THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES 1

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Informatique*

Par
Malo MANINI

Allocation de ressources et ordonnancement dans les réseaux de 5ème génération.

Thèse présentée et soutenue à Cesson-Sévigné, le 28 janvier 2021
Unité de recherche : IRISA

Rapporteurs avant soutenance :

Lila BOUKHATEM	Maître de conférence à l'Université Paris-Sud XI
Nadjib AIT SAADI	Professeur des universités à UVSQ Paris-Saclay

Composition du Jury :

Président :	Marceau COUPECHOUX	Professeur des universités à Télécom Paris
Examinateurs :	Lila BOUKHATEM	Maître de conférence à l'Université Paris-Sud XI
	Nadjib AIT SAADI	Professeur des universités à UVSQ Paris-Saclay
	Abderrezak RACHEDI	Professeur des universités à l'Université Paris-Est Marne-la-Vallée
Dir. de thèse :	Xavier LAGRANGE	Professeur des universités à IMT Atlantique
Encadrants :	Cédric GUEGUEN	Maître de conférence à l'Université de Rennes 1
	Rodolphe LEGOUABLE	Chercheur à l'IRT b<>com

Acknowledgement

The support and help from many people played a role in this thesis. This section is an opportunity to thank them.

Firstly, I would like to thank my supervisors Xavier Lagrange, Cédric Guéguen and Rodolphe Legouable who helped me during these last three years. Their knowledge and experience, from both the academic and industrial point of view, provided a beneficial balance. I am also gratefull for their patience and disponibility, even during difficult time.

Secondly, I would like to thank the members of the jury Lila Boukhatem, Nadjib Ait Saadi for reviewing the manuscript and providing relevant insight; as well as Marceau Coupechoux and Abderrezak Rachedi who took the time to assist the thesis defense. The exchange with all jury members was a real positive experience.

The thesis would not have been possible without the Institute of Research and Technology b<>com. For over 3 years, I learned a lot and discovered new interests in a lab full of friendly people, always willing to help and interact. A special mention to Luc and Matthieu, with whom I shared the same space and more, hopefully this will continue. During this time at b<>com, I also had the opportunity to rediscover Aikido with Phillipe in a privileged way.

This PhD is also a consecration of the years spent at the University of Rennes 1, with passionate professors and interesting projects. A special thanks to Isabelle Puaut and Cédric Guéguen who encouraged me to continue after my Master's degree.

I would also like to thank my friends, with whom I spend unforgettable times during these years. In particular, Napoun and Christophe who shared the same experience, being able to relate to those up and down was really comforting.

And finally, I also deeply thank my family who supported me in the most difficult moments, without ever failing, even if the subject could be a bit mysterious for them. And Noëmie, without whom, this thesis would not have been possible, with countless hours helping me and supporting me.

Résumé français

L'augmentation du nombre d'utilisateurs des réseaux sans-fil et la diversification de leurs usages amène à faire évoluer les méthodes de gestion des ressources. L'allocation de ressources consiste à distribuer les ressources disponibles de la station de base aux différents utilisateurs dans la zone de couverture du réseau. L'utilisation de techniques sans-fil induit des performances changeantes en fonction des propriétés du lien radio. Ces propriétés dépendent notamment de la position de l'utilisateur, des obstacles qui l'entoure ou encore des interférences venant d'autres stations ou utilisateurs. Combiné avec la variations des demandes en quantités de données, la variété des situations devient importante et la qualité des décisions d'allocation de ressources peut jouer un rôle déterminant. Cette thèse porte sur les stratégies d'allocation des ressources dans les réseaux sans-fils et profite des nouvelles techniques proposées par la 5G.

Dans les réseaux sans-fil classiques, les cellules sont considérées comme indépendantes et les utilisateurs ne peuvent interagir qu'avec une seule cellule. C'est un sujet de recherche où beaucoup de stratégies ont été proposées et ces solutions sont principalement axées sur des résolutions de problèmes spécifiques. Cependant les conditions rencontrées dans les réseaux sont rarement figées et peuvent dépendre de la période de la journée ou encore d'évènements spéciaux à proximité. Deux contraintes principales apparaissent dans les réseaux modernes. Premièrement, les besoins en débit sont grandissant et doivent s'adapter aux nouvelles applications. De plus, il est important de garantir un niveau de service équitable entre les différents utilisateurs afin de garantir un même niveau de satisfaction. Deuxièmement, les contraintes liées à la consommation énergétique doivent être prises en compte. Un réseau trop gourmand engendrerait des coûts importants à l'opérateur. De même pour l'utilisateur, plus l'impact est important sur l'autonomie de sa batterie, plus les sessions d'utilisation seront courtes. Dans ce contexte, nous proposons un algorithme d'allocation de ressources dont l'objectif est de garantir équitablement le meilleur service aux utilisateurs, tout en préservant leurs batteries. Lorsque la charge de la cellule est suffisamment faible pour garantir un service nécessaire, l'algorithme réoriente dynamiquement ses priorités vers l'économie d'énergie. Ce comportement permet un

compromis efficace entre la capacité et la consommation énergétique à différents niveaux de charge. Les évaluations de performances montrent que notre algorithme est au moins aussi efficace que les algorithmes dédiés dans leurs zone de prédilection. Il a cependant l'avantage de s'adapter lorsque la situation change contrairement aux algorithmes avec un objectif fixe.

Afin d'étendre la capacité des réseaux, l'ajout de nouvelles cellules permet d'élargir la bande passante et de réduire l'atténuation du signal liée à la distance. Grâce à des solutions permettant de centraliser le contrôle des stations de bases, les cellules peuvent être gérées simultanément pour optimiser les ressources du réseaux à une plus large échelle. Cette coordination permet une meilleure gestion des interférences et de la mobilité des utilisateurs. De plus, différentes techniques de réseau sans-fil peuvent être utilisées pour augmenter la densité tout en limitant les interférences, par exemple par l'ajout de micro-cellules dans des cellules plus large. Ces micro-cellules ont des propriétés propres, comme une zone de couverture plus restreinte ou l'utilisation de bandes plus hautes en fréquence. Ces cellules sont déployées principalement dans des zones avec une forte densité d'utilisateurs (stade de sport, aéroport, rue passagère...). Ce gain en flexibilité améliore les capacités et performances des réseaux, mais nécessite une adaptation des systèmes classiques. Nous présentons un algorithme de répartition des utilisateurs dans ce contexte multicellulaire qui intervient avant l'étape d'allocation de ressources. Cette répartition aura de fortes répercussions sur l'équilibre des charges des cellules et sur la qualité de service générale du système. En priorisant les allocations des micro-cellules, notre algorithme décharge au maximum la cellule principale, lui permettant ainsi de gérer les utilisateurs qui ne sont couverts que par celle-ci. Après cette étape, l'ensemble des algorithmes d'allocation de ressources des systèmes monocellulaires peuvent être utilisés et ainsi profiter de leurs performances et spécialités.

Bien que la densification des réseaux augmente fortement leurs capacités, le nombre de cellules dans un secteur est limité par sa géographie. Afin de encore améliorer les performances, l'amélioration des performances des stations elles-mêmes reste nécessaire. Une de ces solutions, le Massive-MIMO, permet d'accroître les fonctionnalités des cellules en permettant une directivité de l'énergie et ainsi ajoute la composante spatiale à l'allocation de ressources. Cette technique, appelée beamforming, augmente la distance de propagation et le débit, tout en réduisant les interférences. Le beamforming nécessite un accroissement du nombre d'antennes, ainsi qu'un traitement du signal complexe. Plusieurs faisceaux peuvent être formés dans des directions différentes en utilisant la même fréquence, permettant aux utilisateurs de recevoir des informations simultanément sur la même fréquence. Les utilisateurs partageant les mêmes ressources doivent être sélectionnés avec soin pour limiter leurs interférences mutuelles. Dans ce contexte particulier, nous proposons un nouvel indicateur de compatibilité spatiale des utilisateurs qui se base sur les allocations passées. Une fois intégré dans un algorithme d'allocation de ressources, il profite des capacités supérieures du Massive-MIMO et ses performances sont supérieures aux systèmes auxquels il est comparé. Cet indicateur pourrait être associé à d'autres métriques, notamment de qualité de service (QoS), pour répondre au mieux à des besoins spécifiques dans les réseaux modernes.

Dans tous les types de réseaux et particulièrement dans les réseaux sans-fil, l'allocation de ressource peut jouer un rôle déterminant sur les performances. L'allocation de ressource dans les réseaux classiques est un sujet très exploité dont les nouveaux systèmes devraient pouvoir bénéficier. Afin d'adapter les stratégies classiques dans ces nouveaux systèmes, il faut comprendre quels sont leurs indicateurs et leviers qui leurs sont propres. Les objectifs de cette thèse sont de d'abord présenter les systèmes classiques et les stratégies d'allocation de ressources qui y sont adaptées, pour ensuite proposer des solutions pour adapter ces stratégies aux futures techniques de réseaux sans-fils.

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivations

Network usage is continuously evolving towards a wireless access with a more nomad usage of the networks. Mobile devices are expected to be used by 70 percent of the population by 2023 [1], with up to 5.7 billion mobile subscribers. The network usage of those nomad users ranged from light social networks notifications to very high quality video streaming. Higher definition video formats such as 4K double the necessary bit rate compared to High Definition (HD) video. This new format is now commonly used on streaming platforms and 8K contents are now emerging [2]. The combination of an increased number of wireless users with larger bit rate requirements necessitates an adaptation or a change of existing wireless networks. These changes must be in line with the mobile nature of users. They rely on their batteries to transmit, receive and consume data, therefore energy-intensive networks lead to a low Quality of Experience (QoE). In addition to users' QoE, energy saving helps to reduce the impact of wireless networks on global warming. With more and more people using a smartphone to access the Internet [3], the impact of wireless communications is not negligible. To respond to these growing challenges, in this thesis, we explore different resource allocation strategies to increase the networks capacity while preserving the energy consumption

Resource allocation aims at distributing available network resources to reachable users. In wireless networks, those users are said to be inside the coverage area. The use of Radio Frequency (RF) induces unequal performance across the coverage area: users close to the Base Station (BS) have a higher achievable bit rate than users far from the BS, due to a loss increasing with the distance (pathloss). Combined with the fluctuation of the services' requested resources, the variety of user characteristics is very important. With this user diversity, the choice of the resource allocation strategy has a strong impact on the users' QoE.

The basic resource allocation system is a large single-cell (macro cell), where users can interact with only one BS in Single Input Single Output (SISO) configuration. This single cell is managed independently of the other cells in the network. This is a well known studied topic, where solutions aim at providing dedicated strategies for specific use cases. However, the traffic in wireless networks is not constant and fluctuates, depending on the time period of the day or special events nearby. Two major constraints can emerge from modern use of wireless networks. Firstly and foremost, the constraints such as the maximum delay and minimum data bit rate should be fulfilled according to the specific application requirements. To ensure that all users reach an equivalent level of QoE, they should be considered with fairness according to their needs and constraints. This is particularly challenging in crowded places or for edge users, where user disparity is strong. Secondly, another major constraint is the energy consumption, users being mainly nomad, battery life is crucial to maintain their services. With this basic wireless network design, additional macro cells must be added to increase the overall capacity. However, due to geographical constraints and interference between cells, the creation of new macro cell is very limited [4].

To allow a larger connectivity than a typical BS coverage area, wireless networks are composed of several BS, allowing to densify the network with the aim of capacity increase. Thanks to cloud based solutions (C-RAN) and their centralized management approach, cells are managed jointly to optimize resource allocation on a larger scale. This coordination allows a better inter-cell interference management, as seen in [5], and limit handovers [6]. In addition, different cell types can be used to increase network density, without raising the interference level, by adding micro cells within the coverage area of a macro cell. Those cell types have different properties, such as different ranges and/or different frequencies and are mostly deployed in very crowed places (stadium, airport, major streets...). This flexibility increases the wireless networks' capability and improves performance, at the expense of a new layer of complexity. On top of the traditional problem of resource allocation, new challenges have emerged, such as the balancing of cell loads. By using different types of cells, one may be more profitable than the others (better wireless conditions) and therefore be more chosen, resulting in unbalanced cell loads. This aspect is crucial and has to be considered during the resource allocation process.

Network densification can dramatically increase network capacity. However, depending on the geographical situation and in regard to the environmental issue, increasing the number of micro BS can reach a limit. A solution is to improve the BS's capability.

The Massive-Multiple Input Multiple Output (MIMO) technique allows the BS, by the use of several antennas and complex signal processing, to accurately focus the energy in space. This main feature, called beamforming, increases the range and bit rate while reducing interference. Several beams can be formed simultaneously on the same frequency, allowing different users to receive data on the same time and frequency resource. This technique called Multiple Users Multiple Input Multiple Output (MU-MIMO) greatly improves network performance without increasing the number of BSs or the energy spent. However, in some cases the transmissions interfer to each other and this decreasees the overall capacity. Users sharing the same resource need to be selected with care and most MU-MIMO resource allocation processes are based on the same indicator of compatibility: the channel matrix correlation. The MU-MIMO technique increases the BS performance and improves the system capacity while preserving the energy consumption and without increasing the number of BS deployed.

Regardless of the wireless network deployed or the BS characteristics, resource allocation plays an important role in system performance and users' Quality of Service (QoS) fairness. Traditional resource allocation schemes is a topic well covered and new systems should benefit from it. To adapt classical resource allocation strategies to different wireless network deployments, new algorithms should consider the right indicators and parameters.

1.2 Contributions and outline of the thesis

In this thesis, divided in 4 chapters, we propose resource allocation solutions for different system constraints. The main results are summarized in the following:

1.2.1 Spectral efficiency and energy efficiency dynamic trade-off for SISO system

Chapter 2 presents the general principles of resource allocation and the main issues encountered. Guaranteeing the QoS and the energy efficiency is essential in wireless networks, but those objectives are often opposed and rarely simultaneously considered by existing solutions. We present several classical resource allocation algorithms and their strategies for solving these problems. In a real scenario, networks are not always saturated and QoS requirements are easily met, allowing for more energy efficiency considerations. We propose a new resource allocation scheduler that dynamically balances energy and capacity according to system load and fulfills the gap between existing solutions. Our solution is compared to others in a performance evaluation to enlight its versatile capabilities. Main results show that our scheduler reaches the same system capacity as spectral efficiency oriented schedulers while saving energy when possible.

1.2.2 Resource allocation adaptation from single-cell to multi-cell systems

Chapter 3 enlarges the resource allocation description to a multi-cell context. Wireless networks were traditionally composed of several base stations managing their resources independently. In new system designs, those stations operate conjointly to reduce interference and optimize users' distribution [7, 8]. Those users' repartition can lead to unbalanced cells and furthermore have a strong impact on scheduler performance and policy. A first example is given on a naive approach of distributing users in two different cells as a basis of our solution. We then propose a pre-scheduler able to adapt existing single cell schedulers to multi-frequency multi-cell systems. The pre-scheduler distributes users to the most suitable cell. This distribution depends on the cells' capacity and prevents uneven loads. After this pre-scheduling stage, any scheduler can be used and retain its specific properties. The performance evaluation shows that our solution improves performances in all tested context.

1.2.3 Resource allocation indicator for MU-MIMO system

Chapter 4 introduces the specificities of the MU-MIMO technique and its strong impact on resource allocation due to the fact that several users share the same resource. The spatial allocation of MU-MIMO is a new paradigm for resource allocation, opening to a third dimension for resource allocation. Users sharing the same resource need to be associated carefully. MU-MIMO grouping strategies are often based on the users' channel correlation [9, 10]. While presenting some advantages, this indicator has significant limitations in terms of updating rate or the accuracy of losses. To overcome the limits of the existing strategies, we propose a strategy using a new indicator that is based on the past group allocations. This users' compatibility indicator is more stable in time with overall better system capacity.

1.2.4 Conclusion and perspectives

Chapter 5 summarizes the main contributions and provides perspectives for future work and the development of resource allocation.

Chapter 2

SISO RESOURCE ALLOCATION

2.1 Introduction and motivations

In this chapter we focus on resource allocation in Single Input Single Output (SISO) systems [1]. Many research efforts have been done in this field and solutions already exist to increase capacity, fairness or energy efficiency. The resource allocation process needs to conjugate between medium management of the physical layer and application constraints of higher layers. In this process, the strategy to choose the right user at the right time might be crucial for the system and frequency stability. These strategies, implemented in the schedulers, try to solve problems encountered in wireless networks. To measure the efficiency of a scheduler we can use different Key Performance Indicators (KPIs). Those metrics allow to evaluate different aspects of a network:

- Throughput capacity: the number of user bits that a system is able to manage in a given time. A higher value will lead to a stronger congestion resistance.
- Spectral efficiency: the number of user bits that the system is able to transmit on a Resource Unit (RU). Directly linked to the throughput capacity, a higher value means a better usage of system resources.
- Bandwidth usage ratio: the ratio between the number of RUs used and the total number of RUs made available by the access point. A ratio of 100 % indicates a congested system. Schedulers must reach this limit as late as possible: once it has been reached, users' Quality of Service (QoS) can no longer be guaranteed.
- Delay: the time that is needed for a packet to arrive at its destination. A value lower than the application constraints is required for the user to be satisfied.
- Energy efficiency: the consumption in mWatt for a bit. We can also look at the total energy consumption of a system. To reduce the impact of wireless networks

1. This work has been published in [11, 12]

on global warming or to extend the life of users' batteries, it is always preferable to minimize the energy consumption.

— User capacity limit with Quality of Experience (QoE) guarantee: the number of users the system can manage while guaranteeing a global system mean delay less than the application delay threshold.

— The QoE fairness index: the system ability to ensure fairness regarding users requirements. This one is calculated according the following Raj Jain method (Jain's fairness index) [13, 14, 15]:

$$J\left(user_1, user_2, ..., user_n\right) = \frac{\left(\sum\limits_{i=0}^{n} x_i\right)^2}{n\sum\limits_{i=0}^{n}\left(x_i\right)^2}, \tag{2.1}$$

where x_i represents the evaluated parameter, $x_i = Delay_i$ or $x_i = qoe_i$ depending of the case. As described in the section 3.3.3, QoE is determined with the mean ratio between the number of packets out of delay and the total number of packets.

By focusing on one of these problems, schedulers often neglect other aspects of the network. For example a scheduler, which focuses on the total throughput may not be energy efficient. Even if a trade-off exists, it still mainly focuses on one aspect and does not consider the system globally. Overall system energy and spectral efficiency are major issues in wireless network, achieving these objectives jointly seems very difficult and requires the usage of trade-offs. Moreover, depending on the context, the importance of the two objectives may differ. In an underloaded context, guaranteeing high QoS is easily achievable due to a large amount of available radio resources, so the focus should be put on energy rather than system throughput. On the other hand, in an overloaded context, the lack of available radio resources requires that scheduling algorithms focus on system capacity in order to preserve QoS. Since the major issue of the network is to satisfy users, in this specific case, energy consumption must become less important. Many specialized solutions that focus either on energy saving or throughput maximization have been proposed [16, 17, 18, 19, 20]. They provide high performance on their specific network traffic load context but are not optimized outside. Other solutions that proposed static trade-offs provide average performance but cannot be fully efficient in all scenarios.

We propose a dynamic trade-off between energy and throughput efficiency that adapts the scheduler priorities to the network context and particularly to the traffic load. Con-

n_s	number of subcarriers by RB
n_t	number of antennas at the eNB
n_r	number of antennas at the UE
K	total number of RB
$\mathcal{U}$	set of all active UEs

Table 2.1 – Main system parameters

sidering the context, the scheduler is able to adjust its behavior in order to maintain a high QoS while reducing the energy as much as possible. Performance evaluation shows that the proposed solution succeeds in minimizing the energy consumption better than the energy-focused schedulers in an underloaded context while being able to reach the same spectral efficiency as throughput-oriented schedulers in a highly loaded context. In this chapter we propose one solution called the Dynamic Trade-off (DT), which dynamically adjusts the trade-off between throughput and energy efficiency according to the bandwidth usage.

2.2 General scheduling description

Each evolved Node B (eNB) is equipped with n_t antennas. We thus assume that all eNBs have the same number of antennas. Similarly, each User Equipment (UE) has n_r antennas (same number for all UEs). For the sake of simplicity the system studied is single cell, with only one eNB. We consider that there are n_s subcarriers in one Resource Block (RB) and K is the total number of RB. Let $\mathcal{U}$ be the set of users attached to the eNB with an active session. In table 2.1 we can find the main system parameters.

At a given time, scheduling can be viewed as an indicator function: $\delta_{i,k}(t)$ where i the UE index is, k the RB index and $\delta_{i,k}(t) \in [0, \min(n_t, n_r)]$ gives the number of streams on resource k allocated to i.

We have the following constraint:

$$\sum_{i \in \mathcal{U}} \delta_{i,k}(t) \leq n_t, \ \forall k \in \{1, .., K\} \tag{2.2}$$

The potential number of usable resources is given by $n_t \times K$. We assume that there is no limitation regarding the number of RF chains, the processing capacity. To illustrate the previous equation, we can consider a number of UEs equal to $n_t \times K$, for example

100. In this case UEs will only have 1 RB among 100 in average per Time Slot (TS). N Note that in most cases, $\delta_{i,k}(t) = 0$: only a small subset of $\mathcal{U}$ is really served at a given slot t.

2.2.1 SISO channel model

In a SISO system $n_t = 1$ and $n_r = 1$, meaning that only one user i can use a resource k on a given TS. With no interference between users of the same cell, the user throughput is computed independently from other users. At each frame allocation, the scheduler computes the maximum number of bits $q_{i,k}$ that can be transmitted using a resource k if assigned to user i [21, 22], while keeping below its Bit Error Rate (BER) target ($BER_{target,i}$), for all i and all k:

$$q_{i,k} \leq \left\lfloor \log_2 \left(1 + \frac{3P \times T_s \times \left(\frac{1}{d_i}\right)^{\beta} \times \alpha_{i,k}^2}{2N_0 \left[\mathrm{erfc}^{-1}\left(\frac{BER_{target,i}}{2}\right)\right]^2} \right) \right\rfloor, \tag{2.3}$$

where P is the transmission power, N_0 is the spectral density of noise, T_s is the Orthogonal Frequency-Division Multiplexing (OFDM) symbol duration, d_i is the distance to the access point of the user i and $\alpha_{i,k}^2$ represents the flat fading experienced by this user on resource k. In the following, $\alpha_{i,k}$ is Rayleigh distributed with an expectation equal to unity. The exponent β corresponds to the experienced path loss and goes from 2 to 4 considering environment density level. Due to multi-path fading, the potential number of bits that a user can transmit on a RU will fluctuate around this value over the time. Measurements show that the channel stability set this value over a 50ms period[23]. Because of this stability, resource allocation decisions rarely change from one TS to another if they are only dependent on wireless conditions. Figure 2.1 shows an example of an horizontal allocation due to the channel stability. Indeed, once a user is selected to transmit on a RB, the decision is likely to remain the same until the wireless fading changes.

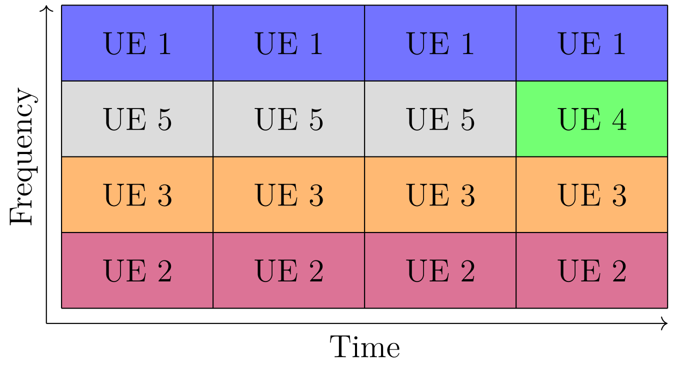


Figure 2.1 – Stability of resource allocation decisions

We further assume that the supported Quadrature Amplitude Modulation (QAM) modulation orders are limited such as q belongs to the set $S = \{0, 2, 4, \ldots, q_{max}\}$. Hence, the maximum number of bits $m_{i,k}$ that will be transmitted on a TS of resource k if this RU is allocated to the user i is:

$$m_{i,k} = \max\{q \in S, q \leq q_{i,k}\}. \tag{2.4}$$

2.3 State of The Art (SoTA) for SISO resource allocation

2.3.1 Non-opportunistic resource allocation scheme

A non-opportunistic scheduler does not take into account the radio condition of the users before allowing RB. It may lead to a case where the scheduler assigns a RB to a user that has a momentary low achievable throughput ($m_{i,k}$). However, a great advantage is that most of the time it is easier to compute and implement.

Round Robin (RR) [24, 25, 26] is one of the well known algorithm of this type. Even if there are different ways to implement it, the main principle remains the same. It successively attributes the RB to users. It attributes n units to a first user i, then n units to a second user $i+1$ and when it reaches the last user, it starts over to the first user i. By doing so the RR is fair regarding the fact that it distributes the same number of RB for all users. But this fairness is relatively low because users do not have the same mean achievable throughput, due to the losses, and therefore with the same number of RB they are not all able to transmit the same amount of data. Because it does not consider the radio conditions of the users, the spectral efficiency is close to the mean achievable throughput of the users.

Random Access (RA) [27] follows the same principle than RR and attributes the RB randomly to users that need to transmit data. In average, all users have the same number of resource units. It has the same level of fairness that RR. Depending on the platform, it might be easier to implement.

2.3.2 Opportunistic resource allocation scheme

Because they do not consider the radio conditions of users, non-opportunistic schedulers quickly reach their limits in a wireless context. To use the radio conditions' variability at their advantage, opportunistic schedulers add the $m_{i,k}$ of a user in the decision process. It will then prefer users with good radio conditions.

Throughput focus algorithm

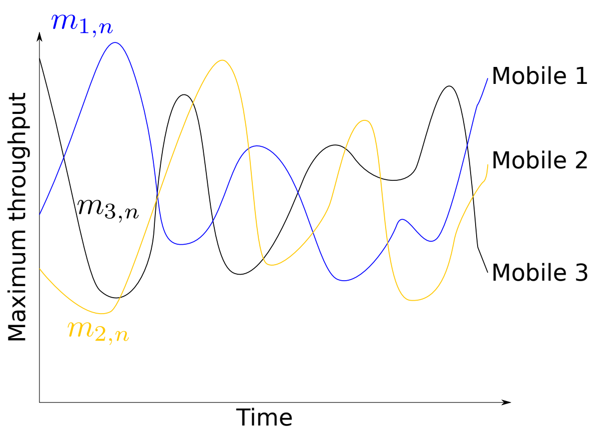
One of the main concerns in networks in general is the total throughput, or how much information a system is able to transmit.

Maximum Signal to Noise Ratio (MaxSNR) [16, 17] is the simplest opportunistic scheduler, which only focuses on throughput. The aim is to use radio conditions to attribute RB to the user with the best $m_{i,k}$. The user selection formula is :

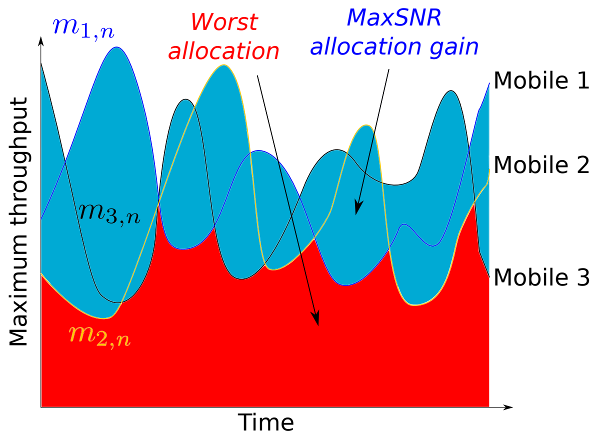
$$u = \text{argmax}(m_{i,k}) \tag{2.5}$$

By doing so, MaxSNR avoids all the bad allocations of a non opportunist algorithm (Fig. 2.2) and the spectral efficiency is higher than the mean $m_{i,k}$ of the users.

MaxSNR is particularly efficient when the number of users grows and outperforms RR. In fact, when the number of users is growing, the chance to obtain good conditions also grows. An opportunistic scheduler needs a great multi-user diversity to exploit its real potential. By always selecting the user with the best achievable throughput the MaxSNR can be really unfair. For example if we consider two users, one near the base station and one at the edge of the coverage area, the closest user will, most likely, receives the RBs before the far user. This results in a big difference regarding the QoS for those two users.



(a) Throughput/RU variation over the time due to multi-path fading.



(b) Opportunistic system capacity increase.

Figure 2.2 – Benefit of opportunistic scheduling strategies on spectral efficiency and system capacity.

Fairness oriented resource allocation scheme

As seen previously, focusing only on throughput may lead to unfair situations. To maintain a good QoS within a cell regardless of the users distance from the Base Station (BS), algorithms have to compensate the differences between users.

Proportional Fair (PF) [28, 29, 30, 31, 32] solves this issue by compensating the average obtained or achievable throughput. Indeed, users close to the access point have a better average throughput per RU than far users. This implies that, with MaxSNR scheduler, close users have statistically more chances to have access to the medium. In consequence, far users will often obtain radio resources after close users making them overpassing their QoS requirement and being unsatisfied. The basic principle of PF-based algorithms is to allocate resources to one specific user when its channel conditions are the most favorable with respect to its time-averaged conditions. The strategy is to calculate a mean value $M_{i,k}$ of the user throughput m_i, k. The calculation of $M_{i,k}$ depends on the implementation and there is a lot of variants of this algorithm. One way to do it is to calculate the mean number of bits per RB obtained by a user i over a certain period of the time. The user selection formula is then:

$$u = \text{argmax}(\frac{m_{i,k}}{M_{i,k}}) \tag{2.6}$$

This approach is more fair than MaxSNR since all users statistically have the same probability of accessing radio resources. Therefore, PF increases the benefits of multiuser diversity, which reinforce the opportunistic resource allocation behavior. By considering all users when their reach their higher achievable throughput, PF-based algorithms are more spectral efficient than MaxSNR-based algorithms.

FairMaxSNR [33] proposes to solve the fairness problem by focusing on user distance of the BS. FairMaxSNR compensates for the difference in average throughput between users by considering a compensation factor (CF_i), which decreases as the distance increases. This compensation factor gives users the same chance of accessing RBs regardless of their distance from from the BS.

$$u = \text{argmax}(\frac{m_{i,k}}{CF_i}) \tag{2.7}$$

Performance is comparable to the best designed schedulers based on PF.

2.3.3 Energy oriented resource allocation scheme

In order to offer more battery autonomy to users, solutions focusing on energy saving have been developed. With Power-based Proportional Fairness (PPF) [18], the authors propose a PF-based scheduler that avoids the inefficient allocations, with low Signal to Noise Ratio (SNR), and buffers transmissions that have high average energy consumption. This slightly increases energy efficiency since this gives access to the medium only to users with good SNR, allowing to always use higher modulation orders which are the most profitable, but potentially could segregate users with high traffic load (that will use more radio resources and consequently use more energy). In addition, the best way to minimize energy consumption is not only to optimize the modulation but mainly to maximize the sleep time. The Opportunistic Energy Aware (OEA) [19] is built on this principle. It exploits active-sleep mode and channel condition together. While other schedulers can potentially activate all users, the OEA limits this number. This allows to compress the transmission time (i.e. active mode), which is greedy in energy. Considering the channel condition in the allocation process, only allocations with high modulations are also conserved. T-MAC [20] is another strategy that can be considered as an extreme version of OEA. It only schedules a single user by TS, which strongly maximizes sleep time but, by losing multiuser diversity benefit, provides lower throughput. All these energy-specialized schedulers lack fairness and have limited spectral efficiency. Therefore this limits their scope of usage to underloaded context. Since energy efficiency guarantee must not evade QoS requirement and the system capacity optimization, new approaches must be developed in order to bring together: high spectral efficiency, fairness and energy consumption minimization whatever the considered traffic load.

2.4 Dynamic trade-off proposal

Even though system energy and spectral efficiency are major issues in wireless network, reaching these objectives conjointly seems very difficult and requires the usage of trade-offs. In under-loaded context, guaranteeing high QoS is easily achieved due to a large amount of available radio resources and the focus should be put on energy rather than on system throughput. At the opposite, in an overloaded context, the lack of available radio resources required that resource allocation algorithms focus on system capacity in order to preserve QoS. Since the major issue of the network is to satisfy users, in this specific case, energy consumption must become less important. As seen in section 2.3, many specialized solutions that focus either on energy saving or throughput maximization have been proposed. They provide high performance in their specific traffic load context, previously described, but are not optimized outside. Other solutions that proposed static trade-offs provide average performance, but cannot be fully efficient in all scenarios. To fill the gap between throughput and energy oriented schedulers, we propose a new algorithm called DT that dynamically adjust its behavior to the traffic load context. In an under-loaded system, radio resources are abundant and the system can easily satisfy all users. Consequently, in these contexts, DT detects the surplus of unused radio resources and orients its scheduling strategy to be energy aware. It makes a better usage of multiuser diversity than OEA: this allows to preserve more energy than this specialized algorithm even in the scope of usage of OEA. In an overloaded system, radio resources are highly valued and the system is experiencing great difficulties to satisfy all users. In these contexts, DT detects the lack of available radio resources and orients its scheduling strategy to increase the spectral efficiency in order to withstand the load increase. It offers the same system capacity than PF and outperforms MaxSNR. Between these two extreme contexts, DT takes into account the bandwidth usage ratio to smoothly adapt and adjust its energy-throughput trade-off to the traffic load. Performance evaluation shows that user are always satisfied with fairness as well as PF while always preserving as much energy as possible.

The DT scheduling algorithm relies on weights that set the dynamic priorities for allocating the radio resources. These weights are built in order to satisfy three major objectives that are explained separately in the following: system capacity maximization, fairness and energy consumption minimization. Then we present a calibration of the function that adds an ability for the scheduler to adequately tune the multiuser diversity usage considering the context and relative objectives, merging previous weights in a balanced DT solution.

2.4.1 System Throughput Maximization

The main KPI to consider in a resource-constrained network is capacity. As seen in section 2.3.2, MaxSNR-based schemes allocate the RU to the terminals that have the greatest $m_{i,k}$ (eq. 2.4) values. This strategy maximizes the system capacity at short time scale. To provide an opportunistic behavior to DT and to maximize its spectral efficient, DT is inspired by MaxSNR. However, it is highly unfair considering users far to the access point that are often treated outside their delay requirement. In order to provide more fairness considering users' locations while preserving the system throughput maximization, a fairness parameter is introduced in the DT formula.

2.4.2 Fairness guarantee

DT integrates in its scheduling process the fairness parameter proposed in [33]. Called Compensation Factor (CF) and denoted by CF_i, this parameter takes into account the current path loss impact on the average achievable bit rate of mobile i:

$$CF_i = \frac{b_{ref}}{b_i}. \tag{2.8}$$

b_{ref} is a reference number of bits that may be transmitted on a subcarrier considering a reference free space path loss a_{ref} for a reference distance d_{ref} to the access point and a multipath fading equal to unity:

$$b_{ref} = \log_2 \left(1 + \frac{3P_{max} \times T_s \times a_{ref}}{2N_0 \left[\mathrm{erfc}^{-1}\left(\frac{BER_{target}}{2}\right)\right]^2} \right). \tag{2.9}$$

b_i represents the same quantity, but considering a distance d_i to the access point:

$$b_i = \log_2 \left(1 + \frac{3P_{max} \times T_s \times a_{ref} \times \left(\frac{d_{ref}}{d_i}\right)^{\beta}}{2N_0 \left[\mathrm{erfc}^{-1}\left(\frac{BER_{target}}{2}\right)\right]^2} \right), \tag{2.10}$$

with β the experienced path loss exponent.

Adequately combining and taking into account both $m_{i,k}$ and CF_i in the allocation process ($m_{i,k} \times CF_i$), DT considers all mobiles virtually at the same position in the scheduling decision. CF_i adequately compensates the lower spectral efficiency of far mobiles bringing high fairness in the allocation process. An equal throughput can be provided for each mobile while keeping the MaxSNR opportunistic scheduling advantages thanks to the $m_{i,k}$ parameters, which take into account the Channel State Information (CSI). Moreover, in contrast with MaxSNR, which satisfies preferably the mobiles close to the access point, DT keeps more mobiles active but with a relatively low traffic backlog. Satisfaction of delay constraints is more uniform and, by better preserving the multiuser diversity, a more efficient usage of the bandwidth has been highlighted. This jointly ensures fairness and system throughput maximization. If two mobiles have an equal priority for a RU, one is given to the mobile that has the highest Buffer Occupancy (BO) in order to further strengthen fairness. At this step, DT optimizes the throughput and guarantees high fairness, but highly suffers of an inefficient energy management at a same level than PF. In order to provide energy consumption minimization while preserving the system throughput maximization and fairness, an energy parameter is introduced as explained in the next section.

2.4.3 Energy consumption minimization

The third major objective of the DT is to provide efficient energy management in addition to the system throughput optimization and fairness. Existing opportunistic resource mapping (as MaxSNR or PF for example) basically overexploit multiuser diversity, which induces horizontal allocation. Indeed, due to flat fading during a frame, often same user experienced the best channel condition on each TS of a given RB. Consequently, with classical opportunistic schedulers, same user often receives all the TS of a RB and needs to stay in active mode during a long time. We can potentially have as many selected users as the number of RB. Therefore, during all TSs, many selected users cannot be set in sleep mode. They consume a lot of power to transmit few bits during a long time (with many allocated TS but on few RBs).

The DT scheduler integrates a modified version of the energy efficient OEA solution [19], keeping its energy benefit without its fairness and system capacity failure. Energy consumption is minimized particularly by increasing the sleeping mode duration. In order to achieve this goal, DT extends the classical OEA opportunistic cross-layer design to obtain a new vertical opportunistic resource mapping.When a user is in active mode, DT tries, like OEA, to benefit from its activation in order to compress its time of activity and to transmit more bit per “used" TS. Like this, DT allows to significantly increase sleeping mode duration and energy preservation. Originally, OEA scheduler computed an “Energy Transmission Cost" (ETC_i) parameter (in Watt). It is based on the energy cost of user i to transmit on a RU:

$$ETC_i = A_i \times Cn_i + (1 - A_i) \times (C_i + Cn_i), \tag{2.11}$$

When the user k is in active mode, $A_i = 1$ otherwise $A_i = 0$ (i.e. sleep mode). In addition, Cn_i and C_i are two constants (in Watt). C_i represents the energy needed to wake up the user i from the sleep mode to the active mode. Cn_i represents the energy needed to transmit on a n^{th} allocated RB. The energy cost to transmit on the first RU (C_i) is higher than the cost to transmit on n^{th} (Cn_i) since the cost to move to sleep mode to active mode and transmit is greatly higher than just transmit some supplementary bits while user is already active.

ETC_i is used in OEA scheduler but has the negative side effect to highly reduce the exploitation of the multiuser diversity. This drastically and negatively impacts the OEA system capacity optimization. In order to keep its energy minimization properties while fixing this throughput issue, DT integrates a modified ETC_i parameter that we called “Throughput-Energy trade-off" parameter TET_i:

$$TET_i = A_i \times Cn_i + (1 - A_i) \times (\frac{C_i}{MD} + Cn_i), \tag{2.12}$$

where MD is a Multiuser Diversity factor. The higher MD is, the more the system increases the number of active users at the same time, intensifying the multiuser usage and jointly the global system throughput at the expense of the energy consumption (infinite MD value makes TET_i constant and induces DT similar to a PF resource allocation). At the opposite, low MD value makes DT decreasing the number of active users at the same time, reducing energy consumption at the expense of the multiuser diversity usage that provides a resource allocation close to OEA scheduling (excepting that this version is strongly more fair due to section 2.4.2).

2.4.4 DT merging of priorities

The DT scheduler allocates the radio resource k to the mobile i that has the greatest $DT_{i,k}$ value where:

$$DT_{i,k} = \frac{m_{i,k} \times CF_i}{TET_i} \tag{2.13}$$

$$DT_{i,k} = \frac{m_{i,k} \times CF_i}{A_i \times Cn_i + (1 - A_i) \times \left(\frac{C_i}{MD} + Cn_i\right)} \tag{2.14}$$

Taking into account $m_{i,k}$ allows to optimize system capacity, avoiding unprofitable radio resource allocation, CF_i allows to stay fair in the allocation process regarding user location and the other parameter allows to fight versus energy waste. Particularly, by adjusting the multiuser diversity usage thanks to good function of MD, DT could select the minimum number of users per TS to have a good energy efficiency while respecting the QoS requirements. However, when the system is more loaded, DT could increase the multiuser diversity thanks to an higher value of MD in order to obtain a better spectral efficiency to support the load.

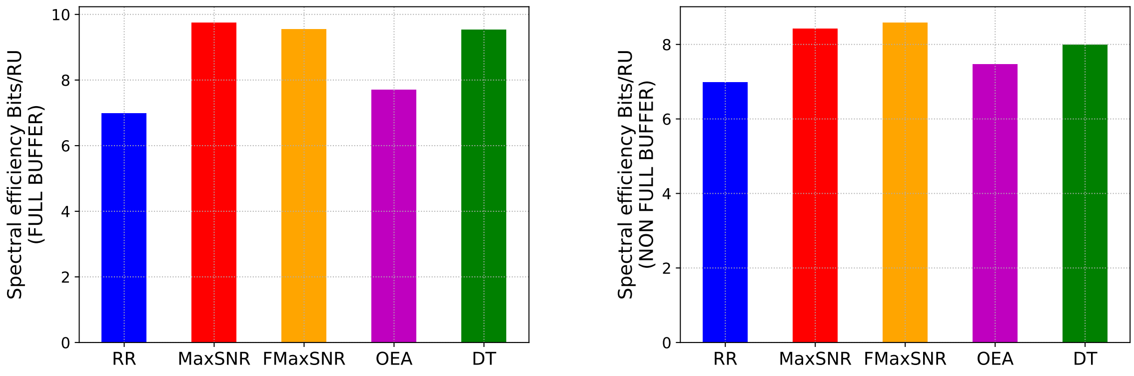
2.5 Simulation setup

2.5.1 Simulator

The simulations are computed using our own simulator written in C++. The main reason for this choice is to be able to run only the necessary features to accelerate the simulations. Indeed, to simulate 16 minutes of real-time, it takes 1 million iterations of the main loop which corresponds to 100 hours on an 8 core computer using multi-threads processing. To ensure our simulator reliability, we validated our simulations using results from the state-of-the-art.

2.5.2 Traffic model

An important aspect of a wireless network simulation is the users' traffic. In fact, most of the simulations are done in full buffer, which means that all users have always something to transmit. Using this kind of simulations, it is not possible to observe the delay of any packet, because there is an infinite number of packets waiting to be transmitted. This



(a) Full buffer traffic.

(b) Non full buffer traffic.

Figure 2.3 – Spectral efficiency in full buffer and in non full buffer traffics

situation corresponds to a particular case where the system is overloaded where normal users would leave the network. This scenario can be used when trying to optimize the physical layer or the overall capacity, but not the QoE of the users. A non full buffer system allows us to access information like delay, jitters or buffer occupancy, which are useful to accurately determine the QoS of our algorithm. This allows us to change the load from an unsaturated to a saturated system and see how schedulers adapt in all these situations.

In addition, a full buffer system may lead to false interpretation. For example, MaxSNR is very efficient (regarding the total throughput) in a full buffer system, but not in the general case where FairMaxSNR is more efficient (Figure 2.3). MaxSNR considers only close users at first and then far users, while the FairMaxSNR considers everyone in the same time, increasing the multi-user diversity. The result is the opposite in full buffer, because MaxSNR can only choose close users due to their infinite buffers and has better spectral efficiency than FairMaxSNR which still has to manage both groups.

There are several ways to design the generation of packets. One way is to do a Constant Bit Rate (CBR), which means that all users create a fixed number of packets at each frame. It has the advantage of being easy to build, and the results can be obtained in a few frames. Indeed, when the scheduler can manage the volume of bits once, it will be able to manage it every time if nothing changes.

But to be more realistic and make things more complex for the scheduler, we chose to implement a Variable Bit Rate (VBR). The traffic source is a multi-phase process. At the beginning of each phase, a data block is generated. The size of the block is an exponential random variable, allowing very high burst. The duration is n times a frame duration, where n is a geometrically distributed random integer. This type of generation is a good compromise between CBR and very complicated traffic models[34] regarding simulation time and realism needs.

2.5.3 User deployment

To evaluate the fairness of the schedulers, users are divided into two groups: close and far users. Users inside the same group are placed at the same distance from the BS. Those distances are set to respectively obtain a $m_{i,k}$ of 8 and 6 for close and far users. Each user application requires 560 kbits/s in mean. The 100% of bandwidth usage is reached around 20 users, using RR and varies depending on schedulers' spectral efficiency. To keep the system stable, users are added two by two, one close and one far. By doing so, there is always the same number of users in both groups.

2.5.4 Frame design

Our simulation uses a scaled down LTE frame. The size of a RB is one subcarrier over 7 symbols, while the frame is divided in 32 RB over 50 TS. This configuration is based on one configuration encountered in the literature[19]. In the 4G LTE, a frame has a duration of 10ms and is divided in 20 time slots that are divided in 7 symbols. The bandwidth can be divided up to 100 RBs with 100×12 subcarriers (for a 20 MHz bandwidth in downlink). A smaller system, like the one we use, reduces the simulation time without changing the performance differences between schedulers. Figure 2.4 shows the spectral efficiency difference between two simulations computed with our simulator: one using the LTE parameters and one using our scaled down parameters. Indeed, even if the gap between two schedulers may change with two different configurations, the order between them will not. A more spectral efficient scheduler will also be more efficient in a standard LTE configuration.

2.6 Study of the Multiuser Diversity factor

The multiuser diversity has an important impact on the load resistance and on the energy consumption. Finding an efficient way to adapt its usage to the context thanks to

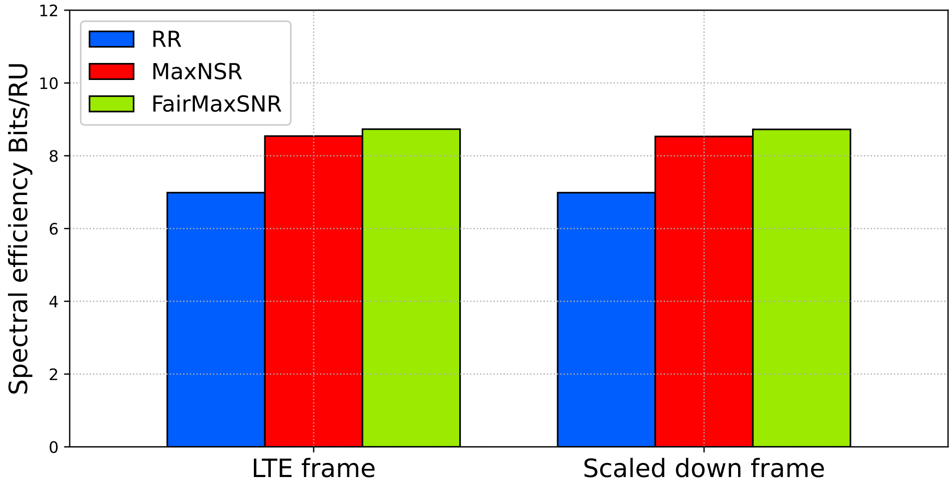


Figure 2.4 – Spectral efficiency compared between LTE frame and scaled down frame

a well tuned MD factor is challenging:

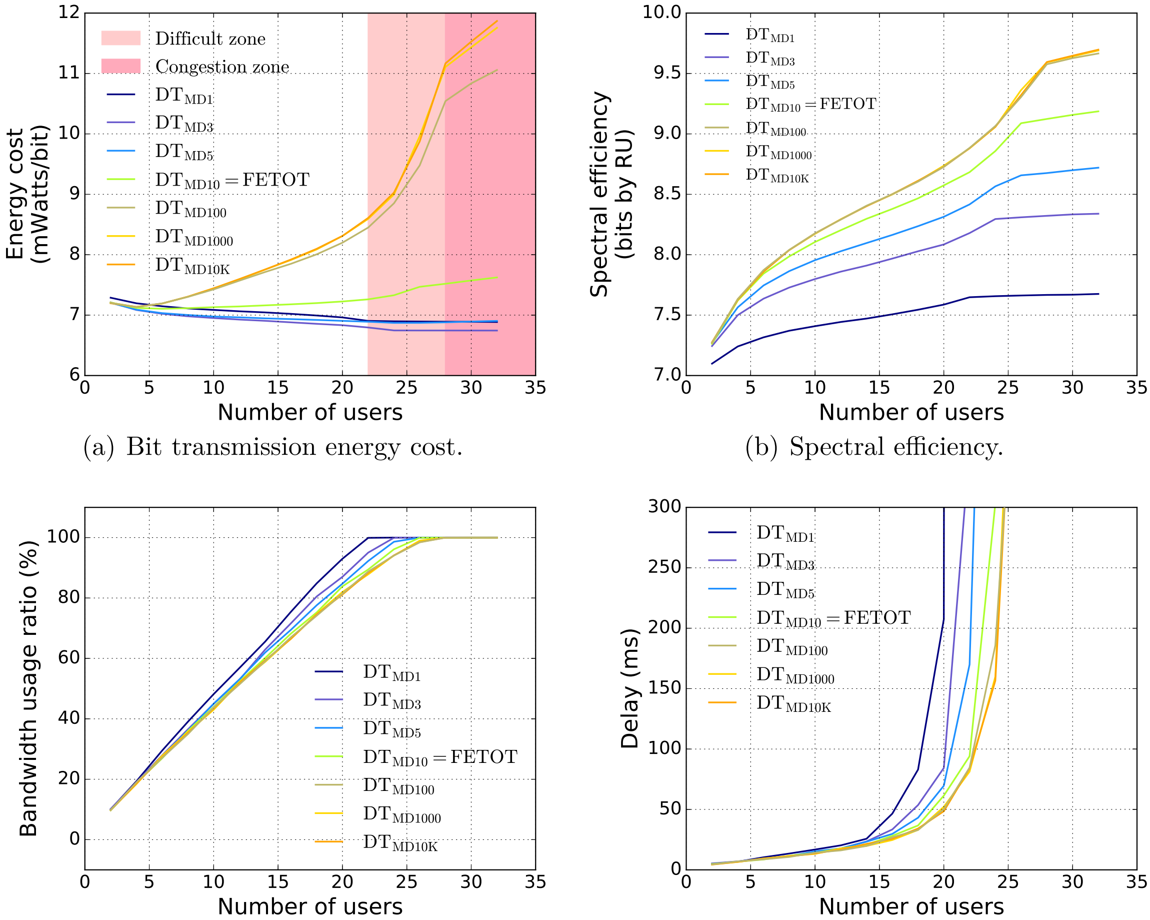
— The first step is to determine the extreme values of the MD which correspond to the best extreme configurations : when the system is clearly underloaded, the only concern is the energy consumption, when the system is largely overloaded, the main focus has to be on the QoS requirements.

— The second step is to find a smooth and adequate transition between those two extreme values based on adapted inputs.

2.6.1 Static MD value proposition

In a previous extensive study we found that $MD = 10$ provides a very efficient static trade-off between energy consumption minimization and spectral efficiency. This study had lead to a proposition of a new scheduler called Fairness-Energy-Throughput Optimized Trade-off (FETOT) [2]. FETOT uses a static MD factor fixed at 10. It allowed to make an adequate usage of the multiuser diversity in order to provide, the same system capacity than MaxSNR, same Fairness than PF and an energy minimization very close to the OEA results. However, we are convinced that the usage of a static MD value is not optimal. Even if FETOT provides a very good static overall trade-off, this can be highly improved with a solution able to adapt and tune the MD (and therefore the trade-off

2. This work was part of a prior master degree internship and has been published in [11]



(a) Bit transmission energy cost.

(b) Spectral efficiency.

(c) Percentage of RU used.

(d) Mean packet delay.

Figure 2.5 – System capacity and spectral efficiency study obtained with different static MD values.

settings) to the network traffic load context. Indeed, in very low traffic load context, energy minimization must be the only objective. With the increase of the traffic load, more attention must be done on spectral efficiency in adequate trade-off. In high traffic load, to improve spectral efficiency becomes the primary goal in order to continue to satisfy users and energy minimization priority must be relegated. The main contribution of this chapter is to propose a new scheduler that combines all previously described parameters and use a dynamic MD parameter to adapt priority to the context.

2.6.2 Study of extreme static MD values

Figure 2.5 shows the performance of preliminary versions of DT using static values of MD factor. For different traffic loads it shows the energy transmission cost per bit (Fig. 2.5(a)), the spectral efficiency (Fig. 2.5(b)), the bandwidth usage ratio (Fig. 2.5(c)) and the packet delay (Fig. 2.5(d)). The bandwidth usage ratio is the number of allocated RU divided by the total number of radio RUs in the system, in average, per frame.

In a non-congested system (i.e. when delay and bandwidth usage ratio are low, here with a number of users lower than 15 users), the focus should exclusively be put on the energy efficiency. As we can see in figure 2.5(a) a too high value of MD (> 10) induces excessive consumption due to several users simultaneously active on same TSs. However, choosing the smallest value is not a good option either. Indeed, if the MD is too small, opportunistic behavior is drastically reduced, and the spectral efficiency (Fig. 2.5(b)) is not good enough to evacuate the necessary amount of information in a short time. Even if the scheduler could appear to be more energy efficient due to a drastically limited number of active users at a same time, it is not at the long time scale since users will transmit during longer periods due to very low spectral efficiency. Consequently, in extreme and very low loaded context, $MD = 3$ seems to be the most adequate value in order to reach the minimization energy consumption objective (Fig. 2.5(a)).

In a congested system (i.e. when delay is high, bandwidth usage ratio very close or equal to 100 %, here with a number of users is over 20 users), the focus should exclusively be put on the spectral efficiency since the priority is to maintain a good level of QoS. Concerning bandwidth usage ratio, figure 2.5(c) underlines that all MD values greater than or equal to 100 allow to better withstand extreme traffic loads providing same spectral efficiency (Fig. 2.5(b)) and best delays (Fig. 2.5(d)). However, having a look at the energy efficiency (Fig. 2.5(a)), we notice a slight advantage to $MD = 100$ over larger values. In conclusion, a MD value around 100 is the most adequate values in this extremely high loaded context.

2.6.3 Dynamic MD function calibration

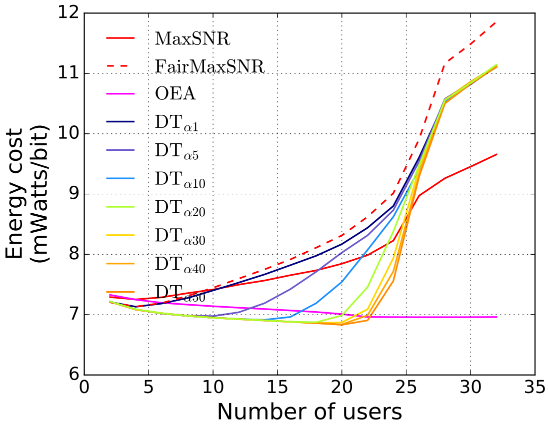
Originally in preliminary works (FETOT is described in section 2.6.1), we have shown that a fixed MD value set at 10 could represent an average good trade-off. However, it is not the best suitable solution for extreme cases as shown above. An adaptive solution can be developed to outperform FETOT in those situations with a dynamic usage of multiuser diversity that can be obtained thanks to a dynamic MD according to the context and particularly to the traffic load (that should define the scheduler priorities/goals). We propose in DT to define MD as an increasing function of the bandwidth usage ratio. This parameter simply and accurately informs about the state of the system and on the difficulties or not for the scheduler to maintain the QoS to the user. Low bandwidth usage ratio values, inducing low MD value (MD=3), underlines to DT to focus on energy. High bandwidth usage ratio, which required to focus on spectral efficiency, will induce high MD value (MD around 100) that will improve multiuser diversity usage. In order to link these two extreme limits, we proposed the intuition-based formula:

$$MD_x = C + \beta x^{\alpha} \tag{2.15}$$

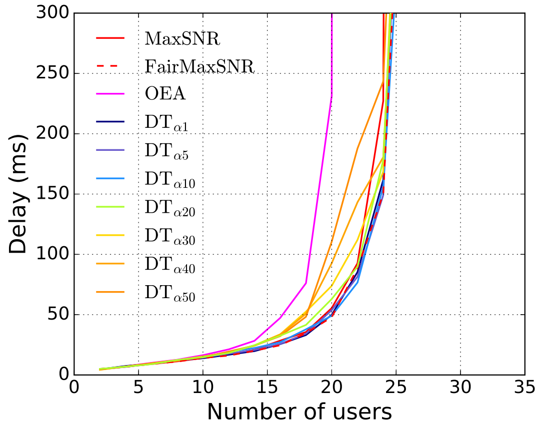
where x is the bandwidth usage ratio, C is a constant that defined the starting value of the MD function when the system is underloaded and β corresponds to the other extreme when the system is overloaded. In the following, we set C to 3 and β to 100 according to the results obtained is section 2.6.2. The parameter α allows to set the reactivity of the function to the traffic load variation. An appropriate calibration of α is highly important and is studied hereafter.

2.7 Study of the reactivity parameter α

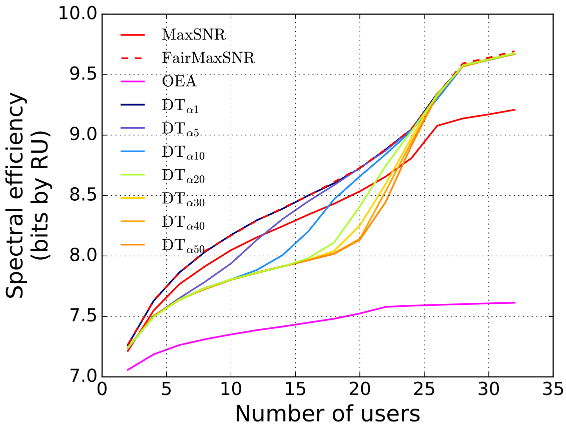
It is important that the MD function gives low values when the bandwidth usage ratio is low. Since QoS is easily guaranteed, DT has to limit the multiuser diversity usage in order to focus on energy consumption minimization. When the traffic load increases, the MD function must increase its output in adequacy with the difficulties met by the scheduler to conserve high QoS. Figure 2.7 represents MD variation depending on traffic load (measured with the bandwidth usage ratio) for different value of α. As we can notice in this figure, the α parameter directly impacts how this MD value will increase from the traffic load. If α is set equal to 1, the MD function is linear and multiuser diversity usage will be constantly increased with the bandwidth usage ratio. It is no optimal since no QoS difficulties are met with low bandwidth usage ratio values and problems are experienced only when we they come closer to 100%. At the opposite, a high value of α (typically $\alpha = 40$) make MD function growing too late in order to satisfy the QoS. Indeed, in realistic scenarios, with the variability of the traffic, even with an average measurable



(a) Bit transmission energy cost.



(b) Mean packet delay.



(c) Spectral efficiency.

Figure 2.6 – Study of α on system capacity and delay.

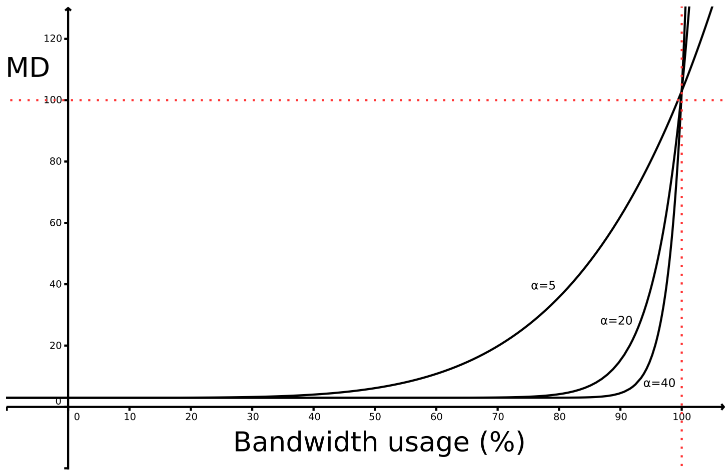
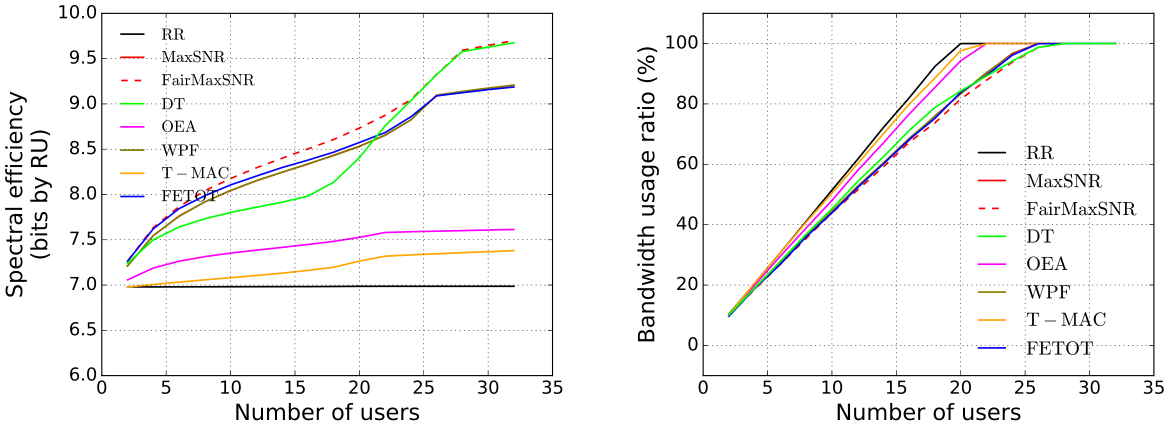


Figure 2.7 – Variation of MD according to α value.

bandwidth usage ratio inferior to 100% but close to this limit, temporary short term congestion can occur decreasing QoS. In these cases multiuser diversity usage must be intensified and this can be done by the DT scheduler if MD function is well calibrated. Detect when MD function must begin to grow is a difficult task and relies on the elasticity of the traffic. In order to define the best value of α, we decide to evaluate all possible α values performance in extensive simulations (Fig. 2.6).

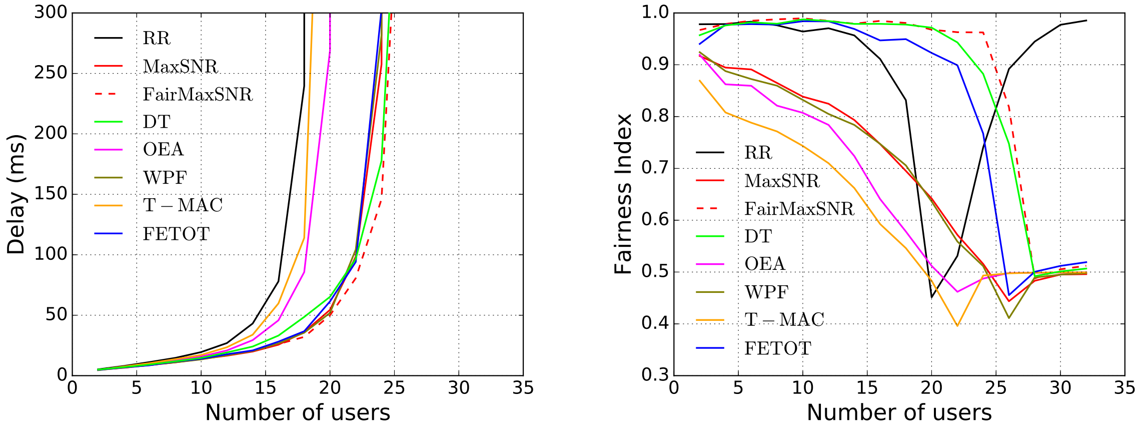
Figure 2.6(a) shows the impact of α regarding the energy efficiency. It is the most important objective for all the left parts of the figure since delay values are very low (Fig. 2.6(b)). Choosing a small α value such as 1 has a very bad impact on the energy consumption that increases quickly. This is due to the fact that the algorithm is too much reactive on the traffic load increase, uselessly exploits a supplementary of multiuser diversity and futilely tries to focus on the QoS. Indeed, very good value of delay is obtained for all α values inferior or equal to 20 (Fig. 2.6(b)). Higher values than 20 increase MD too late and provide worst delay, whereas lower value provides same delay but more energy consumption. If we consider that a final goal is to be able to maintain the best QoS while minimizing energy as much as possible, the most suitable α value for the MD function is equal to 20.



(a) Spectral efficiency.

(b) Percentage of RU used.

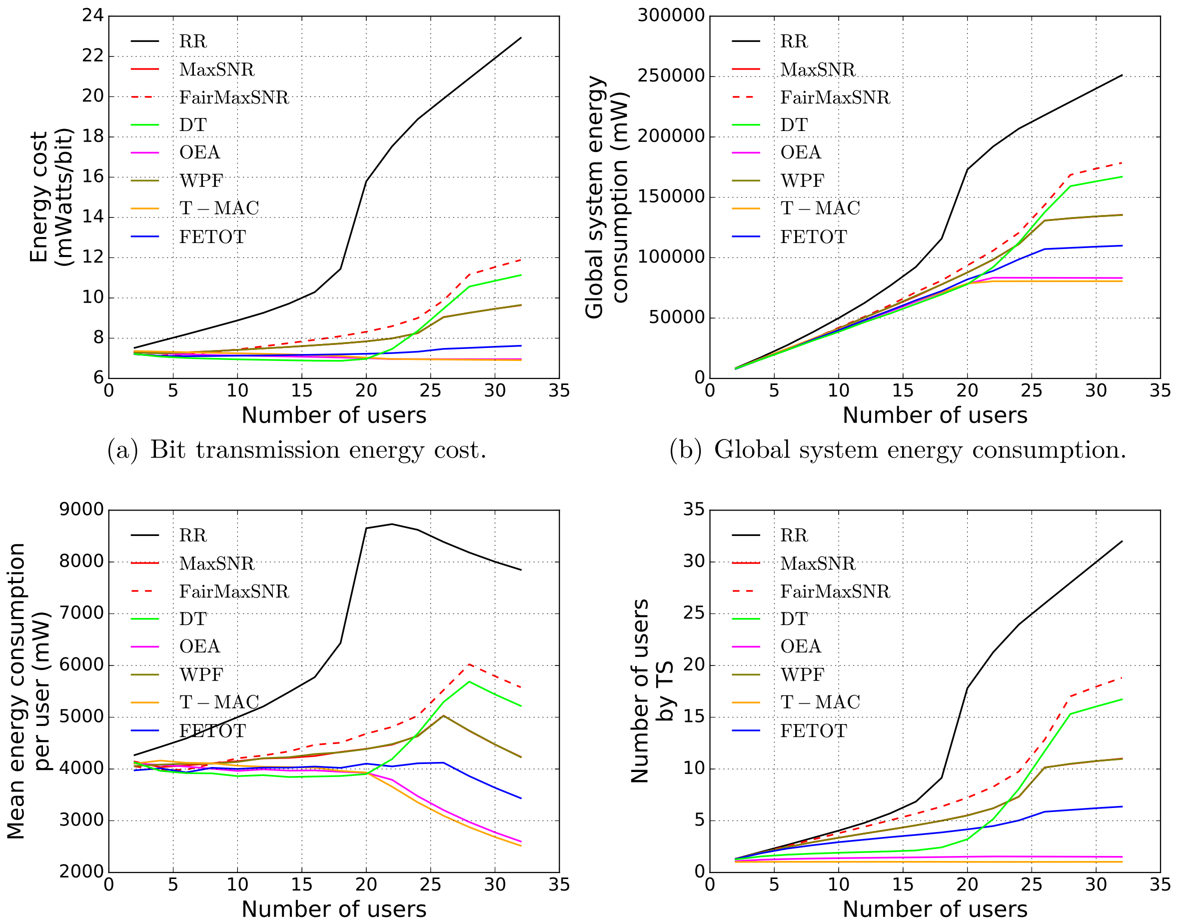
Figure 2.8 – Schedulers system capacity study.



(a) Mean packet delay.

(b) Jain's fairness index.

Figure 2.9 – Schedulers abilities to guarantee high QoS.



(a) Bit transmission energy cost.

(b) Global system energy consumption.

(c) Mean users energy consumption.

(d) Mean number of simultaneously active users (by Time Slot).

Figure 2.10 – Scheduler energy efficiency.

2.8 Performance evaluations

2.8.1 Context and simulation setup

Performance evaluation results are obtained using discrete event simulations. In the simulations, we assume C_i and Cn_i fixed respectively equal to 110.2 mW and 46.8 mW, for all i in accordance with measured hardware consumption[19]. The BER target is taken equal to 10^{-3}. We also consider that all users run VBR applications that generates a high volume of data with high sporadicity and require tight delay constraints, which substantially complicates the task of the scheduler. In order to study the influence of the distance of users on the scheduling performance, a first half of mobiles are situated close to the BS and have a mean $m_{i,k}$ equal to 8 bits. The second half are farthest from the BS such as their mean $m_{i,k}$ equal to 6 bits. All performance criteria are done by studying the influence of the traffic load. This one varies adding users 2 by 2 (each time, 1 close user and 1 far user). Simulation parameters are summarized in Table 2.2 and are described in section 2.5.

	Close users	**Far users**
Mean $m_{i,k}$	8 bits	6 bits
Throughput requirements	560 kbits/s	
Traffic model	VBR	
C_i	110.2 mW	
Cn_i	46.8 mW	

Base Station	
Number of RB	32
Number of TS	50

Table 2.2 – Simulation parameters

2.8.2 Spectral efficiency and throughput

Figure 2.8(a) shows the spectral efficiency obtained with each scheduler for different traffic load in the system. Since RR does not take into account radio conditions and therefore is not opportunistic, it does not take any advantage of multiuser diversity and its spectral efficiency is constant and low. SoTA energy focused schedulers (T-MAC, OEA), drastically limit the usage of the multiuser diversity in their allocation process offering slightly better results. On the contrary MaxSNR, highly opportunist provides a large gain. However, as explained in section 2.4.1, MaxSNR has a lack of fairness and is not able to take all the benefits of the multiuser diversity and is highly outperformed by PF. FETOT makes a trade-off between energy and throughput and provides spectral efficiency results close to MaxSNR.

Thanks to its dynamic MD parameter based on the bandwidth usage ratio, DT has lower spectral efficiency in low traffic load context using a moderate usage of the multiuser diversity focusing its efforts on energy. However, when it becomes necessary, and while energy specialized scheduler approach congestion (Fig. 2.8(b)), its MD factor adequately increases and raises the DT usage of the multiuser diversity, improving the spectral efficiency at the same level than PF reaching the same overall maximum system capacity (Fig. 2.8(b)).

2.8.3 Delay and fairness

A major QoS key performance indicator is the latency. Figure 2.9(a) represents the mean packet delay experienced in the system in milliseconds according to the number of users showing the traffic load. We can notice that 2 groups emerged:

— First, RR, T-MAC and OEA that have the worst results. Having a low spectral efficiency (Fig. 2.8(a)), they failed to support a large amount of traffic load with good QoS.

— Secondary, MaxSNR, PF, FETOT and DT are able to better sustain higher load increase with acceptable delay.

Figure 2.9(b) focus on fairness computed thanks to the jain's fairness index applied on mean packet delay, With $\mathcal{U}$ defined as the current number of users in the system. Jain's fairness index can vary between $\frac{1}{i}$ and 1 respectively associated to the most unfair scheduling with the most fair. T-MAC, OEA and MaxSNR significantly penalize user far from the BS and have decreasingly fairness results with the traffic load increase. On the contrary more fair solutions as RR, PF, FETOT and DT achieve to reach a high fairness index. Note that after congestion fairness cannot be guaranteed since the global mean packet delay is infinite. Consequently, the capacity of these schedulers to maintain high fairness is directly related to their spectral efficiency and when they have no longer available radio resources, fairness disappeared. Note that when the system capacity is highly overpassed, jain's fairness index can increase due to comparison of close but unacceptable huge value of delay. Respectively the most fair schedulers are: PF, DT, FETOT, RR, MaxSNR, OEA, T-MAC.

2.8.4 Energy consumption

Figure 2.10 shows the abilities of each scheduler to be energy efficient. RR widely provides the worst results. This is due to its non opportunistic behavior that makes possible highly inefficient resource allocation in term of bit per RU and therefore a significant energy and RU wastes. In addition, due to a cycling user selection, many users can be simultaneously activated (Fig. 2.10(d)) increasing again the energy waste since more users pay a higher transmission activation price C_i. With more than 20 users, the system is overloaded and RR fails to provide the sufficient amount of RUs required by each user. They are often forced to stay in sleep mode even with data to transmit due to the lack of RUs. More often forced into sleep mode, the users consumed less energy over the time. This explains why with more than 20 users, the RR curve decreases (Fig. 2.10(c)). Limiting the usage of the multiuser diversity to a low value whatever the context (Fig. 2.10(d)), T-MAC and OEA provide very good energy consumption (Fig. 2.10(c)). Note that these good results must be put into perspectives. Indeed, those solutions continue to search to minimize energy consumption, even when the traffic loads increase and this stubborn behavior conducts these schedulers to quickly reach congestion (Fig. 2.8(b)) with high delay (Fig. 2.9(a)). At the opposite, PF, fully exploiting the multiuser diversity (Fig. 2.10(d)), consumes more energy (Fig. 2.10(c)) but less than RR thanks to strongly better spectral efficiency. Focusing on MaxSNR, its energy results are slightly better than PF. Indeed, this scheduler has a tendency to segregate a part of users (far from the BS) and therefore obtains reduced benefits of multiuser diversity usage. This is a weakness in order to improve spectral efficiency, but an advantage to increase user sleep duration. FETOT provides better energy efficiency than MaxSNR, very close to the OEA, when the traffic load is low (below 20 users). Using an adequate static trade-off, energy consumption stays reasonable, even when traffic reaches higher values but, when necessary and contrary to the OEA, this is less done at the expense of spectral efficiency that stay closer than MaxSNR (Fig. 2.8(a)).

Considering underloaded contexts (number of users inferior to 20), guaranteeing high QoS is easily achievable by DT (Fig. 2.9(a)) due to a large surplus of available radio RUs (Fig. 2.8(b)) and focus should be put on energy rather than system throughput. Figures 2.10(a), 2.10(b) and 2.10(c) underline that DT is the scheduler that better optimizes the multiuser diversity usage in this context. Few users are simultaneously activated per TS (close to T-Mac and OEA (Fig 2.10(d))) but, contrary to the specialized SoTA energy aware schedulers, DT provides an adequate spectral efficiency forbidden inefficient resource allocation. This combination allows to better compress the transmission time and therefore better optimize energy consumption. Considering a highly loaded context (number of users greater than 20), the lack of available radio resources (Fig. 2.8(b)) required that schedulers focus on system capacity in order to preserve QoS: energy consumption must become a lower priority. In this context, DT behavior slightly sacrifices energy in order to sustain the network viability and then favors high spectral efficiency that reach values closer than PF (Fig. 2.8(a)), which provides acceptable delay as long as possible (close to PF).

2.9 Conclusion on SISO resource allocation

Reaching both low system energy consumption and high spectral efficiency are very difficult tasks in a wireless network. Specialized solutions as MaxSNR, PF or T-MAC have been designed to well answer one of these criteria failing for the second. Other solutions propose static trade-offs that provide better average results on these two metrics without outperforming specialized schedulers in their focused domain. In this chapter, we underline that the network objectives must be dependent on the context and particularly to the traffic load. In underloaded context, guaranteeing high QoS is easily achieved due to a large surplus of available radio resources and the focus must be put on energy rather than system throughput. At the opposite, in a high traffic loaded context, the lack of available radio resources required that resource allocation algorithms focus on system capacity in order to preserve QoS, satisfy users, thus energy consumption must become less important. The main contribution of this chapter is to propose a DT scheduler able to tune its priorities, by adjusting the multiuser usage, according to the network traffic load context. It provides a better energy efficiency than specialized energy aware scheduler when it is feasible while providing the same spectral efficiency and delays than throughput oriented scheduler when it is required. This is achieved thanks to a special focus on fairness, which is also guaranteed. Surpassing specialized schedulers both in their preferred context and in their ability to adapt, the main limitation of DT is wireless network technology. In a macro cell SISO network, the main way to improve the spectral efficiency is to increase the available bandwidth. However, in frequencies below 6 GHz, bandwidth is rather scarce, opening the way for smaller, higher frequency cells and the need to manage them. In the following chapter, a strategy for adapting existing SISO resource allocation schemes to multi-cell networks is presented.

Chapter 3

Multi-cell resource allocations

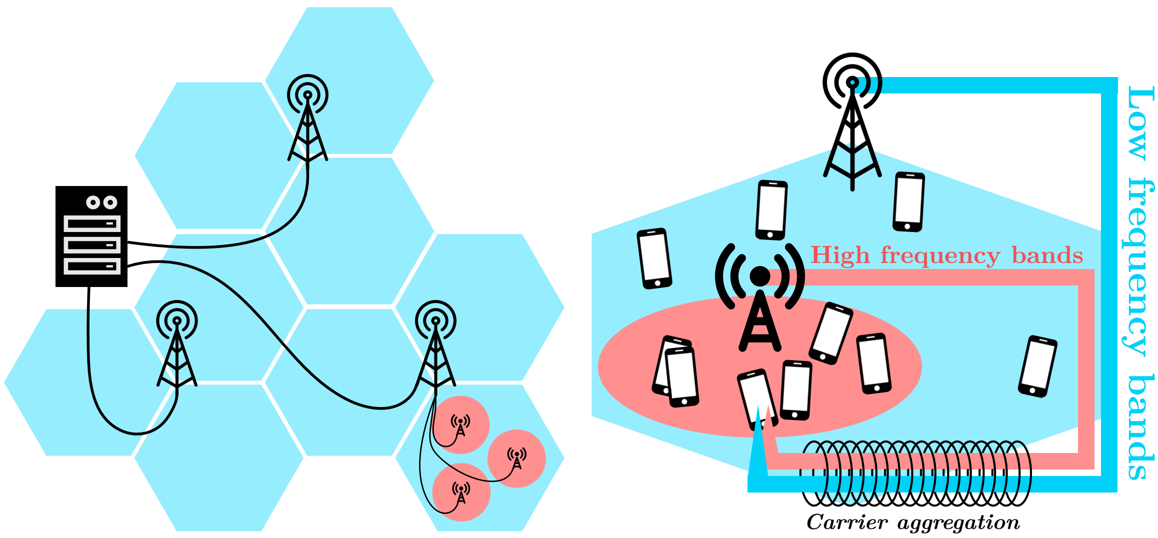
3.1 Introduction and motivation

Network resources such as bandwidth are often scarce and need to be shared between users inside the same coverage area. In order to manage this traffic, network densification constitutes a major evolution of network deployment. This system, called Hierarchical Cell Structure (HCS) [35], lies in the multiplication of the number of access points or Base Station (BS) with which users can establish a connection. In a fifth generation (5G) multicell context, we generally consider that micro cells (μgNB) are deployed in addition to the traditional 5G-based macro cell (next generation Node B (gNB))[36]. These cells allow to locally offload the system in dense areas where the demand in data may be higher (shopping centers, stadiums, train stations...). Access points are then divided into two categories whose characteristics are presented as follows:

— A macro cell layer provides a wide-area coverage due to their low frequencies band (<6GHz), facilitating a geographical continuous connection to users. Traditionally a macro cell BS can cover up to 10 km in rural areas and up to 1 km in dense urban area.

— Micro cell access points are low-power hotspots providing short range coverage area. Operating in high frequency bands (millimeter waves for 5G), these access points allow to deliver high data throughput thanks to large spectrum bandwidth. However, these frequencies suffer from stronger pathloss, shadowing and multipath fading significantly affecting the coverage area.

In this context, the management of all these access points can be done by a central processing unit. This unit (Figure 3.1) performs all the scheduling decisions and allocates resources to the users. This hierarchical management [37] optimizes the decision-making on how the network operates and allocates resources depending on user demands and channel quality.



(a) Centralized management (b) User deployment

Figure 3.1 – Typical 5G multicell scenario

In this type of wireless network, a new problematic lies in the allocation process. Indeed, in a typical 5G multicell scenario, radio layer evolution allows a user to be connected to several access points at the same time. This leads to the emergence of new allocation strategy issues. One of them lies in optimizing user traffic repartition between micro cell and macro cell access points. This depends on a lot of parameters such as users channel quality, services type and system load. Hereinafter, the study proposes different allocation strategies when the system suffers from high traffic load and tends to be overload.

A well studied topic, Coordinated MultiPoint (CoMP) transmissions, tackles the users repartition in multi-cell context, as seen in [7] and [8]. However, our solution takes benefit from different frequency bands for each station type. The macro cell and the micro cell work at different frequencies, eliminating the interference problem. Moreover, our pre-scheduler proposal is compatible with all allocation strategies (schedulers). Four of them are summarized in section 2.3.

This chapter aims to present our pre-scheduling solution called Multi-Cell Pre-Scheduler (MCPS) and evaluate its performance impact when applied upstream these four schedulers[1] [24, 17, 32, 19]. MCPS computes user-expected throughput with all the access points they can establish a connection. It assigns to each user a priority index. Depending on the micro cell maximal capacity, users with the higher priority index are assigned to

1. This work was done in association with an intern that I co-supervised and the work has been published in [38] and is part of the One5G project [39]

the micro while the rest of the users are assigned to the macro. This allows to assign to the macro cell the users who take the highest benefit. By dynamically arrange users between the two access points, the user traffic repartition is optimized guaranteeing a high system capacity.

The chapter is organized as follows. Section 3.2 describes related work concerning multi-cell algorithms. Section 3.3 provides a detailed description of the system under study and gives an important preliminary result attesting the use of a pre-scheduler. Section 3.4 introduces MCPS pre-scheduler principles. Section 3.5 presents two users deployment scenarii in order to evaluate the performance of our pre-scheduling solution. Section 3.6 concludes the chapter.

3.2 State of The Art (SoTA) for multi-cell algorithms

Using the same frequency for all access points leads to the need of advanced interference management. CoMP [7, 8, 40, 41] is a topic well covered in the literature and aims at covering this problem. The objective is to provide a fair throughput regarding the position of the users, where edge users may suffer from interference of nearby cells. In our context, edge users are covered by the micro-cell and do not suffer from interference thanks to the use of different frequencies.

One solution to limit interference for edge users is to identify users type (edge or center) during the resource allocation process. In [7], authors propose to divide the allocation process in three steps. The first step aims at classifying users in two groups: center users and edge users. The second step divides the resources available on these two user groups. The third step allocates resources to users in need. While MCPS also classifies users into two categories, resources are not shared between user groups, thanks to the use of different frequencies.

Another approach is to communicate with all surrounding BSs on the same resource to increase the signal and prevent those BSs from allocating that resource to another user and interfering with the first user. This technique is used by [40] and is divided into three steps. During the first step, users estimate the average power from surrounding BSs. In the second step, users rank the BSs according to the received power. In the third step, if two BSs have similar power, the user enables the CoMP-mode and authorizes the user to be served by both stations. If one BS has a higher power than others, the user does not activate the CoMP-mode and is only served by a single BS. Using MCPS, users are also able to receive from both cells. However, to prevent a cell from being selected by all the users, the distribution of users is done according to cell loads and not just user preferences.

Another main advantage of our solution is to be compatible with different types of schedulers, without changing their resource allocation policy, where CoMP algorithms often integrate a predefined scheduler.

3.3 System description

3.3.1 Deployment scenario

In the studied deployment scenarii, the entire bandwidth is split into several subcarriers. We assume that each subcarrier has independent Channel State Information (CSI) values [42]. Depending on the numerology used [43], Time Slot (TS) duration and subcarrier bandwidth may differ. The packets issued from the backhaul network are buffered in the access points, which schedules the downlink transmission. The radio resources are further divided in the time domain in frames. Each frame is itself divided in TSs, with a duration depending of the numerology. The TS duration is an integer multiple of the Orthogonal Frequency-Division Multiplexing (OFDM) symbol duration. The number of subcarriers is chosen so that the width of each Physical Resource Block (PRB) is less than the coherence bandwidth of the channel. Moreover, the frame duration is fixed to a value much smaller than the coherence time of the channel. We consider two access points with available Resource Units (RUs). Each RU is characterized by a {frequency ; time} pair and represents an unbreakable element, which is entirely made available to one user (Figure 3.2).

We assume a simple deployment scenario by considering a unique micro cell access point whose coverage area (represented in red on figures) is fully included in the macro cell one (represented in blue). We also assume that a user achievable throughput is linked to the power he receives and is limited due to his distance from the access point, the BS power and multipath fading [21, 22]. We consider that each BS has a full knowledge of the channel for each user. Each of the two access points made available its own RUs. By using different frequency bands, they do not interfere with each other.

We also consider an admission control that denies the connection between a user and an access point if its received power and therefore its achievable throughput is below

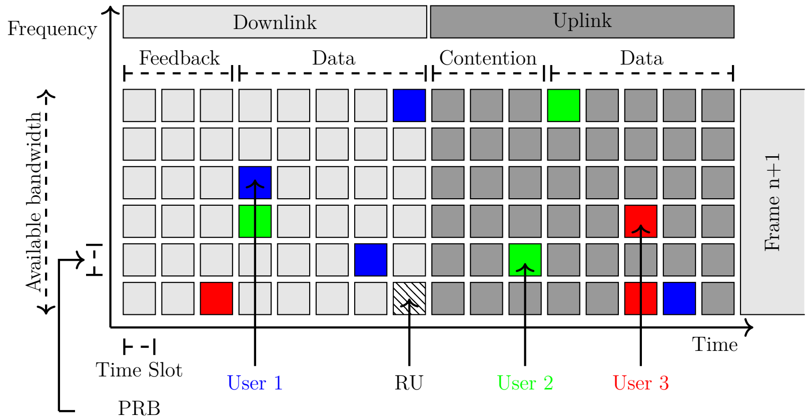


Figure 3.2 – 5G frame structure in TDD mode

a given threshold. This situation may happen when a user is far from a BS and need too much RUs to guarantee a sufficient Quality of Experience (QoE). In the rest of the chapter, we also consider that users covered by the micro cell are also covered by the macro cell.

3.3.2 Simulation parameters

In order to evaluate the performance of the multi-cells strategies we decide to simulate a simple user deployment scenario. We consider a random uniform user deployment in which some users are inside the micro cell coverage and some users outside. Simulation results were averaged over 100 different user position patterns. One of them is presented in Figure 3.3.

The potential number of bits that can be transmitted on a given RU is given by equation 2.3 and fluctuates over time. For the sake of simplicity, both BS types use the same numerology. In the following, $\alpha_{i,k}$ is Rayleigh distributed with an expectation equal to unity. In the rest of the chapter we consider that all users get the same BER_{target} of 10^{-3}.

We also consider a variable traffic, which is described in section 2.5.2. During a simulation, system load increases by adding users 2 by 2. BSs and users characteristics are given in table 3.1.

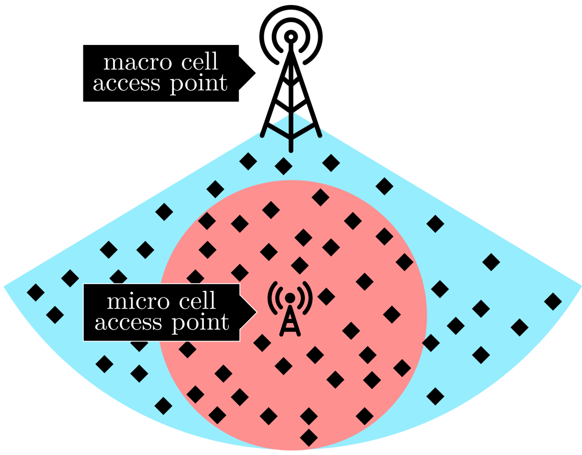


Figure 3.3 – Random user position deployment

	Users
Throughput requirements	560 kbits/s
Traffic model	Variable Bit Rate (VBR)
C_i	110.2 mW
Cn_i	46.8 mW

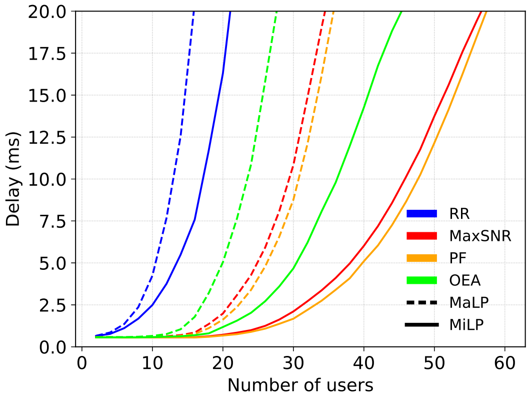
	Micro cell BS	**Macro cell BS**
Frequency	26.6 GHz	4 GHz
Power	33 dBm	44 dBm
Numerology	0	

Table 3.1 – BSs and users characteristics

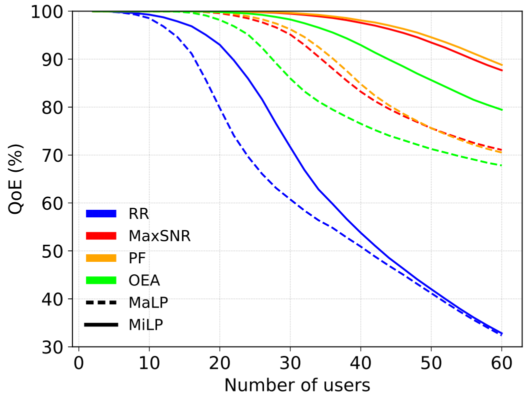
3.3.3 Preliminary result

The main objective of this subsection is to validate an important preliminary assumption necessary to set the implementation of the pre-scheduler. In a typical 5G multicell scenario in which a user may be in the coverage area of several BSs, the main problem is to find the optimal way to share the traffic between all access points. This lies in giving a priority order to each access point in the allocation process. This priority order allows to always select the best access point to establish a connection. Three different states can describe this multicell scenario:

— *State 1*: The system is not overloaded, both access points can provide RUs.



(a) Mean delay in the global system



(b) QoE indicator

Figure 3.4 – MiLP/MaLP performance results comparison

— <u>*State 2*</u>: The system is in the process of congestion. One of the two access points can't provide RUs anymore. Two different subcases are presented:

  (a) The micro cell is congested, the macro cell is not.

  (b) The macro cell is congested, the micro cell is not.

— <u>*State 3*</u>: The system is totally congested, no more access points can provide RUs.

As the load in a system increases, it goes from *State 1* to *State 2* (2(a) or (2(b)) and finally to *State 3*. Two allocation strategies acting during *State 1* are studied. These two strategies are called "MiLP" and "MaLP" for "Micro Load Priority" and "Macro Load Priority" respectively. MiLP solution prioritizes State 2(a) by allocating users to the micro in priority (if they are in its coverage area) while MaLP solution favors State 2(b) by allocating users to the macro in priority. In other words, with MiLP strategy, traditional schedulers (Round Robin (RR), Maximum Signal to Noise Ratio (MaxSNR),...) firstly allocate users with the micro cell access point. When this one overloads, the latter macro cell deals with users which couldn't be assigned to the micro (due to overloading) but also with users outside micro coverage. With MaLP solution, traditional schedulers firstly allocate users with the macro cell access and then to micro cell when the first one overloads. An important result lies in finding the best allocation strategy allowing the system to be more robust to the system congestion keeping user QoE at its highest level.

Figure 3.4 gives the performance results in terms of global system mean packet delay (3.4(a)) and QoE (3.4(b)) for the four SoTA schedulers. Delay indicator represents the time duration between a packet creation in the backhaul network and its reception by the

User Equipment (UE). A packet is considered out of delay if the time constraint imposed by the user service (streaming video, web,...) is not respected. The QoE is expressed as the proportion of packets that experience a delay below a threshold. At the end of each communication, we determine the ratio between the number of packets out of delay and the total number of packets sent to the user.

Mean delay result shows that by applying MiLP strategy, the mean delay experienced by users is more robust to the system load increase. For example, with MaxSNR scheduler, about 35 users are needed to reach a mean delay of $20ms$ with MaLP solution and up to 56 users for MiLP strategy (Figure 3.4(a)). In other words MiLP deployment allows to accept 60% more users while guaranteeing a QoE beyond 90% (Figure 3.4(b)).

By keeping the MiLP solution, the system can improve the probabilities that all users are in the radio coverage to an access point, which is not overloaded. Indeed, if all micro-cell users are firstly assigned to the macro cell leading to its congestion (MaLP solution), a user outside the micro cell coverage can no longer receive RU from it (State 2b). This leads to a sub-optimal solution knowing that several users are inside micro coverage (which is not overloaded) but assigned to the macro cell. These same users should be assigned to the micro cell releasing the macro bandwidth to users that are outside the micro coverage. This result constitutes an important preliminary result before applying the MCPS pre-scheduler (introduced in the next section 3.4). In the rest of the document, MCPS pre-scheduler is implemented by considering MiLP strategy, which allows to be more robust to traffic congestion.

3.4 Multi-Cell Pre-Scheduler (MCPS)

MCPS pre-scheduler objective is to avoid the system congestion by solving an allocation optimization problem when the micro cell tends to be overloaded. This situation may occur when too many users are located in the micro cell coverage area causing a high traffic peak demand leading to its congestion. In this case some traffic (i.e. some users) need to be moved to the macro cell. MCPS pre-scheduler allows to adequately choose the users located in the micro cell coverage area, which are assigned to the macro cell in the allocation process since some users are more profitable to be assigned to the macro cell. MCPS allows to effectively choose these users by allocating to each one a priority index. The higher a user index is, the higher its probability to be allocated by the micro cell is. The main objective is to keep in the micro cell the more profitable users and thus exclude

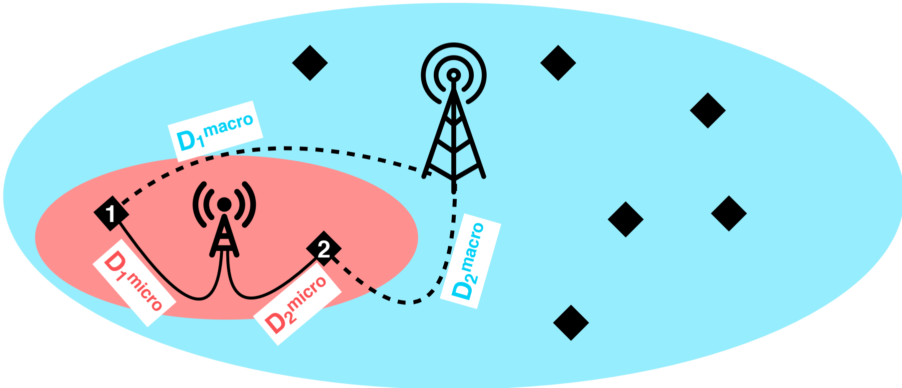


Figure 3.5 – MCPS throughput calculation representation

them for the macro (considering throughput). MCPS pre-scheduler, whose main objective is to increase system capacity, is based on user-expected throughput calculation and is described in subsection 3.4.1.

3.4.1 Algorithm description

MCPS pre-scheduler is based on the calculation of the user-expected throughput with the access points they can establish a connection. To each mobile i, a $MCPS_i$ is associated whose value is updated at each allocation decision instant. $MCPS_i$ is calculated following equation (3.1).

$$MCPS_i = \overline{D_i^{micro}} - \overline{D_i^{macro}}, \tag{3.1}$$

where $\overline{D_i^{micro}}$ and $\overline{D_i^{macro}}$ represent the mean user i expected throughput on a RU with the micro cell and the macro cell access points, respectively. So $MCPS_i$ values is mainly linked to user location (pathloss) and the implemented scheduler.

All users are then classified according to their $MCPS$ value with an index between 1 and i_{max} (where i_{max} is the number of users in the micro BS coverage and then concerned by $MCPS$). The user with higher $MCPS$ value gets index 1 and the user with lower $MCPS$ value gets index i_{max}. A positive (correspondingly negative) $MCPS_i$ value reflects the interest of connecting user i to the micro cell (correspondingly macro cell) access point with which its expected throughput is higher. Unfortunately, because of micro cell limited capacity, all users with a positive $MCPS$ value could not necessarily be allocated by the micro cell. Users with higher $MCPS$ value get higher index priority. The second step is then to determine the maximal number of users the micro cell can manage. This limit is calculated as follows:

$$Limit\ [\%] = 100 \times \frac{\overline{D^{micro}} \times N_{RB}^{micro}}{\sum_{i=1}^{i_{max}} BO_i(t)}, \tag{3.2}$$

where:

— $\overline{D^{micro}}$ *[bit/RU]* is the mean spectral efficiency of the micro cell and is measured from the previous allocations.

— N_{RB}^{micro} *[RU]* is the number of Resource Blocks (RBs) made available by the micro cell access point and is constant.

— $\sum_{i=1}^{i_{max}} BO_i(t)$ *[bit]* corresponds to the Buffer Occupancy (BO) of user i at time t. The BO is the number of bits that have to be transmitted to a given user. i_{max} is the total number of users in the coverage area of the micro cell BS.

This limit corresponds to the ratio between the number of RUs the micro cell can provide and the total traffic (bits) created by users in the micro cell at a given time. In other words, this limit value gives the ratio of RUs the micro can handle until its congestion which corresponds to a bandwidth usage ratio equal to 100%. According to this limit ratio of users, the micro cell manages the percentage of the i_{max} users that have the higher $MCPS$ value letting the macro cell the other users with lower $MCPS$. By doing so, the system allows the macro cell to allocate the users, which are the most beneficial to it. So the system dynamically manages all the users by adequately arrange them between the two access points in order to increase the global system capacity. Once the MCPS has made this pre-allocation by assigning each user to an access point, traditional schedulers (RR, MaxSNR,...) take over their own assigned users in both cells. This pre-scheduling process is summarized by the Figure 3.6 with its flow chart.

3.4.2 MCPS example

In order to well understand MCPS operation, this subsection is dedicated to give an example presented on Figure 3.7(a). This example considers a group of 10 users all covered by both micro and macro cell access points. They are therefore concerned by MCPS pre-scheduling. In this implementation, $MCPS$ value (fourth column) is calculated from expected throughput in micro (second column) and in macro (third column) respectively.

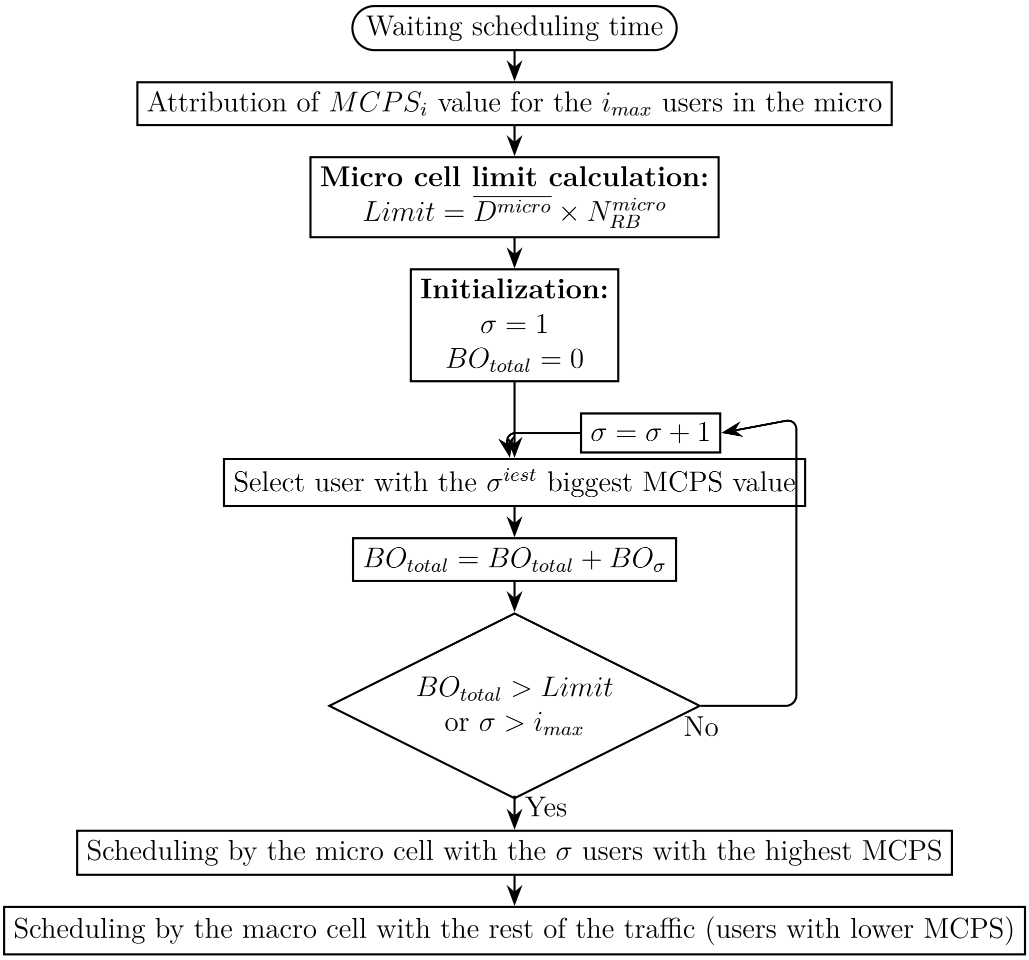


Figure 3.6 – MCPS pre-scheduler algorithm flow chart

In this example, micro cell that has a fixed capacity, can provide RU to the six first users with high MCPS values (users 7, 2, 1, 5, 9 and 10). The other four users (users 3, 8, 6 and 4) with lower MCPS are allocated by the macro cell. Figure 3.7(b) shows the importance to sort users in order to maximize the mean system throughput. The upper table represents the previous situation 3.7(a) for which users are classified (MCPS applied). In this case system throughput reaches 21.8 MBit/sec. Bottom table of Figure 3.7(a) considers that MCPS is not applied and therefore users are not classified. So the excess traffic is randomly chosen. In this case, the mean global throughput reaches 19.5 MBit/sec. It shows that MCPS users sort efficiently assigns a given user to the access point allowing him to reach a good throughput.

Note that the MCPS limit (i.e. the number of users considered by the micro cell) is closely linked to the generated traffic at the decision-making instant. Each user produces its own variable traffic and so get a weighting that depends on what he generates. At a given time t, some users may generate high traffic peak while some others very stable one. This impacts the MCPS limit value and so the number of users assigned to the micro cell. Figure 3.7(c) gives an example of two scenarii impacted by different traffic peaks. Scenario A is a visualization of the table issued from Figure 3.7(a)

In scenario B, user 7 with a high MCPS value produces a high traffic peak that occupy a large portion of the microcell bandwidth. This impacts the rest of the micro cell in bandwidth availability. The result is that the number of users assigned to the micro cell may decrease. This is the case in Scenario B for which only 5 users are assigned to the micro cell. The traffic generated by the k_{max} users being totally independent to the micro cell capacity, micro cell limit is rarely fixed between two distinct users. This is the case in Scenario A with user 10 and scenario B with user 9. Their generated traffic allows to achieve $100+\epsilon\%$ of the total micro cell bandwidth. MCPS considers that, if the micro cell bandwidth usage ratio is less than 100%, it assigned to it one more user by slightly exceed its maximal capacity. In that case, the micro cell capacity is fully exploited allowing to achieve better performance due to higher multi-user diversity in the micro cell. Then, the $\epsilon\%$ overflowed traffic is transferred to the macro cell.

3.4.3 Discussion concerning MCPS value

As explained in previous subsection 3.4.1, MCPS is linked to the user throughput and so closely depends on their channel condition. Experienced throughput is also linked to the implemented scheduler and on the way it favors users compared to some others.

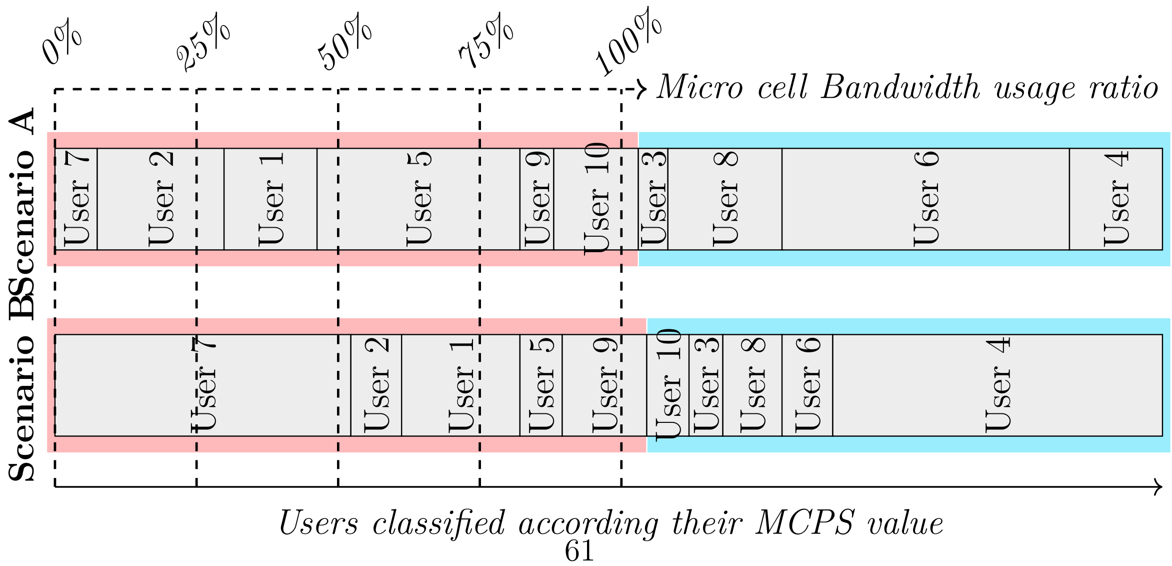
User index	micro bit rate *Mbit/sec*	macro bit rate *Mbit/sec*	MCPS value	Priority	Cell
7	30.3	15.2	15.1	1	←micro cell→
2	33.2	20.9	12.3	2	
1	20.6	11.8	8.8	3	
5	21.5	12.8	8.7	4	
9	31.7	24.6	7.1	5	
10	18.8	14.9	3.9	6	
Micro cell limit capacity					
3	15.1	13.0	2.10	7	macro cell
8	20.8	22.3	-1.5	8	
6	8.9	12.3	-3.4	9	
4	6.5	14.1	-7.6	10	

(a) User classification according their MCPS value

If users are classified : MCPS is applied										
User index	7	2	1	5	9	10	3	8	6	4
Realized throughput	30.3	33.2	20.6	21.5	31.7	18.8	13.0	22.3	12.3	14.1
Total mean throughput	21.8 MBit/sec									

If users are not classified : MCPS is not applied										
User index	6	7	1	8	2	3	5	4	10	9
Realized throughput	8.9	30.3	20.6	20.8	33.2	15.1	12.8	14.1	14.9	24.6
Total mean throughput	19.5 MBit/sec									

(b) Impact of user classification on total system mean throughput



(c) Impact of VBR traffic on micro cell limit

Figure 3.7 – MCPS example

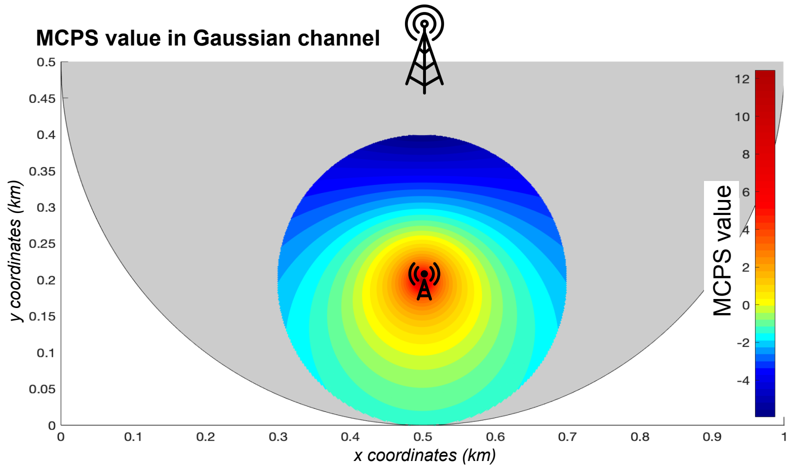


Figure 3.8 – MCPS

According to the channel conditions, path loss affects the user achievable throughput. If we consider a Gaussian channel without fading attenuation, $\alpha_{i,k} = 1$ in equation (2.3) so $MCPS$ value can be written as:

$$MCPS = \log_2 \left(\frac{1 + \eta \frac{P_{micro}}{d_{i_{micro}}^{3.5}}}{1 + \eta \frac{P_{macro}}{d_{i_{macro}}^{3.5}}} \right), \tag{3.3}$$

where η is a constant, which is common to all users and equal to $\dfrac{3 \times T_s}{2N_0 \left[\mathrm{erfc}^{-1} \left(\frac{BER_{target,k}}{2} \right) \right]^2}$.

In this case, $MCPS$ mainly depends on BS radiating power and user location. Figure 3.8 shows that $MCPS$ value can take value from 12 (red part close to the micro cell base station) to -6 (blue part close to the macro cell BS). Circles of same $MCPS$ values are notable around micro cell BS. This phenomenon is closely linked to the distance as shown in equation 2.3. The theoretical locations where $MCPS = 0$ represent the neutral user locations and there is no advantage to assigned the user either to the micro or to the macro. We can easily show that $MCPS = 0$ under the condition that $P_{micro} \times d_{i_{macro}}^{3.5} = P_{macro} \times d_{i_{micro}}^{3.5}$. This shows that assigning a user to an access point depends on its location, on BSs power and also on relative position of the two access points. Some users deployments bring more benefits to MCPS pre-scheduler than some others. Indeed, given the presence of only two BSs, it is clear that MCPS value admits a symmetry behavior along the axis of the two access points.

3.5 Performance evaluation

3.5.1 Simulation setup and studied Key Performance Indicators (KPIs)

This section aims to study the performance of the four SoTA schedulers, introduced in first section, when they are applied or not downstream to the *MCPS* pre-scheduler. All simulation parameters are described in section 3.3.1 and summarized in table 3.2.

	Micro cell BS	**Macro cell BS**
Frequency	26.6 GHz	4GHz
Power	33dBm	44 dBm
Access Point	5G	
Numerology	0	

Table 3.2 – Base stations parameters

In the next two subsections 3.5.2 and 3.5.3, two users deployment scenarii and their respective obtained performance are presented. The base station deployment (BS power, frequency and position) remains unchanged from the preliminary results presented in subsection 3.3.3. Performance results are obtained using discrete event simulations. The KPIs (described in section 2.1) are studied according to the load in the system and more precisely following the number of users connected to the system. Application delay threshold is fixed to 100 ms in the following. For each discrete point, simulation results are averaged over a communication duration of 1000 Frames.

3.5.2 Scenario 1: Proof of concept deployment

In order to clearly underline the MCPS pre-scheduler behavior, we first study the simple users deployment scenario presented on Figure 3.9. This lies in considering two groups of users located at the same distance from the micro cell BS ($100m$). One group is far ($400m$) and the other one close ($200m$) from the macro cell BS. A random position variation is assigned to each user in its group. These horizontal and vertical variations Δ_x and Δ_y follow a uniform distribution on $[-20m, 20m]$. This allows users to be assigned to uncorrelated fading allowing to have a better multi-user diversity. So opportunistic

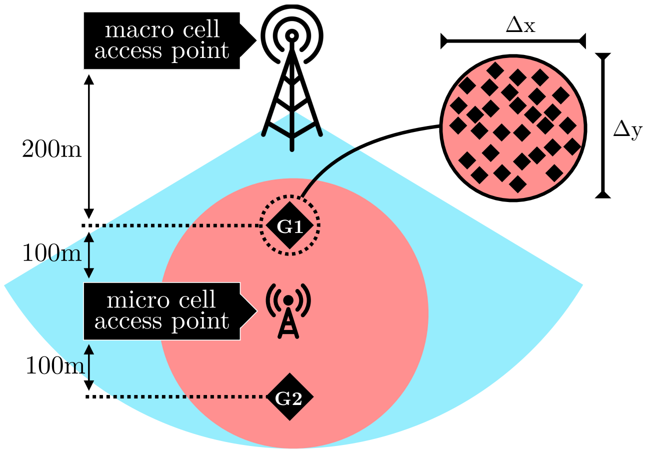
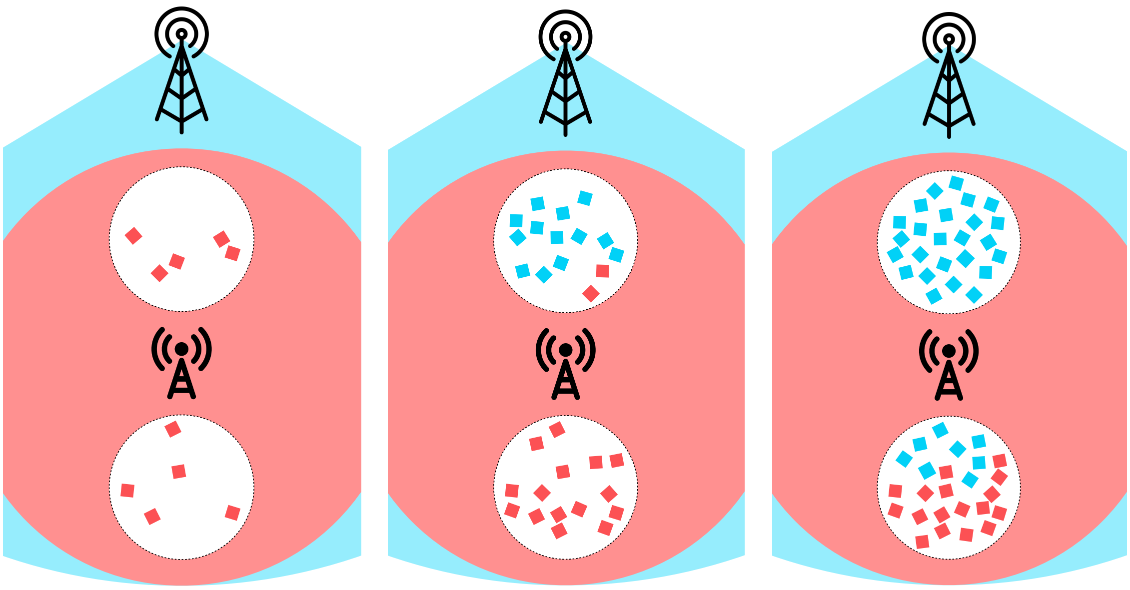


Figure 3.9 – Grouped position deployment

schedulers and MCPS can operate by always having a wide range of choices. We also consider that each user generates a realistic variable traffic [44, 45].

First results are exposed on Figure 3.11 and present the performance of the studied deployment scenario for the four SoTA schedulers when they are applied (full line), or not (dashed line) downstream to our MCPS pre-scheduling solution. In order to well understand the system behaviour facing the increase of the traffic load, bandwidth usage ratio graph (Figure 3.11(a)) gives the result both for micro and macro cells. We consider that the full system (micro+macro) is totally congested when both cells admit a bandwidth usage ratio of 100%. Before starting any technical explanation, it is important to remember that MiLP allocation strategy has been implemented: users are firstly assigned to micro cell, if possible, before the macro cell.

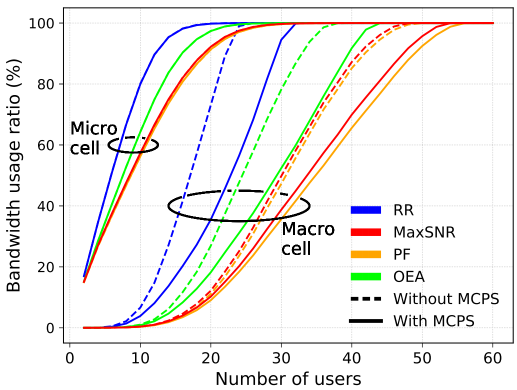
Concerning the bandwidth usage ratio, results (Figure 3.11(a)) highlights that the global system resists to a higher load (higher number of users) when the MCPS pre-scheduler is applied. With MCPS, schedulers are able to absorb more traffic and are more robust to unexpected traffic peaks. For example MCPS allows Round Robin to accept 24 users in the system without MCPS and 32 with MCPS. This lead to accept 33.3% more users with MCPS. In the same way MCPS allows the system to deal with 17.4% more users with MaxSNR. Figure 3.11(b) is linked to this result and shows the number of users considered by schedulers in the micro cell when applied with MCPS (and in dashed line without MCPS). For very low traffic loads the total numbers of users considered by the micro cell is equal to the total number of users in the system. When traffic load increases, the MCPS attributes the most beneficial users to the micro and macro cell respectively.



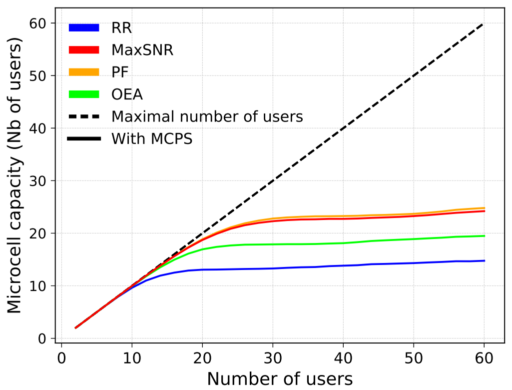
(a) 10 users (b) 30 users (c) 50 users

◆ : *Users allocated by macro cell*
◆ : *Users allocated by micro cell*
Considering a micro cell maximal capacity of 17 users

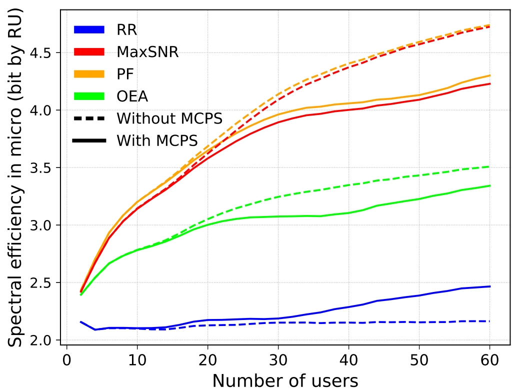
Figure 3.10 – MCPS behavior with grouped users deployment



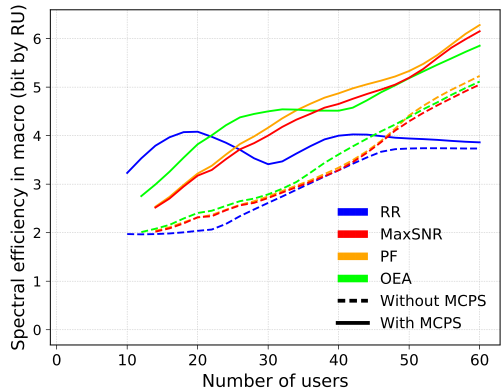
(a) Bandwidth usage ratio



(b) Micro cell users capacity

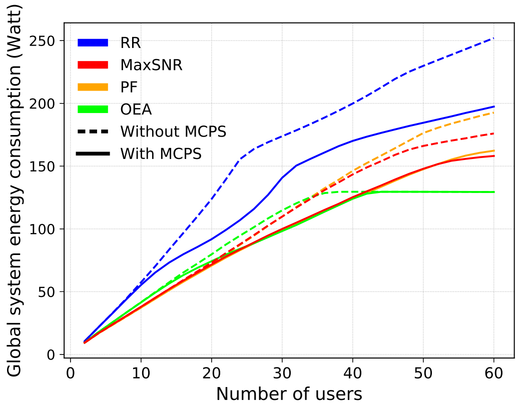


(c) Spectral efficiency in the micro cell

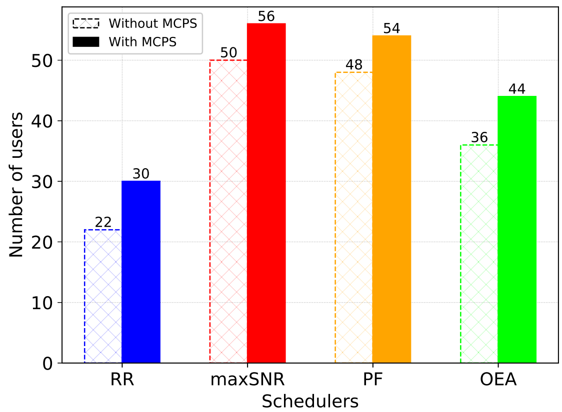


(d) Spectral efficiency in the macro cell

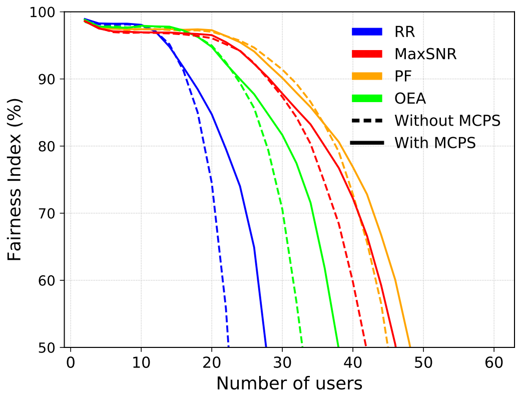
Figure 3.11 – Performance results for scenario 1



(a) System energy consumption



(b) User capacity limit with QoE guarantee



(c) Fairness Index

Figure 3.11 – Performance results for scenario 1

Maintaining the maximum number of user manageable by the micro, curve flat appears when micro cell reaches maximum capacity and macro begins to manage the overloaded traffic. The asymptotic behavior testifies of the maximal number of users the micro cell can manage depending of the scheduler.

Bandwidth usage behaviour is closely linked to the spectral efficiency one. By applying MCPS pre-scheduler, a significant decrease of the spectral efficiency for opportunistic schedulers (MaxSNR, Proportional Fair (PF), Opportunistic Energy Aware (OEA)) is notable in the micro cell (Figure 3.11(c)). This loss in the micro cell is mainly due to this type of scheduler that needs a high multi-users diversity in order to always choose the best user allowing to offer him the highest system capacity. By offloading some users to the macro cell (case (b) on Figure 3.10), MCPS pre-scheduler restrains the optimal operation of schedulers when selecting the best user. This degradation is perceptible from a given number of users ($\simeq$ 14 for OEA, $\simeq$ 16 MaxSNR and PF), that underlines when MCPS begins to operates. Indeed, below this number of users, congestion probabilities are negligible and MCPS is inactive. Round Robin which is a non opportunistic scheduler admits a slight gain in the micro by applying MCPS. Round Robin does not take into account channel conditions and user diversity. By applying MCPS with Round Robin, the loss of multi-user diversity has therefore no consequence on spectral efficiency. The impact of MCPS is even positive on spectral efficiency. Indeed, according to Figure 3.10(b) when the micro cell saturates, close users from the macro access point (and the farthest to the micro) are the first assigned to the macro (MCPS behaviour). Users staying in the micro cell admit a higher mean achievable throughput due to better path loss, spectral efficiency for RR increases in the micro cell.

However, this micro cell spectral efficiency loss is largely compensated by a gain in the macro cell (Figure 3.11(d)). As soon as the load exceeds a given threshold specific to each scheduler, MCPS detects the micro cell congestion risk and adequately share the most beneficial users between micro and macro cells. This threshold is reached when the micro cell has managed the number of users corresponding to its maximal capacity (Figure 3.11(b)). Macro cell spectral efficiency behaviour can be divided into four parts:

— For very low loads (<10 users), system is widely underloaded and macro spectral efficiency does not exist since all users are allocated by the micro cell (Figure3.10(a)).

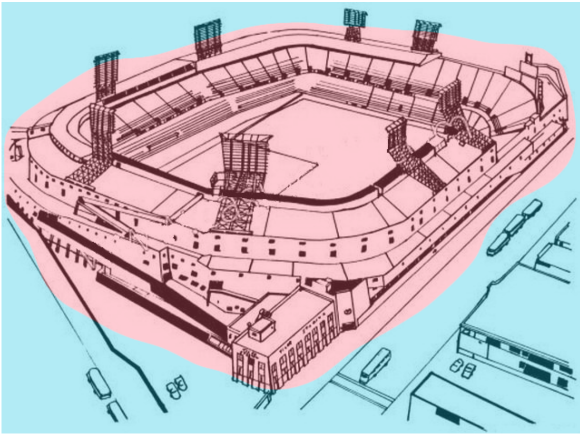
— When the micro cell overloads, first users that are assigned to the macro cell are the one with good throughput (close users - low MCPS). Spectral efficiency in the macro increases (Figure3.10(b)). The spectral efficiency gain with MCPS is linked to a better management of users. Macro cell receives the more profitable users to it.

— When the micro reached its maximal users capacity, the macro cell has to manage far users, which are less advantageous in terms of throughput. This slowdown is noticeable on Figure 3.11(d) from 20 to 32 users for RR, 30 to 42 users for OEA, and from 40 to 50 users for MaxSNR and PF. The curve with MCPS solution tends to get closer from the curve without MCPS solution (compared to Figure3.10(d)). But the increase still persists due to a better users-cells repartition.

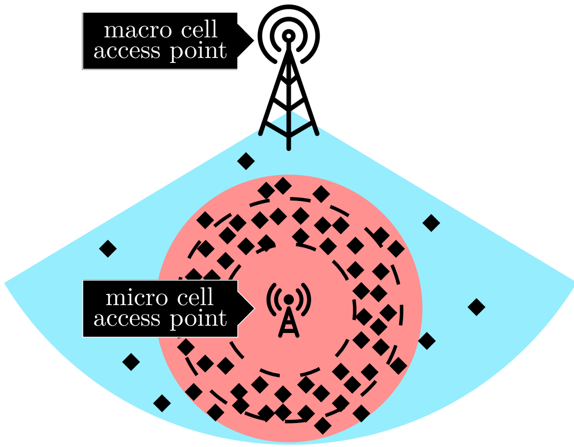
— For very high loads this gap seems to be maintained or even increased. With MCPS, opportunistic schedulers resume their normal operation reaching good performance due to the supplementary of multi-users diversity (Figure3.10(e)). A significant throughput gain with MCPS is obtain over non-MCPS solutions thanks to more close users assignation.

In addition to increase the overall capacity of the system (bandwidth usage ratio and spectral efficiency previously studied), some additional KPIs are improved by applying MCPS pre-scheduler (fairness, energy saving). For example Figure 3.11(a) shows that the global system energy consumption are improved with MCPS. Indeed, MCPS effectively assigned users with the access point they can achieve a good throughput and therefore they are able to finish their connection quicker. Users are then in an active mode for a shorter time and then consume less.

Figure 3.11(c) also shows the MCPS ability to keep QoE fairness between users as the load increases. Considering MaxSNR scheduler at 40 users in the system, the fairness index reaches 60% and 72% without and with MCPS respectively. This gain is perceptible for the four schedulers. When users are assigned to BS, they share the access point with other users whose throughput difference is not significant. More users obtain a good QoE satisfying their application requirements (3.11(b)).



(d) Realistic stadium deployment



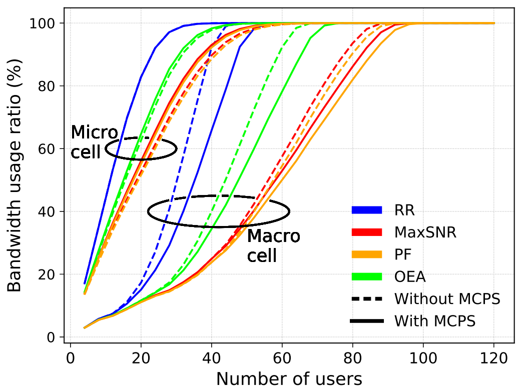
(e) dessin

Figure 3.12 – Realistic scenario representation

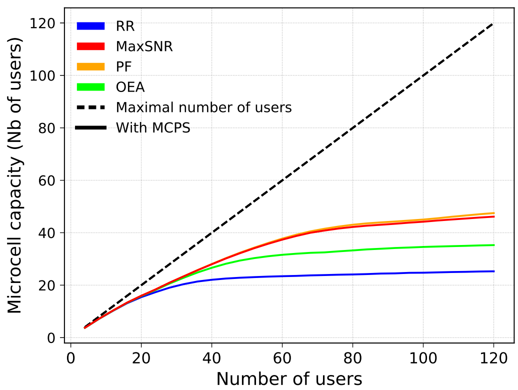
These first simulation results constitute a proof of concept testifying the performance bring by MCPS pre-scheduling. A more realistic deployment is studied in the following subsection.

3.5.3 Scenario 2: Realistic user deployment

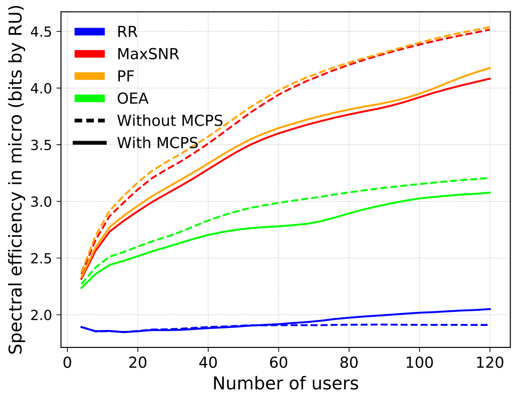
In order to confirm the previous performance results bring by MCPS pre-scheduling, we simulate a more realistic users deployment scenario. This deployment is inspired by 3GPP deployment specifications [46] and is represented on Figure 3.12. We consider an ultra-dense deployment scenario for which a user group is distributed around an access point. This access point allows to offload the macro cell in this dense area where the demand is high. Typically, we can consider a stadium deployment (Football stadium, concert arena). A stadium is typically a place where the traffic is high (video uploading,...) and where a micro cell hotspot deployment makes sense. This deployment represents a concrete case for which MCPS pre-scheduler can bring a benefit compared to a traditional solution. This realistic deployment considers a hotspot located in the center of the stadium (Figure 3.12(d)) providing high data rate broadband coverage to the users located around it. We evaluate the MCPS performance by studying the same KPIs than previous users deployment. In order to get reliable KPIs, simulation results were average over 100 different users deployment patterns. The traffic generated by users and BSs characteristics, remains the same as previous studies. Results are presented on Figure 3.13.



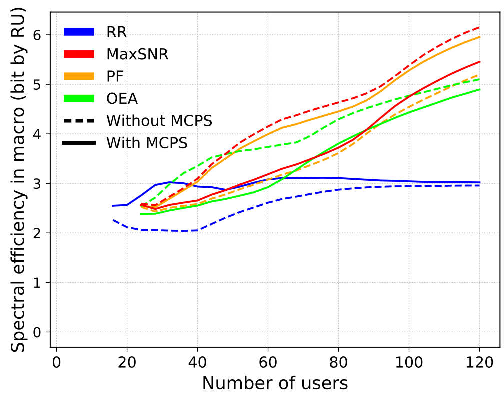
(a) Bandwidth usage ratio



(b) Micro cell users support

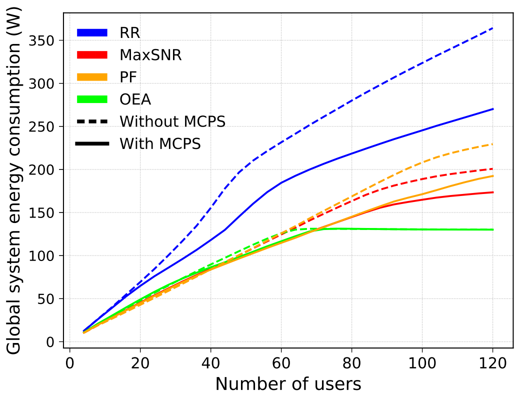


(c) Spectral efficiency in the micro cell

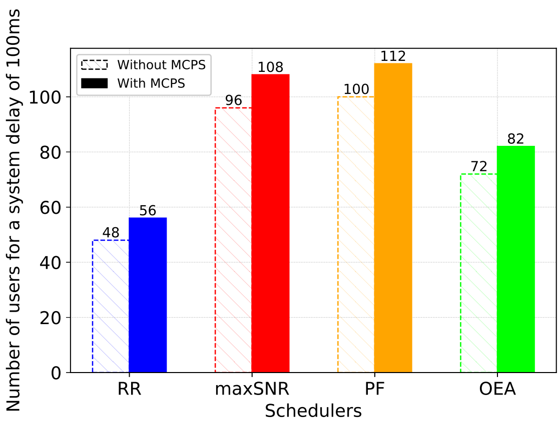


(d) Spectral efficiency in the macro cell

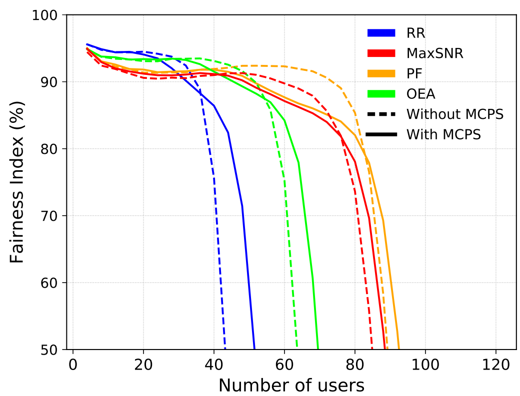
Figure 3.13 – Performance results for scenario 2



(a) System energy consumption



(b) User capacity limit with QoE guarantee



(c) Index Fairness

Figure 3.13 – Performance results for scenario 2

According to Figure 3.13 this scenario confirms MCPS pre-scheduler performance. Results concerning bandwidth usage ratio show that the system can allocate 9.6% more users for MaxSNR and 16% more for RR when applying MCPS. We can also notice the same spectral efficiency trends, meaning a loss in the micro cell compensated by a gain in the macro. User QoE (Figure 3.13(b)) guarantees and QoE fairness (3.13(c)) are also improved.

Figure 3.13 (b) shows that the numbers of users considered by the micro cell is always lower than the total number of users in the system. This shows that some users are located outside the micro cell BS coverage who necessarily have to be allocated by the macro cell access point. We also note the same asymptotic behavior testifying the micro cell maximal capacity and so the number of users the micro cell can manage. The two scenarios previously studied show the benefit brings by MCPS pre-scheduler in a multicell scenario.

3.6 Conclusion on Multi-cell resource allocations

In the literature many research efforts have been done focusing on scheduling optimization in a simple deployment scenario with one cell. However these scheduling techniques are often not optimized to a multicell scenario. This chapter proposed a new pre-scheduling solution called “MCPS" for multicell wireless networks. Taking into account users achievable throughput, MCPS effectively share the traffic load between the different access points increasing the global system capacity. Performance results have shown that the benefits of this new approach allow the system to be more robust to the congestion and traffic peaks. Although the capabilities of MCPS could be improved by including opportunistic, fairness or energy consumption in the decision process, it will be at the expense of the versatility of the algorithm. However, MCPS could be paired with existing algorithms that dynamically activate BSs depending on the system load to reduce power consumption [47]. The main solution to improve the performance of multi-cell systems is to increase the number and variety of micro cells. However, the number of deployable BSs or the available bandwidth may be geographically or technically constraining. The next chapter presents a technique able to increase the capability of BS in order to overcome deployment constraints.

Chapter 4

Massive-MIMO resource allocations

4.1 Introduction and motivation

One of the key technologies for 5G radio transmission is Massive-Multiple Input Multiple Output (MIMO) [48]. Having a larger number of antennas at the Base Station (BS) can significantly improve the link budget by focusing the energy in the chosen direction [49]. Focusing the energy can also be used to serve Multiple Users (MU) on the same time and frequency resource. This opens the way to conceiving new scheduling algorithms in order to take a real advantage from Massive-MIMO.

In the MU mode, several users are served on the same time-frequency unit, they are said to be in the same group. The total system rate is expected to be higher compare to Single User (SU) mode. However, there is intracell interference, which increases with the number of users by group and reduces the bit rate. Reducing interference is crucial but done at the expense of individual throughput. Indeed, an efficient grouping strategy is necessary to find the best compromise between interference reduction and individual throughput. The development of such strategies has to rely on efficient indicators. Papers from the literature propose to use as a grouping indicator the correlation of users' channel matrices [9, 10, 50].

The objective of this chapter is to propose a new indicator for group selection in a Multiple Users Multiple Input Multiple Output (MU-MIMO) context. First, we propose a study on correlation-based systems. Then, we study different user selection algorithms and propose a new solution based on previous user selections. Finally, we evaluate the impact of the group size on overall system throughput. Our solution, called Efficient System Capacity User Selection (ESCUS) [1], progressively built its indicator upon previous users

1. This work under the publication process

selections by storing users throughput and select group of users by considering them couple by couple. With the capability of the system to manage more users simultaneously, the algorithms complexity increases.

The chapter is organized as follows. Section 4.2 provides a detailed description of the system in question. Section 4.4 presents a study of the channel matrix correlation as interference management. Section 4.5 presents our new algorithm proposal for user selection. Section 4.6 introduces the main parameters of our simulation set up and the performance evaluation of our new solution. Section 4.7 concludes the chapter.

4.2 Description of the transmission chain

We consider one BS or next generation Node B (gNB) and several User Equipment (UE) and we study the downlink transmission. In order to maintain our simulation time under a reasonable limit, each simulation frame has a duration of 10ms with a usable bandwidth of 2 MHz. This frame is divided into Resource Unit (RU) with 60 kHz bandwidth and 1 ms length. We define B as the bandwidth of the RU, $\mathcal{U}$ as the set of all UEs in the cell, $\mathcal{A}$ as the set of all UE allocated simultaneously to the same RU, and n_{t}, n_{r} the number of antennas at the transmitter (BS) and at the receiver (UE), respectively. The channel between the set of antennas at the BS and the set of antennas at UE i is modeled by complex matrix $\mathbf{H}_i$, as shown in figure 4.1.

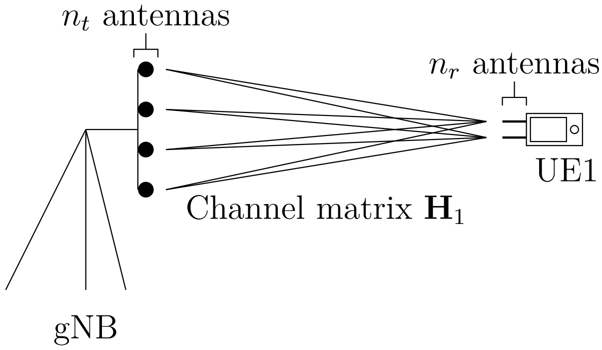


Figure 4.1 – Massive-MIMO description

The precoding matrix at the BS is represented by matrix $\mathbf{F}_i$ and the digital combining matrix at UE i is represented by $\mathbf{W}_i^*$. The white noise average power is denoted by σ_{N}^2. Note that the interference from neighbouring BSs can be integrated in σ_{N}^2.

The channel matrix of a UE i is given by $\mathbf{H}_i[t] \in \mathbb{C}^{n_\mathrm{r} \times n_\mathrm{t}}$ and is perfectly known at the gNB. The total number of parallel transmissions that can be made on the same RU by one gNB is given by n_t. The maximum number of transmissions on the same RU to a given UE is $\min(n_\mathrm{t}, n_\mathrm{r})$. The time is considered discrete and Δt is equivalent to a Time Slot (TS). We assume that the coherence time of the channel is greater than Δt and therefore greater than the scheduling time.

4.2.1 Transmission model

At a given time, scheduling can be viewed as an indicator function: $\delta_i(t)$ where i is the UE index, k the RU index and $\delta_i(t) \in [0, \min(n_\mathrm{t}, n_\mathrm{r})]$ gives the number of streams on the resource k allocated to i.

We have the following constraint:

$$\sum_{i \in \mathcal{U}} \delta_i(t) \leq n_\mathrm{t}. \tag{4.1}$$

The potential number of usable resources is given by $n_\mathrm{t} \times K$, where K is the number of resource units. We assume that there is no limitation regarding the number of Radio Frequency (RF) chains and the processing capacity. To illustrate the previous equation, we can consider a number of UEs equal to $n_\mathrm{t} \times K$, for example $n_\mathrm{t} \times K = 100$. In this case UEs will only have 1 RU among 100 on average per TS. It is then safe to say that in most cases, $\delta_i(t) = 0$.

For the sake of clarity we omit t for this part. We denote the precoding matrix associated with channel $\mathbf{H}_i$ as $\mathbf{F}_i \in \mathbb{C}^{n_\mathrm{t} \times \delta_{i,k}}$. The dimension of $\mathbf{F}_i$ depends on the number of antennas at the gNB and the selected streams. The signal transmitted to UE i is :

$$\mathbf{x}_i = \mathbf{F}_i \mathbf{s}_i \tag{4.2}$$

where $i \in \mathcal{U}$ and $\mathbf{s}_i[k]$ is the $\delta_i \times 1$ transmitted vector of symbols at a given subcarrier k. Note that $\mathbf{s}_i[k]$ is null if $\delta_i = 0$. The overall signal transmitted by the BS is:

$$\mathbf{X} = \sum_{j \in \mathcal{U}, \delta_j > 0} \mathbf{F}_j \mathbf{s}_j \tag{4.3}$$

At the receiver, the signal is affected by the channel matrix $\mathbf{H}_i$, noise corrupting the received signal $\mathbf{n}$ of power σ^2 and $\mathbf{I}$ external cell interference:

$$\mathbf{y}_i = \underbrace{\mathbf{H}_i\mathbf{X}}_{signal} + \underbrace{\mathbf{I}+\mathbf{n}}_{external\ interference+noise} \tag{4.4}$$

Due to the fact that we are considering a MU-MIMO transmission the transmitted signal for other UEs scheduled on the same resource unit has to be considered as internal interference:

$$\mathbf{y}_i = \underbrace{\mathbf{H}_i\mathbf{F}_i\mathbf{s}_i}_{signal} + \underbrace{\sum_{j\neq i,\delta_j>0} \mathbf{H}_i\mathbf{F}_j\mathbf{s}_j + (\mathbf{I}+\mathbf{n})}_{interference+noise} \tag{4.5}$$

This internal interference depend on the precoders $\mathbf{F}_j$ orthogonality to the channel matrix $\mathbf{H}_i$. The signal at the receiver after combining is finally given by:

$$\mathbf{z}_i = \mathbf{W}_i^*\mathbf{H}_i\mathbf{F}_i\mathbf{s}_i + \mathbf{W}_i^* \sum_{j\neq i,\delta_j>0} \mathbf{H}_i\mathbf{F}_j\mathbf{s}_j + \mathbf{W}_i^*(\mathbf{n}+\mathbf{I}) \tag{4.6}$$

where $\mathbf{W}_i^*[k]$ is the digital combining matrix [51] and $(.)^*$ the conjugate transpose of a complex matrix.

4.2.2 Precoding techniques

Fully digital precoding

We consider two precoding methods. Singular-Value Decomposition (SVD), which aims at maximizing the throughput in SU context and Block Diagonal (BD) which focus on interference management in MU context.

singular-value decomposition

A MIMO channel $\mathbf{H}_i$ of a user i can be decomposed using the SVD as follow:

$$\mathbf{H}_i = U_i\Sigma_i V_i^* \tag{4.7}$$

The optimal precoder in Single User Multiple Input Multiple Output (SU-MIMO) for i is $\mathbf{F}_i$[49]:

$$\mathbf{F}_i = V_i\Lambda^{1/2} \tag{4.8}$$

BD precoding [52][53]

In MU-MIMO several users can be scheduled simultaneously on the same RU. In such conditions, users can experience interference from other users. The objective of the Block Diagonal precoder is to eliminate UEs interference when they are scheduled in a MU-MIMO context. Therefore, all users that are simultaneously scheduled have to be considered in the precoding process.
We define $\tilde{\mathbf{H}}_i^{\mathrm{T}}$ the concatenation of channel matrices of all users in $\mathcal{A}$ except i:

$$\tilde{\mathbf{H}}_i^{\mathrm{T}} = [\mathbf{H}_1^{*}...\mathbf{H}_{i-1}^{*}\mathbf{H}_{i+1}^{*}...\mathbf{H}_A^{*}]^{*} \tag{4.9}$$

where $A = \mathrm{card}(\mathcal{A})$. Note that $\tilde{\mathbf{H}}_i^{\mathrm{T}}$ is a $n_{\mathrm{r}} \times (A-1)$ rows and n_{t} columns complex matrix. We are using the SVD on $\tilde{\mathbf{H}}_i^{\mathrm{T}}$.

$$\tilde{\mathbf{H}}_i^{\mathrm{T}} = U_i^{\mathrm{T}} \Sigma_i [V_i^{(1)} V_i^{(0)}]^{*} \tag{4.10}$$

where $[V_i^{(1)} V_i^{(0)}]$ is a $n_{\mathrm{t}} \times n_{\mathrm{t}}$ matrix. $V_i^{(1)}$ contains vectors corresponding to nonzero singular values and $V_i^{(0)}$ contains vectors corresponding to zero singular values.

$$\tilde{\mathbf{H}}_i^{\mathrm{T}} \tilde{V}_i^{(0)} = \tilde{U}_i \tilde{\Sigma}_i [\tilde{V}_i^{(1)} \tilde{V}_i^{(0)}]^{*} \tag{4.11}$$

The total precoding matrix is given by:

$$T^{BD} = [\tilde{V}_1^{(0)} V_1^{(1)} \tilde{V}_2^{(0)} V_2^{(1)} ... \tilde{V}_A^{(0)} V_A^{(1)}] \Lambda^{1/2} \tag{4.12}$$

where $\Lambda^{1/2}$ is a diagonal matrix of which the element λ_i scales the power transmitted.

4.2.3 UE throughput in MU mode

For the MU mode, we consider the BD precoding technique because the BD precoder is focused on interference management and therefore limits the reduction of the bit rate due to the intracell interference. The bit rate for UE i is given by [54] :

$$R_i = B \log_2 \left| \mathbb{1}_{n_{\mathrm{r}}} + \frac{\mathbf{W}_i^{*}\, \mathbf{H}_i\, \mathbf{F}_i\, (\mathbf{W}_i^{*}\, \mathbf{H}_i\, \mathbf{F}_i)^{*}}{\mathbf{W}_i^{*}\, (\sigma^2 \mathbb{1}_{n_{\mathrm{r}}} + \sum\limits_{j \neq i, \delta_j > 0} \mathbf{H}_i\, \mathbf{F}_j\, (\mathbf{H}_i\, \mathbf{F}_j)^{*}) \mathbf{W}_i} \right| \tag{4.13}$$

where $\mathbb{1}_{n_\mathrm{r}}$ is an identity matrix of size n_r.

In MU, $n_\mathrm{t}/n_\mathrm{r}$ is the limit on users served at the same time. We define n_s as the number of simultaneously scheduled users, where $n_s \leq n_\mathrm{t}/n_\mathrm{r}$. The mean system rate R is:

$$R = E\left[\sum_{i=1}^{n_s} R_i\right], \tag{4.14}$$

where E is the mathematical expectation. We consider a large set of random deployment of terminals, which are deployed randomly thanks to the properties of the channel model. Thus, $\mathbf{H}_i$ is a random matrix. In the following, for different configurations, we study R, which is our main performance indicator.

4.2.4 UE throughput in SU mode

In the in SU mode, δ_i is non-zero for only one value of i. Thus, there is no intracell interference and the rate C_i is maximized. Equation (4.13) is simplified as:

$$C_i = B\log_2\left|\mathbb{1}_{n_\mathrm{r}} + \frac{\mathbf{W}_i^*\,\mathbf{H}_i\,\mathbf{F}_i\,(\mathbf{W}_i^*\,\mathbf{H}_i\,\mathbf{F}_i)^*}{\mathbf{W}_i^*\,(\sigma^2\mathbb{1}_{n_\mathrm{r}})\mathbf{W}_i}\right| \tag{4.15}$$

User throughput in in SU mode is used as an indicator for classical opportunistic resource allocation schedulers. It is the main indicator for opportunistic scheduler, such as Maximum Signal to Noise Ratio (MaxSNR) [16].

4.3 State of The Art (SoTA) for MU-MIMO resource allocation

The objective of an algorithm for users selection is to build the most profitable set of users to share a Resource Block (RB). In MU-MIMO a good interference management is crucial to fully benefit from the technique. The channels matrices correlation is commonly used as an indicator of the interference level between UEs. The correlation of two UEs channel matrices indicates the interference level if they are using simultaneously the same RU simultaneously. The correlation of two UEs is defined by [55]:

$$\xi(i,j) = \frac{|\mathbf{tr}(\,\mathbf{H}_i\,\mathbf{H}_j^*)|}{||\,\mathbf{H}_i||_F||\,\mathbf{H}_j||_F} \tag{4.16}$$

where $(.)^*$ denotes the conjugate transpose operation and $||.||_F$ the Forbenius norm of a matrix.

4.3.1 Common strategy for user selection

Correlation Based User Selection (CBUS) [9] proposes to construct the grouping strategy of users on this correlation parameter. The first step is to select the first user in the group. To maximize the system capacity, the first user is selected depending on the best potential throughput, similarly to MaxSNR 2.3.

$$1.\ u = argmax(C_i) \tag{4.17}$$

To include the following users, CBUS relies on the correlation as an indicator. Users are added until the total throughput start to decrease.

$$2.\ u = argmin(\xi(i,j)) \tag{4.18}$$

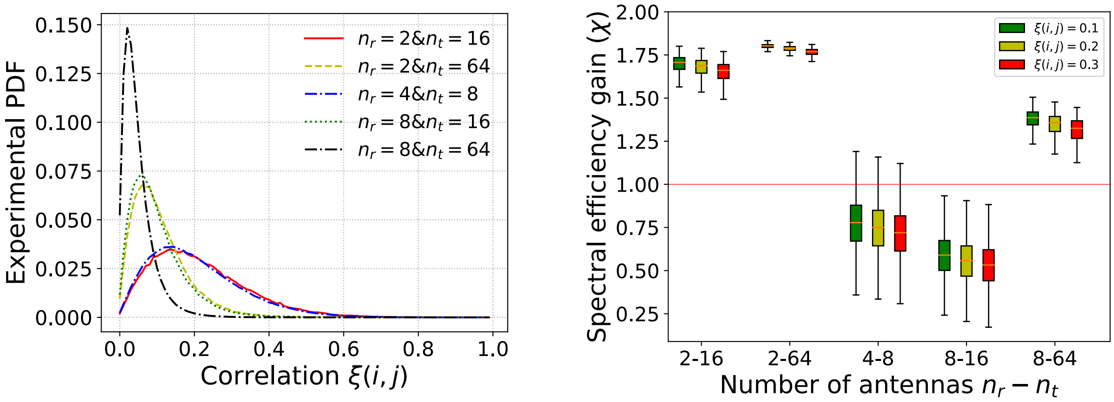
By sharing the same properties as MaxSNR, CBUS encounters the same flaws, such as the lack of fairness or energy concerns.

4.3.2 Fairness aware user selection

Proportional Fair (PF)-based user selection are proposed to overcome MaxSNR lack of fairness.

Orthogonality Probing based UserSelection (OPUS) [50] selects the first user depending on a throughput fairness criterion. This criterion considers the obtained throughput of all the users. The probability of a user being selected decreases as the user's throughput increases. All other users are selected using a strategy MaxSNR on previously probed orthogonal directions.

Weighted Users' Correlation and FairnesS (WUCFS) [10] proposes to integrate the fairness criterion in each user selection. Users that are added to the group needs to maximize a criterion based on both a fairness factor and channels' correlation. By using



(a) Correlation distribution depending on antenna numbers

(b) Spectral efficiency gain of MU-MIMO over SU-MIMO

Figure 4.2 – Algorithms impact on total system capacity.

a fairness indicator and then reducing the impact of interference, WUCFS prevents users to starve from bad geographical (high interference) or radio (strong pathloss) conditions.

4.4 Channel matrix correlation as interference management

4.4.1 Impact of the number of antennas on the correlation distribution

The number of antennas is a main parameter of Massive-MIMO. The more antennas there are, the more precise the spatial separation will be. However, due to technical limitations, such as precoding complexity or room space available for mobile devices, the number of antennas is limited. 3GPP fixes the number of antennas at the transmitter and at the receiver in the system evaluation framework. In our performance evaluation, we study different configurations to understand the impact on the correlation. In the following example we show the impact of the number of antennas, at the transmitter n_{t} and at the receiver n_{r}, on UEs correlation distribution. For each sample we compute the channel correlations of two UEs.

Figure 4.2(a) shows the empirical PDF of the correlation. The in Cumulative Distribution Function (CDF) can be easily deduced from the in Probability Density Function (PDF). Several observations can be highlighted from figure 4.2(a). The product of the number of antennas determines the correlation distribution. Two antenna configurations having the same product have a close correlation distribution: configurations $n_{\mathrm{r}} = 2$ & $n_{\mathrm{t}} = 64$ and $n_{\mathrm{r}} = 8$ & $n_{\mathrm{t}} = 16$ have a nearly similar correlation distribution. Having a greater number of antennas allows the system to have more narrow beams, resulting in a lower average correlation. With $n_{\mathrm{r}} = 8$ & $n_{\mathrm{t}} = 64$ there is a larger number of low correlation than with $n_{\mathrm{r}} = 2$ & $n_{\mathrm{t}} = 16$.

4.4.2 Impact of the correlation on throughput

The main objective when designing a resource allocation algorithm is to increase the capacity of the system. To understand if the correlation is a good indicator for a scheduling process, we study the correlation impact on throughput. For each sample, we calculate the throughput gain from the MU-MIMO allocation, that is given by 4.14.

Figure 4.2(b) shows χ values, computed in (4.14), for different antenna configurations. Each antenna configuration is represented tree boxes with 0.1, 0.2 and 0.3 correlations. Concerning the throughput gain, for example, with $n_{\mathrm{r}} = 2$ & $n_{\mathrm{t}} = 16$, between 0.1 and 0.3 correlations, less than 3% of capacity is lost on average. The Block Diagonal precoding technique is built to withstand the increase of interference. When the correlation increases, even if the spectral efficiency decreases, the precoder is able to contain the degradation. With an antenna configuration $n_{\mathrm{r}} = 2$ & $n_{\mathrm{t}} = 16$ at a correlation of 0.2 they may be more than 10% between the highest and lowest value of throughput. This difference depends on the degree of freedom available in the precoding process. To reduce the deviation the ratio $\frac{n_{\mathrm{t}}}{n_{\mathrm{r}}}$ should be larger. The main issue in this example, is the ability to predict the gain. In all configurations there is an intersection between throughput values, meaning that the same throughput gain can be experienced with different correlation values. Note that in the case of $n_{\mathrm{r}} = 8$ & $n_{\mathrm{t}} = 16$ and $n_{\mathrm{r}} = 4$ & $n_{\mathrm{t}} = 8$, it is not even profitable to schedule two UEs at the same time. They will both experience a smaller throughput than if they were scheduled in Single-User mode.

4.4.3 Mobility impact on the correlation between UEs

In order to be able to use the correlation as an indicator, the correlation should be stable in frequency and in time when users are moving. The more variable the channel, the more frequent the correlation should be updated. There is thus a feasibility issue: requiring

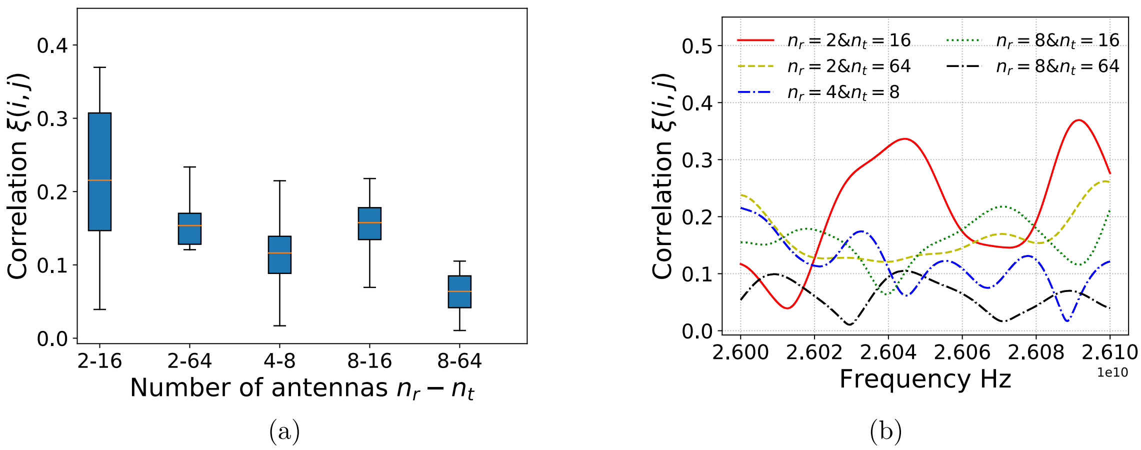


Figure 4.3 – Correlation over 100 MHz bandwidth at 26GHz

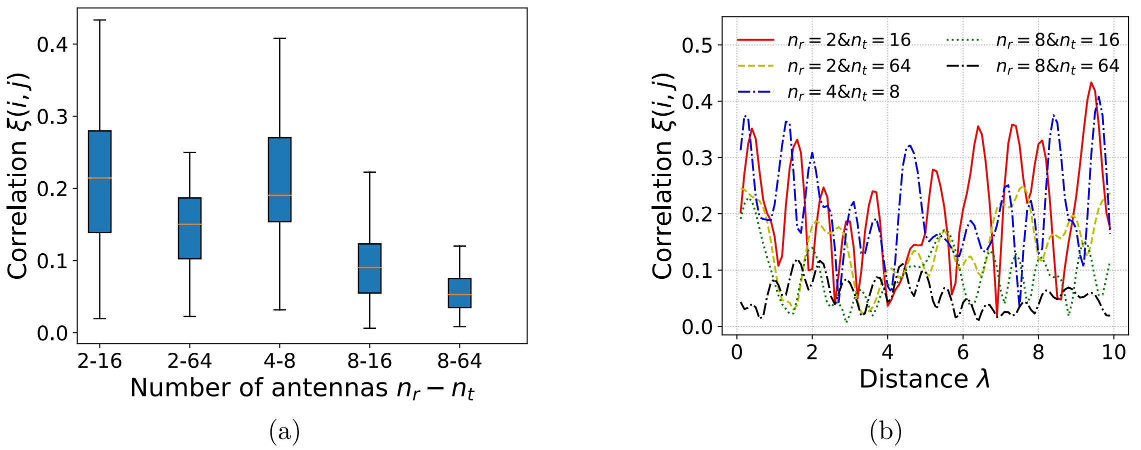


Figure 4.4 – Correlation over 10λ

more Channel State Information (CSI) feedbacks and more CPU resources. As seen in the simulation setup, we are considering two users with independent channel matrices. The results are computed on one snapshot in each configuration. One user is static during the simulation, while the other user is moving and its channel adjusts accordingly. The purpose of this experiment is to determine if the correlation is stable enough to be calculated as often as possible for two UEs during a period of time and over a given bandwidth.

Figure 4.3 shows the correlation evolution between two users depending on the frequency. Over 100 MHz bandwidth, the correlation significantly varies in amplitude but slowly in time. This shows that the variation of phase offset is not negligible over a large bandwidth. However, the rate variation is slow enough and only a few correlation calculations are needed by TS. Another variable factor is the displacement over time. Figure 4.4 shows the correlation evolution when a user moves from its original position. Distance is represented in λ, which is the wavelength, and 10λ at 26 GHz is traveled in roughly 84 ms. We can observe a high variability of the correlation with the distance changes, with a similar amplitude gap compares to figure 4.3. When the variation in frequency and distance are combined, the correlation stability might not be sufficient for an accurate calculation between scheduling decisions.

4.4.4 Discussion on the correlation as an indicator

Correlation is often used in the user selection process as an indicator to maximize the global throughput. In this study, we show the correlation distribution, the relation between correlation and throughput and the mobility impact on the correlation when using a Block Diagonal precoder. The distribution of the correlation depends on the product of the number of antennas at the transmitter and at the receiver. A higher product reduces the probability of high correlation values and diversity. The ratio between the number of receive and transmit antennas determines the variability of the throughput gain. A large number of transmitting antennas compared to number of receiver antennas makes the throughput gain more predictable. Finally, the correlation variability quickly changes over frequency and time, when a user is mobile. There is a substantial number of correlation calculations needed over time. As a summary, a user selection process, using Block Diagonal precoding, can barely benefit from a correlation indicator in different configurations, due to high variability with minimum effect on the throughput.

4.5 New algorithm proposal for user selection

We propose a solution, called ESCUS, based on previous group allocations. Our solution is based on a matrix $\mathbf{M}$ that stores the capacity values obtained when 2 users are served in the same group. This matrix is gradually built from previous allocations. ESCUS presents several advantages over correlation based solutions [9, 10, 56]. It does not rely on complex calculation for each pair of users. Indeed, correlation-based solutions must periodically update the value of correlation between users (equation 4.16) in order to correctly group users, due to the instability of correlations (see section 4.4.3). While ESCUS only relies on a $\text{card}(\mathcal{U}) \times \text{card}(\mathcal{U})$ matrix which is updated at the end of the allocation process. Compare to a full MaxSNR strategy, our solution does not increase complexity, the metrics compared are not updated during the user selection process. This solution is very flexible and will be easy to adapt to any MIMO system thanks to the channel hardening [57] consideration.

ESCUS is divided in two main phases, the exploration and the exploitation. When the system starts, users' compatibility (group quality) are unknown. During the exploration phase, ESCUS associates users in a certain order, in order to populate its database with throughput of different pairs of users. When the database completion reaches a certain level, ESCUS enters in an exploitation phase. During this phase, ESCUS selects users to be served depending on the best obtained throughput in the database. In each cases, the database is always updated if the obtained throughput evolves. In this way, users can be served with the highest compatibility, based on their previous association.

The indicator used by ESCUS is the cumulated throughput of all pairs of users in previous groups. When the number of users per group increases, the number of possible combination increases as well. Storing the obtained throughput for each combination would require a large database, which would grow with the number of users per group (a n-dimensional matrix, with n the number of users per group). The larger the database is, the longer it takes to fill it, leading to more exploration than exploitation. To avoid such constraints, ESCUS database only stores the throughput obtained by a pair of users. The data to be stored corresponds to a $\text{card}(\mathcal{U}) \times \text{card}(\mathcal{U})$ matrix. Another advantage of using a bi-dimensional matrix is its flexibility concerning the number of users per group, allowing the same database structure to be used regardless of the group size. The following example illustrates the database structure. We consider a group of 3 users $\mathcal{A} = \{a, b, c\}$. ESCUS stores the sum of the rate for each possible pair set in $\mathcal{A}$. For example, $\mathbf{M}_{a,b} = \mathbf{M}_{b,a} = R_a + R_b$.

Independently from the current phase of ESCUS, the first user of the group is always selected according to a MaxSNR strategy [16] in order to maximize the total system rate. The objective of ESCUS is only to maximize the system rate. If fairness is considered, then another first-user selection strategy can be applied, such as proportional fair [28, 29].

The first phase aims at discovering unknown shared throughput between users. This exploration phase selects secondary users, to complete the group, having the best instantaneous in SU throughput ($\mathcal{C}_i$). As stated previously, this MaxSNR strategy might be replaced by another strategy to reach different goals. During this selection, only users with unknown throughput, when associated with the first user, are selected. This process stops when the number of selected users reaches the maximum group size.

When the database contains enough information, ESCUS switches to the exploitation mode. The threshold for this transition depends on the system configuration and is set to $\frac{1}{4} \times \mathrm{card}(\mathcal{U})$ in our examples, as it was found to be efficient in all antenna configurations and allows for sufficient group diversity.

During this second phase ESCUS uses the data gathered previously. Secondary users are selected according to the best value in the database when associated with the first user. To prevent allocating low value, the selected value is compared to the mean of the database. When the value is lower, ESCUS switches to exploration in order to find a more suitable pair of users.

At the end of this selection process, the resource is allocated to selected users. After this allocation, the matrix $\mathbf{M}$ is updated with experienced throughput from each pair of users. This full process is described in Figure 4.5.

4.6 Performance evaluation

4.6.1 Simulation set up

Twenty UE, randomly distributed in the cell, are considered. The UEs channels are mutually independent. A channel for a user i is generated using the Saleh-Valenzuela model [58] extended to millimeter waves, where the Obstructed Line Of Sight (OLOS) parameters are given in table 4.1.

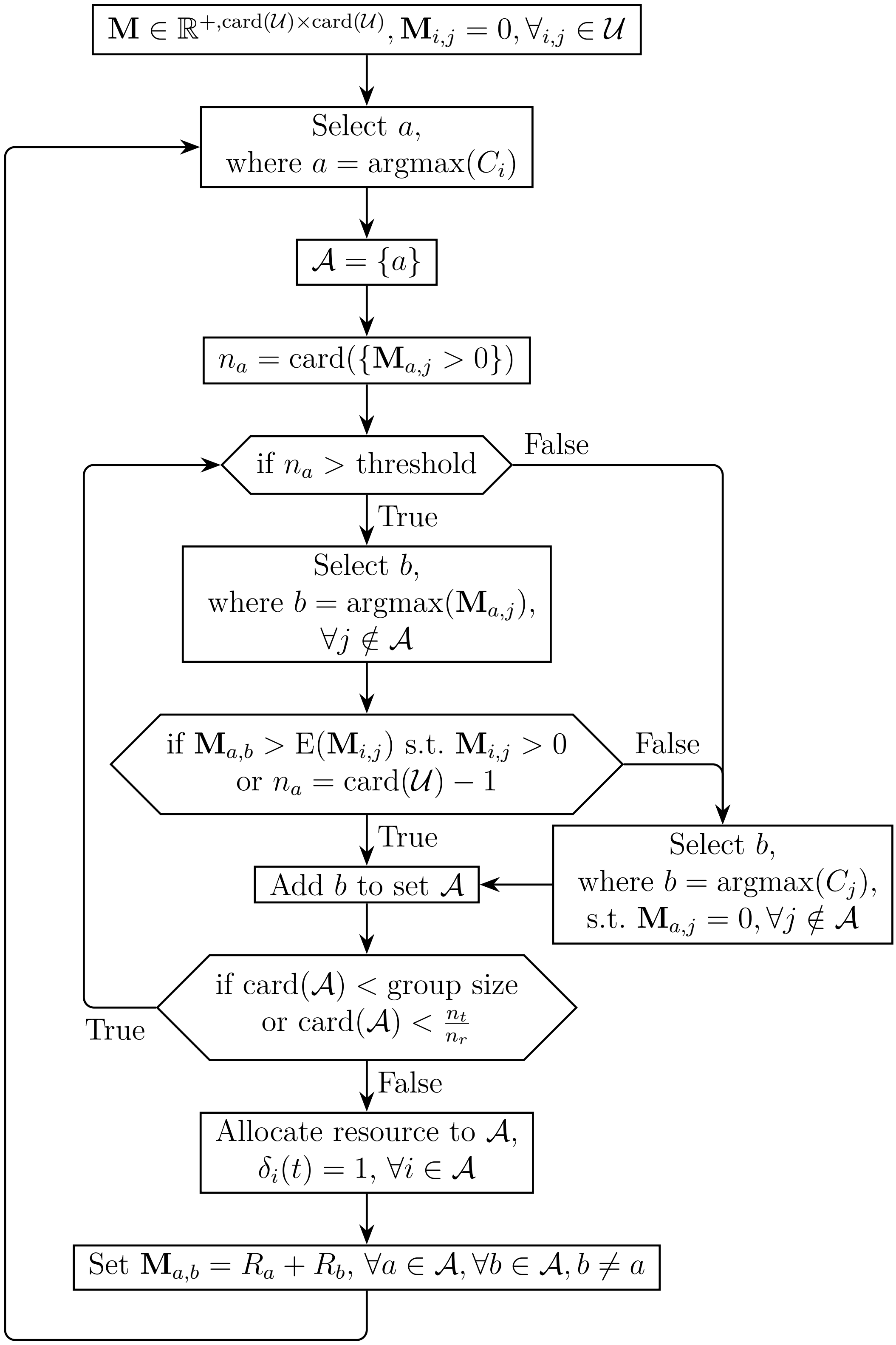
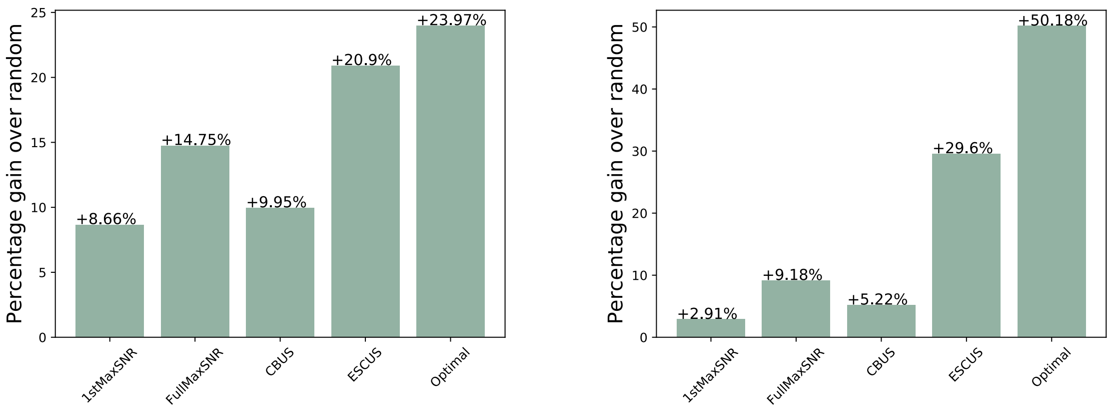


Figure 4.5 – Diagram of the proposed solution



(a) 2 UEs per group

(b) 3 UEs per group

Figure 4.6 – Algorithms impact on total system capacity.

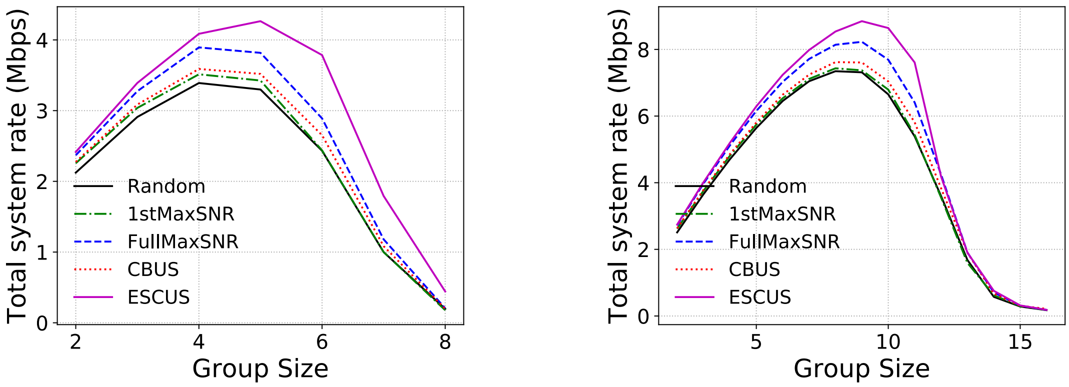
f	26	GHz	Frequency
Λ	5	ns	Cluster arrival rate
λ	1	ns	Ray arrival rate
Γ	8.7	ns	Cluster decay rate
γ	4.7	ns	Ray decay rate
σ	0.1	rad	Intra-cluster angles standard deviation

Table 4.1 – Channel model parameters, from [58]

We consider one sector in a typical 3-sector configuration: the angle between the terminal and the BS is between 0° and 120° horizontally and between −45° and 45° vertically. Antennas arrays at the BS and on the user side are Uniform Linear Array (ULA), with $\frac{\lambda}{2}$ distance between antenna elements, where λ is the wavelength of the central frequency. We assume that the transmitter has a full knowledge of the channel for each UE.

ESCUS is compared to several simple and known algorithms. As a representative correlation based algorithm, CBUS [9] presents the advantage of focusing only on system rate maximization, as opposed to fairness-oriented strategies [10, 50] (which are not directly comparable to our solution):

— 1stMaxSNR, is a MaxSNR for the first user and then all following users of the group are chosen using a random function.



(a) Antenna configuration n_r:4 and n_t:32 (b) Antenna configuration n_r:4 and n_t:64

Figure 4.7 – Group sizes impact on the system rate (for a 2 MHz bandwidth)

- FullMaxSNR, is a MaxSNR for the first user and then all following users of the group are chosen depending on their Signal to Noise Ratio (SNR).
- CBUS, is a MaxSNR for the first user and then all following users of the group are chosen according to the lowest correlation with the first user. This solution from the literature is described in [9].
- ESCUS, is our proposed solution and is described in Section 4.5
- Optimal, is the exhaustive search of the best group. This optimal solution guarantees the maximum system throughput.

Two types of performance evaluation were conducted, either with a small system or with larger systems (regarding the number of users and possible group sizes). A first study with a small system allowed us to compare our solution with all previously described strategies, including the optimal one. A second study is focused on Massive-MIMO systems, where large groups of users can be simultaneously allocated. However, maximizing the number of users per group did not always provide the best system capacity, due to a strong increase in inter-group interference. Another limiting factor is the number of RF chains, which defines the limit of group size. A 3GPP technical report [59] sets the maximum number of RF chains at 12. At this group size, it is not possible to compute an optimal solution in reasonable time, due to its polynomial complexity depending on the number of users per group. The optimal strategy is then excluded in section 4.6.3. Results are obtained with the mean of one hundred frame samples, to avoid non-representative statistical event.

4.6.2 Performance evaluation with small number of users per group

Figure 4.6 shows the percentage gain over a full random solution for different algorithms on the total system rate. With two users per group (see Figure 4.6(a)) the gain for all the solutions is significant over a random selection. The use of MaxSNR increases the capacity by 8.7% when used only on the first user and 14.8% when used on both users. The use of the correlation on the second user (CBUS) only increases the gain by 1.3% compared to a random second user. Our ESCUS solution based on previous allocations selection, represents an important gain over all the other solutions, 6.2% over full MaxSNR, and is only 3.1% below the optimal strategy.

With three users per group (see Figure 4.6(b)) the gain compared to random allocation is lower for all the classical solutions. MaxSNR applied only on the first user is only 2.9% higher than random allocation and 9.2% when applied to all users. The solution using the correlation, CBUS, benefits from the high interference environment and is now 2.3% higher than a random allocation on the following users. However its performance is still 4% below full system MaxSNR allocation. Our solution is the only one with an increase in its performance, with 20.4% higher than full MaxSNR, but is now 20.6% below the optimal strategy.

When the number of users per group increases, the level of interference becomes harder to manage. ESCUS allows to maintain its performance when we increase the group size thanks to the knowledge of previous allocations, whereas other solutions see their performance decrease due to their lack of consideration for group quality. Finally, the gap between our solution and the optimal selection for three users per group is still large and further work is needed to achieve the optimal performance.

4.6.3 Performance evaluation with large number of users per group

In this section, results with large user group sizes are studied. Users are chosen among a set of 60 users randomly distributed in the cell, thus increasing the group diversity compared to the previous setup.

Figure 4.7 shows the evolution of the total system rate depending on the group size for two antenna configurations. All algorithms show a slow increase up to their best system rate and a fast decrease. Communicating simultaneously with an increasing number of users, increases the overall data rate, but also causes interference between users in the same group. When the number of users reaches a given size, interference has a greater impact than the gain from MU-MIMO, leading to a decrease in the total system rate. Depending on the user selection algorithm, this achievable size may differ.

Algorithm performance matches those obtained in section 4.6.2: random allocation is the worst solution, followed by 1stMaxSNR and then by CBUS [9] (solution from the SoTA based on the often used channel matrix correlation) and finally by FullMaxSNR and our ESCUS solution.

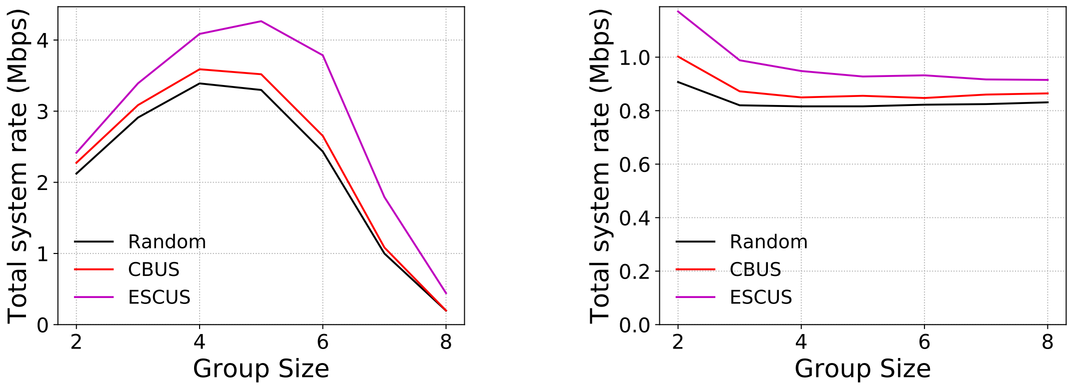
As shown in Figure 4.7(a) the system experiences a strong gain by using ESCUS over FullMaxSNR with an increase of 11.7% with a group size of 5 users and an increase of 30.9% with a group size of 6 users. In Figure 4.7(b) the gain over FullMaxSNR is slightly less with a gain of 7.5% with a group size of 9 users and 12.4% with a group size of 10 users.

Compared to correlation based algorithms, ESCUS performs better for all group configurations with a higher total system rate. The results in Figure 4.7(a) show an increase of 18.8% for ESCUS compared to CBUS regarding the highest value of the total system rate for any group size with n_r:4 and n_t:32. Figure 4.7(b), with n_r:4 and n_t:64, shows a lower increase of 13.4%.

4.6.4 Impact of the precoder on user selection strategies

The precoder BD has a strong impact on interference mitigation. Because ESCUS does not take into account the channel matrix correlation, one can imagine that when it is associated with a precoder not oriented towards interference management, performance decreases. To avoid this misunderstanding, we propose to compare ESCUS to CBUS and a random distribution, using the precoder BD and the precoder SVD. While BD focuses on interference management, SVD focuses only on maximizing SU throughput, without considering other users within the group.

Figure 4.8 shows the total system rate evolution for both a BD precoder and a SVD precoder, depending on the number of users per group. The results of figure 4.8(a) are similar to the results of figure 4.7(a) and are described in section 4.6.3. These results show that ESCUS outperforms CBUS in all user configurations, while CBUS is slightly better than a random distribution. With the SVD precoder, as shown in figure 4.8(b), results are confirmed and ESCUS is still presenting better performance than CBUS. Note that the system throughput is lower using SVD than BD, indeed the focus on MU of the precoder



(a) Antennas n_r:4 and n_t:32 with BD

(b) Antennas n_r:4 and n_t:32 with SVD

Figure 4.8 – System rate depending on the precoder (for a 2 MHz bandwidth)

BD allows a better interference cancellation and thus a better system throughput when the number of users increases.

4.7 Conclusion on Massive-MIMO resource allocations

New techniques such as MU-MIMO greatly improve performance over Single-User techniques. To enable the BS to smartly select users, user selection algorithms are often based on users channel matrix correlation. In this chapter we showed that the correlation as an indicator was not suitable for our context. We then propose a solution called ESCUS based on a new principle which uses previous group allocations as an indicator. This indicator depends on the obtained throughput between pairs of users and does not increase in complexity when the number of users per group increases. Thanks to channel stability due to the channel hardening, this indicator is stable in time and frequency. Results show a strong increase in the system throughput compare to correlation based algorithm. This solution is easier to calculate than correlation based strategies when the number of users per group increases. In terms of performance ESCUS presents an important improvement compared to classical solutions. As the potential throughput of the MaxSNR is used as the main indicator for opportunistic Single Input Single Output (SISO) algorithms, the indicator of ESCUS could be used to transpose schedulers from a SISO to a MU-MIMO context. Due to the large number of antennas required for MU-MIMO, the algorithm is more suitable for a micro cell deployment. However, ESCUS is easily adaptable in a multi-cell context on top of the Multi-Cell Pre-Scheduler (MCPS) described in chapter 3.

Chapter 5

Conclusion

5.1 Summary of the main results

The results presented in this thesis focus on resource allocation problems encountered in modern wireless networks. These solutions are designed for three different wireless network systems with their own specificities and challenges.

5.1.1 Dynamic SISO scheduler

Chapter 2 introduces the resource allocation principle in classic Single Input Single Output (SISO) systems, these systems consider a user to be covered by only one Base Station (BS). We then describe some of the best-known schedulers and their main characteristics. In these schedulers, spectral efficiency and power consumption are often opposed. To overcome this problem, we propose a new scheduler called Dynamic Trade-off (DT). DT addresses this issue using a dynamic trade-off between spectral efficiency and energy consumption. Thanks to an adequate heuristic, DT adapts its priorities to the system load and provides the best performance compared to dedicated schedulers in their field of interest.

5.1.2 Multi-cell pre-scheduler

In chapter 3, we extend the resource allocation description to a multi-cell context. In such a system, coordination between cells is primordial and the user's distribution algorithms aim to optimally distribute the load between cells and provide a homogeneous Quality of Service (QoS). An unbalanced distribution can lead to unsatisfied users while some system resources are not used. Our solution called Multi-Cell Pre-Scheduler (MCPS) attempts to maximize the efficiency of the micro cell to discharge as much as possible the macro cell. This technique allows the macro cell to serve more users who are not covered by the micro cells, and the performance evaluation shows a significant increase in system capacity. Another major advantage of the MCPS is its compatibility with existing schedulers, maintaining their features.

5.1.3 Indicator for MU-Massive-MIMO resource allocation

Finally, in Chapter 4, we introduced a new paradigm in the allocation of resources, the Multiple Users Multiple Input Multiple Output (MU-MIMO). Indeed, when schedulers traditionally distribute time and frequency resources to a single user, with the massive-Multiple Input Multiple Output (MIMO) more users can receive that same resource. However, sharing the same resource can introduce strong interference between users and their selection must be done correctly. We first proposed an evaluation of the channel matrix correlation as an indicator for user selection, this indicator is frequently used in this context. The results show that it is not well suited for this purpose due to its lack of precision and stability. We then introduced a new algorithm, called Efficient System Capacity User Selection (ESCUS), which uses past allocations as an indicator. The performance evaluation shows a significant increase in system capacity compared to a correlation-based solution.

5.2 Perspectives

Wireless networks are constantly evolving towards greater functionality and efficiency. In this respect, our contributions could be extended to different system configurations or scenarios. Two main axes that emerge from our work and which could be treated as future work are presented below, along with a long term objective.

5.2.1 Reinforcement learning based MU-Massive-MIMO algorithm

Our solution ESCUS presented in chapter 4 is divided into two parts. The first part is to explore the possible combinations of users among all users. The second part is to take advantage of experienced groups and increase overall capacity. This approach is very close to a known algorithm of Reinforcement learning (RL) called multi-armed bandit [60, 61]. Future work could focus on adapting ESCUS to the RL problem and benefit from related knowledge.

5.2.2 Multi-cell Massive-MIMO

Since MU-MIMO is mainly associated with small high-frequency cells, it is supposed to be included within a macro cellular coverage area, similar to the systems described in Chapter 3. However, due to the high complexity of massive-MIMO and multi-cell alone, when combined, the complete system will be very complex to manage. Therefore, future work could focus on designing accessible solutions for a massive-MIMO multi-cell system.

5.2.3 Service-oriented multi-cell user placement

Due to the variety of uses of wireless network (smart cities, autonomous cars, industrial IoT...) services could also be included in a multi-cell resource allocation process. Macro and micro cells do not offer the same properties and users must be placed according to their current applications. Fast-moving users, typically autonomous cars, may benefit from the large coverage area of the macro cell, while static users who download large files may benefit from the wide bandwidth of the micro cell. Considering user services in the user selection process dramatically increases the variability due to the large number of existing and future services. These service-oriented user placement algorithms could be combined with cached content delivery strategies [62] to reduce latency. An ideal solution would be to provide a fair QoS to the users of the system, regardless of the services they use, while guaranteeing the necessary resources for critical applications.

BIBLIOGRAPHY

[1] Cisco, « Cisco Annual Internet Report (2018-2023) », *in*: *White paper* (Mar. 2020).

[2] Yaser Al Mtawa, Anwar Haque, and Bassel Bitar, « The mammoth Internet: Are we ready? », *in*: *IEEE Access* 7 (2019), pp. 132894–132908.

[3] Jacob Poushter et al., « Smartphone ownership and internet usage continues to climb in emerging economies », *in*: *Pew research center* 22.*1* (2016), pp. 1–44.

[4] Marceau Coupechoux, Ivan Davydov, and Stefano Iellamo, « Tabu search heuristic for competitive base station location problem », *in*: *Operations Research Proceedings 2015*, Springer, 2017, pp. 325–330.

[5] M. Ezzaouia et al., « Autonomous and dynamic inter-cell interference coordination techniques for future wireless networks », *in*: *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2017, pp. 1–8.

[6] J. Pedreno-Manresa et al., « Improved User Experience by Dynamic Service Handover and Deployment on 5G Network Edge », *in*: *2019 21st International Conference on Transparent Optical Networks (ICTON)*, 2019, pp. 1–4.

[7] Yu-Han Chen et al., « Resource allocation with CoMP transmission in ultra dense cloud-based LTE small cell networks », en, *in*: *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Montreal, QC: IEEE, Oct. 2017, pp. 1–5, DOI: `10.1109/PIMRC.2017.8292432`, (visited on 03/06/2020).

[8] Ling Liu et al., « Load Aware Joint CoMP Clustering and Inter-Cell Resource Scheduling in Heterogeneous Ultra Dense Cellular Networks », en, *in*: *IEEE Transactions on Vehicular Technology* 67.*3* (Mar. 2018), pp. 2741–2755, ISSN: 0018-9545, 1939-9359, DOI: `10.1109/TVT.2017.2773640`, (visited on 03/06/2020).

[9] Tianxiong Ji et al., « Low Complex User Selection Strategies for Multi-User MIMO Downlink Scenario », en, *in*: *2007 IEEE Wireless Communications and Networking Conference*, Kowloon, China: IEEE, 2007, pp. 1532–1537, ISBN: 978-1-4244-0658-6, DOI: `10.1109/WCNC.2007.289`, URL: `http://ieeexplore.ieee.org/document/4224534/` (visited on 11/13/2018).

[10] Zhao Li, Peifeng Li, and Kang G. Shin, « MU-MIMO downlink scheduling based on users' correlation and fairness », en, *in*: *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, Washington DC, USA: IEEE, Sept. 2014, pp. 407–412, ISBN: 978-1-4799-4912-0, DOI: `10.1109/PIMRC.2014.7136199`, URL: `http://ieeexplore.ieee.org/document/7136199/` (visited on 11/13/2018).

[11] Cédric Gueguen and Malo Manini, « Fairness-Energy-Throughput Optimized Trade-off in Wireless Networks », *in*: *ICNC 2018-International Conference on Computing, Networking and Communications*, IEEE, 2018, pp. 1–8.

[12] Cédric Gueguen and Malo Manini, « Dynamic tradeoff between energy and throughput in Wireless 5G Networks », *in*: *Wireless communications and mobile computing* 2018 (2018).

[13] H. SHI et al., « Fairness in Wireless Networks:Issues, Measures and Challenges », *in*: *IEEE Communications Surveys Tutorials* 16.*1* (First 2014), pp. 5–24, ISSN: 1553-877X, DOI: `10.1109/SURV.2013.050113.00015`.

[14] Raj Jain, Arjan Durresi, and Gojko Babic, « Throughput fairness index: An explanation », *in*: *ATM Forum contribution*, vol. 99, 45, 1999.

[15] Rajendra K Jain, Dah-Ming W Chiu, William R Hawe, et al., « A quantitative measure of fairness and discrimination », *in*: *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA* (1984).

[16] Cheong Yui Wong et al., « Multiuser OFDM with adaptive subcarrier, bit, and power allocation », *in*: *IEEE Journal on selected areas in communications* 17.*10* (1999), pp. 1747–1758.

[17] Xudong Wang and Weidong Xiang, « An OFDM-TDMA/SA MAC Protocol with QoS Constraints for Broadband Wireless LANs », *in*: *ACM/Springer Wireless Networks* 12.*2* (2006), pp. 159–170.

[18] Marcos Rubio del Olmo et al., « Energy-efficient user scheduling algorithm for LTE networks », *in*: *Proc. 5th Joint WIC/IEEE SP Symp. Inf. Theory Signal Process. Benelux (WIC)*, 2015, pp. 1–8.

[19] Cedric Gueguen, « Opportunistic Energy Aware Scheduler for Wireless Networks », *in*: *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, IEEE, 2013, pp. 1–5.

[20] Tijs Van Dam and Koen Langendoen, « An adaptive energy-efficient MAC protocol for wireless sensor networks », *in*: *Proceedings of the 1st international conference on Embedded networked sensor systems*, ACM, 2003, pp. 171–180.

[21] John David Parsons, *The Mobile Radio Propagation Channel*, Wiley, 1992.

[22] John G. Proakis, *Digital Communications*, New York: McGraw-Hill, 1995.

[23] Thomas E Truman and Rober W Brodersen, « A measurement-based characterization of the time variation of an indoor wireless channel », *in*: *Proceedings of ICUPC 97-6th International Conference on Universal Personal Communications*, IEEE, 1997, pp. 25–32.

[24] John Nagle, « On packet switches with infinite storage », *in*: *IEEE Transactions on Communications* 35.*4* (April 1987), pp. 435–438.

[25] A Kuurne and A.P Miettinen, « Weighted round robin scheduling strategies in (E)GPRS radio interface », *in*: *Proc. IEEE Int. Vehicular Technology Conference (VTC)*, vol. 5, Sept. 2004, pp. 3155–3159.

[26] Viet L. Do and Kenneth Y. Yun, « An efficient frame-based scheduling algorithm: credit round robin », *in*: *Workshop on High Performance Switching and Routing, 2003, HPSR*, June 2003, pp. 103–110.

[27] Gokhan Mergen and Lang Tong, « Random scheduling medium access for wireless ad hoc networks », *in*: *Proc. IEEE Int. Conf. on MILCOM*, vol. 2, October 2002, pp. 868–872.

[28] Pramod Viswanath, David N. C. Tse, and Rajiv Laroia, « Opportunistic Beamforming Using Dumb Antennas », *in*: *IEEE Transactions on Information Theory* 48 (June 2002), pp. 1277–1294.

[29] Hoon Kim et al., « An efficient scheduling algorithm for QoS in wireless packet data transmission », *in*: *Proc. IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, vol. 5, Sept. 2002, pp. 2244–2248.

[30] Wang Anchun et al., « Dynamic resource management in the fourth generation wireless systems », *in*: *Proc. IEEE Int. Conference on Communication Technology (ICCT)*, vol. 2, April 2003, pp. 1095–1098.

[31] Patrick Svedman, Sarah Kate Wilson, and Björn Ottersten, « A QoS-aware proportional fair scheduler for opportunistic OFDM », *in*: *Proc. IEEE Int. Vehicular Technology Conference (VTC)*, vol. 1, Sept 2004, pp. 558–562.

[32] Hoon Kim et al., « A Proportional Fair Scheduling for Multicarrier Transmission Systems », *in*: *Proc. IEEE Int. Vehicular Technology Conference (VTC)*, vol. 2, Sept 2004, pp. 409–413.

[33] Cedric Gueguen and Sebastien Baey, « A Fair MaxSNR Scheduling Scheme for Multiuser OFDM Wireless Systems », *in*: *Proc. IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2009.

[34] S. Tanwir and H. Perros, « A Survey of VBR Video Traffic Models », *in*: *IEEE Communications Surveys Tutorials* 15.*4* (Fourth 2013), pp. 1778–1802, ISSN: 1553-877X, DOI: 10.1109/SURV.2013.010413.00071.

[35] Jeffrey G. Andrews, « Seven ways that HetNets are a cellular paradigm shift », *in*: *IEEE Communications Magazine* 51.*3* (Mar. 2013), pp. 136–144, ISSN: 0163-6804, DOI: 10.1109/MCOM.2013.6476878, URL: http://ieeexplore.ieee.org/document/6476878/ (visited on 10/06/2017).

[36] Call:H2020-ICT-2016-2, « Deliverable D2.1 Scenarios, KPIs, use cases and baseline system evaluation », *in*: *E2E-aware Optimizations and advancements for Network Edge of 5G New Radio (ONE5G)* (Dec. 2016).

[37] O. Bulakci et al., « RAN Moderation in 5G Dynamic Radio Topology », *in*: *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, 2017, pp. 1–4, DOI: 10.1109/VTCSpring.2017.8108537.

[38] Nicolas Guérin et al., « High system capacity pre-scheduler for multi-cell wireless networks », *in*: *Wireless Networks* (2020), pp. 1–13.

[39] One5G, « Deliverable D3.2Recommended Multi-Service Performance Optimization Solutions for Improved E2E Performance », *in*: *E2E-aware Optimizations and advancements for Network Edge of 5G New Radio(ONE5G)* (May 2019).

[40] Tareq M. Shami et al., « User-centric JT-CoMP clustering in a 5G cell-less architecture », en, *in*: *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Bologna, Italy: IEEE, Sept. 2018, pp. 177–181, ISBN: 978-1-5386-6009-6, DOI: `10.1109/PIMRC.2018.8580792`, (visited on 06/12/2020).

[41] Seung-Yeon Kim and Choong-Ho Cho, « Call Blocking Probability and Effective Throughput for Call Admission Control of CoMP Joint Transmission », en, *in*: *IEEE Transactions on Vehicular Technology* (2016), pp. 1–1, ISSN: 0018-9545, 1939-9359, DOI: `10.1109/TVT.2016.2541172`, (visited on 06/12/2020).

[42] M. Andrews et al., « Providing quality of service over a shared wireless link », *in*: *IEEE Communications Magazine* 39.*2* (Feb. 2001), pp. 150–154, ISSN: 0163-6804, DOI: `10.1109/35.900644`.

[43] Ali A Zaidi et al., « Waveform and numerology to support 5G services and requirements », *in*: *IEEE Communications Magazine* 54.*11* (2016), pp. 90–98.

[44] Sébastien Baey, « Modeling MPEG4 video traffic based on a customization of the DBMAP », *in*: *Proc. Int. Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)*, 2004, pp. 705–714.

[45] A. K. A. Tamimi, R. Jain, and C. So-In, « Modeling and Prediction of High Defninition Video Traffic: A Real-World Case Study », *in*: *Proc. Second International Conferences on Advances in Multimedia*, June 2010, pp. 168–173, DOI: `10.1109/MMEDIA.2010.10`.

[46] ETSI 3rd Generation Partnership Project, *5G; Service requirements for next generation new services and markets, 3GPP TS 22.261 version 15.5.0 Release 15*, 2018-07.

[47] Mohammed Yazid Lyazidi, Nadjib Aitsaadi, and Rami Langar, « A dynamic resource allocation framework in LTE downlink for Cloud-Radio Access Network », *in*: *Computer Networks* 140 (2018), pp. 101–111.

[48] Cheng-Xiang Wang et al., « Cellular architecture and key technologies for 5G wireless communication networks », en, *in*: *IEEE Communications Magazine* 52.*2* (Feb. 2014), pp. 122–130, ISSN: 0163-6804, DOI: `10.1109/MCOM.2014.6736752`, URL: `http://ieeexplore.ieee.org/document/6736752/` (visited on 02/25/2019).

[49] Emre Telatar, « Capacity of Multi-antenna Gaussian Channels: Capacity of Multi-antenna Gaussian Channels », en, *in*: *European Transactions on Telecommunications* 10.*6* (Nov. 1999), pp. 585–595, ISSN: 1124318X, DOI: `10.1002/ett.4460100604`, URL: `http://doi.wiley.com/10.1002/ett.4460100604` (visited on 01/22/2019).

[50] Xiufeng Xie and Xinyu Zhang, « Scalable user selection for MU-MIMO networks », *in*: *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, IEEE, 2014, pp. 808–816.

[51] Ahmed Alkhateeb and Robert W. Heath, « Frequency Selective Hybrid Precoding for Limited Feedback Millimeter Wave Systems », *in*: *IEEE Transactions on Communications* 64.*5* (May 2016), pp. 1801–1818, ISSN: 0090-6778, DOI: `10.1109/TCOMM.2016.2549517`, URL: `http://ieeexplore.ieee.org/document/7448873/` (visited on 04/05/2018).

[52] Federico Boccardi and Howard Huang, « A Near-Optimum Technique using Linear Precoding for the MIMO Broadcast Channel », en, *in*: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, Honolulu, HI: IEEE, Apr. 2007, pp. III–17–III–20, ISBN: 978-1-4244-0727-9, DOI: `10.1109/ICASSP.2007.366461`, URL: `http://ieeexplore.ieee.org/document/4217635/` (visited on 01/28/2019).

[53] Md Hashem Ali Khan et al., « A simple block diagonal precoding for multi-user MIMO broadcast channels », *in*: *EURASIP Journal on Wireless Communications and Networking* 2014.*1* (2014), p. 95.

[54] Adam Mohamed Ahmed Abdo et al., « MU-MIMO Downlink Capacity Analysis and Optimum Code Weight Vector Design for 5G Big Data Massive Antenna Millimeter Wave Communication », en, *in*: *Wireless Communications and Mobile Computing* 2018 (2018), pp. 1–12, ISSN: 1530-8669, 1530-8677, DOI: `10.1155/2018/7138232`, URL: `https://www.hindawi.com/journals/wcmc/2018/7138232/` (visited on 10/02/2018).

[55] Nicolai Czink et al., « Spatial separation of multi-user MIMO channels », *in*: *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, IEEE, 2009, pp. 1059–1063.

[56] Yukiko Shimbo et al., « Control Overhead Reduction Method Employing Frequency Correlation for MU-MIMO-OFDM THP with User Scheduling », en, *in*: *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, Porto: IEEE, June 2018, pp. 1–5, ISBN: 978-1-5386-6355-4, DOI: `10.1109/VTCSpring.2018.8417814`, URL: `https://ieeexplore.ieee.org/document/8417814/` (visited on 03/18/2019).

[57] Matthieu Roy et al., « MIMO Channel Hardening for Ray-based Models », *in*: *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, IEEE, 2018, pp. 1–7.

[58] Carl Gustafson et al., « On mm-Wave Multipath Clustering and Channel Modeling », en, *in*: *IEEE Transactions on Antennas and Propagation* 62.*3* (Mar. 2014), pp. 1445–1455, ISSN: 0018-926X, 1558-2221, DOI: `10.1109/TAP.2013.2295836`, URL: `http://ieeexplore.ieee.org/document/6691924/` (visited on 07/09/2018).

[59] 3GPP, *Study on New Radio (NR) access technology*, Technical report (TR) 38.912, Version 15.0.0, 3rd Generation Partnership Project (3GPP), July 2018, URL: `https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3059`.

[60] Michael N Katehakis and Arthur F Veinott Jr, « The multi-armed bandit problem: decomposition and computation », *in*: *Mathematics of Operations Research* 12.*2* (1987), pp. 262–268.

[61] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer, « Finite-time analysis of the multiarmed bandit problem », *in*: *Machine learning* 47.*2-3* (2002), pp. 235–256.

[62] Nhan Nguyen-Thanh et al., « Multimedia content popularity: Learning and recommending a prediction method », *in*: *2018 IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2018, pp. 1–7.

List of publications

List of international published papers

- [1] Cédric Gueguen and Malo Manini, « Fairness-Energy-Throughput Optimized Trade-off in Wireless Networks », *in: ICNC 2018-International Conference on Computing, Networking and Communications, IEEE, 2018, pp. 1–8.*
- [2] Cédric Gueguen and Malo Manini, « Dynamic tradeoff between energy and throughput in Wireless 5G Networks », *in: Wireless communications and mobile computing 2018 (2018).*
- [3] Malo Manini, Cédric Gueguen, Rodolphe Legouable and Xavier Lagrange, « Study of MIMO Channel Matrices Correlation to Optimize Resource Allocation Algorithms in Multi-Users 5G. », *in: 2019 12th IFIP Wireless and Mobile Networking Conference (WMNC), IEEE, Sept. 2019.*
- [4] Nicolas Guérin, Malo Manini, Rodolphe Legouable and Cédric Gueguen, « High system capacity pre-scheduler for multi-cell wireless networks », *in: Wireless Networks (2020), pp. 1–13.*

List of national published papers

- [1] Malo Manini, Cédric Gueguen, Xavier Lagrange and Rodolphe Legouable, « Etude de la Corrélation des Canaux Massive MIMOpour l'Optimisation d'Algorithmes d'Allocation deRessources Multi-Users 5G », *in: CoRes 2019.*

Participation in a European project

- [1] One5G, « Deliverable D3.2Recommended Multi-Service Performance Optimization Solutions for Improved E2E Performance », *in: E2E-aware Optimizations and advancements for Network Edge of 5G New Radio(ONE5G) (May 2019).*SS

Acronyms

5G fifth generation. 6, 43, 44, 47, 48, 57

BD Block Diagonal. 72, 73, 86, 87

BER Bit Error Rate. 18, 38

BO Buffer Occupancy. 25, 52

BS Base Station. 8–11, 21, 22, 29, 38–40, 43, 45–48, 51, 52, 56, 57, 63, 64, 67, 69–71, 83, 87, 89

CBR Constant Bit Rate. 28, 29

CBUS Correlation Based User Selection. 75, 83–86

CDF Cumulative Distribution Function. 76

CF Compensation Factor. 24

CoMP Coordinated MultiPoint. 44–46

CSI Channel State Information. 25, 46, 79

DT Dynamic Trade-off. 3, 17, 23–27, 32, 33, 35, 39–42, 89

eNB evolved Node B. 17

ESCUS Efficient System Capacity User Selection. 69, 80, 82–88, 90

FETOT Fairness-Energy-Throughput Optimized Trade-off. 30, 32, 39, 40

gNB next generation Node B. 43, 70, 71

HCS Hierarchical Cell Structure. 43

HD High Definition. 9

KPI Key Performance Indicator. 4, 15, 24, 57, 63, 64

MaxSNR Maximum Signal to Noise Ratio. 20–25, 28, 30, 39–41, 49, 50, 52, 58, 62, 63, 67, 74, 75, 80, 82–87

MCPS Multi-Cell Pre-Scheduler. 4, 6, 44, 45, 50–59, 62–64, 67, 88–90

MIMO Multiple Input Multiple Output. 5, 11, 69, 72, 76, 80, 84, 87, 90, 91

MU Multiple Users. 4, 69, 72–74, 86

MU-MIMO Multiple Users Multiple Input Multiple Output. 3, 5, 11, 12, 69, 72–77, 86, 87, 90, 91

OEA Opportunistic Energy Aware. 22, 23, 26, 30, 39–41, 62, 63

OFDM Orthogonal Frequency-Division Multiplexing. 18, 46

OLOS Obstructed Line Of Sight. 82

OPUS Orthogonality Probing based UserSelection. 75

PDF Probability Density Function. 76

PF Proportional Fair. 21–23, 25, 26, 30, 39–41, 62, 63, 75

PPF Power-based Proportional Fairness. 22

PRB Physical Resource Block. 46

QAM Quadrature Amplitude Modulation. 19

QoE Quality of Experience. 9, 10, 16, 28, 47, 49, 50, 63, 67

QoS Quality of Service. 6, 11, 15–17, 20, 21, 23, 27, 28, 30, 32, 33, 35, 36, 39, 41, 89, 91

RA Random Access. 20

RB Resource Block. 17–22, 25, 26, 29, 38, 52, 74

RF Radio Frequency. 9, 71, 84

RL Reinforcement learning. 90

RR Round Robin. 19, 20, 29, 38–40, 49, 52, 62, 63, 67

RU Resource Unit. 15, 18, 19, 21, 24–26, 31, 32, 36, 40, 41, 46–52, 54, 70, 71, 73, 74

SISO Single Input Single Output. 10, 15, 18, 42, 87, 89

SNR Signal to Noise Ratio. 22, 83

SoTA State of The Art. 3–5, 19, 21, 39, 41, 45, 49, 57, 58, 74, 75, 86

SU Single User. 4, 72, 74, 82, 86

SU-MIMO Single User Multiple Input Multiple Output. 72, 76

SVD Singular-Value Decomposition. 72, 73, 86, 87

TS Time Slot. 18, 19, 22, 25–27, 29, 32, 38, 41, 46, 71, 79

UE User Equipment. 4, 5, 17, 18, 50, 70–74, 76, 77, 79, 82, 83

ULA Uniform Linear Array. 83

VBR Variable Bit Rate. 28, 38, 48

WUCFS Weighted Users' Correlation and FairnesS. 75, 76

Titre : Allocation de ressources et ordonnancement dans les réseaux de $5^{ème}$ génération.

Mot clés : Allocation de ressources, réseau sans-fil, MU-MIMO, 5G, QoS

Résumé : L'augmentation du nombre d'utilisateurs des réseaux sans-fil et la diversification de leurs usages amène à faire évoluer les méthodes de gestion des ressources. Cette thèse porte sur les techniques d'allocation des ressources dans les réseaux de 5G.

Dans le contexte des systèmes classiques de réseaux sans-fil, nous proposons un algorithme d'allocation de ressources dont l'objectif est de garantir équitablement le meilleur service aux utilisateurs. Lorsque la charge de la cellule est suffisament faible pour garantir un service suffisant, l'algorithme réoriente dynamiquement ses priorités vers l'économie d'énergie. Ce comportement permet un compromis efficace entre la capacité et la consommation energétique à différents niveaux de charge.

Afin d'étendre la capacité des réseaux, l'ajout de nouvelles cellules permet d'élargir la bande passante et de reduire l'atténuation du signal liée à la distance. Nous présentons un algorithme de répartition des utilisateurs dans un contexte multi-cellulaire qui intervient avant l'étape d'allocation de ressources. Cette répartition aura de fortes répércussions sur l'équilibre des charges des cellules et sur la qualité de service générale du système.

Le nombre de cellules dans un secteur est limité par sa géographie. Le Massive-MIMO permet d'accroître les fonctionnalités des cellules en permettant une directivité de l'energie et ainsi ajoute la composante spatiale à l'allocation de ressources. Nous proposons un nouvel indicateur de compatibilité spatiale des utilisateurs en se basant sur les allocations passées. Une fois intégré dans un algorithme d'allocation, il profite des capacités supérieures du Massive-MIMO.

Title: Resource allocation and scheduling in 5^{th} generation networks.

Keywords: Resource allocation, wireless network, MU-MIMO, 5G, QoS

Abstract: The increasing number of wireless network users and the diversification of their usage call for an evolution of the resource management methods. This thesis is based on 5G resource allocation techniques.

In regular wireless networks, cells are managed independently. In this context, we propose a resource allocation algorithm with the aim of fairly guaranteeing the best service to users. When the cell charge is low enough to guarantee a sufficient quality of service, the algorithm redirects dynamically its priorities towards energy saving. This behavior allows to obtain an efficient compromise between capacity and energy consumption at different charge levels.

In order to extend network capacity, the adding of new cells allows to broaden the available bandwidth and to reduce the distance induced signal attenuation. We present an algorithm of user repartition in a multi-cell context, which intervenes before the resource allocation stage. A user can be covered by several cells using different frequencies, thus its repartition will have strong repercussions on the cell charge balance and on the general quality of service in the system.

The maximal number of cells in a sector is limited by its geographical environment. The Massive-MIMO allows to increase the cell functionalities while allowing a better energy directivity, and thus adding the spatial component to the resource allocation. We propose a new indicator evaluating the spatial compatibility of users based on past allocations. Once integrated in an allocation algorithm, it takes advantage of the superior capacities of Massive-MIMO.