User Association: Convergence

B.1 Proof of Theorem 3.4.1

Proof: Update (3.36) generates a sequence $\{\mathbf{x}_k, k = 1, 2, \cdots\}$. By the convexity of $g^{(i)}$ we have that that $g^{(i)}(\mathbf{x}) \geq g^{(i)}(\mathbf{x}_k) + \partial g^{(i)}(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k)$ and so $C_{\mathbf{x}_k} \subset C$ (for any $\mathbf{x} \in C_{\mathbf{x}_k}$ we have $h^{(i)}(\mathbf{x}) - g^{(i)}(\mathbf{x}) \leq h^{(i)}(\mathbf{x}) - g^{(i)}(\mathbf{x}_k) - \partial g^{(i)}(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) \leq 0$, $i = 1, \cdots, m$). Hence, $\mathbf{x}_k \in C$, $k = 1, 2, \cdots$. Further, $f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k)$ and so the sequence $\{f(\mathbf{x}_k)\}$ is decreasing. Since f is convex (so continuous) and $C \subset B$ is bounded then $f(\mathbf{x}_k)$ is bounded. Hence, by the monotone convergence of bounded sequences (*e.g.* [86, Theorem 16.2]), sequence $\{f(\mathbf{x}_k)\}$ converges to a finite limit. Let f_∞ denote this limit and let $C_\infty = \{\mathbf{x} \in C : f(\mathbf{x}) = f_\infty\}$ denote the corresponding set of limit points in C.

From (3.36) we have that for every $\mathbf{x} \in C_\infty$ then $\mathbf{x} \in \arg\min_{\mathbf{x} \in C_{\mathbf{x}}} f(\mathbf{x})$ (else we could find a point $\mathbf{y} \in D_{\mathbf{x}}$ such that $f(\mathbf{y}) < f(\mathbf{x})$ contradicting the fact that $\mathbf{x}$ corresponds to a limit point of monotonic sequence $\{f(\mathbf{x}_k)\}$). It follows that $\mathbf{x}$ satisfies the Fritz John conditions,

$$K_0(\mathbf{x}) := \lambda_0 \partial f(\mathbf{x}) + \sum_{i=1}^{m} \lambda^{(i)}(\partial h^{(i)}(\mathbf{x}) - \partial g^{(i)}(\mathbf{x})) = 0 \qquad \text{(B.1)}$$

$$K_i(\mathbf{x}) := \lambda^{(i)}(h^{(i)}(\mathbf{x}) - g^{(i)}(\mathbf{x})) = 0, \ i = 1, \cdots, l \qquad \text{(B.2)}$$

with multipliers $\lambda^{(i)} \geq 0$, $i = 0, \cdots, l$. But these are also the Fritz John conditions for optimisation P, and so it follows that $\mathbf{x}$ is a stationary point of P. Since f is convex it is Lipschitz continuous on compact set C and so $f(\mathbf{x}_k) \to f_\infty$ implies that for any $\delta > 0$ and $k \geq k_\delta$, $\exists \mathbf{x} \in C_\infty$ s.t. $\|\mathbf{x}_k - \mathbf{x}\| < \delta$ provided k_δ is sufficiently large. By assumption, ∂f, $\partial h^{(i)}$ and $\partial g^{(i)}$ are continuous, so K_0, K_i are continuous. Hence, $K_0(\mathbf{x}_k) \to 0$, $K_i(\mathbf{x}_k) \to 0$, $i = 1, \cdots, l$ and we are done. $\square$

B.2 Proof of Theorem 3.4.2

Proof: Since C is compact, then for every open covering there exists a finite subcovering. Let $\cup_{\Gamma \in \Psi \subset 2^C} \Gamma$ be a covering of C consisting of balls of radius r and let $\cup_{\Gamma \in \Psi' \subset \Psi} \Gamma$ be a finite subcovering. Each set $\Gamma \in \Psi'$ either covers C or has at least one neighbour $\Sigma \in \Psi'$ such that $\Gamma \cap \Sigma \neq \emptyset$ since C is connected. Now expand each set $\Gamma \in \Psi'$ to be a ball of radius $2r$ and define a new set $\tilde{\Gamma}$ obtained by taking the union of the expanded set Γ with all of its expanded neighbours. In the new expanded covering $\tilde{\Psi}'$ so obtained neighbours $\tilde{\Gamma}$ and $\tilde{\Sigma}$ have at least a ball of radius r in common, so $\tilde{\Gamma} \cap \tilde{\Sigma}$ has volume at least $\nu(r)$ where $\nu(r)$ is the volume of a hypersphere of radius r in $\mathbb{R}^n$. Since C is connected and the sets in $\tilde{\Psi}'$ form a finite covering, between any two points $\mathbf{x}$, $\mathbf{y} \in C$ there exists a path traversing a sequence of at most $|\tilde{\Psi}'|$ neighbouring sets from $\tilde{\Psi}'$. Given $\mathbf{x}_k \in \tilde{\Gamma}$ then the probability that $\mathbf{x}_{k+1}$ lies in a neighbour $\tilde{\Sigma}$ is at least $\epsilon \eta \nu(r)$ and so every path is traversed with probability at least $(\epsilon \eta \nu(r))^{|\tilde{\Psi}'|}$. That is, starting from any initial condition for every optimum $\mathbf{x}^*$ of problem P a set W of size no greater than $\nu(2r)$ containing $\mathbf{x}^*$ is visited with probability at least $(\epsilon \eta \nu(r))^{|\tilde{\Psi}'|}$. This holds for every $r > 0$. Selecting $\xi = \nu(2r)$, then for any initial condition and every $\xi > 0$, for every optimum x^* of optimisation P a ball $B_\xi(x^*)$ containing x^* is visited with positive probability and the stated result now follows. $\quad\square$