UNIVERSITY OF TWENTE.

Master Thesis

Assessing the benefits of spectrum sharing in wireless access networks

Author:
Ivo Noppen, BSc

Supervisors:
Prof. Dr. J.L. van den Berg
Dr. Ir. G.J. Heijenk
Dr. H. Zhang
Dr. R. Litjens, MSc

In collaboration with:
Netherlands Organisation for Applied Scientic Research (TNO)

TNO

September 28, 2012

Abstract

In this thesis, we analyse the potential gain in capacity and performance of the non-orthogonal SAPHYRE project transmission schemes when simulated in a system-level simulator with a realistic model including multiple users, propagation models and traffic models. We compare the results of the simulations with results for the uncoordinated orthogonal scenario, and the coordinated orthogonal scenario, as well as for ZF beamforming in the coordinated non-orthogonal scenario. We also introduce some methods to deal with coordinated scheduling for MSR, PF and MM scheduling. Furthermore, we show in a sensitivity analysis how sensitive the SAPHYRE transmission schemes are with regards to feedback delay, feedback error and interference of surrounding cells.

We show that, with the SAPHYRE transmission schemes, an almost twofold increase in average user throughput and 10^{th} percentile throughput can be reached when compared to the uncoordinated orthogonal scenario. For the coordinated orthogonal scenario the results are lower, but still a decent improvement. Furthermore, we show that we can also increase the system throughput almost twofold when the system is fully loaded with the SAPHYRE schemes. With lower loads, the throughput decreases to the same values as the orthogonal scenarios. With respect to ZF, we show that the SAPHYRE schemes are of similar performance.

Lastly, we show that the MSR scheduling algorithm is more resilient to feedback error and interference of surrounding cells than the PF algorithm with SAPHYRE transmission schemes. Both scheduling algorithms are not affected by a delay of up to 8 Transmission Time Intervals (TTIs), for the pedestrian users included in our model.

Acknowledgements

A number of people have been very important for the realization of this thesis. First of all, my daily supervisors Dr. Haibin Zhang and Dr. Remco Litjens from TNO have been very helpful in the process and were always available as a sparring partner and as valuable colleagues in the SAPHYRE project. Prof. Dr. Hans van den Berg and Dr. Ir. G.J. Heijenk from the University of Twente have also been of great help with their guidance from the university side and good suggestions on the subject matter. The guidance and insightful comments of all four committee members were of great value and made the completion of this thesis possible.

From TNO, I would specifically like to thank Dick van Smirren and Frits Klok for their guidance on a personal- and career level. I have learnt a lot about myself and my ambitions in this period at TNO.

Last but not least, I would like to thank my family for their ongoing support during my bachelor and master studies, my friends for being there for me in times I needed it and for the fun times we had. Especially, I would like to thank my girlfriend, Sanne van Aerts, who stood by me during my whole period at University notwithstanding the physical distance between us. I could always depend on her for moral or emotional support.

Ivo Noppen
Delft, September 28, 2012

Contents

1 Introduction **1**
- 1.1 Research questions . . . 2
- 1.2 Outline . . . 3

2 A brief overview of cellular networks **5**
- 2.1 History . . . 5
- 2.2 Basic principles . . . 5
  - 2.2.1 Spectrum . . . 6
  - 2.2.2 Multiple access . . . 6
  - 2.2.3 Signal propagation . . . 7

3 State of the art in spectrum sharing **9**
- 3.1 Taxonomy of spectrum allocation . . . 9
  - 3.1.1 Exclusive use . . . 11
  - 3.1.2 Hierarchical access . . . 11
  - 3.1.3 Spectrum commons . . . 12
- 3.2 SAPHYRE . . . 12
  - 3.2.1 Orthogonal spectrum sharing . . . 12
  - 3.2.2 Non-orthogonal spectrum sharing . . . 13
  - 3.2.3 Adaptive and robust signal processing in multi-user and multi-cellular environments . . . 14
- 3.3 Conclusions . . . 15

4 Scheduling **17**
- 4.1 Concept of scheduling . . . 17
- 4.2 Throughput-optimal scheduling . . . 18
  - 4.2.1 Maximum Sum Rate scheduling . . . 19
- 4.3 Fair scheduling . . . 20
  - 4.3.1 Average historical rate . . . 20
  - 4.3.2 Proportional Fair scheduling . . . 22
  - 4.3.3 Max-Min scheduling . . . 25
  - 4.3.4 Example of Max-Min (MM) scheduling . . . 29

5 Modelling **33**
- 5.1 System model . . . 33
  - 5.1.1 Operators and users . . . 33
  - 5.1.2 Network topology . . . 34
- 5.2 Traffic model . . . 35

5.3 Bandwidth, power and interference . . . 37
5.4 Physical layer abstraction . . . 38
5.4.1 Propagation model . . . 38
5.4.2 Physical layer traces . . . 40
5.4.3 Transmission schemes . . . 41
5.4.4 From abstraction to bit rate . . . 43
5.4.5 Exponential Effective Signal to Noise Ratio Mapping (EESM) 43

6 Simulation results & analysis 47
6.1 Simulation scenarios . . . 47
6.2 Simulation parameters . . . 50
6.3 Overview of metrics . . . 51
6.4 Spectrum sharing analysis . . . 53
6.4.1 Uncoordinated orthogonal sharing (FSA) . . . 53
6.4.2 Uncoordinated non-orthogonal sharing . . . 54
6.4.3 Coordinated orthogonal sharing . . . 56
6.4.4 Coordinated non-orthogonal sharing . . . 58
6.4.5 Sharing scenario comparison . . . 61
6.4.6 Scheduling algorithm comparison . . . 64
6.5 Sensitivity analysis (coordinated non-orthogonal sharing) . . . 66
6.5.1 Sensitivity to interference of surrounding cells . . . 67
6.5.2 Sensitivity to feedback delay . . . 69
6.5.3 Sensitivity to feedback error . . . 69

7 Conclusions and future work 73
7.1 Conclusions . . . 73
7.1.1 Answer to the research questions . . . 74
7.2 Future work . . . 76

List of acronyms

3GPP	3rd Generation Partnership Project
AWGN	Additive White Gaussian Noise
BLER	Block Error Rate
BS	Base Station
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CSI	Channel State Information
CQI	Channel Quality Indicator
DSA	Dynamic Spectrum Access
DySPAN	IEEE Symposium on new frontiers in Dynamic Spectrum Access Networks
IC	Interference Channel
ISM	Industrial, Scientific and Medical
ISY	Institutionen för Systemteknik
LOS	Line of Sight
LTE	Long Term Evolution
EESM	Exponential Effective Signal to Noise Ratio Mapping
FDD	Frequency Division Duplexing
FDMA	Frequency Division Multiple Access
FhG	Fraunhofer-Gesellschaft
FSA	Fixed Spectrum Allocation
GSM	Global System for Mobile communications
ITU	International Telecommunication Union
MAC	Medium Access Control

MCS	Modulation and Coding Scheme
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
M-LWDF	Modified Largest Weighted Delay First
MM	Max-Min
MMSE	Minimum Mean Square Error
MSR	Maximum Sum Rate
NGMN	Next Generation Mobile Networks
NB	Nash Bargaining
NE	Nash Equilibrium
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PDSCH	Physical Downlink Shared Channel
PF	Proportional Fair
PRB	Physical Resource Block
QoS	Quality of Service
RR	Round Robin
SAPHYRE	Sharing Physical Resources
SB	Spectrum Broker
SINR	Signal to Interference plus Noise Ratio
SNR	Signal to Noise Ratio
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
UE	User Equipment
UHF	Ultra High Frequency
UMTS	Universal Mobile Telecommunications System
WiFi	Wireless Fidelity
WLAN	Wireless Local Area Network
WRC	World Radiocommunication Conference
ZF	Zero-forcing

Chapter 1

Introduction

The demand for mobile data communications is ever increasing. To cope with the data growth, either more spectrum is needed, or operators need to make more efficient use of the spectrum. The current way of licensing spectrum exclusively for extended periods of time does not enable operators to keep up with the data demand. Furthermore, this licensing method promotes inefficient spectrum usage because operators are bound to exclusively use the spectrum allocated to them. With the ever increasing demand for spectrum, fixed spectrum allocation does not allow or promote operators to share their excess spectrum with other operators. Spectrum sharing is the idea to make more efficient use of the spectrum by simultaneous usage of the spectrum by multiple operators. Spectrum sharing also presents an opportunity to reform the way of thinking about spectrum allocation for both operators and regulators alike.

Various methods have been developed to share spectrum between operators. Most of these methods can be categorised as orthogonal sharing. The common denominator of the orthogonal methods is that the spectrum is shared in an interference avoidance way; at any point in time and space, different users are allotted different frequencies for transmission. In this way users and base stations do not have to cope with interference. Frequency reuse is still possible in different cells. Relatively newer are the non-orthogonal sharing methods, based on interference cancellation. With these methods, frequencies can be used by multiple users at the same time. The non-orthogonal methods provide a way of dealing with the interference that is generated by simultaneous usage of frequencies. The European funded Sharing Physical Resources (SAPHYRE) FP7 project also developed an interference cancellation-based method based on joint beamforming [1]. This method aims to steer the transmission power towards the receiver and thus away from other users. Operators in this method have to be aware of the resources, demands and users of other operators in order to aid the signal processing needed for this beamforming technique. Another non-orthogonal beamforming technique is Zero-forcing (ZF) [2]. This technique is well known in literature and effectively cancels interference on channels without noise when used by multiple users.

Since operators and regulators alike are sceptical about sharing the spectrum between parties without exclusively licensing it, this research may help convince these parties of the benefits of spectrum sharing. Hence, it might change the way mobile networks are operated and help spectrum regulators to radically

alter the way they license the available spectrum.

1.1 Research questions

The transmission schemes developed by the SAPHYRE project have been evaluated at the link-level, which means that the performance of the transmission schemes has been evaluated at one communication link between a base station and a user. These link-level assessments have been proven quite promising over orthogonal sharing techniques but lack realistic aspects of network operation like scheduling, feedback delay, multi-user traffic, propagation environments and network layout. In order to realistically assess the performance of the non-orthogonal sharing methods in a real-life environment, a system-level evaluation is required. Furthermore, this enables the comparative assessment of the performance of different forms of spectrum sharing and scheduling of the whole system instead of one link (e.g. system throughput, spectral efficiency, capacity gain). In other words we try to answer the following question: *what can we gain in terms of performance and capacity at the system level, by applying the advanced transmission schemes for non-orthogonal sharing, as developed in the SAPHYRE project, with respect to Fixed Spectrum Allocation, orthogonal sharing, and non-orthogonal sharing with the ZF transmission scheme?*

We can identify the following tasks that ultimately lead to the answer to the research question:

- develop scheduling algorithms to divide the available resources over mobile users in a way that is near-optimal in its scheduling goal. Furthermore, these scheduling algorithms should be applicable to both orthogonal and non-orthogonal sharing of the spectrum and align well with the transmission schemes used at the physical layer as developed in the SAPHYRE project;
- implement the developed scheduling algorithms into a system-level simulator;
- define relevant scenario(s) and model the system parameters like propagation and a traffic model, to compare the different forms of spectrum sharing and the developed scheduling algorithms on a system level;
- define relevant metrics to compare the spectrum sharing techniques in terms of performance and capacity gain;
- evaluate the performance- and capacity gain of the non-orthogonal spectrum sharing developed in the SAPHYRE project, for selected scenarios and parameters;
- evaluate the sensitivity of the SAPHYRE transmission schemes to interference, feedback error, and feedback delay.

To evaluate the SAPHYRE spectrum sharing and signal processing techniques on a system level, a TNO proprietary system-level simulator is used. This system level simulator simulates downlink traffic (i.e. from the Base Station (BS) to the user), and includes models for propagation, transmission, scheduling, user traffic and mobility. To use this simulator for non-orthogonal

sharing, we will build on the existing simulator to include scheduling for multiple users at the same frequencies and to include the relevant model parameters which we will introduce in this study.

Close collaboration with the partners is necessary to evaluate their transmission schemes to their full potential while retaining real-life simulation parameters. Alignment between the choices made at the physical layer by SAPHYRE partners and the scheduling regarding their goals is important for fair simulation (i.e. do not use a transmission scheme with maximum throughput as the underlying goal at the physical layer while promoting fairness higher up in the scheduling algorithm). These choices are also used as a framework for the comparison of the different techniques to ensure fair comparison.

To evaluate the spectrum sharing and advanced signal processing techniques developed by the SAPHYRE at a high level, we will arrange an abstraction of the physical layer in close consultation with the SAPHYRE partners. This allows us to abstract from the implementation of the physical layer while retaining the possibility of evaluating the performance- and capacity gain at a system level.

1.2 Outline

This thesis consists of seven chapters:

Chapter 2 continues the introductory part of this thesis with a brief overview of cellular networks including main concepts and a very short history.

In Chapter 3, we take a look at the current state of the art in spectrum sharing techniques and taxonomise the different solutions according to the main literature on this subject. In the same chapter, we introduce the SAPHYRE project and outline the work regarding spectrum sharing done by this project so far. Finally, this chapter describes the interference avoidance based solution to spectrum sharing developed in the SAPHYRE project.

Chapter 4 introduces scheduling concepts and ultimately leads to the scheduling algorithms as used in the simulator.

In Chapter 5, the used models and the decisions about model parameters are outlined as a first step to the system-level evaluation of the different spectrum sharing solutions. Furthermore, the input needed from partners in the SAPHYRE project is outlined and a physical layer abstraction is established.

Subsequently in Chapter 6, the complete scenarios are outlined and the results of the simulations for these scenarios are analysed. Furthermore, a sensitivity analysis is included for selected parameters, scenarios and scheduling algorithms.

Finally, in Chapter 7, this research project report will be concluded with some final remarks about this research and a recommendation of possible future work.

Chapter 2

A brief overview of cellular networks

2.1 History

The base for all wireless communications was established by James Clerk Maxwell with his theory about computational electromagnetics. In 1887, Maxwell's theory was verified by Heinrich Hertz when he discovered electromagnetic radiation at ultra high frequencies (UHF). Maxwell's equations have since been studied over a century and are one of the most successful theories in radio science. Even Einstein found that Maxwell's theories were already relativistically correct and needed no adjustment as was the case with for instance Newtons dynamics.

The first demonstration of wireless transmission was carried out by Nicola Tesla in 1893 and subsequently by Guglielmo Marconi in 1895 and 1896. Marconi deemed Morse code sufficient for ship-to-shore and shore-to-ship communications and saw no need for voice transmissions. He did not foresee the development of radio broadcasting and left early experiments with wireless telephony to others like Reginald Aubrey Fessenden.

Fessenden continued the work of Tesla, among others, in the field of continuous wave propagation as he recognized the need of this for voice transmission. The first continuous wave transmission, however, would not take place until 1906. The first steps towards modern radio communication systems were made, abandoning Marconi's ideas that Morse code would be sufficient. Just a few months after the 1906 transmission, Fessenden and his assistants broadcast the first radio transmission including a speech by Fessenden and Christmas music played live by Fessenden on the violin. The broadcast was heard on ships from the US navy and United Fruit Company equipped with Fessenden's wireless receivers all over the Atlantic Ocean [3].

2.2 Basic principles

Fundamental to the operation of wireless networks is *spectrum*. *Multiple access* schemes divide the spectrum somehow in channels and allow multiple terminals to access the network. In order to provide wireless communications, electromag-

netic waves *propagate* through the wireless medium from transmitter to receiver influenced by omnipresent noise and faded by reflection, shadowing, etc.

2.2.1 Spectrum

The electromagnetic spectrum is a range of possible frequencies of electromagnetic radiation. This includes everything from low frequencies with a wavelength of kilometres to very short wavelengths like gamma radiation. Since spectrum is a limited resource that cannot be renewed or replenished, spectrum is typically allocated by national governments. Lately, it is often coordinated by the European union or even globally within International Telecommunication Union (ITU)'s World Radiocommunication Conference (WRC), to allow for low-cost production of communication equipment, to allow international roaming, and to manage interference between the various wireless services worldwide.

Spectrum assigned for cellular networks is typically licensed to mobile network operators for a period of ten to fifteen years in order to grant the operator the opportunity to make large investments in the networks and be able to make a long-term profit of it. Typically, these licenses are bound to rules regarding coverage and actual usage of the licensed frequencies to prevent unused frequencies.

In both Global System for Mobile communications (GSM) (900 and 1800 MHz bands) and Universal Mobile Telecommunications System (UMTS) (2 GHz band), the uplink and downlink channels are separated by Frequency Division Duplexing (FDD). With this separation in the frequency domain, the downlink channel is typically the one with the higher frequency in cellular networks because transmission over higher frequencies takes more power, which is more freely available at the BS than at the User Equipment (UE). The UMTS standard also contains spectrum where Time Division Duplexing (TDD) is used, where both up- and downlink transmissions can happen and are separated in the time domain.

Not all spectrum is licensed to operators for a specific technology: a so-called *unlicensed* band can be used without governmental permission although there are restrictions on e.g. transmission power (e.g. the Industrial, Scientific and Medical (ISM) band - 2.4 GHz). The primary advantage that this band can be used free of charge can be shown by the popularity of technologies used in this band e.g. Wireless Fidelity (WiFi), Bluetooth, baby phones and microwave ovens. Due to the power restrictions, communication technologies that coexist in this band are relatively short range and have to be able to cope with the interference caused by other technologies in this band.

2.2.2 Multiple access

For a cellular system, it is necessary to enable multiple users to be served simultaneously. In light of this requirement, several schemes have been developed to enable this. The four main schemes developed for multiple access are Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), Code Division Multiple Access (CDMA) and Orthogonal Frequency Division Multiple Access (OFDMA). These schemes range from the first generation of cellular networks to those that are being developed for future fourth generation networks.

In the FDMA scheme, users coming onto the system are assigned a frequency or a channel and their transmissions are thus physically separated. This scheme is mainly used by first generation analogue systems.

As cellular systems become digital, data can suddenly be split up in time and sent as bursts. Digitised voice data is eligible for partitioning in short bursts as the small delay does not affect speech quality. This characteristic enables organising transmissions in a number of time slots. Each subscriber that enters the system is now assigned certain timeslots in which transmissions can be scheduled. By using TDMA on top of FDMA, multiple users can be served per channel.

In the CDMA scheme, information signals are spread onto a wideband carrier using semi-orthogonal spreading codes. One of the major advantages of using CDMA is universal frequency reuse. This means that because of the spreading codes, frequencies can be reused in adjacent cells, where TDMA and FDMA interfere too much to do the same thing. This leads to more efficient frequency usage and thus to more capacity per cell.

OFDMA is a multiple access scheme that is considered for fourth generation cellular technologies as well as evolutions of third generation cellular systems. It is based around Orthogonal Frequency Division Multiplexing (OFDM), which uses a large number of closely spaced sub-carriers modulated with orthogonal low data rates into one high-rate channel eliminating interference between the sub-carriers [4]. In OFDMA, users are now associated with specific sub-carriers that carry their data.

2.2.3 Signal propagation

When a signal is transmitted wirelessly, the signal degrades during propagation from the transmitter to the receiver. This degradation of the signal is caused by three main components influencing the propagation: *attenuation*-, *shadowing*- and *multipath* losses.

Attenuation is the gradual loss of intensity of a signal we experience with transmissions over increasing distance between the transmitter and receiver. The greater the distance between the two, the greater the attenuation loss. Effects of attenuation are usually modelled by an average attenuation loss over distance according to a power law.

Shadowing is caused by objects like buildings or mountains obstructing the path between the transmitter and the receiver. Since electromagnetic signals propagate differently through these objects, this loss is experienced when there is no Line of Sight (LOS). Shadowing is frequently referred as *slow fading* because the shadowed areas tend to be quite large and the rate of change is quite slow.

Multipath fading effects are caused by the observation that usually multiple copies of the same signal are received. These multiple signals are caused by reflection, diffraction and scattering of the signal against objects. The term *fast fading* is frequently used for this type of loss because the rate of change of multipath loss is quite fast: usually only a half wavelength of movement can change the degree in which this type of loss is experienced.

Chapter 3

State of the art in spectrum sharing

Spectrum usable for communication purposes is a limited and government regulated resource that cannot be renewed or replenished. In most mobile markets, several stakeholders play a role in the allocation of the spectrum like service providers, network operators and the government. Blocks of spectrum are typically leased by an auction organised by the government to interested parties for a typical duration of ten to fifteen years. This Fixed Spectrum Allocation (FSA) scheme has two significant problems [5]:

- *Efficiency*
  The amount of usable spectrum is finite. As more services get their own fixed spectrum allocated, at some point in the future there will be no unallocated spectrum left, yielding the need for more efficient spectrum usage.

- *Deployment difficulty*
  Since allocated frequencies differ from country to country, coordination is required between stakeholders for the deployment of services. This adds to the complexity of deployment and prevents rapid deployment.

The problems in FSA both stem from the static nature of spectrum allocation. Although FSA effectively controls interference between different networks by limiting the spectrum usage, this approach lacks the ability to reuse allocated spectrum over space and time between stakeholders. This results in poor utilization and perceived scarcity of spectrum resources. Also, capacity demand of network operators typically fluctuates over time due to human patterns, calling for the need to be able to flexibly share resources.

3.1 Taxonomy of spectrum allocation

To solve the problems FSA schemes impose on wireless communication, Dynamic Spectrum Access (DSA) techniques are widely sought after in the research community. The extent of DSA techniques can easily be illustrated by

the diversity of ideas submitted to the past five editions of the IEEE Symposium on new frontiers in Dynamic Spectrum Access Networks (DySPAN).

Based on literature review, we can divide spectrum access into four models [6, 7]:

- *Command and control*
  Users get near-eternal access to the spectrum under strict usage conditions. Usually, this model is exempt of market mechanisms and therefore mainly used for military and governmental services. This model is outside the scope of DSA since there is no sharing possible whatsoever [6, 8, 9].

- *Exclusive use*
  In this model, an entity can obtain exclusive use of the spectrum under certain rules. Two variants can be distinguished: the *long-term exclusive use* model in which exclusive ownership is guaranteed for longer time and the *dynamic exclusive use* model where spectrum is managed in a finer granularity of time, space, frequency and use [10, 8, 9, 11].

- *Shared use of primary licensed spectrum* or *hierarchical access*
  The spectrum is owned by a primary user and shared with a secondary user that does not have a license. This type of sharing is designed to have minimal impact on primary users by either making use of temporal and spatial whitespace (*spectrum overlay*) or by severely limiting the transmission power of the secondary user to remain under the noise floor of the primary user (*spectrum underlay*) [12, 7, 9].

- *Open sharing* or *spectrum commons*
  While the word 'commons' suggests an open spectrum usable by everyone without government regulation (*uncontrolled commons*), this model also encompasses *cooperative and managed commons*, where the spectrum is controlled and restricted by a group of entities, and *private commons*, where the ownership of the spectrum is centralized but other entities may use the spectrum under conditions set by the owner [6, 9, 7, 13, 11, 8].

To give a more consistent overview of the existing DSA techniques, we will evaluate these schemes in terms of coordination (distributed or centralized), orthogonality (is the spectrum used exclusively by one entity at a certain point in time) and access priority (horizontal or vertical). Because spectrum sharing is the topic of this report, we will not look at the command and control and the long-term exclusive models. Table 3.1 gives an overview of the characteristics of the various spectrum sharing schemes, which will be discussed in the following sections.

Regarding *access priority*, we can distinguish between two general scenarios: horizontal sharing and vertical sharing. In vertical sharing, the spectrum is shared in a hierarchical way with different access priorities. A primary user of the spectrum can rent its excess spectrum to secondary users on a certain timescale. Spectrum pooling [14] is a good example of this approach. In horizontal sharing the spectrum is shared on an equal-priority base as is the case in e.g. Wireless Local Area Networks (WLANs).

	Coordination		Orthogonality		Access priority	
	Centralized	Distributed	Orthogonal	Non-orthogonal	Horizontal	Vertical
Dynamic exclusive use	✓	✗	✓	✗	✓	✓
Spectrum overlay	✗	✓	✓	✗	✗	✓
Spectrum underlay	✗	✓	✗	✓	✗	✓
Uncontrolled commons	✗	✗	✗	✓	✓	✗
Managed commons	✓	✓	✓	✗	✓	✗
Private commons	✓	✗	✓	✓	✗	✓

Table 3.1: Characteristics of the various spectrum access schemes.

3.1.1 Exclusive use

Under the *dynamic exclusive use model*, at any point in time and space only one entity has exclusive rights to a distinct part of spectrum. Therefore, all techniques within the dynamic exclusive use model are orthogonal schemes. A secondary spectrum market is needed for this model to be able to divide the spectrum. In this secondary market, spectrum can be bought or sold when there is under- or overcapacity for a certain operator. Coordination is centralized by the primary licensee, who acts as a spectrum broker. Depending on the activities of this spectrum broker, the type of sharing is either horizontal when the spectrum broker does not deploy own activities within the owned spectrum, or vertical when the spectrum broker is a network operator. The latter can be called vertical because the spectrum broker can decide not to share resources when those are needed for himself.

3.1.2 Hierarchical access

As the name *hierarchical access model* implies, all techniques in this model can be classified as a form of vertical sharing for the spectrum owned by a primary user will be shared in this model with a secondary user. In *spectrum underlay*, the sharing is non-orthogonal because secondary users are allowed to transmit at frequencies already in use by primary users with a very low transmit power that stays under a certain interference cap. *Spectrum overlay* however, uses a form of orthogonal sharing where secondary users only make use of spectrum not being used by primary users in time and space (i.e. white space). For both schemes, control is distributed since there is no central authority that regulates the sharing in any way. For both spectrum underlay and overlay, the secondary users have to comply with the etiquette of respectively the power requirements to keep under the noise floor of primary users and checking if the whitespace is still unused.

3.1.3 Spectrum commons

The term *spectrum commons* is not a well-defined term. Commons implies after all that the spectrum belongs to each and everyone and that it can be shared at will. However, we can define three types of spectrum commons schemes: in an *uncontrolled commons*, no entity has an exclusive license to the spectrum. Typically, only transmission power is constrained by a regulatory body. No coordination is further required to use the spectrum, making the sharing in this scheme non-orthogonal. A good example of such an uncontrolled commons is the ISM band used for WiFi, Bluetooth, etc. On the other hand, a *managed commons* is restricted by some form of coordination. This coordination can be either centralized or distributed. The coordination takes care of orthogonality of different services broadcasting in the spectrum by synchronising the right to transmit. Furthermore, the spectrum is not licensed exclusively and therefore primary users do not exist. The last subcategory, *private commons*, is a concept aimed at gradually allowing advanced technologies into licensed bands. It is a managed commons where the ownership of the spectrum lies with the licensee. This licensee, the primary user, can set its own rules with regard to usage of the spectrum. Depending on the set of rules, sharing can be orthogonal or non-orthogonal. Furthermore, sharing will most likely take place in a vertical manner since the licensee paid for the spectrum and therefore wants to exercise control over the spectrum.

3.2 SAPHYRE

The SAPHYRE project is a European Union funded FP7 project that aims to demonstrate how equal priority resource sharing in wireless networks improves spectral efficiency, enhances coverage, increases user satisfaction, leads to increased revenue for operators and decreases capital and operating expenditures [15].

The objective of the SAPHYRE project is to investigate approaches to make better use of the spectrum resources available for mobile communication services. The different options investigated are infrastructure sharing, new adaptive spectrum sharing models, efficient co-ordination and high spectral efficiency. In order to achieve spectrum sharing in the SAPHYRE project, both orthogonal sharing and non-orthogonal sharing are explored.

3.2.1 Orthogonal spectrum sharing

Orthogonal spectrum sharing or *interference avoidance-based spectrum sharing* is the case when at a specific point in time and space the same spectrum is never simultaneously used by different users. The assignment of the spectrum over operators can be done at varying timescales with direct implications on the complexity of implementation and the attainable performance. This assignment timescale ranges from years (FSA) to more dynamic forms where the spectrum is re-assigned each minute or even each millisecond.

The easiest way, but also the most inflexible, of orthogonal spectrum sharing is frequency planning. By analysing the environment and the relation of one BS to other BSs, operators carefully plan where frequencies are reused in the network to avoid interference between cells. Together with FSA, this scheme takes

care of orthogonal access to the spectrum. However, the practice of dynamically reusing frequencies in the spatial dimension is not actively researched since the introduction of 3G networks. This is mainly because with CDMA modulation one can reuse frequencies with a factor one, meaning that each cell can reuse the frequencies of the adjacent cells. In other words: all cells can use all available frequencies with CDMA modulation.

The academic community has published a considerable amount regarding orthogonal spectrum sharing. Vertical sharing is a topic frequently published about as this adapts well on the short term when spectrum is still allocated in a fixed manner [16]. Horizontal sharing is however also possible within the FSA framework, but it will be a hard task to convince operators to share their spectrum when the access to their already licensed spectrum is on an equal sharing base.

Horizontal sharing can be enabled in an orthogonal fashion through centralized coordination by a so called Spectrum Broker (SB). However, because the decision making process is dependent on many factors, this centralized approach is very likely to become unrealistic with larger network size. Two solutions are envisioned for a decentralized approach: *fully autonomous and uncoordinated* and *collaborative and distributed* [17].

In the fully autonomous and uncoordinated case, bandwidth brokering happens at individual devices in an interference avoiding way. Therefore, devices have to sense the spectrum and identify opportunities to transmit. Since opportunities can manifest in different forms (time, frequency, power, space and codes), this is quite a complex approach. For fairness purposes an etiquette is desired. Since autonomous and uncoordinated sharing depends on the characteristics of the transmission technologies used, it will be most feasible for homogeneous networks.

In the collaborative and distributed approach, collaborative groups are formed that jointly identify opportunities. Therefore, the coordination is always between small groups and is thus manageable in comparison to the centralized approach. In comparison to the fully autonomous and uncoordinated approach, some signalling is needed to coordinate devices.

To overcome the complexity of the centralized approach with SBs, an approach envisioned in [11] introduces a hierarchic trading scheme where multiple levels of SBs are defined. On the global level, spectrum is traded for long time like with FSA. However, in this approach regional markets and local markets take care of the trading of regionally or locally excess spectrum from operators on a smaller time scale. This hybrid model takes away some of the complexity of completely centralized approaches while retaining the original FSA model.

3.2.2 Non-orthogonal spectrum sharing

Non-orthogonal spectrum sharing or *interference cancellation-based spectrum sharing* is the exact opposite of orthogonal spectrum sharing. Instead of focusing on the avoidance of interference by exclusively using parts of the spectrum at a given moment in time, non-orthogonal spectrum sharing focuses on cancelling the interference between devices when frequencies are used simultaneous.

Publications about non-orthogonal spectrum sharing mainly focus on hierarchical spectrum sharing, to make spectrum owned by primary users available to secondary users. Key in these techniques is the cognitive radio [12]. Cognitive

radios are smart radios that have built-in sensing, enabling dynamic spectrum access by using their ability to observe and asses the medium and learn from their environment. Secondary users may opportunistically access the primary licensed spectrum using their cognitive radio to dynamically adapt the transmission power to keep under the maximum interference level of primary users.

To solve the problem of allowing secondary users to the spectrum, many methods base themselves on game theory to find an optimal solution to this problem [18, 19, 20]. Other optimization methods are also used to find a solution [21]. Furthermore, the approach with respect to the secondary user differs. Early approaches show the secondary users as individual entities that individually make the decision to transmit. Later approaches include joint power control and / or beamforming for multiple secondary users to make even more efficient use of the available spectrum [22].

However, most of the techniques only involve opportunistic spectrum access by the secondary user. The primary user is rarely involved in non-orthogonal sharing because there is no incentive for the primary user to be actively involved in the decision making. A new approach to include primary users is introduced in [23]. This approach combines the dynamic exclusive use and the spectrum underlay techniques: primary users do not lease whole blocks of resources exclusively to secondary users, but can adjust how much of the resource they are willing to lease by adjusting for instance the maximum allowable interference on a certain frequency. In this scheme, primary operators get rewarded for leasing more spectrum and penalized for degrading their target Quality of Service (QoS).

One prominent technique for non-orthogonal spectrum sharing is *beamforming*, enabled by the availability of multiple transmit antennas at modern BSs. The main idea behind beamforming is to steer the transmission power towards the UE and thus away from other UEs by individually scaling the transmitted signal at different antennas of the BS. Effectively the interference is managed in space instead of time or frequency like with orthogonal sharing and FSA. Most publications about beamforming show techniques for vertical spectrum sharing in a spectrum underlay fashion [22, 24, 25, 26] and few for horizontal sharing [27].

3.2.3 Adaptive and robust signal processing in multi-user and multi-cellular environments

Within work package three of the SAPHYRE project, task 3.1 focuses on adaptive and robust signal processing in multi-user and multi-cellular environments [1]. Instead of looking at orthogonal sharing of the available spectrum, the work focuses on developing advanced signal processing techniques on the physical layer to enable non-orthogonal sharing.

To aid the signal processing, a method is proposed to share information between operators through shared backhaul links. Information operators should be aware of includes the existence of other operators, their resources, their willingness to share these resources and their currently active users and demands. In the study, transmitters are assumed to be perfectly aware of local Channel State Information (CSI) and also aware of the channel from itself to all its (un)intended receivers. It is unrealistic to assume perfect CSI, but these

assumptions are used nonetheless to provide an upper bound to the potential gain.

To mitigate interference between users, a joint beamforming mechanism is proposed for Multiple Input Multiple Output (MIMO) systems using decentralized coordination to share CSI between transmitters. In order to do so, interference alignment based strategies are considered that render interference cancellable. Often this takes the form of maximizing Signal to Interference plus Noise Ratio (SINR) or minimizing the Minimum Mean Square Error (MMSE). This provides good rates in symmetric networks where all links are subjected to noise and interference of similar level. However, [1] argues that a better sum rate can be obtained when the egoistic and altruistic objectives are properly weighed at link level. The proposed coordinated beamforming technique achieves close to (Pareto) optimal sum rate maximization without pricing feedback from users. Simultaneously, this technique outperforms interference alignment based methods in terms of sum rate in asymmetric networks.

For the two-user Multiple Input Single Output (MISO) Interference Channel (IC), a distributed beamforming mechanism is also proposed. It is an iterative algorithm that uses the interference each transmitter generates towards the receiver of the other user as a bargaining value. Beamforming vectors are herein chosen in a distributed manner decreasing the generated interference mutually as long as both users' rates keep increasing. This algorithm can also be applied when transmitters have either instantaneous or statistical CSI at their disposal. In the former, the core optimization problem is solved in closed-form whereas in the latter the problem is solved numerically. For instantaneous CSI, the possible fractional gain is almost two throughout the measurements, meaning that the rate is almost doubled. For full-rank statistical CSI, the fractional gain is less but still higher than the orthogonal case with 1.4 to 1.7. The only exception is when low-rank statistical CSI is used, in which case the fractional gain linearly decreases from 1.7 to values below one (loss) for high Signal to Noise Ratio (SNR) above 20 dB. Compared to the Nash equilibrium, which is the overall best achiever in orthogonal sharing, this mechanism is in all cases better.

3.3 Conclusions

As we can see from literature research and the research by SAPHYRE, a variety of solutions to the spectrum sharing problem have been proposed. The general direction in spectrum sharing seems to be towards non-orthogonal forms of spectrum sharing. Where most research focuses on spectrum underlay techniques or opportunistic access, the SAPHYRE project focuses on a coordinated form of spectrum sharing by mitigating interference by means of advanced transmission schemes. This is the subject where SAPHYRE really adds value to spectrum sharing research.

What lacks in literature is a good comparison of the different forms of spectrum sharing. Most spectrum sharing schemes have been individually assessed for one or two links, but most assessments include non-realistic system settings such as artificial interference levels and lack of path loss models. We can add value with a good system-level simulation in which real-life system parameters are taken into account including good channel models, more users, and an ap-

plicable traffic model. This way we can proof not only the theoretical gain on one link, but the performance gain when a spectrum sharing scheme is used in a system as well. This includes effects caused by scheduling multiple users, amongst others, which cannot be observed when only simulating one or two links.

Chapter 4

Scheduling

Many users compete for resources in mobile networks to get their data or voice transferred. It is important that the assignment of resources is fair since there are many users, but it is also important that the resources are used efficiently because of the limited availability of spectrum. Since the channel quality differs with external influences and also differs on a per user basis, scheduling of the resources poses significant challenges.

Although scheduling algorithms have been widely researched, most research focuses on scheduling users in an orthogonal manner over the spectrum. Non-orthogonal sharing of the spectrum poses specific problems as this paradigm forces decisions to take multiple users per resource into account. This means that we need a way of comparing combinations of scheduled users to align with the scheduling goal. We extend the ideas of various scheduling algorithms to take this into account.

In this chapter, we will introduce the concept of scheduling. Subsequently, scheduling goals will be introduced, followed by the problems that arise when we need to schedule multiple users according to this goal. Finally, this will lead to specific algorithms, taking the problems into account.

4.1 Concept of scheduling

To grasp the concept of scheduling, we need to know what resources are available to the users in an Long Term Evolution (LTE) network. As mentioned before, the scarce resource we use as a medium for communication is called spectrum. The spectrum in LTE networks is divided in a number of sub-carriers; frequencies that carry signals. These sub-carriers are 15 kHz wide and make up the total spectrum assigned to LTE. A Physical Resource Block (PRB) consists in its turn of 12 of these carriers for 0.5 ms, making the total spectral width of one PRB 180 kHz. Due to the allocation of guard carriers to prevent the PRB from interfering with each other, 25 PRBs is the maximum allocation per 5 MHz of spectrum. The PRB is the smallest unit of allocation and will always be allocated in time-consecutive pairs (1 ms) to one user. For this reason, when we use PRB in the rest of the text, we refer to a time-consecutive pair (1 ms) of allocation, the Transmission Time Interval (TTI).

To allow communication from a BS to the User Equipment (UE), we need to

divide the available PRB over the users that have active queues for transmission. In essence, this could be as simple as assigning all PRBs to a user that needs it at random for a certain period of time. While this would theoretically work fine, some communication might be more urgent, or users might just not be satisfied having to wait for a certain period of time to send or receive their data. To solve this problem, we need an algorithm that divides the available PRBs in time over the users in a smart way. As a large number of permutations exist to divide the spectrum, a scheduling algorithm needs to have a certain goal either from a system perspective or from a user perspective. We can roughly divide the scheduling algorithms into two approaches, according to their goal: *throughput-optimal scheduling* and *fair scheduling* [28].

In order to make the best scheduling decision according to the scheduling goal, the scheduler needs to be aware of the quality of the channel between the BS and UE, the CSI. This CSI is expressed as a Signal to Interference plus Noise Ratio (SINR), a value indicating the strength of the signal over the sum of the noise and interference and is measured by the UE. This CSI is subsequently mapped to a PRB-specific aChannel Quality Indicator (CQI) and reported by the user to the BS. The CQI is a simplification of the CSI, and can be mapped to a bit rate that can be attained with such channel quality. A higher CQI value indicates better channel quality, and translates to higher attainable bit rates. Based on the CQIs for the different users, the scheduler will make a scheduling decision. To help the decision, the scheduler can also make use of a historic average throughput.

4.2 Throughput-optimal scheduling

Throughput-optimal scheduling algorithms aim to maximise system throughput [28, 29]. This goal is reached by assigning network resources to the least "expensive" flows from a system perspective, meaning that the users with the best channel quality will get scheduled. However, this also means that users with lower channel quality may be starved because they cannot obtain high bit rates and thus do not contribute significantly to the system throughput. However, since users with better channel qualities have higher rates, their buffers are emptied faster giving room for other users during the time that these users are idle.

Although the general aim of the throughput-optimal algorithms is to schedule the user with the highest throughput, different scheduling algorithms have been invented to tackle specific problems. The *Maximum Sum Rate (MSR)* scheduling algorithm [30] aims to maximise the sum of the rates for scheduled users when scheduling in a MIMO environment where multiple antennas are used for transmission. The *Exponential Rule* algorithm [31] also aims to maximise the sum-rate, but takes the exponentially weighted queue length of each user into account. This way, users with longer queues will be prioritised over users with shorter queues when their attainable bit rate is equal. The *Modified Largest Weighted Delay First (M-LWDF)* algorithm [32] also takes the length of the queues into account, but weights the queues in a different manner.

Most of the throughput-optimal scheduling algorithms rely on the knowledge of channel conditions of the active users and their queue length. With non-orthogonal sharing all this information should be known by the scheduling entity.

Furthermore, we need a scheduling algorithm that can take care of multiple users being scheduled at one channel. Because the MSR scheduling algorithm fulfils the latter restriction, and the information exchanged between operators is minimal, we select this algorithm for further evaluation in the throughput-optimal category.

4.2.1 Maximum Sum Rate scheduling

The Maximum Sum Rate (MSR) scheduling algorithm is not very complex, and relies on little information to make its scheduling decisions. The algorithm can schedule multiple users or antennas, making it a suitable scheduling algorithm for both orthogonal and non-orthogonal scheduling. The basic idea behind the scheduling algorithm is to make the scheduling decisions in a way that the sum of attainable bit rates for a combination of scheduled users is the maximum sum of bit rates for all combinations of users in a given TTI for a given PRB. Equation 4.1 shows this mathematically: for users i and j from different operators, choose the maximum combined rate $r_i + r_j$ at time t.

$$\arg\max_{i,j}(r_i(t) + r_j(t)), i \in \{1, \ldots, N_A\}, j \in \{1, \ldots, N_B\} \tag{4.1}$$

Note that we can reduce the formula to an orthogonal scheduling decision by only selecting one user, and setting the remaining rate to zero.

To calculate the best scheduling combination, the scheduler considers all combinations of users with non-empty buffer and looks up their attainable bit rates for the current TTI and PRB. The pair of users with the maximum joint rate (the sum of the attainable rates) is saved in a vector with scheduling decisions. This process is repeated for each PRB, and the final list with scheduling choices is used to adjust all users' PRB assignments, which will be used when we calculate the final bit rates and the Block Error Rate (BLER).

As the scheduling algorithm purely relies on attainable bit rates and not on user-dependent average rates or queue lengths, we can straightforwardly schedule the scheduling combination with the maximum sum-rate on a given PRB. Note that we do not take power constraints at the UE into account as we only simulate downlink traffic i.e. from BS, where power is abundant, to UEs.

Example of MSR scheduling

Assume we have a system with three PRBs and two users per operator with an active transmission queue. The spectrum is used non-orthogonally, so all users can be scheduled either in isolation on a certain PRB or in each combination of one user of both operators. Table 4.1 gives the attainable bit-rates for the users in the current TTI. The following steps are taken to schedule the users:

- Calculate the sum of each scheduling combination for PRB 1;
- Select the highest of these sums (1154 for the combination (A_1, B_2));
- Assign PRB 1 to user A_1 and B_2;
- Calculate the sum of each scheduling combination for PRB 2;
- Select the highest of these sums (915 for the combination (A_2, B_2));

Combi / PRB	$\emptyset$ B_1	$\emptyset$ B_2	A_1 $\emptyset$	A_2 $\emptyset$	A_1 B_1	A_1 B_2	A_2 B_1	A_2 B_2
1	N/A	N/A	745	406	600	550	250	350
	615	804	N/A	N/A	260	604	243	706
2	N/A	N/A	632	576	496	451	491	526
	459	754	N/A	N/A	179	462	186	389
3	N/A	N/A	634	382	486	541	244	348
	647	561	N/A	N/A	334	433	387	485

Table 4.1: Attainable bit rates in kbps for one TTI. Bit rates are given for each user in the scheduling combinations (denoted by "combi") as outlined in the header row.

- Assign PRB 2 to user A_2 and B_2;
- Calculate the sum of each scheduling combination for PRB 3;
- Select the highest of these sums (974 for the combination (A_1, B_2));
- Assign PRB 3 to user A_1 and B_2;
- As all PRBs are assigned, the scheduling is done.

4.3 Fair scheduling

An obvious drawback of throughput-optimal scheduling is the lack of fairness between users as the user with the highest rate will always get the channel, i.e. typically the user nearest to the BS. Fair scheduling algorithms differ from throughput-optimal scheduling algorithms in the sense that they explicitly promote some degree of fairness between users. This does not mean that at any point in time the allocation of resources should be equal, but on the longer term the resources will be fairly distributed. The fairness can either be related to the number of PRBs assigned to users, or to the bit rates users can attain. The former is used in *Round Robin (RR)* scheduling, in which a PRB is assigned to the top of the stack of users after which this user is appended to the bottom of the stack. The latter is used within *Max-Min (MM)* scheduling. *Proportional Fair (PF)* scheduling tries to balance two competing interests: maximising system throughput and providing a minimum level of service to users. In this section, we will discuss both PF and MM scheduling.

4.3.1 Average historical rate

Both the Proportional Fair and Max-Min scheduling algorithms make use of the average historical rate $\hat{R}$ of the user to divide the spectrum over the active users. While this rate could be calculated over a certain time window, this approach imposes severe implementational complexity. When we would calculate the average rate over a window of say 1000 TTI, we would have to store 1000 experienced rates for all UEs. Instead, we can use exponential smoothing to calculate the average historic rate. Exponential smoothing takes all history

into account, but applies more weight to recent values (Equation 4.2). The main difference is that only one value has to be stored per UE: the average historical rate calculated in the last TTI. As is visible from Equation 4.2, the exponential smoothing formula uses the average historical rate updated from last TTI and adds the attained rate r. Both variables are weighted by the α parameter that controls the smoothing. When we set the α parameter close to 0, we get a smooth average which really depends on the long term, and when we set it closer to 1 the average is less smooth and reflects more the shorter term. Usually, this parameter is set to 0.001.

$$\hat{R}(t) = \alpha r(t-1) + (1-\alpha)\hat{R}(t-1) \tag{4.2}$$

One of the problems of working with this smoothed average takes place during the scheduling itself. As the scheduling algorithm depends on the smoothed average to calculate priority of one user over another, the results will be different when we only update the smoothed average historical rate once per TTI versus after each scheduling decision (each PRB assignment). After all, when we decide to schedule a certain user, its average rate will increase while decreasing that of other users. When we do not update the smoothed average historical rate in-between scheduling steps, the algorithm is solely dependent on the attainable rates of users while not taking the implications of its scheduling into account. There are three ways to deal with the smoothed average historical rate in between PRB assignments:

- *Calculate $\hat{R}(t)$ only once for each UE at the beginning of each TTI*
  The smoothed average rate is updated only once for each UE at the beginning of each TTI, with the experienced rates in the last TTI. A drawback of this method is when a UE has a significantly low smoothed rate and strong channels, most PRBs will be assigned to this UE, decreasing short-term fairness. As an advantage however, in the following TTI, the smoothed average rate will have been corrected and the UE scheduled accordingly.

- *Calculate $\hat{R}(t)$ after each PRB assignment for all UEs*
  Instead of updating once at the beginning of each TTI, we can update the smoothed average rate at the beginning of each TTI and after each PRB assignment with the expected instantaneous rate in the TTI that is currently being scheduled, taking all scheduled PRBs for the UE into account i.e. use $\sum r(t)$. Unfortunately, this method does not reflect the actual average smoothed rate the UE will have at the end of the TTI as for users that are not scheduled, the average smoothed rate should decrease. However, an advantage is that the method is computationally inexpensive.

- *Calculate $\hat{R}(t)$ with EESM after each PRB assignment for all UEs*
  A computationally complex method is to use the EESM model, which is explained in Chapter 5, to calculate the actual bit rate the UE will likely experience in the TTI and use this value instead of the sum of the instantaneous rate. This method takes all scheduled PRBs into account and calculates the bit rate for the current assignment for each individual user. It has the advantage of generating an accurate prediction of the expected smoothed average rate. This in turn means that the priority

level calculated with this average rate will also be more accurate, yielding a better division of resources over UEs. The drawback of this method is added computational complexity as the recalculation of the smoothed average is more involved than with the other methods.

As calculating the smoothed average historical rate with the last method gives the most accurate prediction and is thus expected to schedule most optimally towards the goal of the scheduling algorithm, we choose this method to calculate the smoothed average historical rate. Note that not only the smoothed average historical rates for scheduled users are updated, but also for all others as an expected instantaneous rate of zero also has impact on the smoothing average.

4.3.2 Proportional Fair scheduling

Proportional Fair (PF) scheduling aims to maximise the system throughput while retaining fairness between users [28, 33, 34, 35, 36]. In order to do so, the scheduling algorithm makes use of a priority index. After computing this priority index for the total set of active users $i \in \{1, 2, \ldots, N\}$, the scheduling algorithm will choose the user with the highest priority index to be scheduled (Equation 4.3). The priority index is a ratio of the attainable bit rate r_i and the average historical rate $\hat{R}_i$. As such, a user that can provide a good attainable rate over its average historical rate will have a better chance to be scheduled than a user with low attainable rate compared to its average historical rate. The PF scheduling algorithm can be tuned with the α and β parameters to strike a balance between the throughput and the fairness objective of this algorithm. With $\alpha = 0$ and $\beta = 1$ we get a Round Robin (RR) scheduler, and with $\alpha = 1$ and $\beta = 0$ the algorithm will always choose the user with the best channel conditions. As we are interested in the compromise of the algorithm, we choose $\alpha = 1$ and $\beta = 1$ to strike the best balance between the two objectives.

$$\arg\max_i \frac{r_i(t)^\alpha}{\hat{R}_i(t)^\beta}, i \in \{1, \ldots, N\} \tag{4.3}$$

After a scheduling decision is made, all the priority indices will change as the average rate of all the active users are corrected with the attainable bit rate. This means that users that were not scheduled will see their average historical rate drop slightly, rendering the chance for them to be scheduled a little bit higher for the next scheduling decision. Because the average historical rates change after each scheduling choice, we need to consider all scheduling combinations on each currently unassigned PRB after a scheduling decision has been taken. This means that the complexity of the algorithm will be higher than the maximum sum-rate algorithm as we loop over the PRBs multiple times for each TTI.

In practice this means that the algorithm considers all possible combinations of users for all unscheduled PRBs, and selects the combination of a PRB and scheduled users that yields the highest priority. After having made this decision, the historical averages of all the users in the system are updated with the instantaneous rate for their current PRB assignment. Note that this is only an expectation of the progression of the historical average as the scheduling is not definitive and we do not know yet whether the transmission of the user succeeds.

To schedule the next PRB, the algorithm again considers all unscheduled PRBs and repeats this process until all PRBs are associated with a scheduling decision. This means that the order in which the PRB are assigned is not pre-determined, but is dependent on the priority indices.

When the scheduling is uncoordinated (each BS schedules their own users), or when the spectrum sharing is orthogonal, each scheduling choice is straightforward for the scheduling algorithm as we only have to regard one user from one operator at a time. For coordinated non-orthogonal scheduling however, we end up with a priority index for each user involved in a certain scheduling combination. As it is not directly apparent how to schedule according to the proportional fair philosophy when we have two priority indices, we consider various options regarding the calculation of a single integrated priority index for these combinations of users.

- *Consider the highest priority index (PFMax)*
  One way to get rid of the fact that we have multiple priority indices in each scheduling decision, is to just consider the maximum value $P = max(P_a, P_b)$ of the two priority indices, where P_a is the priority index of the user of operator A and P_b for the user of operator B. For each scheduling combination, the highest priority index would be considered as the actual priority index for this scheduling combination. The drawback of this method is that the scheduling decision is not based on both priority indices and thus might not be fair towards users of both BSs if there are several users with high channel quality on one of the two BSs.

- *Consider the multiplication of priority indices (PFProduct)*
  If we multiply the priority indices of the UEs involved in scheduling combinations, we get a combined priority $P = P_a * P_b$. If the instantaneous rate for a certain UE is lower than the average rate, this decreases the priority index of that scheduling decision. Conversely, if the instantaneous rate is higher than the average rate, the priority increases.
  A problem might be the systematic decreasing of the joint priority index by users with a low priority index. This problem is apparent when we look at a scheduling combination of a user with a high priority index and a user with a very low priority index (below 1 due to a bad attainable rate). The low priority index will have a big effect on the single integrated priority index as the product of these two priority indices will be lower than the index of the high priority user. When all possible scheduling combinations for the high priority user are combinations with low priority users, the algorithm might prefer to schedule the user orthogonally, decreasing the system throughput as the advantage of non-orthogonal sharing is not used.

- *Consider the sum of priority indices (PFSum)*
  Another method is to take the sum of the two priority indices $P = P_a + P_b$. Using this method, a user with a very low priority index will never decrease the single integrated priority index as is possible in the PFProduct scheme. As a result, it is more likely in this scheme that the spectrum is shared between users of different BSs with a case as described in the PFProduct method. The fact that this method seems to focus more on both users than

the PFProduct and PFMax method, raises the expectation that PFSum will outperform PFMax and PFProduct.

- *Consider a combined ratio (PFCombi)*
  Another possible way to calculate the priority for scheduling multiple users in one PRB is to compute a combined ratio of the instantaneous rates and the average historical rates. The formula $\frac{r_a+r_b}{\hat{R}_a+\hat{R}_b}$, takes a more weighted approach to calculating the priority index than multiplication of the ratios. Most importantly, this formula is true to the original idea of the proportional fair algorithm in the sense that the formula weighs the instantaneous rate and the average rate of both UEs.

The PFMax scheme where only the priority index of one UE is taken into account disregarding the other UE's priority is not selected for evaluation because this scheme focuses too much on only one user. Also, as it is yet unknown which of the methods PFSum, PFProduct, and PFCombi will yield the best result, all three proportional fair algorithms are included in the simulation scenarios (see Figure 6.1).

One possible problem of PF scheduling in general from a system-level point of view is the focus on fairness instead of system throughput. Although the fairness between users will most likely be better than with MSR scheduling, the system throughput might suffer from the improved fairness.

Theoretically, two scheduling combinations could yield the same single integrated priority index. If we keep one UE at a fixed priority index P_a and two other UEs have the same priority index but different values for their rate and average historical rate, the scheduling can be a tie, e.g. $P_a * \frac{2}{1} = P_a * \frac{4}{2}$. From a system perspective it would make more sense to schedule the option with the UE that has a higher instantaneous rate to maximise system throughput. In practice, the situation that two scheduling combinations yield the exact same single integrated priority index is very small. In the unlikely case that it does happen, the scheduling algorithms choose one of these highest rates at random.

Example of PF scheduling

To provide a consistent example of the various scheduling algorithms, we build on the same example we used for MSR scheduling. Therefore, we use the attainable bit rates as shown in Table 4.1. Furthermore, as the PF scheduling algorithms depend on the average historical rates of the users, we use the historical average rates for the four users as outlined in table 4.2a. Scheduling for the different PF algorithms is largely the same, but depends on different input data. In the following example, we will outline the scheduling steps and indicate the differences for each PF algorithm.

- First, we identify which PRBs are unscheduled as of yet. As we begin with all PRBs unscheduled, this is the set $\{1, 2, 3\}$.
- For all PRBs in the set of unscheduled PRBs, we now calculate the priority indices for all scheduling combinations.
  - For PFMax, we calculate the individual priority indices for all users (Table 4.2b).

  - For PFSum, we use the individual priority indices like in PFMax and take the sum of these priority indices for each scheduling combination, yielding the single integrated priority indices outlined in Table 4.2c.
  - For PFProduct, we take the product of both priority indices as calculated for PFMax (Table 4.2d).
  - For PFCombi, we skip the intermediary step of generating the separate priority indices as we do not need the individual priority indices but are instead interested in the combined priority index, yielding Table 4.2e.

- Now that we have all priority indices for all scheduling combinations, we can choose the best one depending on the algorithm we selected. In order to do so, we search for the highest priority index over all scheduling combinations over all unscheduled PRBs.
  - With PFMax, this yields the scheduling combination $(\emptyset, B_1)$ at PRB 3.
  - With PFSum, we get the scheduling combination (A_1, B_1) at PRB 1.
  - PFProduct: $(\emptyset, B_1)$ at PRB 3.
  - PFCombi: $(\emptyset, B_1)$ at PRB 3.

- As we have our first scheduling decision, we can remove the selected PRB from the set of unscheduled PRBs. This results in the set $\{2, 3\}$ for the PFSum algorithm, or $\{1, 2\}$ for the other algorithms.

- With the first scheduling choice in hand, we can update the average historical rates for all users to reflect what we expect the average historical rates to be after scheduling. The attainable bit rates used for calculation of this average can be found in Table 4.1. It is important to note that in each scheduling step, the original average historical rates are used, and the sum of the PRB-specific attainable bit rate and the bit rates for already scheduled PRBs, to calculate the priority indices. For example, for PFSum this means the following update:

$$\begin{aligned}
\hat{R}_{A1} = \quad & 0.001 * 600 + 0.999 * 550 && = 550.05 \\
\hat{R}_{A2} = \quad & 0.001 * 0 + 0.999 * 512 && = 511.49 \\
\hat{R}_{B1} = \quad & 0.001 * 260 + 0.999 * 389 && = 388.87 \\
\hat{R}_{B2} = \quad & 0.001 * 0 + 0.999 * 830 && = 829.17
\end{aligned}$$

- Now, the algorithm repeats itself for the set of unscheduled PRBs, with the new expected average historical rates as input for the calculation of the priority indices.

4.3.3 Max-Min scheduling

MM scheduling is achieved when an algorithm aims to maximise the minimum long term average throughput of users [34]. It does so by allocating network

A_1	A_2	B_1	B_2
550	512	389	830

(a) Average historical rates

Combi / PRB	$\emptyset$ B_1	$\emptyset$ B_2	A_1 $\emptyset$	A_2 $\emptyset$	A_1 B_1	A_1 B_2	A_2 B_1	A_2 B_2
1	N/A	N/A	1.35	0.79	1.09	1.00	0.49	0.68
	1.58	0.97	N/A	N/A	0.67	0.73	0.62	0.85
2	N/A	N/A	1.15	1.13	0.90	0.82	0.96	1.03
	1.18	0.91	N/A	N/A	0.46	0.56	0.48	0.47
3	N/A	N/A	1.15	0.75	0.88	0.98	0.48	0.68
	1.66	0.68	N/A	N/A	0.86	0.52	0.99	0.58

(b) Individual priority indices $P_A = \frac{r_A}{\bar{R}_A}$ and $P_B = \frac{r_B}{\bar{R}_B}$ for all scheduling combinations

Combi / PRB	$\emptyset$ B_1	$\emptyset$ B_2	A_1 $\emptyset$	A_2 $\emptyset$	A_1 B_1	A_1 B_2	A_2 B_1	A_2 B_2
1	1.58	0.97	1.35	0.79	1.76	1.73	1.11	1.53
2	1.18	0.91	1.15	1.13	1.36	1.38	1.44	1.50
3	1.66	0.68	1.15	0.75	1.74	1.50	1.47	1.26

(c) Single integrated priority indices $P = P_A + P_B$ (PFSum)

Combi / PRB	$\emptyset$ B_1	$\emptyset$ B_2	A_1 $\emptyset$	A_2 $\emptyset$	A_1 B_1	A_1 B_2	A_2 B_1	A_2 B_2
1	1.58	0.97	1.35	0.79	0.73	0.73	0.30	0.58
2	1.18	0.91	1.15	1.13	0.41	0.46	0.46	0.48
3	1.66	0.68	1.15	0.75	0.76	0.51	0.48	0.40

(d) Single integrated priority indices $P = P_A * P_B$ (PFProduct)

Combi / PRB	$\emptyset$ B_1	$\emptyset$ B_2	A_1 $\emptyset$	A_2 $\emptyset$	A_1 B_1	A_1 B_2	A_2 B_1	A_2 B_2
1	1.58	0.97	1.35	0.79	0.92	0.84	0.55	0.79
2	1.18	0.91	1.15	1.13	0.72	0.66	0.75	0.68
3	1.66	0.68	1.15	0.75	0.87	0.71	0.70	0.62

(e) Single integrated priority indices $P = \frac{r_A + r_B}{\bar{R}_A + \bar{R}_B}$ (PFCombi)

Table 4.2: An overview of the average historical rates of the users and their priority indices for one TTI, given for the different PF scheduling schemes.

resources in such a way that the bit rate of a flow cannot be increased without decreasing the bit rate of another flow with a smaller bit rate [35]. In a sense, this means that the algorithm gives some form of priority to users with a smaller bit rate. MM scheduling does not promote throughput for individual users nor does it promote efficient usage of the whole spectrum. This means that this method will most likely yield an overall lower throughput with higher fairness between the users. The following procedure represents the general idea of the algorithm:

1. Start from a bit rate equal to zero for all flows;
2. Increase the rates of all flows by assigning resources to the users until the bit rate of one of the flows cannot be increased any more; freeze the bit rate of this flow;
3. Apply step 2 to non-frozen flows until the bit rate one of these flows is constrained, and repeat until all resources are divided.

For MM scheduling, the same complexity arises to find the most optimal scheduling decision as is the case with PF scheduling: we need to consider all unscheduled PRBs after each scheduling decision. Instead of using a priority index as the PF scheduler does, the MM algorithm is directly applied to the bit rates of the users. For each scheduling decision, the algorithm has to find the scheduling decision that maximises the minimum bit rate over all users.

We can identify two choices in MM scheduling regarding the timescale for which we want the algorithm to be fair. On the one hand we can aim to maximise the minimum instantaneous bit rate in the current TTI, providing short-term fairness. On the other hand, we can maximise the minimum average historical rate over all users for longer term fairness. With the latter option, newly arriving downloads of users have a higher chance to be scheduled as their initial average rates are non-existent. This implied priority will last until the rates of all users converge again as the algorithm will first dedicate all resources to the arriving user as this user has the minimum rate at that moment. With the short term option, this priority effect for new spurts does not exist as in each TTI all users start with a zero bit rate again. When a user with significantly weak SINR is present in the system, MM will devote many resources to this user as it is trying to maximise its bit rate since it is the smallest bit rate in the system. This might result in lower average throughput for all users. As we will compare the MM scheduling algorithm with PF, it is adamant that we make similar choices in fairness objective. Therefore, we choose to promote fairness on the longer term by making use of the average historical bit rates.

Within the MM objective, we identify two different algorithms to schedule the users. One computationally complex algorithm and a simpler algorithm. The simple algorithm might not deliver the optimal solution, but still aims to maximise the minimum rate.

- *Simple MM algorithm*
  In the simple algorithm for MM scheduling, we first select the user with the lowest average historical rate. Subsequently, we consider for all unscheduled PRBs the attainable bit rate for this user in each scheduling combination containing this user. Subsequently, we assign the PRB to

the user in which the user has the highest attainable bit rate. These steps are subsequently repeated, after recalculating the expected average historical rate for all users based on the already scheduled PRBs. This process stops when no PRBs are available any more that were not already associated with a scheduling combination.
The simplicity of this algorithm lies in the fact that we do not consider the impact of the different scheduling choices on the average rates of other users. It is possible that a scheduling decision with this simple algorithm decreases the rate of another user. Furthermore, the scheme does not take the attainable bit rate of the secondary user into account with non-orthogonal sharing. As a consequence, this algorithm will not necessarily yield the optimal MM scheduling goal as different combinations of users might have yielded a higher minimum rate over all users. Overall complexity will be lower than the advanced algorithm as we only consider the scheduling combinations that include the preselected user instead of all scheduling combinations.

- *Advanced MM algorithm*
  Instead of considering only one user for our scheduling decision, we can also consider a combination of users. Furthermore, instead of only considering the attainable bit rate when we search for the scheduling decision, we can also take into account the effect of the various scheduling decisions on the bit rates of all users. This can be done by pre-calculating the expected average historical bit rates for all users in each scheduling combination at each PRB. This will show the negative effects on the average historical bit rates when a user does not appear in the scheduling combination or when the attainable bit rate is lower than the historical average bit rate, as well as the positive effects when a user appears in the scheduling combination and the attainable bit rate is higher than the historical average. From these pre-calculated expected average historical bit rates, we can select the lowest average historical rate over all users, for each scheduling combination. Subsequently, from the selected information, we can choose the scheduling combination that yields the highest minimum rate. This way, we take the effect of scheduling combination on all users into account when selecting the best scheduling decision. This process is done for all unscheduled PRBs at once, and repeated until all PRBs are associated with a scheduling decision. In-between the scheduling decisions, the average historical rate is recalculated, just as in the simple algorithm.
  Although this scheme is quite complex, it will likely perform better in maximising the minimum rate than the simple algorithm as this algorithm also chooses the best option when the average historical rates are really close. Furthermore, the algorithm does not only consider the instantaneous rate to make its scheduling decision, but rather considers the consequences of scheduling a certain combination of users on the average historical bit rates.

As the advanced algorithm is more involved than the simple algorithm and truly considers all user combinations, we choose this algorithm for the implementation of the MM scheduling. It can be expected that simulations done with the MM algorithm will run longer due to their complexity and to their potentially lower rates, which is compensated by increased fairness.

4.3.4 Example of MM scheduling

Advanced algorithm For the MM scheduling example, we again use the same attainable bit rates as used for the MSR example as outlined in Table 4.1. Furthermore, we use the average historical rates as outlined in Table 4.3a. Next, we will show an example for the advanced algorithm.

- First, we identify which PRBs are unscheduled as of yet. As we begin with all PRBs unscheduled, this is the set $\{1, 2, 3\}$.
- For all PRBs in the set of unscheduled PRBs, we now calculate the expected average historical rates for each user for all scheduling combinations. These average historical rates are shown in Table 4.3b.
- For each scheduling combination and PRB, we now select the minimum average historical rate, shown by the bold rates.
- Of all the minimum rates, we now select the scheduling combination and PRB that yields the maximum of the minimum rates. In this case, the maximum minimum rate is reached with scheduling (A_2, B_2) on PRB 2.
- As we have our first scheduling decision, we can remove the selected PRB from the set of unscheduled PRBs, resulting in the set $\{1, 3\}$.
- Now, for the next scheduling decision, we recalculate the table of average historical rates for the unscheduled PRBs. As with PF scheduling, we use the original average historical rate, combined with the sum of the scheduled bit rate and the attainable bit rate, e.g. for scheduling choice (A_2, B_1) on PRB 1, this yields the following average historical rates:

$$\begin{aligned} \hat{R}_{A1} = &\quad 0.001 * (0 + 0) + 0.999 * 550 &&= 549.45 \\ \hat{R}_{A2} = &\quad 0.001 * (526 + 250) + 0.999 * 389 &&= 389.39 \\ \hat{R}_{B1} = &\quad 0.001 * (0 + 243) + 0.999 * 512 &&= 511.73 \\ \hat{R}_{B2} = &\quad 0.001 * (389 + 0) + 0.999 * 389 &&= 389.00 \end{aligned} \tag{4.4}$$

- With the new table of average historical rates (Table 4.3c), we repeat the process.
- As we can see, in this case scheduling choice (A_2, B_2) on both PRB 1 and 3 yield the maximum minimum rate. In such cases, the algorithm randomly selects one of the choices. If the choice were between a scheduling choice where users of both operators are scheduled and a scheduling choice with only one scheduled user, the algorithm would choose the former.

Simple algorithm For the simple algorithm, we will show a case where it does not select the maximum minimum rate because the algorithm does not work with the expected average rates.

- Assume that we are in the begin situation as in the beginning of the advanced algorithm. Our average historical rates are as specified in Table 4.3a.

- We select the user with the lowest average historical rate: either user B_2 or user A_2. We randomly pick user A_2.
- For this user, we select the highest attainable bit rate for the unscheduled PRBs from Table 4.1. This yields a bit rate of 576 on PRB 2 for scheduling decision $(A_2, \emptyset)$.
- As we have our first scheduling decision, we can remove the selected PRB from the set of unscheduled PRBs, resulting in the set $\{1, 3\}$.
- Now, the average historical rates are recalculated, taking the scheduling decision into account:

$$\begin{aligned} \hat{R}_{A1} = & \quad 0.001 * 0 + 0.999 * 550 & = & 549.45 \\ \hat{R}_{A2} = & \quad 0.001 * 576 + 0.999 * 389 & = & 389.19 \\ \hat{R}_{B1} = & \quad 0.001 * 0 + 0.999 * 512 & = & 511.49 \\ \hat{R}_{B2} = & \quad 0.001 * 0 + 0.999 * 389 & = & 388.61 \end{aligned} \tag{4.5}$$

- For the second decision, we select the user with the just recalculated lowest average historical rate from Equation 4.6: user B_2.
- For this user, we select the highest attainable bit rate for the unscheduled PRBs from Table 4.1. This yields a bit rate of 804 on PRB 1 for scheduling decision $(\emptyset, B_2)$.
- Again, we can remove the selected PRB from the set of unscheduled PRBs, resulting in the set $\{3\}$.
- Subsequently, the average historical rates are recalculated again, taking all scheduling decisions up until now into account:

$$\begin{aligned} \hat{R}_{A1} = & \quad 0.001 * (0 + 0) + 0.999 * 550 & = & 549.45 \\ \hat{R}_{A2} = & \quad 0.001 * (0 + 576) + 0.999 * 389 & = & 389.19 \\ \hat{R}_{B1} = & \quad 0.001 * (0 + 0) + 0.999 * 512 & = & 511.49 \\ \hat{R}_{B2} = & \quad 0.001 * (804 + 0) + 0.999 * 389 & = & 389.42 \end{aligned} \tag{4.6}$$

As we can see in Equation 4.6, the simple algorithm has a minimum average historical rate of 389.19 after just two scheduling decisions where the advanced algorithm chooses a scheduling combination where the minimum average historical rate is 389.49 (see Table 4.3c at scheduling combination (A_2, B_2). Furthermore, the simple algorithm tends to choose scheduling combinations with only one user because the rates for scheduling in an orthogonal fashion will likely be higher for users due to diminished interference, but lower in total sum rate for system throughput.

A_1	A_2	B_1	B_2
550	389	512	389

(a) Average historical rates

PRB/User \ Combi		$\emptyset$ B_1	$\emptyset$ B_2	A_1 $\emptyset$	A_2 $\emptyset$	A_1 B_1	A_1 B_2	A_2 B_1	A_2 B_2
1	A_1	549.45	549.45	550.20	549.45	550.05	550.00	549.45	549.45
	A_2	**388.61**	**388.61**	**388.61**	389.02	**388.61**	**388.61**	388.86	**388.96**
	B_1	512.10	511.49	511.49	511.49	511.75	511.49	511.73	511.49
	B_2	**388.61**	389.42	**388.61**	**388.61**	**388.61**	389.22	**388.61**	389.32
2	A_1	549.45	549.45	550.08	549.45	549.95	549.90	549.45	549.45
	A_2	**388.61**	**388.61**	**388.61**	389.19	**388.61**	**388.61**	389.10	389.14
	B_1	511.95	511.49	511.49	511.49	511.67	511.49	511.67	511.49
	B_2	**388.61**	389.37	**388.61**	**388.61**	**388.61**	389.07	**388.61**	**389.00**
3	A_1	549.45	549.45	550.08	549.45	549.94	549.99	549.45	549.45
	A_2	**388.61**	**388.61**	**388.61**	388.99	**388.61**	**388.61**	388.86	**388.96**
	B_1	512.14	511.49	511.49	511.49	511.82	511.49	511.88	511.49
	B_2	**388.61**	389.17	**388.61**	**388.61**	**388.61**	389.04	**388.61**	389.10

(b) Expected average historical rates for the first scheduling decision. Rates in bold are the minimum rates for the given scheduling combination on the given PRB

PRB/User \ Combi		$\emptyset$ B_1	$\emptyset$ B_2	A_1 $\emptyset$	A_2 $\emptyset$	A_1 B_1	A_1 B_2	A_2 B_1	A_2 B_2
1	A_1	549.45	549.45	550.20	549.45	550.05	550.00	549.45	549.45
	A_2	389.14	**389.14**	389.14	389.54	389.14	**389.14**	389.39	**389.49**
	B_1	512.10	511.49	511.49	511.49	511.75	511.49	511.73	511.49
	B_2	**389.00**	389.80	**389.00**	**389.00**	**389.00**	389.60	**389.00**	389.71
3	A_1	549.45	549.45	550.08	549.45	549.94	549.99	549.45	549.45
	A_2	389.14	**389.14**	389.14	389.52	389.14	**389.14**	389.38	**389.49**
	B_1	512.14	511.49	511.49	511.49	511.82	511.49	511.88	511.49
	B_2	**389.00**	389.56	**389.00**	**389.00**	**389.00**	389.43	**389.00**	**389.49**

(c) Expected average historical rates for the second scheduling decision after the first yielded (A_2, B_2) on PRB 2. Rates in bold are the minimum rates for the given scheduling combination on the given PRB

Table 4.3: An overview of the average historical rates of the users at the beginning, for the first scheduling decision, and for the second scheduling decision.

Chapter 5

Modelling

In order to realistically assess the benefits of the non-orthogonal sharing method on a system level, we need to carefully model scenario aspects to reflect a realistic, real-life setup. This does not only include parameters like the used network topology and the propagation model, but also includes the radio resource management taking the advanced transmission schemes developed in the SAPHYRE project into account.

In Table 5.1, we provide an overview of the parameters as used in the simulations. In this chapter, we explore the underlying assumptions of the selected parameters and the reasoning behind the applicability of the parameters to a real-life simulation scenario. Furthermore, the models used for the simulation are explained in detail to provide a thorough understanding of the inner workings of the simulator.

5.1 System model

5.1.1 Operators and users

To keep the simulation simple yet realistic on a system level, we include two operators having one BS each in the simulation to allow a performance assessment of both orthogonal and non-orthogonal spectrum sharing methods. Each operator will have its own set of users as is normal in real life environments, connecting exclusively to their own operator's BS.

To consider sufficiently different scenarios in terms of the realized number of active users per operator, up to ten users ($N = 10$) will maintain an active session with their operator's BS. Users can be denoted by $\mathrm{UE}_{i,j}$ where i denotes the operator ($i = A, B$) and j denotes the distinct user $j = 1, 2, ..., N$. Consequently, we can consider scenarios where the number of active users (N_A, N_B) is equal to e.g. $(1, 1)$, $(1, 9)$, $(0, 3)$, etc. Note that the performance gains with more active users at both operators might differ a lot from cases where less active users are available for the advanced transmission schemes as the former yields more combinations of users, potentially yielding higher bit rates. The users are assumed to be moving at an average pedestrian speed of three kilometres per hour in a straight line. A user's trajectory is modelled by an initial location, a moving speed, and the direction of movement. From these parameters, the user location can be calculated at any point in time.

Model parameter	**Value**
Number of operators	2
Number of BSs	1 per operator
Number of active users per operator	10
Number of user constellations	150
User mobility	Pedestrian speed, 3 km/h
Network topology	Overlapping cells with co-sited BSs
Inter-site distance	500 m
Number of transmit antennas per BS	4
Antenna spacing on BS	10x wavelength
Number of receive antennas per UE	2
Antenna spacing on UE	0.5x wavelength
Channel model	WINNER II
Path loss model	WINNER II
Trace length	1000 TTIs (1 second)
Traffic model	Data only, downlink
Spurt size	Exponentially distributed with $\mu = 561721$ bytes
Frequency band	2600 MHz
Total bandwidth	10 MHz (50 PRB) (5 MHz per operator)
Power per BS	40 W
Propagation environment	Urban
Surrounding cells	Included for interference (50% activity level)

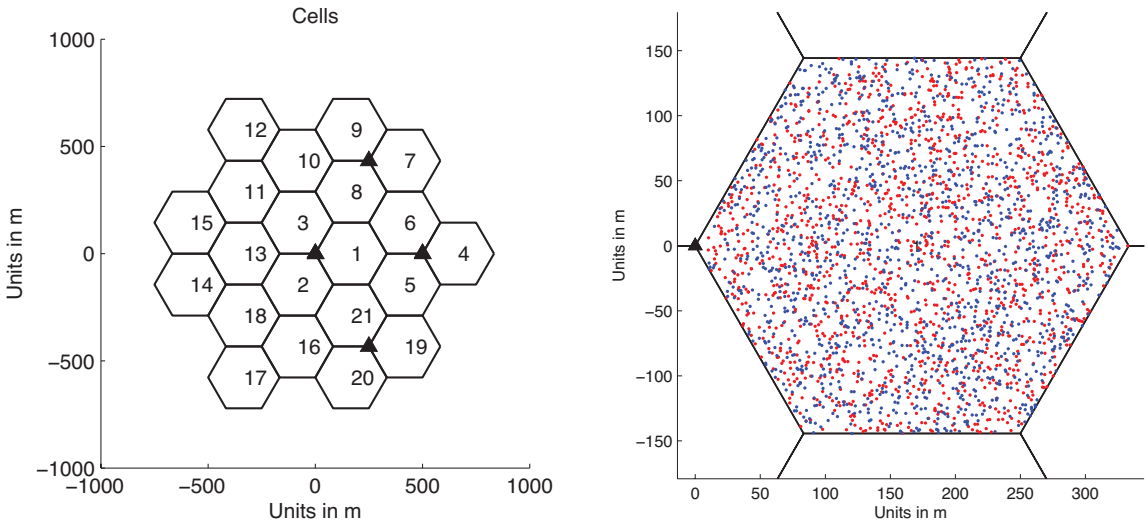
Table 5.1: Model parameters

Our simulations will include $2N = 20$ users, randomly sampled from a uniform spatial distribution (Figure 5.1b). However, as these 20 users only move with pedestrian speed, one such collection of users is not enough to calculate any statistical relevant measures as the effects might very well be caused by the user distribution. To counter this problem, we consider 150 different users constellations of 20 users for each simulation scenario. In effect, we simulate 3000 distinct users in combinations of 20. This will ensure statistical relevancy as we make sure that enough different spatial distributions of the users are simulated.

5.1.2 Network topology

For the topology of the network, we consider two overlapping cells of two operators, A and B, that are each part of a multi-cellular network. This network is depicted as a hexagonal layout (Figure 5.1a). In this study, we focus on two cells that overlap completely (cell one in Figure 5.1a). This cell of focus is served by both operators, so if a user is in range of operator A, it logically follows that the user is also in range of operator B. All the cells outside the cell of focus are only included to establish some realistic interference towards the cell of focus. The inter-site distance between BSs is set to 500 meters.

Two choices are available for BS placement: *co-siting* and *non co-siting* (Figure 5.2). The former reflects a current trend in cellular networks where operators share the sites where antennas are deployed to share the cost. However, sometimes co-siting is not possible due to physical limitations of the site or sub-



(a) (b)

Figure 5.1: Network topology consisting of 21 cells (a), where cell one is the cell of focus with co-sited antennas at the black triangle at (0,0). In (b), the cell of focus is shown in closeup with all 3000 users depicted by coloured dots. Different colours depict users of different operators.

optimal placement in the network, so the non co-siting case is also a relevant one. As co-siting antennas of different operators is common practice and due to dependability on third party data (see also Section 5.4), we only consider the co-siting scenario in this study. Because the BSs are at the same location due to the co-siting scenario and identical azimuths and tilts are assumed, the cells of focus are perfectly overlapping. Furthermore, it makes sense to coordinate between two BSs that serve the same geographical area rather than coordinating between BSs that are located in adjacent cells as the gains of using the same spectrum are potentially higher than coordination between disjoint cells.

To support MIMO, both BSs and UEs should have multiple antennas. A system with 4 tx antennas per BS and 2 rx antennas per UE will be considered. This antenna configuration is one of the standardized antenna configurations in LTE [37].

5.2 Traffic model

The traffic model for the simulator represents the flow of data from the BSs to the users as we simulate downlink traffic, and the dynamics such as load. As we simulate a limited number of users per constellation, we use persistent calls in the simulation with a dynamicity in the spurts users generate within such calls. This means that each call generates new spurts until the replication for that user constellation is ended. One such replication consists of 200 finished spurts. To incorporate also the unfinished spurts at the time the 200^{th} spurt is finished, we calculate the throughput for these spurts in the time that they

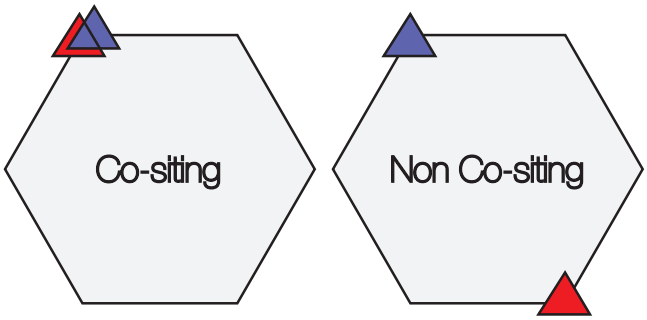


Figure 5.2: The difference between co-siting and non co-siting, where operators respectively share a site or are located at disjoint locations.

were active.

Since many users use their devices (phones, laptops) for web surfing, we simulate web traffic with the simulator. For definition of the distribution of the size of web pages, we use data from the HTTP archive [38], a project that provides a permanent repository of web performance information such as web page sizes. The information in this repository is collected by using a browser to access a subset of the one million most popular sites according to alexa.com. Not only does the HTTP archive measure the size of the resulting HTML page, but also all referenced files like images, JavaScript and css. The database we use in this report contains data about web pages loaded with a mobile phone, using the Mobitest [39] measurement tool. This dataset is more applicable to our scenario compared to the dataset gathered with a desktop setup, as we simulate mobile users moving at pedestrian speed. Furthermore, many websites have optimised versions of their pages for mobile, so the average download size of a web page is smaller on mobile phones as well. Analysis of the dataset [38] shows that the total request size including referenced assets is exponentially distributed with a mean of 561721 bytes. We can use this distribution to calculate the spurt size for a user when a spurt is generated. Note that another spurt can only be generated after the previous spurt has been fully downloaded to the UE. As we use size to define a spurt, the spurt duration is effectively dependent on its size and the experienced bit rate.

In-between two downloaded web pages, the user takes some time to read the information on the web page. This time is the time between two spurts of one user, and can be varied in the simulator to increase or decrease the load. We call this time between two spurts the *inter-spurt time* (Figure 5.3). The longer the inter-spurt time, the lower the load on the system, and the shorter the inter-spurt time, the higher the load. With an inter-spurt time of zero, each user will immediately generate a new spurt after finishing its previous one, resulting in the maximum load the system can be offered with the given model.

The traffic model as described above builds on the traffic models from Next Generation Mobile Networks (NGMN) and 3rd Generation Partnership Project (3GPP), as described in [40] and [41]. Both models define certain distributions for the size of web pages. However, as web page sizes increase fast, calling for recent measurements, and the NGMN and 3GPP models do not explicitly say

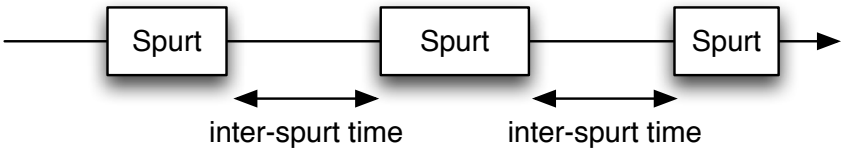


Figure 5.3: Representation of the traffic model for one user. After each spurt, a new spurt is generated after a certain inter-spurt time. Spurt duration depends on the size of the spurt and on the experienced bit rate.

whether the page size has been measured on mobile devices, we use the HTTP archive data. The inter-spurt time is used in both the NGMN and the 3GPP model, although under the name 'reading time'. However, we do not fix the inter-spurt time to the values described in the models as we want to use the inter-spurt time to vary the load.

When a spurt fails to get allocated resources, the spurt is terminated and a new spurt is generated for the user after an inter-spurt time interval, just as when the spurt would have finished in a normal manner. The termination criterion for the spurts is as follows: when a spurt is running for 2000 TTIs (2 seconds) or more, and the average throughput experienced during this spurt is below 1 kbps, terminate the spurt. This termination criterion is mainly put in place to prevent MM scheduling to allocate all resources to poorly performing UEs, so it should be regarded as a network-based termination rather than a user-based termination.

5.3 Bandwidth, power and interference

The most important resources managed by the radio resource management are the available bandwidth and the power. The bandwidth determines the number of PRBs available to the system. For 10 MHz bandwidth for example, 50 PRBs of 180 KHz utilize about 90% of the available bandwidth. The PRB is the smallest unit of allocation in LTE, consisting of 12 sub-carriers during one TTI (see Section 4.1). The reason that only 90% of the bandwidth is available for transmission stems from the guard-space between the PRBs. For transmission power, typically 20 Watt is used for 5 MHz of bandwidth and 40 Watt for bandwidths greater than 5 MHz. For this study, we simulate 10 MHz bandwidth in total, which means that 50 PRBs are available to the system. As the bandwidth exceeds 5 MHz per operator in all cases except FSA, 40 Watt is used at the BSs for transmission, evenly distributed over the PRBs.

As mentioned before, surrounding cells are included only to establish interference. These interfering cells are assumed to transmit at a fixed power level and the links of these BSs towards UEs will be characterized purely by path loss (see Section 5.4.1). The interfering cells are assumed to be transmitting at a fixed activity level, simulating a certain load at the interfering cells. This activity level is defined to be the percentage of the maximum power of the interfering cells. This means that the setting of this activity level defines a fixed

transmit power with which the interfering cells transmit. The default setting is set at 50%, and another option is an activity level of 30% to study the sensitivity of the performance of the system to interference of the surrounding cells. The maximum transmission power of the interfering cells is set to be the same as the BS transmit power in the cell of focus: 40 Watt.

5.4 Physical layer abstraction

As the focus of this study lies in the comparison of the physical layer transmission schemes, as developed by the SAPHYRE project, with other sharing methods, we need input from the project partners to incorporate this into the simulator. This way, we can abstract from the physical layer transmission schemes and focus on the scheduling and analysis of the spectrum sharing.

This section serves as a clarification how we get from a propagation model, through the physical layer abstraction, to an actual bit rate for the users. In this section, we will introduce the propagation model and the associated channel traces. Furthermore, we will introduce the physical layer traces from the transmission schemes developed by the SAPHYRE project to get a good understanding of the data we are dealing with. Subsequently, we will shortly introduce the transmission schemes themselves, as the theory is not an integral part of this study, but serves well for further clarification. Finally, we will outline how the values obtained from the physical layer abstraction are used to calculate an actual bit rate for the users involved.

5.4.1 Propagation model

To know the quality of the channel between transmitter and receiver, we need to know how the signals behave when they travel from the transmitter to the receiver: the channel response. As mentioned in Chapter 2, the channel response generally consists of three components: path loss, slow fading and fast fading. For the fast fading component we need channel traces for a number of TTI and for a certain number of transmit- and receive antennas. The fast fading component of the MIMO channel between a transmitter and receiver is represented by a matrix H that defines for n receive- and m transmit antennas all possible channel responses between the transmit and receive antennas denoted by $h_{i,j}$ at time t [42]. Equation 5.2 shows such a channel matrix for 2 receive- and 4 transmit antennas, conforming to our model parameters. Figure 5.4 shows a graphical representation of the paths between the transmit- and receive antennas, as specified by our model parameters.

$$H(t) = \begin{pmatrix} h_{1,1}(t) & h_{1,2}(t) & h_{1,4}(t) & h_{1,4}(t) \\ h_{2,1}(t) & h_{2,2}(t) & h_{2,3}(t) & h_{2,4}(t) \end{pmatrix} \tag{5.1}$$

The channel response matrices can be generated by the use of standardized models such as SCM, SCME or WINNER (I and II) [42, 43, 37, 44]. The main difference between the three models is that both WINNER models and the SCME model support larger bandwidth and both WINNER models are applicable for more scenarios. All models can theoretically be applied to our system-level simulations because they all provide random delay and angle spreads for

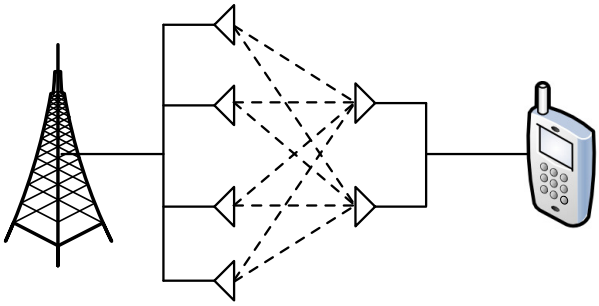


Figure 5.4: Paths between BS and UE antennas in a system with four transmit antennas and two receive antennas

different users [37], making the choice of the models largely dependent on the availability of the knowledge how to generate the channel traces.

The channel matrix is generally determined by path loss, shadowing and multipath fading. The path loss component is calculated by the use of a path loss model, such as Okumura-Hata, COST-231 or WINNER II [37, 44, 45]. The path loss models are sets of algorithms, mathematical expressions and diagrams representing the radio characteristics of a certain environment. These models are either empirical or deterministic and are targeted to a certain environment like urban or rural. Empirical models are based on real-life measurements, taking the environmental influences implicitly into account. The correctness of an empirical model depends on the quality of the measurements and the applicability of the original measurement environment. Most empirical models provide different parameters for usage in different environments e.g. WINNER II provides parameters for rural-, urban- and suburban environments, among others. Deterministic models, however, are based on the physical properties of signal propagation in different environments and can as such be applied to different environments without affecting accuracy. The downside is that the algorithms are computationally complex and are thus most used in small scale simulations or for indoor propagation simulation [46]. Shadowing is applied separately, and is often modelled using a log-normal distribution. The WINNER II includes different formulas for shadowing in various environments. Multipath fading is applied next, and is included in the SCM, SCME and WINNER models.

Spectrum-wise there are many bands that can be chosen to deploy LTE in. In Japan, 800 MHz, 1500 MHz and 1700 MHz bands are in planned or deployed phase; the United States and Canada mainly use 700 MHz and 2100 MHz while in Europe the deployed and planned networks are mainly situated at 800 MHz and 2600 MHz. Furthermore, the idea of reusing the spectrum originally assigned to GSM are also frequently raised. As we aim to simulate a European urban environment, we choose to simulate the 2600 MHz frequency. This frequency is more applicable in an urban environment than for instance 800 MHz since in urban areas operators prefer to use the higher capacity of 2600 MHz over the larger coverage of 800 MHz as there is larger population density than in rural areas.

The choice of a propagation environment and frequency band limits the choice of path loss models. The Okumura-Hata and COST-231 path loss models are not applicable in this scenario because their frequency range is limited to respectively 50 MHz - 1500 MHz and 1500 MHz - 2000 MHz [47, 48, 49]. For the 2600 MHz band, the Stanford University Interim (2500 MHz - 2700 MHz) model or the WINNER II (2000 MHz - 6000 MHz) path loss model is applicable [44]. Both are applicable to urban terrain, but as the WINNER II is more advanced and Fraunhofer-Gesellschaft (FhG) is willing to calculate the channel matrices for this model, we use this model in our simulations. These channel matrices are generated for the parameters of our system model, and this full set of channel matrices will be referred to as channel traces, as they describe the channel for our users in the modelled time.

As calculation of the channel traces is computationally complex and resource intensive, we cannot model too many TTIs. The length of the channel trace has to be chosen carefully as it has to be long enough to be able to see effects of slow- and fast fading, yet it must not be too long as the data size increases linearly with the number of TTIs. Furthermore, this would lead to delay in the data delivery, surely delaying this study. Therefore, we set the number of TTIs included in the channel trace to 1000. This is enough to see effects from slow- and fast fading. As the a replication of the simulation typically takes more than one second, we can repeatedly loop over the channel traces from front to back and back again to simulate longer timespans.

5.4.2 Physical layer traces

We need to obtain the CSI from BS to UE when we use one of the transmission schemes schemes to transmit data. This data could for instance be obtained in the form of a formula which we can apply to the WINNER II channel trace, or in the form of pre-computed traces adapted from the channel trace. As the calculation in our simulator with a formula turns out to be very computationally complex, Linköping University pre-calculated the traces for the physical layer for us. These traces serve as input for the simulator, and serves as the input information for the scheduling algorithms.

In the traces of the physical layer, all model aspects have been taken into account to create the experienced SINR values for all the UEs. This includes the interference from surrounding cells. Like the channel traces provided by FhG, the traces for the physical layer abstraction include SINR values for each combination of TTI and PRB. However, the difference between the channel traces and the traces for the physical layer abstraction is that the latter take the non-orthogonal usage of the spectrum into account. Instead of one value for each user, the trace now consists of SINR values for each user for all possible scheduling combinations. This includes the possibility that no other user is scheduled at the other BS, and the combinations with the users of the other BS. In Table 5.2, we can see the SINR values for one TTI, PRB and user constellation combination. For users UE_A, i and UE_B, j where $i \in \{1, 2, \ldots, N\}$ and $j \in \{1, 2, \ldots, N\}$, we can see the SINRs for both users when the corresponding users as signified in the first column and row are scheduled together in the current TTI at the PRB this table signifies. Furthermore, the column and row denoted with 'IDLE' signify the cases where a PRB is just used by one user from one BS at a time. Note that scheduling for the orthogonal scenarios solely uses this

BS_B \ BS_A	**IDLE**	$\mathbf{UE}_{A,1}$	$\mathbf{UE}_{A,2}$	$\cdots$	$\mathbf{UE}_{A,n}$
IDLE		$SINR_{A,1}$	$SINR_{A,2}$	$\cdots$	$SINR_{A,n}$
$\mathbf{UE}_{B,1}$		$SINR_{A,1}$	$SINR_{A,2}$	$\cdots$	$SINR_{A,n}$
	$SINR_{B,1}$	$SINR_{B,1}$	$SINR_{B,1}$		$SINR_{B,1}$
$\mathbf{UE}_{B,2}$		$SINR_{A,1}$	$SINR_{A,2}$	$\cdots$	$SINR_{A,n}$
	$SINR_{B,2}$	$SINR_{B,2}$	$SINR_{B,2}$		$SINR_{B,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\mathbf{UE}_{B,n}$		$SINR_{A,1}$	$SINR_{A,2}$	$\cdots$	$SINR_{A,n}$
	$SINR_{B,n}$	$SINR_{B,n}$	$SINR_{B,n}$		$SINR_{B,n}$

Table 5.2: Table for serving one out of n users per BS at a certain TTI at a certain PRB from a certain user constellation

column and row.

5.4.3 Transmission schemes

Linköping's Institutionen för Systemteknik (ISY) provides the abstraction layer for the physical layer transmission schemes so we can evaluate their gain. To be sure to keep the input format and the way of calculation of the different schemes the same, they also provide reference transmission schemes that serve as a reference for evaluation. All schemes are based on game theory, trying to solve resource conflicts in wireless networks. Game theory falls apart in two categories: non-cooperative- and cooperative game theory. In non-cooperative games, players directly compete with each other and cannot strike deals. In cooperative games, however, players can form joint strategies and strike deals with each other. [50] The supplied transmission schemes used in the simulation contain both non-cooperative and cooperative strategies: Nash Equilibrium (NE), Zero-forcing (ZF), Maximum Sum Rate (MSR), Nash Bargaining (NB), and Max-Min (MM).

All the transmission schemes will be described in the next sections. Furthermore, the applicability constraints of the schemes are introduced for the forms of spectrum use and coordination between operators.

Nash Equilibrium (NE)

A non-cooperative scheme is supplied with the Nash Equilibrium (NE) transmission scheme. The NE solution is a beamforming technique that maximizes transmit diversity. It is also known as maximum-ratio (because it maximizes the SNR) or as matched-filter (because the SNR maximization is achieved by matching the direction of the beamforming vector with the one of the channel vector). Since the solution maximizes transmit diversity, the solution gives an upper bound on the gain when compared with LTE's transmit diversity scheme, and therefore serves as a reference scheme. The NE solution is Pareto optimal for orthogonal channels. For regular channel realizations however, the NE solution is close to the Pareto boundary for low SNR, but for high SNR the cost of the anarchy of a non-cooperative game is unbounded.

NE is the reference scheme for LTE and therefore will be used in the FSA scenario. Also, as NE excels with orthogonal spectrum use, we will use this scheme as well in more dynamic orthogonal sharing. Furthermore, as we are interested in what will happen when we use the spectrum in a non-orthogonal fashion without coordination between operators, we will also use NE for that scenario.

Zero-forcing (ZF)

The Zero-forcing (ZF) reference scheme can be considered a cooperative one. The ZF beamformers assure that the transmitter generates no interference on the other system. To accomplish this, the ZF method requires knowledge of the CSI of the users of the other BSs and thus needs information exchange between BSs about CSI of their respective users. The beamforming vector is designed such that the interference in the link is cancelled, so it maximizes the numerator of its own SINR, while minimizing the denominator of the other SINR. The ZF scheme is also considered a reference scheme as this scheme is well-known and well-described in literature and not developed in the SAPHYRE project.

As ZF requires knowledge of the CSI of the other operator's users, the ZF is used in a coordinated scenario. Furthermore, because of the nature of this scheme, the CSI only differs for non-orthogonal cases from the NE scheme. Therefore, this reference scheme is used with non-orthogonal spectrum use.

SAPHYRE schemes

The cooperative schemes are the remaining MSR-, NB- and MM- schemes, derived from methods developed in the SAPHYRE project. All these tables deal with cooperation between BSs to increase the utility of the system. As the schemes are more advanced than the ZF scheme, they are expected to outperform ZF in the evaluation. It is crucial to understand that a player in a cooperative game can be cooperative and rational at the same time. That is, being cooperative does not mean the same thing as being altruistic. Players may be willing to accept a bargaining solution that is found to be good enough for both when they are interested in maximizing their own outcome. The three transmission schemes provided each have their own objective: MSR aims to maximise the sum-rate in the game, NB aims to divide resources in a fair way, and MM aims to maximise the minimum SINR to promote fair throughput.

For usage in different scenarios, the same restrictions apply to these schemes as with ZF. Furthermore, as each transmission scheme has its own objective, we will use the transmission schemes with a scheduling algorithm that matches in the objective. Herewith, we provide a consistent translation of the operator's perspective into both the transmission scheme at the physical layer and the scheduling at the Medium Access Control (MAC) layer. We match the MM scheme with MM scheduling because of the similar objective to maximise the minimum rate/SINR. The NB scheme is matched with PF scheduling as they both try to divide resources in a fair way, and the MSR scheme is matched with MSR scheduling because of their objectives to maximise the rate/SINR.

5.4.4 From abstraction to bit rate

With the traces from the schemes used in the physical layer abstraction, we have all the input we need to make the scheduling decisions. At the beginning of each new TTI, the simulator calculates the bit rate attained in the last TTI for each active user, and deducts the transferred bits from the size of their active spurts. To do so, we need to convert the SINR values experienced by the user in the PRBs the user has been assigned into the attained bit rate.

5.4.5 Exponential Effective Signal to Noise Ratio Mapping (EESM)

The Exponential Effective Signal to Noise Ratio Mapping (EESM) method is used to obtain an effective SINR over a whole transmission, which can be used to map to a single Modulation and Coding Scheme (MCS), and then to the BLER. EESM is a simple mapping method used when all the PRBs of a user are modulated using the same MCS. The effective SINR (γ_{eff}) is obtained by combining the SINRs exponentially with the EESM method by performing the following formula:

$$\gamma_{eff} = EESM(\gamma_i, \beta) = -\beta * ln(\frac{1}{N} * \sum_{i=1}^{N} e^{\frac{\gamma_i}{\beta}}) \tag{5.2}$$

In this formula, N denotes the number of PRBs to be averaged, and β is a calibrated value unique to each MCS, obtained from link-level simulations executed by [51, 52, 53, 54].

Adaptive modulation and coding

Now that we have a single SINR, we can employ adaptive modulation and coding (often called link adaptation) to convert the experienced SINR into a single Modulation and Coding Scheme (MCS) for the transmission. The MCS is a combination of a modulation scheme and a code rate. Lower-order modulation is more robust and can tolerate higher levels of interference but provides lower transmission bit rates. Higher-order modulation offers a higher bit rate, but is more prone to errors due to higher sensitivity to interference and noise. Therefore, it is only useful when the SINR is sufficiently high. For a given modulation, we can choose the code rate based on the channel conditions: a lower code rate in poor channel conditions and a higher code rate in case of high SINR. By adapting the modulation and coding techniques used in the communication at a certain PRB, a BS can choose the best MCS for the quality of the channel, rendering better channel utilisation and spectral efficiency [37, 55].

The principle is simple: based on the SINRs experienced in the last TTI, calculate and select a MCS that is optimal for the SINR according to Figure 5.5. The MCSs (numbered 0 to 15) signify different modulation schemes (QPSK, 16QAM, 64QAM) with different code rates [37].

Conversion to a bit rate

With the MCS we can look up the MCS- and PRB-specific attainable bit rate in Figure 5.6. To obtain the actually experienced bit rate we need to know the

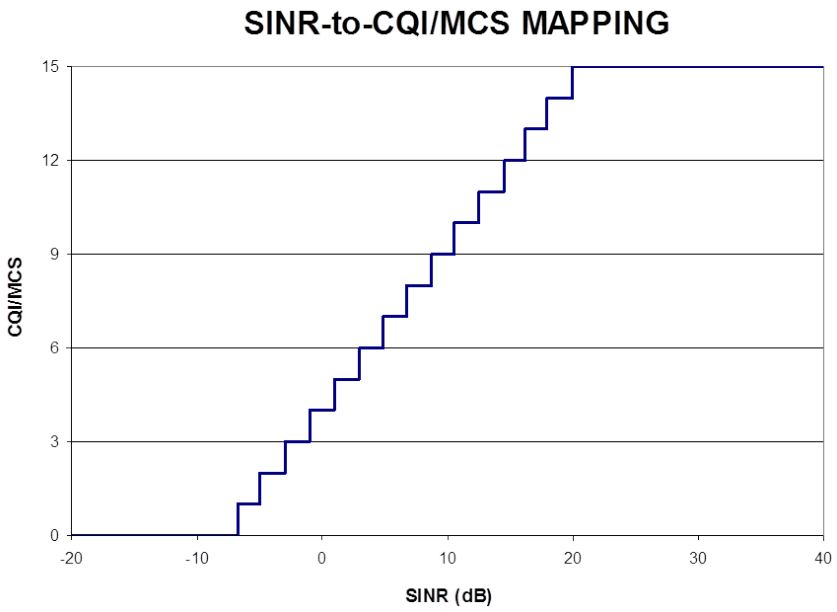


Figure 5.5: Mapping of SINR to MCSs

Block Error Rate (BLER) for the user. The BLER is used to determine success or failure of a certain transmission and is MCS dependent. To assess this BLER, we employ a set of Additive White Gaussian Noise (AWGN) curves as shown in Figure 5.7, derived from link level simulations in which the performance of different MCSs has been validated [56, 57]. With this BLER, we can flip a biased coin to decide whether the transmission has succeeded or failed. When the transmission succeeds, we can look up the bit rate in Figure 5.6 and convert it to the actual transferred bits. The remaining spurt size is subsequently updated according to the calculated number of transferred bits.

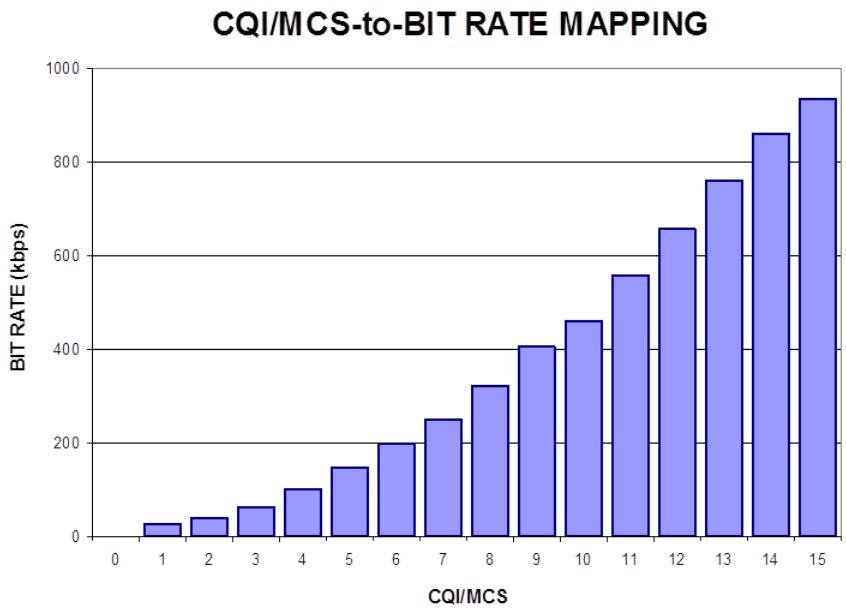


Figure 5.6: Mapping of MCS to attainable bit rate. Note that these rates have to be corrected by $\varphi_D = 0.7788$ since not all resources are available for PDSCH bits.

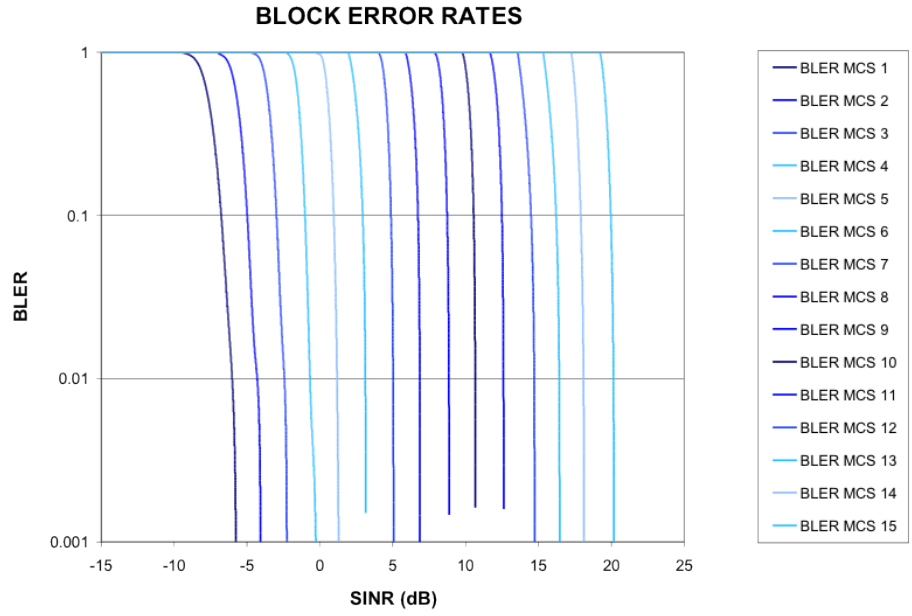


Figure 5.7: BLER, approximated with an 8^{th} degree polynomial

Chapter 6

Simulation results & analysis

In Chapter 1, we posed the following question: *what can we gain in terms of performance and capacity at the system level, by applying the advanced transmission schemes for non-orthogonal sharing, as developed in the SAPHYRE project, with respect to Fixed Spectrum Allocation, orthogonal sharing, and non-orthogonal sharing with the ZF transmission scheme?* This chapter discusses the results from the system-level simulations and tries to find answers to the research question.

As was stated in the research question we want to know what could be the performance- and capacity gain of SAPHYRE when compared to other forms of spectrum use and transmission schemes. As a base question this is a valid question as spectrum is a scarce resource and we can make better use of this available spectrum when we take global scheduling into account. However, it is also important to know how these techniques behave when we introduce feedback delay and error, as was stated in one of the sub questions. Will the system hold its desirable properties (if any)?

To answer these questions, first the simulation scenarios and simulation parameters will be introduced, followed by an overview of the metrics used in analysis. When the constraints of the simulation and the metrics are set, we will discuss the differences between the different sharing methods and their associated scheduling algorithms using the simulation results. With a smaller set of desirable scheduling forms, we will perform a sensitivity analysis to see if the performance degrades when delay and error is introduced.

6.1 Simulation scenarios

In the preceding chapters we wrote about scheduling and spectrum sharing techniques. In order to analyse meaningful results, we need to introduce the simulation scenarios which we will use for evaluation of the physical layer transmission schemes. Apart from the modelled parameters, we can distinguish between two very important factors that determine the simulation scenarios: use of the spectrum and coordination between BSs. In this section we will provide an overview of the simulation scenarios that will be used for evaluation of the combinations

of scheduling algorithms and transmission schemes. Figure 6.1 shows the different simulation scenarios, clearly categorised in combinations of spectrum use and coordination.

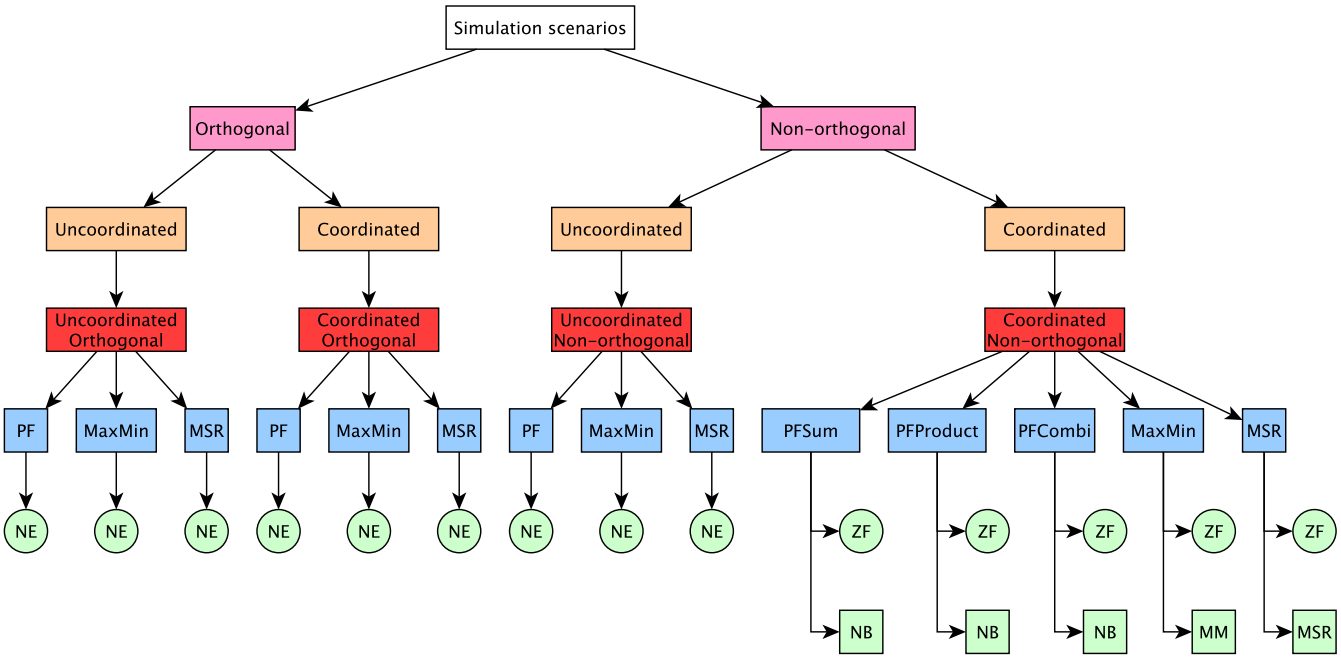


Figure 6.1: Tree of simulation scenarios, scheduling algorithms, and their associated transmission schemes. From the top level of the tree, we first see the division between orthogonal and non-orthogonal sharing (pink), followed by a division of coordination between BSs (orange), and the short name of this scenario (red). Working our way down the tree, we see the scheduling algorithms (blue) and their associated transmission schemes (green). Reference transmission schemes are depicted by green circles, SAPHYRE transmission schemes by green squares.

The way in which the spectrum is used by the operators is the first main distinction we can identify in classifying the scenarios. As discussed in Chapter 3, we can distinguish between orthogonal and non-orthogonal use of the spectrum.

The second classification aspect of the simulation scenarios is the coordination between BSs. Within uncoordinated scenarios, no communication is needed between operators to divide the spectrum or to make scheduling decisions. Each BS manages the spectrum it is allowed to use. Since there is no coordination between the BSs, each BS is only aware of its own users. Consequently only scheduling combinations of the operator's own users are considered, assuming the other operator does not use the spectrum. Within the coordinated scenarios, the CSI of the users of both operators is shared. This enables scheduling on a level that ascends over the individual scheduling decisions of the operators: the system level. Instead of making the scheduling decisions individually per operator, the best scheduling combination of the users of both operators is

selected by a global scheduler.

Within orthogonal use of the spectrum, the distinction between uncoordinated and coordinated scheduling is the distinction where the spectrum is allocated for longer periods to operators (*Uncoordinated orthogonal*) and where the spectrum is allocated to operators more dynamic (*Coordinated orthogonal*). These two scenarios are extremes within orthogonal spectrum use in how dynamic the spectrum is allocated. In the uncoordinated orthogonal scenario, which is reminiscent of FSA, each operator is assigned a consecutive 50% of the spectrum, or 25 PRBs. Scheduling decisions are taken by the operator taking only its own spectrum into account as there is no chance of interference from other operators. In the coordinated orthogonal scenario, the scheduling decisions are made on a global level and the spectrum is assigned to the operators each TTI according to the results of the scheduling algorithm for all users of both operators. This means that, in case that only one operator has active users, all spectrum will be used by this operator. Besides these two extremes in orthogonal spectrum sharing, many mix-forms can be applied like partly fixed spectrum with a dynamic pool for usage by both operators. These mix-forms will not be evaluated.

Within non-orthogonal use of the spectrum, we again differentiate between two extreme forms with regard to coordination. In the *uncoordinated non-orthogonal* scenario, both operators are allowed to use the whole spectrum, but since there is a lack of coordination, it is possible that they use the same PRBs at the same time. This increases interference and will lead to suboptimal scheduling decisions as we cannot use the SAPHYRE transmission methods due to the lack of coordination, but are limited to the reference transmission method. In the *coordinated non-orthogonal* scenario, the scheduling is coordinated on a system level and either the ZF transmission scheme is used or the SAPHYRE transmission schemes are used. Therefore, the scheduler will take the users of both operators into account and yield more optimal scheduling decisions. Furthermore, due to the use of advanced transmission schemes, a large part of the interference will be cancelled leading possibly to better rates than the uncoordinated non-orthogonal scenario.

For all four scenarios, we can employ the scheduling algorithms MSR, PF, and MM. However, because the PF algorithm has to deal with two priority indices per scheduling decision in the coordinated non-orthogonal scenario, for this scenario we simulate the three different options PFSum, PFProduct, and PFCombi, as described in Chapter 4.

As described in Chapter 5, we can use different transmission schemes to transmit from the BSs to the UEs. Only for the coordinated non-orthogonal scenario, we can use the advanced transmission schemes, where ZF is the reference transmission scheme and NB, MM, and MSR are the SAPHYRE transmission schemes. The SAPHYRE transmission schemes are used with the scheduling algorithms that match in goal. For the other scenarios, we use the reference transmission scheme NE, as the scenarios are either orthogonal or uncoordinated.

Parameter	Possible values	Default value
Inter-spurt time (seconds)	0, 5, 10, 15	5
Activity level of interfering cells (%)	30, 50	50
Feedback delay (TTI)	0, 4, 8	0
SINR error standard deviation (dB)	0, 1, 2, 3, 4, 5	0

Table 6.1: Simulation parameters and default values

6.2 Simulation parameters

Apart from the model parameters defined in Chapter 5, a few more parameters can be set to run simulations (Table 6.1). These simulation parameters control parameters that introduce feedback delay, introduce error in SINR values and control the inter-spurt time.

The first parameter we use in our simulations is the inter-spurt time. As described in Chapter 5, this parameter defines the time (in seconds) between the end of a spurt and the start of a new spurt for a certain user. This parameter effectively controls the offered load to the system, which depends on this inter-spurt time and the attained bit rates. Each simulated scenario is run with different inter-spurt time settings to simulate different offered loads. The parameter is set at the values 0, 5, 10 and 15 to simulate full load, and lower loads, respectively. For the sensitivity analysis, we fix the inter-spurt time at 5.

To vary the interference caused by the neighbouring cells, we have a parameter that defines the activity level of these cells. Unless otherwise stated, this parameter is set at 50%. There is one other possible setting: 30%. Other values are not possible at this moment as the neighbouring cell interference is calculated by third parties and is included in the final SINRs in the SINR-tables. This parameter could have significant impact on the performance, as this is a parameter that cannot be controlled in a real life environment. Furthermore, the noise from surrounding cells is a given parameter that is outside the scope of influence for the BSs. This effectively implies that the attainable rates of users could be different then expected in the scheduling step, leading to diminished performance.

For introducing delay in the SINR values, a parameter exists that influences the delay in SINR values, expressed in TTIs. This delay can be interpreted as a feedback delay from the user to the BS, so the CSI arrives later. The default value is 0, meaning that there is no delay in the SINR values used for decision making during scheduling, adaptive modulation and coding, and the corresponding SINR values experienced by the users. A higher value means that a certain amount of TTIs delay is introduced, causing the scheduling and the adaptive modulation and coding to use the outdated SINR value. This means that a discrepancy exists between the delayed SINR and the actual experienced SINR. With the selected MCS based on the delayed SINR, a higher than expected BLER could be experienced as the MCS is not tailored to the actual experience SINR. As described in [58] and [59], for pedestrian traffic, the feedback delay should not significantly change the performance of the system as long as the delay is less than 10 ms. Thus, as a delay of 4 TTIs is considered normal, and we estimate the delay at 8 TTIs for coordinated scenarios, we evaluate the impact of this parameter to confirm the stability of the system.

Another parameter is defined for the error in the SINR values: the standard deviation (in dB). This parameters defines the Gaussian error applied to the SINR values used for scheduling, and can be regarded as a channel estimation error. The default value is 0, meaning that there is no error in the SINR. As with the delay parameter, the error only affects scheduling and adaptive modulation and coding. The mean value of the SINR error is always 0, so it is possible for the system to either over- or underestimate the SINR during scheduling. As we can observe from [60], the performance of mobile networks will decrease with increasing error in the channel estimation. The reason for diminished performance is quite straightforward: by introducing error in the values on which the scheduling is based, the scheduling algorithm will not be able to make the most optimal scheduling combination, reducing system performance.

6.3 Overview of metrics

To evaluate the performance of the system in the various scenarios and to be able to compare these scenarios, we define the following metrics which will be introduced in this section:

- Offered load;
- Average PRB utilisation;
- Average UE throughput;
- Average UE throughput by distance from the BS;
- 10^{th} percentile UE throughput;
- Fairness.

An important system-level metric is the offered load. The offered load is defined as the total number of bits offered to the system divided by the simulated time. This metric depends on inter-spurt time, as a higher inter-spurt time means that less load is offered to the system and thus less throughput will be reached. Note that the offered load is similar to the system throughput as users cannot start a new spurt before the last one ended. Therefore, the system can never be overloaded, meaning that the maximum system throughput is reached with an inter-spurt time of 0. This metric will be useful in comparison of the capacity of the different sharing techniques and scheduling algorithms. With full load (inter-spurt time of 0), we can see the theoretical upper bound for the system throughput with all scheduling choices as all users keep receiving data. With lower offered loads, we can see how different scheduling algorithms perform when we cannot necessarily make the best combinations (in terms of the scheduling goal) of scheduled users as less users are available simultaneously. The offered load is an average of the offered load in all the 150 replications.

A metric that is closely related to the load is the average PRB utilisation. This metric signifies the average PRB usage from the perspective of the BSs. With full load, it gives an indication whether the scheduling algorithm makes efficient use of the available spectrum as a value less than 100% would mean not all PRBs are used in each TTI. With decreasing load, it gives an idea of the share of time the channel is unused as the scheduling algorithms will always

try to fully utilize the available spectrum. For orthogonal sharing scenarios, it is expected that the maximum PRB utilisation is 50%, as only half of the spectrum will be used by each operator.

The average UE throughput is a user-centred metric signifying the average throughput over all users. In order to compute this average throughput, the total transferred bits of all spurts of a user are divided by the sum of the transmission time of the spurts. Note that this translates into a weighted average over all spurts of a user since we take the transmission times into account. To generate the average UE throughput metric, these user throughput for all 3000 users (150 replications times 20 users) are averaged.

Not only are we interested in the average UE throughput over the whole system, but also analysed over distance between the user and the BS. As, on average, the channel quality should be better near the BS than far away, we expect this measure to decrease over distance. In order to provide this metric, we categorize the users into ten distance zones from the BS. As the maximum distance is 300 meters, each zone is 30 meters wide. The reason to categorize the distance from the BS is to make sure that enough users are evaluated for a certain range of distance to get a representative average throughput.

Closely related to the average UE throughput, we calculate a Cumulative Distribution Function (CDF) of user throughput. In the analysis we use the 10^{th} percentile, to give an idea of the minimum rates achieved. This metric means that 90% of the users has experienced a minimum average throughput of the depicted rate. As such, we can easily see if the user experience would be acceptable for the majority of users. The higher the 10^{th} percentile rate, the better the experience for the users. The 10^{th} percentile user throughput is generated by taking the average throughputs of all users and generating the CDF from that data.

A metric that is important for the user, but even more important for the operator is the fairness of the user throughput. As a user you would want to have a certain quality of service even when you are located on the cell edge, and as an operator you would want to promote fairness between users because these users have these quality of service constraints. To quantify fairness, we have to realise that the even allocation of the user throughput over all users is deemed fair. The more uneven this allocation is, the more unfair the scheme. To calculate the fairness, we use Jain's fairness index [61]. Jain's fairness index is a metric calculating fairness over a set of metrics $(x_1, x_2, \ldots, x_n)$, see Equation 6.1. In our case, we use the user throughputs as input for the calculation. The fairness index is bounded between 0 and 1, and can as such be interpreted as a percentage. A value of 0.1 means that the scheme is unfair to 90% of the users, whereas a value of 0.9 means that the scheme is unfair to only 10% of the users. In general, a higher fairness index means that the user throughputs are more evenly distributed over all users.

$$J(x_1, x_2, \ldots, x_n) = \frac{(\sum_{i=1}^{n} x_i)^2}{n * \sum_{i=1}^{n} x_i^2} \tag{6.1}$$

6.4 Spectrum sharing analysis

According to Figure 6.1, we have in total 19 leaves at the tree, which means that 19 distinct sets of simulations have been executed to form the basis of the analysis in this section. For each of the 19 combinations, we simulated according to the default simulation parameters (Table 6.1), including all options for the inter-spurt time. In this section, we take a look at the performance of these 19 combinations of spectrum use, scheduling algorithms and transmission schemes. To prevent confusing graphs due to the number of schemes, the simulations have been grouped by the different scenarios. The results of the simulations are depicted in Figure 6.2 until Figure 6.5.

6.4.1 Uncoordinated orthogonal sharing (FSA)

One of the reference scenarios, FSA, is the scenario depicting how the spectrum is used nowadays (Figure 6.2a - 6.2f). In our simulations each operator owns a consecutive part of 50% of the spectrum. The operators do not have to deal with interference from other operators on their share of the spectrum as the spectrum is used orthogonally. The physical layer scheme used is NE, as this is an uncoordinated scheme that gives an upper bound to using transmit diversity.

Figure 6.2f shows the offered load on the vertical axis, with varying inter-spurt times on the horizontal axis. As we can see, the maximum load the system can cope with is just over 30 Mbps in the case of MSR scheduling. The PF algorithm performs worse than MSR with about 24 Mbps under full load, and the MM algorithm can only handle just shy of 20 Mbps of traffic. As we increase the inter-spurt time, the offered load decreases. With increasing inter-spurt times the algorithms approach one another in offered load, indicating that the algorithms can all transmit the surplus in load that differentiated them from each other in the 0 inter-spurt time case within the inter-spurt times of the other users. Furthermore, there might even be time that the BSs are idle, i.e. do not have active user spurts. However, the MM scheduling algorithm remains the worst in terms of offered load even with longer inter-spurt times, suggesting that the MM algorithm is inefficient in scheduling even under lower system load. As the objective of the MM scheduling algorithm is to maximise the minimum rate, the scheduling algorithm will often prefer users with lower bit-rates over users with higher bit-rates to increase the rates of the former. The effect is that the spectrum is used less efficient than with MSR and PF, which both also take the system throughput into account.

As expected, we can confirm that BSs have more idle time in Figure 6.2d. In this figure, the average PRB utilisation is shown for the offered loads as in Figure 6.2f. For full system load, all algorithms reach 50% PRB usage, which is expected as each operator can only use its own share of the spectrum in the uncoordinated orthogonal scenario. For lower loads, the PRB utilisation decreases as the system fails to use all PRB as it is waiting certain amounts of time for new spurts to arrive due to the inter-spurt times.

As for the 10^{th} percentile throughput and the average user throughput (Figure 6.2a and 6.2b), we can observe that the MSR algorithm assures highest average throughput and MM the lowest, but the 10^{th} percentile throughput is dominated by the PF algorithm. It deserves notice that the 10^{th} percentile throughput for full load in the MSR algorithm is 0.0, meaning that at least 10%

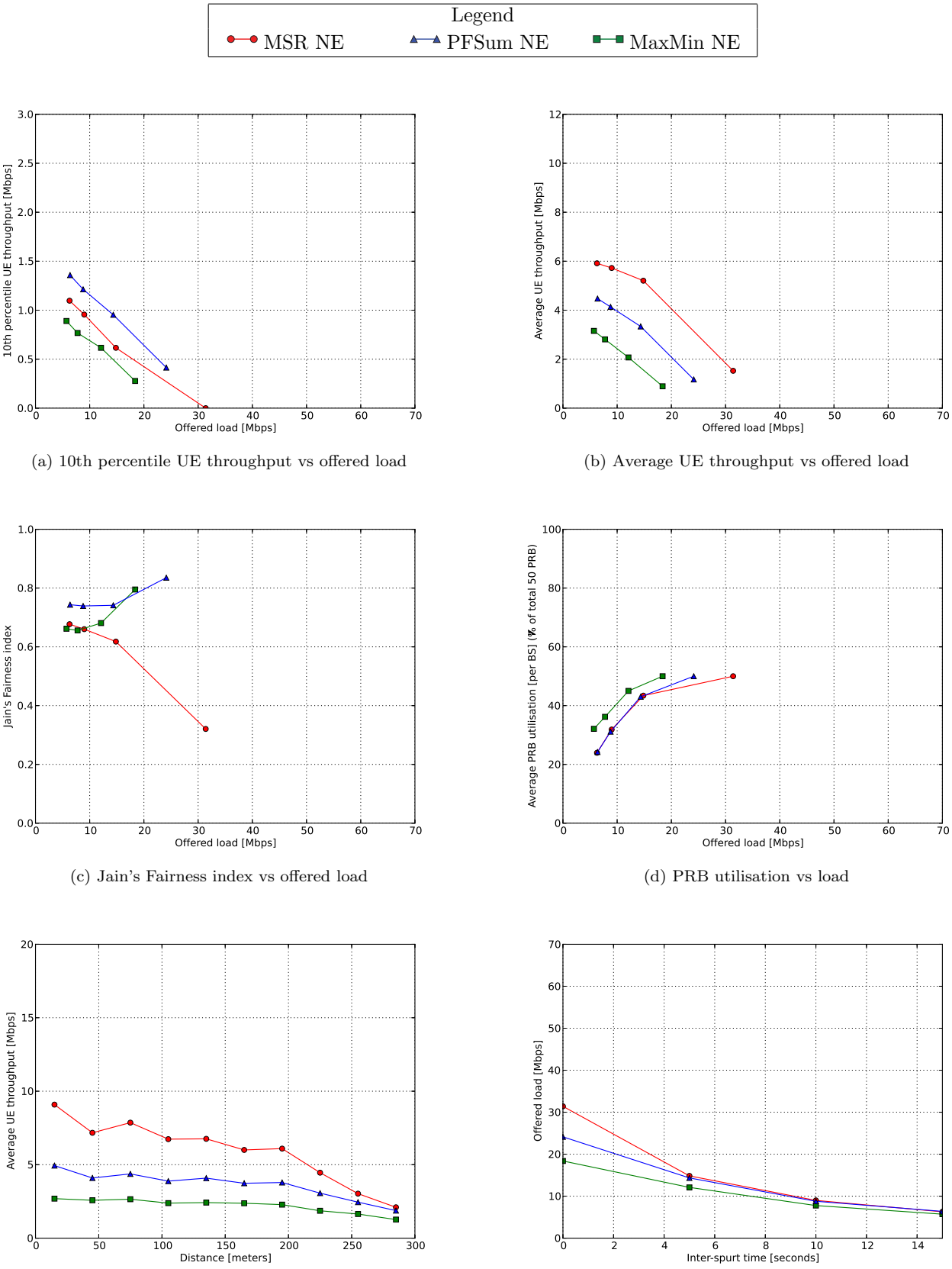
of the users experience no throughput at all. This is due to the fact that MSR always schedules the best users, and since each user generates new spurts instantly, users with lower attainable bit rates do not get a chance to be scheduled. This improves with longer inter-spurt times as users with less favourable channel conditions are scheduled in-between spurts of users with more favourable conditions. As said, MM performs the worst in the average user throughput. For the 10^{th} percentile throughput, it is a bit more complicated. When we translate the data points in the graph to inter-spurt times, we can see by looking at Figure 6.2f that the data points are in decreasing order for inter-spurt time. This means that the rightmost data points signify the maximum load that the algorithms can cope with. Therefore, we can say that the MM algorithm actually dominates the MSR algorithm for full load as the MM algorithm manages 0.25 Mbps 10^{th} percentile UE throughput, where the MSR algorithm manages to offer nothing to these users. However, as the load decreases, we can see that for all other inter-spurt times the MM algorithm is on par, or below the other algorithms making it still the worst performing algorithm in this scenario.

Fairness-wise, we can clearly see in Figure 6.2c that the PF algorithm dominates in terms of fairness between users. The higher the load, the better the PF algorithm can provide fairness between users, up until over 80%. The MM algorithm follows suit with a marginally smaller at full load, and a somewhat bigger gap at the lowest load. Both the PF and the MM algorithm have curves that increase with increasing offered load. This is due to the fact that both algorithms will equalize the rates more when more users are available. With lower loads, it is possible that only one user is active at a time, yielding him the maximum rate as he will be assigned all PRBs. For the MSR algorithm, the overall curve is contrary to the other algorithms. When the system experiences full load, the fairness is low due to the tendency to only serve the very best users for the same reasons outlined before. With lower offered loads, users that would not be served with full load can be served in-between the spurts of higher performing users, thus increasing fairness as they will actually be allowed to receive data instead of being ignored by the scheduling algorithm due to bad channels. Note that the MSR algorithm will never be more fair than the PF and MM algorithm, as in the case when only one user has an active transmission at a time, the fairness will be equal for all algorithms. In cases with more active users however, the MSR algorithm will choose the best performing user, decreasing fairness.

Figure 6.2e shows the average user throughput over distance from the BS. The graph shows the throughput with an inter-spurt-time of five seconds. As expected, this follows a clear downward trend with increasing distance from the BS for the MSR algorithm. This clearly indicates that users closer to the BS on average have better channel conditions. For PF and MM, the downward trend is less pronounced as the fairness of these schemes is higher, meaning that UEs far away (with also get a chance to receive data) will get more attention by the algorithm.

6.4.2 Uncoordinated non-orthogonal sharing

Figure 6.3a - 6.3f show similar graphs as we just analysed, for non-orthogonal uncoordinated sharing of the spectrum. As discussed earlier, in this scheme BSs can use the whole spectrum and select users regardless of the choice of the other



(a) 10th percentile UE throughput vs offered load

(b) Average UE throughput vs offered load

(c) Jain's Fairness index vs offered load

(d) PRB utilisation vs load

(e) Average UE throughput vs distance (at 5 seconds inter-spurt time)

(f) System throughput vs inter-spurt time

Figure 6.2: Graphs for different aspects of uncoordinated orthogonal sharing (FSA).

BS. This means that the expectation of the channel quality during scheduling does not align with the experienced channel quality due to added interference of users scheduled on the same PRBs from the other BS.

The offered load generated with the varying inter-spurt-times is about two thirds of the load generated with the uncoordinated orthogonal scheme for MSR, and a less drastic decrease for the PF and MM algorithms. The lower offered load at an inter-spurt time of 0 is likely to be caused by overestimation of the channel due to the added interference by the other BS transmitting on the same spectrum. This means that the BLER will be higher due to lower SINR, effectively increasing the error rate due to the choice of a wrong MCS. At lower inter-spurt times however, the offered load is similar to the load in the uncoordinated orthogonal sharing, suggesting that this scenario is capable of handling the same loads when the system is not fully loaded.

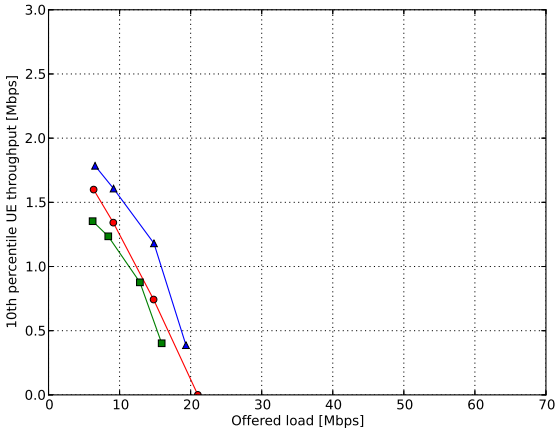
The PRB usage is 100% at full load as expected since each BS can use the whole spectrum. With lower load, we see the PRB usage decrease indicating an idle channel at times. The PRB usage drop of the MM algorithm is less than the other algorithms, which can be explained by the lower average user throughput of this algorithm due to the increased fairness. This lower throughput implies that the transfers of spurts are longer in duration, leading to higher PRB usage as the idle time is less.

For the average user throughput, this sharing method is overall worse than uncoordinated orthogonal sharing with MSR scheduling. For both PF and MM scheduling, the average user throughput is similar at higher loads, but higher at lower loads. The 10^{th} percentile UE throughput is overall higher for all schemes, but also has the highest gain at lower loads, just as the average throughput for PF and MM scheduling. A reason for this observation is that the total number of PRBs available is bigger, resulting (in the fair scheduling algorithms) in more assigned PRBs for users with lower channel qualities, which corrects for the diminished SINR due to added interference. Because MSR focuses on the best performing users, the drop in average user throughput is higher than with the other algorithms, as a user with higher attainable decreases relatively more in throughput than a user that would already have had a lower rate. With lower loads the system has more available bandwidth, and if no concurrent users at the other BS are scheduled, the bit rates will be higher across the board due to the ability to schedule more PRBs.

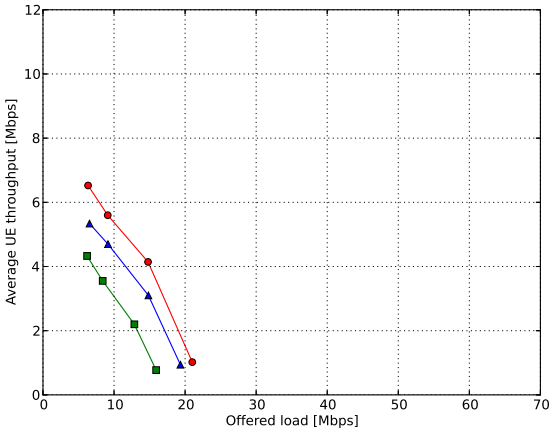
Over distance, the order of the different algorithms remains the same (MSR the best and MM the worst). However, the average throughputs are spaced more closely together. Notably, MM virtually has the same throughput over distance as in the FSA scheme. The tails of the throughput over distance creep more together, and at the edge of the cell, the MSR algorithm is even ever so slightly worse than the PF algorithm.

6.4.3 Coordinated orthogonal sharing

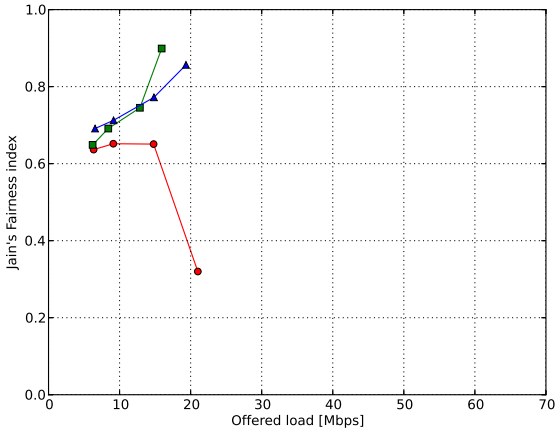
Coordinated orthogonal sharing (Figure 6.4a - 6.4f) is similar to uncoordinated orthogonal sharing in terms of the spectrum sharing: the spectrum is used in an orthogonal fashion. However, the BSs do not have their own fixed piece of spectrum, but the spectrum is coordinated in usage and shared in full between both BSs. The scheduling algorithms decide which PRB is assigned to which user on a system level and thus effectively also assigns the spectrum to the



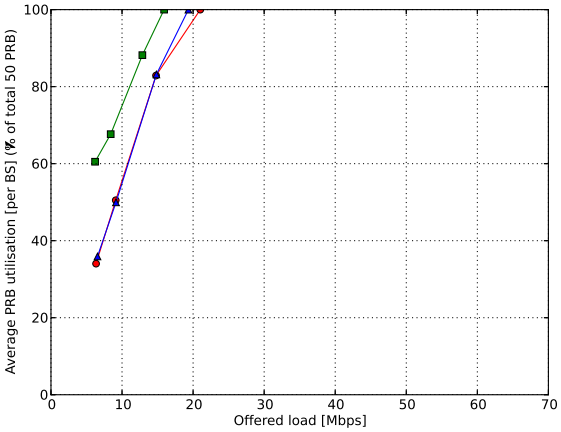
(a) 10th percentile UE throughput vs offered load



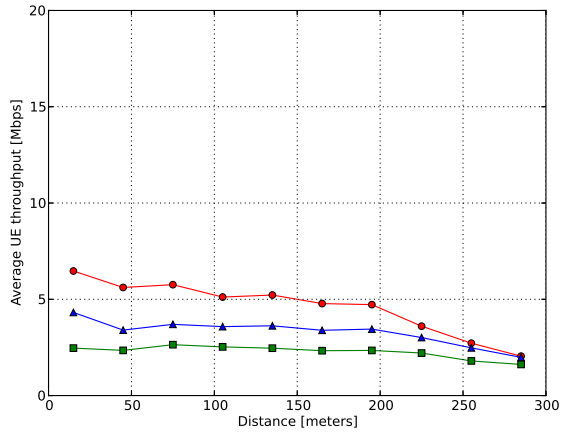
(b) Average UE throughput vs offered load



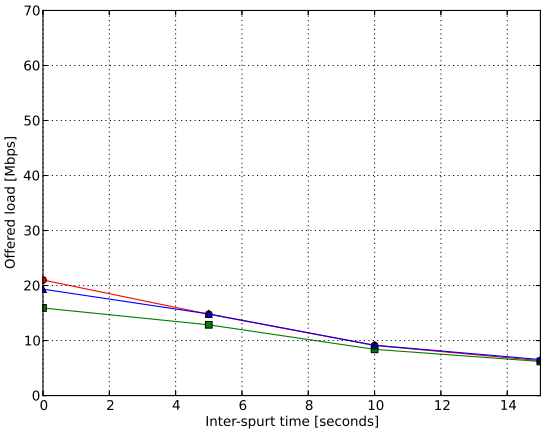
(c) Jain's Fairness index vs offered load



(d) PRB utilisation vs offered load



(e) Average UE throughput vs distance (at 5 seconds inter-spurt time)



(f) Offered load vs inter-spurt time

Figure 6.3: Graphs for different aspects of uncoordinated non-orthogonal sharing.

operators. As the spectrum is used orthogonally, we can not take advantage of the desirable properties of the SAPHYRE schemes as there is no interference on the channel generated by other users. However, as the SINR of NE is the same as SAPHYRE for orthogonal cases, this is a baseline in what the coordinated SAPHYRE schemes can provide when used orthogonally.

When we look at the offered load generated with different inter-spurt-times, we can see that the curves look almost like a replica of the same graph for the uncoordinated orthogonal scenario. The scenario does introduce more flexibility in scheduling as the whole spectrum is usable when only one operator has active users, which can be confirmed with the average UE throughput and the 10^{th} percentile throughput. In both the average UE throughput and the 10^{th} percentile throughput, we can observe overall higher throughputs for lower loads in PF and MSR scheduling. For maximum load the average user throughput and 10^{th} percentile throughput remains the same, which is an effect of that the users keep generating spurts, and the maximum capacity of scheduling over 50 PRBs is the same as scheduling over two times 25 PRBs. The MM scheduling algorithm does not gain as much from the coordination as the other scheduling algorithms do, as the scheduling algorithm now takes all users from both operators into account. This means that if one lower performing user is active in the system, all users will see lower rates as the system tries to maximise the rate of the lower performing user. Therefore, instead of only affecting the users of one BS, this now affects users of all BSs.

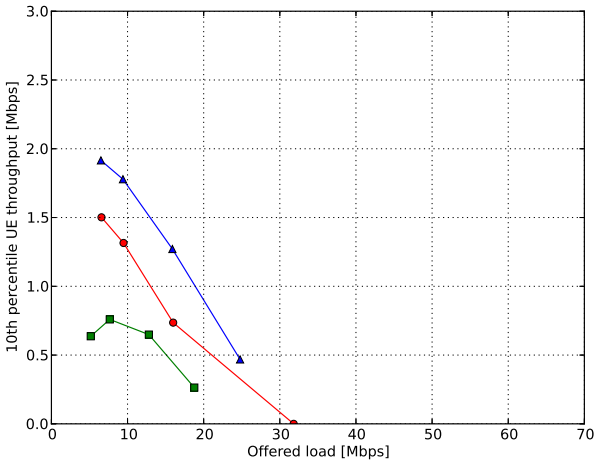
Higher average UE throughputs with similar offered load should translate to higher PRB usage, which is indeed the case for the MSR and PF scheduling algorithm. For MM, the PRB usage is lower on average for lower load, which is an effect of the decreased average UE throughput at lower load due to the fact that the scheduling takes place for the whole system and thus the slowest user in the system slows down the other users as well.

From the fairness graph, we can observe that PF and MSR are a little bit less fair, and MM slightly more fair for non-maximum loads compared to the uncoordinated orthogonal scenario. The latter can be explained as the MM now slows down the whole system instead of just the users of one BS when there is a slower user present, yielding a higher fairness as the rates over all users are more equal, and the former can be explained by the observation that the average rates are higher due to the possibility to exploit more PRBs for scheduling.

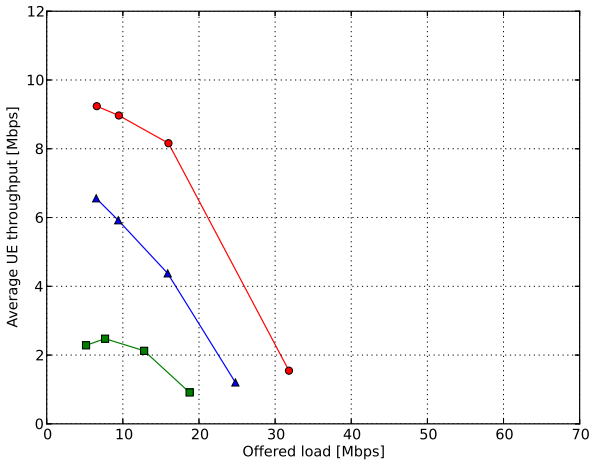
6.4.4 Coordinated non-orthogonal sharing

Lastly, we consider the coordinated non-orthogonal scenario, in which the scheduling is executed on a system level and the spectrum is used non-orthogonally. Besides the simulation of the advanced physical layer transmission schemes from the SAPHYRE project, we also simulate the ZF transmission scheme, a beamforming scheme that can be considered a reference scheme. The SAPHYRE schemes are expected to outperform the ZF transmission scheme, as outlined before. Furthermore, because of the specific challenges of non-orthogonal scheduling, we evaluate all three PF algorithms: PFSum, PFProduct and PFCombi. The results of the coordinated non-orthogonal scenario can be seen in Figure 6.5a - 6.5f.

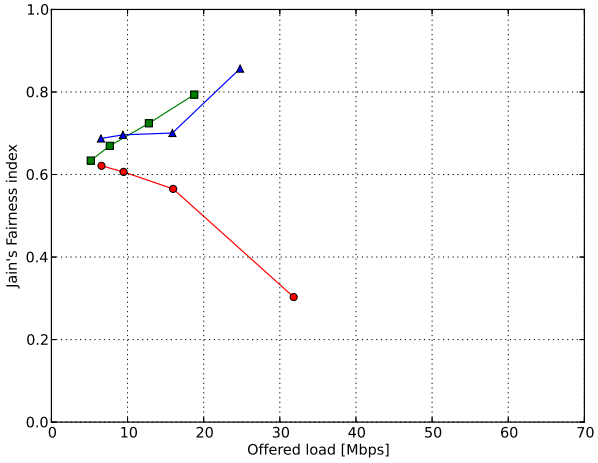
When we look at the offered load, we can observe an almost twofold increase of the maximum load (with 0 inter-spurt time) when compared to the orthog-



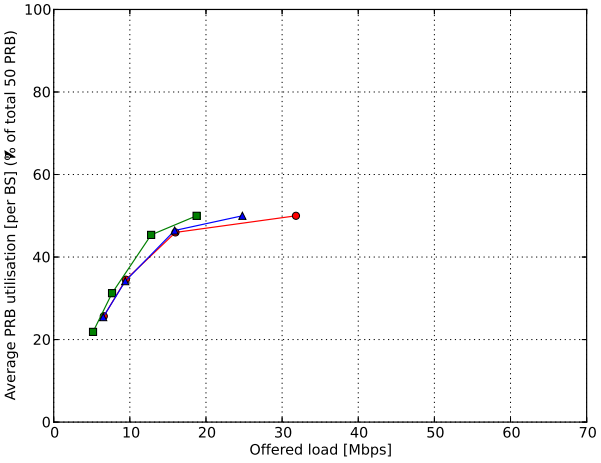
(a) 10th percentile UE throughput vs offered load



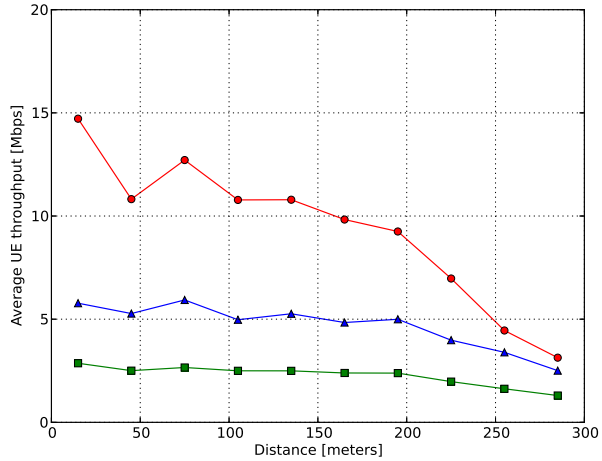
(b) Average UE throughput vs offered load



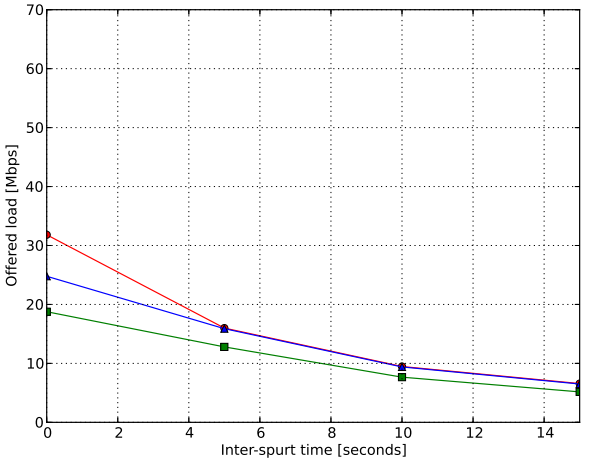
(c) Jain's Fairness index vs offered load



(d) PRB utilisation vs offered load



(e) Average UE throughput vs distance (at 5 seconds inter-spurt time)



(f) Offered load vs inter-spurt time

Figure 6.4: Graphs for different aspects of coordinated orthogonal sharing.

onal scenarios. For the lower offered loads due to a longer inter-spurt time, we can see that the coordinated non-orthogonal scenario sustains only slightly higher offered load than the orthogonal scenarios. This observation leads us to believe that most of the gain in throughput in this scheme is coming from the scheduling of combinations of two users at one PRB, combined with the interference cancellation in the transmission schemes. With longer inter-spurt times, the number of combinations we can make between active users during scheduling becomes less, making it less likely to schedule a good combination of users. Furthermore, when increasing the inter-spurt time even higher, the coordinated non-orthogonal scenario will behave the same as the coordinated orthogonal scenario as in most cases only one user will be active in the system due to the long inter-spurt times, forcing the system to schedule orthogonally. For the inter-spurt times of 5 and 10 seconds, we can still see a marginal increase over the coordinated orthogonal scheme as apparently sometimes the system can make good combinations of scheduled users.

Globally, the offered load for the different scheduling algorithms behaves similar as in the other scenarios, but we can see here that the elaborate PFProduct and PFCombi algorithms for PF scheduling perform significantly worse than the simple PFSum algorithm. In the PRB usage graph, we can see that at full load only 50% of the PRBs is used per BS for PFProduct and PFCombi scheduling, indicating that the bad performance stems from the way in which the priority indices are combined. In fact, both PFProduct and PFCombi scheduling algorithms seem to use the spectrum orthogonally in the majority of scheduling cases. This can be caused by users that have a priority index of in fact are always lower than the priority index of an orthogonally scheduled user, reducing the global scheme to an orthogonal scheme. As the performance of these scheduling algorithms is now essentially the same as PF in coordinated orthogonal sharing, we will not discuss these scheduling algorithms further here.

The expectation that the SAPHYRE transmission schemes would outperform ZF cannot be observed from the graphs. In fact, the MM transmission scheme with MM scheduling even performs significantly worse than ZF with MM scheduling. This can be explained by the fact that the MM transmission scheme specifically tries to equalize the SINR for the users by performing smart beamforming, resulting in lower SINR for the user with better channel quality in the ZF transmission scheme and marginally higher SINR for the weaker users. For the other transmission schemes, the curves for all graphs are similar if not almost the same. The observation that the MSR and PF scheduling algorithms have similar performance under both ZF and SAPHYRE transmission schemes can be explained because the interference of the other user is quite high, and effectively nulled out by both algorithms. However, the SAPHYRE schemes are also applicable in other scenarios like when each BS serves multiple users per PRB by spatial multiplexing. The interference situation will then change drastically, and the SAPHYRE schemes will probably outperform ZF in that scenario. Also, in this light a non co-sited scenario might as well show improvement. Furthermore, the SAPHYRE schemes might perform better in this scenario with less users, as described in [2].

Both the average UE throughputs and the 10^{th} percentile throughputs are consistently higher than all other scenarios evaluated. Combined with a slightly higher system throughput at longer inter-spurt times, this means that the users have consistently better rates due to the non-orthogonal use of the spectrum

and the coordinated scheduling. This in turn means that spurts of comparable size are downloaded to the UE faster, providing a better user experience.

In Figure 6.5c, we can see the fairness of the different combinations of scheduling algorithm and transmission scheme in the coordinated non-orthogonal scenario. Because we work with more scheduling algorithms in this scenario, and all scheduling algorithms are evaluated for two transmission schemes, the graph is a bit crowded in the 60% to 80% area. Therefore, we included an enlargement of this part in Figure 6.6. We can immediately see that, although the MM algorithm with the MM transmission scheme performs worst in the average user throughput and offered load metrics, it performs stronger than MM with the ZF transmission scheme in fairness. In fact, when we look at the fairness of MM scheduling with MM transmission, than it seems to be more in line with the fairness in other scenarios for MM scheduling than with the ZF transmission scheme. Again, this is caused by the optimization in the MM transmission scheme that tries to equalize the SINR of the scheduled users, leading to better fairness between users. Another observation is that both the PFCombi and PFProduct scheduling algorithms perform best in fairness with 85% although they in fact schedule more orthogonally and perform bad in the other metrics.

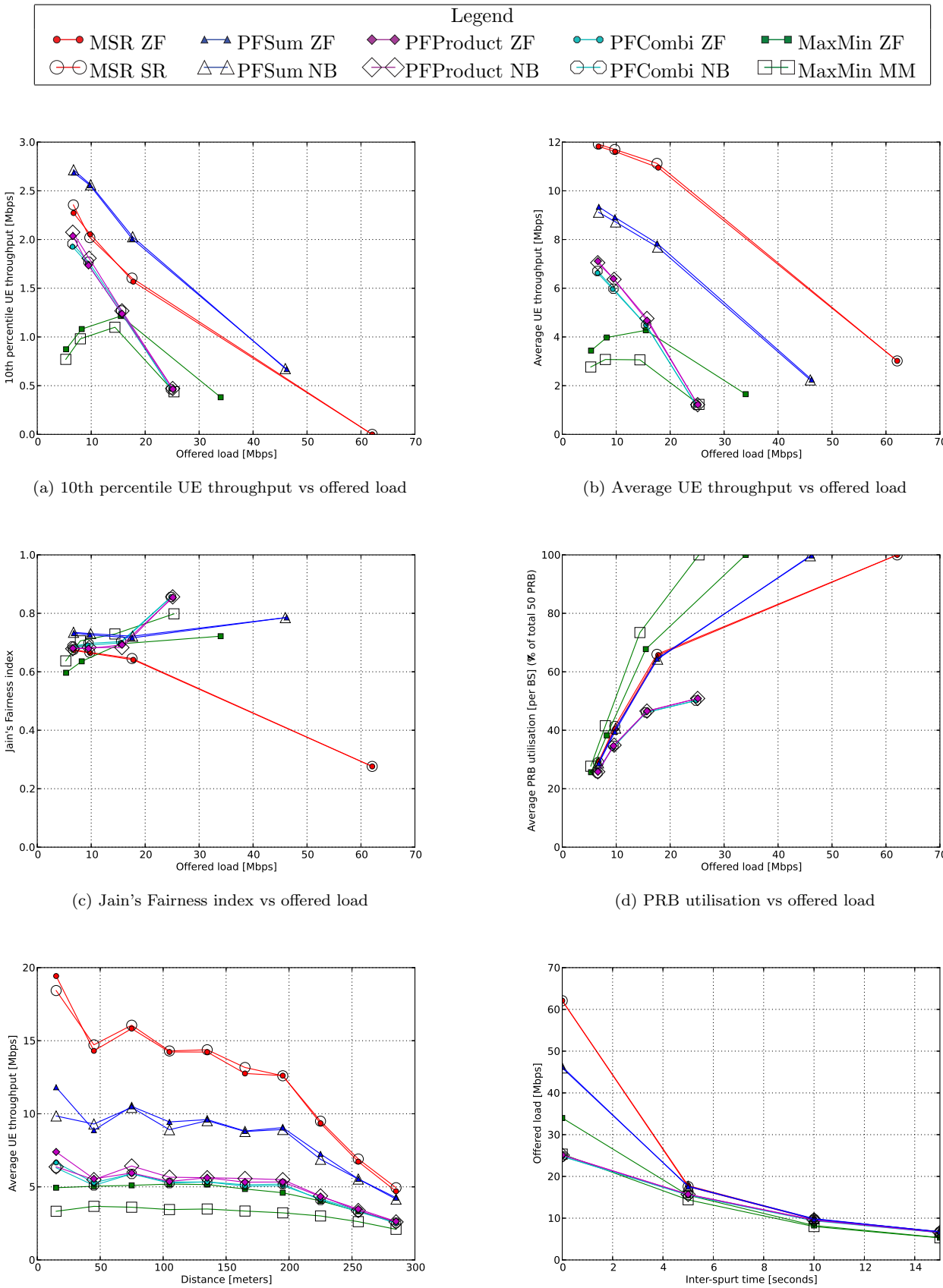
6.4.5 Sharing scenario comparison

Now that we have evaluated the various sharing scenarios with respect to the metrics, we can compare these sharing scenarios with each other. From a scenario perspective, we can see that the coordinated non-orthogonal scenario has the best performance across the board. This comes however at the cost of coordination between operators, so we will analyse the sensitivity to the delay caused by this coordination later in this report. Furthermore, the coordinated orthogonal scenario shows promise over the uncoordinated orthogonal scenario in average user throughput and 10^{th} percentile throughput. As the uncoordinated non-orthogonal scenario performs worst, we cannot say that non-orthogonal spectrum sharing alone is the holy grail for the imminent spectrum crunch. However, it can be a solution when combined with coordination between operators. In this section we will compare the different scenarios for each metric.

Offered load

The offered load metric is an important metric to compare the schemes with each other. Due to the design of the system, this metric can be directly related to the system throughput, and with an inter-spurt time of 0, we can show the theoretical upper limit of the system throughput.

As we have observed before, only in the simulations with an inter-spurt time of 0 the offered load really differs in the different sharing scenarios. With higher inter-spurt times, the load becomes more or less the same for all scenarios, although the coordinated non-orthogonal scenario consistently reaches a little bit more load. The convergence of the offered load at higher inter-spurt times is caused by the fact that the inter-spurt times are multiples of the transmission time for most users. As this is the case, marginal average speed differences do not show in the offered load metric. For 0 inter-spurt time however, all users directly generate new spurts after finishing the previous spurt, so we can see the maximum maintainable system throughput. When the load is maximised,



(a) 10th percentile UE throughput vs offered load

(b) Average UE throughput vs offered load

(c) Jain's Fairness index vs offered load

(d) PRB utilisation vs offered load

(e) Average UE throughput vs distance (at 5 seconds inter-spurt time)

(f) Offered load vs inter-spurt time

Figure 6.5: Graphs for different aspects of coordinated non-orthogonal sharing.

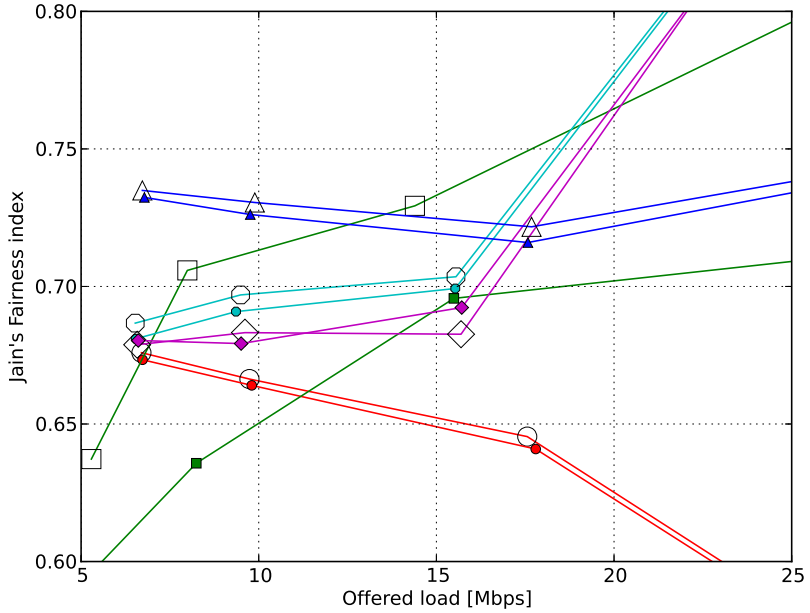


Figure 6.6: Enlargement of Figure 6.5c: Jain's Fairness index vs offered load.

the uncoordinated non-orthogonal scenario performs worst. This is caused by the added interference of independently scheduling BSs, increasing the BLER and thus decreasing the experienced bit rates. The coordinated non-orthogonal scenario on the other hand, performs best of all sharing scenarios under full load with an almost twofold increase over the orthogonal scenarios and even more over the uncoordinated non-orthogonal scenario. The coordinated non-orthogonal scenario benefits well from the possibility to schedule multiple users on the same PRB and from the coordination of the scheduling between operators.

Average user throughput and 10^{th} percentile

When evaluating the average user rates and the 10^{th} percentile, we can immediately confirm our findings about the benefits for the coordinated non-orthogonal scenario of scheduling multiple users per PRB. The experienced average user throughput is significantly higher with gains ranging from 20% of up to 80% when compared to the coordinated orthogonal scenario. This is a gain that is important to the end users as they will see better rates. The highest gains are visible under full load, which is as expected because the scheduling algorithms can then make the best combinations as most users are active simultaneously. When we compare the coordinated non-orthogonal to the uncoordinated orthogonal scheme, we can also directly observe an increase in the average UE throughput of nearly 100% at maximum.

Similar gains as the increase in average user throughput can be observed in the 10^{th} percentile throughput for the coordinated non-orthogonal scenario.

For full load however, the MSR algorithm does not show an increase in 10^{th} percentile throughput as the algorithm will also in this scenario choose the users with the best channel quality.

An important observation from the average UE throughput over distance is that the curves of the better performing scenarios are steeper than the curves of the scenarios that perform worse. However, the average UE throughput at the cell edge is also increased, so not all of the performance gain we see in the average UE throughput comes at the expense of users at the cell edge.

Fairness

Fairness-wise, the general trend of the PF and MM algorithms is that they become more fair towards users when the system has more offered load. This is due to the fact that the scheduling algorithms are able to choose between different users when multiple users are active simultaneously. With lower loads however, it is less common that multiple users are active, so the algorithms cannot do anything different than allocating as much spectrum as possible. This means that the fairness will become similar to the fairness in attainable bit rates between users in extreme low loads. The MSR algorithm follows a downwards trend as the algorithm allocates most of its PRBs to users with good channel quality when multiple users are active. This means some users will get little or no PRBs assigned, leaving an unfair distribution of the resources.

PRB usage

The main difference in PRB usage is the difference between non-orthogonal and orthogonal sharing, where the BSs will on average use either maximum 50% or 100%, respectively. With maximum offered load, all scenarios reach their maximum PRB usage, but PFCombi and PFProduct in the coordinated non-orthogonal scenario. The reason for the lower PRB usage with the PFCombi and PFProduct algorithm is that the way in which the priority indices are combined in these algorithms seems to promote orthogonal scheduling combinations, although the scenario enables non-orthogonal combinations as well. With lower offered loads, the PRB usage decreases in all scenarios as the BSs are idle for a certain amount of time due to times when no active users are available for scheduling.

6.4.6 Scheduling algorithm comparison

From an operator's perspective user experience is a very important issue. If you consistently provide bad user experience, your clients will eventually find another operator as the playing field is very competitive.

Overall, for systems which are not under full load, the MSR algorithm performs best. The PF algorithm does provide a better overall experience as this algorithm provides better bit rates when the system is fully loaded and is not the worst performer in the rest of the cases. MM can be regarded as the weakest scheme, as the fairness is not that much better than PF, and the trade-off in average user throughput is quite dramatic as well as the trade-off in system throughput. Furthermore, MM is the most complex algorithm, making the MM

algorithm even less favourable. Because of the desirable properties of both PFSum and MSR, we select these scheduling algorithms for sensitivity analysis to see how the algorithms perform under less favourable conditions.

Of the three forms of PF scheduling algorithms, we already shortly analysed that the PFSum algorithm was the best. This is mainly due to the fact that both PFProduct and PFCombi make almost orthogonally use of the spectrum, leaving no room for the gains we can get from non-orthogonal sharing. This reduces the algorithms to the same performance as the PFSum algorithm in the coordinated orthogonal scenario. This leads to the conclusion that the way in which the two priority indices of the users are combined in PFProduct and PFCombi are suboptimal, leading to lower priority indices than the orthogonal decision would.

In the remainder of this section, the different scheduling algorithms will be compared with regards to the various metrics.

Offered load

In terms of offered load, the MSR scheduling algorithm consistently provides the highest offered load, indicating that it can handle the most traffic. This is quite logical when we remember that the objective of MSR scheduling is to maximize the sum-rate of the system. Yet, this high throughput is only reached with higher system loads. With lower loads it eventually decreases to the throughput of PF scheduling. The lowest system throughput is consistently generated by the MM scheduling algorithm, as this algorithm will give a fair amount of attention to users with bad channel conditions.

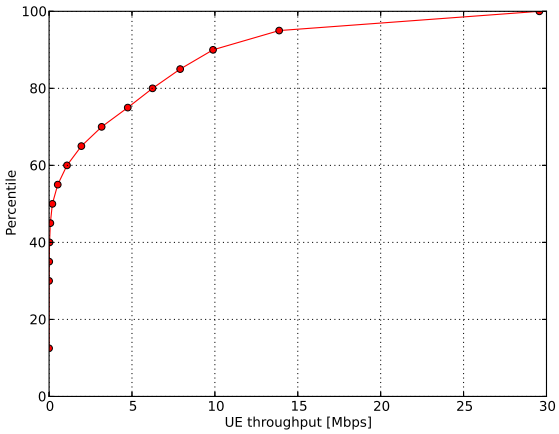
Average user throughput and 10^{th} percentile

The average user throughput shows the same order of the scheduling algorithms as does the offered load metric; the MSR scheduling algorithm reaches the highest average user throughput. This also remains true with lower offered loads, suggesting a higher spread of experienced average user throughput than the PF or MM algorithm. Indeed, in the 10^{th} percentile throughput graphs, we can observe that this larger spread is the case as the PF (PFSum for the coordinated non-orthogonal scenario) takes the lead in this metric. Furthermore, 10% of the users actually experience a rate of zero when the system is fully loaded with MSR scheduling, indicating that the MSR scheduling algorithm takes its objective very seriously at the expense of users experiencing bad channel quality.

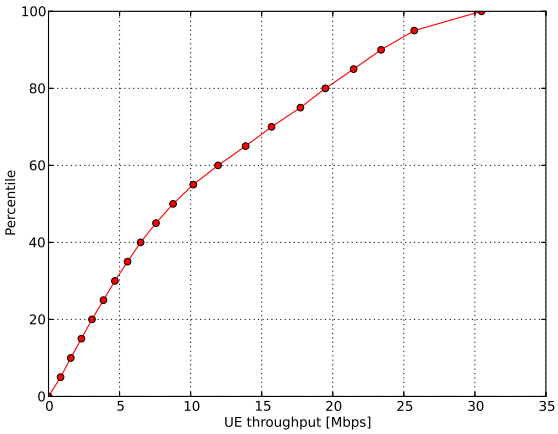
With even longer inter-spurt times than we simulated, the average rates for the non-orthogonal scenarios will likely decrease to the coordinated orthogonal scenario as scheduling combinations become scarce when less users are active simultaneously. We can see this effect already with MM in the coordinated non-orthogonal scenario. Furthermore, PF and MSR already show signs of stabilization of the average user throughput with lower loads. Although MSR provides higher average rates, from an operator's perspective the PF algorithm is better for the user experience when the system gets fully loaded.

Fairness

Fairness-wise, the MM and PF algorithms both score similarly with increasing fairness when the load increases and between 60% up to almost 90% fairness.



(a) CDF for an inter-spurt time of 0



(b) CDF for an inter-spurt time of 5

Figure 6.7: Cumulative Distribution Function (CDF) of UE throughput for the MSR scheduling algorithm in the coordinated non-orthogonal scenario

MM seems to flourish with uncoordinated non-orthogonal sharing, in which it reaches its top fairness. For the other scenarios, the PF algorithm takes the lead. The MSR scheduling algorithm is less fair with its fairness ranging from 25% under full load up to 65% under lower loads. The low fairness of MSR scheduling under full load makes you wonder how many users actually experience zero bit rate. Figure 6.7a shows the CDF for MSR scheduling under full load in the coordinated non-orthogonal scenario. As can be observed from this graph, around 40% of the users experience zero bit rate and 75% of the users experience maximum 5 Mbps on average, making the MSR algorithm quite unsuitable for operators when they experience full load. For reference, Figure 6.7b shows the CDF of user throughput for an inter-spurt time of 5. We can observe from this graph that the average UE throughput is spread more even over the users. In contrast to the other scheduling algorithms, with the MSR algorithm fairness increases under lower loads. This can be explained by the fact that the users with high rates will be served fast, and the users with lower rates can be served in-between these high-performing users, which is not possible under full load as the high performing users keep being active. As the MSR algorithm does not deserve a 'fair' predicate, this round is a tie between PF and MM scheduling. However, we need to remember that MSR fairness increases radically when the system is not fully loaded.

6.5 Sensitivity analysis (coordinated non-orthogonal sharing)

In this section we analyse the sensitivity of the PFSum and the MSR scheduling algorithm in the coordinated non-orthogonal scenario, with respect to interference, error in the SINRs and feedback delay. To get a fair comparison, we

consider one deviating parameter from the default values at a time, with all other parameters kept equal. This ceteris paribus assumption rules out the possibility that other factors influence the observed effects, and allows us to focus on the sensitivity to one varying parameter at a time. Because redoing all the simulations for all inter-spurt times would be very time-sensitive, we choose to simulate with an inter-spurt time of five seconds, as this is the lowest setting where the system throughput of the different algorithms is similar. Further parameters are kept at their default values as described earlier in this Chapter. As the error, delay and interference could likely have some impact on the spread of measured values, we include a 95% confidence interval in all graphs for the sensitivity analysis.

We are mainly interested in the sensitivity of the transmission schemes in the coordinated non-orthogonal scenario, as this includes the transmission schemes developed by the SAPHYRE project. Furthermore, as we argued before, the coordinated non-orthogonal scenario is the best performing scenario. For the scheduling algorithms we choose to evaluate the MSR and the PFSum algorithm as they are the best performing algorithms in all scenarios.

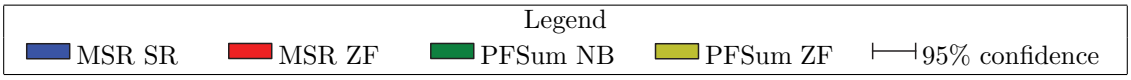
6.5.1 Sensitivity to interference of surrounding cells

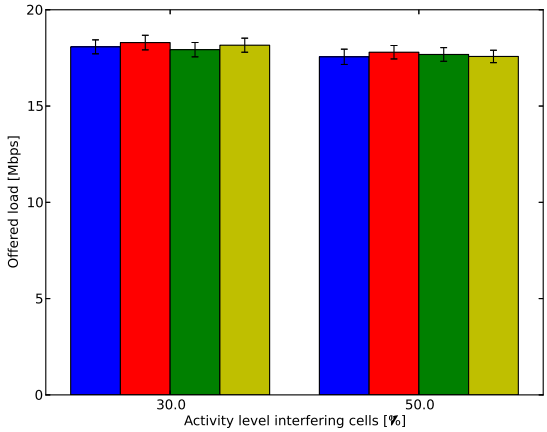
To analyse the effect of interference of surrounding cells on the system, extra input data from the SAPHYRE partners is needed as the interference is included in the traces for the transmission schemes. The default parameter of the interference generated by the surrounding cells is based on an activity level of 50%. Due to the constraints of the input data, we only have one other option at our disposal: an activity level of 30%. Figure 6.8 shows the results of the simulations with an activity level of 30% and the corresponding simulations with the default setting of 50%.

As can be observed from Figure 6.8a, the total offered load increases only marginally with lowered interference. As the 95% confidence interval overlap partly with both interference activity levels, we cannot be sure that the marginally lower offered load, we cannot be confident that the effect is caused by the change of the interference level or just by the specific simulation. If the increase is caused by the interference level, we should be able to see an increase in the average user throughput as well.

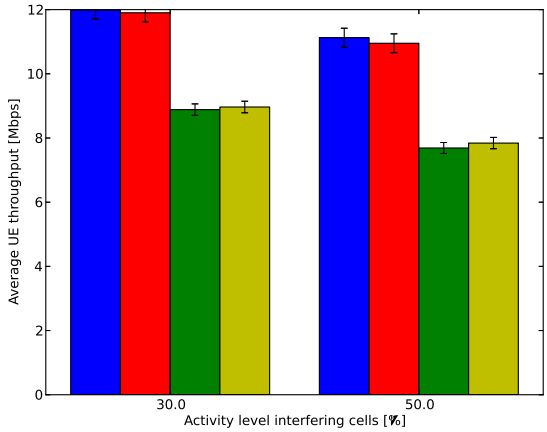
Figure 6.8b confirms the increased average user throughput with quite some difference between the levels of interference. We can observe from the average user throughput graph that the PFSum algorithm has more gain (about 15%) than the MSR algorithm (about 8%). An explanation for this effect can be that the MSR algorithm already focuses on scheduling the users with the highest attainable bit rates. Users that would already would have had the highest MCS in the default scenario will not see a bit rate increase. Therefore, these users with high bit rates are still the most likely to be scheduled. Only the users with lower bit rates see the real advantage of the lowered interference. In the PFSum algorithm, users with lower bit rates in the default scenario gain more from the increased SINRs as their increased attainable bit rates increases their priority indices making it more likely that they are scheduled.

Figure 6.8c shows with the 10^{th} percentile user throughput that the reduction in interference is beneficial to all users in the system. For MSR this is valid as well, because the increased bit rates of the users makes more room for users

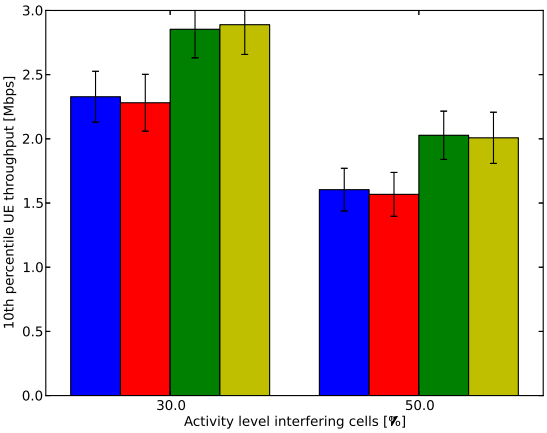




(a) Impact of activity level on offered load



(b) Impact of activity level on average UE throughput



(c) Impact of activity level on 10th percentile UE throughput

Figure 6.8: Impact of the activity level of interference generated by surrounding cells on offered load, average UE throughput, and 10^th percentile UE throughput.

with smaller bit rates as all spurts are finished faster. In the PFSum algorithm, the increase in 10^{th} percentile throughput stems from the fact that users with lower attainable bit rates will have more chance to be scheduled due to their increased bit rates when compared to the default scenario.

6.5.2 Sensitivity to feedback delay

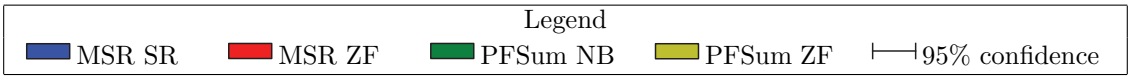
The effects of introducing a 4- and 8-TTI delay in the SINRs used for scheduling are displayed in Figure 6.9. As can be observed from these graphs, the difference between no delay and 4- or 8-TTI delay is negligible. As explained before, this is as expected as the users move at pedestrian speed and thus the channel quality changes only marginally within this time window. As the results are very close and within each others 95% confidence intervals, we cannot observe any trend. This leads us to the conclusion that a 4- to 8-TTI delay in the channel quality does not impact users moving at pedestrian speed. However, as is mentioned in [59], pedestrian users should be able to tolerate up to a 10-TTI delay without significant impact. It might therefore be a good idea to test this assumption in a follow-up study by simulating even longer delays.

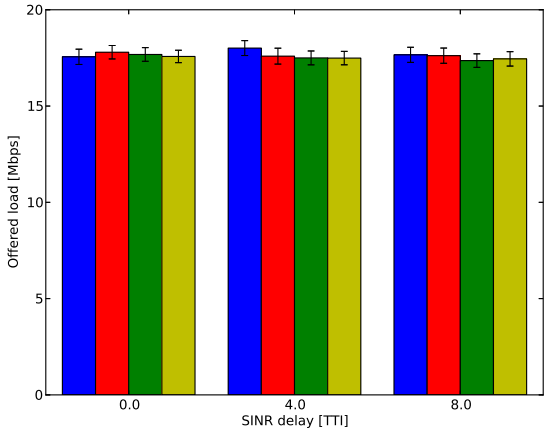
6.5.3 Sensitivity to feedback error

For the SINR error, we introduced three different values of the error. The error is sampled from a Gaussian distribution with a standard deviation of 1 dB, 2 dB, 3 dB, 4 dB or 5 dB and a mean of 0 dB. Figure 6.10 shows the effect of introducing the error in the SINRs. As the SINR error has a mean of 0 dB, the SINR can both be underestimated or overestimated with each scheduling decision for each combination of users. This means that, in the case of underestimation, the channel quality of the user will be better than expected during scheduling and MCS selection, and in case of overestimation, the rate of the user will be worse than expected. Selecting the wrong MCS for a certain channel quality causes the BLER to be higher when the channel quality is overestimated, leading to lower bit rates. When the channel quality is underestimated however, the BLER is lower resulting in a good transmission, though possibly a better MCS could have been chosen that would have increased the bit rate. Furthermore, due to working with uncertain data in the scheduling, the scheduling algorithms likely make suboptimal decisions in the scheduling combinations, leading to lower throughputs. Due to these concerns, we expect that the performance of the system in both user throughput and total offered load will decrease with increasing error.

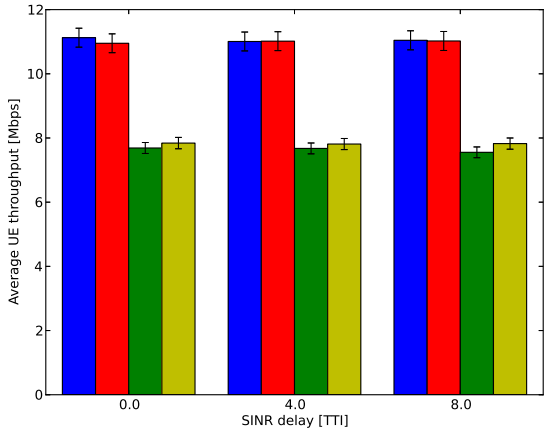
Looking at the offered load (Figure 6.10a), a decrease is clearly visible with errors of 3 dB and higher. With 1 dB and 2 dB of error, the total offered load is similar to the default of 0 dB error. A slight downwards trend seems visible, but because the 95% confidence intervals are overlapping, purely based on that data we cannot be certain of a decrease. With 3 dB, 4dB and 5 dB error, a downwards trend in offered load begins speeding up. With 5 dB error, the offered load is only 60% for PF scheduling, and 68% for MSR. Such a decrease in the offered load means that the it takes the users significantly longer to finish their spurts.

We can confirm that the average spurt will take longer to complete by looking at the average user throughput (Figure 6.10b). We can observe the same trend

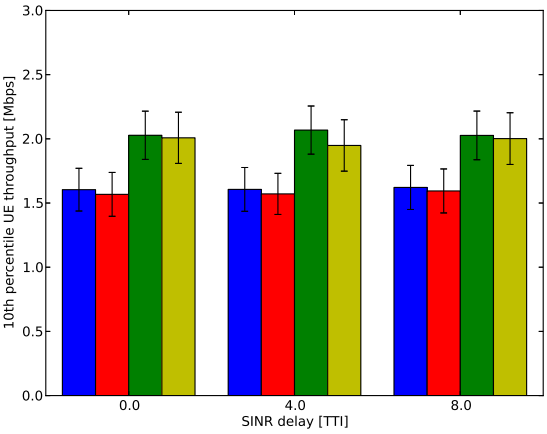




(a) Impact of delay on offered load

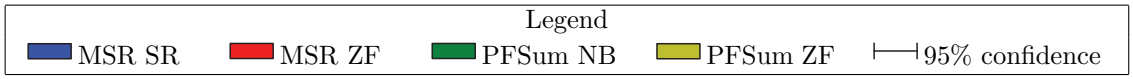


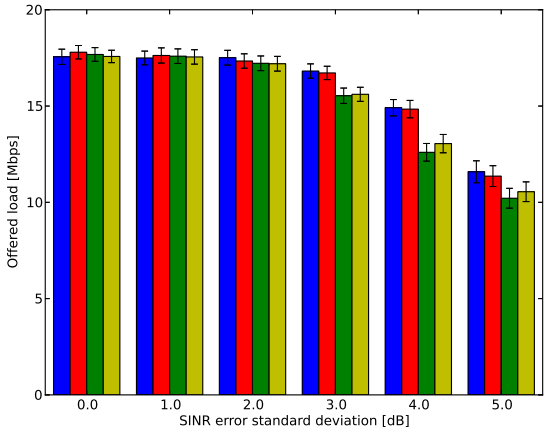
(b) Impact of delay on average UE throughput



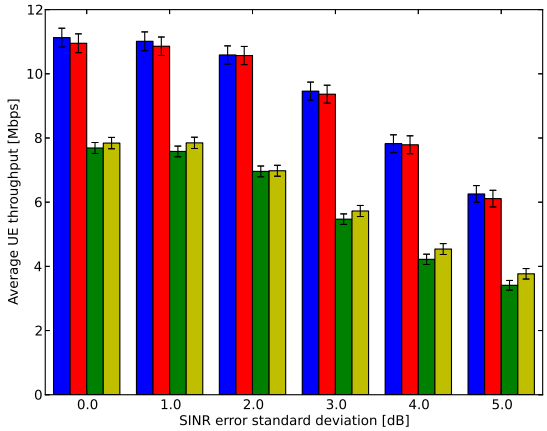
(c) Impact of delay on 10th percentile UE throughput

Figure 6.9: Impact of feedback delay on offered load, average UE throughput, and 10^th percentile UE throughput.

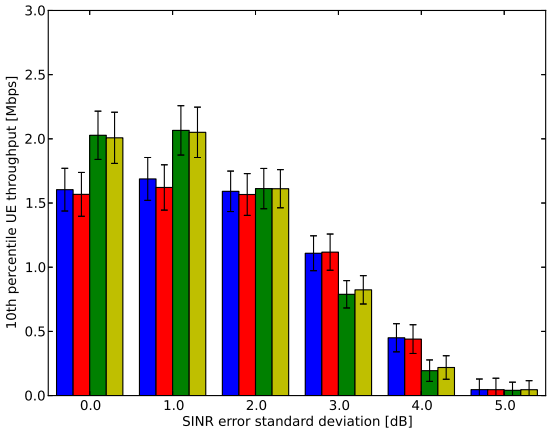




(a) Impact of SINR error on offered load



(b) Impact of SINR error on average UE throughput



(c) Impact of SINR error on 10th percentile UE throughput

Figure 6.10: Impact of feedback error on offered load, average UE throughput, and 10^th percentile UE throughput.

as in the offered load with the real decrease beginning at an error of 3 dB. However, in this graph, we can also observe that the PF algorithm starts its descent already clearly at an error of 2 dB.

The 10^{th} percentile throughput (Figure 6.10c) shows again the same downwards curve as the the average user throughput. In the 10^{th} percentile throughput however, we can observe that the PF algorithm performs higher than the MSR algorithm with 0 dB and 1 dB error; it performs similar at an error of 2 dB, and worse than MSR at errors of 3 dB and higher. Due to this effect and the increased rate of decrease for PF in the other metrics, we can say that the MSR algorithm is more resilient to SINR error than the PF algorithm. This can be explained by the focus of the MSR algorithm at users with good channel quality. The high channel quality users in a scenario with error will most of the time still be users in the higher channel quality regions that either overestimated their channel quality to unrealistically high values or have underestimated a little, causing a higher BLER but on a high bit rate as well. So the remaining bit rate might be higher than a user in the lower channel quality regions that the PF algorithm is more likely to serve, that have overestimated their channel quality to the point where they can actually receive data but in reality are below the threshold for any successful data transfer. Furthermore, a higher BLER on lower bit rates in effect still means even lower experienced bit rates. So the focus on the higher channel quality users of the MSR algorithm is the quality that makes it more resilient to SINR error.

Taking all the metrics into account, we can conclude that the MSR algorithm is more tolerant to smaller SINR error than the PF algorithm. The MSR algorithm will not see significant impact on offered load, average user throughput and 10^{th} percentile throughput for errors of 1 dB or 2 dB, while the PF algorithm only tolerates a 1 dB error before decreasing significantly in the average user throughput and the 10^{th} percentile user throughput.

Chapter 7

Conclusions and future work

This thesis quantified the performance of different scenarios of spectrum use on a system level including model parameters that reflect a realistic environment. In this chapter, we conclude this thesis with general conclusions about the results and analysis presented in the preceding chapter. Furthermore, we present recommendations for future work.

7.1 Conclusions

The transmission schemes developed by the SAPHYRE project have been evaluated at link-level assessments, lacking realistic aspects of network operation like scheduling, feedback delay, multi-user traffic, propagation environments and network layout. To provide a realistic system-level simulation, we modelled a realistic urban environment with WINNER II empirical channel traces, two co-sited operators and 3000 users, homogeneously spread over the cell and divided in 150 groups of 20 randomly selected users which are active simultaneously. We introduced a traffic model consisting of multiple spurts per user with an exponentially distributed inter-spurt time, and the size of the spurts distributed as the empirical website size data as found in the HTTP archive gathered with mobile phones. Furthermore, we described a physical layer abstraction and used it in the simulator to incorporate the transmission schemes of the SAPHYRE project as well as the NE transmission scheme serving as a reference for LTE and the ZF transmission scheme serving as a reference for non-orthogonal spectrum sharing. Furthermore, we described the MSR, PF and MM scheduling algorithms and suggested a way to deal with scheduling in a coordinated non-orthogonal scenario because scheduling two users on the same resource is a significantly different problem from scheduling only one user. For PF scheduling this led to three possible scheduling algorithms: PFSum, PFCombi and PFProduct.

In order to assess the results of the spectrum sharing scenario simulations, we analysed the metrics regarding the load offered to the system, the average user throughput, the 10^{th} percentile user throughput, fairness, and PRB usage. Based on this analysis, we selected the coordinated non-orthogonal scenario with

PFSum and MSR scheduling for sensitivity analysis to analyse how the system holds under feedback delay, feedback error, and interference of surrounding cells.

7.1.1 Answer to the research questions

In the introduction of this thesis, we established the following research question: *what can we gain in terms of performance and capacity at the system level, by applying the advanced transmission schemes for non-orthogonal sharing, as developed in the SAPHYRE project, with respect to Fixed Spectrum Allocation, orthogonal sharing, and non-orthogonal sharing with the ZF transmission scheme?* In order to answer this research question, we break it up into three separate questions:

- What can we gain with respect to uncoordinated orthogonal sharing (FSA)?
- What can we gain with respect to coordinated orthogonal sharing?
- What can we gain with respect to coordinated non-orthogonal sharing with the ZF transmission scheme?

First, we will focus on above questions, followed by a focus on the sensitivity of the SAPHYRE scheme to feedback delay, feedback error and interference of surrounding cells.

SAPHYRE gain

When we compare the SAPHYRE transmission schemes in the coordinated non-orthogonal scenario with uncoordinated orthogonal sharing, we can immediately observe an improvement across the board. Not only can the operators use the whole spectrum in the coordinated non-orthogonal scenario, but they can also use it at the same time (with coordinated scheduling between the operators). When we look at both scenarios under full load, we can observe an almost twofold increase in offered load and average user throughput for all scheduling algorithms. This comes at a small cost in fairness for both PFSum and MSR scheduling. At lower loads the offered load to the system converges to the uncoordinated orthogonal scenario. However, the almost twofold increase in average user throughput remains. For the 10^{th} percentile user throughput it deserves mention that the MSR scheduling algorithm does not increase its metric with the SAPHYRE transmission schemes at full load, but does at lower loads, like the PFSum algorithm. The PFProduct and PFCombi scheduling algorithm perform similar to the PFSum algorithm in the uncoordinated orthogonal scenario in offered load albeit with a little bit higher average user throughput and 10^{th} percentile user throughput.

Compared to the coordinated orthogonal scenario, the SAPHYRE transmission schemes in the coordinated non-orthogonal scenario also show improvement. The coordinated orthogonal scenario extends the uncoordinated orthogonal scenario with coordination between operators, and thus with the ability to use the whole spectrum instead of only their own spectrum share in a coordinated fashion. As it is still an orthogonal scheme, the maximum offered load does not change, so the situation remains the same: an almost twofold increase for the SAPHYRE transmission schemes. However, the ability to coordinate the spectrum usage between operators has its effect on the average user throughput and

the 10^{th} percentile throughput in comparison with uncoordinated orthogonal, as operators are allowed to use 100% of the spectrum when the other operator has no active users. Therefore, the gain in average user throughput and 10^{th} percentile throughput is less dramatic between the SAPHYRE transmission schemes in the coordinated non-orthogonal scenario and the coordinated orthogonal scenario than above comparison with the uncoordinated orthogonal scenario. At full load, the increase is almost twofold in average user throughput, but at lower loads the gain decreases to around 35%.

The last comparison is between the SAPHYRE transmission schemes and the ZF transmission scheme, both in the coordinated non-orthogonal scenario. Although the SAPHYRE transmission schemes are expected to outperform the ZF transmission scheme, the performance of both is very similar if not almost the same for the MSR, PFSum, PFProduct, PFCombi, and MM scheduling algorithms. Only the MM algorithm with the SAPHYRE transmission scheme (MM) shows consistently worse performance than the MM algorithm with the ZF transmission scheme. The similar performance between the ZF and SAPHYRE transmission schemes might be caused by, as suggested by SAPHYRE project partners, the co-sited setup of the scenario. Furthermore, when more than one user per PRB per operator is served, the interference situation will change dramatically, likely giving the SAPHYRE schemes better performance than the ZF transmission scheme.

Sensitivity

For sensitivity analysis, we selected the MSR and the PFSum scheduling algorithm to evaluate the sensitivity of the SAPHYRE schemes to feedback delay, feedback error and interference of surrounding cells. Furthermore, for comparison we included the ZF transmission scheme to see how the SAPHYRE schemes stack up to this transmission scheme in suboptimal conditions.

The interference level of surrounding cells (expressed in an activity percentage) does not have a significant effect on the offered load to the system when comparing between 30% and 50% interference. However, with a higher interference level, the average user throughput drops by 15% for the PFSum algorithm and by 8% for the MSR algorithm. In the 10^{th} percentile throughput we can observe a drop as well, with the PFSum algorithm again taking a higher drop than the MSR algorithm.

The effect of feedback delay on the average user throughput, 10^{th} percentile user throughput and offered load is negligible for a 4- or 8-TTI delay. This is mainly caused by the fact that we simulate pedestrian users with low speed, which are suggested in literature to tolerate a delay of up to 10 TTIs. However, for coordinated spectrum sharing an 8-TTI delay seems enough to account for delays in LTE (4 TTIs) and delays in coordination (another 4 TTIs).

The effect of feedback error in the SINR values has a somewhat larger impact on the system. An error of 1 dB is manageable by both the MSR and the PFSum algorithm, but with 2 dB the PFSum algorithm begins decreasing average user throughput and 10^{th} percentile user throughput. The MSR algorithm is stable under errors up to 2 dB and therefore a little bit more resilient to the error than the PFSum algorithm. With an error of 5 dB, the PFSum algorithm takes an almost 40% decrease in offered load, where the MSR algorithm takes a smaller decrease of 32%. The average user throughput decreases

by around 50% for the PFSum algorithm and 45% for the MSR algorithm. The 10^{th} percentile throughput really suffers with a meagre less than 0.1 Mbps 10^{th} percentile throughput left for both scheduling algorithms.

The MSR scheduling algorithm is slightly more resilient in the sensitivity analysis. However, this comes at the cost of selecting only the best performing users to be scheduled, which is disastrous for the 10^{th} percentile user throughput under full load, as we have demonstrated before. No observable differences can be found between the ZF transmission scheme and the SAPHYRE transmission schemes in sensitivity to feedback delay, feedback error and interference of surrounding cells.

7.2 Future work

Whilst working on this thesis, it was suggested by the SAPHYRE project partners that the gain of the SAPHYRE schemes might be even higher when not using a co-sited scenario. Furthermore, serving even more than one user per PRB per operator might also drastically change the interference, giving the SAPHYRE schemes a gain over the ZF transmission scheme. In this light, we recommend to run simulations for the non co-sited scenario as well to verify this claim. Furthermore, we recommend extending the simulator to provide a way of simulating multiple users per PRB per operator. With minor modifications, the PFSum, MSR and MM scheduling algorithm will work with this new extension.

Furthermore, we recommend to also investigate a more generic way to analyse the effects of spectrum sharing. For instance, this could be a generic algorithm that generates SINR values instead of the complex physical layer abstraction for which we need third party input. This generic process could be used to quantify results regarding questions relating to the impact of the realistic aspects of a system-level simulator on for instance an average increase of SINR values of 3 dB.

Bibliography

[1] M. Haardt, Z. K. M. Ho, E. Jorswieck, E. Karipidis, J. Kibilda, J. Li, J. Lindblom, C. Scheunert, J. Sýkora, D. Gesbert, E. G. Larsson, and J. Zhang, "Adaptive and robust signal processing in multi-user and multi-cellular environments (initial) D3.1a," tech. rep., SAPHYRE, 2011.

[2] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 528 – 541, March 2006.

[3] J. Belrose, "Reginald Aubrey Fessenden and the birth of wireless telephony," *IEEE Antennas and Propagation Magazine*, vol. 44, pp. 38 –47, April 2002.

[4] J. Cimini, L., "Analysis and simulation of a digital mobile channel using Orthogonal Frequency Division Multiplexing," *IEEE Transactions on Communications*, vol. 33, pp. 665 – 675, July 1985.

[5] M. Pereirasamy, J. Luo, M. Dillinger, and C. Hartmann, "Dynamic inter-operator spectrum sharing for UMTS FDD with displaced cellular networks," in *IEEE Wireless Communications and Networking Conference, 2005*, vol. 3, pp. 1720 – 1725, March 2005.

[6] M. Buddhikot, "Understanding dynamic spectrum access: Models,taxonomy and challenges," in *2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2007. DySPAN 2007*, pp. 649 –663, April 2007.

[7] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, pp. 79 –89, May 2007.

[8] D. Hatfield and P. Weiser, "Property rights in spectrum: taking the next step," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005*, pp. 43 –55, November 2005.

[9] W. Lehr and J. Crowcroft, "Managing shared access to a spectrum commons," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005*, pp. 420 –444, November 2005.

[10] P. Leaves, K. Moessner, R. Tafazolli, D. Grandblaise, D. Bourse, R. Tonjes, and M. Breveglieri, "Dynamic spectrum allocation in composite reconfigurable wireless networks," *IEEE Communications Magazine*, vol. 42, pp. 72 – 81, May 2004.

[11] T. Yamada, D. Burgkhardt, I. Cosovic, and F. K. Jondral, "Resource distribution approaches in spectrum sharing systems," *EURASIP Journal on Wireless Communications and Networking*, 2008.

[12] I. Mitola, J., "Cognitive radio for flexible mobile multimedia communications," in *IEEE International Workshop on Mobile Multimedia Communications, 1999. MoMuC 1999*, pp. 3 –10, 1999.

[13] Q. Zhao and A. Swami, "A survey of dynamic spectrum access: Signal processing and networking perspectives," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, vol. 4, pp. 1349–1352, April 2007.

[14] T. Weiss and F. Jondral, "Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency," *IEEE Communications Magazine*, vol. 42, pp. 8–14, March 2004.

[15] E. Jorswieck, L. Badia, T. Fahldieck, D. Gesbert, S. Gustafsson, M. Haardt, K.-M. Ho, E. Karipidis, A. Kortke, E. Larsson, H. Mark, M. Nawrocki, R. Piesiewicz, F. Rö andmer, M. Schubert, J. Sykora, P. Trommelen, B. van den Ende, and M. Zorzi, "Resource sharing in wireless networks: The SAPHYRE approach," in *Future Network and Mobile Summit, 2010*, pp. 1 –8, June 2010.

[16] G. Salami, A. Quddus, D. Thilakawardana, and R. Tafazolli, "Nonpool based spectrum sharing for two UMTS operators in the UMTS extension band," in *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008*, pp. 1 –5, September 2008.

[17] S. Kumar, G. Costa, S. Kant, F. Frederiksen, N. Marchetti, and P. Mogensen, "Spectrum sharing for next generation wireless communication networks," in *First International Workshop on Cognitive Radio and Advanced Spectrum Management, 2008. CogART 2008*, pp. 1 –5, February 2008.

[18] W. Wang, Y. Cui, T. Peng, and W. Wang, "Noncooperative power control game with exponential pricing for cognitive radio network," in *IEEE 65th Vehicular Technology Conference, 2007. VTC2007-Spring*, pp. 3125 –3129, April 2007.

[19] L. Grokop and D. Tse, "Spectrum sharing between wireless networks," in *The 27th Conference on Computer Communications. INFOCOM 2008*, pp. 201 –205, April 2008.

[20] Z. Ji and K. Liu, "Cognitive radios for dynamic spectrum access - dynamic spectrum sharing: A game theoretical overview," *IEEE Communications Magazine*, vol. 45, pp. 88 –94, May 2007.

[21] W. Wang, T. Peng, and W. Wang, "Optimal power control under interference temperature constraints in cognitive radio network," in *IEEE Wireless Communications and Networking Conference, 2007. WCNC 2007*, pp. 116 –120, March 2007.

[22] H. Islam, Y. chang Liang, and A. Hoang, "Joint power control and beamforming for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 2415 –2419, July 2008.

[23] S. Jayaweera and T. Li, "Dynamic spectrum leasing in cognitive radio networks via primary-secondary user power control games," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 3300 –3310, June 2009.

[24] D. Cabric and R. Brodersen, "Physical layer design issues unique to cognitive radio systems," in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005*, vol. 2, pp. 759 –763, September 2005.

[25] R. Zhang and Y.-C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, pp. 88 –102, Februay 2008.

[26] K. Phan, S. Vorobyov, N. Sidiropoulos, and C. Tellambura, "Spectrum sharing in wireless networks via QoS-aware secondary multicast beamforming," *IEEE Transactions on Signal Processing*, vol. 57, pp. 2323 –2335, June 2009.

[27] Z. Ka, M. Ho, and D. Gesbert, "Spectrum sharing in multiple-antenna channels: A distributed cooperative game theoretic approach," in *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008*, pp. 1 –5, September 2008.

[28] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," in *24th Annual Joint Conference of the IEEE Computer and Communications Societies. INFOCOM 2005. Proceedings*, vol. 3, pp. 1794 – 1803, March 2005.

[29] K. Kar, X. Luo, and S. Sarkar, "Throughput-optimal scheduling in multichannel access point networks under infrequent channel measurements," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 2619 –2629, July 2008.

[30] R. Louie, M. Mckay, and I. Collings, "Maximum sum-rate of MIMO multiuser scheduling with linear receivers," *IEEE Transactions on Communications*, vol. 57, pp. 3500 –3510, November 2009.

[31] S. Shakkottai and A. L. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: The exponential rule," tech. rep., University of Illinois, 2000.

[32] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "Scheduling in a queuing system with asynchronously varying service rates," *Probability in the Engineering and Informational Sciences*, vol. 18, no. 02, pp. 191–217, 2004.

[33] H. Kim, K. Kim, Y. Han, and S. Yun, "A proportional fair scheduling for multicarrier transmission systems," in *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, vol. 1, pp. 409 – 413, September 2004.

[34] J.-Y. le Boudec, "Rate adaptation, congestion control and fairness: A tutorial," tech. rep., Ecole Polytechnique Fédérale de Lausanne (EPFL), 2008.

[35] T. Bonald, L. Massoulié, A. Proutière, and J. Virtamo, "A queueing analysis of max-min fairness, proportional fairness and balanced fairness," *Queueing Systems*, vol. 53, pp. 65–84, 2006. 10.1007/s11134-006-7587-7.

[36] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *25th IEEE International Conference on Computer Communications. INFOCOM 2006. Proceedings*, pp. 1 –12, April 2006.

[37] S. Sesia, I. Toufik, and M. Baker, *LTE: The UMTS Long Term Evolution: From theory to practice*. John Wiley & Sons Ltd., 2009.

[38] HTTP archive, "June 15 2012 mobile web performance data (iphone), http://httparchive.org/downloads/httparchive_mobile_jun_15_2012.gz," August 2012.

[39] Akamai, "Mobitest - free mobile web performance measurement tool, http://mobitest.akamai.com/m/index.cgi," September 2012.

[40] R. Irmer, G. Liu, S. Xiadong, J. Kramer, S. Abeta, T. Salzer, E. Jacks, A. Buldorini, and G. Wannemacher, "Next Generation Mobile Networks radio access performance evaluation methodology," tech. rep., NGMN Alliance, January 2008.

[41] 3GPP, "3GPP TR 36.814: Further advancements for E-UTRA physical layer aspects (release 9)," tech. rep., 3rd Generation Partnership Project, March 2010.

[42] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyösti, D. Laurenson, G. Matz, A. F. Molisch, C. Oestges, and H. Özcelik, "Survey of channel and radio propagation models for wireless MIMO systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, p. 56, January 2007.

[43] M. Narandzic, C. Schneider, R. Thoma, T. Jamsa, P. Kyosti, and X. Zhao, "Comparison of SCM, SCME, and WINNER channel models," in *IEEE 65th Vehicular Technology Conference, 2007. VTC2007-Spring*, pp. 413 – 417, April 2007.

[44] P. Kyösti, J. Meinilä, L. Hentilä, X. Zhao, T. Jamsa, C. Schneider, M. Narandzic, M. Milojevic, A. Hong, J. Ylitalo, V.-M. Holappa, M. Alatossava, R. Bultitude, Y. d. Jong, and T. Rautiainen, "WINNER II channel models," Tech. Rep. IST-4-027756, WINNER, 2007.

[45] V. Abhayawardhana, I. Wassell, D. Crosby, M. Sellars, and M. Brown, "Comparison of empirical propagation path loss models for fixed wireless access systems," in *IEEE 61st Vehicular Technology Conference, 2005. VTC 2005*, vol. 1, pp. 73 – 77, May 2005.

[46] A. Neskovic, N. Neskovic, and G. Paunovic, "Modern approaches in modeling of mobile radio systems propagation environment," *IEEE Communications Surveys Tutorials*, vol. 3, no. 3, pp. 2 –12, 2000.

[47] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Transactions on Vehicular Technology*, vol. 29, pp. 317 – 325, August 1980.

[48] A. Medeisis and A. Kajackas, "On the use of the universal Okumura-Hata propagation prediction model in rural areas," in *IEEE 51st Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo*, vol. 3, pp. 1815 –1818, 2000.

[49] J. Milanovic, S. Rimac-Drlje, and K. Bejuk, "Comparison of propagation models accuracy for WiMAX on 3.5 GHz," in *14th IEEE International Conference on Electronics, Circuits and Systems, 2007. ICECS 2007*, pp. 111 –114, December 2007.

[50] E. Larsson, E. Jorswieck, J. Lindblom, and R. Mochaourab, "Game theory and the flat-fading Gaussian interference channel," *IEEE Signal Processing Magazine*, vol. 26, pp. 18 –27, September 2009.

[51] 3GPP, "3GPP TR 25.892: Feasibility study for Orthogonal Frequency Division Multiplexing (OFDM) for UTRAN enhancement (release 6)," tech. rep., 3rd Generation Partnership Project, June 2004.

[52] M. Döttling, "Assessment of advanced beamforming and MIMO technologies," Tech. Rep. IST-2003-507581, WINNER, 2005.

[53] Ericsson, "System-level evaluation of OFDM - further considerations," Tech. Rep. R1-031303, 3GPP, 2003.

[54] D. Huy, R. Legouable, D. Ktenas, L. Brunel, and M. Assaad, "Downlink B3G MIMO OFDMA link and system level performance," in *IEEE Vehicular Technology Conference, 2008. VTC Spring 2008.*, pp. 1975 –1979, May 2008.

[55] S. Catreux, V. Erceg, D. Gesbert, and J. Heath, R.W., "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Communications Magazine*, vol. 40, pp. 108 –115, June 2002.

[56] C. Mehlfuhrer, M. Wrulich, J. C. Ikuno, D. Bosanska, and M. Rupp, "Simulating the Long Term Evolution physical layer," in *17th European Signal Processing Conference. EUSIPCO 2009*, 2009.

[57] J. Ikuno, M. Wrulich, and M. Rupp, "System level simulation of LTE networks," in *IEEE 71st Vehicular Technology Conference, 2010. VTC 2010-Spring*, pp. 1 –5, May 2010.

[58] S. Schwarz, M. Wrulich, and M. Rupp, "Mutual information based calculation of the precoding matrix indicator for 3GPP UMTS/LTE," in *2010 International ITG Workshop on Smart Antennas (WSA)*, pp. 52 –58, February 2010.

[59] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Transactions on Information Theory*, vol. 52, pp. 2203 –2214, May 2006.

[60] S. Zhou and G. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over rayleigh MIMO channels?," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 1285 – 1294, July 2004.

[61] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," 1984.