Design and performance of coordination schemes in cellular networks

Nivine Abbas

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2016-ENST-0068

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité Informatique et Réseaux

présentée et soutenue publiquement par

Nivine Abbas

le 9 novembre 2016

Conception et performance de schémas de coordination dans les réseaux cellulaires

Directeur de thèse: **M. Thomas BONALD**
Co-directeur de thèse: **Mme. Berna SAYRAC**

Jury

M. Mohamad ASSAAD, Maître de Conférence, Supélec, France	Rapporteur
M. Olivier BRUN, Chargé de Recherche, LAAS-CNRS , France	Rapporteur
M. Eitan ALTMAN, Directeur de Recherches, INRIA , France	Examinateur
M. Tijani CHAHED, Professeur, Télécom SudParis, France	Examinateur
Mme Nidhi HEGDE, Ingénieur de Recherche, Nokia Bell Labs, France	Invitée
M. Thomas BONALD, Professeur, Télécom ParisTech, France	Directeur de thèse
Mme. Berna SAYRAC, Ingénieur de Recherche, Orange Labs, France	Co-directrice de thèse

TELECOM ParisTech
école de l'Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

THÈSE

PhD Thesis

TELECOM ParisTech and Orange Labs

presented and defended in public by

Nivine Abbas

On November 9 2016

Design and performance of coordination schemes in cellular networks

PhD Director: **Thomas BONALD**

PhD co-Director: **Berna SAYRAC**

À mes chers parents.

Acknowledgements

I would like to express my deepest gratitude to my advisors Thomas Bonald and Berna Sayrac. I am sincerely grateful for their support and their guidance.

I am truly grateful to my academic advisor Thomas. I thank him for his valuable insights, for his constructive advices and for his availability. It has been an honor for me to do my Ph.D under his supervision and to learn from his scientific rigor.

I am enormously thankful to my Orange Labs advisor Berna. I would like to sincerely thank her for giving me the opportunity to do this Ph.D, for believing in me, for supporting me, for her repeated encouragements and for the kindness she has shown me during these three years. I had the honor and pleasure to work with her.

I am extremely thankful to my manager Pierre Dubois for his support, his encouragement and the opportunities he has given me. I am sincerely grateful for the time he has spent and for the effort he has made to help me during the last period of my Ph.D.

I would like to acknowledge my committee members: Mohamad Assaad, Olivier Brun, Eitan Altman, Tijani Chahed and Nidhi Hegde for having accepted to be part of my Ph.D. examination committee.

I would like to thank all my fellow labmates at LINCS with whom I shared great moments. In particular, I wish to thank Rim, Ahlem, Mira, Céline, Yu-Ting, Ghida, Jose, Nihel, Sameh, Stéphane, Marco, Jordan, Natalia, Andrea and Sara.

I thank the members of the RIDE team who have provided a very pleasant working environment. I thank particularly my office mates and my dear friends Sinda, Rita, Thomas, Abdulaziz, Wassim, Stefan, Mohamad, Sandeep and Vaggelis. Thank you for the great moments and the good humor.

Mostly, I want to thank my parents who made me what I am today and to whom I would like to dedicate this thesis. Thank you for everything. Very special thanks go to my sister and my two brothers.

Abstract

Interference is still the main limiting factor in cellular networks, especially at the cell edge. We focus, in this thesis, on the different coordinated multi-point schemes (CoMP) proposed in the LTE-Advanced standard to cope with interference and to increase cell-edge user throughput, taking into account the dynamic aspect of traffic and users' mobility. The results are obtained by the mathematical analysis of Markov models and system-level simulations.

In the first part, we show the important impact of the scheduling strategy on the network performance in the presence of mobile users considering elastic traffic and video streaming. We propose a new scheduler that deprioritizes mobile users located at the cell edge, in order to improve the overall system efficiency.

Then, we consider the coordination scheme called Joint Processing, which consists in either allowing several base stations to transmit simultaneously to the same user (JT for Joint Transmission) or in silencing some base stations (DPB for Dynamic Point Blanking). We show that it is interesting to activate this feature only in a high-interference network (in urban areas), its activation in a low-interference network (rural or network supporting beamforming) may lead to performance degradation. We propose a new coordination mechanism, where a cell cooperates only when its cooperation brings a sufficient mean throughput gain, which compensates the extra resource consumption. Then, we show that the DPB technique is generally more performant than the JT technique.

Finally, we consider Coordinated Beamforming scheme. We show that coordination is not necessary when a large number of antennas is deployed at each base station; a simple opportunistic scheduling strategy provides optimal performance. For a limited number of antennas per base station, coordination is necessary to avoid interference between the activated beams, allowing substantial performance gains.

KEY-WORDS: Cellular data networks, mobility, opportunistic scheduling, coordinated multipoint CoMP, joint transmission, dynamic point blanking, 2D beamforming, Coordinated beamforming, flow-level modeling, queuing theory, system-level simulations.

Résumé

L'interférence entre stations de base est considérée comme le principal facteur limitant les performances des réseaux cellulaires, notamment en bord de cellule. Nous nous intéressons dans cette thèse aux différents schémas de coordination multi-point (CoMP) proposés dans la norme LTE-Advanced pour y faire face, en tenant compte à la fois de l'aspect dynamique du trafic et de la mobilité des utilisateurs. Les résultats sont obtenus par l'analyse mathématique de modèles markoviens et par des simulations du système.

Dans une première partie, nous montrons l'importance de l'algorithme d'ordonnancement sur les performances du réseau en présence d'utilisateurs mobiles, pour des services de téléchargement de fichier et de streaming vidéo. Nous proposons un nouvel algorithme d'ordonnancement basé sur la dé-priorisation des utilisateurs mobiles se trouvant en bord de cellule, afin d'améliorer l'efficacité globale du système.

Nous nous intéressons ensuite à la technique de coordination dite Joint Processing consistant à permettre à plusieurs stations de base de transmettre simultanément à un même utilisateur (JT pour Joint Transmission) ou à réduire au silence certaines stations interférentes (DPB pour Dynamic Point de Blanking). Nous montrons qu'il est intéressant d'activer cette fonctionnalité uniquement dans un réseau à forte interférence (en zone urbaine), son activation dans un réseau à faible interférence (zone rurale ou réseau avec formation de faisceaux) pouvant conduire à une dégradation des performances. Nous proposons un nouveau mécanisme de coordination où une cellule ne coopère que lorsque sa coopération apporte un gain moyen de débit suffisant pour compenser les pertes de ressources engendrées, puis nous montrons que la technique DPB est généralement plus efficace que la technique JT.

Nous considérons enfin la technique de formation de faisceaux coordonnée (coordinated beamforming). Nous montrons notamment que la coordination n'est pas nécessaire lorsque l'on dispose d'un grand nombre d'antennes par station de base, un simple mécanisme d'ordonnancement opportuniste permettant d'obtenir des performances optimales. Pour un nombre limité d'antennes par station de base, la coordination est nécessaire afin d'éviter l'interférence entre les faisceaux activés, et permet des gains de performance substantiels.

MOTS-CLEFS: Réseaux cellulaires, mobilité, ordonnancement opportuniste, coordination multi-point CoMP, technique de formation de faisceaux, modélisation niveau flot du trafic, théorie des files d'attente, simulations systèmes.

Résumé en Français

Introduction

Les réseaux cellulaires et sans fil ont commencé à être utilisés à grande échelle dans les années 90. Depuis, ils ont connu un développement soutenu grâce à la prise en compte progressive de différentes évolutions de techniques.

Dans les systèmes cellulaires classiques, les fréquences de transmission descendante entre les cellules voisines sont différentes. L'interférence entre les cellules devient ainsi un problème mineur dans ces systèmes. Ceci améliore le rapport signal sur interférence plus bruit, SINR (Signal to noise and interference ratio). Cependant, le gain obtenu avec cette amélioration du SINR reste plus faible que le gain équivalent obtenu dans le cas où les fréquences sont complètement réutilisées dans les cellules voisines.

La contrainte de la rareté de la ressource fréquence dans les systèmes sans fil a conduit à la nécessité de systèmes de communication efficaces permettant de maximiser la capacité. LTE est ainsi conçu pour fonctionner avec un facteur de réutilisation de fréquences de un. L'interférence entre les cellules voisines devient, ainsi l'un des problèmes les plus graves liés au déploiement de la technologie LTE.

Motivation

L'interférence étant le principal facteur limitant dans les réseaux cellulaires, plusieurs approches et fonctionnalités nouvelles ont été considérées comme des éléments-clés pour y faire face dans les réseaux mobiles futurs. La motivation principale de ces nouvelles fonctionnalités est d'améliorer le débit des utilisateurs en bordure de cellule, augmentant ainsi la couverture en haut débit. Dans ce contexte, CoMP est considéré comme une technique prometteuse dans LTE-Advanced pour satisfaire les exigences en termes de capacité et de débit des utilisateurs aux frontières des cellules. En downlink, deux schémas de coordination CoMP sont principalement considérés: Joint processing (JP) et coordinated scheduling / beamforming (CS / CB).

Le schéma Joint Processing JP, est considéré comme une technique prometteuse pour réduire les interférences intercellulaires, soit en réduisant au silence les stations interférentes (Dynamic Point de Blanking DPB), soit en permettant à plusieurs stations de base de transmettre simultanément à un même utilisateur (Joint Transmission JT). Cependant, ceci est au détriment d'une consommation excessive de ressources, d'une augmentation du trafic de signalisation, et de nombreux problèmes techniques, engendrant plusieurs défis pratiques. Ainsi, on peut s'attendre à une dégradation des

performances, dans certains cas.

Dans la littérature, la sélection des cellules coopérantes est basée sur la différence entre la puissance reçue par l'utilisateur de la cellule serveuse et celle reçue par l'utilisateur de la cellule voisine candidate à coopérer. Lorsque cette différence est inférieure à un certain seuil fixe prédéfini δP, la cellule voisine est définie comme une cellule coopérante. Cependant, le choix de ce seuil n'est pas bien défini.

La technique JT est généralement considérée comme la technique JP la plus prometteuse et peu d'importance a été accordée à la technique DPB dans la littérature. La mobilité des utilisateurs peut également avoir un impact critique sur la performance de ces systèmes de coordination car les utilisateurs peuvent passer d'une zone où la coordination est effectuée à une autre où la coordination n'est pas effectuée et vice versa.

D'autre part, le schéma Coordinated Scheduling/Beamforming, où les décisions d'ordonnancement et les faisceaux (beam) sont coordonnés entre les cellules coopérantes pour contrôler l'interférence générée, semble être un schéma efficace. Mais ceci est au prix d'une plus grande complexité.

La littérature actuelle sur CoMP n'a pas abouti à un consensus sur l'applicabilité de cette fonctionnalité. En plus, il est courant de supposer un nombre fixe d'utilisateurs statiques ou semi-statiques dans l'évaluation de performance.

Donc, il est intéressant d'avoir une étude qui permet de comparer les différents schémas de coordination, leurs performances et leur complexité, en considérant différentes stratégies d'ordonnancement, tout en tenant compte de l'impact des variations de la qualité de canal : les variations rapides ainsi que les variations lentes dues à la mobilité.

Contributions

Dans cette thèse, nous considérons d'abord le cas d'un réseau sans coordination. Nous montrons que la performance d'une politique d'ordonnancement dépend principalement de la mobilité des utilisateurs, ainsi que du type de service: élastique, streaming. Étant donné que les utilisateurs mobiles, se trouvant dans des mauvaises conditions radio, sont susceptibles de se déplacer et d'être servis dans de meilleures conditions, nous proposons une politique d'ordonnancement qui dé-priorise ces utilisateurs. La politique proposée, en adaptant à la mobilité observée des utilisateurs actifs, améliore les performances des politiques d'ordonnancement habituelles. Nous analysons aussi les performances, sous plusieurs stratégies d'ordonnancement, en présence de trafic streaming adaptatif à travers des modèles basés sur la théorie des files d'attente. Nous montrons qu'il y a un compromis entre la qualité de la vidéo et le surplus de buffer et nous suggérons, ainsi, une politique d'ordonnancement selon la métrique ciblée.

Nous considérons, ensuite, le cas d'un réseau supportant l'un des schémas de coordination: JT, DPB et CS / CB. Nous évaluons d'abord les performances du schéma JP, et nous identifions les scénarios où il est intéressant de déployer cette fonctionnalité. L'évaluation est basée sur l'analyse des modèles mathématiques ainsi que sur des simulations réalisées au niveau système. Nous établissons des modèles, tenant compte de la mobilité des utilisateurs, essentiellement pour le scénario de plusieurs cellules coopérantes avec une zone commune de coordination et pour le cas d'un cluster statique constitué de trois cellules coopérantes.

Nous étudions les deux cas, le cas d'un scénario avec haut niveau d'interférence et celui d'un scénario avec bas niveau d'interférence. En effet, les gains dépendent principalement du niveau d'interférence dans le réseau. Nous proposons, ainsi, un nouveau mécanisme de coordination où une cellule ne coopère que quand sa coopération apporte un gain moyen de débit suffisant pour compenser les pertes de ressources engendrées. Nous montrons que la technique DPB est plus prometteuse que JT. Bien que la coordination intra-site soit beaucoup plus simple à mettre en œuvre, la coordination

dynamique est nécessaire afin de réduire davantage l'interférence.

D'autre part, nous montrons qu'il n'est pas recommandé d'activer la fonctionnalité JP, dans un scénario à faible niveau d'interférence. Si malgré cela, la coordination est activée pour les utilisateurs statiques, souffrant d'un débit très dégradé, il vaut mieux de ne pas l'activer pour les utilisateurs mobiles qui sont susceptibles de se déplacer et d'être servis dans de meilleures conditions radio. Ce qui évite la consommation de ressources supplémentaires pour les utilisateurs cell-edge mobiles.

La technique de formation de faisceaux (beamforming) étant la solution la plus efficace permettant de réduire les interférences, nous étudions, dans cette thèse, un système supportant le beamforming 2D. Nous considérons d'abord le cas où l'activation des faisceaux est effectuée d'une manière non-coordonnée dans les différentes cellules. Nous évaluons pour différents nombres d'antennes, le gain de capacité maximal, par rapport à un système sans formation de faisceaux. Les performances, en termes de débit moyen des utilisateurs, sont aussi évaluées en considérant des ordonnanceurs non-coordonnés et des ordonnanceurs coordonnés qui permettent d'activer les faisceaux d'une manière coordonnée. Nous constatons, ainsi, que les politiques d'ordonnancement telle que max C/I, dans un système avec formation de faisceaux, où un grand nombre d'antenne est déployé, permettent d'obtenir des performances similaires à celles obtenues en utilisant des ordonnanceurs coordonnés. Dans les cas où un nombre limité d'antennes est déployé à la station de base, la coordination peut être nécessaire afin d'éviter l'interférence entre les faisceaux activés. Cependant, ceci est au détriment d'une plus grande complexité. Il y a donc un compromis à considérer entre le nombre d'antennes déployées et la complexité de la politique d'ordonnancement.

Analyse de performance

Afin d'évaluer les performances des différents schémas de coordination et des différentes stratégies d'ordonnancement, nous nous appuyons dans cette thèse sur l'analyse mathématique de modèles markoviens et sur des simulations du système.

a-Modèles markoviens

Une cellule est représentée comme un ensemble de régions concentriques autour de la station de base, le débit physique disponible étant de plus en plus faible au fur et à mesure qu'on s'en éloigne. Dans chaque région, les utilisateurs génèrent de nouveaux flots de trafic suivant un processus de Poisson. Ce modèle est équivalent à un ensemble de processeurs couplés. Chaque file d'attente représente une région. Le taux de service de chaque file dépend de l'état global du système à travers la politique d'ordonnancement.

b-Simulation du système

Afin de valider les résultats obtenus par l'analyse des modèles mathématiques, nous nous basons sur des simulations du système grâce à un simulateur développé pour les besoins de la thèse, utilisant la technologie LTE avec une fréquence porteuse de 2GHz et une bande passante de 10 MHz. Nous considérons les 21 cellules hexagonales formées par 7 sites tri-sectoriels (un site de référence entouré par 6 sites interférants). Nous considérons un environnement urbain en outdoor avec des macro cellules uniquement et une distance inter-site de 500 m.

Le trafic consiste de transferts de fichiers. Les flux arrivent uniformément dans le réseau selon un processus de Poisson. Les tailles des fichiers sont générées à partir d'une distribution exponentielle. Les décisions d'ordonnancement sont prises dans chaque stattion de base chaque milliseconde, selon la stratégie d'ordonnancement utilisée.

Chapitre II: Réseaux cellulaires en l'absence de coordination inter-cellules

Ce chapitre est consacré à l'évaluation de performance de différentes politiques d'ordonnancement dans les réseaux LTE en l'absence de coordination inter-cellules.
Une cellule est représentée comme un ensemble de régions, où chaque région représente une condition radio liée essentiellement à la qualité du signal reçu par l'utilisateur.

La politique d'ordonnancement détermine la répartition des ressources radio entre les différents utilisateurs servis par la même station de base. Ces ressources sont soit partagées d'une manière complètement équitable indépendamment des conditions radio des utilisateurs, en adoptant une stratégie RR (round robin), soit partagées d'une manière opportuniste tenant compte des conditions radio des différents utilisateurs, en adoptant une stratégie PF (équité proportionnelle) ou max C/I. Ainsi, le débit instantané de chaque utilisateur dépend de ses conditions radio ainsi que de la politique d'ordonnancement.

La **mobilité** intra-cellule des utilisateurs est modélisée en introduisant des taux de transition supplémentaires dans la chaîne de Markov, qui expriment qu'un utilisateur actif peut passer d'une région à une région adjacente en cours de communication.

Chaque cellule est ainsi vue comme un réseau de files d'attente couplées avec routage. Le taux de service instantanée dans chaque file d'attente dépend essentiellement de la nature des flux. Nous traitons alors séparément les cas des flux élastiques et des flux de streaming adaptatif.

1-Analyse de performance des flots élastiques

En supposant que les flots ont une taille exponentiellement distribuée, nous établissons tout d'abord des résultats explicites sur les conditions de stabilité et le débit moyen par utilisateur, en l'absence de mobilité, des politiques RR et max C/I. L'analyse avec mobilité permet d'établir des résultats analytiques dans deux régimes asymptotiques (en faible trafic et quand les taux de mobilité entre les régions tendent vers l'infini). En dehors de ces deux cas, il faut alors essentiellement recourir à des méthodes numériques de résolution de chaînes de Markov. Ces derniers résultats sont ensuite généralisés pour prendre en compte plusieurs classes de flots et plusieurs classes de mobilité.

Grâce à ces modèles, nous montrons l'effet bénéfique de la mobilité sur les performances (niveau flot) des différentes politiques d'ordonnancement considérées. Pour des utilisateurs statiques, une politique équitable comme PF est préférable, alors que pour des utilisateurs mobiles, il vaut mieux utiliser une politique plus opportuniste telle que max C/I, comme le montre les résultats illustrés par la Figure 1. Cette figure montre le débit moyen des utilisateurs en fonction de la charge du réseau pour le RR, le max C/I et le PF (on prend comme charge de référence la charge d'un réseau avec une stratégie d'ordonnancement RR).

En conséquence, nous proposons un ordonnanceur (Mob-aware) exploitant l'information liée à la mobilité des utilisateurs. L'idée de cet ordonnanceur est de dé-prioriser les utilisateurs mobiles se trouvant dans de mauvaises conditions radio car ces utilisateurs sont susceptibles de se déplacer pour être servis dans de meilleures conditions radio.

Ordonnancement prenant compte de la mobilité

La stratégie d'ordonnancement est un élément clé dans les réseaux cellulaires. LTE utilise la technique d'accès multiple OFDMA où l'accès multiple est obtenu en attribuant des sous-ensembles de sous-porteuses à différents utilisateurs.

La stratégie la plus simple consiste à servir les utilisateurs en RR, indépendamment de leurs conditions radio. La simplicité de cette stratégie est au prix d'une performance dégradée.

La stratégie d'ordonnancement max C/I associe les ressources à l'utilisateur ayant le débit instantané le plus élevé:

$$u(t) = \arg\max_i r_i(t),$$

$u(t)$ étant l'utilisateur servi à l'instant t et $r_i(t)$ le débit instantané de l'utilisateur i à l'instant t. Cette stratégie est la plus efficace en terme de débit. Cependant, elle peut négliger les utilisateurs ayant de mauvaises conditions radio, qui ne sont, alors, jamais servis à forte charge. Néanmoins, avec cette stratégie, la mobilité est très bien exploitée, étant donné que les utilisateurs mobiles seront plus susceptibles d'être servis dans de bonnes conditions radio.

La stratégie PF sert l'utilisateur ayant le meilleur débit instantané sur débit moyen:

$$u(t) = \arg\max_i \frac{r_i(t)}{\bar{R}_i(t)}.$$

Cette stratégie offre un bon compromis entre l'efficacité et l'équité en l'absence de mobilité.

La stratégie d'ordonnancement Mob-aware que nous proposons s'adapte à la mobilité des utilisateurs. L'idée de cette stratégie est de dé-prioriser les utilisateurs mobiles ayant de mauvaises conditions radio, car dans ce cas, ils sont susceptibles de se déplacer et de bénéficier de meilleures conditions radio. Pour ce faire, nous utilisons deux métriques d'ordonnancement différentes en fonction de la mobilité de chaque utilisateur. Ainsi, une classification des utilisateurs en mobile ou statique est nécessaire.

Nous désignons par $\mathbb{S}$ l'ensemble des utilisateurs statiques et par $\mathbb{M}$ l'ensemble des utilisateurs mobiles. L'utilisateur servi à l'instant t est l'utilisateur ayant la métrique $\eta_i(t)$ la plus élevée:

$$u(t) = \arg\max_i \eta_i(t),$$

la métrique des utilisateurs statiques est identique à la métrique classique de l'équité proportionnelle:

$$\eta_i(t) = \frac{r_i(t)}{\bar{R}_i(t)} \qquad \forall i \in \mathbb{S},$$

La métrique des utilisateurs mobiles est évaluée comme suit:

$$\eta_i(t) = \frac{r_i(t)}{\bar{R}_{\text{mob}}(t)} \frac{r_i(t)}{\bar{R}(t)} \qquad \forall i \in \mathbb{M},$$

$\bar{R}_{\text{mob}}(t)$ est le débit moyen perçu par l'utilisateur mobile ayant les meilleures conditions radio (SINR) à l'instant $t-1$. Notez qu'au sein de l'ensemble des utilisateurs mobiles l'ordonnancement est similaire à celui assuré par la stratégie max C/I classique. À chaque utilisateur mobile est également associé un poids $r_i(t)/\bar{R}(t)$, où $\bar{R}(t)$ est le débit moyen expérimenté dans le réseau, ce débit dépend principalement des conditions radio dans le réseau. Il est évalué par une moyenne exponentiellement lissée du débit moyen des utilisateurs actifs:

$$\bar{R}(t) = \left(1 - \frac{1}{t_a}\right)\bar{R}(t-1) + \frac{1}{t_a}\frac{\sum_i \bar{r}_i(t)}{n(t)},$$

$n(t)$ étant le nombre total des utilisateurs actifs à l'instant t et t_a représente l'échelle de temps sur laquelle la moyenne est effectuée. Ce paramètre de temps doit être suffisamment grand (par exemple, de l'ordre de secondes) pour refléter les conditions radio moyennes de la cellule. Le débit moyen de

l'utilisateur i est également évalué par une moyenne exponentiellement lissée, uniquement sur les intervalles de temps où l'utilisateur est servi:

$$\bar{r}_i(t) = \bar{r}_i(t-1) + \frac{1}{t_a}\left(r_i(t) - \bar{r}_i(t-1)\right)\mathbb{1}_{\{u(t-1)=i\}}.$$

Notez que le poids $r_i(t)/\bar{R}(t)$ attribué aux utilisateurs mobiles permet de déprioriser les utilisateurs mobiles qui expérimentent des conditions radio pires que les conditions radio moyennes du réseau. L'algorithme d'ordonnancement fonctionne comme suit (t_i indique l'instant d'arrivé de l'utilisateur i):

Data: CQI (indice de qualité de canal) et classe de mobilité de chaque utilisateur
Result: Utilisateur à servir
// Ordonnancement
MAX $\leftarrow 0$
for each utilisateur i **do**
  if utilisateur i est statique **then**
    $\eta \leftarrow \frac{r_i}{\bar{R}_i}$
  else
    $\eta \leftarrow \frac{r_i}{\bar{R}_{\text{mob}}}\frac{r_i}{\bar{R}}$
  end
  if MAX $< \eta$ **then**
    MAX $\leftarrow \eta$
    $u \leftarrow i$
  end
end

// Mises à jour
$\bar{R} \leftarrow \left(1 - \frac{1}{t_a}\right)\bar{R}$
for each utilisateur i **do**
  $\bar{R}_i \leftarrow \left(1 - \frac{1}{t-t_i}\right)\bar{R}_i + \frac{1}{t-t_i} r_i \times (u == i)$

  $\bar{r}_i \leftarrow \bar{r}_i + \frac{1}{t_a}(r_i - \bar{r}_i) \times (u == i)$

  $\bar{R} \leftarrow \bar{R} + \frac{1}{t_a}\frac{\bar{r}_i}{n}$
end
if u est mobile **then**
  $\bar{R}_{\text{mob}} \leftarrow \bar{R}_u$
end

Cet ordonnanceur se comporte comme max C/I lorsque tous les utilisateurs sont mobiles et comme PF lorsque tous les utilisateurs sont statiques. Cette stratégie est un bon compromis dans un réseau avec plusieurs classes de mobilité.

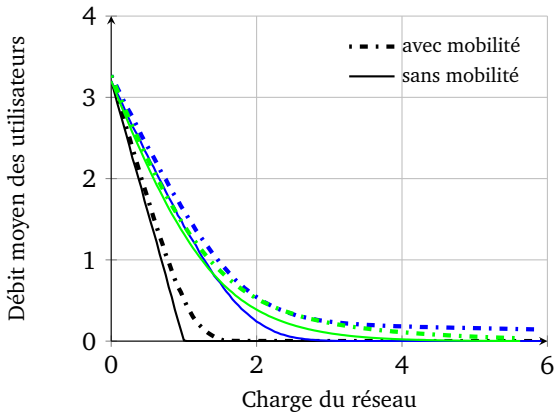


Figure 1: Débit moyen des utilisateurs pour le RR (noir), max C/I (bleu) et PF (vert).

Résultats numériques

Les résultats sont obtenus par la résolution numérique des modèles markoviens et par des simulations du système.

a-Résolution numérique de chaines de Markov

La Figure 2 montre le débit moyen d'un utilisateur pour le PF, le max C/I ainsi que pour la stratégie proposée en supposant que 75% des utilisateurs sont mobiles. Les résultats montrent qu'il y a un gain significatif apporté par la stratégie d'ordonnancement proposée, notamment à charge élevée. Plus le taux de mobilité (qui est proportionnel à la vitesse) et le pourcentage des utilisateurs mobiles sont élevés, plus le gain est élevé. Cette stratégie peut être une stratégie prometteuse dans les réseaux mobiles chargés.

b-Simulation du système

Dans les simulations présentées dans la Figure 3, nous supposons qu'il existe deux comportements de mobilité: des utilisateurs statiques et des utilisateurs mobiles qui se déplacent à une vitesse constante de 300 km/h. Nous supposons que 75% des utilisateurs sont mobiles. Nous simulons 100 minutes et nous estimons le débit moyen d'un utilisateur comme étant le rapport entre la taille moyenne du fichier et la durée moyenne de séjour d'un utilisateur pour une taille moyenne de fichier de 1.25 MBytes, pour les différentes stratégies d'ordonnancement (PF, max C/I et mob-aware). Nous évaluons le gain de débit relatif apporté par la stratégie d'ordonnancement proposée par rapport aux ordonnanceurs classiques PF et max C/I, en fonction de la charge du réseau. Les résultats sont très proches de ceux obtenus par l'analyse (présentés dans la Figure 2) et montrent que la stratégie d'ordonnancement proposée est un bon compromis dans un réseau avec plusieurs classes de mobilité. Les performances sont améliorées surtout à forte charge. Plus la proportion d'utilisateurs mobiles est élevée, plus le gain relatif à la stratégie PF est élevé et plus les performances sont proches de celles obtenues pour une stratégie max C/I.

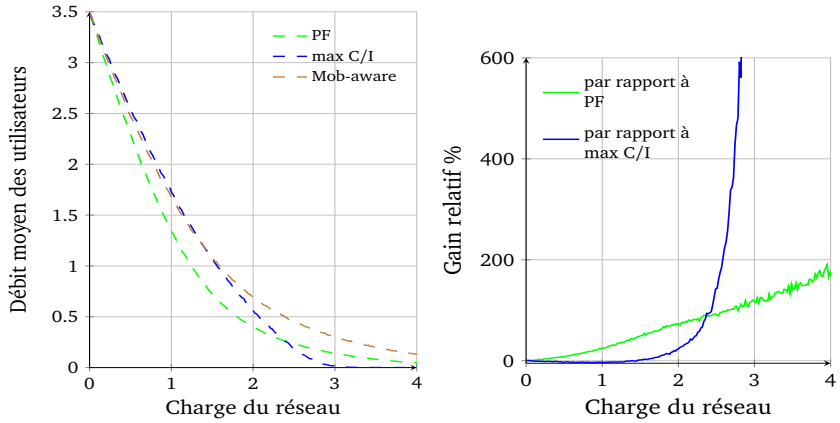


Figure 2: Le débit moyen des utilisateurs dans un réseau où 75% des utilisateurs sont mobiles.

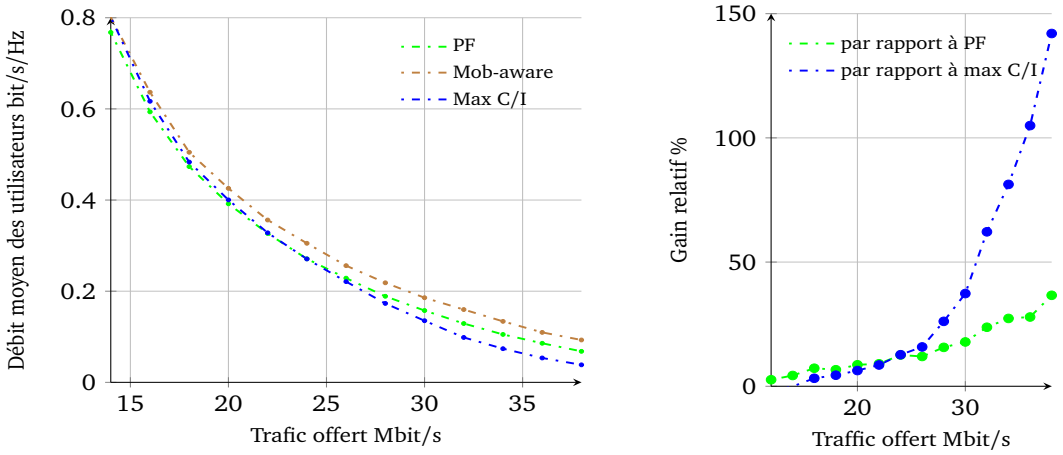


Figure 3: Le débit moyen des utilisateurs dans un réseau où 75% des utilisateurs sont mobiles, obtenu par des simulations du système.

2-Applications de streaming vidéo adaptatif

Ce type d'application peut réguler la qualité vidéo en fonction du débit disponible, la quantité totale de données téléchargées dépendant alors des conditions de trafic. Dans le modèle markovien, nous supposons des durées de vidéo exponentiellement distribuées. Nous considérons principalement deux métriques : le débit vidéo, qui contrôle la qualité de l'image, et le buffer surplus, qui contrôle le taux d'interruption de la vidéo. Nous explicitons les conditions de stabilité pour les politiques RR, max C/I et max-min. Nous obtenons des résultats explicites en condition de faible trafic. Les résultats numériques obtenus montrent l'importance de prendre en compte la mobilité des utilisateurs. Pour des utilisateurs mobiles, il apparaît ainsi que si la politique max C/I offre un meilleur débit vidéo moyen, elle peut conduire à une dégradation du buffer surplus en condition de charge modérée par rapport à RR. Au vu de ces résultats, nous concluons qu'il y a un compromis à faire entre max C/I et RR, et proposons une politique DPS. Les résultats de simulation montrent la pertinence de cette proposition.

Chapitre III: Joint Processing CoMP

Le chapitre III est consacré aux techniques CoMP de traitement conjoint qui permettent de coordonner les cellules pour éliminer les problèmes d'interférence inter-cellules. La coordination entre cellules est réalisée soit en transformant un signal interférant en signal utile (JT : Joint Transmission), soit en rendant silencieuse une cellule interférente (DPB : Dynamic Point Blanking). Alors que la plupart des études se sont focalisés sur la première technique, au détriment de la seconde, et ont utilisé des modèles ignorant la dynamique aléatoire de la demande, ce chapitre évalue l'intérêt de ces deux techniques et les compare en utilisant des modèles de niveau flot ainsi que des simulations du système. Pour cela, trois configurations sont étudiées dans la modélisation:

- plusieurs cellules interférentes avec une zone de coordination commune
- un cluster statique constitué de trois cellules coopérantes (coordination intra-site)
- une cellule hexagonale coopérant avec les cellules l'entourant

Chaque configuration est étudiée en considérant des scénarios à forte ou faible interférence, avec différentes techniques de clustering, et plusieurs politiques d'ordonnancement.

Ordonnancement

.

Lorsqu'une cellule est impliquée dans un processus de coordination, ses ressources sont partagées entre ses utilisateurs et les utilisateurs CoMP associés aux cellules voisines et nécessitant une coopération de la cellule en question. Dans chaque cellule appartenant à un cluster (groupe de cellules coopérantes), la décision d'ordonnancement doit dépendre des décisions d'ordonnancement des cellules coopérantes.

Ainsi, la stratégie d'ordonnancement est un élément clé dans les réseaux avec une coordination inter-cellules. Afin d'allouer efficacement les ressources aux utilisateurs d'un cluster, les informations liées à la qualité de canal de tous les utilisateurs associés aux cellules du cluster doivent être prises en compte par l'ordonnanceur. Cela peut être facilement réalisé en utilisant un ordonnanceur centralisé dans chaque cluster. Cette approche est plus adoptée à la coordination intra-site où tous les secteurs sont contrôlés par la même station de base macro. Elle peut également parfaitement fonctionner avec une architecture C-RAN. Dans le cas général, il est possible de définir une cellule du cluster en tant que cellule "master". Toutes les autres cellules du cluster agissent ainsi comme étant des "slaves".

Dans ce chapitre, nous considérons un ordonnanceur centralisé dans chaque cluster. Cet ordonnanceur dispose des informations liées aux conditions radio de tous les utilisateurs dans le cluster et peut ainsi prendre des décisions d'ordonnancement en choisissant l'utilisateur à servir dans chaque cellule faisant partie du cluster considéré.

Nous nous concentrons sur le partage des ressources entre les utilisateurs CoMP et les utilisateurs non-CoMP. Ce partage peut être soit équitable en traitant les utilisateurs CoMP et non-CoMP également ou basé sur une stratégie de priorisation qui donne la priorité à une catégorie d'utilisateurs (CoMP ou non-CoMP). Nous montrons que la performance de ces stratégies de priorisation dépend principalement de la mobilité des utilisateurs et nous proposons une stratégie d'ordonnancement qui s'adapte à la mobilité des utilisateurs.

- **Allocation statique**:
  La stratégie la plus simple consiste à dédier une sous-bande de fréquence aux utilisateurs CoMP.

Cette stratégie offre des performances limitées à cause d'une utilisation inefficace des ressources radio. Nous considérons cette stratégie à titre de comparaison.

- **Ordonnancement itératif**:
  Une autre façon de procéder à l'ordonnancement de manière centralisée consiste à le faire de façon itérative, en traitant les utilisateurs non-CoMP et les utilisateurs CoMP également sans affectation d'une sous-bande de fréquence à une catégorie ou à une autre.

  Dans ce cas, la complexité d'ordonnancement augmente linéairement avec la taille du cluster N. Elle nécessite au plus N itérations. Cela est dû au fait qu'au plus N utilisateurs peuvent être servis simultanément sur les mêmes ressources à l'intérieur d'un même cluster à un instant donné.

  Le premier utilisateur choisi est celui ayant la métrique d'ordonnancement (PF, max C/I, RR ...) maximale, bloquant ainsi la cellule serveuse et toutes les cellules coopérantes (dans le cas d'un utilisateur CoMP). Le deuxième meilleur utilisateur selon la même métrique d'ordonnancement est choisi seulement si sa cellule serveuse et toutes ses cellules coopérantes sont toujours disponibles après la première sélection. Et ainsi de suite jusqu'à ce que toutes les cellules du cluster deviennent indisponibles ou après avoir effectuer N itérations...

- **Priorisation des utilisateurs non-CoMP**:
  Avec cette politique d'ordonnancement, les utilisateurs non-CoMP dans chaque cellule sont servis en premier et sont alloués toutes les ressources radio chaque fois qu'ils sont actifs. Les utilisateurs CoMP attendent jusqu'à ce que les ressources de leurs cellules serveuses et celles de toutes leurs cellules coopérantes soient disponibles. L'ordonnancement se fait en deux étapes. A la première étape, les ressources sont affectées aux utilisateurs non-CoMP. La deuxième étape consiste à affecter les ressources restantes aux utilisateurs CoMP appropriés.

  Les utilisateurs sont choisis en fonction d'une métrique d'ordonnancement particulière (PF, max C / I ...) ou même aléatoirement (pour RR).

  Cette stratégie offre les meilleures performances lorsque les utilisateurs sont mobiles. Au contraire, en l'absence de mobilité les performances sont très dégradées.

- **Priorisation des utilisateurs CoMP**:
  Cette politique consiste à servir les utilisateurs CoMP en premier. Les utilisateurs non-CoMP sont servis uniquement lorsqu'il n'y a pas d'utilisateurs CoMP actifs ou quand il y a encore de cellules disponibles après la planification des utilisateurs CoMP.

- **Ordonnanceur proposé: Mobility-aware**:
  Cet ordonnanceur exploite la mobilité comme information supplémentaire dans les décisions d'ordonnancement. Il consiste à déprioriser les utilisateurs CoMP qui sont susceptibles de se déplacer et d'être servis dans de bonnes conditions radio où la coordination cellulaire n'est pas nécessaire. Cette stratégie améliore les performances des stratégies précédentes dans un réseau comportant plusieurs types de mobilité.

Impact de la coordination inter-cellule sur le gain d'opportunisme

Nous montrons que les techniques de coordination améliorent les performances et la capacité du réseau quand le déclenchement de la coordination est effectuée au bon moment. Cependant, une dégradation des performances peut se produire à forte charge pour des stratégies d'ordonnancement sous-optimales.

La stratégie d'ordonnancement est un élément essentiel qui affecte fortement les performances en présence des techniques de coordination.

Considérons le cas d'un réseau classique sans coordination avec un ordonnanceur classique PF. Dans ce cas, il y a un gain d'opportunisme significatif apporté par le PF. Mais, lorsque les techniques de coordination sont appliquées avec une stratégie d'ordonnancement itérative (sous-optimale), ce gain d'opportunisme est réduit. Si le gain de coordination n'est pas suffisant pour compenser cette réduction du gain d'opportunisme, une dégradation des performances peut se produire.

Pour que le gain en coordination ne se fasse pas au détriment du gain opportuniste, nous proposons un ordonnanceur global choisissant les utilisateurs à servir en résolvant un problème d'optimisation non-linéaire en variables binaires, et évaluons numériquement les gains qu'il permet.

Nous étudions d'abord la technique de transmission JT, puis la technique DPB

1-Joint Transmission

JT a été démontré comme une méthode pour améliorer le débit des utilisateurs en bordures de cellules. Cependant, ceci est au prix d'une consommation plus élevée de ressources. Il a été récemment montré que CoMP et plus précisément le schéma JT peut être nuisible à forte charge, en raison de la dégradation de la condition de stabilité. Il y a donc un compromis entre la performance des utilisateurs se trouvant en bordures de cellules et la capacité du réseau à traiter tout le trafic.

Pour une cellule coopérante avec les cellules l'environnant, nous montrons qu'une condition suffisante pour que la coopération ne dégrade pas la capacité du système dans le cas d'une topologie de réseau symétrique est que le gain en débit soit supérieur au nombre de cellules coopérantes. Au vu de ces résultats, nous proposons une nouvelle définition d'une cellule coopérante : au lieu d'utiliser un seuil fixe sur la puissance reçue, une cellule coopérante est définie comme permettant un gain en débit de 100 % pour l'utilisateur dans le cas d'une topologie de réseau symétrique et un gain en débit supérieur à l'augmentation de la charge en bordure de cellule dans le cas général.

Cette technique est étudiée pour deux schémas de transmission conjointe différents. Dans le premier, toutes les cellules coopérantes transmettent simultanément les mêmes données vers l'utilisateur, alors que dans le second, les cellules transmettent simultanément des données différentes.

Étude de la condition de stabilité

Afin d'étudier les conditions qui permettent d'éviter la dégradation de la condition de stabilité, nous considérons le cas général d'une cellule entourée par un nombre donné de cellules voisines. La Figure 4 montre le cas d'un réseau homogène avec une topologie hexagonale.

Nous modélisons chaque cellule par deux zones principales: une zone de non-coordination (cell center) et une zone de coordination (cell edge). p_{center} est la probabilité d'arrivé d'un nouvel utilisateur dans le centre de la cellule. $p_{\text{edge},n}$ est la probabilité qu'un nouvel utilisateur arrive en bordure de la cellule impliquant n cellules dans la coordination. $p'_{\text{edge},n}$ est la probabilité d'arrivé d'un nouvel utilisateur associé à une cellule voisine, nécessitant la coopération de la cellule considérée et impliquant n cellules dans la coordination.

μ_{center} est le taux de service moyen d'un utilisateur au centre de la cellule. $\mu_{\text{edge,n}}$ est le taux de service moyen d'un utilisateur en bordure de la cellule lorsqu'il est servi sans coordination, et qui peut impliquer n cellules dans la coordination lors de l'obtention du statut CoMP. Nous désignons par λ le taux d'arrivée total dans la cellule.

Dans le cas où aucune coordination n'est effectuée entre les différentes cellules, la charge de la

cellule considérée pour une politique d'ordonnancement RR est:

$$\rho = \frac{p_{\text{center}}\lambda}{\mu_{\text{center}}} + \sum_{n} \frac{p_{\text{edge,n}}\lambda}{\mu_{\text{edge,n}}}, \tag{1}$$

La condition de stabilité en absence des techniques de coordination n'est autre que:

$$\rho < 1.$$

La condition de stabilité lorsque la cellule considérée coopère avec ses cellules voisines, c'est-à-dire en activant le schéma de coordination JP, dépend de la technique (JT ou DPB) où elle peut être plus restrictive lors de l'activation de JT par rapport au cas d'activation du DPB .

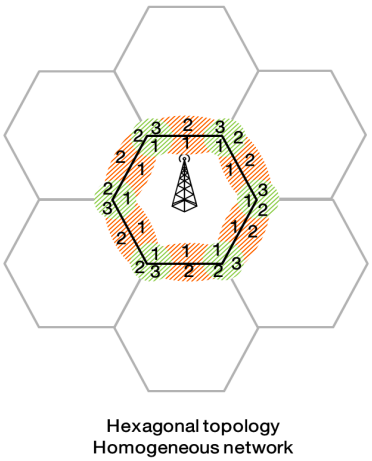


Figure 4: Augmentation de trafic en bordure de cellule.

Supposons maintenant que la technique JT soit activée dans le réseau. Soit β_n le gain de coordination moyen d'un utilisateur CoMP impliquant n cellules dans la transmission, c'est-à-dire le gain de débit moyen d'un utilisateur CoMP par rapport au cas lorsqu'il est servi sans coordination. Par conséquent, le taux de service moyen d'un utilisateur CoMP impliquant n cellules dans la transmission devient $\beta_n \mu_{\text{edge,n}}$.

Lors de l'activation du JT, les utilisateurs qui sont associés aux cellules voisines et qui requièrent la coopération de la cellule considérée créent une charge supplémentaire à cette cellule puisqu'ils constituent des utilisateurs en plus pour cette cellule, contrairement aux utilisateurs localisés dans cette cellule et qui font partie de ses propres utilisateurs.

Donc, le trafic des utilisateurs en bordure de cellule impliquant n cellules dans la transmission augmente d'un facteur: $\alpha_n = (p_{\text{edge,n}} + p'_{\text{edge,n}})/p_{\text{edge,n}}$. Mais, ces utilisateurs (en bordure de cellule) bénéficient d'un gain de débit grâce à la coopération. Ce qui leur permet de compléter plus rapidement leur service. Donc, si le gain de débit apporté aux utilisateurs en bordure de cellule est suffisamment élevé pour compenser la consommation supplémentaire de ressources, l'état de stabilité de la cellule peut être maintenu comme dans le cas sans coordination.

La nouvelle charge de la cellule en activant le JT avec une stratégie d'ordonnancement itérative RR devient:

$$\rho' = \frac{p_{\text{center}}\lambda}{\mu_{\text{center}}} + \sum_{n} \frac{p_{\text{edge,n}}\lambda + p'_{\text{edge,n}}\lambda}{\beta_n \mu_{\text{edge,n}}} = \frac{p_{\text{center}}\lambda}{\mu_{\text{center}}} + \sum_{n} \frac{\alpha_n}{\beta_n} \frac{p_{\text{edge,n}}\lambda}{\mu_{\text{edge,n}}} \tag{2}$$

Afin de maintenir la même charge de cellule, la condition suivante doit être remplie:

$$\sum_n \frac{\alpha_n}{\beta_n} \frac{p_{\text{edge,n}}}{\mu_{\text{edge,n}}} = \sum_n \frac{p_{\text{edge,n}}}{\mu_{\text{edge,n}}}.$$

Ainsi, une façon d'éviter la dégradation de la condition de stabilité est de garantir que:

$$\beta_n \geq \alpha_n \qquad \forall n.$$

En d'autres termes, l'implication d'une cellule donnée dans la coopération augmente son trafic dans une zone impliquant n cellules dans la coordination par un facteur α_n. Il n'y a pas d'impact sur la capacité de la cellule si ce trafic peut être servi α_n fois plus rapidement.

Dans le cas d'un réseau symétrique homogène, on peut supposer que $\alpha_n = n$. Ainsi, le gain de débit moyen d'un utilisateur CoMP impliquant n cellules dans la transmission doit être au moins égal à n:

$$\beta_n \geq n \qquad \forall n. \tag{3}$$

Choix du seuil de puissance δP

Considérons le cas d'une seule cellule coopérante $\mathbb{C} = \{c\}$, $\mathbb{C}$ étant l'ensemble de cellules coopérantes. Le gain de coordination moyen dans ce cas doit être au moins égal au nombre de cellules impliquées dans la transmission, soit $n = 2$.

P_s est la puissance reçue par l'utilisateur de la cellule serveuse et P_c est la puissance reçue de la cellule coopérante. Nous désignons par $I = \sum_{i \neq c} P_i$ l'interférence reçue par l'utilisateur de toutes les cellules voisines non coopérantes et $\mathcal{N}$ est le bruit. Le rapport signal sur interférence plus bruit SINR en l'absence de coopération est dans ce cas:

$$SINR = \frac{P_s}{P_c + I + \mathcal{N}}$$

Supposons que le premier schéma de transmission, où les mêmes données sont transmises à l'utilisateur des différentes cellules coopérantes, soit utilisé. Lorsque la cellule c coopère avec la cellule serveuse, le SINR devient:

$$SINR' = \frac{P_s + P_c}{I + \mathcal{N}}$$

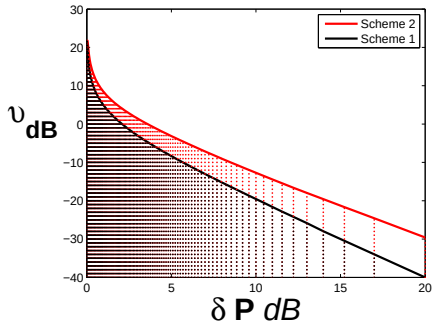


Figure 5: Région d'interférence pour un gain en débit de 100% en fonction de δP_{dB}

Étant donné que cet utilisateur utilise deux ressources, son nouveau débit doit être deux fois plus élevé que son débit initial en l'absence de coordination:

$$\frac{\log_2(1+SINR')}{\log_2(1+SINR)} \geq 2$$

En posant: $\zeta = P_c/P_s$ and $\upsilon = (I+\mathcal{N})/P_s$, l'expression précédente devient:

$$\frac{1+\zeta}{\upsilon} \times (\zeta+\upsilon) \geq \frac{1}{\zeta+\upsilon} + 2.$$

Cette inéquation est vérifiée lorsque:

$$\upsilon \leq \frac{\zeta^2}{1-\zeta}. \tag{4}$$

La zone hachurée de la Figure 5 représente la solution à cette inéquation ($\delta P_{dB} = P_s - P_c = 10\log_{10}(1/\zeta)$, $\upsilon_{dB} = 10\log_{10}(\upsilon)$). Cette zone hachurée illustre le fait qu'il n'y a pas une valeur unique de δP qui permet de garantir que la cellule coopérante apporte un gain de 100%. Cette figure met en valeur que l'augmentation de débit apporté par la coordination dépend du δP et en outre de l'interférence résiduelle. Plus l'interférence résiduelle $I+\mathcal{N}$ est élevée, plus la puissance reçue de la cellule coopérante P_c doit être proche de la puissance P_s (δP plus petit) pour respecter la garantie d'un gain de débit de 100%.

Dans le cas où le second schéma de transmission qui consiste à transmettre simultanément des données différentes des différentes cellules coopérantes est utilisé, l'inéquation devient:

$$\left(1+\frac{1}{\upsilon}\right)\left(1+\frac{\zeta}{\upsilon}\right) - \left(1+\frac{1}{\upsilon+\zeta}\right)^2 \geq 0 \tag{5}$$

Par conséquent, il ne faut pas faire coopérer une cellule si l'utilisateur est localisé dans une zone qui est fortement interférée par une ou plusieurs autres cellules. Dans ce cas, il est préférable que toutes les cellules interférentes coopèrent ou qu'aucune d'entre elles ne coopèrent.

Contrairement aux techniques connues qui définissent une cellule coopérante en se basant sur la différence entre les puissances reçues et qui nécessitent de prédéfinir un seuil δP, le procédé que nous proposons est basé sur le gain en débit pour déterminer les cellules coopérantes.

Nouvel algorithme de définition de cellules coopérantes

Soit $\mathbb{S} = \{s_1, s_2, \ldots s_{K_{max}-1}\}$ l'ensemble des cellules candidates à devenir cellules coopérantes avec la cellule serveuse lors de la transmission pour un utilisateur donné, classées par ordre de puissance décroissante. K_{max} est le nombre maximal de cellules coopérantes pour un seul utilisateur y compris sa cellule serveuse.

Le débit d'un utilisateur lorsqu'il est servi sans coordination est donné par:

$$r = \boldsymbol{F}\left(1+\frac{P_s}{\sum_i P_i + \mathcal{N}}\right).$$

$\boldsymbol{F}$ est une fonction croissante qui permet d'obtenir le débit à partir du SINR.

Nous désignons par r'_n le débit de l'utilisateur étant donné que les n premières cellules de l'ensemble $\mathbb{S} = \{s_1, s_2, \ldots s_{K_{max}-1}\}$, sont impliquées dans la transmission. Lorsque le premier schéma de transmission est appliqué:

$$r'_n = \boldsymbol{F}\left(1+\frac{P_s + \sum_{c=s_1,s_2,\ldots s_n} P_c}{\sum_{i\neq s_1,s_2,\ldots s_n} P_i + \mathcal{N}}\right).$$

L'algorithme de sélection de cellules coopérantes fonctionne comme suit:

Data: Ensemble $\mathbb{S}$ de cellules candidates à devenir cellules coopérantes
Result: Ensemble $\mathbb{C}$ de cellules coopérantes
$r'_0 \leftarrow r$
$i \leftarrow 1$
$j \leftarrow 0$
while $(i < K_{max})$ **do**
  if $r'_i/r - r'_j/r > (i-j) * \beta_T$ **then**
    $\mathbb{C} \leftarrow \mathbb{C} \cup \{s_{j+1}, \ldots s_i\}$
    $j \leftarrow i$
  end
  $i \leftarrow i+1$
end

Algorithm 1: Algorithme de sélection des cellules coopérantes.

β_T est la contrainte d'augmentation de débit par la coordination, qui doit être 100% dans le cas d'une topologie de réseau symétrique afin de maintenir la même condition de stabilité que dans le cas sans coordination.
Dans le cas général (topologie de réseau non symétrique, distribution de trafic non uniforme ...), l'algorithme peut être combiné avec un algorithme SON (Self Organizing Network), de sorte que la contrainte de gain de coordination qui correspond le mieux aux caractéristiques du réseau puisse être reconfigurée à chaque période.

Dans ce chapitre, nous étudions la technique de coordination JT pour deux schémas de transmission et différents types de clustering (intra-site, inter-site, clustering dynamique....)

Dans les cas d'une zone de coordination commune et de la coordination intra-site, nous proposons des modèles de niveau flot pour étudier les conditions de stabilité et le débit moyen de plusieurs politiques d'allocation de ressources. Nous proposons également une politique exploitant la mobilité des utilisateurs généralisant celle décrite au chapitre II.

De nombreux résultats numériques et des résultats de simulation sont présentés, conduisant aux conclusions suivantes: pour des scénarios à interférence faible la politique exploitant la mobilité proposée améliore les performances des utilisateurs et la capacité du système. Pour des scénarios à interférence forte, la technique JT apporte de vrais gains de performance lorsqu'elle est utilisée notamment dans le second schéma de transmission.

Nous proposons une nouvelle définition d'une cellule coopérante : au lieu d'utiliser un seuil fixe sur la puissance reçue, une cellule coopérante est définie comme permettant un gain en débit de 100 % pour l'utilisateur. Les résultats numériques confirment la pertinence de cette nouvelle définition.

Pour que le gain en coordination ne se fasse pas au détriment du gain opportuniste, nous proposons de plus un ordonnanceur global choisissant les utilisateurs à servir en résolvant un problème d'optimisation non-linéaire en variables binaires.

2-Dynamic Point Blanking

A l'aide de simulations numériques, nous montrons tout d'abord que pour toutes les techniques de clustering, la technique DPB permet un gain en terme de SINR, particulièrement significatif avec du clustering dynamique. Les résultats analytiques obtenus dans le cadre de la technique JT pour

différentes politiques d'allocation de ressources au moyen de modèles de niveau flot s'étendent directement à DPB quand il y a une zone de coordination commune.

Pour une cellule hexagonale interférant avec les cellules l'entourant, nous établissons une condition sur le gain de coordination qui est en général moins restrictive que pour JT, sous laquelle il y a un gain de capacité par rapport au cas sans coordination

Les résultats de simulation montrent des gains relatifs assez importants par rapport au scénario sans coordination, y compris en régime de fort trafic, notemment avec du clustering dynamique.

Mais ce schéma reste intéressant même pour une technique simple à implémenter comme la coordination intra-site, ce qui nous amène à conclure que la technique DPB offre un meilleur compromis entre performance et complexité d'implémentation que la technique JT.

Chapitre IV: Coordinated Beamforming CoMP

Le chapitre IV est consacré à la formation coordonnée de faisceaux au sein d'un même cluster. La formation de faisceaux est un procédé qui permet la transmission directionnelle de signaux en utilisant un réseau d'antennes à commande de phase. C'est l'une des techniques qui peuvent être utilisées pour atténuer l'interférence inter-cellule (ICI), un des principaux problèmes des réseaux cellulaires actuels.

Le signal peut être dirigé exactement sur l'utilisateur si la station de base peut localiser précisément chaque utilisateur (Position-based beamforming PB), ou bien de manière approchée en choisissant le déphasage dans un ensemble prédéfini (Codebook-based beamforming CB). Ce procédé permet donc déjà de réduire significativement les interférences, même en l'absence de coordination. La probabilité que les faisceaux activés dans les cellules voisines interfèrent entre eux est d'autant plus faible que les faisceaux formés sont étroits. Ces faisceaux sont d'autant plus étroits que le nombre d'éléments d'antennes est grand.

Si elle peut permettre d'envisager une réduction d'interférences supplémentaire, la formation coordonnée de faisceaux va se traduire par une complexité d'implémentation accrue. Nous cherchons alors à évaluer les gains supplémentaires attendus des techniques de coordination en comparant à l'aide de simulations à événements discrets les performances de la formation aveugle de faisceaux 2D (sans coordination) et celles de la formation coordonnée de faisceaux 2D.

Nous montrons que les ordonnanceurs distribués non-coordonnés donnent des performances similaires à celles des ordonnanceurs coordonnés lorsque le nombre d'antennes est grand, tout en ajoutant beaucoup de complexité au réseau.

Évaluation du SINR

La Figure 6 montre la fonction de distribution cumulative du SINR en fonction du nombre d'antennes N, tant pour la formation de faisceau CB que pour la formation de faisceaux PB et pour plusieurs nombres de faisceaux $N_B = N + 1, 3N/2 + 1, 2N + 1, \ldots \infty$. Elle illustre l'amélioration du SINR dans un réseau avec formation de faisceaux par rapport à un réseau sans formation de faisceaux. Le SINR moyen augmente significativement avec le nombre d'antennes. Le SINR moyen est d'environ 5 dB dans le cas sans formation de faisceau. Dans le cas d'un réseau avec formation de faisceaux, le SINR moyen augmente à 8-9 dB lorsque $N = 4$ et atteint 26-27 dB lorsque $N = 256$. Notez que le SINR moyen augmente d'environ 3 dB lorsque le nombre d'antennes est doublé. La probabilité d'expérimenter un SINR <20 dB dans un réseau avec $N = 256$ antennes est inférieure à 10%.

Évaluation du gain en capacité

La Figure 7 montre le gain en capacité d'un réseau avec formation de faisceaux par rapport à un réseau sans formation de faisceaux, pour un ordonnanceur RR. En d'autres termes, ce gain représente le gain en débit moyen d'un utilisateur quand toutes les ressources cellulaires lui sont allouées.

Des gains en capacité très significatifs sont obtenus dans un réseau avec formation de faisceaux. La capacité du réseau augmente significativement avec le nombre d'antennes. Elle augmente entre 80% et 100% approximativement lorsque le nombre d'éléments d'antennes est doublé.

Avec un nombre raisonnable de changements de phase la technique de formation de faisceaux CB a un gain en capacité très proche de celui de la technique de formation de faisceaux PB.

Ces gains en capacité ne tiennent pas compte de la surcharge de signalisation qui peut être ajoutée afin de déterminer le meilleur faisceau à activer pour servir un utilisateur donné.

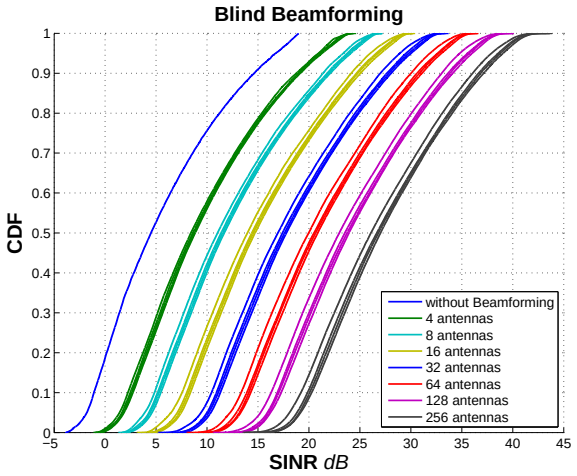


Figure 6: CDF du SINR dans un réseau avec formation aveugle de faisceaux.

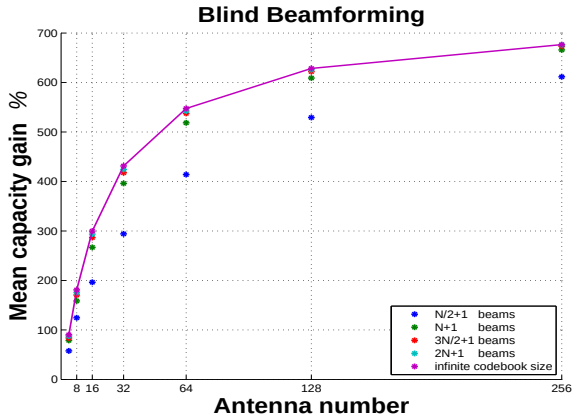


Figure 7: Gain en capacité dans un réseau avec formation de faisceaux.

Formation aveugle de faisceaux

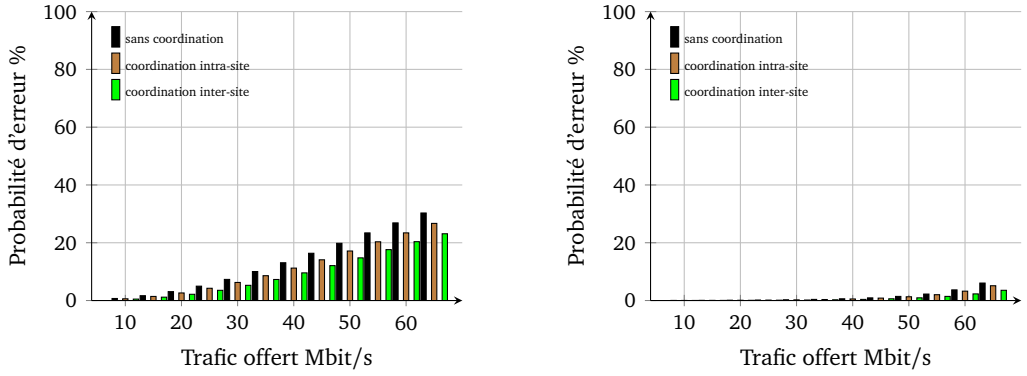
Nous comparons numériquement les débits moyens obtenus sous différentes politiques d'allocation de ressources. Des gains spectaculaires sont obtenus avec la politique RR même avec un nombre d'antennes relativement réduit, les politiques PF et max C/I permettant en outre un gain opportuniste en fort trafic.

Formation coordonnée de faisceaux

Nous supposons un ordonnanceur centralisé qui choisit les utilisateurs à servir dans chaque cellule en résolvant un problème d'optimisation non-linéaire en variable 0-1 afin de maximiser la somme des métriques des politiques opportunistes.

Avec un grand nombre d'antennes, la politique max C/I permet de réduire significativement la probabilité d'erreur de réception due à un mauvais choix du schéma de modulation et de codage (MCS). Ceci met en lumière qu'avec cette politique, même si les utilisateurs à servir sont choisis

sans aucune coordination par les différentes stations de base, il y a quand même une coordination implicite, les décisions locales tenant compte de la qualité du canal, et donc des interférences. Les résultats de simulation, dans ce cas, montrent des gains très modestes par rapport à la formation aveugle de faisceaux, qui ne justifient pas la complexité supplémentaire induite par la coordination.



(a) PF (b) max C/I

Figure 8: La probabilité d'erreur de réception dans un réseau avec formation aveugle de faisceaux avec et sans coordination, pour le PF et le max C/I.

Donc, un réseau entièrement coordonné garantit une probabilité d'erreur nulle. Cependant, nous pouvons voir que la probabilité d'erreur, quand $N = 64$ antennes sont déployées à la station de base, est très faible, en particulier avec un ordonnanceur max C/I, même dans un réseau avec formation aveugle de faisceau, comme le montre la Figure 8.

Ceci conduit à la conclusion suivante: dans un réseau avec formation de faisceaux où un grand nombre d'antennes est déployé, les ordonnanceurs standards non-coordonnés, et plus spécifiquement les ordonnanceurs max C/I, peuvent garantir une probabilité d'erreur négligeable sans rajouter de complexité au réseau.

Cependant, dans le cas où un nombre limité d'antennes est déployé, une coordination peut être nécessaire car les faisceaux sont plus larges dans ce cas. Ainsi la probabilité d'erreur devient plus grande.

Il y a donc un compromis entre le nombre d'antennes à déployer et la complexité de l'ordonnanceur.

Conclusion

Dans la première partie de la thèse, nous avons étudié l'impact de la mobilité dans les réseaux cellulaires, en présence de trafic élastique et de trafic streaming adaptatif, en absence des mécanismes de coordination inter-cellule.

Nous avons montré que les variations lentes des conditions radio dues à la mobilité peuvent être exploitées en présence de trafic élastique. Nous avons quantifié les gains induits par la mobilité dans les réseaux de données cellulaires pour les stratégies d'ordonnancement RR, max C/I et PF.

La performance niveau flot s'améliore au fur et à mesure que la mobilité des utilisateurs augmente. Cependant, le choix optimal de la stratégie d'ordonnancement dépend principalement de la mobilité.

Ainsi, nous avons proposé un ordonnanceur qui s'adapte à la mobilité des utilisateurs: il tend à être plus opportuniste lorsque tous les utilisateurs sont mobiles et plus équitable quand tous les utilisateurs sont statiques. Les utilisateurs mobiles en bordure de cellule sont dépriorisés étant donné qu'ils sont susceptibles de se déplacer et d'être servi dans de meilleures conditions radio. Nous avons montré que cet ordonnanceur offre un bon compromis dans un réseau avec plusieurs types de mobilité en améliorant les performances globales, en particulier à forte charge.

Nous avons examiné également la performance des différentes stratégies d'ordonnancement en présence du trafic streaming adaptatif, en se basant sur deux indicateurs de performance que nous avons défini. Nous avons proposé un ordonnanceur qui offre un bon compromis entre ces deux indicateurs de performance.

La deuxième partie de cette thèse concerne les deux schémas de coordination en downlink, le Joint Processing JP et le Coordinated Scheduling/Beamforming CS/CB.

Nous avons d'abord étudié les techniques JP, plus précisément JT et DPB. Nous avons évalué les performances de ces deux techniques à l'aide de modèles markoviens et des simulations du système.

En effet, la performance dépend fortement du niveau d'interférence dans le réseau par le gain de coordination qui est le gain de débit apporté par la coopération d'une cellule à un utilisateur en bordure de cellule.

Nous avons étudié un environnement à forte interférence, où contrairement à un scénario à interférence faible, nous avons montré que JT pourrait être intéressant en appliquant un schéma de transmission qui garantie un gain de coordination moyen suffisamment élevé pour compenser la consommation supplémentaire de ressources engendrée. Les gains dépendent principalement du type de clustering et du schéma de transmission ainsi que de la charge du réseau.

Nous avons analysé les problèmes de capacité en présence du JT et avons proposé un nouvel algorithme de sélection de cellules coopérantes où au lieu d'utiliser un seuil fixe sur la puissance reçue, une cellule coopérante est définie comme permettant un gain de débit de 100 % pour l'utilisateur dans le cas d'une topologie de réseau symétrique. Dans le cas général où le réseau n'est pas symétrique, un schéma de coordination plus avancé combiné avec un algorithme SON peut être utilisé. L'algorithme SON sera chargé de calculer et de mettre à jour la contrainte de gain de coordination qui convient le mieux aux caractéristiques du réseau.

Nous avons étudié différentes stratégies d'ordonnancement. Nous avons montré que la politique d'ordonnancement est un élément essentiel qui affecte fortement les performances en présence des techniques de coordination. Pour que le gain en coordination ne se fasse pas au détriment du gain opportuniste, nous avons proposé d'effectuer l'ordonnancement d'une façon optimale.

La performance du DPB est également évaluée dans plusieurs cas selon différentes méthodes de clustering. Nous avons montré que la plus grande partie du gain de débit d'un utilisateur se trouvant en bordure de cellule est obtenue par l'élimination de l'interférence et que le DPB peut présenter une

condition de stabilité moins restrictive que celle en présence du JT. Nous avons montré que DPB est une technique simple et performante.

Nous avons également étudié le cas d'un scénario à faible interférence où nous avons montré qu'il n'est pas recommandé d'activer la fonction JP. Toutefois, si on tient à assurer l'uniformité de service notamment pour les utilisateurs statiques en bordure de cellule souffrant de très mauvaises conditions radio, il vaut mieux de ne pas effectuer la coordination pour les utilisateurs mobiles qui sont capables de se déplacer et d'améliorer leurs conditions.

Nous avons ensuite considéré le cas d'un réseau avec formation de faisceaux afin d'évaluer les gains attendus de la coordination de faisceaux. Nous avons d'abord étudié le cas de la formation de faisceau aveugle, où aucune coordination n'est effectuée entre les différentes cellules. Les gains en capacité maximaux prévus dans le cas d'un ordonnanceur RR ont été quantifiés compte tenu des différents nombres d'antennes. Nous avons également étudié les performances en termes de débit moyen des utilisateurs. Ce procédé permet donc déjà de réduire significativement les interférences.

Nous avons ensuite étudié la performance de la coordination de faisceaux effectuée à travers des ordonnanceurs coordonnés. Si elle peut permettre d'envisager une réduction supplémentaire de l'interférence, la formation coordonnée de faisceaux va se traduire par une complexité d'implémentation élevée. Nous avons montré que les ordonnanceurs distribués non-coordonnés donnent des performances similaires à celles des ordonnanceurs coordonnés lorsque le nombre d'antennes est grand, tout en ajoutant beaucoup de complexité au réseau. Il y a donc un compromis à considérer entre le nombre d'antennes à déployer et la complexité de l'ordonnanceur.

Contents

Chapter I

Introduction

I.1 Wireless technology evolution

Advances in wireless communications have played an important role in making it become ubiquitous. In the recent years, the number of mobile phones and wireless Internet users has risen notably. This is mainly due to the developments in wireless networks since its first generation which only handled voice and data communications occurring at low data rates.

The evolution in user needs has led to optimizing telecommunication technologies. As a matter of fact, voice calls were the most demanded service in fixed and wireless networks in the mid-1980s. At the beginning of the 1990s, the Global System for Mobile Communications (GSM) standard emerged and was initially designed and optimized for voice transmission. It was at the basis of an precedented change in telecommunication between people. The Second-generation (2G) cellular networks were launched using this standard and brought to the world, digital voice transmission, SMS and mobile Internet. This is why GSM had a great success worldwide with over 3 billion subscribers in 2010 compared to analog wireless systems which were used by a few people. Several applications emerged thanks to these advances like email which was enabled and reinforced by mobile Internet, and people-to-machine connectivity which was its first appearance even if the speed was much lower than that of today.

Initially, GSM network was designed as a circuit-switched system. In such a system, a direct and exclusive connection is set up between two users on every interface involving all the network nodes. Gradually, this physical circuit-switched systems have been virtualized and, today, the interconnection of many network nodes is being done over IP-based broadband connections.

Starting mid-1990s, The Internet started increasing in importance. To address this new need, General Packet Radio Service (GPRS) was developed to enhance the GSM standard with regard to the efficiency in data transport and to enable Internet access for wireless devices. Later, speed and latency were further optimized giving birth to the so-called Enhanced Data rates for GSM Evolution (EDGE).

After that, the third generation wireless telecommunication system emerged implementing the Universal Mobile Telecommunications System (UMTS). Indeed, GSM was just the beginning and at the end of the 1990s, designers constructed a new system that went far beyond the capabilities of GSM and GPRS. This sytem using the so-called UMTS combines the properties of a circuit-switched voice network with those of a packet-switched data network. Thanks to its properties, UMTS offers a multitude of new functionalities compared to 2G systems. Despite this fact, at first, these function-

alities were not valued and the adoption of 3G was slow. Indeed, primary applications consisted in Voice and SMS and 2G speeds were still acceptable for email. Later, however, the increased adoption of smartphones made the consumers have a more developed taste for nomadic computing especially for video streaming. From that time on and over years, the UMTS radio network system continued to evolve and to be further enhanced to the point of offering broadband speeds far beyond the original design. These high-speed enhancements are referred to, in chronological order of appearance, as High-Speed Packet Access (HSPA) and then HSPA+.

Although UMTS has been evolving, it has a number of inherent design limitations similar to that of GSM and GPRS which are a decade older. Hence, the Third Generation Partnership Project (3GPP) decided to rethink the design of both the radio network and the core network once again. The result of their efforts produced a new standard commonly referred to by Long-Term Evolution (LTE).

For the LTE Release 8, the key requirements can be summarized as follows:

- delays are reduced with regard to connection establishment and transmission latency
- user data rates are raised
- cell-edge bit-rate are enhanced for a more uniform provisioning of the service
- cost per bit are decreased for more spectral efficiency
- new and pre-existing bands dispose a greater flexibility of spectrum usage
- network architecture is simplified
- mobility is made seamless, including between different radio-access technologies
- the mobile terminal should consume power reasonably

LTE presents several improvements over UMTS. For instance, UMTS had a specified air interface named Wideband Code Division Multiple Access (WCDMA) with a carrier bandwidth of 5 MHz. This interface had a very good performance within this limit. However, scalability was an issue: the time between two transmission steps has to be lowered every time the transmission rate is increased. With LTE, the air interface has been designed differently to address mutlipath fading. To deal with this issue, LTE uses the so-called **Orthogonal Frequency Division Multiplexing** (OFDM). With this technology, data is transmitted over many narrowband carriers of 180 kHz each instead of spreading one signal over the complete carrier bandwidth (e.g., 5 MHz). Thus, a data stream is split and is transmitted by simultaneously sending slower data streams splits rather than a single high data rate transmission. LTE has several bandwidths that can be used ranging from 1.25 MHz up to 20 MHz. Even though one bandwidth will be used in practice, devices supporting LTE must be able to handle all the bandwidth range. Then, it will be up to the network operator to set the one that will be used based on the frequency band and the amount of spectrum available within its network. For instance, if the signal conditions are good, data rates can be higher than 100 Mbit/s in a 20 MHz carrier.

In addition to this bandwidth related flexibility, LTE allows devices that supports it to handle Multiple Input Multiple Output (MIMO) transmissions. Spatial-domain is a new dimension that can be exploited as well thanks to the use of **multiple antenna technology**. This allows to achieve an increased spectral efficiency. Indeed, it has been proven that in theory, the scaling of achievable spectral efficiency is linear with respect to the minimum of the number of transmit and receive antennas used. Even though, sometimes, exploiting the variety of features offered by a multiple antenna technology seems promising from a theoretical perspective, it does not always perform as expected when implemented in practical systems.

Indeed, Figure I.1 illustrates how multiple antennas can be exploited through three main principals. These main benefits can be summarized as follows:

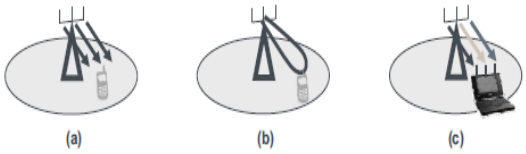


Figure I.1: Multiple antenna technology benefits: (a) diversity gain; (b) array gain; (c) spatial multiplexing gain.

- The *Diversity gain* is obtained by using the spatial diversity provided by the multiple antennas. It allows to enhance transmission robustness with respect to multipath fading.
- The *array gain* is achieved by concentrating energy in one or more given directions via precoding or **beamforming**. This also allows multiple users located in different directions to be served simultaneously through the so-called multi-user MIMO (MU-MIMO).
- The *spatial multiplexing gain* is related with the capability to use multiple spatial layers to deliver multiple streams to a single user. Spatial multiplexing allows to achieve higher data rates compared to those of a single-stream transmission, under very good signal conditions.

Uplink and downlink transmissions are separated using Frequency Division Duplex (FDD) in the implementation of LTE in most countries. However, some use time to separate the uplink and downlink transmissions which use the same carrier. This done by assigning what is known by Time Division Duplex (TDD) to network operators. Specifications for both FDD and TDD coexist in the same LTE standard and its was at the same time that both air interface types began to be deployed. All layers higher than 1 and 2 on the air interface are not influenced by the implementation choice of one of the two modes.

Compared to older systems and standards, the **all-IP approach** is another major change that is introduced in LTE. Its consists in relying only on an IP-based core network. This is different from the traditional circuit-switched packet core supported by UMTS to deliver voice services, SMS or any other GSM related services, in the sense that it is simpler to implement. Indeed, in LTE, the design and implementation of the air interface, the radio network and the core are simplified with an all-IP network architecture.

Quality of Service (QoS) mechanisms have been standardized on all interfaces to ensure that the requirements of voice calls for a constant delay and bandwidth can still be met when capacity limits are reached. Despite the significance of these efforts from an architectural perspective, offering voice services over LTE is still not clear having all the different possible options and their pros and cons.

I.2 LTE Network architecture

LTE has a high-level network architecture made up of three main components: the User Equipment (UE), the Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) and the Evolved Packet Core (EPC). This architecture is illustrated in Figure I.2.

While the EPC allows the communication with the outside world such as the Internet, the access network called E-UTRAN, interconnects the UE and the EPC.

The E-UTRAN is a network of evolved Base stations called eNodeBs (see Figure I.2). Its architecture is said to be flat for normal user traffic (in contrast to broadcast) due to the absence of any centralized controller.

The eNodeBs are interconnected with each other by means of an interface named X2. This interface is used for all signaling information and data exchange required by the different functionalities.

The E-UTRAN is responsible among all other radio-related functions for, Radio resource management (RRM). RRM includes all functions having connection with radio bearers, scheduling and dynamic allocation of resources to UEs. These functions reside in the eNodeB which can be in charge of managing multiple cells. A site is generally used to refer to a group of cells managed by the same eNodeB.

In the LTE architecture, the access network (E-UTRAN) is the main focus of this thesis.

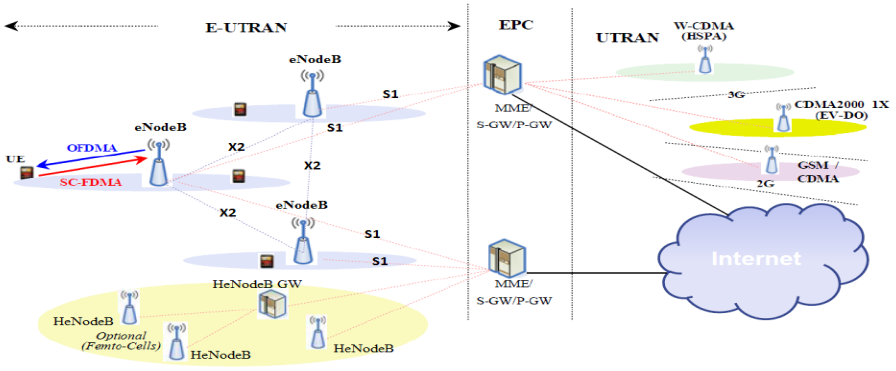


Figure I.2: LTE Network architecture.

I.3 Resource allocation-Scheduling in LTE

In the eNodeB layer 2 radio protocol stack, a key component is the MAC Scheduler. It is in charge of scheduling radio resources used in the downlink and uplink in a cell. Radio resource scheduling consists in allocating the available radio resources within the cell to specific UEs for the purpose of transmission and reception of Transport Blocks (TB).

The basic LTE downlink physical resource can be represented by a time-frequency grid, as illustrated in Figure I.3.

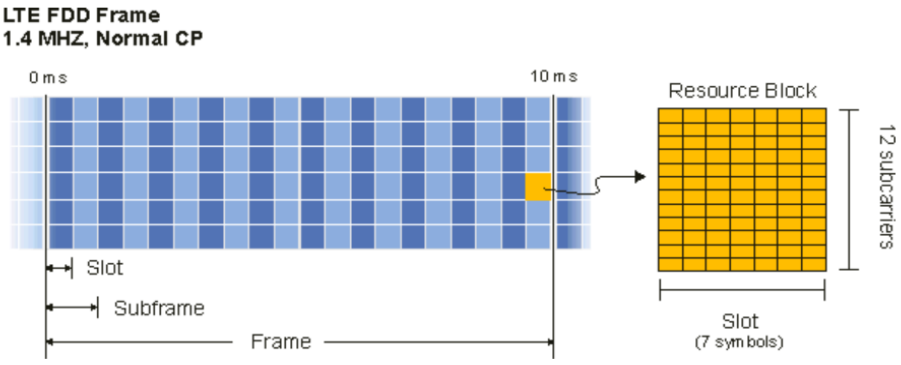


Figure I.3: LTE Frame.

The OFDM symbols are grouped into so-called resource blocks (RB). An OFDM symbol consists of 12 resource elements (RE). The RBs have a total size of 180kHz in the frequency domain and 0.5ms in the time domain; it is the smallest unit of resources that can be allocated to a user. A physical resource block (PRB) consists of two RBs and the size of a PRB is equal to 1ms in the time domain.

Each user is allocated a number of RBs in the time-frequency grid as shown in Figure I.4. Thus, the bit rate of a user depends on the number of RBs this user is allocated, and the modulation order used in the REs. The allocation of RBs depends on advanced scheduling mechanisms in the frequency and time dimensions. OFDMA is the multi-user OFDM technology where, unlike OFDM, a single user does not necessarily need to occupy all the sub-carriers at any given time. In other words, a subset of RE is assigned to a particular user.

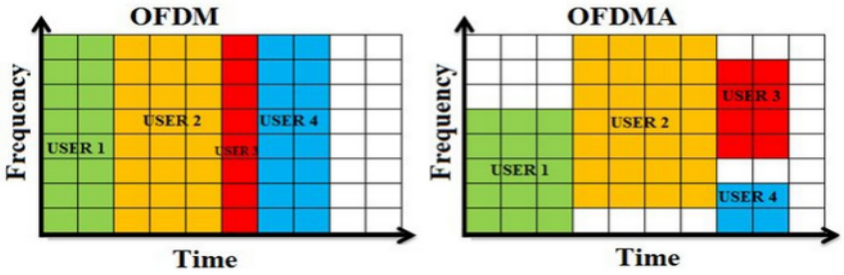


Figure I.4: Resource allocation.

The downlink modulation and coding scheme (MCS), used in the REs, depends on the radio conditions experienced by the receiving UE. The MAC Scheduler receives Channel Quality Indicator (CQI) reports from all UEs in the cell based on their measurements of the downlink channel. The reported CQI is a number between 0 (worst) and 15 (best) indicating the most efficient MCS which would give a Block Error Rate (BLER) of 10% or less.

CQI reports from UEs can take the following forms:

- Wide-band CQI reports based on measurements across the entire downlink channel.
- Sub-band CQI reports based on measurements across subsets of the downlink channel (subset of RBs).

In addition to the MCS, CQI reports can also be used by the scheduler to assign resources under channel-aware scheduling that allocates the radio resources corresponding to the best channel conditions for individual UEs.

The scheduling strategy is a key component of any radio access network. It can range from ensuring complete fairness to complete opportunism by fully exploiting the instantaneous channel state information (CSI).

I.4 LTE-Advanced main features

Despite the fact that the requirements set by ITU-R for the IMT-Advanced designation were largely addressed by LTE Releases 8 and 9, LTE Release 10 was issued to deal mainly with increasing capacity. This new release evolves LTE towards what is known by LTE-Advanced.

LTE–Advanced targets at the same time, a better spectral efficiency, increased bit rates and an enhanced performance at the cell-edge while meeting all the requirements of 4G system defined by ITU. To fulfill these requirements, several functionalities were developed in LTE-A among which five

essential ones that outstands LTE-A (release 10) from previous releases. These five functionalities are:

Carrier Aggregation : This is a major new aspect in LTE-A. It consists in allowing aggregated bandwidth which can reach up to 100MHz by cumulating different bandwidth of at most five component carriers. This feature is an ideal solution to operators that do not possess a contiguous chunk of 100 MHz spectrum.

Higher order MIMO : MIMO is a function that uses multiple transmission and receiver antennas to achieve a higher bit rate. This functionality already existed in LTE. However, LTE can only handle 4x4 MIMO configuration even if 2x2 is the most commonly used one. For LTE-A, it is possible to support 8x8 configurations in the downlink and 4x4 in the uplink. Handling higher order MIMO is a key for enhancing spectral efficiency and throughput. It has been theoretically proved that 8 spatial streams can reach speeds that are 8 time faster than a single input single output (SISO) system .

Relay nodes and Heterogeneous networks : Relay nodes are low power base stations used to enhance coverage and capacity at cell edges. These nodes act as repeaters to enhance the signal quality and rebroadcast the signal. They are used in the context of a heterogeneous network (HetNet).

Enhanced Inter-Cell Interference Coordination : This coordination is a mechanism that manages and mitigates interference occurring typically in a heterogeneous network due to macro and pico cells that transmit and receive data simultaneously.

Coordinated Multipoint (CoMP) Transmission : This functionality is formalized in 3GPP Release 11. It is yet another essential functionality that characterizes an LTE-A network. A Coordinated multipoint transmission and reception scenario consists in a dynamic coordination between multiple eNodeBs in order to prevent interference with other transmission signals. This improves network coverage as well as quality for cell edge users.

I.5 Coordinated mutipoint CoMP

When the frequencies of downlink transmission between neighboring cells are different, like in the conventional cellular systems, the inter-cell interference will be a minor problem. This type of frequency reuse is common in older generation of wireless networks as it improves the Signal to Interference plus Noise Ratio (SINR). However, the gain achieved with this SINR improvement is lower than the equivalent gain in the case of a frequency reuse of one.

The scarcity of bandwidth and spectrum in wireless systems has driven the need for spectrally efficient communication systems. The performance of a communication system is typically measured in terms of spectrum efficiency in bit/s/Hz/unit-area. Thus, LTE is designed to operate with a frequency reuse factor of one. However, inter-cell interference caused by neighboring base stations has been marked as one of the most severe problem towards the deployment of LTE technology, as it can significantly deteriorate the performance of cell edge users.

Dealing with this issue is challenging, especially in the context of high resource reuse. Indeed, recent wireless systems are constantly upgraded with new features that aim to improve both the users' experience and the system capacity while ensuring uniformity of service provision. In this context, CoMP is considered as a promising technique for LTE-A to satisfy the system requirements in terms of capacity and cell edge user throughput.

I.5.1 CoMP concept

The main purpose of this concept is to help those users at the cell edge, see Figure I.5 who are suffering from a lot of interference from the neighboring cells, through a coordination between several cells, mainly the serving cell and the interfering neighboring cell(s).

These cells coordinate between each other in such a way that the transmission signals to other users do not incur serious interference to the so-called *CoMP user*, that is the user at the cell-edge affected by the interference. This can be done either by muting an interfering cell, or by activating a beam that does not interfere with the scheduled CoMP user, in the neighboring cooperating cells in the case of a system with beamforming, or even by converting an interfering signal into a meaningful signal. This coordination is supposed to improve the link performance of the cell edge users so that the spectral efficiency can be increased.

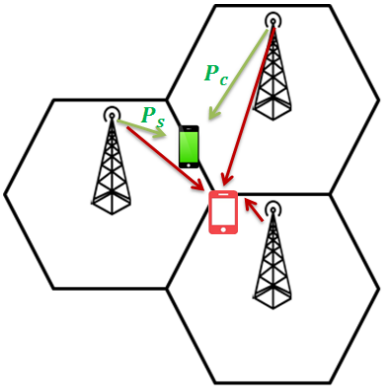


Figure I.5: Cell coordination.

I.5.2 CoMP categories

In this work, we focus on downlink CoMP transmission where two transmission schemes are mainly considered see [6], as shown in Figure I.6: Joint Processing (JP) and Coordinated Scheduling/Beamforming (CS/CB).

JP is further categorized into three main techniques: Joint transmission (JT), dynamic point blanking (DPB) and dynamic point selection (DPS).

In **JT** two or more cells transmit simultaneously to a CoMP user in a coherent or non-coherent manner. The same RB(s) of the Physical Downlink shared Channel (PDSCH) is transmitted from multiple cells to the CoMP user. Thus data should be available at multiple transmission points. The applicability of such a technique depends to a great extent on the backhaul characteristics in terms of latency and capacity as signaling and UE data exchange among coordinated cells is required.

DPS is performed through dynamic switching between the serving cell and the coordinated cell according to channel variation. Only one cell among the coordinated cells, that is the cell corresponding to the best radio link, transmits the RB(s) to the CoMP user while the other cells can allocate their resources to other users. In general, the transmitting cell is dynamically selected through fast scheduling at the central base station. In other words, we can see this technique as a kind of performing handover at small time scale.

DPB consists in muting the interfering cooperating cells in order to eliminate the interference. Only the cell corresponding to the minimum pathloss transmits to the user. Thanks to the handover

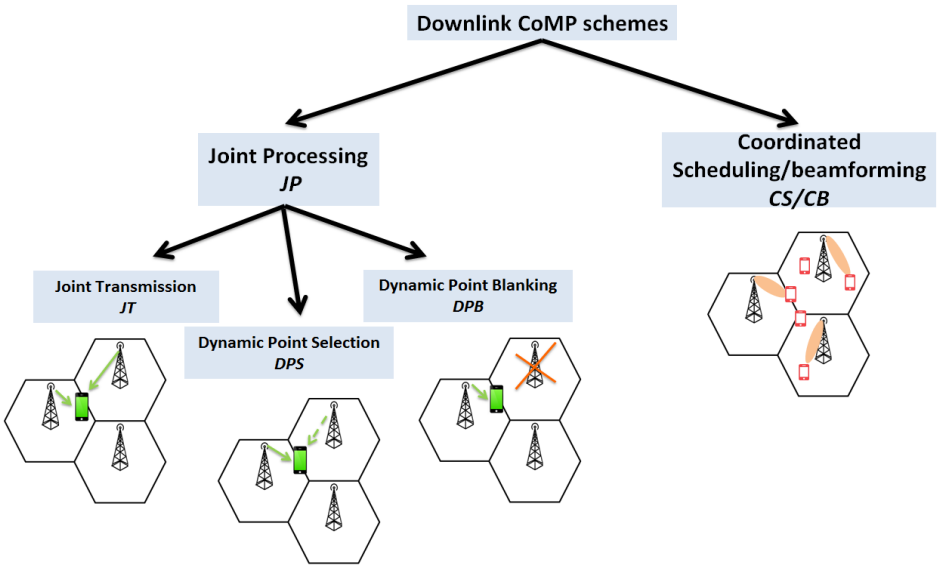


Figure I.6: Downlink CoMP categories.

procedure, the user is generally associated with the cell with the minimum pathloss (and shadowing) that is the maximum mean Reference Signal Received Power (RSRP). The transmitting point is then nothing more than the serving cell. In this case, only signaling information exchange is needed, putting reasonable amount of burden on the back-haul of the system.

On the other hand, **CS/CB** as the name implies, can only be deployed in a system supporting beamforming. The scheduling is done by a coordination between different coordinated cells, so that the set of scheduled users in all coordinated cells corresponds to the best set of users that can be scheduled together when performing beamforming. In other words each cell activates a beam which does not interfere with the scheduled user within the neighboring coordinated cell, reducing the unnecessary interference to other UE scheduled within the coordinated cells.

More advanced CoMP schemes can be used by combining different techniques together. DPB can be used in conjunction with DPS. CS/CB can be for instance combined with JT and so on...

The main weakness of JP coordination techniques, especially JT and DPB is the extra resource consumption that may alter the ability of the system to process all traffic especially at high load. So we should ensure that the introduction of such a new feature does not degrade the existing performance by carefully choosing the best scenario where it is interesting to have these techniques. Unlike CS/CB, JP techniques can be deployed in a system with or without beamforming. However, a beamforming capable network is highly interference limited. In such a network the main purpose for which JP coordination techniques are proposed, no longer exits. We shall see that JP techniques should be preferably used in an interference-dominated environment, that is a network without beamforming.

I.5.3 Cooperating cell definition

As only some neighboring cells should cooperate in order to serve a given cell-edge user, the question is which base stations should cooperate? Thus a definition of a coordinated cell is needed. Usually, this is done based on the averaged RSRP, see [64, 28, 58]. Once the difference between the average

power received from the neighboring cell P_c and that received from the serving cell P_s is less than a given predefined threshold δP (see Figure I.5), the neighboring cell is defined as a cooperating cell for the considered user which is consequently defined as a CoMP user:

$$|P_s - P_c| < \delta P \tag{I.1}$$

This can be done at the base station leveraging the measurement report messages defined in the LTE specifications. Generally, the value of δP can vary between $3dB$ and $18dB$. We shall see that the optimal value depends on many factors: the scenario, the transmission scheme, the clustering method... A neighboring cell fulfilling this condition can cooperate given that it takes part in the cooperative cluster of the serving cell, see Section I.5.4. Thus the cooperating cells of a user are configured and updated over time based on long-term UE power measurement; it is not expected to change over time as long as the location of the UE does not change.

I.5.4 Clustering

In order to know which base stations are allowed to cooperate between each other, clusters should be defined [56, 36, 16]. There are two basic categories how to cluster base stations as shown in Figure I.7 (example of an hexagonal topology): static and dynamic clustering.

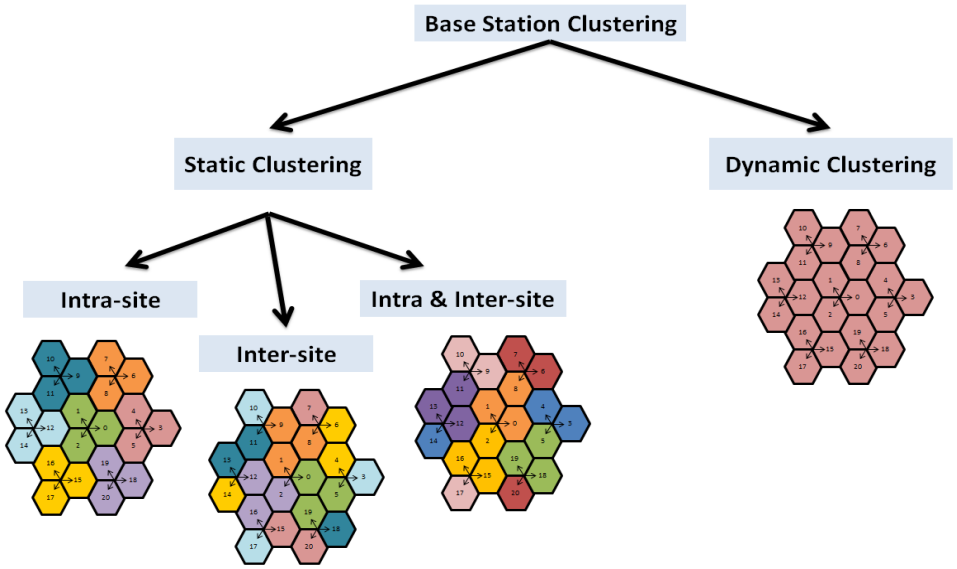


Figure I.7: Base stations clustering categories.

In the **static** approach, clusters do not change over time and are usually chosen based on geographical criteria, such as:

- Intra-site coordination where coordination is performed between the cells (sectors) of the same site controlled by the same macro base station so that no back-haul connection is needed in this case.
- Inter-site coordination where cells belonging to different radio sites can cooperate between each other. In this case data and signaling information exchange is performed through the X2 interface.

- Intra and Inter-site coordination represents a mix of the two previous cases.

There are so many other examples ...We consider that there is no coordination between clusters, which also means that a cell belongs to only one cluster. The main drawback of this approach is that clusters are not optimized for individual users. A user can be strongly interfered by cell which does not belong to the cluster of its serving cell, and thus not allowed to cooperate with its serving cell.

The **dynamic** clustering approach consists in dynamically selecting the best cooperating cells set for a given UE, so that any neighboring cell fulfilling the previous condition explained in I.5.3 cooperates. Ideally, a fully coordinated network could guarantee interference-free operations. However, this creates additional overhead for deciding which set of cells fits best for a certain UE. Moreover, it would require to exchange and process an enormous amount of data and signaling information. Thus, it is more reasonable to assume that only a limited number of cells constitute the cooperative cluster.

I.6 Performance Analysis

In order to evaluate performance of several coordination schemes, scheduling strategies as well as proposed solutions, we rely in this thesis on flow level models based on queuing theory as well as system-level simulations.

I.6.1 Queuing theory

We only present here some key concepts which are useful in the context of this thesis.

A queue is characterized by several parameters, like the number of servers, the queue capacity in number of customers, and the statistical characteristics of customer arrivals and service times. For convenience, the following code invented by Kendall is commonly used: A/S/m[/n] with:

- A, letter denoting the distribution of inter-arrival times;
- S, letter denoting the distribution of service times;
- m, number of servers which is possibly infinite;
- n, maximum number of customers in the queue which is infinite by default.

The typical distributions of the inter-arrival and service times A and S are designated by the following notation: M stands for Markovian and refers to an exponential distribution, D stands for deterministic distribution, i.e. a fixed service/inter-arrival time, E stands for Erlang distribution, it corresponds to a sum of exponentials, H stands for Hyperexponential, it corresponds to a random choice among exponentials, and G stands for general unspecified distribution.

Unless otherwise speci

ed, the customers are served in their order of arrival (First In First Out).

For example, the M/D/1 denotes a single-server queue, without any limit on the number of customers, for which customers arrive according to a Poisson process (exponential inter-arrival times) and have deterministic service times.

We denote by λ the arrival rate of customers in the queue; the mean inter-arrival time is thus equal to $1/\lambda$. Similarly, we denote by μ the departure rate of customers from a busy server, or service rate; the mean service time is thus equal to $1/\mu$.

We denote the load as the ratio of the arrival rate to the total service rate of the queue, that is for m servers:

$$\rho = \frac{\lambda}{m\mu}.$$

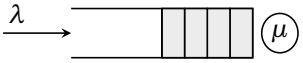


Figure I.8: A single server queue.

This is a dimensionless quantity. The infinite capacity queue ($n = \inf$) is said to be stable whenever $\rho < 1$: the arrival rate must be less than the total service capacity. This condition is necessary to guarantee a low reject probability for finite capacity queues ($n < \inf$).

The service discipline defines the order in which customers are served. Common service disciplines are the following:

- FIFO (First In First Out): customers are served in their order of arrival;
- LIFO (Last In First Out): customers are served in their inverse order of arrival, the service of any customer is interrupted by the arrival of a new customer in the queue;
- PS (Processor Sharing): customers are served simultaneously, with fair sharing of the server(s).

PS service discipline is also useful for modeling IP networks; it consists in a fair bandwidth sharing between active data flows. This service disciplines have the so-called insensitivity property, for which the stationary distribution of the state of the queue is independent of the distribution of service times beyond the mean.

Queues are omnipresent in packet-switched networks. They are at the heart of any computer, switch, router, access point. This is the place where sharing policies are implemented through packet scheduling and active queue management. More generally, a set of data flows sharing the same capacity (of a cell for instance) may be viewed as a virtual queue, the service required by each flow corresponding to the transfer of some data volume.

I.6.1.1 Little's formula

Consider a queue in steady state. If the arrival rate of customers in this queue is equal to λ, we have the following simple relationship between the mean number of customers in the queue and the mean time spent by each customer in the queue, δ:

$$E(X) = \lambda\delta \tag{I.2}$$

This is Little's formula, valid for any queue in steady state.

I.6.1.2 The M/M/1 queue

The simplest and most common queue is the M/M/1. Customers arrive according to a Poisson process of intensity λ and require services whose duration has an exponential distribution with parameter μ. There is a single server and the queue is of infinite capacity. The number of customers in the queue forms a birth-death process, with birth rate λ and death rate μ. The transition graph of this process is represented by Figure I.9. The load of this queue is simply given by the ratio of arrival rate to service rate:

$$\rho = \frac{\lambda}{\mu}.$$

The queue is stable in the sense that the number of customers does not grow indefinitely when the

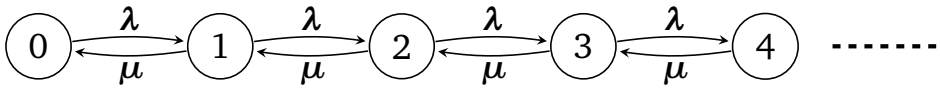


Figure I.9: State diagram of the M/M/1 queue.

birth rate is strictly less than the death rate. The stability condition can thus be expressed as

$$\rho < 1.$$

Under this condition, system reaches a steady state and the stationary probability $\pi(x)$ to have x customers in the system can be derived by writing the balance equations (i.e., for every state, the probability of leaving the state must be equal to the probability of entering the state). Based on these balance equations, we obtain the stationary measure. By adding the normalization equation, we get the stationary distribution:

$$\pi(x) = (1-\rho)\rho^x \qquad \forall x \in \mathbb{N}.$$

In particular, the fraction of time the server is busy, $1-\pi(0)$, is equal to the queue load, ρ. This result is also satisfied by the M/G/1 queue. The mean number of customers in the queue is given by:

$$E(X) = \sum_x x\,\pi(x) = \frac{\rho}{1-\rho}$$

As the load approaches 1, the mean number of customers tends to infinity.

From Little's formula (I.2), we deduce the mean delay or the mean sojourn time of a customer in the queue:

$$\delta = \frac{1}{\mu-\lambda}$$

We have so far referred to the entities arriving and departing from the queue as customers. In communication networks, these customers may be users or a set of flows that share the capacity of a cell.

In this thesis, we focus on flow-level dynamics of users in order to evaluate the performance of a network where a flow represents the transfer of a particular document. The term flow refers to a continuous stream of packets using the same path in a network and characterized by the starting time and the size.

We consider two types of flows: elastic flows and adaptive streaming flows. Elastic flows result from the transfer of digital documents like web pages and peer-to-peer file sharing, characterized by a fixed size (i.e. the volume of data to be transferred) and a variable duration which depends on the network conditions (radio conditions, load). Streaming flows are produced by video applications. They are characterized by a fixed flow duration. The video bit rate can be adapted according to channel conditions.

I.6.2 System-level simulations

In order to evaluate the performance of new mobile network technologies, system-level simulations are crucial. Flow level modeling is a very good tool that allows to predict system performance and to show the relevance of some solutions compared to others. However, the models analysis, when

dealing with complex systems, may be insufficient in some cases and need to be validated by more realistic simulations as models may sometimes ignore important aspects of the network. Along with the standardization process, commercially available LTE simulators have been developed. Equipment vendors, to this effect, have also implemented their own, proprietary solutions. Some universities and research centers have also developed such simulators. In system-level simulations the physical layer is abstracted by simplified models that capture its essential characteristics with high accuracy and simultaneously low complexity. However, unlike the modeling tools, the main issue that can be faced in system simulations is the computational time and memory that would be required. The key point is to find simulation methods and to make the necessary approximations and assumptions where necessary without sacrificing from being realistic. As one of the main objectives of my thesis is to evaluate the performance of the state-of-the-art CoMP solutions from a system level point of view, I have developed a C++ LTE simulator with some simplifications and approximations which allow to have a good trade-off between accuracy of results and simulation time/memory consumption. A more detailed description of this simulator is provided in Appendix A.

I.7 Motivation

As interference is still the main limiting factor in cellular networks, several new approaches and features are being considered as key elements to cope with this issue. The strongest motivation for such new features is to increase the cell-edge user throughput and to ensure high data rate coverage.

Cell coordination [7, 37], more precisely the JP scheme, has been proposed as an efficient way of reducing inter-cell interference by either silencing some base stations or allowing several base stations to transmit data simultaneously to the same user. But this is at the cost of higher resource consumption [43], increased traffic in mobile backhaul networks [18, 17] and many technical issues in terms of synchronization and feedback design leading to several practical challenges [49]. This solution may even lead to performance degradation, see for instance [43]. Moreover, the scheduling strategy is a key element that strongly impacts the performance in the presence of CoMP schemes, see for instance [3, 4, 53, 28, 77].

The current literature on CoMP transmission/ reception has not reached a consensus on the applicability and the capacity issues of this feature, [6, 68, 49, 64]. JT is considered as the most promising JP technique and less attention has been paid to DPB in the literature, see for instance [59]. A complete comparative study between these two JP techniques cannot be found in the literature. It is common to use full buffer traffic models, assuming *static* or *semi static* users because these models are typically easier to simulate. However, it is important to use non-full buffer traffic models when dealing with cooperation techniques because the random nature of traffic can lead to entirely different system behaviors and performance results. Mobility of users (see [22, 27, 26]) may also have a critical impact on the performance of these coordination schemes as users can move from an area where coordination is performed to another where coordination is not performed and vice versa.

CS/CB on the other hand seems to be an efficient scheme that allows cells to choose the beams in such a way that interference is reduced, see for instance [75, 51, 31]. However, coordinated schedulers add significant processing complexity [54], feedback and signaling overhead. The question is then: is it really worth it ? How much gain can be expected from these techniques?

So it will be interesting to have a study that allows to compare the different schemes, their drawbacks, their benefits, their applicability and their complexity while taking into account the impact of channel quality variations including fast variations but also slow variations due to the mobility.

I.8 Thesis contributions

The contributions of this thesis are the following:

Chapter II:

We show in this chapter that the performance of a scheduling policy depends primarily on the mobility of users as well as on the type of service, elastic or streaming. Indeed, slow channel variations due to mobility can be exploited by adequate scheduling strategies.

First, we show that slow channel variations due to mobility can be exploited in the presence of **elastic traffic**. We assess the performance gains of mobility on the downlink of cellular data networks. The gains are only due to the elastic nature of traffic and thus observed even under a blind, fair scheduling scheme: data are more likely transmitted when users are close to the base stations, in good radio conditions. This phenomenon is further amplified by opportunistic scheduling schemes that exploit multi-user diversity. Specifically, since mobile users in poor radio conditions are likely to move and to be served in better radio conditions, we propose a **mobility-aware scheduler** that deprioritizes those users. We compare the performance of this scheduler to that of other usual scheduling schemes in a dynamic setting with a random number of active users and various scenarios of mobility. While the proportional fair scheduler is considered as the best algorithm in the absence of mobility, the system performance improves under more opportunistic schedulers like max C/I in the presence of mobility. It turns out that the proposed mobility-aware scheduler outperforms these two scheduling policies by adapting its behavior to the observed mobility of active users. We show that this scheduler is a good compromise in the presence of multiple mobility behaviors and improves the overall performance, especially at high load. The results are based on the analysis of flow-level traffic models based on networks of coupled queues with routing, and validated by system-level simulations. These results have been presented at Wiopt 2015 [C1]. A journal paper has been also submitted for possible publication in the IEEE Transactions on Wirelesss Communications [J1].

Second, we evaluate the performance under several scheduling strategies in the presence of **adaptive streaming traffic**, through the analysis of flow level traffic models. We consider two performance indicators, mean video bit rate and mean buffer surplus, in various scenarios of mobility and under several scheduling policies. We show that in the static scenario, the max C/I policy pushes users to select higher video bit rate which may degrade the stability condition. In the presence of mobile users, there are performance trade-offs between the video bit rate and the buffer surplus, based on the targeted metric. In order to obtain an intermediate results, discriminatory scheduling is also investigated and suggested in this case. These results have been obtained in collaboration with Yu-Ting Lin and have presented at PIMRC 2016 [C4].

Chapter III:

In this chapter, we evaluate the performance of the JP scheme. The evaluation is based on the analysis of flow level traffic models as well as system-level simulations. We consider mainly two scenarios: several cells with one common coordination zone (overlapping area), and a static cluster constituted of three cooperating cells, which may represents the case of intra-site coordination for instance. We derive analytical expressions for the stability condition and the throughput performance in light traffic, under several scheduling strategies.

We figure out the scenarios where it is interesting to deploy JP. We study both high interference

scenario and low interference scenario with different clustering methods and different transmission schemes (one-multi stream). We analyze the capacity issues of this scheme, and provide a complete comparative study between JT and DPB. Indeed, the performance depends primarily on the mean coordination gain, that is the mean throughput gain brought by the cooperation of a cell to a cell-edge user. This gain depends primarily on the interference level in the network.

It turns out that in an **interference-dominated environment**, where the mean coordination gain is high enough in order to compensate the extra resource consumption, JT could be interesting, particularly at medium load, especially when different streams are sent from different cooperating cells, but at the cost of higher complexity (multiplexing and advanced receivers techniques). In addition, in order to maintain stability condition in this case, we propose a **new coordination scheme** where a cell cooperates only when it brings at least 100% mean throughput gain when cooperating, in a symmetric network topology (a patent application has been filed [P1]).

We evaluate the performance of several scheduling strategies and show that the scheduling strategy strongly impact the performance in the presence of coordination mechanisms. A global scheduler that maximizes the instantaneous utility function, where the coordination can be activated and deactivated dynamically according to the best scheduling decision, provides best performance. However, the complexity of such a scheduler increases significantly with the number of users and the size of the cluster.

The performance of DPB is also evaluated in several cases under different clustering methods assuming DPB scheduling granularity of one sub-frame. We show that most of the mean rate gain of a cell-edge user is achieved through the elimination of the interference (adding a new useful signal provides only limited additional rate gain) and that DPB may present less restrictive stability condition compared to a JT scheme with moderate coordination gain. The relative gain of DPB, compared to the case without coordination, increases with load. This is due to the fact that the gain brought by the elimination of the interference is relatively small at low load since slightly loaded neighboring cells are naturally muted most of the time and thus collaboration is likely to be less triggered. When the network becomes more loaded, interference between different cells becomes more important. In this case, the blanking of interfering cells at the right moment when the scheduled user is strongly affected by the generated interference seems to bring interesting gains. We show that **DPB is more promising than JT** since it is a simpler and performant technique especially at higher loads, where only a limited number of antennas is deployed. Even with only intra-site coordination, the gains are important for the reason that all users, even those who are not defined as CoMP users, can benefit implicitly from the muting of some interfering cells. The scheduling granularity is very important in this case and allows users to be scheduled at the right moment where interfering cells go blank. While intra-site coordination is performant and much easier to implement, dynamic coordination is needed in order to benefit from the full interference reduction.

Synchronization presents a real challenge for both JP techniques, to this must of course be added the CSI feedback design as well as the backhaul characteristics in terms of latency and capacity especially when considering a distributed coordination approach.

In a low interference scenario on the other hand, we show that it is not recommended to activate JP feature. However, if we stick to ensure uniformity of service by performing coordination, more precisely JT, for static cell-edge users suffering from very degraded throughput, it is not worth to perform coordination for a mobile user, who is able to move and to get better radio conditions. Consequently, we propose a **mobility-aware scheduler**, which is the extension of the scheduler considered in Chapter II and which deprioritizes mobile CoMP users. We show that this scheduler improves the performance by giving the chance to mobile cell-edge users to be served in better radio conditions where cell coordination is not required, which limits the extra radio resource consumption for those mobile cell-edge users. This scheduler is suitable for elastic traffic where the delay is tolerable. Part

of these results have been published and presented at GLOBECOM 2015 [C2] and PIMRC 2016 [C3].

Chapter IV:

In this chapter, we study a beamforming system, using uniform linear antenna array.

First, we consider a system with **blind beamforming**, that is a system where beams activation is done in a distributed and non-coordinated manner in different cells. We quantify for different numbers of antenna elements, the maximum expected capacity gains with respect to a system without beamforming, under a round robin scheduler. We evaluate the performance, in terms of throughput, under several scheduling strategies and we show that it is more likely to activate beams that do not interfere each other under channel-aware strategies, where users are scheduled based on their channel conditions.

Then, we considered a system with **coordinated beamforming** under different clustering strategies, in order to quantify the gains brought by more sophisticated coordinated schedulers. We show that coordinated scheduling/beamforming add high complexity to the network, in terms of feedback, signaling, backhaul and processing, and brings only limited further gains compared to blind beamforming, operated under standard non-coordinated (distributed) channel-aware schedulers.

We show that using channel-aware non-coordinated schedulers, and more specifically max C/I strategy, in a beamforming system where a large number of antenna elements is deployed, achieves similar performance as coordinated schedulers, without any additional feedback requirements and without adding any complexity to the system. In cases where only a limited number of antennas is deployed at the BS, coordinated schedulers may be needed in order to avoid interference between activated beams in neighboring cells, but this is at the detriment of higher processing overhead and a larger amount of feedback, which place additional burden on the network. Thus, there is a tradeoff to be considered between the number of deployed antenna elements and the complexity of the scheduler, the feedback and signaling requirements.

The results of this thesis are obtained through the analysis of flow level traffic models as well as system level simulations. The performance evaluation done from a system level point of view allows to have realistic information on the real practical gains of different CoMP schemes, scheduling strategies, as well as beamforming technique and thus helps on deployment decisions.

Publications

[C1] Nivine Abbas and Thomas Bonald and Berna Sayraç. Opportunistic gains of mobility in cellular data networks. *WiOpt 2015, Mumbai, India, May 25-29, 2015*.

[C2] Nivine Abbas and Thomas Bonald and Berna Sayraç. How Mobility Impacts the Performance of Inter-Cell Coordination in Cellular Data Networks. *GLOBECOM 2015, San Diego, CA, USA, December 6-10, 2015*.

[C3] Nivine Abbas and Thomas Bonald and Berna Sayraç. Mobility-aware Scheduler in CoMP Systems. *pimrc 2016, Valencia, Spain, September 4-7, 2016*.

[C4] Nivine Abbas and Yu-Ting Lin and Berna Sayraç. Mobility-driven Scheduler for Mobile Networks Carrying Adaptive Streaming Traffic. *pimrc 2016, Valencia, Spain, September 4-7, 2016*

[C5] Dora Boviz, Nivine Abbas, Gopalasingham Aravinthan, Chung Shue Chen, Mohammed Amine Dridi. Multi-cell Coordination in Cloud RAN: Architecture and Optimization. *WINCOM 2016, Fez, Morocco, October 26-29, 2016*

[J1] Nivine Abbas and Thomas Bonald and Berna Sayraç. Mobility-Aware Scheduling in Cellular Data Networks, *submitted to IEEE Transactions on Wireless Communications*.

[P1] Nivine Abbas and Berna Sayraç. Procédé de transmission CoMP avec contrainte d'augmentation de gain, produits programme et dispositifs correspondants, *patent application no. 1658643, filed on 15 september 2016*.

Chapter II

Network without coordination

II.1 Introduction

LTE (Long Term Evolution) and next generation wireless networks are designed to provide high speed mobile multimedia services. Voice calls and SMSs are no longer the dominant services, especially for smartphone users, who are increasingly making daily use of data-based services. In order to meet this increasing demand of data transmissions, various techniques have been introduced throughout the previous generations to allocate the scarce radio resources on the downlink of cellular data networks and to improve network efficiency and capacity.

The efficiency of wireless data transmissions strongly depends on the variations of the radio channel quality, especially the *fast* variations caused by the constructive and destructive combinations of the multiple signal paths between the transmitter and the receiver. While traditional wireless systems use diversity techniques to mitigate these channel variations and thus provide constant bit rates, advanced wireless systems exploit these variations through so-called opportunistic schedulers, taking advantage of the inherent "elasticity" of data transfers to increase the overall system capacity [71, 15, 19].

Slow channel variations, due to the mobility of the device, can be exploited as well. In fact, they are already exploited by current packet schedulers, even blind strategies like the round-robin policy, where the capacity is naturally increased. The reason is that, due to mobility, data are more likely transmitted when users are close to the base stations, in good radio conditions. This phenomenon is further amplified by opportunistic schedulers, which schedule data transmission to the users when their channel conditions are relatively favorable. The throughput is improved even at the cell edge. The reason is that users with stronger radio channels quickly complete their file transfers, which saves scheduling resources for cell-edge users.

The gains due to the mobility-induced rate variations have been thoroughly explored in the literature in the presence of elastic traffic, see for instance [22, 21, 27, 26, 39, 45, 11]. It has been shown that in a dynamic setting, where users come and go over time, the performance at flow level improves as the mobility of the users increases. It is worth noting that the actual transmission rates may be reduced at higher speeds due to the estimation and prediction problems that may occur, especially in high frequency bands [62, 12, 14]. We do not take into account such issues in the following, like most papers on this topic. We assess in this work the intra-cell and inter-cell mobility-induced gain under a fair scheduling scheme (the round-robin policy) and an opportunistic scheduling scheme (the max C/I policy, see [47] and the proportional fair policy, see [41]) for both the cell center and the

cell edge. We show in particular that under max C/I scheduling policy, mobility improves throughput performance at the cell edge. This is a rather surprising result for the max C/I policy, which gives priority to cell-center users. The reason is that these users quickly complete their file transfers, which saves scheduling resources for cell-edge users.

The scheduling strategy has an important impact on the performance of wireless systems. Under the family of α-fair scheduling policies, the optimal choice of the fairness factor α depends critically on mobility. It has been shown in [26] that in the presence of mobility, the more opportunistic the scheduler is (i.e., $\alpha \to 0$), the higher is the system capacity. In contrast, proportional fairness (i.e., $\alpha = 1$) is usually considered as the best strategy in the absence of mobility. In this chapter, we propose a mobility-aware scheduler that exploits the user mobility as an additional information. Specifically, mobile users at the edge are deprioritized since they are likely to move and to be served in better radio conditions. Moreover, the scheduler adapts to the mobility of users and tends to be more opportunistic (i.e., $\alpha \to 0$) when all users are mobile and more fair (i.e., $\alpha \to 1$) when all users are static. We compare the performance of the proposed scheduler to that of other usual scheduling policies (round-robin, max C/I, proportional fair) through the analysis of flow level traffic models and we validate the results by system level simulations. We show that this scheduler is a good compromise in the presence of multiple mobility behaviors and improves the overall performance, especially at high load.

Slow channel variations due to mobility can be further exploited by opportunistic schedulers in the presence of elastic traffic. However, as video streaming service like YouTube and Netflix account for larger part of system traffic, there is a need to verify whether this conclusion is still valid for streaming services.

Regarding to streaming services, [24] and [25] are two early work that apply flow-level model to study the performance of real-time adaptive streaming. Then, authors of [72] focused on the impacts of the scheduler on the performance of mobile networks carrying HTTP streaming traffic considering constant video bit rate only. It is shown that opportunistic scheduling can provide better throughput with some trade-offs of freezing delay. As more video distribution platforms adopt adaptive streaming technology like MPEG-DASH [1][67] to deliver video content in wireless system, the quality of experience (QoE) of adaptive streaming needs to be examined. In [23], authors have investigated the QoE of static adaptive streaming using flow-level dynamics. However, the impacts of users' mobility and scheduling strategies have not been considered.

In this chapter we analyze the QoE of adaptive streaming, under two fair scheduling schemes (the round-robin policy and the max-min policy see [19]) and an opportunistic scheduling scheme (the max C/I policy, see [47]) in the absence of mobility. We measure the QoE based on two performance metrics: the mean video bit rate and the mean buffer surplus. We study the performance at flow level and we show that opportunistic schedulers may degrade the stability condition in this case. The reason is that users in good radio conditions choose large video bit rate and increase the burden on the network while users in bad radio conditions are completely penalized. We then extend the analysis to the case of mobile users, where we show that there is a trade-off between the video bit rate and the buffer surplus: while the max C/I strategy enables to achieve the best mean video bit rate, it may degrade the buffer surplus at medium load compared to the round-robin strategy. However, in contrast to the scenario without mobility, the max C/I strategy provides the best stability condition in the presence of mobility. Thus the scheduling strategy should be chosen based on the targeted metric. We also give a brief summary on the proper scheduling policy to adopt in each scenario (without and with mobility). The results are based on the analysis of flow-level traffic model.

We show in this chapter that the performance of a scheduling policy depends primarily on the mobility of users as well as the service type: elastic, streaming. The chapter is divided into two main parts.

The first part II.2 focuses on the case of elastic traffic. We describe the proposed mobility-aware

scheduler in Section II.2.1 then we present in Section II.2.2 the flow level models. In Section II.2.3, we show the numerical results obtained though the evaluation of the Markov process and we validate these results by system level simulations.

The second part II.3 concerns the case of adaptive streaming. We present the flow level models and we define the key performance metrics, then we analyze the impacts of intra-cell mobility under different scheduling strategies, in Section II.3.1. The numerical results are then presented and analyzed in Section II.3.2. A summarized suggestion of scheduling policy for different scenarios is proposed in Section II.3.3.

II.2 Elastic traffic

II.2.1 Mobility-aware scheduler

The scheduling strategy is a key component of cellular systems. It is the process through which a base station decides which user should be allocated resources to receive data. The simplest strategy consists in serving users in round robin regardless of their channel conditions. The simplicity of this strategy is at the expense of degraded performance.

In the downlink, LTE uses OFDMA where multiple access is achieved by assigning subsets of subcarriers to individual users, see Section I.3. Scheduling is done per sub-frame level (each 1 ms) and for each subset of resource blocks (subcarriers), which allows to exploit multiuser diversity and instantaneous channel fluctuations (fast fading) through opportunistic scheduling. For the sake of simplicity, we focus on one subset of subcarriers and present the different assignment of the considered subset under different scheduling strategies.

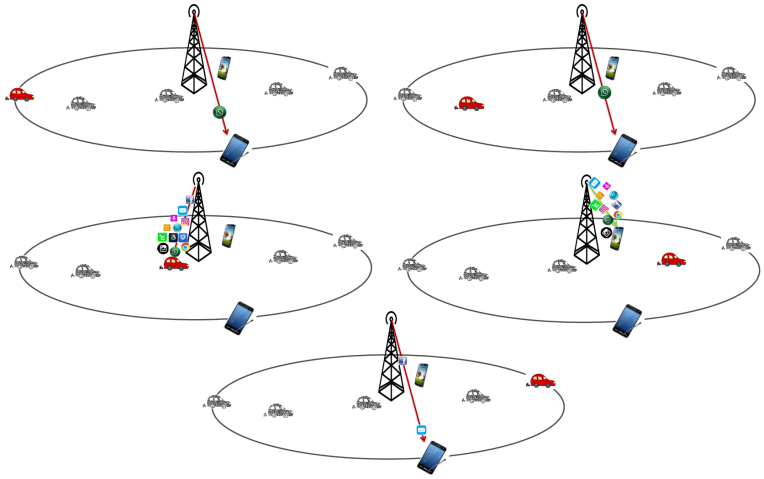


Figure II.1: Mobility-aware scheduling

The max C/I scheduler [47] selects the user with the highest instantaneous data rate

$$u(t) = \arg\max_i r_i(t),$$

where $u(t)$ is the scheduled user at time t and $r_i(t)$ is the instantaneous transmission rate of user i at time t. It is the most efficient in terms of throughput. However, it may neglect users in bad radio conditions, who are never served at high load. This results in degraded performance, unless users

are mobile. We shall see in the following sections that under this strategy, the mobility is very well exploited and mobile users are more likely to be served in good radio conditions [11].

The proportional fair scheduler [41, 30, 60] selects the user with the highest instantaneous data rate relative to its mean data rate:

$$u(t) = \arg\max_i \frac{r_i(t)}{\bar{R}_i(t)}.$$

It is a good trade-off between efficiency and fairness in the absence of mobility. Note that $\bar{R}_i(t)$ is typically evaluated through an exponentially smoothed average:

$$\bar{R}_i(t) = \left(1 - \frac{1}{t_s}\right)\bar{R}_i(t-1) + \frac{1}{t_s} r_i(t-1)\mathbb{1}_{\{u(t-1)=i\}},$$

where the parameter t_s captures the time-scale of the scheduler: large values of t_s make the user wait a long time before being scheduled when its channel quality hits a peak. However we shall see that the optimal value of t_s should be adaptive and of the order of the mean sojourn time of the user; thus we replace t_s by the current sojourn time of user i at time t.

Our mobility-aware scheduler adapts to the mobility of the users. The idea of this strategy is to deprioritize mobile users in poor radio conditions as they are more likely to move and to be served in good conditions. Figure II.1 shows a simple case where the user in the car is deprioritized when at the edge, in poor radio conditions, and served when close to the base station. To do that, we use two different scheduling metrics according to the mobility of each user. Thus a classification of users as mobile or static is needed (for instance, users with a speed less than 4 km/h are considered as static, the others are considered as mobile). Note that the estimation of the speed of each user does not need to be very accurate. Mobility can be easily detected by a simple estimation of the variation of the mean radio conditions, that is the variation of the path loss (with the shadowing), which depends primarily on the distance of the user from the base station. An estimation of the mobile speed as done in [29, 61, 35, 34] can also be used to classify the users in different mobility classes. We denote by $\mathbb{S}$ the class of static users and by $\mathbb{M}$ the class of mobile users. The scheduled user at time t is the user with the highest metric $\eta_i(t)$:

$$u(t) = \arg\max_i \eta_i(t),$$

where the metric of static users is the classical proportional fair metric

$$\eta_i(t) = \frac{r_i(t)}{\bar{R}_i(t)} \qquad \forall i \in \mathbb{S},$$

while the metric of mobile users is evaluated as follows :

$$\eta_i(t) = \frac{r_i(t)}{\bar{R}_{\text{mob}}(t)} \frac{r_i(t)}{\bar{R}(t)} \qquad \forall i \in \mathbb{M},$$

where

$$\bar{R}_{\text{mob}}(t) = \bar{R}_j(t)_{\{j=\arg\max_{k\in\mathbb{M}} r_k(t-1)\}}$$

is the average throughput perceived by the mobile user experiencing the best radio conditions (SINR) at time $t-1$ among all mobile users. Observe that the term $r_i(t)/\bar{R}_{\text{mob}}(t)$ allows to perform the scheduling inside the mobile class according to the classical max C/I metric. Each mobile user is also assigned a weight $r_i(t)/\bar{R}(t)$, where $\bar{R}(t)$ is the actual mean experienced rate in the network which depends primarily on the radio conditions in the network. This actual rate is evaluated through an exponentially smoothed average of the mean actual experienced rate of active users:

$$\bar{R}(t) = \left(1 - \frac{1}{t_a}\right)\bar{R}(t-1) + \frac{1}{t_a}\frac{\sum_i \bar{r}_i(t)}{n(t)},$$

where $n(t)$ is the total number of active users at time t and t_a represents the time-scale over which the average is performed. This time parameter should be sufficiently large (e.g., of the order of seconds) to reflect the average radio conditions of the cell. The mean perceived rate of user i is also evaluated through an exponentially smoothed average, only on those slots where the user is scheduled, that is:

$$\bar{r}_i(t) = \bar{r}_i(t-1) + \frac{1}{t_a}\left(r_i(t) - \bar{r}_i(t-1)\right)\mathbb{1}_{\{u(t-1)=i\}}.$$

Observe that the weight $r_i(t)/\bar{R}(t)$ assigned to mobile users allows to deprioritize mobile users experiencing radio conditions worse than the mean radio conditions of the network. In LTE wireless networks, at each timeslot (1 ms) the algorithm works as follows, where t_i denotes to the arrival time of user i:

Data: CQI (channel quality indicator) and mobility class of each user
Result: User to schedule
// Scheduling
MAX $\leftarrow 0$
for each user i **do**
  if user i is static **then**
    $\eta \leftarrow \frac{r_i}{\bar{R}_i}$
  else
    $\eta \leftarrow \frac{r_i}{\bar{R}_{\text{mob}}} \frac{r_i}{\bar{R}}$
  end
  if MAX $< \eta$ **then**
    MAX $\leftarrow \eta$
    $u \leftarrow i$
  end
end

// Updates
$\bar{R} \leftarrow \left(1 - \frac{1}{t_a}\right)\bar{R}$
for each user i **do**
  $\bar{R}_i \leftarrow \left(1 - \frac{1}{t-t_i}\right)\bar{R}_i + \frac{1}{t-t_i} r_i \times (u == i)$

  $\bar{r}_i \leftarrow \bar{r}_i + \frac{1}{t_a}(r_i - \bar{r}_i) \times (u == i)$

  $\bar{R} \leftarrow \bar{R} + \frac{1}{t_a}\frac{\bar{r}_i}{n}$
end
if u is mobile **then**
  $\bar{R}_{\text{mob}} \leftarrow \bar{R}_u$
end

Observe that the scheduler behaves like max C/I when all users are mobile and like the proportional fair scheduler when all users are static. We shall see that this strategy is a good compromise between these two strategies in the presence of mobility behaviors.

II.2.2 Flow level model

II.2.2.1 Model without mobility

We present in this section the reference model in the absence of mobility and fast fading; the corresponding performance results are derived under round-robin and max C/I scheduling schemes.

a) Cellular network

We model each cell by a set of N regions. In each region, radio conditions are supposed to be homogeneous and thus users are served at the same physical data rate on the downlink. In the simple case of two regions illustrated by Figure II.2, users may be close to the cell center and experience good radio conditions (light gray) or close to the cell edge and suffer from bad radio conditions (dark gray). We model each region by a queue with a specific service rate corresponding to the physical data rate in this region; since all users in the cell share the same radio resources, each cell can be viewed as a set of N parallel queues with coupled processors. The precise coupling depends on the scheduling policy, as explained below.

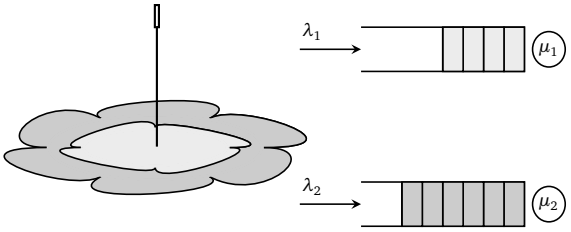


Figure II.2: A simple model with two cell regions.

b) Traffic model

We consider elastic traffic only, corresponding to data transfers. Specifically, we assume that new data flows are generated in region i at the random times of a Poisson process of intensity λ_i. We denote by $\lambda = \sum_i \lambda_i$ the total flow arrival rate in the cell and by $p_i = \lambda_i/\lambda$ the probability that a new data flow is generated by a user in region i.

For the sake of tractability, we use a flow-level model where each data flow is viewed as a fluid of random volume to be transmitted. The volumes have an exponential distribution with mean σ (in bits). When region i is served, flows are completed at rate μ_i in the absence of fast fading, corresponding to the physical rate $R_i = \mu_i\sigma$ (in bit/s). We assume regions are numbered in decreasing order of physical rates, that is $R_1 > R_2 > \ldots > R_N$.

Queue i represents the number of active flows in region i. The load of queue i is the ratio $\rho_i = \lambda_i/\mu_i$ of arrival rate to service rate. Since the radio resources are shared in time, the actual service rate of queue i is modulated by ϕ_i, the fraction of time spent by the scheduler on users in region i. This depends on the system state and the scheduling policy. For work-conserving policies, we have $\sum_i \phi_i = 1$. The overall stability condition is then

$$\rho < 1 \tag{II.1}$$

in the absence of mobility, with

$$\rho = \sum_i \rho_i. \tag{II.2}$$

We shall see §II.2.2.2 that mobility may increase the stability region. Observe that the cell load may be written $\rho = \lambda/\mu$ where μ is the harmonic mean service rate weighted by the distribution of arrivals in the N regions:

$$\mu = \frac{1}{\sum_i p_i/\mu_i}.$$

We denote by R the harmonic mean physical rate corresponding to μ:

$$R = \frac{1}{\sum_i p_i/R_i}. \tag{II.3}$$

c) Throughput metrics

We measure performance in terms of mean throughputs in the different regions. We denote by $X(t)$ the system state at time t. This is an N-dimensional vector whose component i gives the length of queue i at time t. It is an irreducible Markov process with stationary distribution π. In any state x such that $x_i > 0$, each user in region i has throughput $R_i\phi_i(x)/x_i$. Now the distribution seen by users in region i is the size-biased distribution [20]:

$$\pi_i(x) \propto x_i\pi(x).$$

We denote by E_i the corresponding expectation. The mean throughput of users in region i is then given by

$$\gamma_i = E_i\left(\frac{R_i\phi_i(X)}{X_i}\right) = \frac{E(R_i\phi_i(X))}{E(X_i)}. \tag{II.4}$$

By the traffic conservation equation

$$\lambda_i = E(\mu_i\phi_i(X)), \tag{II.5}$$

we deduce

$$\gamma_i = \frac{\lambda_i\sigma}{E(X_i)}. \tag{II.6}$$

Observe that, by Little's law, this is the ratio of mean flow size (in bits) to mean flow duration in region i.

By a similar argument, the mean throughput in the cell is

$$\gamma = \frac{E(\sum_i R_i\phi_i(X))}{E(\sum_i X_i)}. \tag{II.7}$$

By the traffic conservation equation

$$\lambda = \sum_i E(\mu_i\phi_i(X)), \tag{II.8}$$

we deduce

$$\gamma = \frac{\lambda\sigma}{\sum_i E(X_i)}. \tag{II.9}$$

We shall see that, in the absence of mobility, the mean throughput in region i decreases from the physical data rate R_i to 0 when the cell load ρ grows from 0 to 1, while the mean throughput in the cell decreases from the harmonic mean physical data rate R to 0, the precise impact of the cell load depending on the scheduling policy.

The quality of experience is also strongly affected by the throughput variance, a quantity that is often neglected in performance studies. Using the same expectation, the throughput variance of users in region i is given by :

$$v_i = E\left(\left(\frac{\phi_i(X)R_i}{X_i}\right)^2 X_i\right)/E(X_i) - \gamma_i^2. \tag{II.10}$$

Similarly, the cell throughput variance is given by:

$$v = E\left(\left(\frac{\sum_i \phi_i(X)R_i}{|X|}\right)^2 |X|\right)/E(X) - \gamma^2.$$

d) Round-robin policy

Under the round-robin policy, users share the radio resources equally, independently of their radio conditions. Thus users in region i are allocated a fraction

$$\phi_i(x) = \frac{x_i}{\sum_j x_j} \tag{II.11}$$

of radio resources in state x. This corresponds to the case $\alpha = 1$ in the absence of fast fading. The transition graph of the Markov process $X(t)$ is shown in Figure II.3 (omitting the blue arrows, corresponding to mobility). This Markov process is reversible and corresponds to the state of a multi-class processor-sharing queue. The stationary distribution of $X(t)$ is:

$$\pi(x) = (1-\rho)\binom{\sum_i x_i}{x_1, \ldots, x_N}\rho_1^{x_1} \ldots \rho_N^{x_N}. \tag{II.12}$$

We deduce the mean number of flows in region i:

$$E(X_i) = \frac{\rho_i}{1-\rho},$$

and, in view of (II.6), the mean throughput in region i:

$$\gamma_i = R_i(1-\rho). \tag{II.13}$$

Observe that the mean throughputs decrease *linearly* with the cell load in all regions.

In view of (II.9), the mean throughput in the cell is

$$\gamma = R(1-\rho). \tag{II.14}$$

It also decreases linearly with the cell load.

Using (II.10), (II.12) and (II.13) we compute the throughput variance of users in region i under the round-robin strategy:

$$v_i = \sum_{x_i>0}\sum_{x_j\geq 0, j\neq i}\left(\frac{R_i}{x}\right)^2 x_i \pi(x)/E(X_i) - {\gamma_i}^2$$

$$= \frac{R_i^2(1-\rho)^2}{\rho_i}\sum_{x_i>0}\sum_{x_j\geq 0, j\neq i}\frac{x_i}{x^2}\frac{x!}{x_1!x_2!\ldots x_N!}\rho_1^{x_1}\rho_2^{x_2}\ldots\rho_N^{x_N} - R_i^2(1-\rho)^2$$

$$= \frac{R_i^2(1-\rho)^2}{\rho} \sum_{x\geq 0} \frac{\rho^{x+1}}{x+1} - R_i^2(1-\rho)^2$$
$$= -R_i^2(1-\rho)^2 \left(\frac{\log(1-\rho)}{\rho} + 1 \right).$$

We deduce the throughput standard deviation in region i,

$$R_i(1-\rho)\sqrt{\frac{-\log(1-\rho)}{\rho} - 1}. \tag{II.15}$$

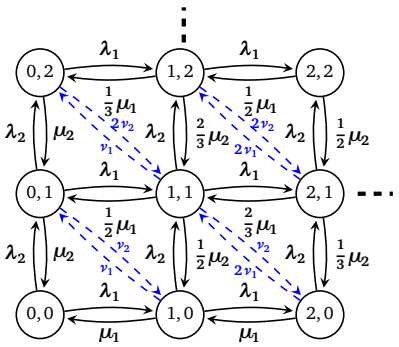


Figure II.3: State diagram for two regions under the round-robin policy.

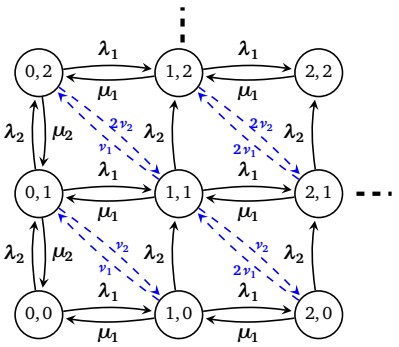


Figure II.4: State diagram for two regions under the max C/I policy.

e) Max C/I policy

The max C/I policy is an opportunistic scheduling strategy that prioritizes those users with the best radio conditions. For two regions for instance, cell-center users are scheduled first and are allocated all the resources whenever active; cell-edge users are served only when there are no active cell-center users. The transition graph of the Markov process $X(t)$ is shown in Figure II.4 (omitting the blue arrows). The Markov process is no longer reversible and corresponds to a preemptive priority queue. Applying known results of queuing theory [46], we get the mean number of flows in region i:

$$E(X_i) = \frac{\rho_i}{1-\sum_{j<i}\rho_j} \left(1 + \frac{\rho}{1-\sum_{j<i}\rho_j} \mu_i \mu \sum_{j<i} \frac{p_j}{\mu_j^2} \right)$$

and, in view of (II.6), the mean throughput in region i:

$$\gamma_i = R_i \frac{1-\sum_{j<i}\rho_j}{1+\frac{\rho}{1-\sum_{j\leq i}\rho_j}\mu_i\mu\sum_{j<i}\frac{p_j}{\mu_j^2}}.$$

Observe that the mean throughput in region i is positive whenever $\sum_{j\leq i}\rho_j < 1$: it decreases from R_i to 0 when the load generated by users in regions $1,\ldots,i$ grows from 0 to 1.

Similarly, we get the mean throughput in the cell

$$\gamma = \mu\sigma \left(\sum_i \frac{\rho_i/\rho}{1-\sum_{j<i}\rho_j} \left(1 + \frac{\rho}{1-\sum_{j\leq i}\rho_j} \mu_i\mu \sum_{j<i} \frac{p_j}{\mu_j^2} \right) \right)^{-1}$$

which decreases from R to 0 when the cell load ρ grows from 0 to 1.

The standard deviation of the throughput of users in region $i = 1$ is given by

$$R_1(1-\rho_1)\sqrt{\left(\frac{-\log(1-\rho_1)}{\rho_1}-1\right)}. \tag{II.16}$$

Note that there is no explicit expressions for $i > 1$.

II.2.2.2 Impact of intra-cell mobility

In this section, we add intra-cell mobility to the previous model, as shown in Figure II.5 for two regions. We suppose that users move from the center of the cell to the edge and vice versa. We still assume that there is no fast fading.

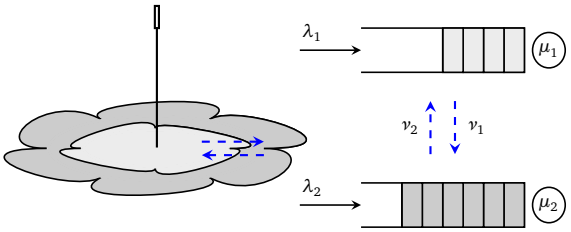


Figure II.5: Queuing model for two regions with intra-cell mobility.

a) Mobility model

We assume that each user in region i moves to region $i-1$ (for $i > 1$) and to region $i+1$ (for $i < N$) after exponential durations, at respective rates $\nu_{i,i+1}$ and $\nu_{i,i-1}$. The probability that a user is in region i then satisfies:

$$q_i \propto \prod_{j=1}^{i-1} \frac{\nu_{j,j+1}}{\nu_{j+1,j}}.$$

Note that this is not the probability that an *active* user is in region i, which is given by

$$p_i' = \frac{E(X_i)}{\sum_j E(X_j)}$$

and depends both on the cell load and on the scheduling policy. In state x, the total mobility rate from region i to region $i+1$ and from region $i+1$ to region i (for $i < N$) are $x_i\ \nu_{i,i+1}$ and $x_{i+1}\ \nu_{i+1,i}$, respectively. For notational convenience, we write $\nu_1 \equiv \nu_{1,2}$ and $\nu_2 \equiv \nu_{2,1}$ for two regions. The corresponding state diagrams are given by Figures II.3 and II.4, where the blue arrows correspond to the transitions due to mobility.

It is worth noting that, for mobile users, the traffic conservation equation (II.5) no longer applies: the traffic arriving in region i is not equal in general to the traffic served in region i. Thus the mean throughput in region i is given by (II.4) but not by (II.6). Now the overall traffic conservation equation (II.8) still applies (in the absence of inter-cell mobility) so that the mean throughput in the cell is given by (II.9). It follows from (II.4) and (II.7) that

$$\gamma = \sum_i p_i' \gamma_i.$$

b) Round-robin policy

Under the round-robin policy, there is no longer an explicit expression for the stationary distribution of the Markov process $X(t)$, except in the limiting regime of infinite mobility where $\nu_{i,i+1}, \nu_{i+1,i} \to \infty$ for all $i < N$. In this case, the N queues are equivalent to one processor-sharing queue with service rate:

$$\bar{\mu} = \sum_i q_i \mu_i.$$

and physical rate

$$\bar{R} = \sum_i q_i R_i.$$

We denote by $\bar{\rho} = \lambda/\bar{\mu}$ the corresponding load. The stationary distribution of $X(t)$ becomes:

$$\pi(x) = (1-\bar{\rho})\bar{\rho}^n \binom{n}{x_1, \ldots, x_N} q_1^{x_1} \ldots q_N^{x_N},$$

with $n = \sum_i x_i$, under the stability condition $\bar{\rho} < 1$. The flow throughput is independent of the region where the flow starts and is given by:

$$\bar{\gamma} = \bar{R}\sigma\,(1-\bar{\rho})\,.$$

The stability condition, which can be written

$$\rho < \frac{\bar{\mu}}{\mu}, \tag{II.17}$$

is actually independent of the mobility rates: at maximum load, flows tend to be infinitely long so that users move in the whole cell during their file transfers and have the mean physical rate $\bar{\mu}\sigma$. In the particular case where $p_i = q_i$ for all i (that is, the flow arrivals are distributed according to the distribution of the user positions), then $\bar{\mu} \geq \mu$ (the weighted arithmetic mean is larger than the corresponding weighted harmonic mean) and the maximum load increases, as observed in [22].

c) Max C/I policy

Under the max C/I policy, it is also possible to derive the stationary distribution of the Markov process $X(t)$ in the limiting regime of infinite mobility. Given n active users, the probability that the user(s) with the best radio conditions is (are) in region i is given by

$$q_i(n) = \left(\sum_{j \geq i} q_j\right)^n - \left(\sum_{j > i} q_j\right)^n.$$

Since the max C/I policy only serves this (these) user(s), the mean service rate becomes:

$$\bar{\mu}(n) = \sum_i q_i(n)\mu_i.$$

Observe that $\bar{\mu}(n)$ is increasing, with

$$\bar{\mu}(1) = \bar{\mu} \quad \text{and} \quad \lim_{n\to\infty} \bar{\mu}(n) = \mu_1.$$

The stationary distribution of $X(t)$ is:

$$\pi(x) = \pi(0) \prod_{k=1}^{n} \frac{\lambda}{\bar{\mu}(k)} \times \binom{n}{x_1, \ldots, x_N} q_1^{x_1} \ldots q_N^{x_N},$$

with $n = \sum_i x_i$, under the stability condition $\lambda < \mu_1$.

This stability condition is in fact independent of the mobility rates. The maximum load is:

$$\rho < \frac{\mu_1}{\mu}, \tag{II.18}$$

which is always larger than in the absence of mobility.

d) Throughput in light traffic

The performance in light traffic (that is, when $\rho \to 0$) is independent of the scheduling policy. Indeed, a user when alone in the system is always allocated all radio resources. The mean throughput in region i is then equal to $\mu_i \sigma$. We denote by

$$\nu_i = \nu_{i,i+1} + \nu_{i,i-1}$$

the total mobility rate of a user in region i. Thus, the probability that a user in region i moves to region $i+1$ before leaving the system is:

$$\alpha_{i,i+1} = \frac{\nu_{i,i+1}}{\mu_i + \nu_i}.$$

Similarly, the probability that a user in region i moves to region $i-1$ before leaving the system is:

$$\alpha_{i,i-1} = \frac{\nu_{i,i-1}}{\mu_i + \nu_i}.$$

We deduce the mean duration of a flow initiated in region i:

$$T_i = \alpha_{i,i-1} T_{i-1} + \alpha_{i,i+1} T_{i+1} + 1/(\mu_i + \nu_i).$$

Solving these equations, we obtain:

$$T_i = \sum_{j=i}^{N} \left((-1)^{i-j} a_j \prod_{k=i}^{j-1} \frac{-\alpha_{k,k+1}}{b_k} \right), \tag{II.19}$$

where

$$b_i = \frac{1 - \sum_{j=1}^{i-1} \left(1 - \sum_{k=1}^{j-2} \alpha_{k,k+1}\alpha_{k+1,k}\right) \alpha_{j,j+1}\alpha_{j+1,j}}{1 - \sum_{j=1}^{i-2} \left(1 - \sum_{k=1}^{j-2} \alpha_{k,k+1}\alpha_{k+1,k}\right) \alpha_{j,j+1}\alpha_{j+1,j}}$$

and

$$a_i = \frac{\sum_{j=1}^{i} \left(\prod_{k=1}^{j-1} b_k\right) \left(\prod_{k=j+1}^{i} \alpha_{k,k-1}\right) \frac{1}{\mu_j + \nu_j}}{\prod_{j=1}^{i} b_j}.$$

We deduce the mean duration of a flow initiated in the cell:

$$T = \sum_{i=1}^{N} p_i T_i.$$

Then, the cell-average flow throughput in light traffic is given by:

$$\gamma = \sigma / T.$$

For two regions for instance, the cell-average throughput in light traffic is:

$$\gamma = \frac{\nu_2\mu_1 + \nu_1\mu_2 + \mu_1\mu_2}{\nu_1 + \nu_2 + p_1\mu_2 + p_2\mu_1}\sigma. \tag{II.20}$$

Observe that this throughput depends on user's mobility: $\gamma = R$ when $\nu_1, \nu_2 \to 0$ (no mobility) while $\gamma \to \bar{R}$ when $\nu_1, \nu_2 \to +\infty$ (infinite mobility).

e) Variance in light traffic

The throughput variance in light traffic is given by:

$$v = \sum_i p_i' R_i^2 - \left(\sum_i p_i' R_i \right)^2 .$$

When $\nu_i \to 0$, $\forall i \leq N$ (no mobility) the throughput variance in light traffic is mainly due to the variation of radio conditions and is written as:

$$v = \sum_i \frac{p_i/\mu_i}{\sum_j p_j/\mu_j} R_i^2 - \left(\sum_i \frac{p_i/\mu_i}{\sum_j p_j/\mu_j} R_i \right)^2 .$$

However, in the limiting regime of infinite mobility $\nu_i \to \infty$ (for all $i \leq N$) the throughput variance is given by:

$$v = \sum_i q_i R_i^2 - \left(\sum_i q_i R_i \right)^2 .$$

For two regions for instance, explicit expressions of the throughput variance as a function of the mobility rates can be written as:

$$v = \frac{(\nu_2 + p_1\mu_2)(\mu_1\sigma)^2 + (\nu_1 + p_2\mu_1)(\mu_2\sigma)^2}{\nu_1 + \nu_2 + p_1\mu_2 + p_2\mu_1} - \left(\frac{\nu_2\mu_1 + \nu_1\mu_2 + \mu_1\mu_2}{\nu_1 + \nu_2 + p_1\mu_2 + p_2\mu_1} \sigma \right)^2 \tag{II.21}$$

II.2.2.3 Impact of inter-cell mobility

We now take hand-overs into account: users may leave the cell during the file transfer. Similarly, some users may arrive from neighboring cells. We assume that handovers occur only in region N (the cell edge) and that the incoming and outgoing handover rates are equal.

Let ν_h be the outgoing handover rate: each user in region N tends to leave the cell at rate ν_h. To balance incoming and outgoing handovers, we assume that incoming handovers are generated according to a Poisson process of intensity $E(X_N)\nu_h$. Note that the net flow arrival rate in region N is still equal to λ_N. In particular, the traffic conservation equation (II.8) applies and the stability condition remains the same as without handovers for both policies.

These properties suggest that inter-cell mobility has a limited impact on throughput performance, in the presence of intra-cell mobility. This is confirmed by the numerical results.

II.2.2.4 Impact of fast fading

We have so far ignored the impact of fast fading. Opportunistic schedulers like max C/I and proportional fair (PF) improve network efficiency by exploiting multi-user diversity. Consider n active users in region i, without any active users in regions $1, \ldots, i-1$ so that only these n users are scheduled. Assuming Rayleigh fading, users' SINR $S_1, S_2 \ldots S_n$ are exponentially distributed. The resulting SINR gain is given by:

$$F(n) = E\left[\max(S_1, S_2, \ldots, S_n)\right] = 1 + \frac{1}{2} + \ldots + \frac{1}{n}.$$

The throughput gain is typically lower due to the concavity of the data rate with respect to SINR. Assuming that the physical data rate is proportional to $\log(1+S)$ for SINR S, we deduce the rate gain of opportunistic scheduling:

$$G(n) = \frac{E[\ \log_2(1 + \max(S_1, S_2, \ldots, S_n))\ ]}{E[\ \log_2(1+S_1)\ ]}. \tag{II.22}$$

In state x, with $x_i > 0$, the service rate of queue i under the max C/I and the PF policies becomes: $\mu_i \phi_i(x) G(x_i)$. Note that under the PF policy, $\phi_i(x)$ are given by (II.11).

We can see that in the absence of mobility all regions benefits from the opportunistic gain under the PF strategy, as they are served in round robin. However the max C/I strategy limit the service to the best region at high load and neglect other users which reduces the capacity of the system.

II.2.2.5 Impact of flow size distribution

Now, we assume that each user belongs to a class of service l. We consider L classes where class l has a mean flow size σ_l. The mean service rate of users of class l in region i is $\mu_i^{(l)}$. However, all users in region i have the same physical data rate $R_i = \mu_i^{(l)} \sigma_l$ $(\forall l \leq L)$. We denote by $p_i^{(l)}$ the probability that a new data flow is initiated by a class-l user. The overall load in the absence of mobility is then given by:

$$\rho = \sum_i \sum_l \rho_i^{(l)} = \lambda/\mu,$$

where

$$\mu = \frac{1}{\sum_i \sum_l p_i^{(l)}/\mu_i^{(l)}}$$

and

$$\rho_i^{(l)} = \frac{p_i^{(l)} \lambda}{\mu_i^{(l)}}$$

is the load of class-l users in region i. In the general case, the mean throughput of class-l users in region i is given by:

$$\gamma_i^{(l)} = \frac{E\left(\mu_i^{(l)} \sigma_l \phi_i^{(l)}(X)\right)}{E\left(X_i^{(l)}\right)},$$

that is in the absence of mobility:

$$\gamma_i^{(l)} = \frac{\lambda \sigma_l}{E\left(X_i^{(l)}\right)}.$$

Similarly, the mean throughput in the cell is:

$$\gamma = \frac{E\left(\sum_i \sum_l \mu_i^{(l)} \sigma_l \phi_i^{(l)}(X)\right)}{E\left(\sum_i \sum_l X_i^{(l)}\right)},$$

that is

$$\gamma = \sum_i \sum_l p'^{(l)}_i \gamma_i^{(l)}$$

where

$$p'^{(l)}_i = \frac{E(X_i^{(l)})}{\sum_j \sum_k E(X_j^{(k)})}.$$

The cell throughput variance can be written as:

$$v = E\left(\left(\frac{\sum_i \sum_l \phi_i^{(l)} \mu_i^{(l)} \sigma^{(l)}}{X}\right)^2 X\right)/E(X) - \gamma^2.$$

a) Static case

Under the round robin policy the mean throughput of any class of flows in region i is: $R_i(1-\rho)$. Thus, the mean throughput in region i is still given by (II.13) and the mean throughput in the cell is given by (II.14). Moreover, the throughput standard deviation of any class of flows in region i is given by (II.10).
Under the max C/I policy, explicit expression for the mean throughput can be written only in region 1, which is for any class l equal to :

$$R_1(1-\rho_1)$$

where

$$\rho_1 = \sum_l \rho_1^{(l)}.$$

Similarly, for any class l the throughput standard deviation in region 1 is given by (II.16).

b) Impact of intra-cell mobility

Now the equivalent service rate in the limiting regime of infinite mobility is given by:

$$\bar{\mu} = \left(\sum_l \frac{p_l}{\sum_i q_i \mu_i^{(l)}} \right)^{-1}$$

under the round robin policy, and by

$$\bar{\mu} = \left(\sum_l \frac{p_l}{\mu_1^{(l)}} \right)^{-1}$$

under the max C/I policy. Note that

$$p_l = \sum_i p_i^{(l)}$$

is the probability that a new data flow is initiated by a user of class l. Consequently, the stability condition is given by

$$\rho < \frac{\bar{\mu}}{\mu}.$$

i. Throughput in light traffic The mean duration of a flow initiated in the cell can be computed as §II.2.2.2d) for each class l of users by replacing μ_i by $\mu_i^{(l)}$:

$$T^{(l)} = \sum_{i=1}^{N} \frac{p_i^{(l)}}{p_l} T_i^{(l)}.$$

Note that $T_i^{(l)}$ is give by (II.19) by replacing μ_i by $\mu_i^{(l)}$. We deduce the average flow throughput in light traffic for class-l users:

$$\gamma^{(l)} = \sigma^{(l)} / T^{(l)}.$$

The cell-average flow throughput is then given by:

$$\gamma = \left(\sum_l \frac{p_l}{\gamma^{(l)}} \right)^{-1}.$$

For two regions for instance the mean flow throughput of class-l users in light traffic is

$$\gamma^{(l)} = \frac{\nu_2 \mu_1^{(l)} + \nu_1 \mu_2^{(l)} + \mu_1^{(l)} \mu_2^{(l)}}{\nu_1 + \nu_2 + p_1^{(l)} \mu_2^{(l)} / p_l + p_2^{(l)} \mu_1^{(l)} / p_l} \sigma_l.$$

ii. Variance in light traffic The throughput variance in light traffic is given by

$$v = \sum_i \sum_l p'^{(l)}_i R_i^2 - \left(\sum_i \sum_l p'^{(l)}_i \mathrm{R}_i \right)^2$$

where

$$p'^{(l)}_i = \frac{E\left(X_i^{(l)}\right)}{\sum_j \sum_k E\left(X_j^{(k)}\right)}.$$

In the absence of mobility, the throughput variation mainly due to the variation of radio conditions is given by:

$$v = \sum_i \sum_l \frac{p_i^{(l)}/\mu_i^{(l)}}{\sum_j \sum_k p_j^{(k)}/\mu_j^{(k)}} \mathrm{R}_i^2 - \left(\sum_i \sum_l \frac{p_i^{(l)}/\mu_i^{(l)}}{\sum_j \sum_k p_j^{(k)}/\mu_j^{(k)}} \mathrm{R}_i \right)^2.$$

However, in the limiting regime of infinite mobility, the throughput variance is written as:

$$v = \sum_l \left(\frac{p_l/\bar{\mu}_l}{\sum_k p_k/\bar{\mu}_k} \sum_i q_i \mathrm{R}_i^2 \right) - \left(\frac{p_l/\bar{\mu}_l}{\sum_k p_k/\bar{\mu}_k} \sum_i q_i \mathrm{R}_i \right)^2$$

Where

$$\bar{\mu}_l = \sum_i q_i \mu_i^{(l)}.$$

II.2.2.6 Multiple mobility behaviors

We have assumed so far that all users have the same mobility behavior. In particular, they are all static or all mobile. We here extend the results to the case of multiple mobility behaviors, where each user may be static or mobile, as illustrated by Figure II.6. Then we propose a mobility-aware scheduler that exploits the mobility as an additional information in order to schedule users. Thus mobile users at the edge are deprioritized since they are likely to move and to be served in better radio conditions.

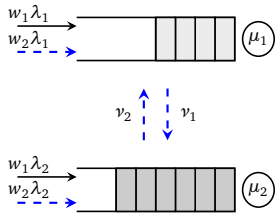


Figure II.6: Queuing model for two regions and two mobility patterns.

a) Mobility model

We consider K classes of mobility. For instance, each class represents a range of speeds (pedestrian, train, etc). Class-1 users are static. For any $k > 1$, a class-k user moves from region i to region $i+1$ and from region $i+1$ to region i (for $i < N$) at respective rates $\nu_{i,i+1}^{(k)}$ and $\nu_{i+1,i}^{(k)}$. We assume that each user belongs to a class of service l. We consider L classes where class-l has a mean flow size σ_l. The mean service rate of users of class l in region i is $\mu_i^{(l)}$. However, all users in region i have the same

physical data rate $R_i = \mu_i^{(l)}\sigma_l$ (for all $l \leq L$). We denote by $w_i^{(k,l)}$ the probability that a new class-l flow in region i is initiated by a class-k user. Let $X_i^{(k,l)}(t)$ be the number of active class-(k,l) flows in region i at time t. The Markov process $X(t)$ has now dimension $N \times K \times L$. The probability that a class-k user is in region i satisfies

$$q_i^{(k)} \propto \prod_{j=1}^{i-1} \frac{\nu_{j,j+1}^{(k)}}{\nu_{j+1,j}^{(k)}}.$$

b) Round-robin policy

In the limiting regime where mobile users have infinite speeds, class-(k,l) users are served at rate

$$\bar{\mu}^{(k,l)} = \sum_i q_i^{(k)} \mu_i^{(l)},$$

for any $k > 1$, while class-$(1,l)$ ($\forall l \leq L$) users are static and thus the total mean service rate of static users is given by:

$$\bar{\mu}^{(1)} = \frac{\sum_i \sum_l w_i^{(1,l)} p_i^{(l)}}{\sum_i \sum_l w_i^{(1,l)} p_i^{(l)} / \mu_i^{(l)}}.$$

Let $\bar{\rho}^{(1)} = \sum_i \sum_l w_i^{(1,l)} p_i^{(l)} \lambda / \bar{\mu}^{(1)}$ and $\bar{\rho}^k = \sum_l \sum_i w_i^{(k,l)} p_i^{(l)} \lambda / \bar{\mu}^{(k,l)}$ for $k > 1$. The queuing system corresponds to a multi-class processor-sharing queue of load $\bar{\rho} = \sum_{k \geq 1} \bar{\rho}^{(k)}$. We deduce the stationary distribution of the Markov process $X(t)$:

$$\pi(x) = (1-\bar{\rho})\bar{\rho}^n \binom{\sum_{i,l} x_i^{(1,l)}}{x_1^{(1,1)}, \ldots, x_N^{(1,L)}}$$

$$\times \left(\prod_{i,l} \left(\frac{w_i^{(k,l)} p_i^{(l)} \bar{\mu}^{(1)}}{\mu_i^{(1,l)}} \right)^{x_i^{(1,l)}} \right) \prod_{k>1} \left[\binom{\sum_{i,l} x_i^{(k,l)}}{x_1^{(k,1)}, \ldots, x_N^{(k,L)}} \prod_{i,l} {q_i^{(k)}}^{x_i^{(k,l)}} \right],$$

with $n = \sum_{i,k,l} x_i^{(k,l)}$ and under the following stability condition:

$$\rho < \frac{1}{\left(\sum_i \sum_l w_i^{(1,l)} p_i^{(l)} \right) \mu / \bar{\mu}^{(1)} + \sum_{k>1} \sum_l \left(\sum_i w_i^{(k,l)} p_i^{(l)} \mu / \bar{\mu}^{(k,l)} \right)}. \tag{II.23}$$

The mean throughput of static users in region i is

$$\gamma_i^{(1)} = R_i(1-\bar{\rho}),$$

while the mean throughput of class-l mobile users depends on their class k (for $k > 1$):

$$\gamma^{(k,l)} = \bar{\mu}^{(k,l)} \sigma (1-\bar{\rho}).$$

The cell-average throughput is

$$\gamma = \bar{\mu}\sigma(1-\bar{\rho}),$$

where

$$\bar{\mu} = \frac{1}{\left(\sum_i \sum_l w_i^{(1,l)} p_i^{(l)} \right) / \bar{\mu}^{(1)} + \sum_{k>1} \sum_l \left(\sum_i w_i^{(k,l)} p_i^{(l)} / \bar{\mu}^{(k,l)} \right)}.$$

c) Max C/I policy

These results can hardly be extended to the max C/I policy. Only the stability condition is explicit. At maximum load, mobile users are served only in region 1; we deduce the stability conditions for static users in region i:

$$\lambda\left(\sum_l\sum_{j\le i} w_j^{(1,l)}p_j^{(l)}/\mu_j^{(l)} + \sum_{k>1}\sum_l\left(\sum_j w_j^{(k,l)}p_i^{(l)}/\mu_1^{(l)}\right)\right) < 1.$$

The overall stability condition is:

$$\rho < \frac{1}{\left(\sum_i\sum_l w_i^{(1,l)}p_i^{(l)}\right)\mu/\bar{\mu}^{(1)} + \sum_{k>1}\sum_l\left(\sum_i w_i^{(k,l)}p_i^{(l)}\mu/\mu_1^{(l)}\right)}. \tag{II.24}$$

d) Mobility-aware scheduler

Without fast fading We propose in this section a mobility-aware scheduler which uses the mobility as an additional information in order to schedule users. For instance, we assign to each cell-edge user a score which is inversely proportional to its speed. This score is used by the scheduler. The more the cell-edge user is mobile the more he is deprioritized. For that purpose, we define a threshold I, so that any user in region $i \ge I$ is considered as a cell-edge user. For any region $i \ge I$, static users are served first and then mobile users. Hence we can enhance the global performance by giving the chance to mobile cell-edge users to be served in better radio conditions. We use a max C/I for both static and mobile users: users in region i are served only when there are no active users in region $j < i$. Thus users in region i, of a mobility class v are served a fraction of time:

$$\phi_i^{(v)}(x) = \begin{cases} \frac{\sum_l x_i^{(v,l)}}{\sum_{k,l} x_i^{(k,l)}} & \text{if } \sum_{j<i,k,l} x_j^{(k,l)} = 0 \text{ and } i < I. \\ 1 & \text{if } \sum_{j<i,k,l} x_j^{(k,l)} + \sum_{k<v,l} x_i^{(k,l)} = 0 \text{ and } i \ge I. \\ 0 & \text{otherwise.} \end{cases}$$

For two regions for instance, cell-center users are served first then static cell-edge users and finally mobile cell-edge users. At maximum load, mobile users are served only in region 1. Thus, the overall stability condition is still given by (II.24).

With fast fading Now we add fast fading to the model as in Section II.2.2.4 since in real networks, channel fluctuations due to fast fading are omnipresent. We modify the mobility-aware algorithm so that static users are served based on a proportional fair strategy and mobile users are served based on a max C/I strategy, in addition to the deprioritization based on mobility. The algorithm becomes as follows:

For $\nu > 1$

$$\phi_i^{(\nu)}(x) = \begin{cases} \frac{\sum_l x_i^{(\nu,l)}}{\sum_{k>1,l} x_i^{(k,l)} + \sum_{l,j} x_j^{(1,l)}} & \text{if } \sum_{j<i,k>1,l} x_j^{(k,l)} = 0 \text{ and } i < I. \\ 1 & \text{if } \sum_{j<i,k>1,l} x_j^{(k,l)} + \sum_{j,l} x_j^{(1,l)} \\ & \quad + \sum_{1<k<\nu,l} x_i^{(k,l)} = 0 \text{ and } i \geq I. \\ 0 & \text{otherwise.} \end{cases}$$

For $\nu = 1$

$$\phi_i^{(\nu)}(x) = \frac{\sum_l x_i^{(\nu,l)}}{\sum_{j,l} x_j^{(\nu,l)}} \times \left(1 - \sum_{j<I,k>1} \phi_j^{(k)}(x)\right).$$

For two regions for instance, cell-center users and static cell-edge users are served in round robin (an opportunistic gain must be taken into account) and finally mobile cell-edge users.

e) Throughput in light traffic

When $\rho \to 0$, the mean throughput of static users in region i is R_i while the mean throughput of class-l mobile users of class $k > 1$ is, for two regions,

$$\gamma^{(k,l)} = \frac{\nu_2^{(k)}\mu_1^{(l)} + \nu_1^{(k)}\mu_2^{(l)} + \mu_1^{(l)}\mu_2^{(l)}}{\nu_1^{(k)} + \nu_2^{(k)} + \left(w_1^{(k,l)}p_1^{(l)}\mu_2^{(l)} + w_2^{(k,l)}p_2^{(l)}\mu_1^{(l)}\right) / \left(w_1^{(k,l)}p_1^{(l)} + w_2^{(k,l)}p_2^{(l)}\right)} \sigma_l.$$

Observe that when $\nu_1, \nu_2 \to 0$, $\gamma = \frac{w_1^{(k,l)}p_1^{(l)} + w_2^{(k,l)}p_2^{(l)}}{w_1^{(k,l)}p_1^{(l)}/\mu_1^{(l)} + w_2^{(k,l)}p_2^{(l)}/\mu_2^{(l)}} \sigma_l$ which is the simple harmonic mean of $\mu_1^{(l)}$ and $\mu_2^{(l)}$, while $\gamma \to \bar{\mu}^{(k,l)}\sigma_l$ when $\nu_1, \nu_2 \to +\infty$.

II.2.3 Numerical results

We present in this section the numerical results in the case of one mobility behavior (without-with mobility) as well as in the case of multiple mobility behaviors.

II.2.3.1 Analysis

a) Case without mobility

We first consider the simple case of a cell with $N = 2$ regions, without mobility. Figure II.7 (solid lines) compares the throughput performance of both policies for two regions, with $p_1 = p_2 = 1/2$, $\mu_1 = 8$, $\mu_2 = 2$ and $\sigma = 1$. Observe that, under max C/I scheduling, the cell-center throughput is positive whenever $\rho_1 < 1$, that is $\rho < 5$, and decreases linearly. The throughput gain compared to the round-robin policy is huge for cell-center users, at the expense of a slight throughput degradation for cell-edge users. The cell-average throughput remains approximately the same since it corresponds to the harmonic mean throughput over the cell and thus is mainly affected by cell-edge users.

Figure II.8 (solid lines) shows the throughput standard deviation of the same particular case, under both policies. Observe that under max C/I policy the cell throughput standard deviation is

slightly higher than that under the round robin policy. The maximum throughput standard deviation for cell-center users is the same under both policies, while it is higher under max C/I policy for cell-edge users. Note that this maximum is reached at medium load.

b) Case with mobility

Consider two regions with $p_1 = p_2 = 1/2$, $\mu_1 = 8$, $\mu_2 = 2$ and $\sigma = 1$. The mobility rates are symmetric, that is $\nu_1 = \nu_2 = \nu$. We consider the static case where $\nu = 0$, a case with moderate mobility $\nu = 0.1$, and the case of very high (nearly infinite) mobility, $\nu = 100$.

i. Intra-cell mobility The results are obtained by the numerical evaluation of the stationary distribution of the Markov process $X(t)$ and shown in Figure II.7 and Figure II.8.

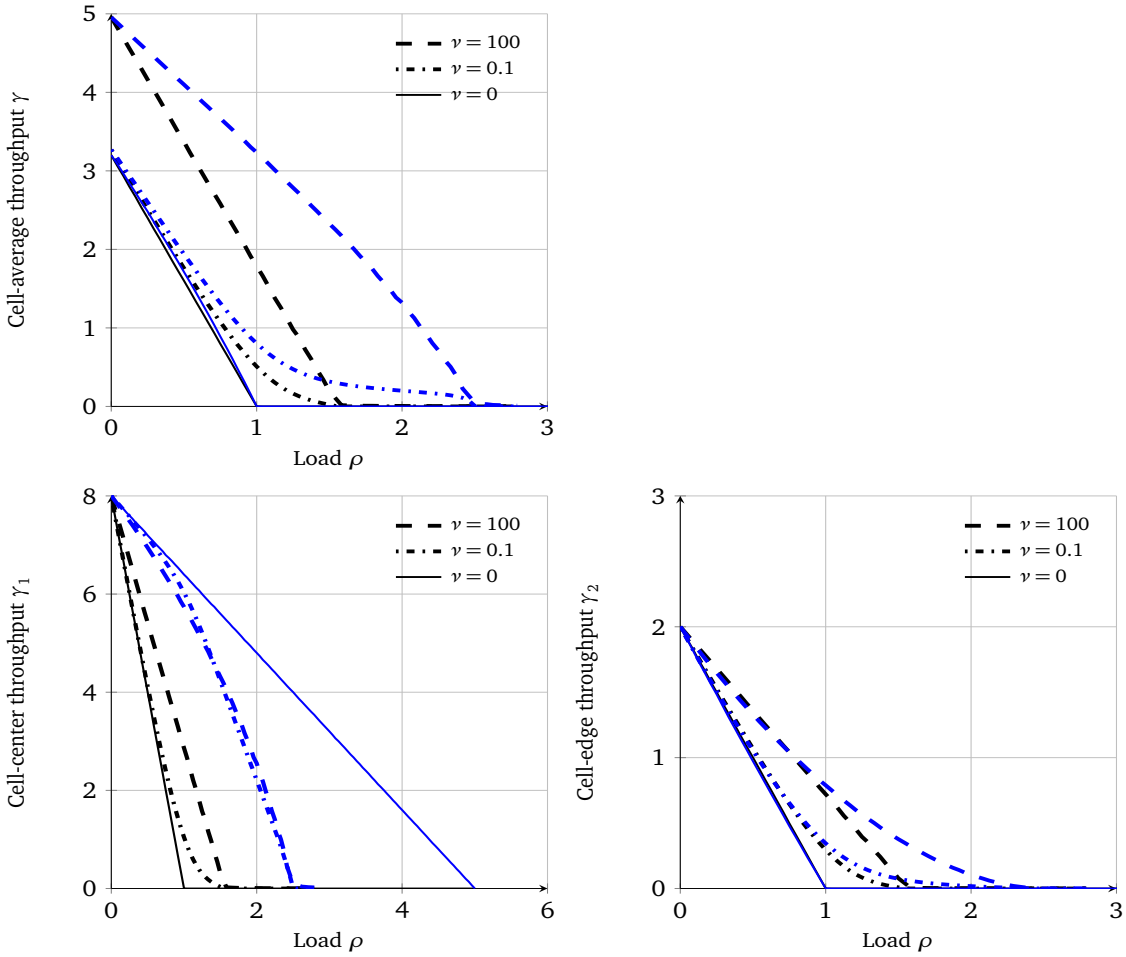


Figure II.7: Throughput performance for two regions with and without mobility, under round-robin (black) and max C/I (blue).

As expected, the maximum load when users move is independent of the mobility rate ν; it is approximately equal to 1.6 under the round-robin policy, in accordance with (II.17), and equal to 2.5

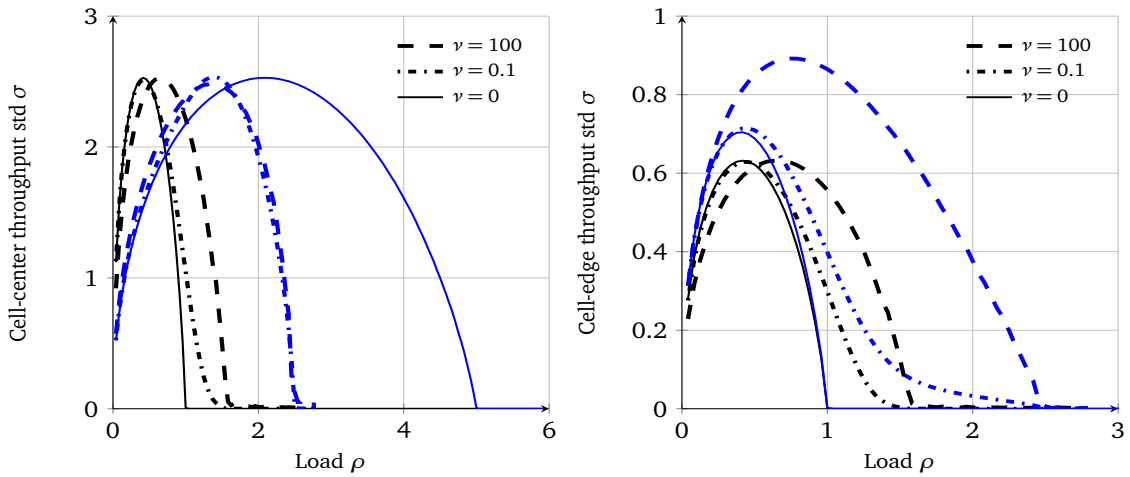


Figure II.8: Throughput standard deviation for two regions, under round-robin (black) and max C/I (blue).

under the max C/I policy, in accordance with (II.18), both corresponding to huge gains compared to the static case. The cell-average throughput in light traffic is independent of the scheduling policy but very sensitive to the mobility rate, as predicted by (II.20).

We observe that mobility improves the cell-average throughput for both scheduling policies, with a higher gain under the max C/I policy. However, mobility may decrease the cell-center throughput (depending on the load) since users may then suffer from bad radio conditions. Surprisingly, the max C/I policy outperforms the round-robin policy for cell-edge users in case of mobility, even in case of moderate mobility $\nu = 0.1$: the fact that cell-edge users are not scheduled (in the presence of cell-center users) is compensated by the fact that cell-center users complete more rapidly their file transfers.

Observe that under both policies, the maximum throughput standard deviation illustrated by Figure II.8 remains the same for cell-center users as the mobility rate increases. However, the maximum throughput standard deviation for cell-edge users increases with mobility under the max C/I policy, while it remains the same under the round robin policy. However the maximum standard deviation under the max C/I strategy remains close to that under the round robin strategy. Therefore, the max C/I improves the mean performance in the presence of mobility with no unfairness issues.

ii. Inter-cell mobility Figure II.9 shows the performance in terms of mean user throughput in the presence of inter-cell mobility when considering a handover rate $\nu_h = \nu = 0.1$, using fixed-point iterations to estimate the incoming handover rate. These results confirm that inter-cell mobility has only limited impact on throughput performance, in the presence of intra-cell mobility.

c) Impact of fast fading

The results obtained for the same scenario as in §b) are shown in Figure II.10. Compared to Figure II.7, max C/I brings a significant throughput gain due to fast fading, in addition to that due to slow fading. Observe that PF ($\alpha = 1$) outperforms max C/I at high load when users are static, while max C/I ($\alpha = 0$) remains the best strategy when users are mobile. The higher the mobility, the better is to decrease the fairness factor [26]. Note that in real networks the actual physical rates R_i depend on the load and they are inversely proportional to this latter. This is due to the fact that when a network is slightly loaded, the interference coming from the neighboring cells is low. However this simple

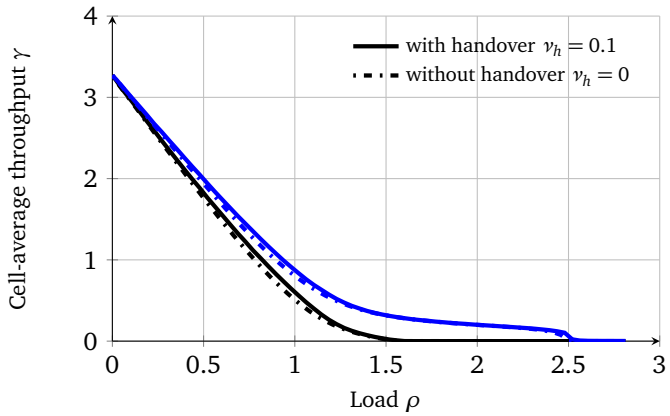


Figure II.9: Throughput performance for two regions with both intra-cell and inter-cell mobility, under round-robin (black) and max C/I (blue).

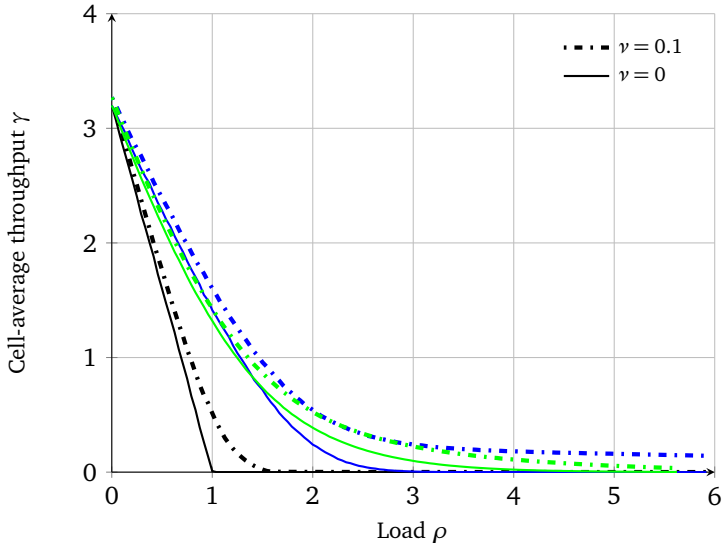


Figure II.10: Throughput performance for two regions with fast fading, under round-robin (black) and max C/I (blue) and proportional fair (green).

model where we consider the same physical rate independently of the load gives a general idea of the relative performance of different scheduling strategies with respect to the load without matching the exact throughput values especially at low load.

d) Impact of mean flow size

Consider two regions with $p_1 = p_2 = 1/2$, $R_1 = 8\,\text{MB/s}$, $R_2 = 2\,\text{MB/s}$ and two equiprobable classes of two different mean flow size: $\sigma^{(1)} = 1\,\text{MB}$ and $\sigma^{(2)} = 10\,\text{MB}$ (the distribution of the flow size is hyperexponential). The mobility rates are symmetric, that is $\nu_1 = \nu_2 = \nu$. As in the previous section, we consider the static case where $\nu = 0$, the case with $\nu = 0.1$, and the case of very high (nearly infinite) mobility, $\nu = 100$.

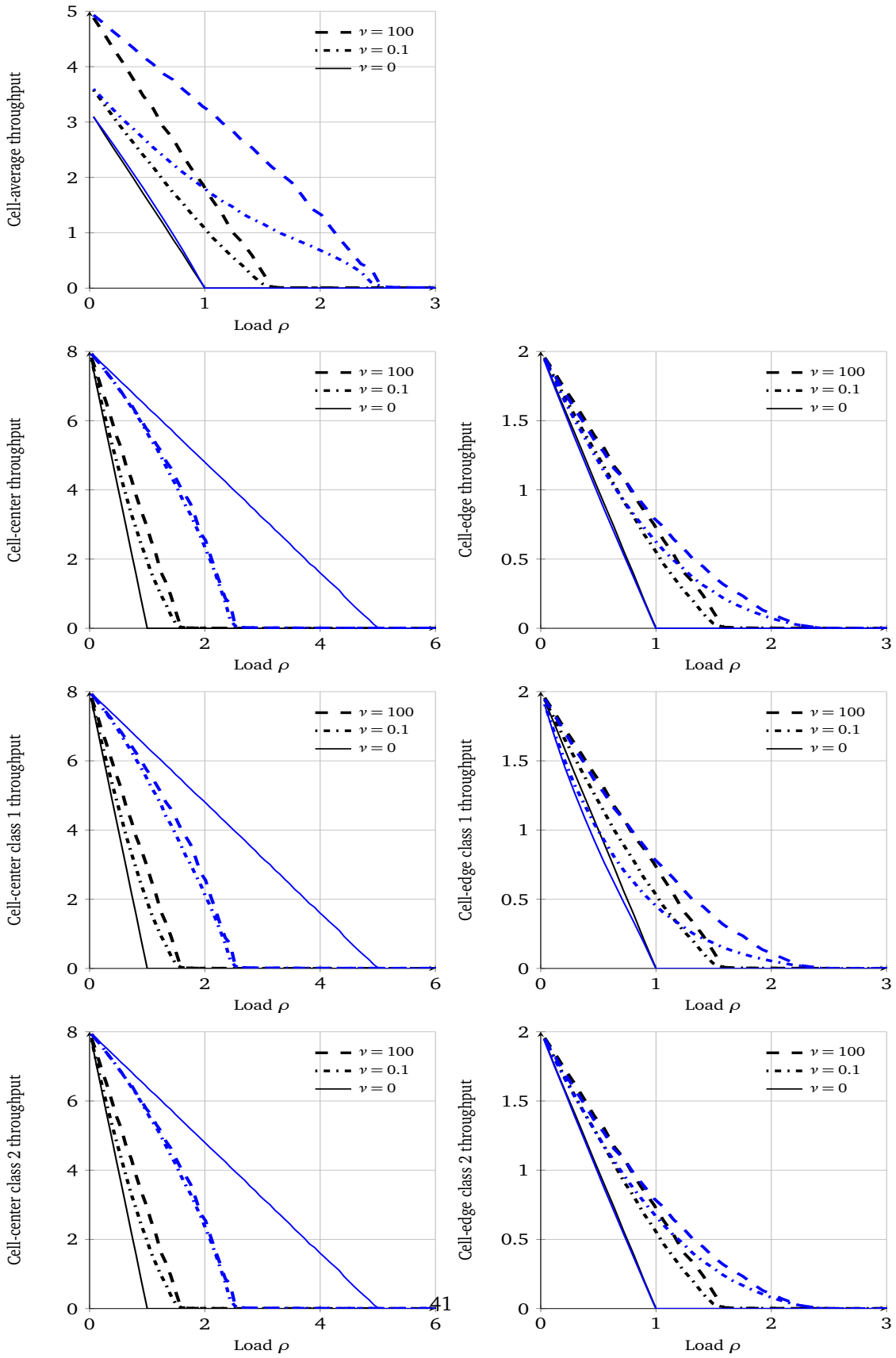


Figure II.11: Throughput performance for two regions and with two classes of data flows, under round-robin (black) and max C/I (blue).

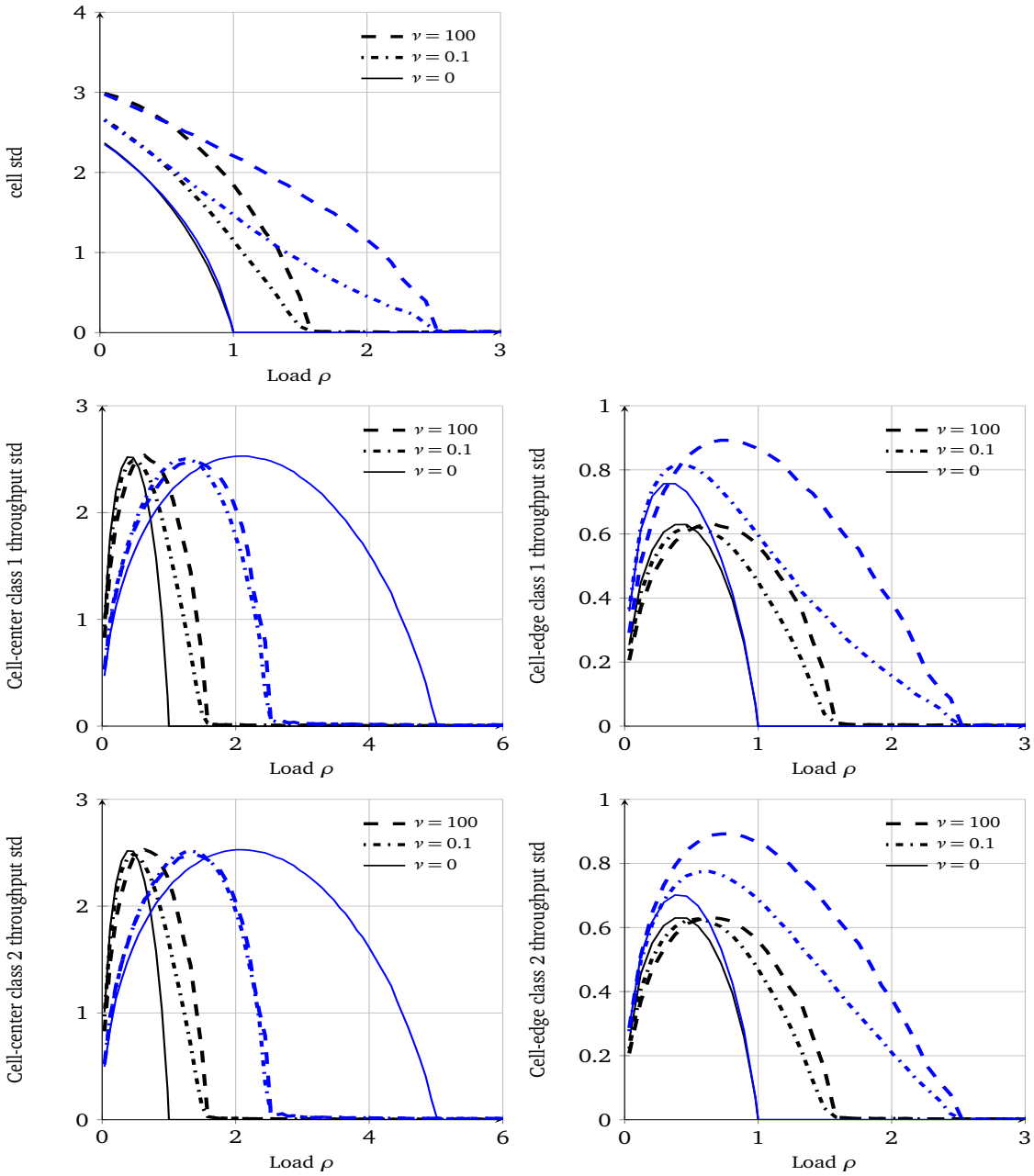


Figure II.12: Throughput standard deviation for two regions and with two classes of data flows, under round-robin (black) and max C/I (blue).

The results are obtained by the numerical evaluation of the stationary distribution of the Markov process and shown in Figure II.11 and Figure II.12. While $\nu = 0.1$ is considered as a moderate mobility rate for data flows of class 1, it is seen as a high mobility rate for data flows of greater mean flow size, since users in this class complete less rapidly their file transfer and have the chance to experiment more radio conditions during their sojourn time. Hence, a mobility rate (speed) can be said to be high or low according to the mean file size. The higher the mean flow size the higher is the impact of a given mobility rate on the performance. But otherwise, there is almost no impact of the flow size distribution on the global performance, in terms of mean throughput as well as throughput variance as shown in Figure II.11 and Figure II.12.

Note that a mobility rate (speed) can be said to be high or low according to the mean flow size. A mobility rate which is considered as a moderate mobility rate for a given mean flow size, is seen as a high mobility rate when the mean flow size is greater. This is due to the fact that users of large mean flow size complete less rapidly their file transfer and have the chance to experiment more radio conditions during their sojourn time. Hence, the higher the mean flow size the higher is the impact of a given mobility rate on the performance. But otherwise, there is almost no impact of the flow size distribution on the global performance, in terms of mean throughput as well as throughput variance.

e) Multiple mobility patterns

Without fast fading Figure II.13 gives the numerical results for the same example as in §b) but with two equiprobable classes of users: a static class and a mobile class ($\nu = 0.1$). The maximum load is approximately 1.2 for the round-robin policy, as given by (II.23), and 1.4 under the max C/I and the mobility-aware policies, as given by (II.24). We observe that the mobility-aware scheduler outperforms the two other strategies, mainly at high load. Note that the relative gain with respect to max C/I is only due to the deprioritization of mobile cell-edge users.

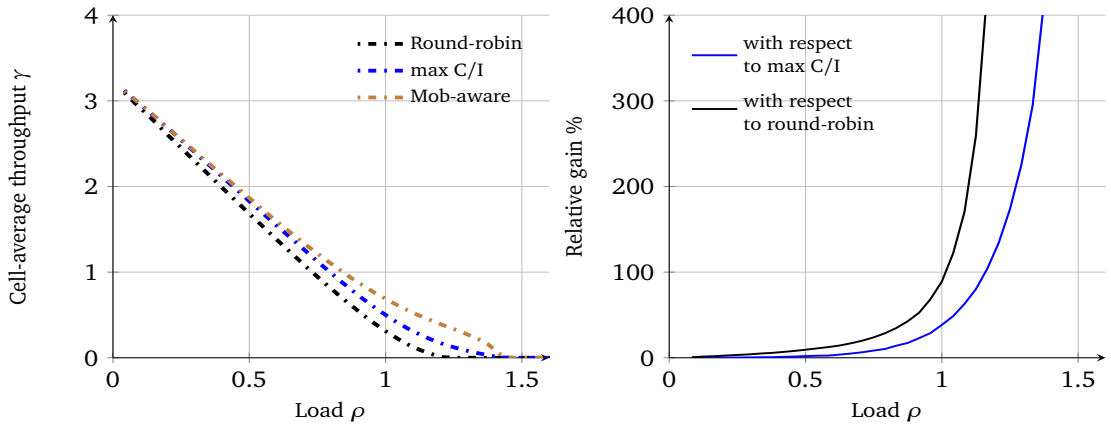


Figure II.13: Throughput performance for two regions with two mobility patterns.

With fast fading Figure II.14, Figure II.15 and Figure II.16 show the mean throughput under the proportional fair, the max C/I and the mobility-aware scheduler as well as the relative gain of the mobility-aware scheduler with respect to proportional fair and max C/I, for two mobility behaviors (static or mobile) and two different mobility rates $\nu = 0.1$ and $\nu = 1$, with 50%, 75% and 90% mobile users respectively.

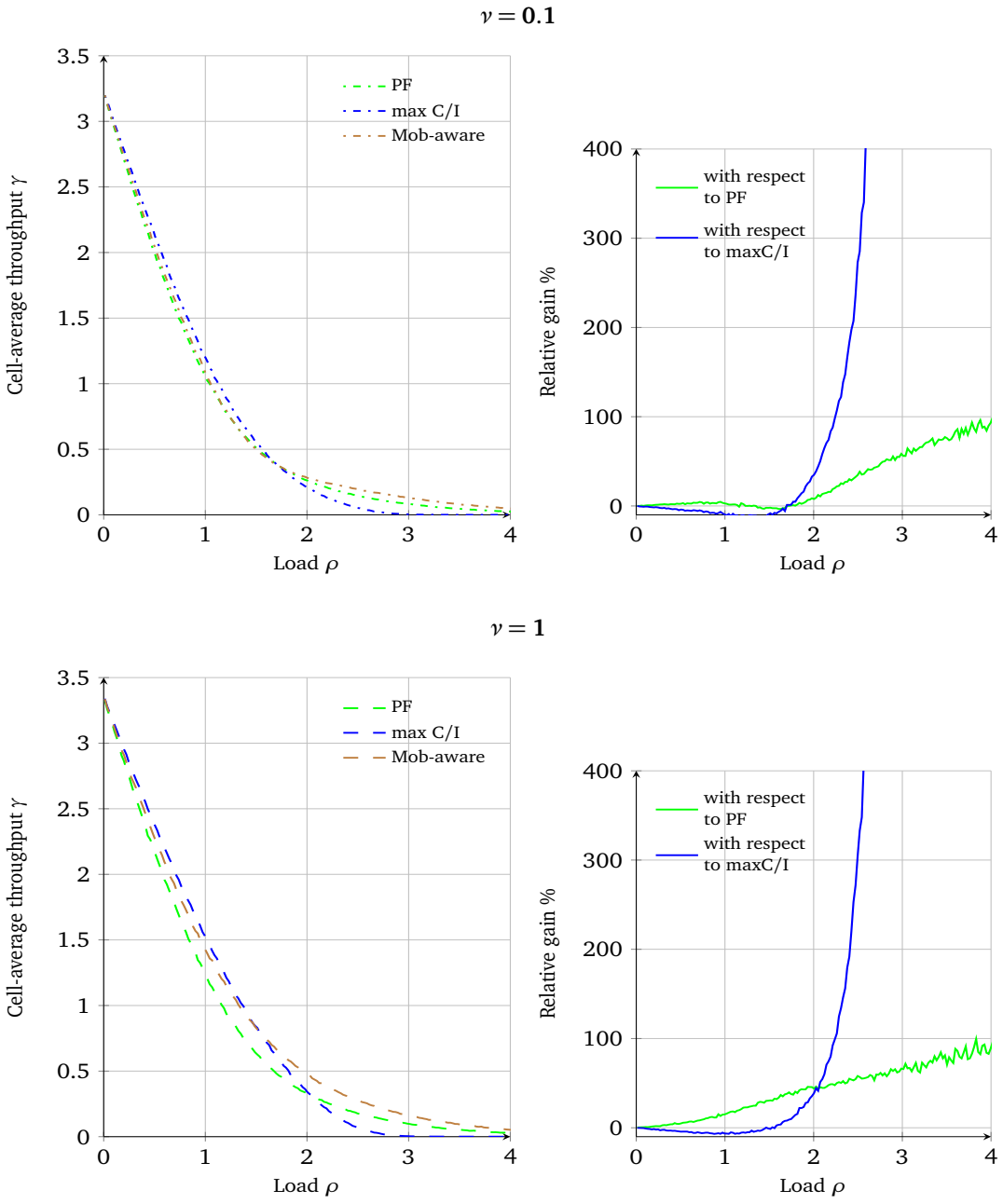


Figure II.14: Throughput performance for two regions with two classes of mobility and 50% mobile users.

Results show that there is a significant gain brought by the mobility-aware scheduler at high load so this strategy may be promising since real cellular networks are high loaded. Note that the higher the mobility rate and the percentage of mobile users the higher the gain.

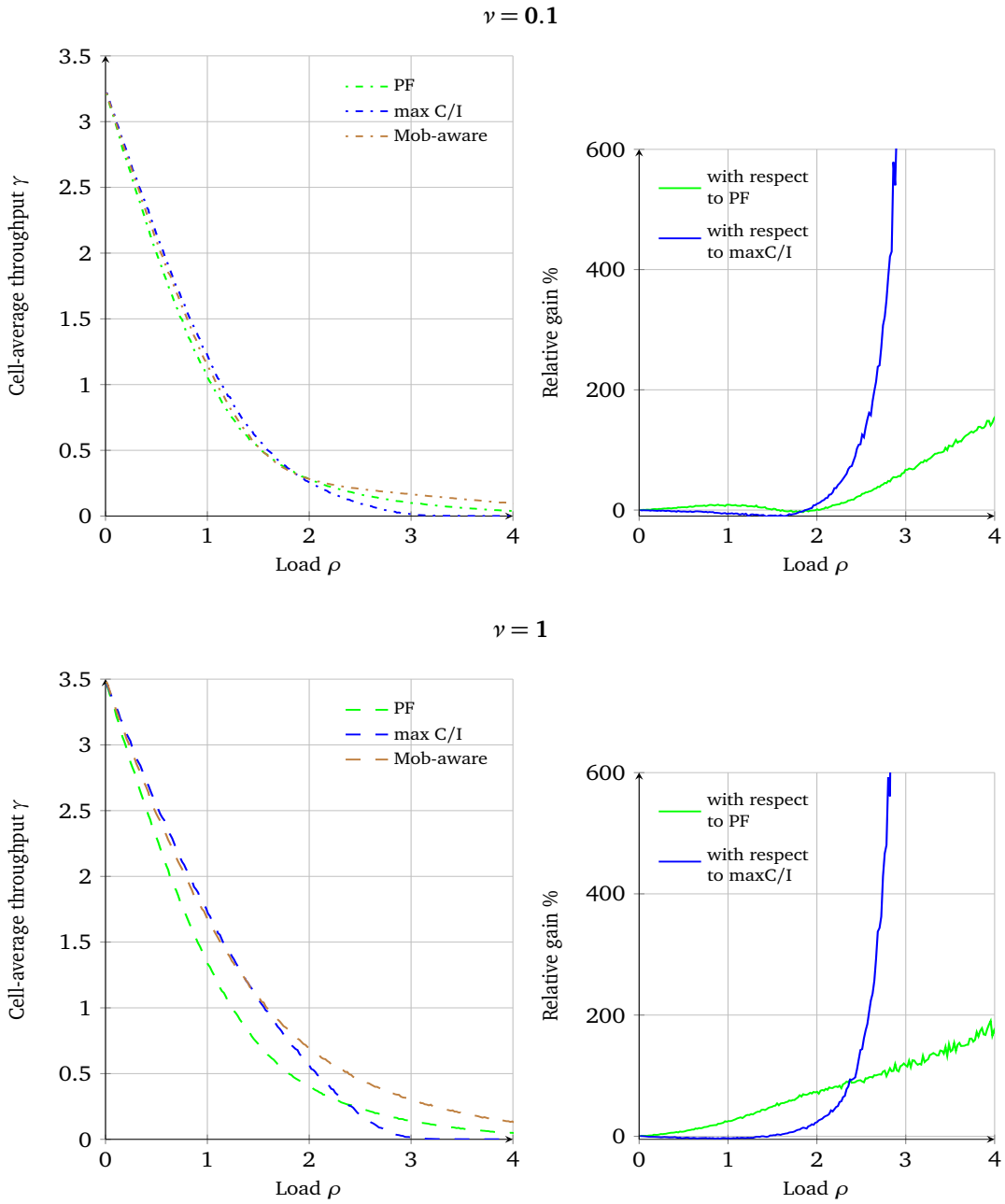


Figure II.15: Throughput performance for two regions with two mobility patterns and 75% mobile users.

II.2.3.2 System simulations

a) Simulation setting

We now validate the results of previous sections by system-level simulations based on the LTE technology. We consider the 21 hexagonal cells formed by 7 tri-sector sites (a reference site surrounded

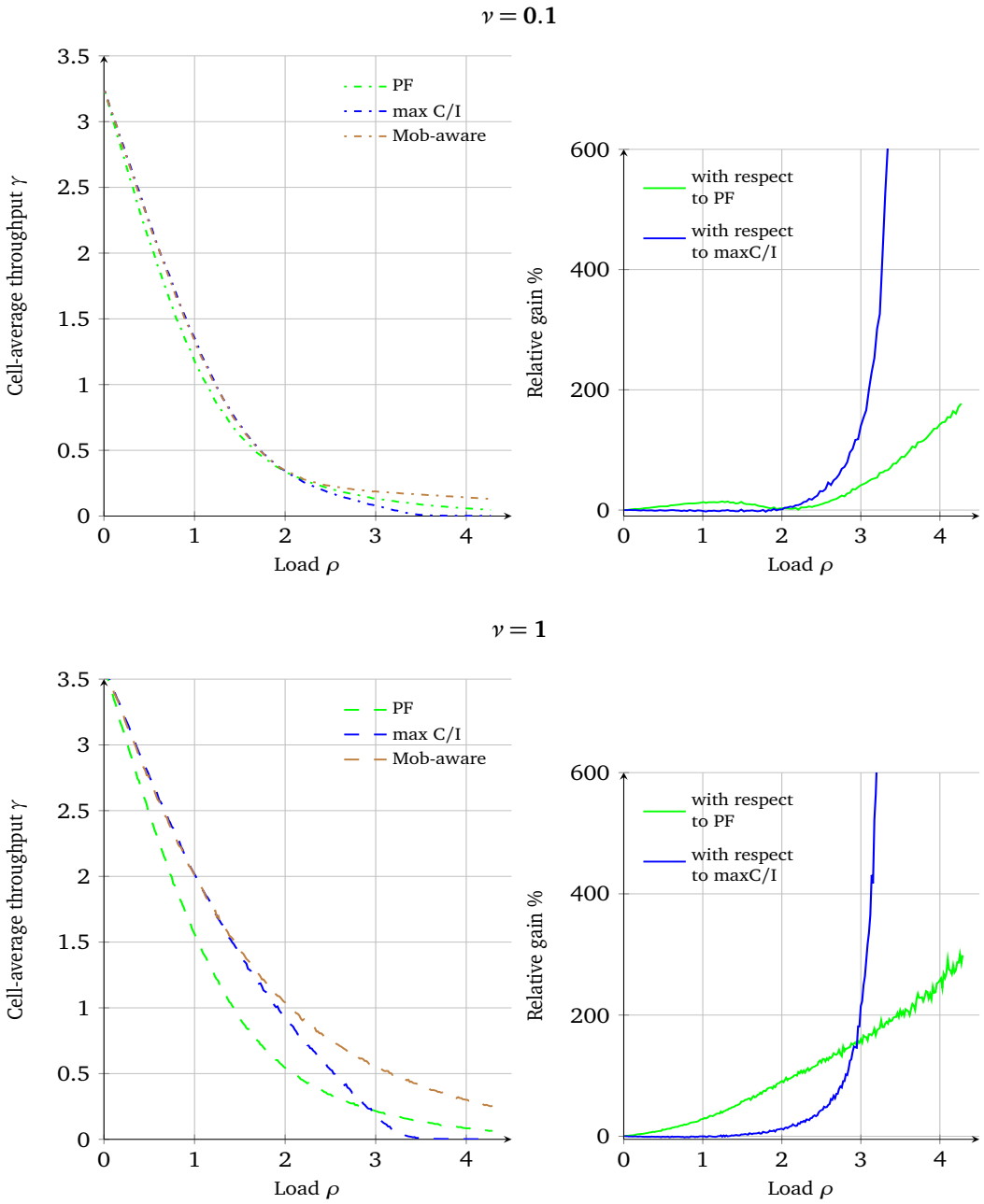


Figure II.16: Throughput performance for two regions with two mobility patterns and 90% mobile users.

by 6 interfering sites). The main parameters are summarized in Table II.1. See Appendix A, for more details.

Traffic consists of file transfers only. Flows arrive according to a Poisson process with uniform spatial distribution. File sizes are generated from an exponential distribution with mean σ (MB). Scheduling decisions are taken in each eNodeB at each TTI. In order to evaluate the proposed model,

Network topology	Macro cells only
Environment	Urban
Context	outdoor
Inter-site distance	500 m
Resolution	5 m
PathLoss	ITU Model
shadowing std	4 dB
shadowing cross correlation (between sites)	0.5
shadowing spatial correlation	$e^{\frac{-\delta d}{d_{corr}}}$
shadowing correlation distance	30 m
Radio access technology	LTE
Number of tx/rx antennas	MIMO (2x2)
number of streams	1
Receiver	MRC
Codebook	3GPP
carrier frequency	2 GHz
Bandwidth	10 MHz
Simulation time	100 minutes=6×10^6 TTI

Table II.1: Simulation setting.

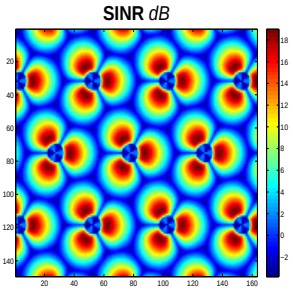


Figure II.17: SINR map

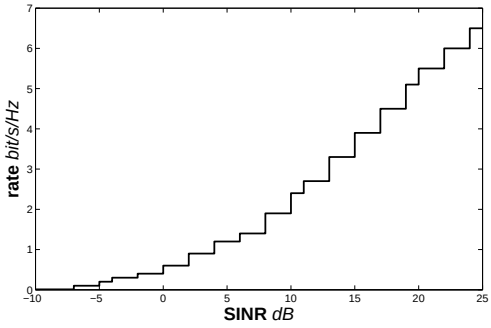


Figure II.18: SINR-Rate Mapping according to LTE CQI table

the three previous schedulers are implemented. The round-robin scheduler serves users in a cycle way, independently of their radio conditions. The mapping between the SINR and the rate (in bit/s/Hz) according to LTE CQI (channel quality indicator) table is given by Figure II.18. Given the rate (in bit/s/Hz), we can obtain the amount of data that can be sent to the considered user in each RB (resource block) allocated to him. Figure II.17 gives the SINR maps using the ITU pathloss model in an urban environment.

In the presence of mobility, each user chooses a direction at random and moves along this direction at a constant speed of 100 and 300 km/h until the file transfer is completed. The network is a torus

and a wrap-around technique is employed in this case to avoid border effects (that is, a user leaving cell 20 from the south arrives in cell 6, 7 or 13 to the north according to the direction). Users can make handover and move from one cell to another. Note that in this work we neglect the throughput decrease for mobile users, which is mainly due to the errors in channel estimation. We suppose that the eNodeB receives a channel feedback from the user and decode it without error each TTI. Based upon this feedback, a particular modulation order and code rate is chosen to transmit to the scheduled user.

b) One mobility behavior

We try out two scenarios, with and without mobility. We simulate 100 minutes and estimate for each scenario the cell-average throughput as the ratio of mean flow size to mean flow duration, that is the average user throughput weighted by the sojourn time.

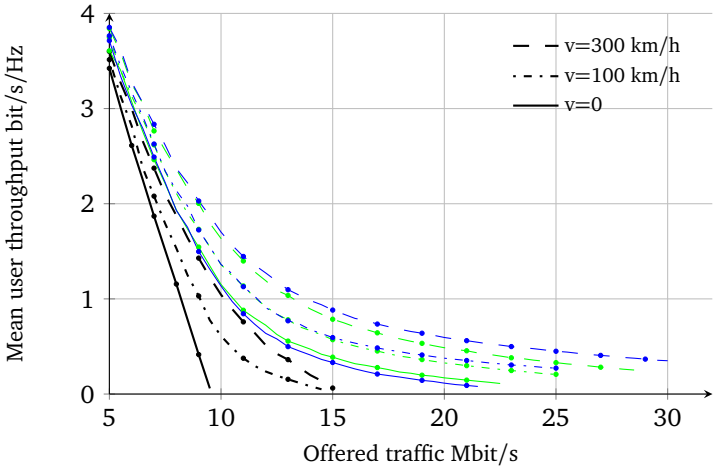


Figure II.19: Throughput performance obtained by system-level simulation for one mobility pattern, under round-robin (black) and max C/I (blue) and proportional fair (green).

Figure II.19 shows the normalized mean user throughput (in bit/s/Hz) with respect to the offered traffic per eNodeB in each scenario for $\sigma = 1.25$ MBytes. Results are very close to those obtained by analysis, compared to Figure II.10. Note that the estimated harmonic mean physical rate given by (II.3) at full interference is $R = 9$Mbit/s thus the offered traffic 9Mbit/s corresponds to load $\rho = 1$. This can be verified based on throughput performance under the round robin strategy, where the mean throughput goes to zero near 9Mbit/s. Moreover, the different stability conditions are almost the same compared to Figure II.10, given $\rho = \lambda\sigma/R$ where $\lambda\sigma$ is the offered traffic.

c) Impact of t_s

Consider now the standard PF scheduler where the average experienced throughput is evaluated as follows:

$$\bar{R}_i(t) = \left(1 - \frac{1}{t_s}\right)\bar{R}_i(t-1) + \frac{1}{t_s} r_i(t-1)\mathbb{1}_{\{u(t-1)=i\}}.$$

We estimate the mean user throughput under PF scheduler with different values of t_s. We consider the case of fixed values of t_s: 10, 100, 500, 1000 and 10000 ms as well as the case where t_s is user-specific and is equal to the sojourn time of the considered user at current time. Figure II.20 shows

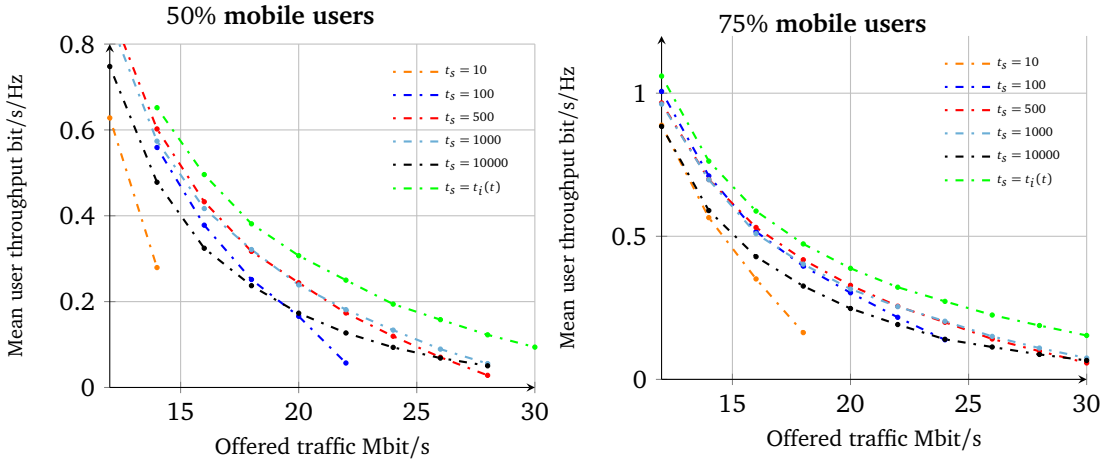


Figure II.20: Impact of t_s under PF scheduler.

the case where 50% of users are mobile and the case where 75% of users are mobile. We can see that the choice of t_s impact strongly the throughput performance. By choosing a fixed value of t_s, we can obtain only limited throughput performance compared to the case where $t_s = t_i(t)$, that is when evaluating $\bar{R}_i(t)$ as the experienced throughput by user i during his current sojourn time. Consequently, we consider in the next section the PF scheduler with $t_s = t_i(t)$ as a reference in order to evaluate the performance of the proposed mobility-aware scheduler.

d) Mobility-aware scheduler

For the sake of simplicity we define only two classes of mobility: a static class and a mobile class. We suppose for instance that any user whose mean radio conditions variation is negligible belongs to the mobile class, while any user with considerable mean radio conditions variation belongs to the static class. According to previous results and as we can see in Figure II.19, the proportional fair is the best scheduling strategy when users are static. However, when users are mobile, the best strategy is the max C/I.

In the simulations presented in Figure II.21 and Figure II.22, we suppose that there are two mobility behaviors: a static class and a mobile class where users move at constant speed 300 km/h. Figure II.21 shows the case where 75% of the users are mobile while Figure II.22 shows the case where 90% of the users are mobile. We simulate 100 minutes and estimate the mean user throughput as the ratio of the mean flow size to mean flow duration for $\sigma = 1.25$ MBytes, under the mobility-aware scheduler as well as under two classical schedulers: proportional fair and max C/I. We evaluate also the mean number of users per eNodeB under the different scheduling strategies and the relative throughput gain under the mobility-aware scheduler compared to the throughput under proportional fair scheduler and max C/I scheduler, with respect to the offered traffic per eNodeB. Results are very close to those obtained by analysis, compared to Figure II.14 and Figure II.15 and show that the mobility-aware scheduler is a good compromise in a network with multiple mobility behaviors, and it improves the performance especially at high load. Observe that the higher the proportion of mobile users the higher the relative gain with respect to the PF scheduler, and the closer to the max C/I performance. So that it will correspond to the max C/I strategy in a network with 100% mobile users. However, in the absence of mobility the proposed scheduler corresponds to the PF strategy.

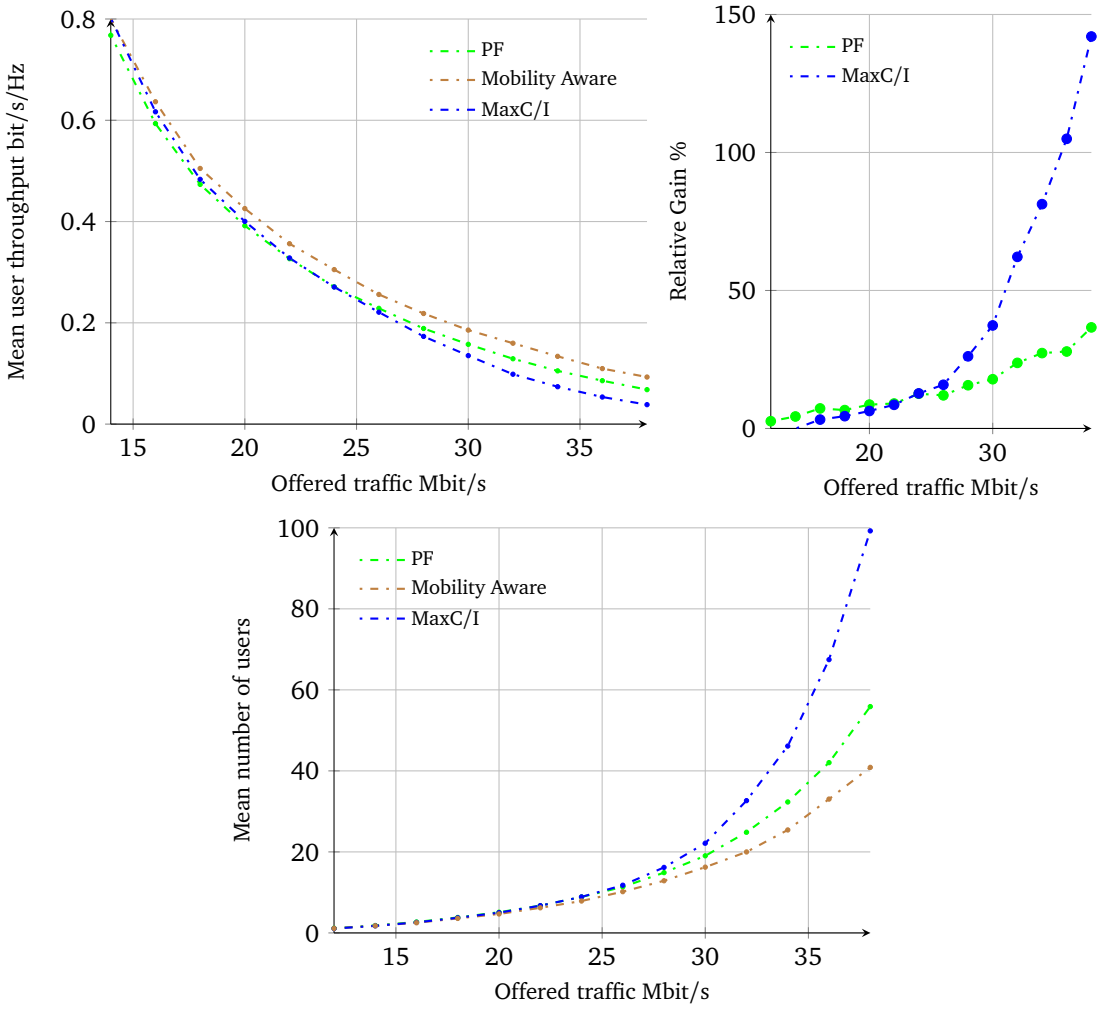


Figure II.21: Throughput performance obtained by system-level simulation for 75% mobile users.

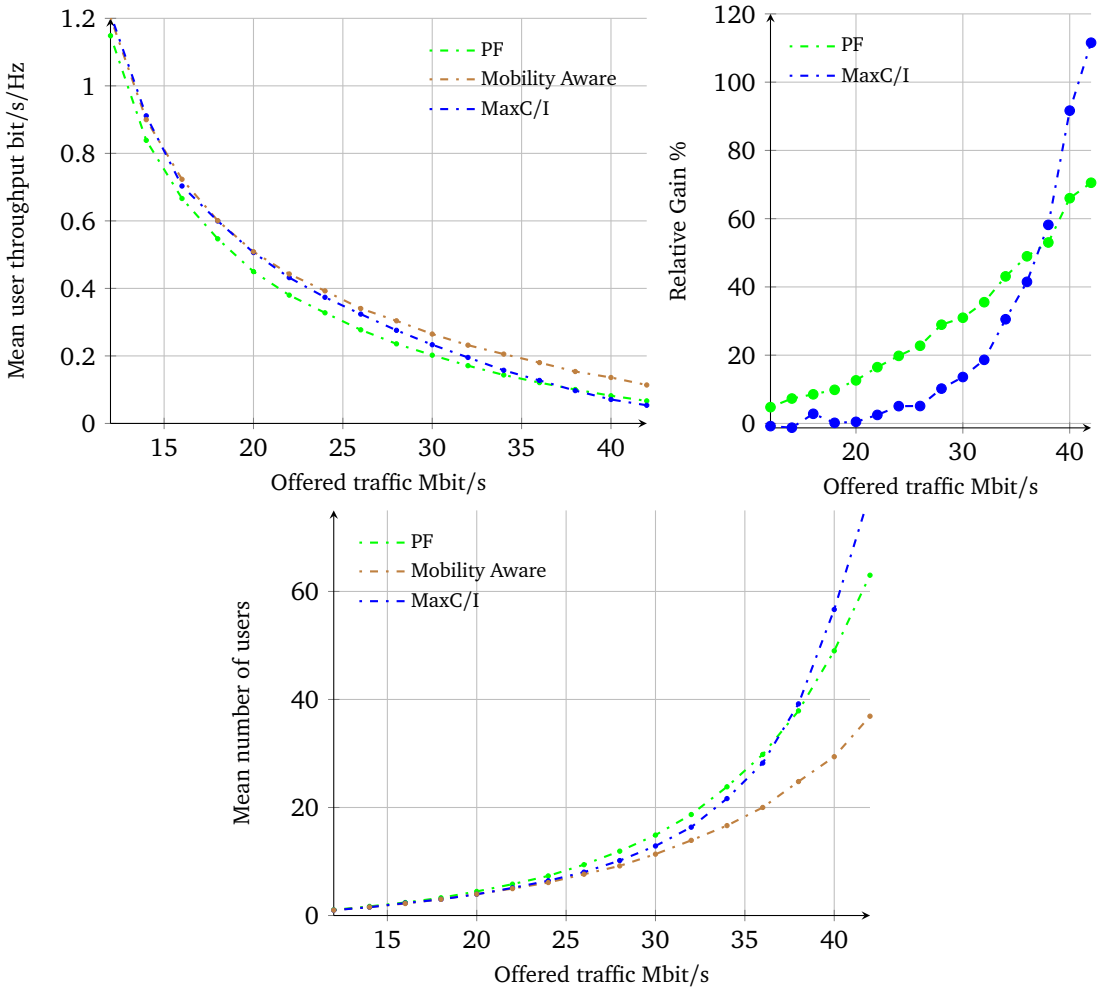


Figure II.22: Throughput performance obtained by system-level simulation for 90% mobile users.

II.3 Adaptive streaming

II.3.1 Flow level model

II.3.1.1 Model without mobility

We first present the model in the absence of mobility and the corresponding performance results under round-robin, max C/I and max-min scheduling schemes.

a) Traffic model

We consider same cell model explained in §II.2.2.1a). Consequently each cell is viewed as a set of N parallel queues with coupled processors as illustrated by Figure II.2.

We consider adaptive streaming only. Adaptive streaming provides chances for users to choose proper video bit rates and to avoid network congestion. The adaptive streaming system like MPEG-DASH [67] shows that video segment encoded by these video bit rate are prepared at the server sides. At the client side, users download the desired video segment by segment and buffer them to wait for video playout. In our studies, we assume that the video segment duration is significantly small. This offers the possibility for users to adapt their video bit rate once their instantaneous throughput changes. In addition, we assume that buffer at client size are significantly large to store all the video files.

For the traffic generation, we assume that new data flows are generated in region i at the random times of a Poisson process of intensity λ_i. We denote by $\lambda = \sum_i \lambda_i$ the total flow arrival rate in the cell and by $p_i = \lambda_i/\lambda$ the probability that a new data flow is generated by a user in region i. The video durations are assumed to be independent and exponentially distributed with mean T. When region i is served, flows are completed at the physical rate R_i (in bit/s). We assume regions are numbered in decreasing order of physical rates, that is $R_1 > R_2 > \ldots > R_N$.

Queue i represents the number of active flows in region i. Since the radio resources are shared in time, the actual service rate of queue i is modulated by ϕ_i, the fraction of time spent by the scheduler on users in region i. This depends on the system state and the scheduling policy. For work-conserving policies, we have $\sum_i \phi_i = 1$.

We denote by $X(t)$ the system state at time t. This is an N-dimensional vector whose component i gives the length of queue i at time t. It is an irreducible Markov process with stationary distribution π. In any state $\boldsymbol{x}$ such that $x_i > 0$, each user in region i has throughput:

$$\frac{R_i \phi_i(\boldsymbol{x})}{x_i}.$$

We assume a continuous set of video bit rates (codec) between two values, $v_{\min}$ and $v_{\max}$. The video bit rate of a user in region i is chosen according to his instantaneous throughput and can be written as:

$$v_i(\boldsymbol{x}) = \max\left(\min\left(\frac{R_i \phi_i(\boldsymbol{x})}{x_i}, v_{\max}\right), v_{\min}\right). \tag{II.25}$$

The higher the chosen video bit rate the higher is the size of a video and the larger is the flow size. Because of the memoryless property of exponential distribution, the mean flow size of a video in region i at state x can be expressed as:

$$\sigma_i(\boldsymbol{x}) = v_i(\boldsymbol{x})T.$$

Then the service rate of queue i at state $\boldsymbol{x}$ is:

$$\mu_i(\boldsymbol{x}) = \frac{\phi_i(\boldsymbol{x})R_i}{\sigma_i(\boldsymbol{x})} = \frac{\phi_i(\boldsymbol{x})R_i}{v_i(\boldsymbol{x})T}.$$

The stationary distribution $\pi(\boldsymbol{x})$ can be obtained by solving the following balance equations:

$$\sum_i (\lambda_i + \mu_i(\boldsymbol{x}))\pi(\boldsymbol{x}) = \sum_i \Big(\lambda_i \pi(\boldsymbol{x} - \boldsymbol{e}_i) + \mu_i(\boldsymbol{x} + \boldsymbol{e}_i)\pi(\boldsymbol{x} + \boldsymbol{e}_i)\Big),$$

where $\boldsymbol{e}_i$ represents a unit vector with value at i-th term. In the general case, the system stability is maintained when the following condition is satisfied:

$$\sum_i \frac{p_i \lambda}{R_i / \Big(\lim\limits_{\max_i E(X_i)\to\infty} v_i(\boldsymbol{x})\, T\Big)} < 1. \tag{II.26}$$

We shall see that the precise stability condition depends on the scheduling strategy.

b) Performance metrics

i. Video bit rate Under the assumption that users watch all the downloaded video.The first performance metrics we measure is the mean video bit rate, which is the average video resolution that a user experiences while watching the video. Noting that the expectation is calculated based on the stationary distribution $\pi(\boldsymbol{x})$, in the static case, we can compute the mean video bit rate of users in each region i as follows:

$$\bar{v}_i = \frac{E\left(\phi_i(X)R_i\right)}{\lambda_i T}. \tag{II.27}$$

The cell mean video bit rate in both scenarios (without and with mobility) is given by:

$$\bar{v} = \frac{E\left(\sum_i \phi_i(X)R_i\right)}{\sum_i \lambda_i T}. \tag{II.28}$$

Intuitively speaking, the mean video bit rate is obtained by dividing the average wireless resource allocated to different classes by the average number of arriving flows. In the case with mobility, only the cell mean video bit rate $\bar{v}$ can be calculated. However, in the absence of mobility it is observed that $\bar{v}$ is nothing then the arithmetic mean of $\bar{v}_i$ weighed by the probabilities p_i:

$$\bar{v} = \sum_i p_i \bar{v}_i. \tag{II.29}$$

ii. Buffer surplus The distribution seen by users in region i is the size-biased distribution. In the following discussion, we denote by E_i the expectation using corresponding biased distribution:

$$\pi_i(\boldsymbol{x}) \propto x_i \pi(\boldsymbol{x}).$$

In addition to the mean video resolution, the quality of experience is also influenced by the video smoothness measured in terms of buffer surplus. By calculating the buffer surplus as Eq. II.30, this performance metrics reflects the average relative buffer variation of each flow.

$$\bar{B}_i = E_i\left(\frac{R_i\phi_i(X)/X_i - v_i(X)}{v_i(X)}\right) \tag{II.30}$$

That is:

$$\bar{B}_i = E\left(\frac{R_i\phi_i(X) - X_i v_i(X)}{v_i(X)}\right)/E(X_i).$$

Similarly, the mean buffer surplus over the cell is:

$$\bar{B} = E\left(\sum_i \frac{R_i\phi_i(X) - X_i v_i(X)}{v_i(X)}\right)/E\left(\sum_i X_i\right). \tag{II.31}$$

Note that

$$\bar{B} = \sum_i p_i' \bar{B}_i, \text{ with } p_i' = \frac{E(X_i)}{\sum_j E(X_j)}$$

where p_i' represents the probability that an active user is in region i. When load becomes large, both $\bar{B}$ and $\bar{B}_i$ values approach -1.

c) Scheduling Policies

In this section, we study the extreme cases of scheduling policies: the round-robin policy, the max C/I policy as well as the max-min policy.

i. round-robin policy Under the round-robin policy, users share the radio resources equally, independently of their radio conditions. Thus users in region i are allocated a fraction

$$\phi_i(\boldsymbol{x}) = \frac{x_i}{\sum_j x_j},$$

of radio resources in state $\boldsymbol{x}$.

At high load all users in all regions select $v_{\min}$ as their video bit rate under the round robin policy. Thus the stability condition follows from (II.26):

$$\rho = \sum_i \frac{p_i\lambda}{\frac{R_i}{v_{\min}T}} < 1 \;\rightarrow\; \lambda_{\max} = \Big(\sum_i \frac{p_i v_{\min} T}{R_i}\Big)^{-1}, \tag{II.32}$$

where $\lambda_{\max}$ is the maximum arrival rate that the system can handle. Observe that this corresponds to the less restrictive stability condition that can be obtained from (II.26).

ii. max C/I policy The max C/I policy is an extreme case of opportunistic scheduling strategies that prioritizes those users with the best radio conditions. For two regions for instance, cell-center users are scheduled first and are allocated all the resources whenever active; cell-edge users are served only when there are no active cell-center users.

Stability condition approximation: Under the max C/I policy, base station first allocate its resources to the cell-center flows. Then it allocates the rest of resources to users having lower physical throughput. The maximum traffic intensity happens when the rest of resource allocated to the cell-edge users can only support for the selection of $v_{\min}$, which is shown as

$$\lambda_{\max} = \Big(\sum_{i\neq N} \frac{p_i \bar{v}_i T}{R_i} + \frac{p_N v_{\min} T}{R_N}\Big)^{-1}, \tag{II.33}$$

where $\bar{v}_i$ is also a function of $\lambda_{\max}$. Therefore, $\lambda_{\max}$ of max C/I policy can be solved as a fixed-point solution of Eq.(II.33) and $\bar{v}_i$ is calculated as

$$\bar{v}_i = \sum_{n=0}^{\infty} \pi_i(n) v_i(n), \tag{II.34}$$

with

$$\pi_i(n) = \pi_i(0) \prod_{j=1}^{n} \frac{p_i \lambda_{\max} R_i'}{v_i(j) T}, \tag{II.35}$$

$$v_i(n) = \max\Big(\min\Big(\frac{R_i'}{n}, v_{\max} \Big), v_{\min} \Big), \tag{II.36}$$

where $R_i' = \prod_{j<i} \pi_j(0) R_i$ and $\pi_i(0)$ is denoted as the probability that there is no flows in class-i.

iii. max-min policy The max-min policy achieves fairness through users throughput equalization. Users in good radio conditions are allocated fewer resources while users in bad radio conditions get the largest share, so that all users get the same throughput:

$$\frac{\phi_1(\boldsymbol{x}) R_1}{x_1} = \frac{\phi_2(\boldsymbol{x}) R_2}{x_2} = \dots \frac{\phi_N(\boldsymbol{x}) R_N}{x_N}.$$

Thus users in region i are allocated a fraction

$$\phi_i(\boldsymbol{x}) = \frac{x_i / R_i}{\sum_j x_j / R_j} \tag{II.37}$$

of radio resources in state $\boldsymbol{x}$. Similarly to the round robin policy, all users get $v_{\min}$ at high load and the stability condition is given by (II.32).

II.3.1.2 Impact of intra-cell Mobility

In this section, we add intra-cell mobility to the previous model, as shown in Figure II.5 for two regions. We suppose that users move from the center of the cell to the edge and vice versa. We still assume that there is no fast fading.

II.3.1.3 Mobility model

We consider the same mobility model as in Section a). The probability that a user is in region i then given by:

$$q_i \propto \prod_{j=1}^{i-1} \frac{\nu_{j,j+1}}{\nu_{j+1,j}}.$$

Based on the mobility rates, the stationary distribution, $\pi(\boldsymbol{x})$, of the Markov process is the solution of the following balance equations:

$$\sum_i (\lambda_i + \mu_i(\boldsymbol{x}) + \nu_{i,i+1} + \nu_{i,i-1}) \pi(\boldsymbol{x}) =$$
$$\sum_i \Big(\lambda_i \pi(\boldsymbol{x} - \boldsymbol{e}_i) + \nu_{i,i+1} \pi(\boldsymbol{x}^+) + \mu_i(\boldsymbol{x} + \boldsymbol{e}_i) + \nu_{i,i-1} \pi(\boldsymbol{x}^-) \Big),$$

where $\boldsymbol{x}^+ = \boldsymbol{x} + \boldsymbol{e}_i - \boldsymbol{e}_{i+1}$ and $\boldsymbol{x}^- = \boldsymbol{x} - \boldsymbol{e}_i + \boldsymbol{e}_{i+1}$.

II.3.1.4 Stability condition

The stability condition follows from the limiting regime of infinite mobility where $\nu_{i,i+1}, \nu_{i+1,i} \to \infty$ and is given by:

$$\lambda/\bar{\mu} < 1$$

that is

$$\rho < \bar{\mu} \sum_i \frac{p_i}{R_i} \nu_{\min} T, \tag{II.38}$$

where $\bar{\mu}$ is the mean service rate at high load. We shall see that this service rate depends only on the scheduling strategy and is independent from the mobility rate in the presence of mobility.

a) round-robin policy

The mean service rate at high load under the round robin strategy is given by:

$$\bar{\mu} = \sum_i q_i \frac{R_i}{\nu_{\min} T}. \tag{II.39}$$

b) max C/I policy

Under the max C/I policy all mobile users are served in the first region (that of the best physical rate R_1) at high load. It follows that:

$$\bar{\mu} = \frac{R_1}{\nu_{\min} T}. \tag{II.40}$$

c) max-min policy

Under the max-min policy, the mean service rate at high load follows from (II.37):

$$\bar{\mu} = \frac{1}{\nu_{\min} T} \sum_i \frac{q_i}{\sum_j q_j / R_j}. \tag{II.41}$$

Observe that in the presence of mobility the less restrictive stability condition is obtained under the max C/I policy. However in the absence of mobility this strategy engender the most restrictive stability condition compared to more fair allocation strategies (round robin, max-min).

II.3.1.5 Performance in light traffic

The performance in light traffic (that is, when $\rho \to 0$) is independent of the scheduling policy. Indeed, a user when alone in the system is always allocated all radio resources. As Eq. (II.25), the video bit rate in region i is $c_i = \max(\min(R_i, \nu_{\max}), \nu_{\min})$, that is $\nu_{\max}$ in all regions i where $R_i > \nu_{\max}$. The mean buffer surplus in region i is then:

$$b_i = \frac{R_i - c_i}{c_i}.$$

Thus the mean buffer surplus in the cell in light traffic is given by:

$$\bar{B} = \sum_i p_i' b_i.$$

When $\nu_i \to 0$, $\forall i \leq N$ (no mobility) the mean buffer surplus in the cell in light traffic can be written as:

$$\bar{B} = \sum_i \frac{p_i/\mu_i}{\sum_j p_j/\mu_j} b_i, \text{ where } \mu_i = \frac{R_i}{c_i T}. \tag{II.42}$$

However, in the limiting regime of infinite mobility $\nu_i \to \infty$ ($\forall i \leq N$) the cell mean buffer surplus is given by:

$$\bar{B} = \sum_i q_i b_i.$$

For two regions for instance, explicit expressions of the cell buffer surplus in light traffic as a function of the mobility rates can be written as:

$$\bar{B} = \frac{(\nu_2 + p_1\mu_2)b_1 + (\nu_1 + p_2\mu_1)b_2}{\nu_1 + \nu_2 + p_1\mu_2 + p_2\mu_1}. \tag{II.43}$$

II.3.2 Numerical results

II.3.2.1 Case without mobility

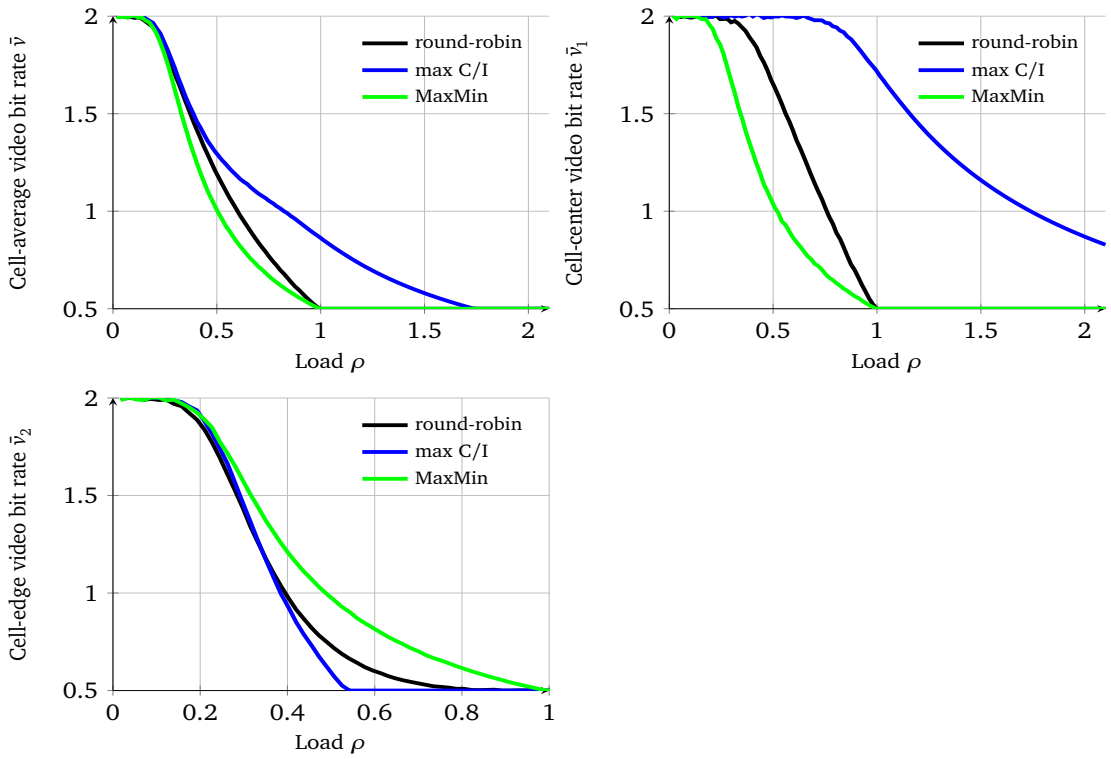


Figure II.23: Video bit rate performance for two regions without mobility.

Under the consideration of two capacity regions standing for cell-center and cell-edge, the simulation results in the absence of mobility are shown in Figure II.23 and Figure II.24 which compare the video bit rate and the buffer surplus performance of different policies for two regions, with

$p_1 = p_2 = 1/2$, $R_1 = 25$ Mb/s, $R_2 = 10$ Mb/s, $v_{\min} = 0.5$ Mb/s, $v_{\max} = 2$ Mb/s and $T = 10$ s. These results are obtained by the numerical evaluation of the stationary distribution of the Markov process $X(t)$

Fig. II.23 shows that under the max C/I policy, we can obtain better cell-average video bit rate compared to the other policies. Observe that the cell-average video bit rate is simply $\bar{v} = p_1\bar{v}_1 + p_2\bar{v}_2$ in accordance with (II.29). However, in terms of stability condition, we observe that max C/I policy provides lower stability condition, $\lambda_{\max}$, than other two policies. Based on Eq. (II.32) and (II.33), we obtain $\lambda_{\max,\mathrm{RR}} = 2.85$, and $\lambda_{\max,\max\ \mathrm{C/I}} = 0.54 \times \lambda_{\max,\mathrm{RR}}$. It is shown that the analytic calculation give the same results as we obtain in simulation, where $\lambda_{\max,\mathrm{RR}}$ and $\lambda_{\max,\max\ \mathrm{C/I}}$ are the minimum ρ that makes $\bar{v}_c$ or $\bar{v}_e$ equal to $v_{\min}$ for respective scheduling policy.

Moreover, in Fig. II.24, bad buffer surplus shows that the starvation event will strongly happen for the cell-edge users under the max C/I policy. In [10], it is concluded that there is no difference to deploy either round-robin or max C/I policy in the absence of mobility. However, for adaptive streaming services, it is showed that a trade-off exists between the two performance metrics and max C/I policy degrades system stability conditions. With little improvement of mean video bit rate and large degradation of system stability, operators are suggested to deploy round-robin policy for adaptive streaming in the static case.

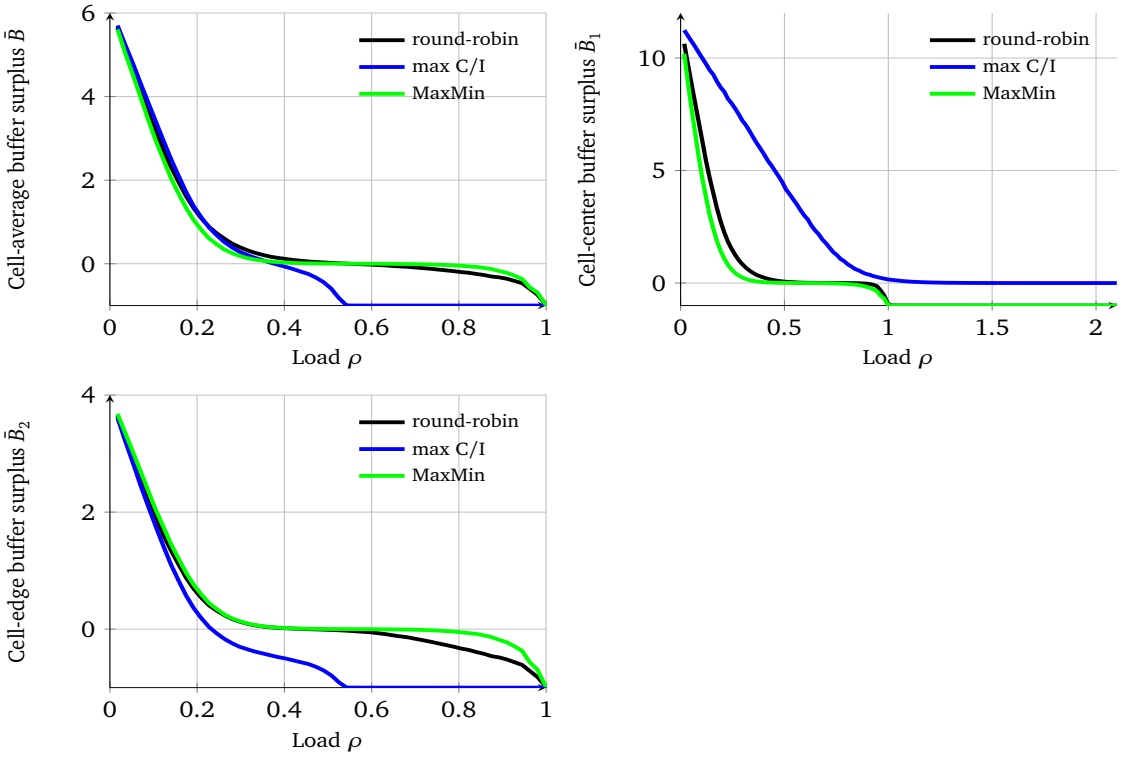


Figure II.24: Buffer surplus performance for two regions without mobility.

II.3.2.2 Case with mobility

Consider two capacity regions standing for cell-center and cell-edge with the same traffic and system configurations as before. We suppose that mobility rates are symmetric, that is $\nu_1 = \nu_2 = \nu$. We

consider the static case $v = 0$ and a case with a mobility rate $v = 1$. The results shown in Figure II.25 and II.26 are obtained by the numerical evaluation of the stationary distribution $\pi(\boldsymbol{x})$ of the Markov process $X(t)$.

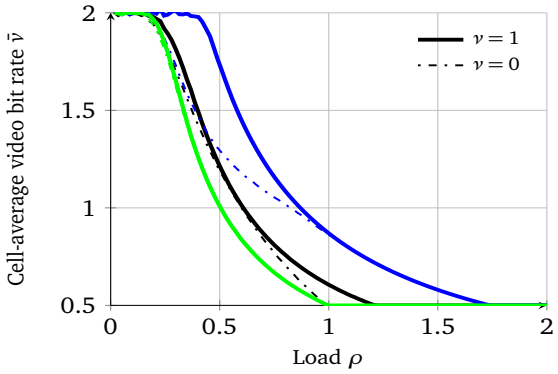


Figure II.25: Mean video bit rate with and without mobility, under round-robin (black), max C/I (blue) and max-min (green).

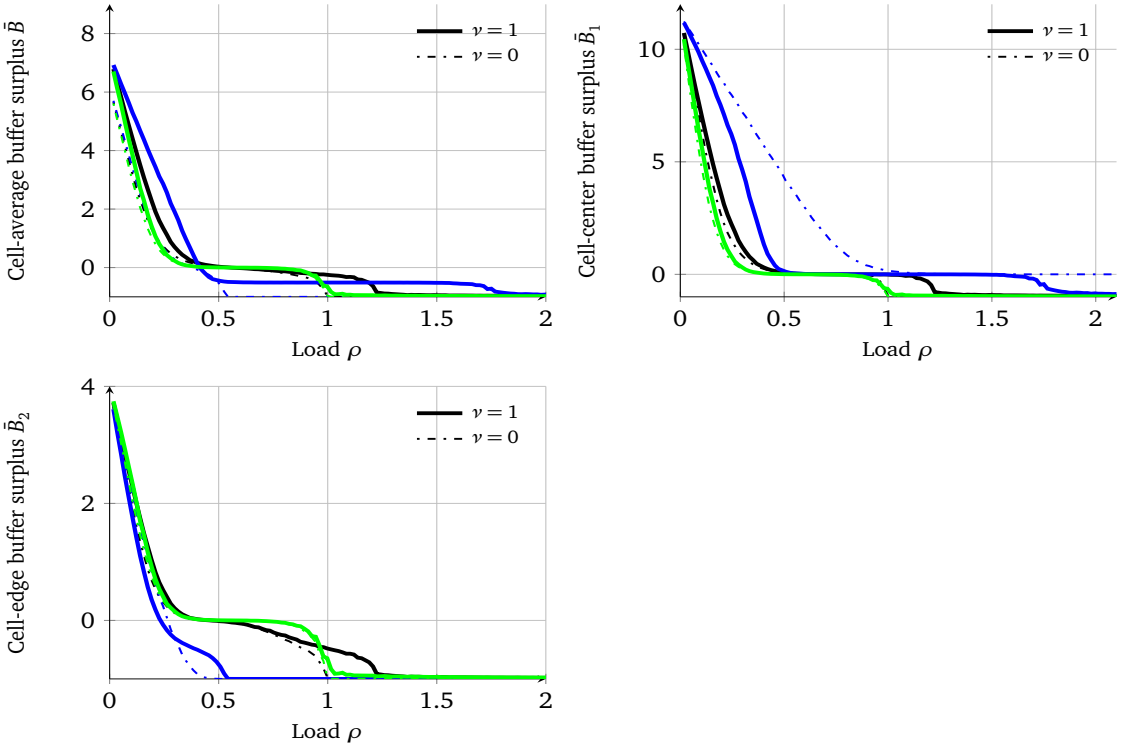


Figure II.26: Buffer surplus for two regions with and without mobility, under round-robin (black), max C/I (blue) and max-min (green).

For the video bit rate, we can only obtain the mean one all over the cell, because when users' mobility is considered, from the flow-level model, no flows belong to only one class. In addition, it

can be shown that max C/I policy provides better mean video bit rate and higher system stability. On the other hand, for the mean buffer surplus, max C/I policy performs better at low load and also at the high load and round-robin performs better at medium load. Observe that the maximum load when users are mobile is approximately equal to 1.2 under the round-robin, 1.75 under the max C/I and 1 under max-min in accordance with (II.38), (II.39), (II.40) and (II.41). The cell-average buffer surplus in light traffic in the absence of mobility is approximately equal to 6 in accordance with (II.42), while it is around 7.3 when users are mobile in accordance with (II.43).

Generally speaking, mobility provides more opportunities for users to exploit diversity gain. However, high video bit rate may degrade the buffer surplus.

II.3.3 Discussions and discriminatory scheduling

In this section, considering the conclusion obtained in the first part of this chapter, we summarize the suggested scheduling policy for different services in Table. II.2. In the static case, all scheduling policies have the same performance for elastic traffic. However, for adaptive streaming round-robin policy is suggested as the max C/I degrades the stability condition. In the presence of mobility, max C/I is recommended for the elastic data. However, for the adaptive streaming suggested policy depends on the desired optimized performance metric.

Services / Scenario	Elastic	Adaptive Streaming
Static	Same for all policies	RR: Better $\bar{B}$, stability condition max C/I: Minor improve $\bar{\nu}$ RR is suggested
Mobile	max C/I	RR: Better $\bar{B}$ at medium load max C/I: Better $\bar{\nu}$, stability condition max C/I: Better $\bar{B}$ at low and high load Depend on the needs

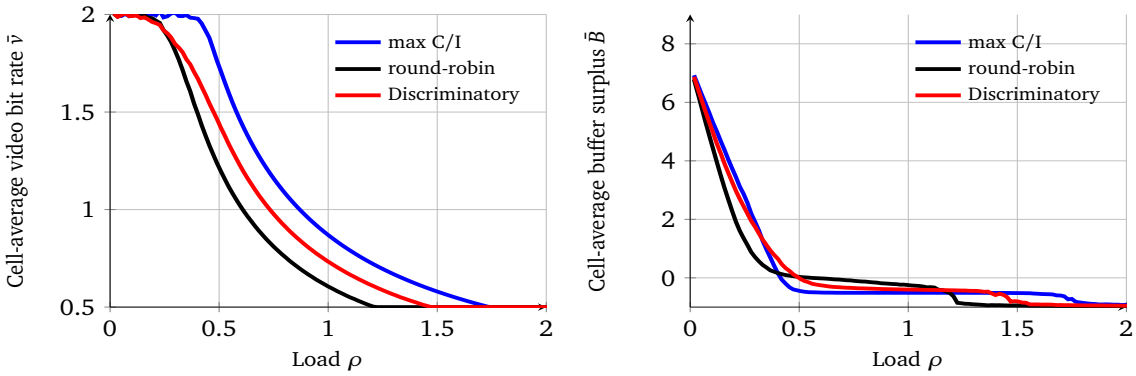
Table II.2: Policies recommended for different cases and services.

In the case of mobile users, there is a trade-off to deploy either round-robin or max C/I policy. Therefore, we examine the performance of discriminatory scheduler [13] to provide some intermediate results between the two scheduling policies. The resource allocation of users in region i under the discriminatory scheduler is calculated as follows:

$$\phi_i(\boldsymbol{x}) = \frac{w_i x_i}{\sum_k w_k x_k} R_i, \tag{II.44}$$

where w_i is the weight value of users in region i.

Fig. II.27a and II.27b shows the simulation results using the same system configuration as previous where users are all mobile. Moreover weighting value are configured as $(w_1, w_2) = (1000, 1)$. Simulation results give an intermediate performance between the round-robin, where $(w_1, w_2) = (1, 1)$ and the max C/I policy, where $(w_1, w_2) = (\infty, 1)$.



(a) Mean video bit rate

(b) Mean buffer surplus

Figure II.27: Performance in the presence of mobility under round-robin (black), max C/I (blue) and discriminatory policy (red).

II.4 Conclusion

We have studied in this chapter the impact of mobility in cellular data networks. We have presented a queuing model accounting for multiple mobility patterns. We have shown that the higher the mobility the lower should be the fairness of the scheduling policy. Based on this observation, we have proposed a mobility-aware scheduler where the fairness depends on the mobility of the users and that deprioritizes mobile users at the edge since they are likely to move and to be served in better radio conditions. We have shown that this scheduler improves the overall performance, through both analysis and system-level simulations. It is important to point out that the gain brought by this scheduling strategy is by exploiting only information related to the speed without any trajectory information. The gains would certainly be higher with this information, a point that we let for future work.

We have also examined the performance under different scheduling strategies in the presence of adaptive streaming traffic, through the analysis of flow level traffic models. We have considered two performance indicators, mean video bit rate and mean buffer surplus which are calculated based on the flow-level dynamics. We show that although the max C/I policy pushes users to select higher video bit rate. Thus, when users are mobile, the max C/I policy provides better mean video bit rate, but may decrease the mean buffer surplus. Therefore, we propose a discriminatory scheduler that allows to obtain an intermediate results to balance the two performance metrics.

Other interesting problems for future research include scheduling policies for mixed service and several mobility behaviors and the context of heterogeneous networks consisting of both picocells and macro cells.

Chapter III

Joint Processing CoMP

III.1 Introduction

Inter-cell interference is one of the key challenges faced in mobile communication systems. It restricts the re-usability of the radio resource and limits spectral efficiency. Since the days of GSM, various techniques are used to cope with interference. Coordinated MultiPoint (CoMP) [7, 37], a main feature on the LTE-A roadmap, is mentioned as a promising approach to mitigate its effects through the coordination of multiple cells.

CoMP allows a group of eNodeBs to cooperate in order to improve the coverage, cell-edge throughput, and system efficiency. These eNodeBs, referred to as cooperating eNodeBs, communicate together through the backhaul network. In the downlink, CoMP schemes range from simple coordinated scheduling to more challenging joint processing approaches.

This chapter focuses on Joint Processing (JP) scheme in the downlink, where coordination is performed either by converting an interference signal into a meaningful signal (Joint Transmission JT) or by muting an interfering cell so that it does not incur serious interference to the served UE (Dynamic Point Blanking DPB) [59], which is often employed in conjunction with DPS (Dynamic Point Selection) [32].

At first glance, the concept of joint processing CoMP may seem always advantageous for cell-edge users [2]. Performance benefits of both downlink and uplink CoMP have been identified in both homogeneous and heterogeneous network deployments [6, 68, 49, 64] and practical implementation have been carried out in [37, 38] showing CoMP as a viable technology while specification support to efficiently realize the benefits of cooperative transmission has been discussed in [50].

Unfortunately, in a scenario with moderate coordination gain (that is the mean throughput gain brought by the cooperation of a cell), JP can be detrimental even for cell-edge users due to the inefficient utilization of radio resources. The case of low interference environment has been studied in [43]. Thus, there is a tradeoff between the performance of cell-edge users and the ability of the network to process all traffic, especially in a highly loaded network. It has been shown in [18, 17] that UE throughput enhancement comes at the expense of increased traffic in mobile backhaul networks. In fact, implementation reveals many practical challenges as CoMP presents a number of technical issues in terms of backhaul characteristics, synchronization and feedback design [49]. If BSs taking part in the cooperation are incapable of satisfying these backhaul and synchronization requirements, performance degradation may be expected. The literature shows no consensus on whether or not it is interesting to deploy such a feature.

The efficiency of CoMP schemes depends critically on the scheduling strategy. It is proposed in [3, 4] to allocate a dedicated frequency band to cell-edge users in order to perform CoMP operations. In [53], the authors propose a joint proportional fairness scheduling algorithm that treats cell-center and cell-edge users equally without any frequency band partitioning. The scheduling algorithm may also depend on the clustering technique, the feedback characteristics and the channel reliability [57]. Several clustering approaches have been studied [56, 36], from static clusters with centralized scheduling [28] to dynamic clusters with distributed scheduling [77]. The impact of backhaul limitations on CoMP clustering has been discussed in [17]. Feedback signaling design and associated achievable system-level performance have been addressed in [55].

Most of the studies are carried out around JT, which is considered as the most promising JP technique and thus little attention has been paid to DPB [73, 52], which seems to be a much simpler and a performant technique compared to JT. DPB technique is less evaluated than JT and a complete comparative study between these two JP techniques cannot be found in the literature. Moreover, full buffer traffic models are commonly used in order to evaluate performance of these techniques. Those models are typically easier to simulate. However, the practical non-full buffer traffic models are very important especially when dealing with cooperation techniques and can lead to entirely different system behavior and performance results.

The literature showing no consensus on the potential benefits or drawbacks of JP CoMP techniques, we try in this work to figure out which scenarios are interesting to deploy JP. We consider both high interference scenario (e.g., network without beamforming) and low interference scenario (e.g., network enabling beamforming) with different clustering methods and different transmission schemes (one-multi stream), under different scheduling strategies. We analyze the capacity issues of this coordination scheme, and provide a complete comparative performance evaluation of JT and DPB considering non-full buffer traffic model. We show that the performance when applying such techniques depends primarily on the coordination gain, that is the mean throughput gain brought by the cooperation of a cell to a user.

It turns out that JT could be interesting in a high interference scenario, particularly at medium load, when different streams are sent from different cooperating cells, but at the cost of higher complexity (multiplexing and advanced receiver techniques). In addition, in order to maintain stability in this case, a cell in a symmetric network topology cooperates only when it brings at least 100% mean throughput gain when cooperating. JT is usually enabled to such cell-edge users that experience an average signal level difference between serving point and strongest interferer of less than a predefined threshold. We propose in this work a new algorithm for triggering JT CoMP based on the user throughput gain. The proposed algorithm is evaluated through flow level traffic models as well as system level simulations. We study different CoMP-specific scheduling strategies and we show the important impact of the CoMP-specific scheduling strategy on the diversity gain and on the global performance.

While steering beams towards users tends to be the most efficient solution to reduce interference and improve performance of cell-edge users, we show that DPB is a simpler and an efficient technique, especially at highe loads, in a high interference scenario where only a limited number of antennas is deployed and where interference management is crucial to ensure proper QoS.

In addition, it was common in previous works to assume *static* or *semi static* users in the performance evaluation of CoMP schemes. Mobility is generally thought as improving throughput performance in the presence of elastic traffic, see for instance [22, 27, 26]. However, we show in this chapter that mobility has a critical impact on the performance of CoMP schemes in the particular low interference scenario, such as a beamforming system. Indeed, in the presence of coordination mechanisms mobility can lead to bad performance if the scheduling strategy is not well chosen. Deprioritizing CoMP users in the presence of mobility outperforms many other strategies in sharp contrast to the

scenario without mobility: this gives the chance to mobile CoMP users to move and to be served in good radio conditions without performing coordination.

Motivated by the above observation, we propose in this chapter a mobility-aware scheduler that exploits the mobility as an additional information in order to schedule users with elastic traffic on the downlink of cellular data networks. This may require an estimation of the mobile speed as done in [61]. Indeed, coordination is performed primarily for users staying at the cell edge, without mobility. Other cell-edge users are likely to move and to be served in better radio conditions where cell coordination is not required. Thus, the extra resource consumption incurred by the joint transmission from several base stations can be avoided for these users. We start by studying the limiting cases where users are either static or mobile. Then we consider the case of several mobility behaviors and we compare the performance of the proposed scheduler to other usual scheduling policies through the analysis of flow-level traffic models. We show that this scheduler improves the global performance.

This chapter is divided into four main sections. In Section III.2, we present flow level models considering different scenarios. In Section III.3, we study the performance and we analyze the issues of JT CoMP. In Section III.4, we provide a detailed study of the proposed JT triggering algorithm. The performance of DPB is evaluated in Section III.5.

III.2 Flow level model

We consider in this section different scenarios. We model mainly the scenario of several cells with one common coordination zone (overlapping area), as well as the case of a static cluster constituted of three cooperating cells, which may represents the case of intra-site coordination for instance. For each scenario, we first present the model without mobility then we describe the model with mobility. Then, in order to study the stability condition in the general case, we introduce the scenario of one cell cooperating with all its surrounding neighboring cells (6 neighboring cells in the case of a hexagonal network topology).

III.2.1 Single coordination zone

III.2.1.1 Model without mobility

a) Cellular network

We consider a cluster composed of K coordinated cells, where each cell is modeled by two main *zones*: a non-coordination zone and a coordination zone, which is part of the global coordination zone of the cluster. The non-coordination zones $k = 1, ..., K$, as illustrated by Figure III.1, are those where users are typically close to the cell center, experience good radio conditions and thus served only by their own serving base station. The coordination zone ($k = 0$) is that where users are close to the cell edge, suffer from bad radio conditions and receive signals from the K coordinated cells. This corresponds to the JT scheme. In fact, the model applies to the DPB scheme as well ; only the transmission rates achieved by the coordination depend on the JP technique, either JT or DPB.

We model each zone k by a set of N_k *regions*. In each region, radio conditions are supposed to be homogeneous and thus users are served at the same physical data rate on the downlink. We model each region by a queue with a specific service rate corresponding to the physical data rate in this region. Consequently, the considered cluster can be viewed as a set of $\sum_{k=0}^{K} N_k$ queues with K coupled processors. The simple case of two coordinated cells and one region in each zone is illustrated by Figure III.1.

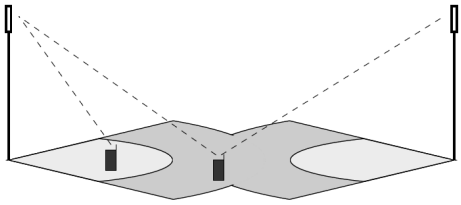


Figure III.1: A simple model with two coordinated cells.

b) Traffic model

We consider elastic traffic only, corresponding to data transfers. Specifically, we assume that new data flows are generated in region i of zone k at the random times of a Poisson process of intensity λ_{ik}. We denote by $\lambda_k = \sum_i \lambda_{ik}$ the total flow arrival rate in zone k and by $\lambda = \sum_k \lambda_k$ the total flow arrival rate in the coordination cluster. The probability that a new data flow is generated by a user in zone k is $p_k = \lambda_k/\lambda$ while the probability that a new flow in zone k is generated by a user in region i is $p_{ik} = \lambda_{ik}/\lambda_k$. Physically, p_k corresponds to the relative area of zone k compared to the area of the cluster, while p_{ik} represents the relative area of region i of zone k compared to the area of zone k, assuming uniform traffic distribution in the cell.

We use a flow-level model where each data flow is viewed as a fluid of random volume to be transmitted. The volumes have an exponential distribution with mean σ (in bits). When region i of zone k is served, flows are completed at rate μ_{ik} in the absence of fast fading, corresponding to the physical rate $\mu_{ik}\sigma$ (in bit/s).

We denote by $X_{ik}(t)$ the number of active flows in region i of zone k at time t and by $X_k(t) = \sum_i X_{ik}(t)$ the total number of active flows in zone k. The vector $X(t) = (X_{ik}(t))_{ik}$ defines a Markov process of dimension $\sum_k N_k$. The load of region i in zone k is $\rho_{ik} = \lambda_{ik}/\mu_{ik}$. The total load of zone k is given by $\rho_k = \sum_i \rho_{ik} = \lambda_k/\mu_k$, where μ_k is the weighted harmonic mean service rate:

$$\mu_k = \frac{1}{\sum_i p_{ik}/\mu_{ik}}.$$

c) Throughput metrics

We measure performance in terms of mean throughputs in the different zones, in the same way we proceed in §II.2.2.1c). In any state x such that $x_{ik} > 0$, each user in region i of zone k has throughput $\mu_{ik}\sigma\phi_k(x)/x_k$. Now the distribution seen by users in region i of zone k is the size-biased distribution:

$$\pi_{ik}(x) \propto x_{ik}\pi(x).$$

We denote by E_{ik} the corresponding expectation. The mean throughput of users in region i of zone k is then given by

$$\gamma_{ik} = E_{ik}\left(\frac{\mu_{ik}\sigma\phi_k(X)}{X_k}\right) = \frac{E(\mu_{ik}\sigma\phi_k(X)\frac{X_{ik}}{X_k})}{E(X_{ik})}. \tag{III.1}$$

By the traffic conservation equation

$$\lambda_{ik} = E\left(\mu_{ik}\phi_k(X)\frac{X_{ik}}{X_k}\right), \tag{III.2}$$

we deduce

$$\gamma_{ik} = \frac{\lambda_{ik}\sigma}{E(X_{ik})}. \tag{III.3}$$

This is the ratio of the traffic intensity in region i of zone k to the mean number of data flows in this region. Observe that, by Little's law, this is also the ratio of mean flow size σ to mean flow duration in region i of zone k.

By a similar argument, we obtain the mean throughput in zone k,

$$\gamma_k = \frac{\lambda_k \sigma}{E(X_k)}, \tag{III.4}$$

and the mean throughput in the cluster,

$$\gamma = \frac{\lambda \sigma}{E(\sum_k X_k)}. \tag{III.5}$$

This is the harmonic mean of the mean throughputs γ_k weighted by p_k, for $k = 0, 1, \ldots, K$.

III.2.1.2 Model with mobility

In this section, we add mobility to the previous model. We use the same mobility model as in Section II.2.2.2. Users move from a non-coordination zone of a cell to a non-coordination zone of another cell through the coordination zone, as shown in Figure III.2 for two coordinated cells.

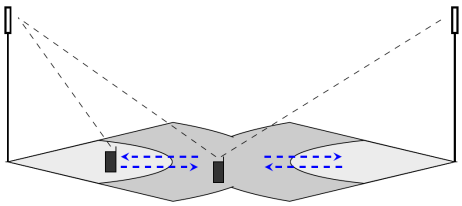


Figure III.2: A simple model with two coordinated cells and mobility.

a) Mobility model

i. Inter-zone mobility We assume that each user in region N_k of non-coordination zone k moves to region N_0 of the coordination zone and vice versa after exponential durations, at respective rates $\nu_{k,0}$ and $\nu_{0,k}$. In state x, the total mobility rate from non-coordination zone k to the coordination zone 0 and from coordination zone 0 to non-coordination zone k are equal to $x_{N_k k}\ \nu_{k,0}$ and $x_{N_0 0}\ \nu_{0,k}$, respectively.

ii. Intra-zone mobility We assume that in each zone k, each user in region i moves to region $i-1$ (for $i > 1$) and to region $i+1$ (for $i < N_k$) after exponential durations, at respective rates $\nu^{(k)}_{i,i+1}$ and $\nu^{(k)}_{i,i-1}$. In state x, the total mobility rate from region i to region $i+1$ and from region $i+1$ to region i (for $i < N_k$) in zone k are $x_{ik}\ \nu^{(k)}_{i,i+1}$ and $x_{i+1\,k}\ \nu^{(k)}_{i+1,i}$, respectively. The probability that a user is in region i of zone k (for $k > 0$) then satisfies

$$q_{ik} \propto \frac{\nu_{0,k}}{\nu_{k,0}} \prod_{j=i}^{N_k-1} \frac{\nu^{(k)}_{j+1,j}}{\nu^{(k)}_{j,j+1}},$$

while the probability that a user is in region i of zone $k = 0$ satisfies

$$q_{i0} \propto \prod_{j=i}^{N_0-1} \frac{\nu_{j+1,j}^{(0)}}{\nu_{j,j+1}^{(0)}}.$$

Note that

$$\sum_{k=0}^{K} \sum_{i=1}^{N_k} q_{ik} = 1.$$

The probability that a user is in zone k is then given by:

$$q_k = \sum_{i=1}^{N_k} q_{ik}.$$

It is worth noting that, for mobile users, the traffic conservation equation (III.2) no longer applies: the traffic arriving in region i of zone k is not equal in general to the traffic served in region i of zone k. Thus the mean throughput in region i of zone k is given by (III.1) but not by (III.3). Similarly, the mean throughput in zone k is given by

$$\gamma_k = \frac{E\left(\sum_i \mu_{ik} \sigma \phi_k(X) \frac{X_{ik}}{X_k}\right)}{E(X_k)}, \tag{III.6}$$

but not by (III.4). Now the overall traffic conservation equation in the cluster still applies so that the mean throughput in the cluster is still given by (III.5). It is still the arithmetic mean of the mean throughputs γ_k weighted by the probability p'_k that an *active* user is in zone k, given by

$$p'_k = \frac{E(X_k)}{\sum_{j=0}^{K} E(X_j)}$$

for all $k = 0, 1, \ldots, K$.

III.2.2 Intra-site coordination

III.2.2.1 Model without mobility

a) Cellular network

In order to study the intra-site joint transmission technique, we consider a site constituted of $K = 3$ sectors, where each sector $k = 1, 2, 3$ is modeled by two main *zones*: a non-coordination zone and a coordination zone. We refer by zone $k, k+1$ (with modulo K notation) to the coordination zone between sectors k and $k+1$. The coordination zones $k, k+1$ (hatched area), as illustrated by Figure III.3, are those where the difference between the signals received from sectors k and $k+1$ does not exceed a given threshold δP, involving both sectors in the transmission. We refer by zone k to the non-coordination zone (area without hatching) where only sector k is involved in the transmission.

We model each zone k and zone $k, k+1$ by a set of N_k and $N_{k\,k+1}$ *regions* respectively. In each region, radio conditions are supposed to be homogeneous and thus users are served at the same physical data rate on the downlink. We model each region by a queue with a specific service rate corresponding to the physical data rate in this region. The considered site shown in Figure III.3 can be viewed as a set of $\sum_{k=1}^{K} (N_k + N_{k\,k+1})$ queues with K coupled processors.

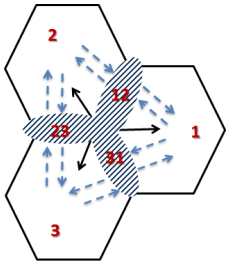


Figure III.3: A tri-sector site with coordination areas.

b) Traffic model

We consider elastic traffic only and we assume that new data flows are generated in region i of zone k and region i of zone $k,k+1$ at the random times of a Poisson process of intensity $\lambda_k^{(i)}$ and $\lambda_{k,k+1}^{(i)}$ respectively. We denote by $\lambda_k = \sum_i \lambda_k^{(i)}$ and $\lambda_{k,k+1} = \sum_i \lambda_{k,k+1}^{(i)}$ the total flow arrival rates in zone k and zone $k,k+1$ respectively. Let $\lambda = \sum_k \left(\lambda_k + \lambda_{k,k+1}\right)$ be the total flow arrival rate in the site. Thus, the probability that a new data flow is generated by a user in zone k is $p_k = \lambda_k/\lambda$ while the probability that a new data flow in zone k is generated by a user in region i is $p_k^{(i)} = \lambda_k^{(i)}/\lambda_k$. Similarly, the probability that a new data flow is generated by a user in zone $k,k+1$ is $p_{k,k+1} = \lambda_{k,k+1}/\lambda$ while the probability that a new data flow in zone $k,k+1$ is generated by a user in region i is $p_{k,k+1}^{(i)} = \lambda_{k,k+1}^{(i)}/\lambda_{k,k+1}$. Each data flow is viewed as a fluid of random volume of exponential distribution with mean σ (in bits) to be transmitted. When region i of zone k is served, flows are completed at rate $\mu_k^{(i)}$ in the absence of fast fading, corresponding to the physical rate $\mu_k^{(i)}\sigma$ (in bit/s). The service rate in region i of zone $k,k+1$ is $\mu_{k,k+1}^{(i)}$.

We denote by $X_k(t) = \sum_i X_k^{(i)}(t)$ the total number of active flows in zone k at time t and by $X_{k,k+1}(t) = \sum_i X_{k,k+1}^{(i)}(t)$ the total number of active flows in zone $k,k+1$, where $X_k^{(i)}(t)$ and $X_{k,k+1}^{(i)}(t)$ are the total numbers of active flows in region i of zone k and region i of zone $k,k+1$ respectively. The vector $X(t) = \left(X_k^{(i)}(t), X_{k,k+1}^{(j)}(t)\right)_{(k,i,j)}$ defines a Markov process of dimension $\sum_k N_k + N_{k,k+1}$. The load of region i in zone k is $\rho_k^{(i)} = \lambda_k^{(i)}/\mu_k^{(i)}$, while the load of region i in zone $k,k+1$ is $\rho_{k,k+1}^{(i)} = \lambda_{k,k+1}^{(i)}/\mu_{k,k+1}^{(i)}$. The total load of zone k is given by $\rho_k = \sum_i \rho_k^{(i)} = \lambda_k/\mu_k$ and that of zone $k,k+1$ is given by $\rho_{k,k+1} = \sum_i \rho_{k,k+1}^{(i)} = \lambda_{k,k+1}/\mu_{k,k+1}$, where μ_k and $\mu_{k,k+1}$ are the weighted harmonic mean service rates:

$$\mu_k = \frac{1}{\sum_i p_k^{(i)}/\mu_k^{(i)}}$$

and

$$\mu_{k,k+1} = \frac{1}{\sum_i p_{k,k+1}^{(i)}/\mu_{k,k+1}^{(i)}}.$$

c) Throughput metrics

Similarly as in §II.2.2.1c), we measure performance in terms of mean throughputs in the different zones. In any state x such that $x_k^{(i)} > 0$, each user in region i of zone k has throughput $\mu_k^{(i)}\sigma\phi_k(x)/x_k$. We denote by $E_k^{(i)}$ and $E_{k,k+1}^{(i)}$ the expectations corresponding to the following size-biased distributions:

$$\pi_k^{(i)}(x) \propto x_k^{(i)} \pi(x)$$

and

$$\pi_{k,k+1}^{(i)} \propto x_{k,k+1}^{(i)} \pi(x).$$

The mean throughput of users in zone k is then given by

$$\gamma_k = E\left(\sum_i \mu_k^{(i)} \sigma \phi_k(X) \frac{X_k^{(i)}}{X_k}\right) / E(X_k). \tag{III.7}$$

This is the arithmetic mean of the mean throughputs $\gamma_k^{(i)}$ weighted by the probabilities $p'^{(i)}_k$, where

$$p'^{(i)}_k = E(X_k^{(i)})/E(X_k).$$

Similarly, the mean throughput in zone $k, k+1$ is given by

$$\gamma_{k,k+1} = E\left(\sum_i \mu_{k,k+1}^{(i)} \sigma \phi_{k,k+1}(X) \frac{X_{k,k+1}^{(i)}}{X_{k,k+1}}\right) / E(X_{k,k+1}). \tag{III.8}$$

The mean throughput in the site is the arithmetic mean of the mean throughputs γ_k (III.7) and $\gamma_{k,k+1}$ (III.8) weighted by the probabilities p'_k and $p'_{k,k+1}$ respectively, given by

$$p'_k = \frac{E(X_k)}{\sum_{j=1}^{K} E(X_j) + E(X_{j,j+1})} \qquad \text{and} \qquad p'_{k,k+1} = \frac{E(X_{k,k+1})}{\sum_{j=1}^{K} E(X_j) + E(X_{j,j+1})}.$$

III.2.2.2 Model with mobility

a) Inter-zone mobility

In a general case we assume that each user in region i of non-coordination zone k moves to region n of a non-coordination zone j and vice versa after exponential durations, at respective rates $\nu_{k\to j}^{(i,n)}$ and $\nu_{j\to k}^{(n,i)}$. In state x, the total mobility rate from region i of zone k to region n of zone j and from region n of zone j to region i of zone k are equal to $x_k^{(i)}\, \nu_{k\to j}^{(i,n)}$ and $x_j^{(n)}\, \nu_{j\to k}^{(n,i)}$ respectively. Similarly, the total mobility rate from region i of a non-coordination zone k to region n of a coordination zone $j, j+1$ and from region n of zone $j, j+1$ to region i of zone k are equal to $x_k^{(i)}\, \nu_{k\to j,j+1}^{(i,n)}$ and $x_{j,j+1}^{(n)}\, \nu_{j,j+1\to k}^{(n,i)}$ respectively. We assume that users can also move from a coordination zone $k, k+1$ to another coordination zone $j, j+1$. Thus the total mobility rate is equal to $x_{k,k+1}^{(i)}\, \nu_{k,k+1\to j,j+1}^{(i,n)}$.

b) Intra-zone mobility

We assume that in each zone k, each user in region i moves to region $n \neq i$ (for i , $n \in \{1 \ldots N_k\}$) after exponential durations, at rate $\nu_k^{(i,n)}$. The total mobility rate from region i to region n is $x_k^{(i)}\, \nu_k^{(i,n)}$.

At high load, the mobility process can be viewed as a Markov process of $\sum_{k=1}^{K} \left(N_k + N_{k,k+1}\right)$ states, where each state represents a region of a given zone. The transitions between the different sates are equal to the corresponding mobility rates. The probability that a user is in region i of zone k is $q_k^{(i)}$

while the probability that a user is in region i of zone $k, k+1$ is $q^{(i)}_{k,k+1}$. These probabilities are given by the stationary distribution and follow from the following balance equations

$$\forall (k,i):$$

$$q^{(i)}_k \times \left(\sum_{\substack{n \leq N_k \\ n \neq i}} \nu^{(i,n)}_k + \sum_{\substack{n \leq N_j \\ j \neq k}} \nu^{(i,n)}_{k \to j} + \sum_{\substack{n \leq N_{j,j+1} \\ j \leq K}} \nu^{(i,n)}_{k \to j,j+1} \right)$$

$$= \sum_{\substack{n \leq N_k \\ n \neq i}} \nu^{(n,i)}_k q^{(n)}_k + \sum_{\substack{n \leq N_j \\ j \neq k}} \nu^{(n,i)}_{j \to k} q^{(n)}_j + \sum_{\substack{n \leq N_{j,j+1} \\ j \leq K}} \nu^{(n,i)}_{j,j+1 \to k} q^{(n)}_{j,j+1}$$

and

$$q^{(i)}_{k,k+1} \times \left(\sum_{\substack{n \leq N_{k,k+1} \\ n \neq i}} \nu^{(i,n)}_{k,k+1} + \sum_{\substack{n \leq N_j \\ j \leq K}} \nu^{(i,n)}_{k,k+1 \to j} + \sum_{\substack{n \leq N_{j,j+1} \\ j \neq k}} \nu^{(i,n)}_{k,k+1 \to j,j+1} \right)$$

$$= \sum_{\substack{n \leq N_{k,k+1} \\ n \neq i}} \nu^{(n,i)}_{k,k+1} q^{(n)}_{k,k+1} + \sum_{\substack{n \leq N_j \\ j \leq K}} \nu^{(n,i)}_{j \to k,k+1} q^{(n)}_j + \sum_{\substack{n \leq N_{j,j+1} \\ j \neq k}} \nu^{(n,i)}_{j,j+1 \to k,k+1} q^{(n)}_{j,j+1},$$

With

$$\sum_{k=1}^{K} \left(\sum_{i=1}^{N_k} q^{(i)}_k + \sum_{n=1}^{N_{k,k+1}} q^{(n)}_{k,k+1} \right) = 1.$$

The probability that a user is in zone k is then given by: $q_k = \sum_{i=1}^{N_k} q^{(i)}_k$, while the probability that a user is in zone $k, k+1$ is: $q_{k,k+1} = \sum_{i=1}^{N_{k,k+1}} q^{(i)}_{k,k+1}$. We denote by q_0 the probability that a user is in a coordination zone, that is: $q_0 = \sum_{k=1}^{K} q_{k,k+1}$.

III.2.2.3 Model with multiple mobility behaviors

We consider in this section the case of multiple mobility behaviors, so that users may be either static or mobile.

Multi-class extension

We have assumed so far that all users have the same mobility behavior. In particular, they are all static or all mobile. In order to evaluate more advanced schedulers that uses mobility as an additional information in order to schedule users, we extend the results to multiple classes of mobility, where each user belongs to a given class of mobility v, defined by a set of mobility rates

$$\left(\nu^{(i,n),v}_{k \to j,j+1}, \nu^{(n,i),v}_{j,j+1 \to k}, \nu^{(i,n),v}_{k \to j}, \nu^{(i,n),v}_{k,k+1 \to j,j+1}, \nu^{(i,n),v}_{k} \right)_{(k,j,i,n)}.$$

We consider V classes of mobility. For instance, each class represents a range of speeds (static, pedestrian, users in train, etc). We assume mobility classes are numbered in decreasing order of speed so that $v = 0$ represents the static class. We denote by w_v the probability that a new flow is initiated by a class-v user. The vector $X(t) = \left(X^{(i),v}_k(t), X^{(n),v}_{k,k+1}(t) \right)_{(k,i,n,v)}$ defines a Markov process of dimension $V \times \sum_k N_k + N_{k,k+1}$. Let $X_k(t) = \sum_v X^{(v)}_k(t)$ and $X_{k,k+1}(t) = \sum_v X^{(v)}_{k,k+1}(t)$, where $X^{(v)}_k(t) = \sum_{i \leq N_k} X^{(i),v}_k(t)$ and $X^{(v)}_{k,k+1}(t) = \sum_{i \leq N_{k,k+1}} X^{(i),v}_{k,k+1}(t)$ are the total numbers of class-v users in zone k and zone $k, k+1$ respectively.

III.2.3 A single cell cooperating with its surrounding neighboring cells

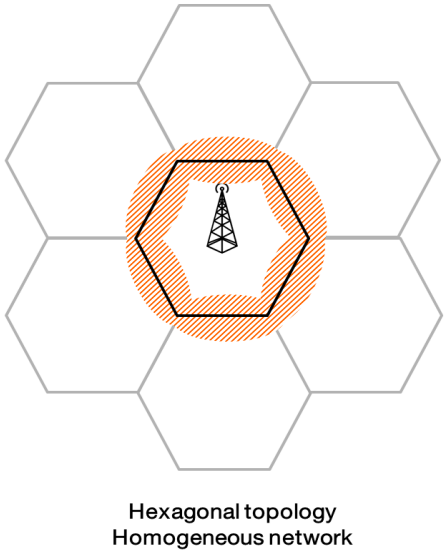


Figure III.4: One cell surrounded by 6 cells.

It has recently been shown that CoMP and more precisely JP CoMP scheme can be detrimental at high load, due to the degradation of the stability condition. In order to study the conditions that allow to avoid the stability condition degradation, we consider the general case of one cell surrounded by a given number of neighboring cells. Figure III.4 shows the case of a homogeneous network with hexagonal topology. We model each cell by two main zones: a non-coordination zone (cell center) and a coordination zone (cell edge). We consider the same traffic model as in previous sections. p_{center} is the probability that a new data flow is generated by a user in the cell center. We denote by $p_{\text{edge},n}$ the probability that a new data flow is generated by a user at the cell edge and involving n cells in the coordination and by $p'_{\text{edge},n}$ the probability that a new data flow is generated by a user associated with a neighboring cell, requiring the coordination of the considered cell and involving n cells in the coordination. μ_{center} is the mean service rate of a user in the cell center. $\mu_{\text{edge,n}}$ is the mean service rate of a user at the cell edge when served without any coordination, and that may involve n cells in the coordination when getting the CoMP status. We denote by λ the total arrival rate in the cell.

For the case when no coordination is performed between the different cells, the load of the considered cell under a round-robin scheduling policy can be written as follows:

$$\rho = \frac{p_{\text{center}}\lambda}{\mu_{\text{center}}} + \sum_n \frac{p_{\text{edge,n}}\lambda}{\mu_{\text{edge,n}}}, \tag{III.9}$$

in the static case according to (II.2). The stability condition is then nothing more than

$$\rho < 1.$$

Note that μ_{center} and $\mu_{\text{edge,n}}$ are the harmonic means of service rates in different radio conditions belonging to each zone (non-coordination zone and coordination zone) in the static case.

We shall see that the stability condition when the considered cell cooperates with its neighboring cells, that is when activating JP coordination scheme, depends on the technique (JT or DPB) where it can be more restrictive when activating JT compared to the case when activating DPB.

III.3 Joint transmission

We study in this section JT scheme, where all cooperating cells of a CoMP user are involved in the transmission and thus not available for other users. JT scheme has been demonstrated as an efficient approach to improve cell-edge user's throughput, especially in high interference environments, but this is at the cost of higher resource consumption. Thus there is a tradeoff between the performance of cell-edge users and the ability of the network to process all traffic.

III.3.1 Transmission schemes

In the absence of coordination, the SINR (Signal-to-Interference-plus-Noise Ratio) experienced by a user is:

$$SINR = \frac{P_s}{\sum_i P_i + \mathcal{N}}, \tag{III.10}$$

P_s is the received power from the serving cell, $\mathcal{N}$ is the thermal noise and $\sum_i P_i$ is the total interference seen by the user from all neighboring cells.
The resulting transmission rate depends on the perceived SINR and is given by:

$$r = \boldsymbol{F}\left(\frac{P_s}{\sum_i P_i + \mathcal{N}}\right). \tag{III.11}$$

We denote by $\boldsymbol{F}$ the mapping function between the SINR and the transmission rate, that is according to Shannon capacity of Gaussian channel:

$$r = W \log_2\left(1 + \frac{P_s}{\sum_i P_i + \mathcal{N}}\right),$$

where W is the channel bandwidth.

We consider two main transmission schemes in the presence of JT CoMP. In the first one, we assume that exactly the same data is scheduled from different cells to the CoMP user. This requires a perfect synchronization of the different coordinated cells. The perceived SINR is then given by:

$$SINR_s + \sum_{c\in\mathbb{C}} SINR_c = \frac{P_s}{\sum_{i\neq c\in\mathbb{C}} P_i + \mathcal{N}} + \sum_{c\in\mathbb{C}} \frac{P_c}{\sum_{i\neq c\in\mathbb{C}} P_i + \mathcal{N}} \tag{III.12}$$

where $\mathbb{C}$ is the set of cooperating cells of the considered user, which is usually determined according to (I.1). Note that the set $\mathbb{C}$ may change over time, especially when the user is moving. P_c is the received power from the cooperating cell c. $SINR_s$ is the serving cell SINR while $SINR_c$ is the SINR associated with cooperating cell c. Note that $\sum_{i\neq c\in\mathbb{C}} P_i$ represents the interference term, that is the total interference seen from all neighboring cells except the cooperating cells. The transmission rate of the CoMP user r' depends on the perceived SINR, and can be written as follows:

Scheme 1

$$r' = \boldsymbol{F}\left(SINR_s + \sum_{c\in\mathbb{C}} SINR_c\right) = \boldsymbol{F}\left(\frac{P_s}{\sum_{i\neq c\in\mathbb{C}} P_i + \mathcal{N}} + \sum_{c\in\mathbb{C}} \frac{P_c}{\sum_{i\neq c\in\mathbb{C}} P_i + \mathcal{N}}\right) \tag{III.13}$$

In the second scheme, we assume that streams with different content are sent from different cells. The UE is capable of an advanced receiver which is able to eliminate the inter-stream interference. Practically, inter-stream interference can not be completely eliminated. However, in this work we

consider the ideal case where the user receives independent streams in order to get an upper bound performance. The transmission rate of the CoMP user is then given by:

Scheme 2

$$r' = \boldsymbol{F}(SINR_s) + \sum_{c\in\mathbb{C}} \boldsymbol{F}(SINR_c) = \boldsymbol{F}\left(\frac{P_s}{\sum_{i\neq c\in\mathbb{C}} P_i + \mathcal{N}}\right) + \sum_{c\in\mathbb{C}} \boldsymbol{F}\left(\frac{P_c}{\sum_{i\neq c\in\mathbb{C}} P_i + \mathcal{N}}\right) \quad \text{(III.14)}$$

This scheme is similar to the so-called "Multiflow" schemes explained in the 3GPP technical report on HSDPA multipoint transmission [5], where the UE is supposed to be capable of a Type 3i receiver per cell. These schemes assume that the application level data is split in the access network thus scheduling different content from different cells.

III.3.2 Scheduling schemes

When a cell is involved in a coordination process, its resources are shared between its users and the additional users associated with neighboring cells, and requiring the cooperation of the considered cell to be served.

On the basis that a set of cells can cooperate in order to serve some users associated with neighboring cells, the scheduling decision made in each cell belonging to the cluster (see Section I.5.4) may affect the decisions of the other cells. Thus, the scheduling strategy is a key component of cellular systems supporting inter-cell coordination, as the scheduling decision of a cell may affect the decisions of all its cooperating cells. In order to allocate the resources efficiently to the users of a cluster, the CSI (channel state information) of all UEs should be taken into account by the scheduler. This can be easily achieved if a central scheduler is applied in the cluster and has knowledge of the CSI/CQI information of all users being served within the cluster. This approach is better suited for intra-site coordination where all sectors are controlled by the same macro BS. It can also perfectly work within a C-RAN (Centralized Radio Access Network) architecture. In the general case, one way to do this is to pre-configure one of the cells of the cluster as master cell. All other cells within the cluster act as slaves.

We consider in this section a centralized scheduler in each cluster. This scheduler has a global view of all the users' channel conditions in the given cluster and thus can make decisions about the scheduled users of all the coordinated cells that belong to the cluster. We focus on how to allocate resources between CoMP and non-CoMP users. The scheduling strategy can either be a fair strategy that treats CoMP and non-CoMP users equally or a prioritization strategy that prioritizes one category of users. We shall see in the following sections that the performance of the scheduling scheme, especially prioritization strategies, depends primarily on users' mobility.

III.3.2.1 Static allocation

The simplest scheduling strategy consists in allocating a fixed fraction of resources to CoMP and non-CoMP users' data flows, as proposed by 3GPP in [3, 4]: a CoMP frequency sub-band is introduced to perform CoMP operation, so that all users selected to be served in CoMP mode use this pre-configured frequency sub-band. Non-CoMP users are scheduled according to classical scheduling strategies, described in Chapter II, on their dedicated frequency sub-band. CoMP users are scheduled in a centralized manner as follows:

Data: set A of coordinated cells with CoMP UEs to be served
Result: Set U_{CoMP} of CoMP UEs to schedule
for each cluster A **do**
  $n \leftarrow |A|/2$
  while $(A \neq \emptyset)\,\&\,(n > 0)$ **do**
    $n \leftarrow n-1$
    max $\leftarrow 0$
    for each CoMP user u in cluster A **do**
      $S \leftarrow \{s\} \cup$ set of cooperating cells
      if $S \subseteq A$ **then**
        Compute metric
        if max < metric **then**
          max $\leftarrow$ metric
          selected user $\leftarrow$ u
        end
      end
    end
    $s \leftarrow$ serving cell of selected user
    $S \leftarrow \{s\} \cup$ set of cooperating cells of selected user
    $A \leftarrow A \setminus S$
    $U \leftarrow U \cup \{$selected user$\}$
  end
end

Algorithm 2: Centralized static allocation scheduling for JT.

A is the set of cells belonging to the same cluster. Observe that the maximum number of iterations for scheduling CoMP users is $|A|/2$, due to the fact that at most $|A|/2$ CoMP users can be scheduled at the same time on the same RBs.

Note that this algorithm is done for every subset of RBs belonging to the frequency sub-band dedicated to CoMP users. This scheduling strategy provides only limited performance due to the inefficient utilization of radio resources. We consider this strategy for the sake of comparison.

III.3.2.2 Iterative scheduling

Another way to do the scheduling in a centralized manner, is to do it in an iterative way, treating non-CoMP users and CoMP users equally without any frequency band partitioning. In this case the scheduling complexity scales linearly with the cluster size N. It requires at most a number of iterations which is equal to the size of the cluster. This is due to the fact that at most N users can be scheduled simultaneously on the same resources within a cluster of N cells. The first user is chosen based on the maximum scheduling metric (PF, max C/I...) or even randomly (for RR) blocking the serving cell and all cooperating cells in the case of a CoMP user. The next best user according to the same metric is chosen only if its serving cell and all its cooperating cells are still available after the first decision and so on till all cells in the cluster become unavailable... In LTE this resource allocation algorithm is done each TTI for every subset of RBs. The algorithm works as follows:

Data: set A of coordinated cells with UEs to be served
Result: Set U of UEs to schedule
for each cluster A **do**
  $n \leftarrow |A|$
  while $(A \neq \emptyset)\,\&\,(n > 0)$ **do**
    $n \leftarrow n - 1$
    max $\leftarrow 0$
    for each user u in cluster A **do**
      $S \leftarrow \{s\} \cup$ set of cooperating cells
      if $S \subseteq A$ **then**
        Compute metric
        if max < metric **then**
          max $\leftarrow$ metric
          selected user $\leftarrow$ u
        end
      end
    end
    $s \leftarrow$ serving cell of selected user
    $S \leftarrow \{s\} \cup$ set of cooperating cells of selected user
    $A \leftarrow A \setminus S$
    $U \leftarrow U \cup \{$selected user$\}$
  end
end

Algorithm 3: Centralized iterative scheduling for JT.

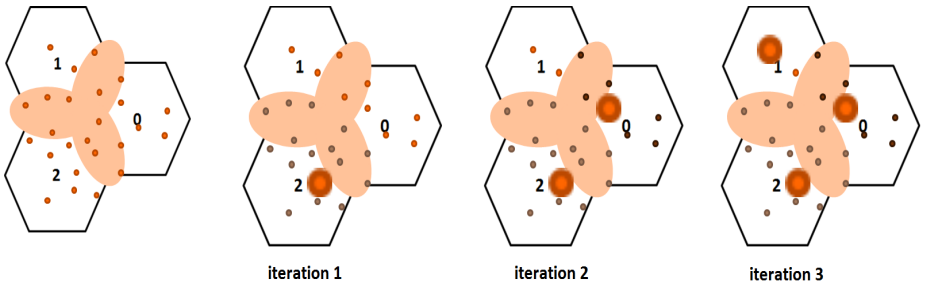


Figure III.5: Iterative scheduling for JT.

Figure III.5 shows an example of a cluster of three cells. The first user is chosen in cell 2 and

does not involve any other cell in the transmission, thus blocking all CoMP and non-CoMP users requiring transmission from cell 2. The second user is then selected from cell 0 blocking all users requiring transmission from this cell. As cell 1 is still available, it can schedule one of the remaining two non-CoMP users requiring only transmission from cell 1.

III.3.2.3 Priority to non-CoMP users

Under this policy, non-CoMP users in each cell are scheduled first and are allocated all the radio resources whenever active; CoMP users wait untill resources in all coordinated cells become available. Scheduling is done in two steps, each TTI and for every subset of resource blocks. Non-CoMP users are selected first according to classical scheduling policies (RR, PF, max C/I). The scheduling decision in each cell is independent of the decisions in the neighboring cells. In some cases, no users are scheduled within a cell due mainly to the fact that there is no non-CoMP users in this cell. The second step consists in assigning remaining resources to appropriate CoMP users. The algorithm describing the second step works as follows:

```
Data: set A of coordinated cells remaining available (after first scheduling step) with CoMP UEs to be
      served
Result: Set U_CoMP of CoMP UEs to schedule
for each set A do
    n ← |A| /2
    while (A ≠ ∅) & (n > 0) do
        n ← n − 1
        max ← 0
        for each CoMP user u in set A do
            S ← {s} ∪ set of cooperating cells
            if S ⊆ A then
                Compute metric
                if max < metric then
                    max ← metric
                    selected user ← u
                end
            end
        end
        s ← serving cell of selected user
        S ← {s} ∪ set of cooperating cells of selected user
        A ← A \ S
        U ← U ∪ {selected user}
    end
end
```

Algorithm 4: Centralized Pri-NC scheduling for JT.

Users are chosen based on a particular scheduling metric (PF, max C/I...) or even randomly (for RR). When JT is activated at most $|A|/2$ CoMP users can be scheduled at the same time in a given cluster due to the fact that each CoMP user is involving at least two cells in the transmission. Thus the maximum number of iterations when scheduling CoMP users is $|A|/2$.

III.3.2.4 Priority to CoMP users

Under this policy, CoMP users are scheduled first and are allocated all radio resources whenever active. Non-CoMP users are served only when there are no active CoMP users or when there are still

available cells after scheduling CoMP users. The algorithm works as follows:

```
Data: set A of coordinated cells with UEs to be served
Result: Set U of UEs to schedule
for each cluster A do
    n ← |A|
    switch To NC ← 0
    while (A ≠ ∅) & (n > 0) do
        n ← n − 1
        max ← 0
        if !switch To NC then
            for each CoMP user u in cluster A do
                S ← {s} ∪ set of cooperating cells
                if S ⊆ A then
                    Compute metric
                    if max < metric then
                        max ← metric
                        selected user ← u
                    end
                end
            end
            if There is no selected user then
                switch To NC ← 1
            end
        end
        if switch To NC then
            for each non-CoMP user u in cluster A do
                S ← {s}
                if S ⊆ A then
                    Compute metric
                    if max < metric then
                        max ← metric
                        selected user ← u
                    end
                end
            end
        end
        s ← serving cell of selected user
        S ← {s} ∪ set of cooperating cells of selected user
        A ← A \ S
        U ← U ∪ {selected user}
    end
end
```

Algorithm 5: Centralized Pri-C scheduling for JT.

III.3.2.5 Mobility-aware scheduler

This scheduler exploits the mobility as an additional information in order to schedule users with elastic traffic.

Deprioritizing CoMP users (see Section III.3.2.3) in the presence of mobility gives them time to move so that they are more likely served in good radio conditions where cell coordination is not

required and thus improves performance. However, this strategy decreases the system stability condition when users are static: the condition of serving a CoMP user is more restrictive than that of serving a non-CoMP user due to the fact that a CoMP user requires the availability of all coordinated cells and thus static CoMP users are never served at high load.

It turns out that in a system with multiple mobility behaviors, static CoMP users should be treated equally as non-CoMP users, while it is better to deprioritize mobile CoMP users; the more the CoMP user is mobile the more this user should be deprioritized. This can be done by assigning a score inversely proportional to the speed to each CoMP user, so that the scheduling algorithm becomes mobility-aware. For instance, the score is equal to one for static CoMP users and less than one for mobile CoMP users. Note that this requires an estimation of the mobile speed. The algorithm works as described below in Algorithm 6.

```
Data: set A of coordinated cells with UEs to be served
Result: Set U of UEs to schedule
for each cluster A do
    n ← |A|
    while (A ≠ ∅) & (N > 0) do
        n ← n − 1
        max ← 0
        for each user u in cluster A do
            S ← {s} ∪ set of cooperating cells
            if S ⊆ A then
                Compute metric
                if user u is a CoMP user then
                    score ← f(speed of user u)
                    metric ← metric × score
                end
                if max < metric then
                    max ← metric
                    selected user ← u
                end
            end
        end
        s ← serving cell of selected user
        S ← {s} ∪ set of cooperating cells of selected user
        U ← U ∪ {selected user}
        A ← A \ S
    end
end
```

Algorithm 6: Mobility-aware algorithm.

III.3.3 SINR map

Figure III.7 shows the coordination zone (black zone), that is the zone where users are in CoMP mode and require the cooperation of some neighboring cells to be served, under different clustering strategies and for different power threshold δP: 6 dB, 12 dB, 18 dB. In other words, black zone corresponds to the locations of users where one or more neighboring cells taking part in the cluster of the serving cell fulfill the condition (I.1). White zone represents the locations of those users who are served only by their serving cells without any coordination. The higher the δP the higher the coordination area.

Figure III.6 shows the variation of the coordination area with respect to δP under different clustering strategies.

Figures III.8 - III.15 show the SINR Maps (in dB) according to (III.12) and the service rate Maps (in bit/s/Hz) for both transmission schemes as given by (III.13) and (III.14) for different power threshold δP: 6 dB, 12 dB, 18 dB. taking into account only the macroscopic pathloss, see Appendix A. The mapping between the SINR and the service rate is obtained according to LTE CQI table.

When $K_{max} = 2$, maximum two cells including the serving cell are involved in the transmission for a CoMP user. When $K_{max} = 3$, maximum three cells can transmit simultaneously to the same CoMP user.

Figures III.8-III.9 illustrate the case of intra-site coordination. The case of inter-site coordination is illustrated by Figures III.10-III.11, while the case of intra and inter-site coordination is shown in Figures III.12-III.13. The case of dynamic clustering is illustrated by Figures III.14-III.15.

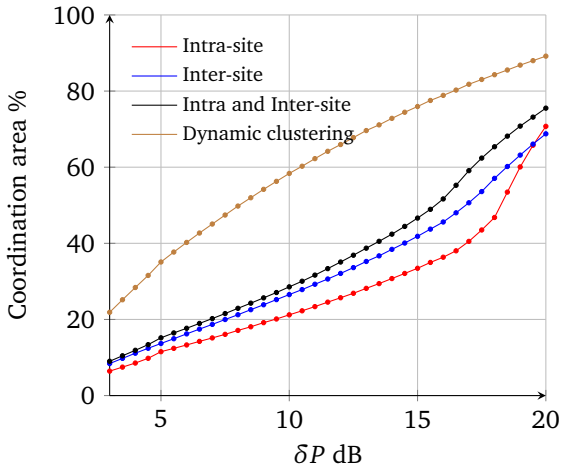
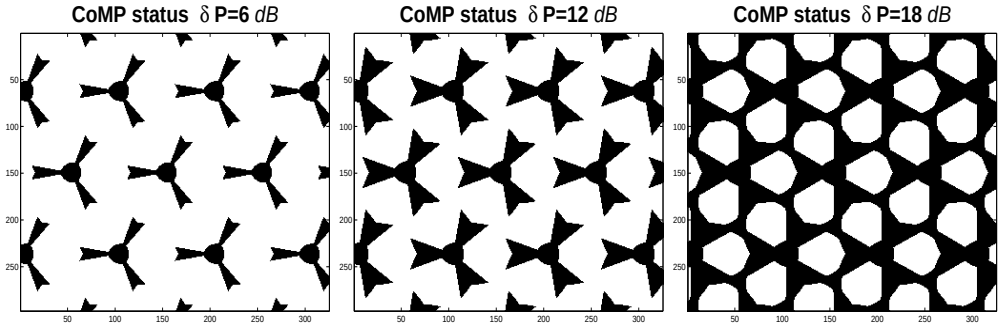


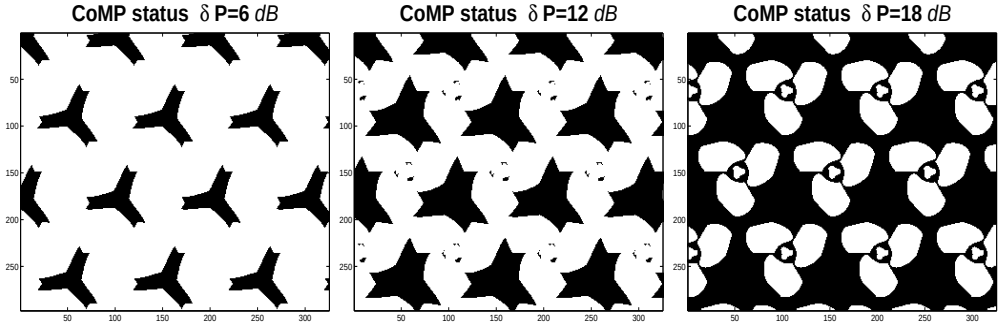
Figure III.6: Coordination area under different clustering strategies.

III.3.3.1 SINR cumulative distribution function

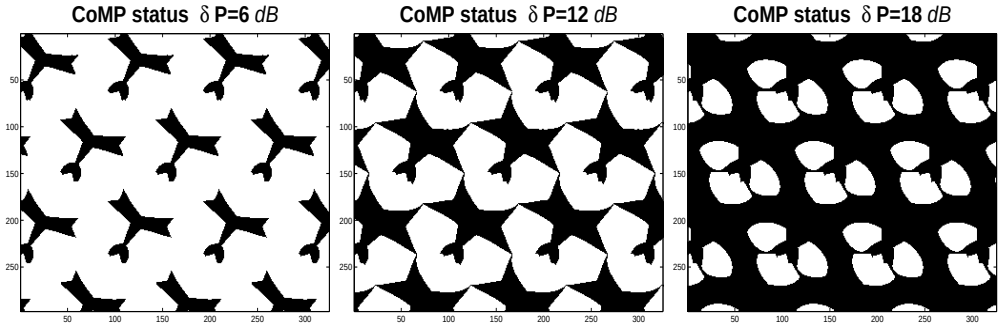
Figure III.17 gives the cumulative distribution function of the SINR in the case without coordination as given by (III.10) and when activating JT with different values of δP as given by (III.12), under different clustering methods and for different values of maximum coordinated cells K_{max}. Observe that there is coverage improvement when JT is activated. In the case of intra-site coordination with $\delta P = 18$ dB, the mean SINR is improved by 1.8 dB when $K_{max} = 2$ and by 2.8 dB when $K_{max} = 3$. The improvement of the mean SINR is higher in the case of inter-site coordination with same $\delta P = 18$ dB, where the increase is around 2.3 dB when $K_{max} = 2$ and 4.5 dB when $K_{max} = 3$. Obviously, dynamic clustering (a fully coordinated network) guarantee the highest SINR improvement where the mean value increases by 7.6 dB when $K_{max} = 3$ and by 10.2 dB when there is no limit on the number of coordinated cells that can be involved simultaneously in the transmission to the same user. Note that this figure gives the performance improvement in terms of coverage but does not show anything on the resource consumption and the mean user throughput in a dynamic setting where users come and go over time.



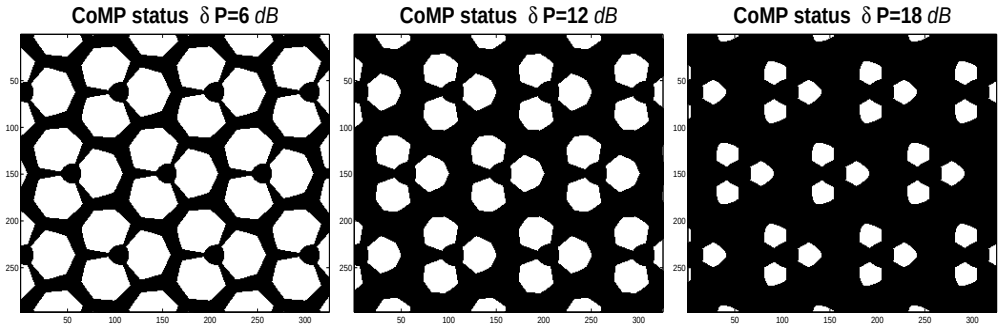
(a) Intra-site



(b) Inter-site



(c) Intra and Inter-site



(d) Dynamic clustering

Figure III.7: Coordination zone (black area).

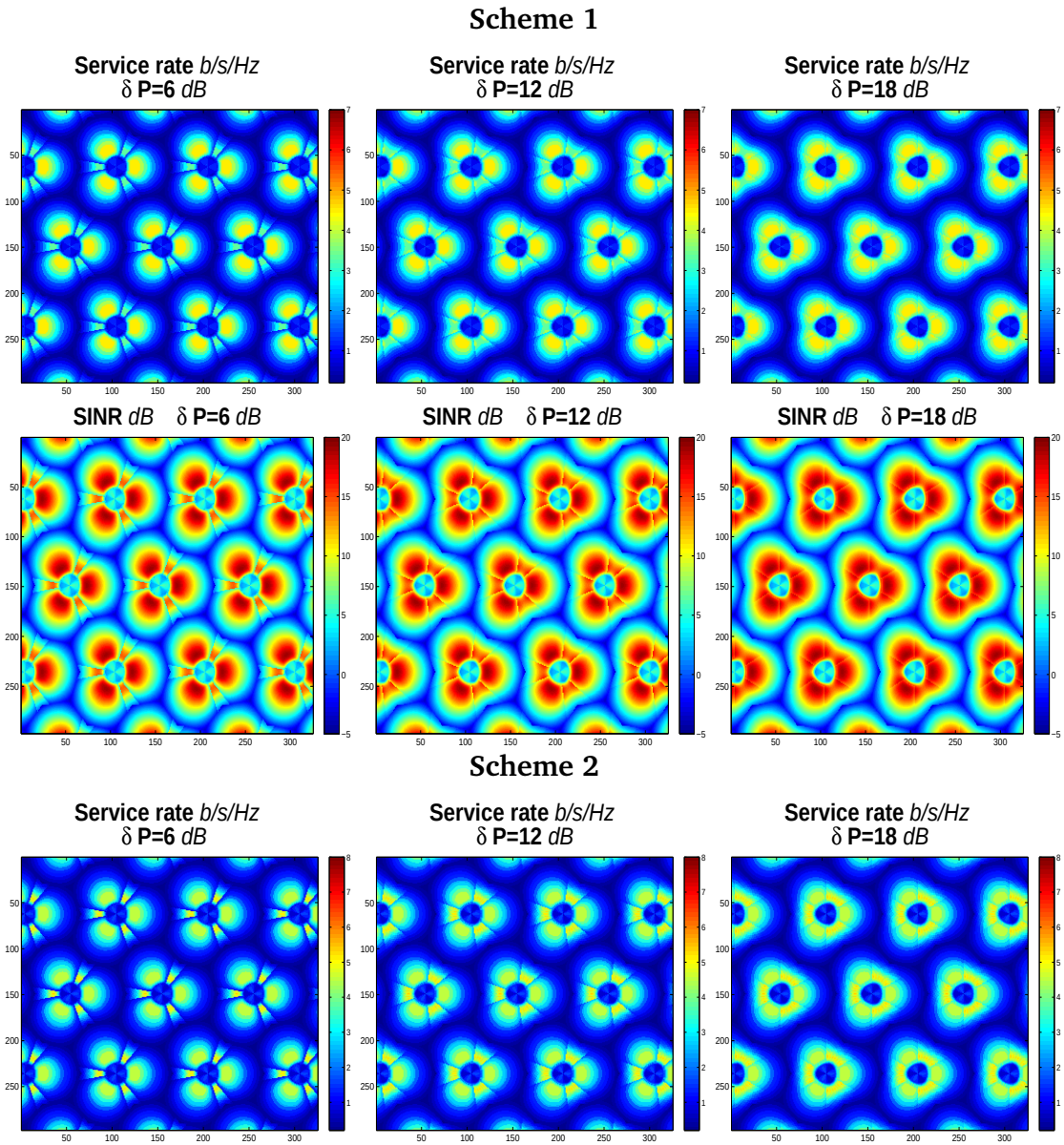


Figure III.8: Service rate and SINR maps for Intra-site coordination with $K = 2$.

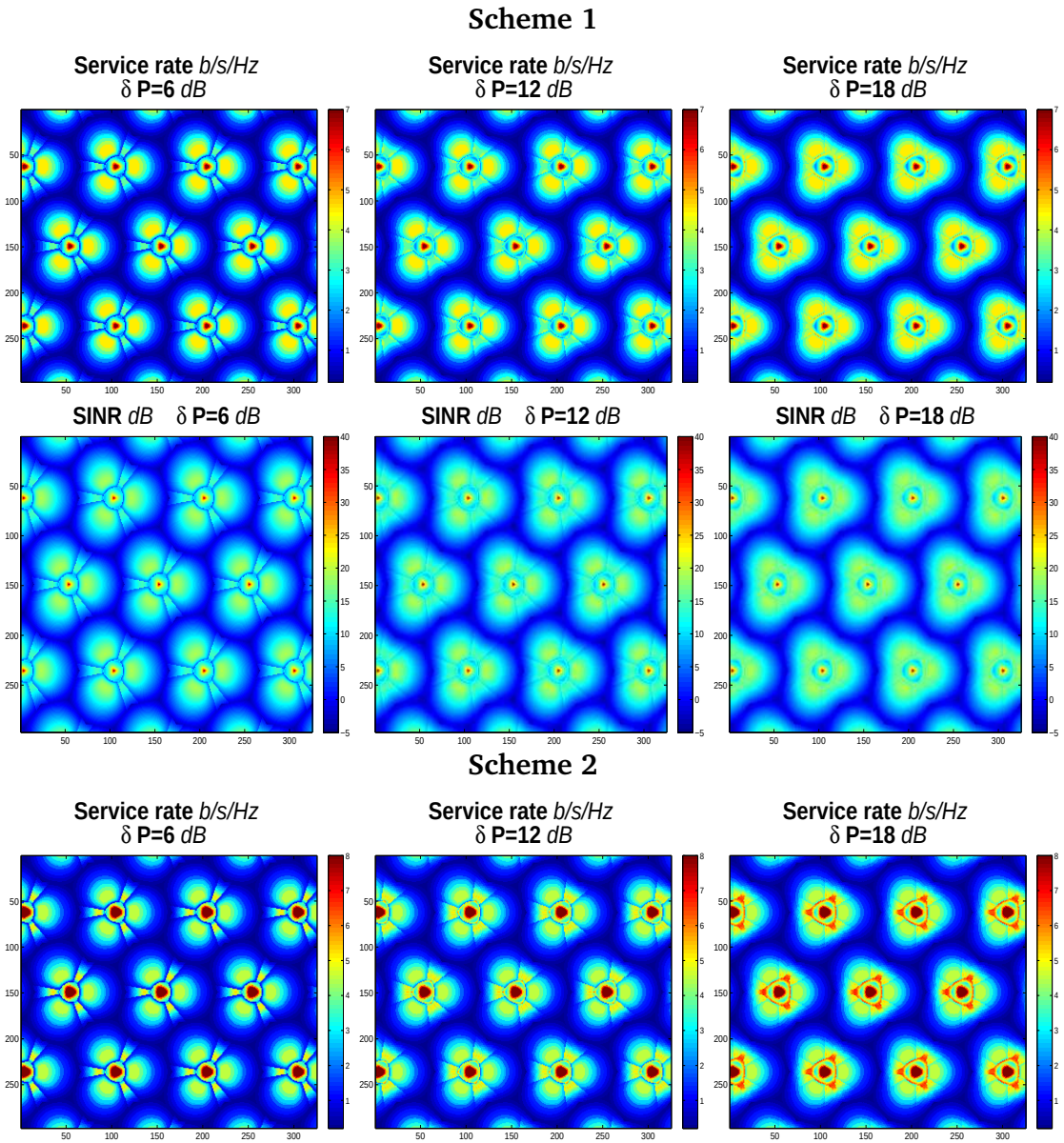


Figure III.9: Service rate and SINR maps for Intra-site coordination with $K = 3$.

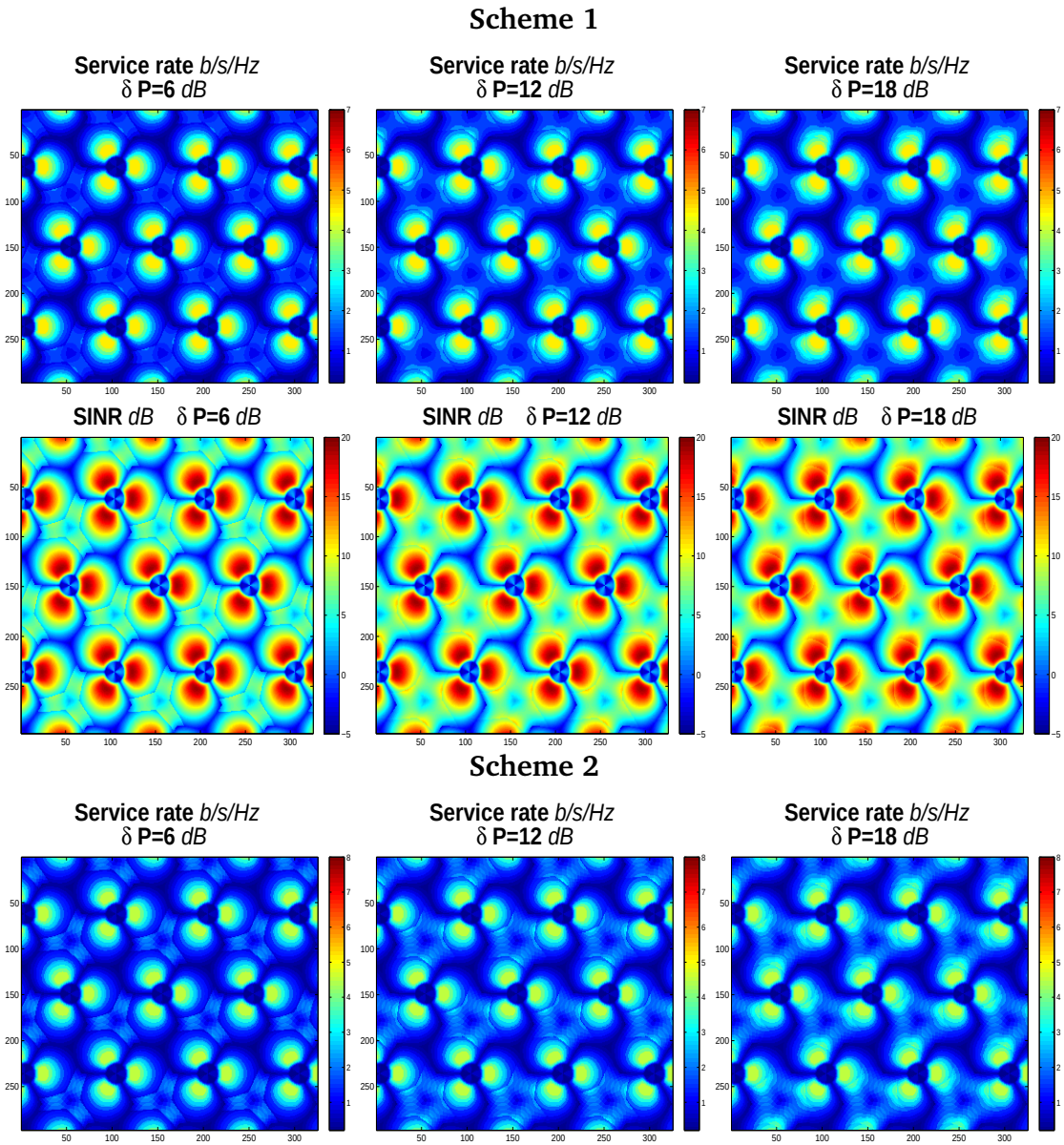


Figure III.10: Service rate and SINR maps for Inter-site coordination with $K = 2$.

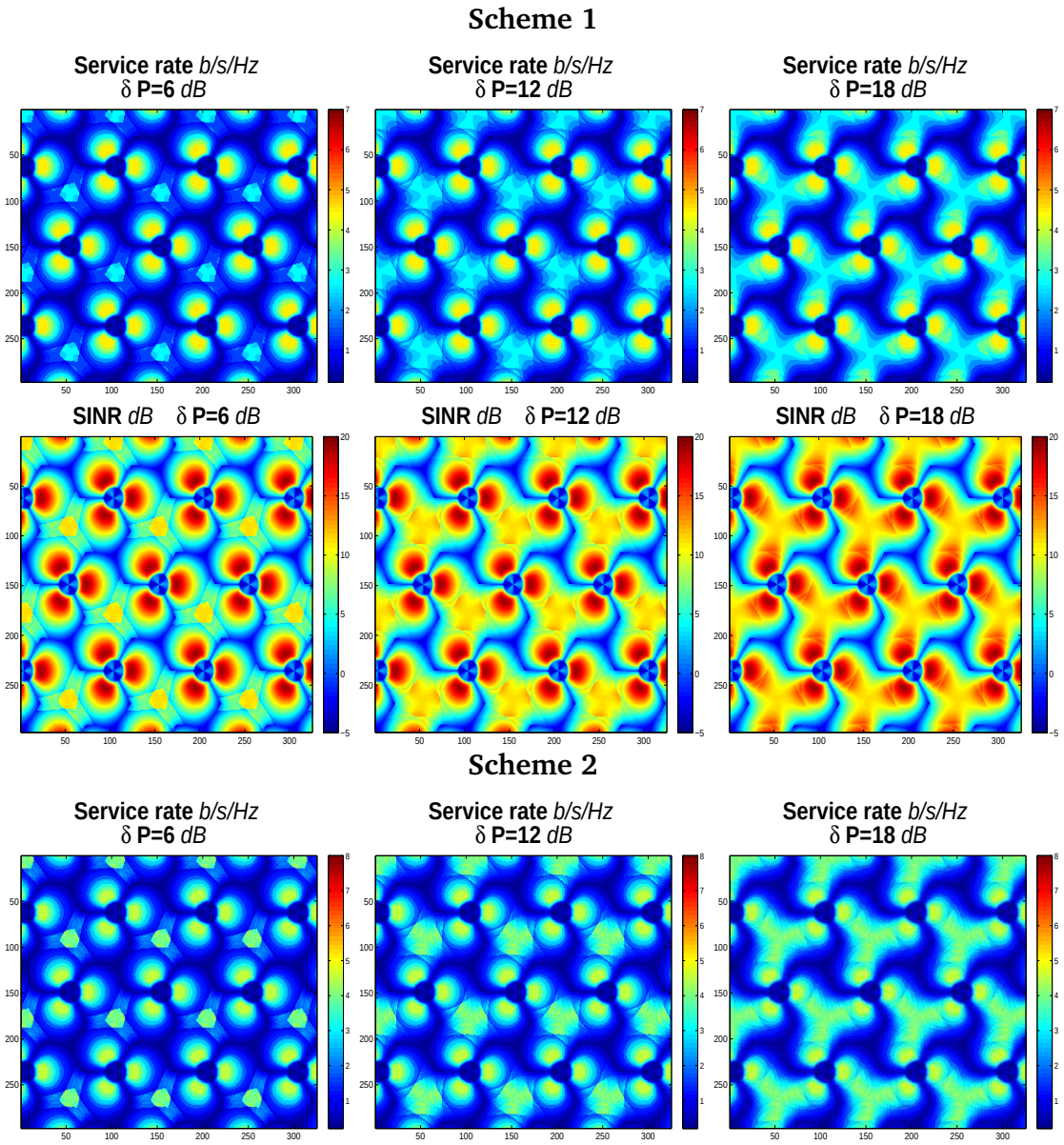


Figure III.11: Service rate and SINR maps for Inter-site coordination with $K = 3$.

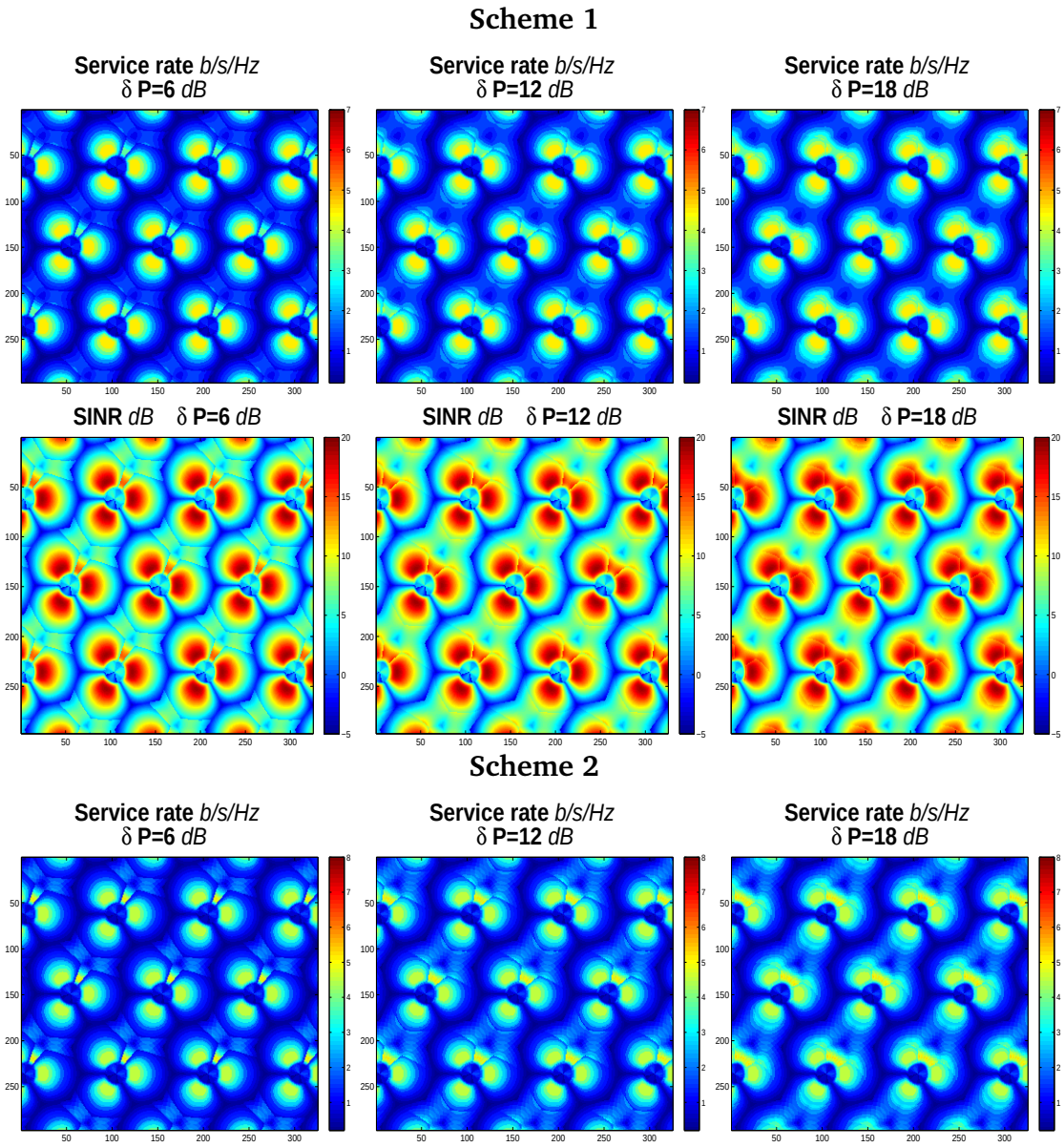


Figure III.12: Service rate and SINR maps for Intra and Inter-site coordination with $K = 2$.

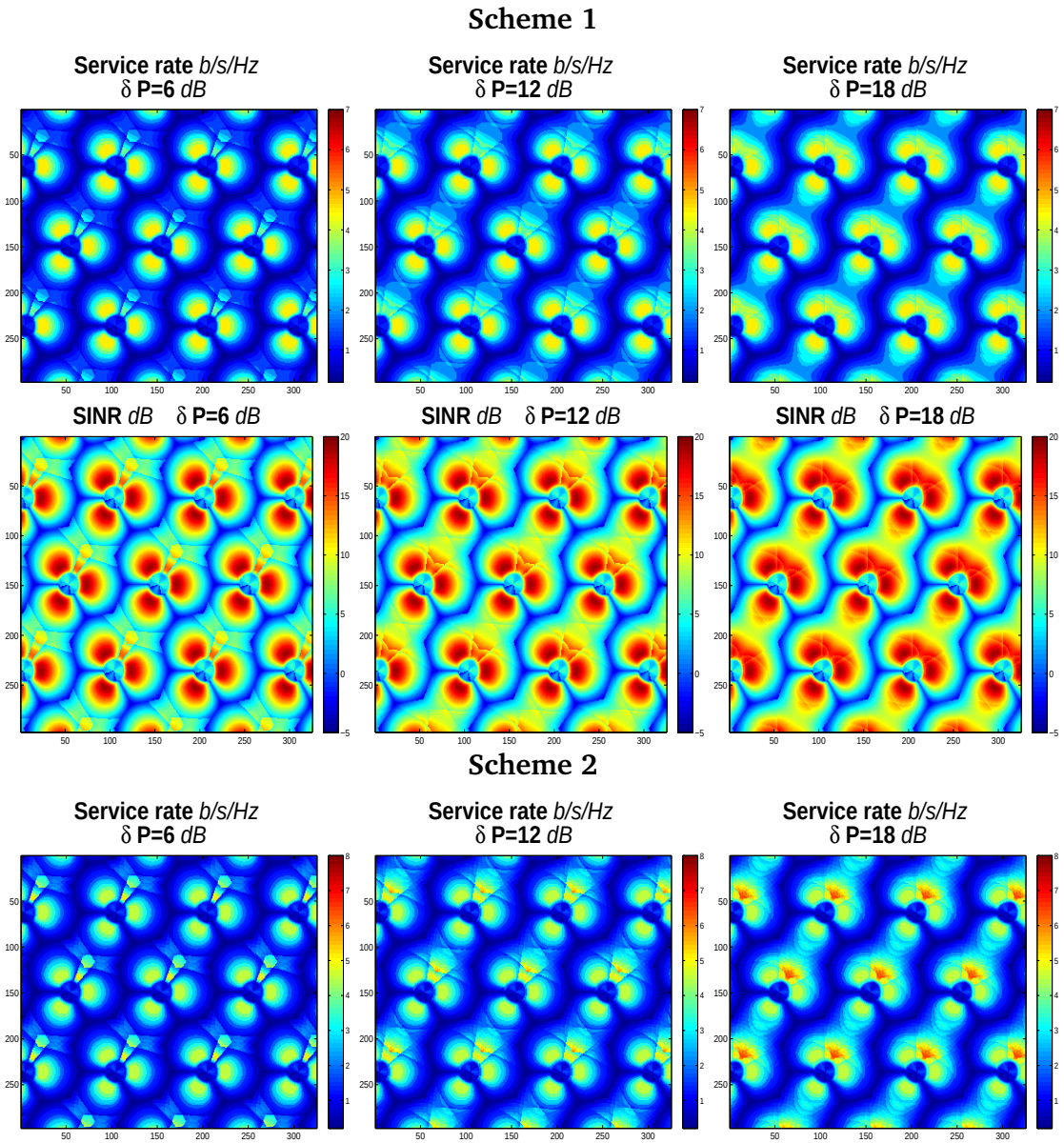


Figure III.13: Service rate and SINR maps for Intra and Inter-site coordination with $K = 3$.

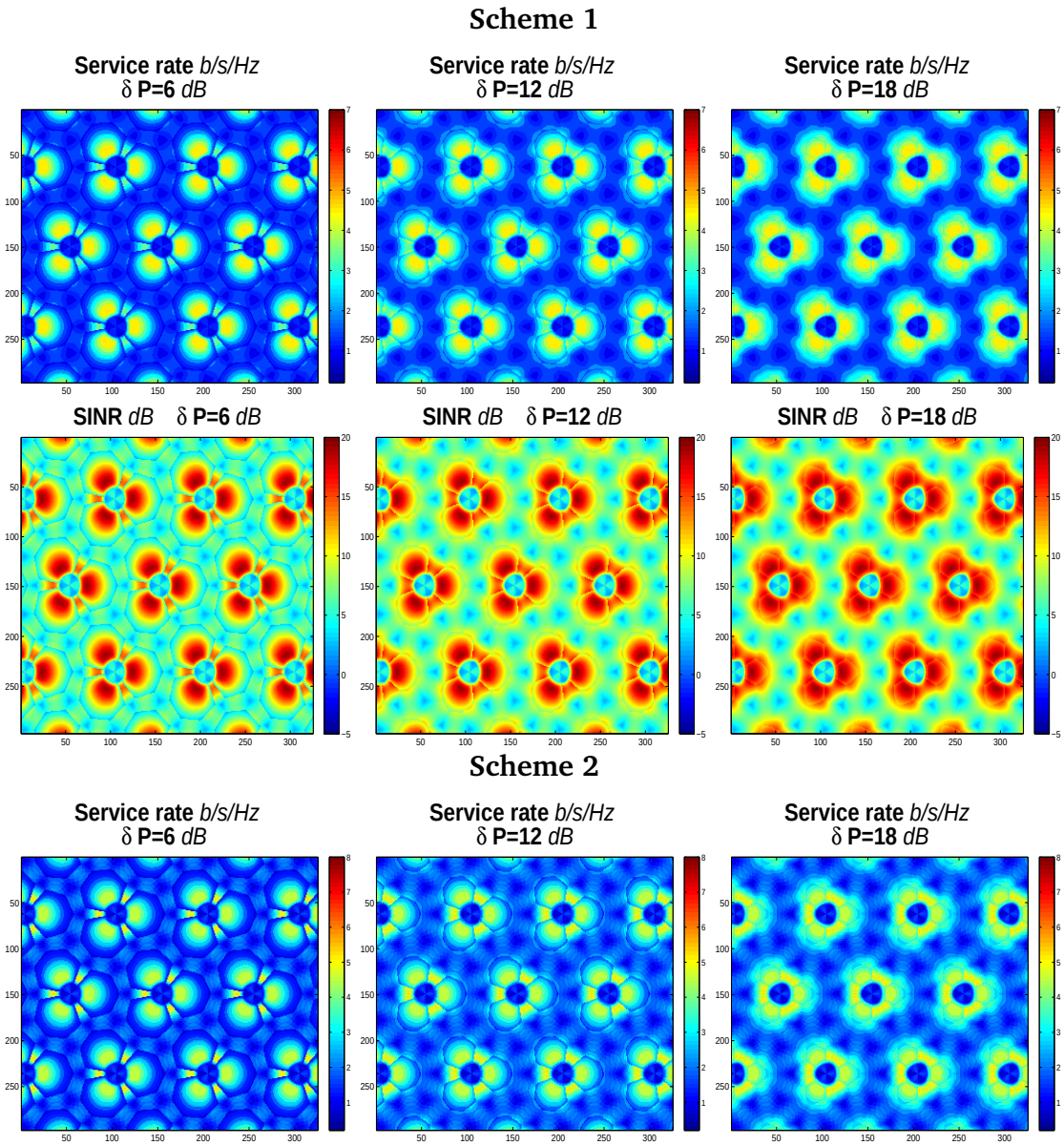


Figure III.14: Service rate and SINR maps for Dynamic clustering with $K = 2$.

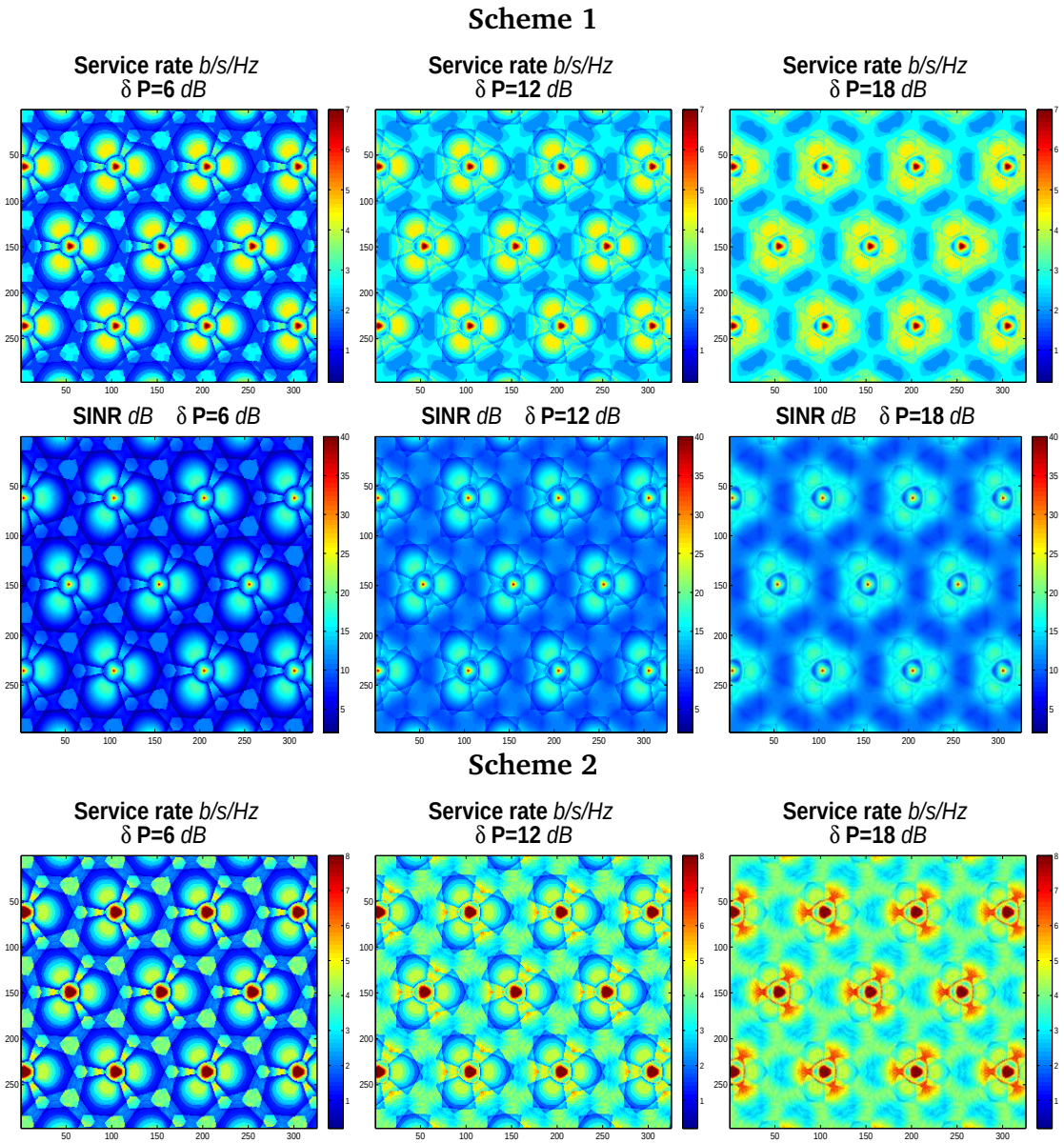
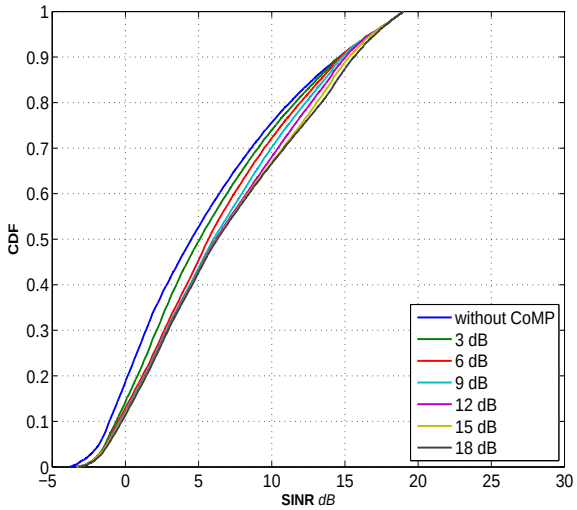


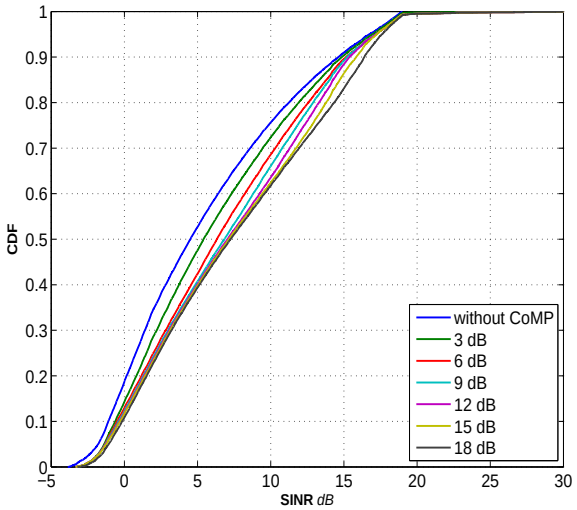
Figure III.15: Service rate and SINR maps for Dynamic clustering with $K = 3$.

Intra-site coordination

$K_{max} = 2$ **maximum coordinated cells**

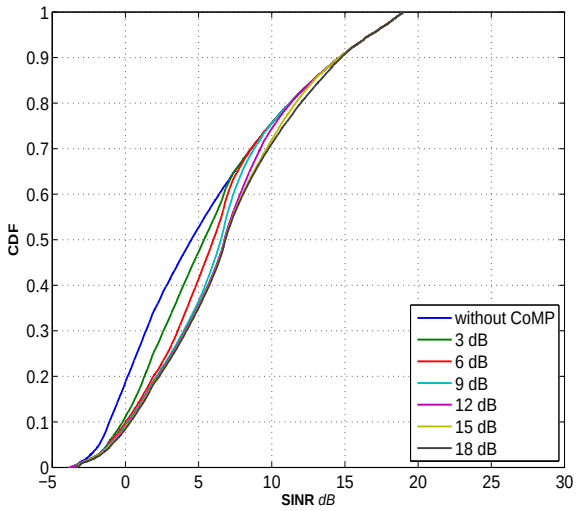


$K_{max} = 3$ **maximum coordinated cells**

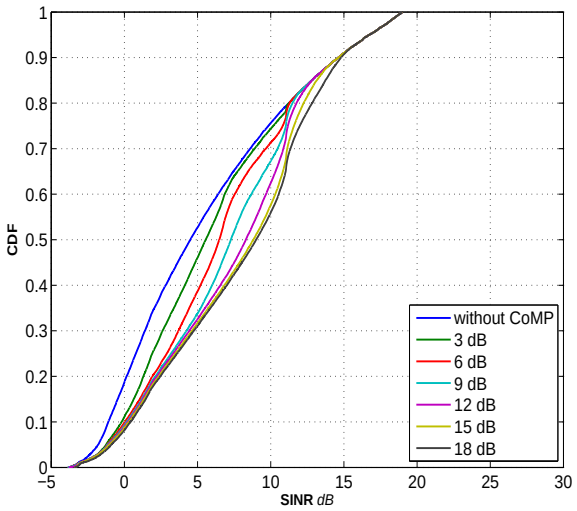


Inter-site coordination

$K_{max} = 2$ **maximum coordinated cells**

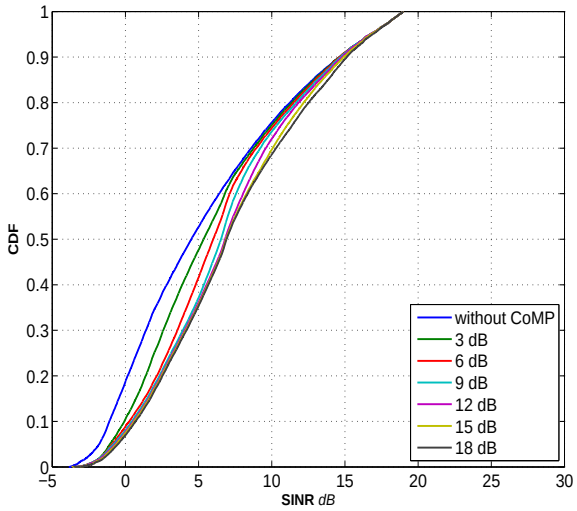


$K_{max} = 3$ **maximum coordinated cells**



Intra-site and Inter-site coordination

$K_{max} = 2$ **maximum coordinated cells**



$K_{max} = 3$ **maximum coordinated cells**

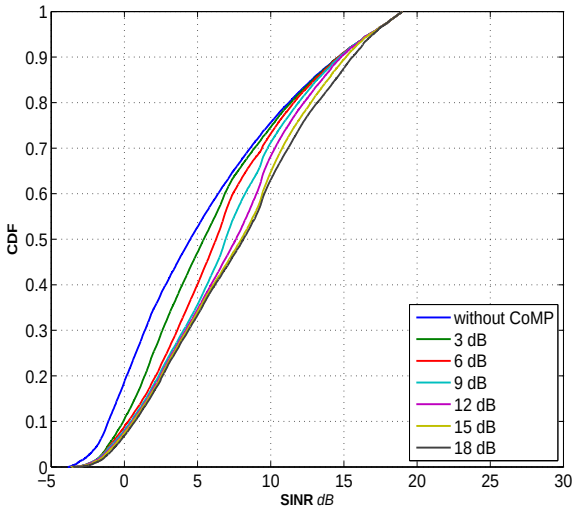


Figure III.16: SINR CDF for fixed clustering JT CoMP

Dynamic clustering

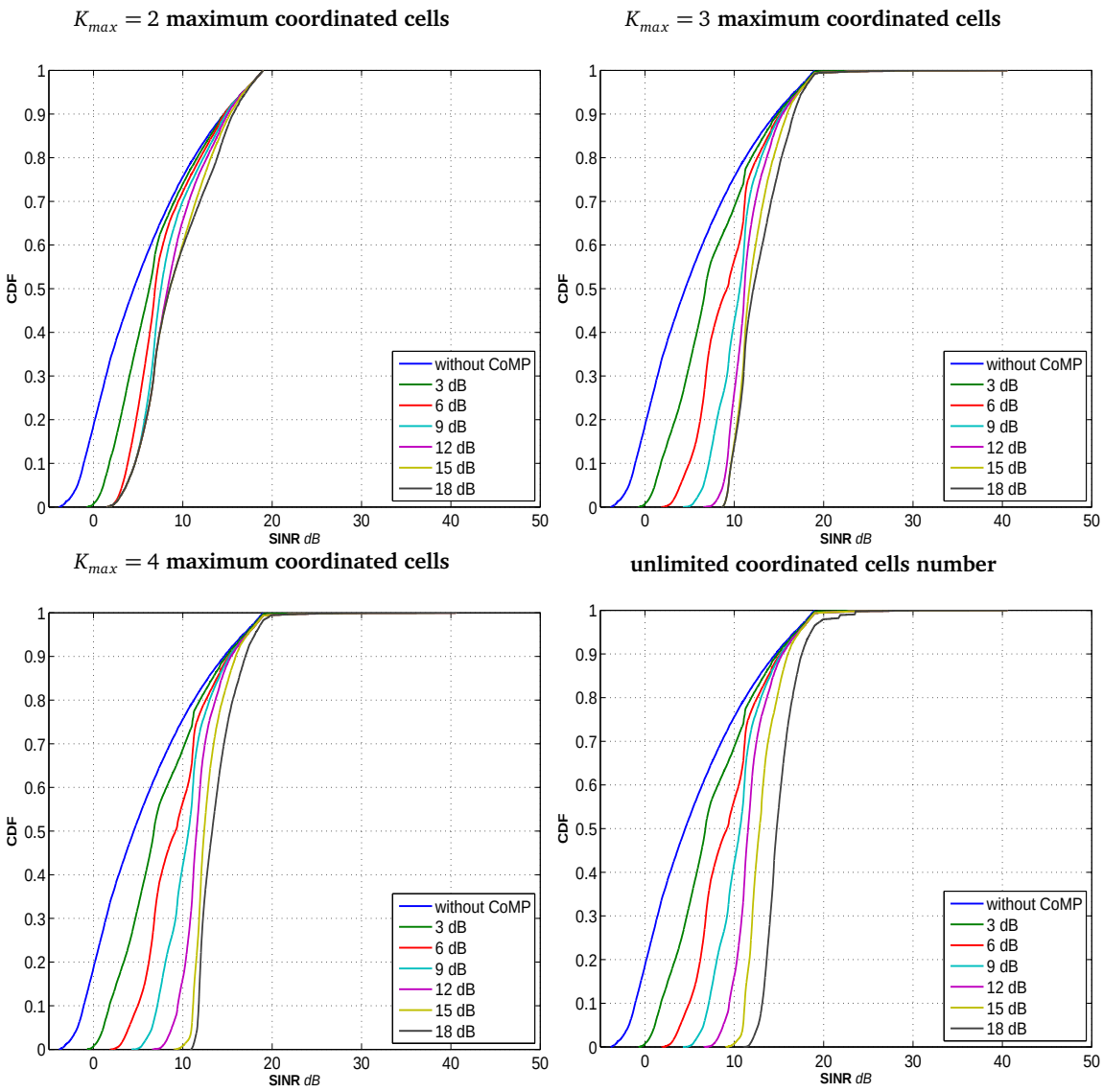


Figure III.17: SINR CDF for Dynamic clustering JT CoMP

III.3.4 Flow level model

III.3.4.1 Single coordination zone

The resources of each cell are shared between CoMP users (in coordination zone $k = 0$) and non-CoMP users (in non-coordination zones $k = 1,\dots,K$), as illustrated by Figure III.18. The actual service rate in zone k is modulated by ϕ_k, the fraction of time spent by the scheduler on users in zone k. This depends on the system state x and on the scheduling policy. For work-conserving policies, we have

$$\forall k : x_k > 0, \quad \phi_k(x) + \phi_0(x) = 1. \tag{III.15}$$

We focus on how to allocate resources between CoMP and non-CoMP users' data flows. In each

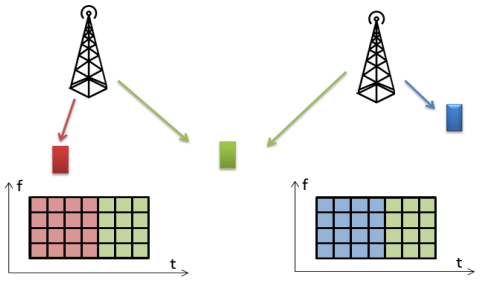


Figure III.18: Resource allocation between CoMP and non-CoMP users.

zone, users are assumed to share the allocated radio resources equally, independently of their radio conditions. However, when the network operates under an opportunistic scheduler which exploits the fast fading, a rate gain, see §II.2.2.4,should be taken into account.

A necessary condition for the ergodicity of the Markov process $X(t)$ is $\rho < 1$ in the absence of mobility, with

$$\rho = \rho_0 + \max_{k=1,\dots,K} \rho_k,$$

that is $\rho < \lambda/\mu$ with

$$\mu = \left(\frac{p_0}{\mu_0} + \max_{k=1,\dots,K} \frac{p_k}{\mu_k} \right)^{-1}.$$

The sufficient stability condition depends on the scheduling policy. We shall see in the following sections that mobility may increase or decrease the stability region, depending on the scheduling strategy.

a) Model without mobility

i. Static allocation CoMP users' data flows are allocated some fixed fraction ϕ_0 of the resources of each coordinated cell while non-CoMP users' data flows in cell k are allocated the fraction $\phi_k = 1-\phi_0$ of the considered cell radio resources.

The corresponding Markov process $X(t)$ is reversible, with stationary distribution

$$\pi(x) = \prod_{k=0}^{K} \frac{(1-\rho_k/\phi_k)}{\phi_k^{x_k}} \binom{x_k}{x_{1k},\dots,x_{N_k k}} \rho_{1k}^{x_{1k}} \cdots \rho_{N_k k}^{x_{N_k k}},$$

provided $\rho_0 < \phi_0$ and $\rho_k < 1-\phi_0$ for all $k = 1,\dots,K$. Observe that this condition may be more restrictive than the natural stability condition $\rho < 1$.

From (III.4), the mean throughput in zone k is

$$\gamma_k = \phi_k \mu_k \sigma (1 - \rho_k/\phi_k).$$

Observe that the mean throughput in zone k decreases *linearly* from the physical rate $\phi_k \mu_k \sigma$ to 0 when the load of the corresponding zone increases from 0 to ϕ_k. Indeed, the stability condition of zone k is $\rho_k < \phi_k$.

ii. Iterative scheduler Resources are shared so as to maximize

$$\sum_{k=0}^{K} x_k \log \phi_k.$$

We get

$$\phi_0(x) = \frac{x_0}{\sum_{k=0}^{K} x_k}$$

and $\phi_k(x) = 1 - \phi_0(x)$ for all k and all x such that $x_k > 0$.

We generalize the model presented in [43] for $K = 2$ coordinated cells to any $K \geq 2$. The stationary distribution of the Markov process $X(t)$ is:

$$\pi(x) = \pi(0) \binom{x_0 + x_1 \ldots + x_K}{x_0} \times \prod_{k=0}^{K} \binom{x_k}{x_{1k}, \ldots, x_{N_k k}} \rho_{1k}^{x_{1k}} \cdots \rho_{N_k k}^{x_{N_k k}},$$

under the stability condition $\rho < 1$, with

$$\pi(0) = \frac{(1-\rho_0-\rho_1)(1-\rho_0-\rho_2)\ldots(1-\rho_0-\rho_K)}{(1-\rho_0)^{K-1}}.$$

By (III.4)–(III.5), we deduce the mean throughput in the non-coordination zone k,

$$\gamma_k = \mu_k \sigma (1 - \rho_0 - \rho_k), \tag{III.16}$$

and the mean throughput in the coordination zone,

$$\gamma_0 = \mu_0 \sigma \left(\sum_{k=1}^{K} \frac{1}{1-\rho_0-\rho_k} - \frac{K-1}{1-\rho_0} \right)^{-1}. \tag{III.17}$$

iii. Priority to non-CoMP users Under this policy, non-CoMP users are scheduled first while CoMP users wait until resources become available. Thus, the stability condition is given by:

$$\rho_0 < \prod_{k=1}^{K} (1 - \rho_k), \tag{III.18}$$

which is more restrictive than the natural stability condition $\rho < 1$. This is due to the fact that some cells may be idle when serving non-CoMP users. The mean throughput in non-coordination zone k is:

$$\gamma_k = \mu_k \sigma (1 - \rho_k).$$

There is no explicit expression for the mean throughput in the coordination zone.

iv. Priority to CoMP users Under this policy, CoMP users are scheduled first; non-CoMP users are served only when there are no active CoMP users. The stability condition is the natural condition $\rho < 1$. Since CoMP users are not affected by non-CoMP users, we have

$$E(X_0) = \frac{\rho_0}{1-\rho_0}.$$

By (III.4), the mean throughput in the coordination zone is:

$$\gamma_0 = \mu_0\sigma(1-\rho_0).$$

Applying known results of queuing theory [46], the mean number of flows in non-coordination zone k is

$$E(X_k) = \frac{\rho_k}{1-\rho_0}\left(1 + \frac{\rho_k + \rho_0\mu_k/\mu_0}{1-\rho_k-\rho_0}\right)$$

We deduce the mean throughput in non-coordination zone k,

$$\gamma_k = \mu_k\sigma\frac{(1-\rho_k-\rho_0)(1-\rho_0)}{(1-\rho_k-\rho_0)+\rho_k+\rho_0\mu_k/\mu_0}.$$

Observe that the mean throughput of CoMP users is positive whenever $\rho_0 < 1$ while the mean throughput of non-CoMP users in zone k is positive whenever $\rho_0 + \rho_k < 1$.

b) Model with mobility

i. Stability condition The stability condition follows from the limiting regime of infinite mobility where $\nu^{(k)}_{i,i+1}, \nu^{(k)}_{i+1,i} \to \infty$ for all $i < N_k$ and $k \geq 0$, and $\nu_{k,0}, \nu_{0,k} \to \infty$ for all $k > 0$ [21]. In this regime, the mean service rate in zone k becomes

$$\bar{\mu}_k = \sum_i q_{ik}\mu_{ik}/q_k.$$

The overall mean service rate is

$$\bar{\mu} = (1-\phi_0)\sum_{k=1}^{K}\bar{\mu}_k + \phi_0\bar{\mu}_0$$

under the SA scheme,

$$\bar{\mu} = (1-q_0)\sum_{k=1}^{K}\bar{\mu}_k + q_0\bar{\mu}_0$$

under the PF scheme,

$$\bar{\mu} = \sum_{k=1}^{K}\bar{\mu}_k$$

under the Pri-NC scheme, and

$$\bar{\mu} = \bar{\mu}_0$$

under the Pri-C scheme. The stability condition under the four schemes is

$$\rho < \bar{\mu}/\mu.$$

Note that in the absence of inter-zone mobility, the stability condition is:

$$\rho < \left(\frac{p_0}{\mu_0} + \max_{k=1,\ldots,K}\frac{p_k}{\mu_k}\right)\min_{k=0,\ldots,K}\frac{\phi_k\bar{\mu}_k}{p_k}$$

under the SA scheme,

$$\rho < \left(\frac{p_0}{\mu_0} + \max_{k=1,\dots,K} \frac{p_k}{\mu_k}\right)\left(\frac{p_0}{\bar{\mu}_0} + \max_{k=1,\dots,K} \frac{p_k}{\bar{\mu}_k}\right)^{-1}$$

under the PF and Pri-C schemes, and

$$\frac{\mu_0}{\bar{\mu}_0}\rho_0 < \prod_{k=1}^{K}\left(1 - \frac{\mu_k}{\bar{\mu}_k}\rho_k\right),$$

under the Pri-NC scheme, with

$$\bar{\mu}_k = \sum_i q_{ik}\mu_{ik},$$

and

$$q_{ik} \propto \prod_{j=i}^{N_k-1} \frac{\nu_{j+1,j}^{(k)}}{\nu_{j,j+1}^{(k)}}.$$

ii. Throughput in light traffic The performance in light traffic (that is, when $\rho \to 0$) is the same for all policies except the static allocation: a user when alone in the system is always allocated all radio resources. For instance, we consider that there is only one region in each zone. The probability that a user in non-coordination zone k moves to the coordination zone before leaving the system is

$$\alpha_{k,0} = \frac{\nu_{k,0}}{\mu_k + \nu_{k,0}}.$$

Similarly, the probability that a user in the coordination zone moves to non-coordination zone k before leaving the system is

$$\alpha_{0,k} = \frac{\nu_{0,k}}{\mu_0 + \nu_0},$$

where

$$\nu_0 = \sum_{j=1}^{K} \nu_{0,j}.$$

The mean duration of a flow initiated in a non-coordination zone is

$$T_k = \frac{1}{\nu_{k,0} + \mu_k} + \frac{\alpha_{k,0}}{1 - \sum_{j=1}^{K} \alpha_{0,j}\alpha_{j,0}} \left[\frac{1}{\nu_0 + \mu_0} + \sum_{j=1}^{K} \frac{\alpha_{0,j}}{\nu_{j,0} + \mu_j}\right],$$

while the mean duration of a flow initiated in the coordination zone is given by:

$$T_0 = \frac{1}{1 - \sum_{j=1}^{K} \alpha_{0,j}\alpha_{j,0}} \left[\frac{1}{\nu_0 + \mu_0} + \sum_{j=1}^{K} \frac{\alpha_{0,j}}{\nu_{j,0} + \mu_j}\right].$$

We deduce the mean duration of a flow initiated in the cluster,

$$T = \sum_{k=0}^{K} p_k T_k,$$

and the average throughput in light traffic $\gamma = \sigma/T$.
Under the SA policy, we get the mean throughput similarly by replacing μ_k by $\phi_k\mu_k$ ($\forall k \le K$).

III.3.4.2 Intra-site coordination

The resources of each sector k are shared between CoMP users in coordination zones $k, k+1$ and $k-1, k$ and non-CoMP users in the non-coordination zone k. The actual service rate in zone k is modulated by ϕ_k while $\phi_{k,k+1}$ represents the actual service rate in zone $k, k+1$. ϕ_k and $\phi_{k,k+1}$ are the fractions of time spent by the scheduler on users in zone k and zone $k, k+1$ respectively. This depends on the system state x and on the scheduling policy. For work-conserving policies, we have

$$\forall k: \quad \phi_k(x) + \phi_{k,k+1}(x) + \phi_{k-1,k}(x) = 1. \tag{III.19}$$

For the sake of simplicity, we assume that users in each zone share the allocated radio resources equally, independently of their radio conditions. When a user in a coordination zone $k, k+1$ is served, all users in all coordination zones are blocked and thus,

$$\sum_{k=1}^{K} \phi_{k,k+1}(x) \leq 1. \tag{III.20}$$

We study different resource sharing strategies between CoMP and non-CoMP users. Note that when a user in a coordination zone $k, k+1$ is served, users in non-coordination zones k and $k+1$ are blocked and vice versa.

While the sufficient stability conditions depend on the scheduling policy, the necessary stability conditions in the absence of mobility are given by:

$$\begin{cases} \forall k: \quad \rho_k + \rho_{k,k+1} + \rho_{k-1,k} < 1 \\ \sum_{k=1}^{K} \rho_{k,k+1} \leq 1 \end{cases} \tag{III.21}$$

Let

$$\rho = \lambda/\mu$$
$$= \max\left(\max_k(\rho_k + \rho_{k,k+1} + \rho_{k-1,k}), \sum_{k=1}^{K} \rho_{k,k+1} \right).$$

We shall see in the following sections that mobility may increase or decrease the stability condition, depending on the scheduling strategy.

a) Model without mobility

i. Static allocation CoMP users' data flows are allocated some fixed fraction ϕ_0 of the resources of each coordinated sector. In particular, CoMP users' data flows in zone $k, k+1$ are allocated the fraction:

$$\phi_{k,k+1} = \frac{x_{k,k+1}}{\sum_{j=1}^{K} x_{j,j+1}} \phi_0.$$

Non-CoMP users' data flows in sector k are allocated the fraction $\phi_k = 1 - \phi_0$ of the considered sector radio resources. From (III.7) and (III.8), the mean throughput in zone k is

$$\gamma_k = \phi_k \mu_k \sigma(1 - \rho_k/\phi_k) \quad k = 1 \ldots K$$

while the mean throughput in zone $k, k+1$ is:

$$\gamma_{k,k+1} = \phi_0 \mu_{k,k+1} \sigma\left(1 - \sum_{j=1}^{K} \rho_{j,j+1}/\phi_0\right).$$

Observe that the mean throughput in zone k decreases *linearly* from the physical rate $\phi_k\mu_k\sigma$ to 0 when the load of the corresponding zone increases from 0 to ϕ_k. Indeed, the stability condition of zone k is $\rho_k < \phi_k$. The stability condition is given by:

$$\begin{cases} \forall k: \quad \rho_k < 1-\phi_0 \\ \sum_{k=1}^{K} \rho_{k,k+1} < \phi_0 \end{cases} \tag{III.22}$$

Observe that this condition may be more restrictive than the natural stability condition (III.21).

ii. Iterative scheduler Under the iterative scheduler [43], users in zone k are allocated the fraction

$$\phi_k(x) = \frac{x_k + x_{k+1,k-1}}{\bar{x}} + \frac{x_{k+1}}{\bar{x}} \frac{x_k + x_{k-1}}{x_k + x_{k-1,k} + x_{k-1}} + \frac{x_{k-1}}{\bar{x}} \frac{x_k + x_{k+1}}{x_k + x_{k,k+1} + x_{k+1}}$$

of sector k resources while users in zone $k, k+1$ are allocated the fraction

$$\phi_{k,k+1}(x) = \frac{x_{k,k+1}}{\bar{x}} + \frac{x_{k-1}}{\bar{x}} \frac{x_{k,k+1}}{x_k + x_{k,k+1} + x_{k+1}}$$

of each coordinated sector. We denote by

$$\bar{x} = \sum_{k=1}^{K} \left(x_k + x_{k,k+1}\right)$$

the total number of active users in the site. Indeed, a centralized scheduler should be implemented in each site. For instance, if a user in zone 1 is selected to be served by the scheduling algorithm, the scheduler can either select a user in zone 2 and a user in zone 3 or select a user in zone 23 to be served simultaneously. Similarly, if a user in zone 12 is selected for instance, another user in zone 3 is served simultaneously. The stability condition is given by (III.21). When

$$\max_k(\rho_k + \rho_{k,k+1} + \rho_{k-1,k}) > \sum_{k=1}^{K} \rho_{k,k+1},$$

explicit expressions for the flow throughputs can be obtained by approximation. We proceed by decoupling the different zones as in [43]. We consider zone k, zone $k+1$ and zone $k, k+1$ and neglect the other zones after taking into account the load induced by zone $k-1, k$ on cell k and that induced by zone $k+1, k-1$ on cell $k+1$. Thus the flow throughput in zone k is approximately:

$$\gamma_k \approx \mu_k\sigma(1-\rho_k-\rho_{k,k+1}-\rho_{k-1,k}) \tag{III.23}$$

while the flow throughput in zone $k, k+1$ is approximately:

$$\gamma_{k,k+1} \approx \mu_{k,k+1}\sigma \left(\sum_{j=k}^{k+1} \frac{1}{1-\rho_{j,j+1}-\rho_{j-1,j}-\rho_j} - \frac{1}{1-\rho_{k,k+1}} \right)^{-1}. \tag{III.24}$$

iii. Priority to non-CoMP users Under this policy, non-CoMP users in each sector are scheduled first and are allocated all the radio resources whenever active: in any state x such that $x_k > 0$ $\phi_k = 1$, that is

$$\bar{\phi}_k = \rho_k.$$

CoMP users in zone $k, k+1$ wait until resources in zones k and $k+1$ become available, these users equally share resources with other CoMP users in the other coordination zones. The fraction of time spent by the scheduler on users in zone $k, k+1$ is:

$$\phi_{k,k+1} = \begin{cases} \frac{x_{k,k+1}}{\sum_{j=1}^{K} x_{j,j+1}} & \text{if } \sum_{k=1}^{K} x_k = 0. \\ 1 & \text{if } \sum_{k=1}^{K} x_k > 0 \,\&\, x_k + x_{k+1} = 0. \\ 0 & \text{otherwise.} \end{cases}$$

Thus the mean throughput in non-coordination zone k is:

$$\gamma_k = \mu_k \sigma (1 - \rho_k).$$

However there is no explicit expression for the mean throughput in the coordination zones. The stability condition is given by:

$$\forall k : \rho_{k,k+1} < (1-\rho_k)(1-\rho_{k+1})\left(\rho_{k-1} + (1-\rho_{k-1})\theta_{k,k+1}\right)$$

such that $\sum_{k=1}^{K} \theta_{k,k+1} < 1$, we get

$$\sum_{k=1}^{K} \left(\rho_{k,k+1} - \rho_k \prod_{j \neq k} (1-\rho_j) \right) < \prod_{k=1}^{K} (1-\rho_k), \tag{III.25}$$

which is more restrictive than the natural stability condition (III.21). This is due to the fact that some cells may be idle when serving non-CoMP users.

iv. Priority to CoMP users Under this policy, CoMP users are scheduled first and are allocated all radio resources whenever active: users in zone $k, k+1$ are allocated the fraction: $\phi_{k,k+1} = x_{k,k+1} / \sum_{j=1}^{K} x_{j,j+1}$. Non-CoMP users in zone k are served only when there are no active CoMP users in zones $k, k+1$ and $k-1, k$. The stability condition is the natural condition given by (III.21). Since CoMP users are not affected by non-CoMP users the mean throughput in the coordination zone $k, k+1$ is given by:

$$\gamma_{k,k+1} = \mu_{k,k+1} \sigma (1 - \sum_{j=1}^{K} \rho_{j,j+1}).$$

b) Model with mobility

i. Stability condition The stability condition follows from the limiting regime of infinite mobility where:

$$\begin{cases} \forall (k,j,i,n) : \nu^{(i,n)}_{k \to j,j+1}, \nu^{(n,i)}_{j,j+1 \to k}, \nu^{(i,n)}_{k \to j}, \nu^{(i,n)}_{k,k+1 \to j,j+1} \to \infty \\ \forall (k,i,n) : \quad \nu^{(i,n)}_{k} \to \infty \end{cases}$$

In this regime the mean service rate in zone k becomes

$$\bar{\mu}_k = \sum_{i=1}^{N_k} q_k^{(i)} \mu_k^{(i)} / q_k.$$

Similarly, the mean service rate in zone $k, k+1$ is:

$$\bar{\mu}_{k,k+1} = \sum_{i=1}^{N_{k,k+1}} q_{k,k+1}^{(i)} \mu_k^{(i)} / q_{k,k+1}.$$

Considering only intra-zone mobility without inter-zone mobility, the stability condition is:

$$\begin{cases} \forall k: \quad \mu_k\rho_k/\bar{\mu}_k + \mu_{k,k+1}\rho_{k,k+1}/\bar{\mu}_{k,k+1} \\ \qquad\qquad +\mu_{k-1,k}\rho_{k-1,k}/\bar{\mu}_{k-1,k} < 1 \\ \\ \sum_{k=1}^{K} \mu_{k,k+1}\rho_{k,k+1}/\bar{\mu}_{k,k+1} \leq 1 \end{cases}$$

under the IT and Pri-C schemes,

$$\begin{cases} \forall k: \quad \mu_k\rho_k/\bar{\mu}_k < 1-\phi_0 \\ \sum_{k=1}^{K} \mu_{k,k+1}\rho_{k,k+1}/\bar{\mu}_{k,k+1}/ < \phi_0 \end{cases}$$

under the SA scheme and

$$\sum_{k=1}^{K}\left(\frac{\mu_{k,k+1}}{\bar{\mu}_{k,k+1}}\rho_{k,k+1} - \frac{\mu_k}{\bar{\mu}_k}\rho_k\prod_{j\neq k}\left(1-\frac{\mu_j}{\bar{\mu}_j}\rho_j)\right)\right) < \prod_{k=1}^{K}\left(1-\frac{\mu_k}{\bar{\mu}_k}\rho_k\right)$$

under the Pri-NC scheme. However, in the presence of inter-zone mobility the stability condition becomes

$$\rho < \bar{\mu}/\mu,$$

where $\bar{\mu}$ is the overall mean service rate. The precise mean service rate depends on the scheduling strategy, where

$$\bar{\mu} = \sum_{k=1}^{K}(1-\phi_0)\bar{\mu}_k + \phi_0\frac{q_{k,k+1}}{q_0}\bar{\mu}_{k,k+1}$$

under the SA scheme,

$$\bar{\mu} = \frac{1}{q_0}\sum_{k=1}^{K} q_{k,k+1}(\bar{\mu}_{k,k+1} + \bar{\mu}_{k-1})$$

under the Pri-C scheme,

$$\bar{\mu} = \sum_{k=1}^{K}\bar{\mu}_k$$

under the Pri-NC scheme and

$$\bar{\mu} = \sum_{k=1}^{K}\phi_k(q)\bar{\mu}_k + \phi_{k,k+1}(q)\bar{\mu}_{k,k+1}$$

under the IT scheduler. Note that

$$\phi_k(q) = q_k + q_{k+1,k-1} + q_{k+1}\frac{q_k+q_{k-1}}{q_k+q_{k-1,k}+q_{k-1}} + q_{k-1}\frac{q_k+q_{k+1}}{q_k+q_{k,k+1}+q_{k+1}},$$

while

$$\phi_{k,k+1}(q) = q_{k,k+1} + q_{k-1}\frac{q_{k,k+1}}{q_k+q_{k,k+1}+q_{k+1}}.$$

ii. Throughput in light traffic The performance in light traffic (that is, when the load is approximately zero) is the same for all policies except the static allocation: a user when alone in the system is always allocated all radio resources. We denote by $T_k^{(i)}$ the mean duration of a flow initiated in region i of zone k:

$$T_k^{(i)}\left(\mu_k^{(i)}+\sum_{\substack{n\le N_k\\ n\neq i}}\nu_k^{(i,n)}+\sum_{\substack{n\le N_j\\ j\neq k}}\nu_{k\to j}^{(i,n)}+\sum_{\substack{n\le N_{j,j+1}\\ j\le K}}\nu_{k\to j,j+1}^{(i,n)}\right)$$
$$=1+\sum_{\substack{n\le N_k\\ n\neq i}}\nu_k^{(i,n)}T_k^{(n)}+\sum_{\substack{n\le N_j\\ j\neq k}}\nu_{k\to j}^{(i,n)}T_j^{(n)}+\sum_{\substack{n\le N_{j,j+1}\\ j\le K}}\nu_{k\to j,j+1}^{(i,n)}T_{j,j+1}^{(n)}, \quad \text{(III.26)}$$

and by $T_{k,k+1}^{(i)}$ the mean duration of a flow initiated in region i of zone $k,k+1$:

$$T_{k,k+1}^{(i)}\mu_{k,k+1}^{(i)}+T_{k,k+1}^{(i)}\sum_{\substack{n\le N_{k,k+1}\\ n\neq i}}\nu_{k,k+1}^{(i,n)}+T_{k,k+1}^{(i)}\sum_{\substack{n\le N_j\\ j\le K}}\nu_{k,k+1\to j}^{(i,n)}+T_{k,k+1}^{(i)}\sum_{\substack{n\le N_{j,j+1}\\ j\neq k}}\nu_{k,k+1\to j,j+1}^{(i,n)}$$
$$=1+\sum_{\substack{n\le N_{k,k+1}\\ n\neq i}}\nu_{k,k+1}^{(i,n)}T_{k,k+1}^{(n)}+\sum_{\substack{n\le N_j\\ j\le K}}\nu_{k,k+1\to j}^{(i,n)}T_j^{(n)}+\sum_{\substack{n\le N_{j,j+1}\\ j\neq k}}\nu_{k,k+1\to j,j+1}^{(i,n)}T_{j,j+1}^{(n)}. \quad \text{(III.27)}$$

We deduce the mean duration of a flow initiated in the considered site:

$$T=\sum_{k=1}^{K}\left(\sum_{i=1}^{N_k}p_k^{(i)}T_k^{(i)}+\sum_{n=1}^{N_{k,k+1}}p_{k,k+1}^{(n)}T_{k,k+1}^{(n)}\right),$$

and the average throughput in light traffic: $\gamma=\sigma/T$.
$T_k^{(i)}$ and $T_{k,k+1}^{(j)}$ follow from the linear system of equations (III.26) and (III.27). For the static allocation strategy, we get the mean throughput similarly by replacing $\mu_k^{(i)}$ and $\mu_{k,k+1}^{(i)}$ by $(1-\phi_0)\mu_k^{(i)}$ and $\phi_0\mu_{k,k+1}^{(i)}$ respectively.

c) Model with multiple mobility behaviors

i. Mobility-aware scheduling Under the mobility-aware scheduler, non-CoMP users and static CoMP users are served first and then mobile CoMP users who are served in the increasing order of speed. Users in zone k are allocated the fraction:

$$\phi_k(x)=\frac{x_k+x_{k+1,k-1}^{(0)}}{\bar{x}}+\frac{x_{k+1}}{\bar{x}}\frac{x_k+x_{k-1}}{x_k+x_{k-1,k}^{(0)}+x_{k-1}}+\frac{x_{k-1}}{\bar{x}}\frac{x_k+x_{k+1}}{x_k+x_{k,k+1}^{(0)}+x_{k+1}}$$

of sector k resources. The share of class-v non-CoMP users ($v\ge 0$) is:

$$\phi_k^{(v)}(x)=\frac{x_k^{(v)}}{x_k}\phi_k(x).$$

Class-v CoMP users of zone $k, k+1$ are allocated the fraction

$$\phi^{(v)}_{k,k+1}(x) = \begin{cases} \frac{x^{(v)}_{k,k+1} \times x_{k-1}/\bar{x}}{x_k + x^{(v)}_{k,k+1} + x_{k+1}} + \frac{x^{(v)}_{k,k+1}}{x} & \textbf{if } v = 0 \\ \frac{x^{(v)}_{k,k+1}}{\sum_j x^{(v)}_{j,j+1}} & \textbf{if } \sum_j \left(x_j + \sum_{s<v} x^{(s)}_{j,j+1} = 0 \right) \\ & \quad v > 0 \\ \phi_{k-1}(x) & \textbf{if } x_k + x_{k+1} + \sum_{s<v} x^{(s)}_{k,k+1} = 0 \\ & \quad x_{k-1} > 0 \\ & \quad v > 0 \\ 0 & \textbf{otherwise.} \end{cases}$$

of each coordinating sector (sector k and sector $k+1$). Observe that the network operates under the IT scheduling strategy for the static users and under the Pri-NC users strategy for the mobile users. These shares of each class of users are used in order to compute the stationary distribution $\pi(x)$ of the Markov process $X(t)$, given the arrival rate, the service rate and the mobility rates in each region of each zone.

III.3.4.3 A single cell cooperating with its surrounding neighboring cells

a) Stability condition analysis

We study in this section the stability condition of a cell cooperating with its surrounding neighboring cells, given that JT is activated in the network. We denote by β_n the mean coordination gain of a user located at the cell edge involving n cells in the transmission, that is the mean rate gain of a CoMP user compared to the case when served without coordination. Consequently, the mean service rate of a CoMP user involving n cells in the transmission is nothing more than $\beta_n \mu_{\text{edge,n}}$.

Observe that when activating JT the traffic at the edge of a cell cooperating with its neighboring cells increases as shown in Figure III.4. All CoMP users associated with neighboring cells and requiring transmission from the considered cell are generating new traffic to the cell. Note that the cell traffic in each particular cell edge zone involving n cells in the transmission increases by a factor:

$$\alpha_n = (p_{\text{edge,n}} + p'_{\text{edge,n}})/p_{\text{edge,n}}.$$

Users located at the edge of this cell benefit from the cooperation so they can complete more quickly their service. So there is a trade-off between the additional traffic burden and the performance improvement of cell edge users. If the rate gain brought to cell edge users is high enough in order to compensate the extra traffic burden, the stability condition of the cell can be maintained as in the case without coordination or can even be improved.

When activating JT, the new cell load under RR iterative scheduler becomes as follows:

$$\rho' = \frac{p_{\text{center}}\lambda}{\mu_{\text{center}}} + \sum_n \frac{p_{\text{edge,n}}\lambda + p'_{\text{edge,n}}\lambda}{\beta_n \mu_{\text{edge,n}}} = \frac{p_{\text{center}}\lambda}{\mu_{\text{center}}} + \sum_n \frac{\alpha_n}{\beta_n} \frac{p_{\text{edge,n}}\lambda}{\mu_{\text{edge,n}}} \tag{III.28}$$

Observe that in order to maintain same cell load given by (III.9), the following condition should be fulfilled:

$$\sum_n \frac{\alpha_n}{\beta_n} \frac{p_{\text{edge,n}}}{\mu_{\text{edge,n}}} = \sum_n \frac{p_{\text{edge,n}}}{\mu_{\text{edge,n}}}.$$

Thus, one way to avoid the stability condition degradation is to guarantee that:

$$\beta_n \geq \alpha_n \qquad \forall n.$$

In other words, the involvement of a given cell in the cooperation increases its traffic in a particular zone at the edge by a factor α_n. Whenever this traffic can be served α_n times faster there is no impact on the capacity of the cell.

In the case of a homogeneous symmetric network, we can assume that $\alpha_n = n$. Thus the mean rate gain of a CoMP user involving n cells in the transmission should be at least equal to n:

$$\beta_n \geq n \qquad \forall n \tag{III.29}$$

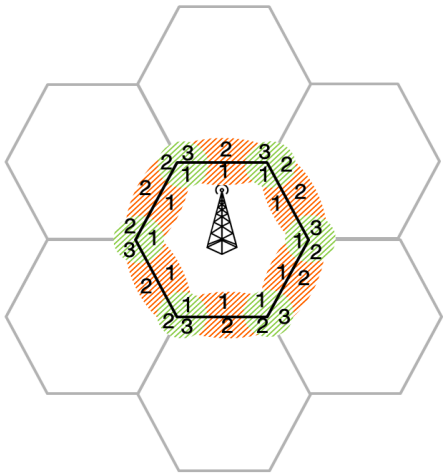


Figure III.19: Traffic increase at the edge.

Consider the case of a hexagonal topology for instance, as illustrated by Figure III.19. Observe that orange area where two cells are involved in the transmission, increases by a factor of two. Green area where three cells are involved in the transmission increases by a factor of three.

Note that this scenario (a single cell surrounded by its neighboring cells) represents the dynamic clustering case.

The static clustering case with three coordinated cells: intra-site, inter-site as well as intra & inter-site coordination, corresponds to the model described in Section III.2.2. Note that in the case of static clustering there is an additional stability condition: the load induced by all CoMP users in the cluster should be less than one, as given by (III.21). However, in general this stability condition is less restrictive than the general one (III.28).

b) Impact of interference on the coordination gain

Consider the case of only one cooperating cell, that is $\mathbb{C} = \{c\}$. The mean coordination gain in this case should be at least equal to the number of cells involved in the transmission, that is $n = 2$.
We denote by $I = \sum_{i \neq c} P_i$ the interference received by the user from all neighboring non-cooperating

cells. The signal-to-noise ratio in the absence of coordination is in this case:

$$SINR = \frac{P_s}{P_c + I + \mathcal{N}}$$

We evaluate in the following two sections both transmission schemes: Scheme 1 and Scheme 2.

i. Scheme 1 Suppose that Scheme 1 is used. By making c cooperate the new SINR becomes:

$$SINR' = \frac{P_s + P_c}{I + \mathcal{N}}$$

let G be the SINR gain:

$$\frac{SINR'}{SINR} = G$$

As the given user is using two resources, the new service rate should be two times more than the original service rate according to (III.29) , that is:

$$\frac{\log_2(1 + SINR')}{\log_2(1 + SINR)} \geq 2$$

$$\frac{\log_2(1 + G \times SINR)}{\log_2(1 + SINR)} \geq 2$$

The SINR gain is then given by:

$$G \geq SINR + 2$$

$$\frac{SINR'}{SINR} \geq SINR + 2$$

let $\zeta = P_c/P_s$ and $\upsilon = (I + \mathcal{N})/P_s$. Note that $0 \leq \zeta \leq 1$ while $\upsilon \geq 0$. We get the following in-equation:

$$\frac{1 + \zeta}{\upsilon} \times (\zeta + \upsilon) \geq \frac{1}{\zeta + \upsilon} + 2.$$

This in-equation is verified when the interference factor υ do not exceed a certain value which is a function of ζ: the ratio of the power received from the cooperating cell to the power received from the serving cell. the solution of this in-equation can be written as follows:

$$\upsilon \leq \frac{\zeta^2}{1 - \zeta} \tag{III.30}$$

ii. Scheme 2 We suppose now that Scheme 2 is used. The new service rate should be two times more than the original service rate:

$$\log_2\left(1 + \frac{P_s}{I + \mathcal{N}}\right) + \log_2\left(1 + \frac{P_c}{I + \mathcal{N}}\right) \geq 2\log_2(1 + SINR)$$

that is,

$$\left(1 + \frac{1}{\upsilon}\right)\left(1 + \frac{\zeta}{\upsilon}\right) - \left(1 + \frac{1}{\upsilon + \zeta}\right)^2 \geq 0 \tag{III.31}$$

Figure III.20 shows the solution (gray areas) for in-equations (III.30) that corresponds to transmission scheme 1 and in-equation (III.31) that corresponds to transmission scheme 1.

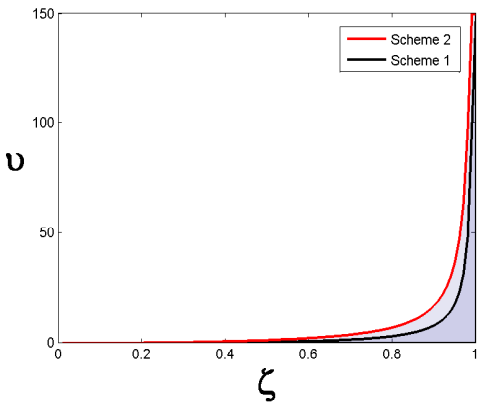


Figure III.20: Interference region for 100% cooperation gain with respect to υ

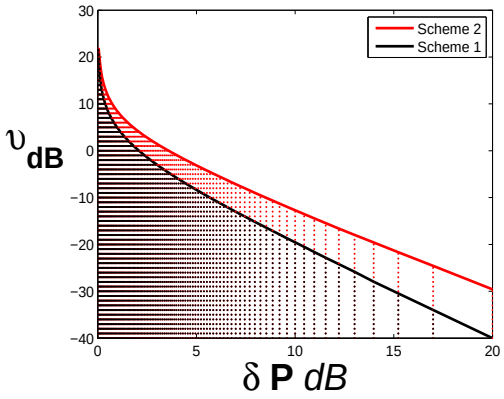


Figure III.21: Interference region for 100% cooperation gain with respect to δP_{dB}

Note that

$$\delta P_{dB} = Ps - Pc = 10\log_{10}(1/\zeta),$$

while

$$\upsilon_{dB} = 10\log_{10}(\upsilon).$$

Figure III.21 shows the interference region represented by υ_{dB} that corresponds to 100% cooperation gain with respect to δP_{dB}. Observe that the coordination gain depends also on the residual interference level (mainly caused by neighboring non-cooperating cells) beyond the power threshold δP. The higher the interference, the closer should be the power received from the cooperating cell to the power received from the serving cell in order to guarantee 100% coordination gain. This leads to the following conclusion: it is not worth to allow one cell to cooperate if the user is located in a zone which is strongly interfered by another cell (or several other cells). In this case, it is better that all strongly interfering cells cooperate or non of them cooperates.

III.3.5 Numerical results

III.3.5.1 Low interference case

We consider in this section a low interference scenario with moderate mean coordination gain which is lower than 100%. We consider in particular the case of $\delta P = 12$ dB in a beamforming system where the mean coordination gain is around 50% and the mean coordination area of each sector is around 30%.

a) Single coordination zone

i. Case without mobility We assume that there is only one region in each zone. The mean coordination area in each cell should be around 30%, thus we consider that $p_0 = p_1 = p_2 = 1/3$ with $\mu_1 = 2$, $\mu_2 = 3/2$ and $\sigma = 1$. We take $\phi_0 = 1/2$ for the static allocation (SA).

Figure III.22 compares the throughput performance of the four policies for $K = 2$ coordinated cells. Observe that iterative scheduler (IT) outperforms the three other strategies. The priority to

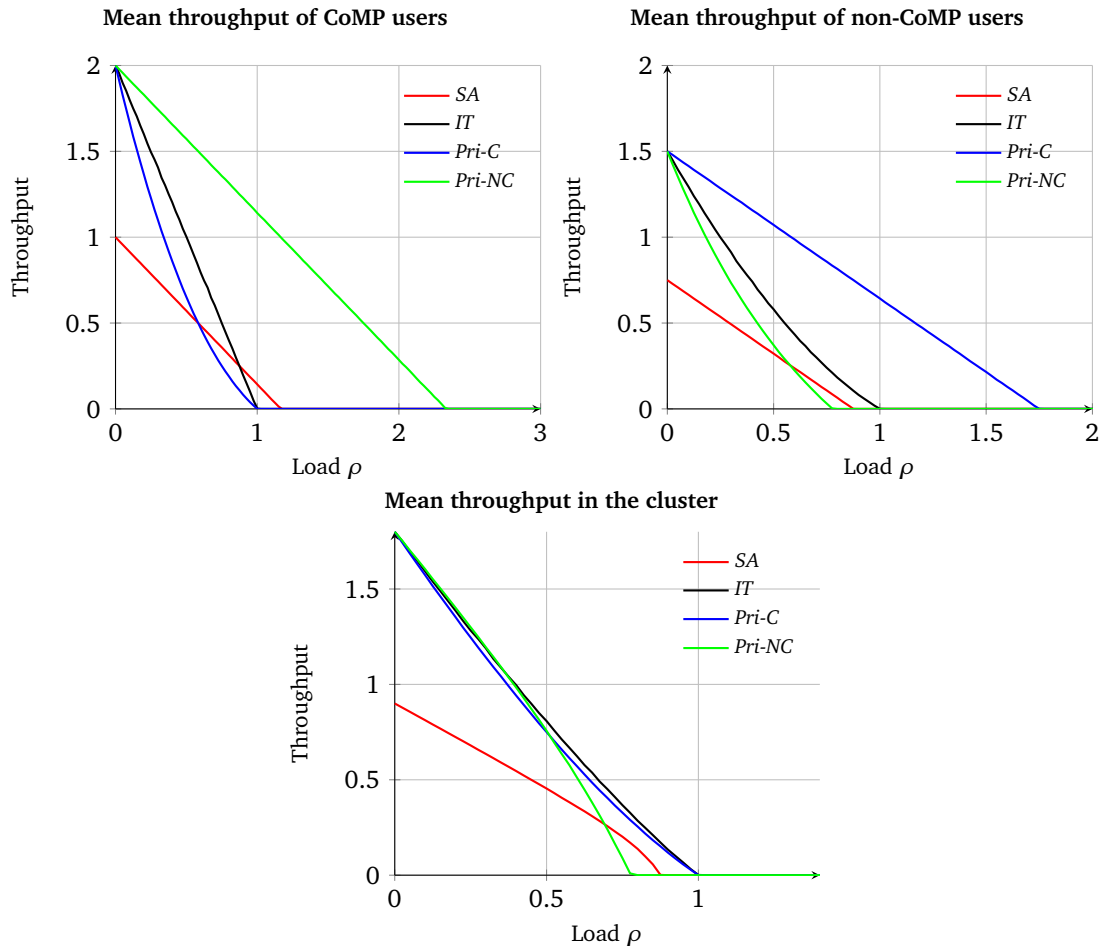


Figure III.22: Throughput performance for two coordinated cells without mobility, single coordination zone.

CoMP (Pri-C) strategy largely outperforms the priority to non-CoMP (Pri-NC) strategy at high load. In terms of stability, the Pri-NC strategy is the worst. This is due to the fact that prioritizing Non-CoMP users adversely affects the stability condition of CoMP users. The SA policy benefits from low complexity at the expense of degraded performance.

The maximum load is approximately equal to 0.77 under the Pri-NC scheme and to 0.88 under the SA scheme, in accordance with the presented results.

ii. Impact of mobility Consider the same scenario of two coordinated cells as in the previous section, with $\nu_{01} = \nu_{10} = \nu_{02} = \nu_{20} = 1$. The results are obtained by the numerical evaluation of the stationary distribution of the Markov process $X(t)$ and shown in Figure III.23.

We observe that inter-zone mobility improves the mean throughput in the cluster for all strategies except for the Pri-C strategy. Under this strategy, mobility leads to a throughput degradation both for non-CoMP users and for CoMP users. This is due to the fact that non-CoMP users may then suffer from bad radio conditions (served with CoMP mode) leading to extra waste of resources. Based on this observation, the following conclusion may be drawn: it is not worth losing a resource to serve a CoMP user if that user is moving to the center and could benefit from better channel conditions. The

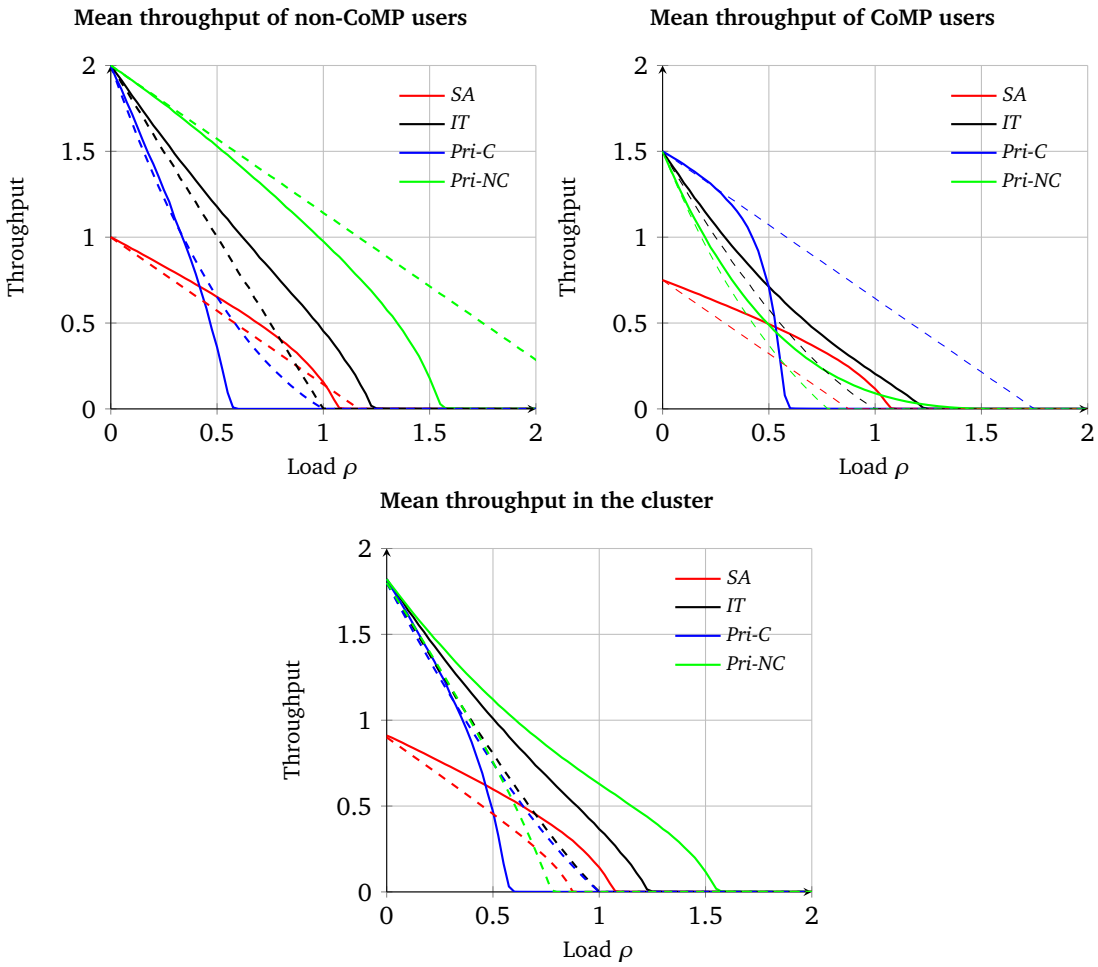


Figure III.23: Throughput performance for two coordinated cells with mobility (solid line) and without mobility (dashed line), single coordination zone.

strategy of prioritizing non-CoMP users proves to be the best strategy when users move. This can be explained by the fact that prioritizing Non-CoMP users gives the chance to cell-edge users to move and to be served in better radio conditions where coordination is not required. However, according to the numerical and analytical results the IT strategy remains a good compromise in a network where the users' mobility cannot be predicted.

In accordance with the presented results, the maximum loads are approximately equal to 1.07, 1.23, 1.55 and 0.6 under the SA, IT, Pri-NC and Pri-C strategies, respectively. This shows that in terms of stability Pri-NC is the best strategy while Pri-C is the worst.

b) Intra-site coordination

i. Case without mobility The mean coordination area of each sector is around 30%, that is $p_1 = p_2 = p_3 = 7/30$ and $p_{12} = p_{13} = p_{23} = 1/10$. We assume that there is only one region in each zone with $\mu_1 = \mu_2 = \mu_3 = 2$ and $\mu_{12} = \mu_{13} = \mu_{23} = 1.5$ and we consider $\phi_0 = 1/2$ for the static allocation (SA). Figure III.24 (dashed lines) compares the throughput performance of the four policies. Observe

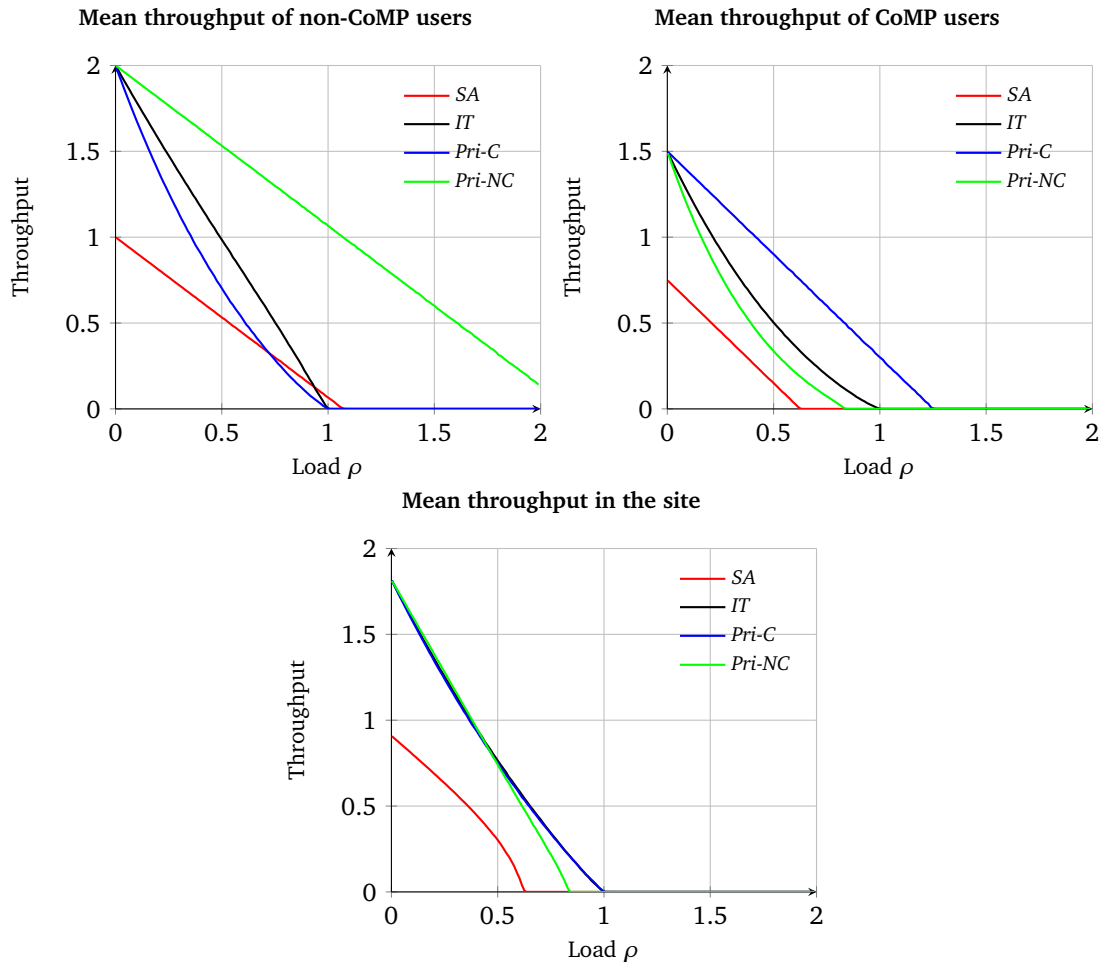


Figure III.24: Throughput performance without mobility, intra-site coordination scenario.

that the iterative scheduler (IT) outperforms the three other strategies. The priority to non-CoMP (Pri-NC) strategy leads to degraded performance. These results show that in the absence of mobility a fair strategy is the best strategy to be applied. The maximum load is approximately equal to 0.83 under the Pri-NC scheme and to 0.625 under the SA scheme, in accordance with (III.22) and (III.25).

ii. Impact of mobility Consider the same scenario as in the previous section, with $\nu_{k-1,k\to k} = \nu_{k\to k,k+1} = 1\ \forall k$. All other mobility rates are supposed equal to zero. The results are obtained by the numerical evaluation of the stationary distribution of the Markov process $X(t)$ and shown in Figure III.25 (solid lines). Inter-zone mobility improves the mean throughput in the site for all strategies except for the Pri-C strategy. Under this strategy, mobility leads to a throughput degradation. However, prioritizing non-CoMP users in the presence of mobility outperforms all other strategies in sharp contrast to the scenario without mobility. This is due to the fact that cell edge users are more likely served in good radio conditions without performing CoMP operations, minimizing the waste of resources. Based on this observation, the following conclusion may be drawn: in low coordination gain scenario, it is not worth losing a resource to serve a CoMP user if that user is moving and can benefit

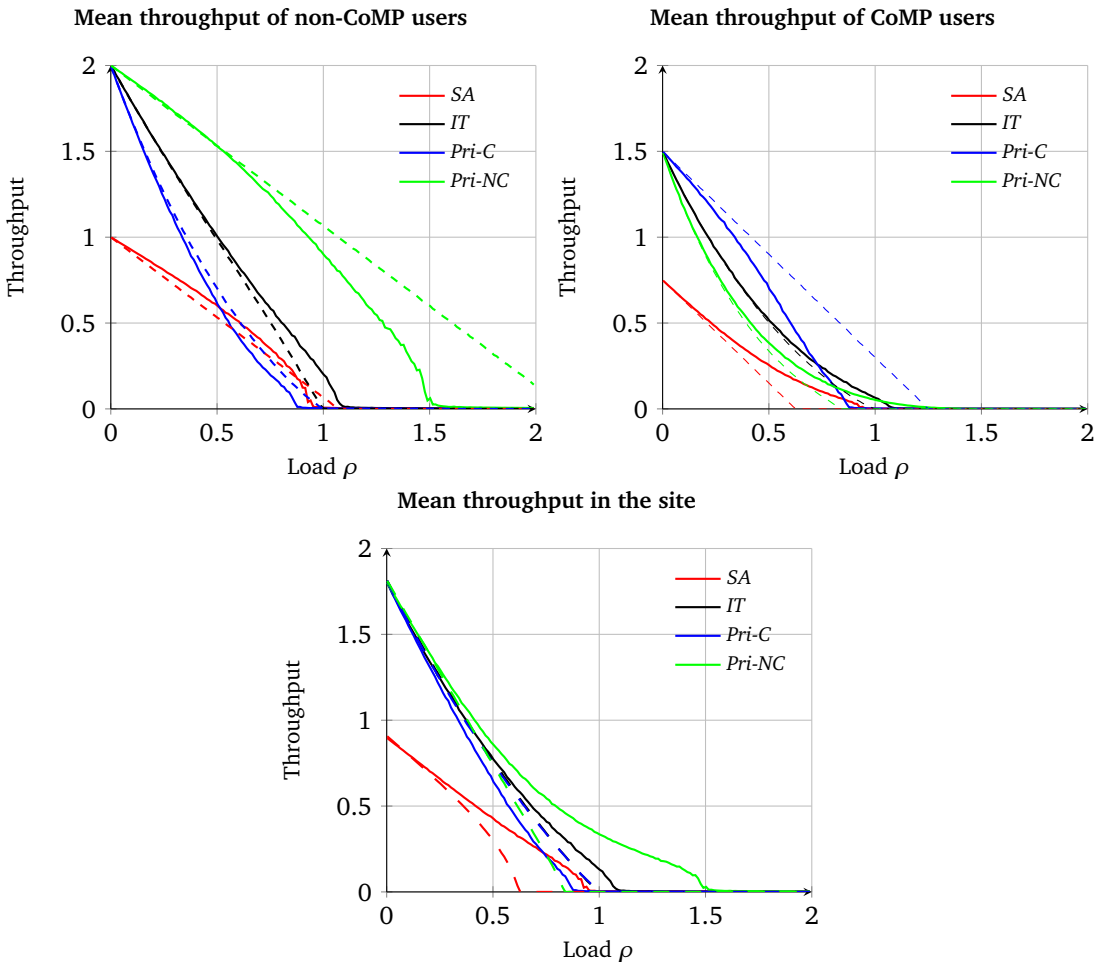


Figure III.25: Throughput performance with mobility (solid line) and without mobility (dashed line), intra-site coordination scenario.

from better radio conditions. So if predicted accurately, mobility can be an interesting property which can include the scheduling strategy. Following this conclusion, we will introduce a more advanced mobility-aware scheduler in the next section.

iii. Mobility aware scheduler The numerical evaluation of the stationary distribution of the Markov process $X(t)$ is shown in Figure III.26 for the same previous scenario but with two classes of users: a static class and a mobile class: $w_0 = w_1 = 50\%$. Results show that the mobility aware scheduler improves the performance as well as the stability condition. Observe that there is an increase of approximately 15% of the stability condition compared to the IT strategy. The reason is that the proposed scheduler, by combining the IT (the best strategy for static users see Figure III.25) and the Pri-NC (the best strategy for mobile users see Figure III.25) strategies, can achieve the best stability condition for each mobility class.

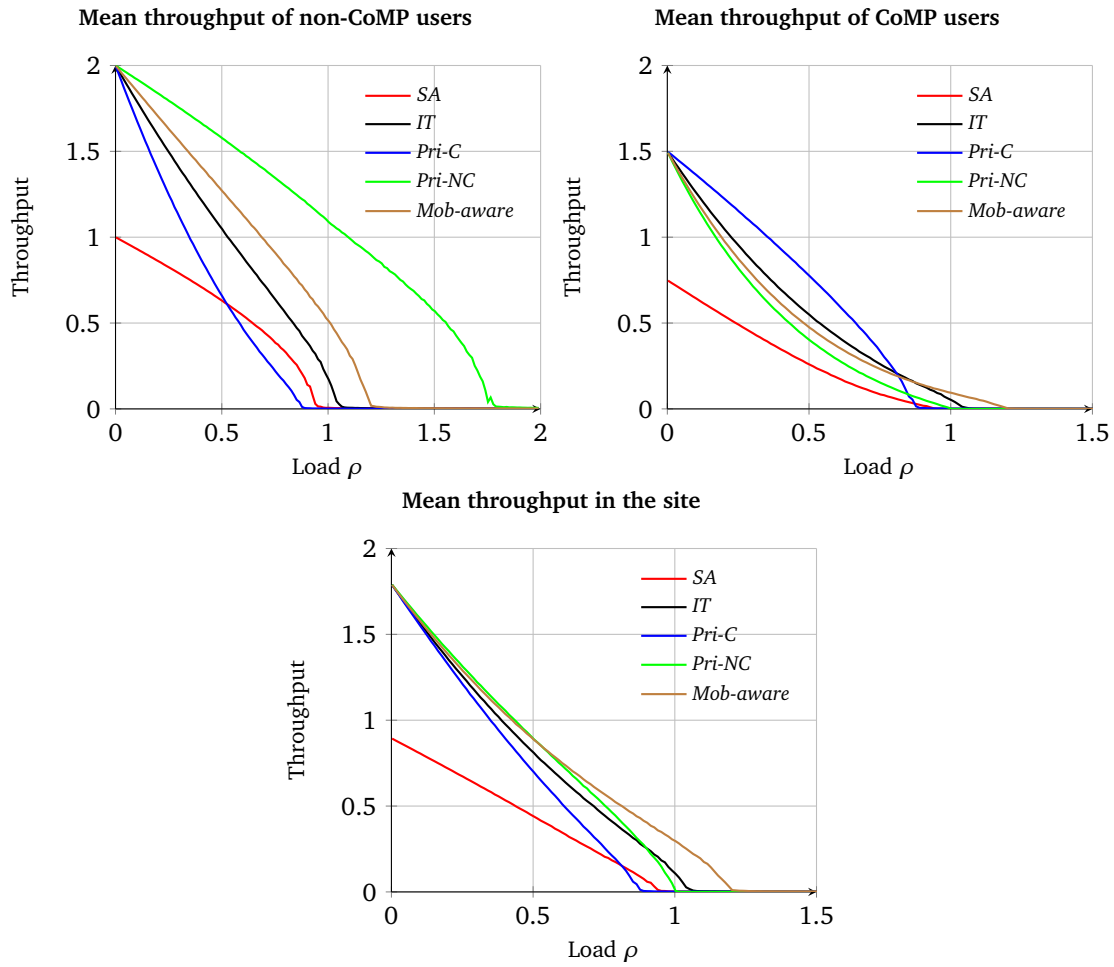


Figure III.26: Throughput performance with two classes of mobility.

III.3.5.2 High interference case

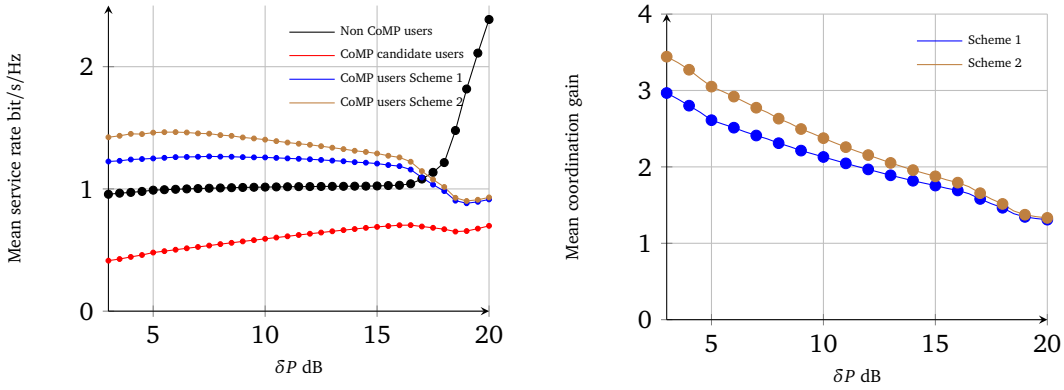
a) Mean parameters evaluation

Figure III.27, Figure III.28 and Figure III.29 show the mean service rates as well as the mean coordination gain under different clustering strategies and different maximum coordinated cells ($K_{max} = 2$ and $K_{max} = 3$), with respect to δP, considering a regular hexagonal topology.

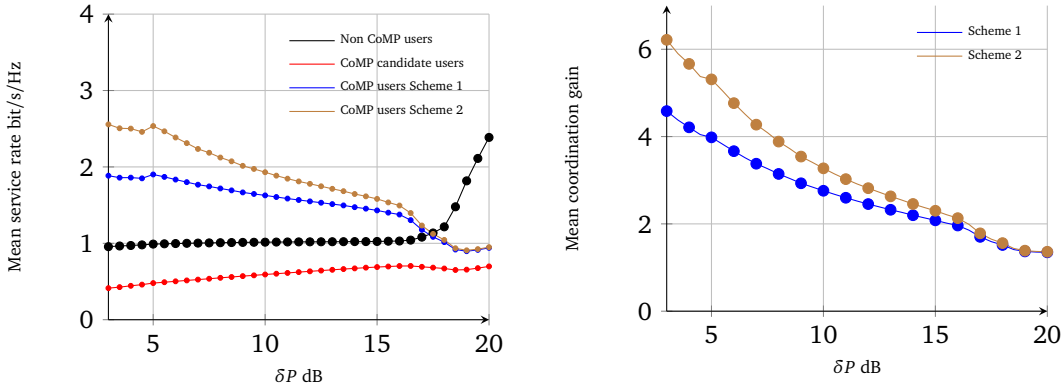
The mean service rate of non-CoMP users is that of those users who are not located in coordination zones. This service rate is evaluated through the harmonic mean of different peak rates that can get a user when allocated all resources in different positions of the cell, located in non coordination zones.

The mean service rate of CoMP candidate users is that of users located in coordination zones when served without coordination. It is also evaluated through a harmonic mean.

The mean service rate of CoMP users corresponds to the mean service rate of those users located in coordination zones when served in CoMP mode, involving multiple cells in the transmission. Note that the mean service rate of CoMP users depends on the transmission scheme: Scheme 1 or Sscheme 2. The mean coordination gain is the mean gain brought to a CoMP user. It is nothing more than



(a) $K_{max} = 2$ maximum coordinated cells



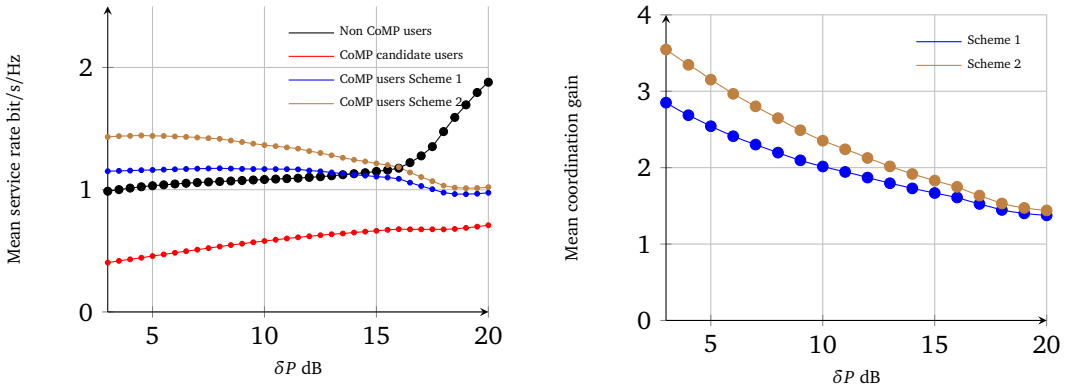
(b) $K_{max} = 3$ maximum coordinated cells

Figure III.27: Mean service rate and mean coordination gain for Intra-site coordination.

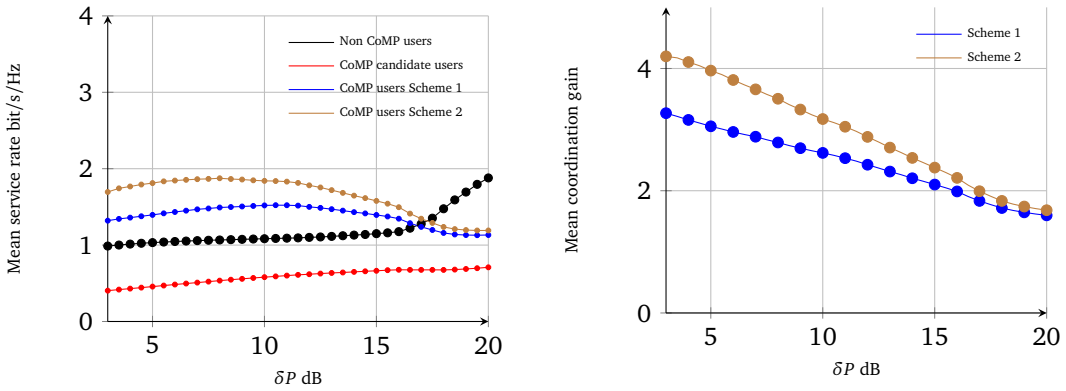
the ratio of the mean service rate of CoMP users to the mean service rate of CoMP candidate users. Observe that this gain decreases with the power threshold δP.

Note that the different service rates are evaluated at full interference, supposing that all neighboring cells are transmitting all the time. These parameters are then used in order to evaluate the stability condition in the different cases. They are also used as input parameters to the model explained in Section III.2.1 in order to evaluate the performance in terms of mean user throughput.

Note that when $K_{max} = 2$, there is no stability condition degradation, under a round robin scheduler, whenever the mean coordination gain is higher then 2, according to (III.29). Thus, no stability degradation is expected in the case of dynamic clustering.

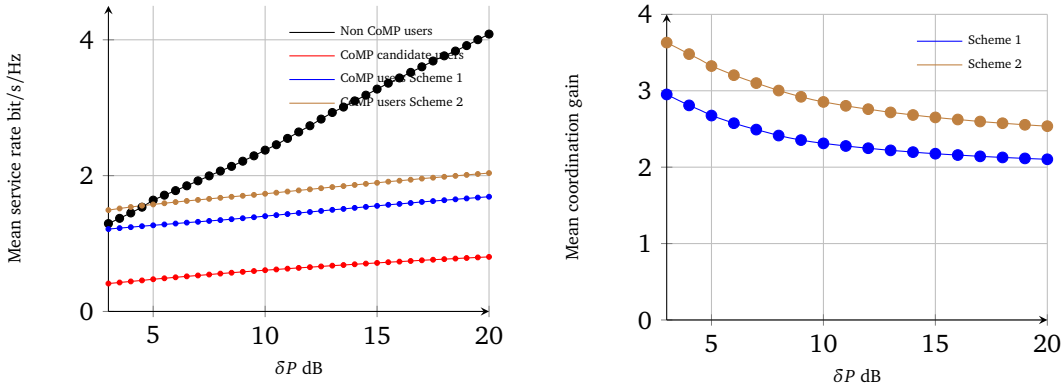


(a) $K_{max} = 2$ maximum coordinated cells

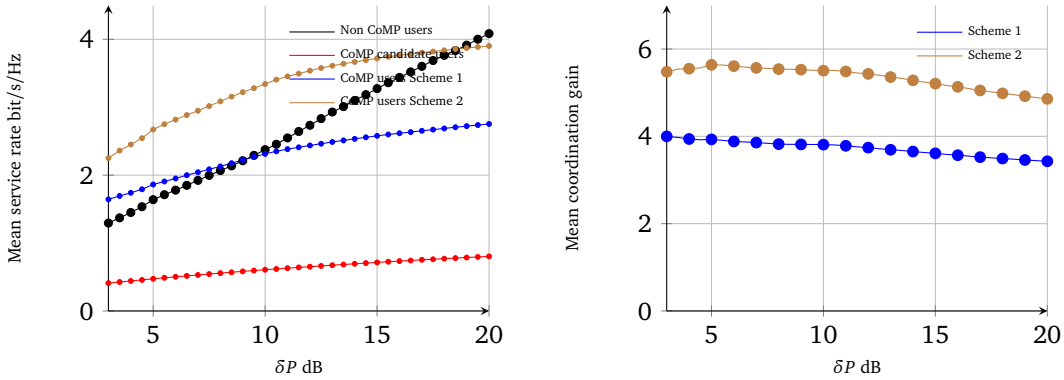


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.28: Mean service rate and mean coordination gain for Inter-site coordination.



(a) $K_{max} = 2$ maximum coordinated cells

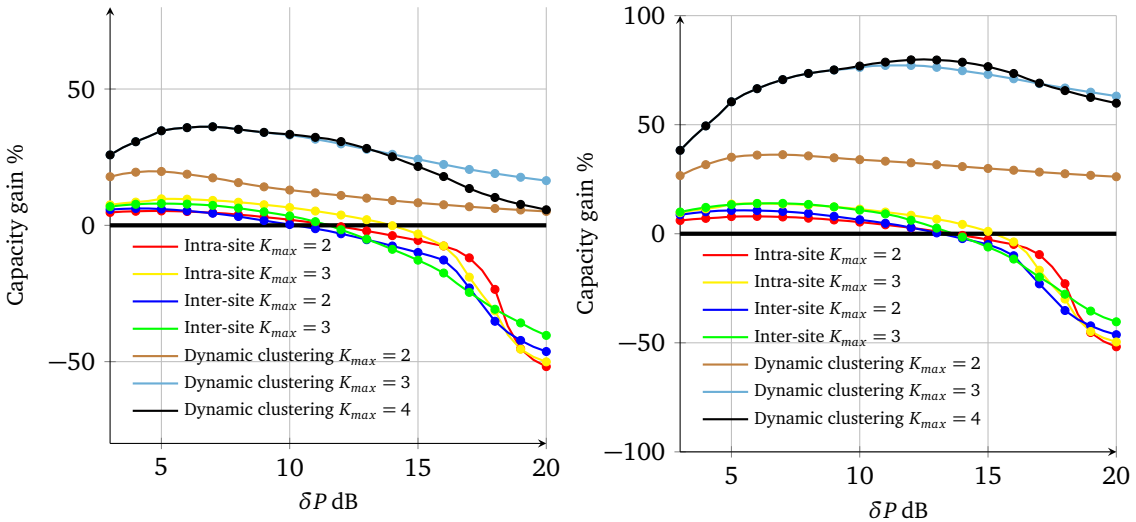


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.29: Mean service rate and mean coordination gain for Dynamic clustering.

b) Capacity

We evaluate the capacity gain when applying the standard cooperating cell definition algorithm I.5.3 with different power threshold δP, under different clustering methods and different maximum coordinated cells: $K_{max} = 2$ and $K_{max} = 3$, for a regular hexagonal network (21 cells) as well as an hexagonal network with eNodeB position and antenna direction dispersion.



(a) Scheme 1

(b) Scheme 2

Figure III.30: Capacity gain with respect to δP under different clustering methods.

We study two transmission schemes (Scheme 1 and Scheme 2) and we evaluate for each case the mean service rate in each cell at full interference, when all neighboring cells are transmitting (generating interference). In the absence of coordination, the mean service rate in the cell is the harmonic mean of the mean service rate of non-CoMP users and that of CoMP candidate users (see the previous section). On the other hand, in the presence of coordination the mean service rate is the harmonic mean of the mean service rate of non-CoMP users and that of CoMP users including those associated with neighboring cells and requiring transmission from the considered cell. We estimate the stability condition under a RR scheduler in the case without coordination (classical RR scheduler) as well as in the case when applying JT (RR iterative scheduler); it is nothing more than the ratio of the total traffic to the mean service rate. Then we compare the stability condition in the presence of JT technique to that in the absence of this technique. Note that in the case when applying JT, the additional traffic induced by the cooperation and coming from neighboring cells is taken into account. The overall capacity gain in the network is evaluated as the arithmetic mean of capacity gains in different cells.

i. Regular hexagonal network Figure III.30 shows the case of a regular hexagonal network. Observe that the highest capacity gains under RR scheduling strategy is obtained in the case of dynamic clustering. Obviously, Scheme 2 provides higher capacity gains than Scheme 1. In the case of fixed clustering, high δP values lead to a degradation of the stability condition which can reach 50% around $\delta P = 20$ dB. The maximum capacity gains are given by Table III.1. Observe that the optimal δP that allows to obtain the maximum capacity gain depends on the clustering method, the

	dP	Maximum Capacity Gain %
Intra-site $K_{max} = 2$	5 dB	5,27%
Intra-site $K_{max} = 3$	5 dB	9,72%
Inter-site $K_{max} = 2$	4 dB	6,18%
Inter-site $K_{max} = 3$	5 dB	7,97%
Dynamic $K_{max} = 2$	4,5 dB	19,83%
Dynamic $K_{max} = 3$	6,5 dB	36,16%

(a) Scheme 1

	dP	Maximum Capacity Gain %
Intra-site $K_{max} = 2$	5,5 dB	7,95%
Intra-site $K_{max} = 3$	6 dB	13,86%
Inter-site $K_{max} = 2$	5,5 dB	10,73%
Inter-site $K_{max} = 3$	6,5 dB	13,96%
Dynamic $K_{max} = 2$	7 dB	36,25%
Dynamic $K_{max} = 3$	11,5 dB	77,18%

(b) Scheme 2

Table III.1: Maximum Capacity gain under different clustering methods, regular hexagonal network.

maximum number of coordinated cells for one user K_{max}, as well as the transmission scheme. The maximum capacity gains goes till 10% in the case of fixed clustering and till 36% in the case of dynamic clustering when transmission scheme 1 is used. However when applying transmission scheme 2, the maximum capacity gain reaches 14% in the case of fixed clustering and around 77% when a dynamic clustering approach is adopted.

ii. Hexagonal network with position and antenna direction dispersion Now, we add dispersion to the position of each eNodeB as well as to the antenna direction of each sector in an hexagonal network. The new SINR map without any coordination between cells is given by Figure III.31. This is closer to a real network.

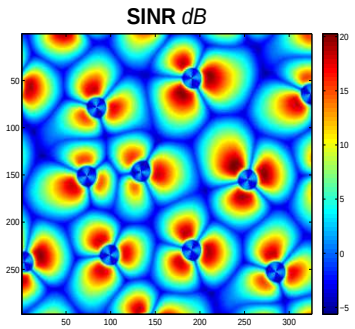


Figure III.31: SINR map without CoMP in an hexagonal network with position and antenna direction dispersion.

	dP	Maximum Capacity Gain %
Intra-site $K_{max} = 2$	4 dB	4,8%
Intra-site $K_{max} = 3$	5 dB	8,81%
Inter-site $K_{max} = 2$	4 dB	6,42%
Inter-site $K_{max} = 3$	4,5 dB	7,37%
Dynamic $K_{max} = 2$	4,5 dB	18,65%
Dynamic $K_{max} = 3$	5,5 dB	29,79%

(a) Scheme 1

	dP	Maximum Capacity Gain %
Intra-site $K_{max} = 2$	5 dB	7,6%
Intra-site $K_{max} = 3$	5,5 dB	13,37%
Inter-site $K_{max} = 2$	5 dB	11,07%
Inter-site $K_{max} = 3$	5,5 dB	13,39%
Dynamic $K_{max} = 2$	6,5 dB	33,79%
Dynamic $K_{max} = 3$	9,5 dB	63,08%

(b) Scheme 2

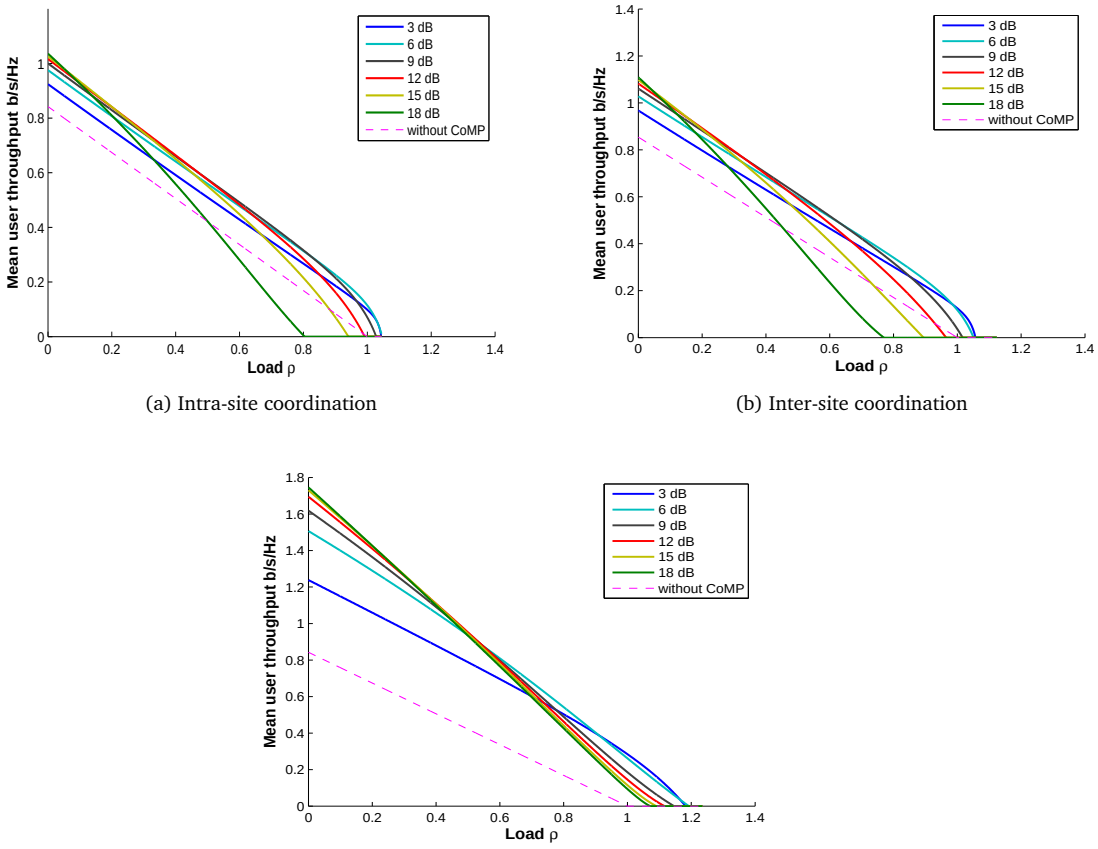
Table III.2: Maximum Capacity gain under different clustering methods, hexagonal network with position and antenna direction dispersion.

The maximum capacity gains under different clustering approaches are given by Table III.2, for Scheme 1 and Scheme 2. Note that these gains are of the same order of these obtained in a regular hexagonal network.

c) Analysis: case of a single coordination zone

We consider in this section a high interference scenario with different values of δP under an iterative scheduler. In order to study the performance in terms of mean user throughput, we consider the simplest scenario which is the scenario of a single coordination zone, see Section III.2.1. Note that the case of fixed clustering: intra-site coordination and inter-site coordination with $K_{max} = 2$ corresponds to the model described in Section III.2.2. However, following the approximation explained in §III.3.4.2 a) ii., we can still use the model corresponding to the scenario of a single coordination zone in order to study performance. We consider static users only. The mean user throughput in a cell is then nothing more than the harmonic mean of the mean user throughput in coordination zones as given by equation (III.17) and that in non-coordination zones as given by equation (III.16). We proceed by decoupling the different zones so that we consider each zone (coordination or non-coordination) alone and neglect the others, after taking into account the load induced by these zones. So, in each cell, ρ_k is the load induced by non-CoMP users which is given by the total arrival rate in non-coordination zone to the mean service rate of non-CoMP users which depends on the δP as shown in Figure III.27 and Figure III.28. On the other hand, ρ_0 is the load induced by CoMP users. This load is given by the ratio of the total arrival rate in coordination zones including the traffic induced by neighboring cooperating cells to the mean service rate of CoMP users as shown in Figure III.27 and Figure III.28. Note that the performance in different coordination zones belonging to a

given cell is similar due to the symmetry of the considered topology.



(a) Intra-site coordination

(b) Inter-site coordination

(c) Dynamic clustering

Figure III.32: Throughput performance obtained through the analysis of flow-level traffic model with $K_{max} = 2$.

We proceed by a similar way to study performance under a dynamic clustering strategy. We decouple different zones after taking into account the load induced by other zones requiring transmission from the same cell. We consider the mean service rates of non-CoMP users and CoMP users given by Figure III.29 in order to compute the load in different zones.

Figure III.32 compares the throughput performance of the case without coordination and the case with coordination considering different values of δP: 3 dB, 6 dB, 9 dB, 12 dB, 15 dB and 18 dB and $K_{max} = 2$ coordinated cells. We consider that same data is transmitted from different transmission points (scheme 1). Observe that in this high interference scenario, JT improves performance without degrading stability condition when the chosen δP can guarantee a sufficient mean coordination gain that should be at least 100% in this case. According to Figure III.27, Figure III.28 and Figure III.29, the mean coordination gain should be higher than two in order to prevent the degradation of the stability condition in accordance with equation (III.29). This condition is satisfied for all δP under dynamic clustering while under static clustering it is not satisfied for δP higher that 12 dB.

Note that these gains are obtained when neglecting opportunistic gains considering a RR scheduler that does not leverage instantaneous radio conditions variation due to fast fading. These results are

also valid in an environment where channel variations are negligible (line of sight only). We shall see that , under sub-optimal scheduling strategies, coordination between different cells may reduce opportunistic gains obtained under classical opportunistic schedulers resulting in a trade-off between coordination gains and opportunistic gains.

Despite the simplicity of the model, it presents some weakness due to the fact that the same service rate is used whatever the load of the network is. This is not accurate since service rate depends on the network load which affects the level of interference coming from neighboring cells. Consequently, the mean service rate at very low load is not the same in the absence and in the presence of JT. Cells are involved in transmissions to users associated with neighboring cooperating cells so they are more often activated, thus generating more interference.

d) System level simulations

Iterative scheduling

We simulate 12×10^6 TTI (which is equivalent to 200 minutes) and estimate the mean user throughput as the ratio of the mean flow size to mean flow duration for $\sigma = 1.25$ MBytes, under different clustering strategies, using the LTE simulator described in Appendix A. We consider static users only. We study the case without coordination under classical PF scheduler as well as the case with JT for different power threshold values δP: 3 dB, 6 dB, 9 dB, 12 dB and 15 dB, under PF iterative scheduler explained in Algorithm 3, where metric is nothing more than the classical PF metric. We evaluate also in each case the mean number of users per eNodeB and the relative throughput gain when applying JT CoMP compared to the case without coordination with respect to the offered traffic per eNodeB.

For each clustering method, we consider two cases. In the first one, we limit the number of cooperating cells for one user to $K_{max} = 2$ while in the second one at most $K_{max} = 3$ cells can cooperate for one user.

i. Static clustering Figure III.33 shows the case of intra-site coordination when same data is scheduled from different coordinated cells (III.13).

Observe that applying JT may degrade performance at high load by reducing the stability condition especially for high values of δP, starting from 6 dB. Scheduling users in neighboring cells might reduce the amount of resources available for users in a given cell so that the system cumulates more users compared to the case without JT. Thus the mean number of users increases quickly at high load when using large values of δP.

Another important point to mention here is that applying JT with high values of δP surprisingly degrades performance at very low load. This degradation has nothing to do with the extra resource consumption induced by the joint transmission from multiple cells as resources are most of the time available at low load. Scheduling data to users from multiple cells has the potential to generate a larger amount of interference in the system than is the case when JT techniques are not present. Such interference could impact in particular users that were not affected by this interference in the absence of JT. The fraction of time where cells are active, generating interference at low load becomes larger thus degrading the performance.

Note that the highest gains are obtained at medium load. However, we can only get limited performance where maximum relative gain is around 10% and obtained for $\delta P = 6$ dB. observe that gains in the case of $K_{max} = 3$ are slightly larger than in the case of $K_{max} = 2$.

Figure III.34 shows the case of intra-site coordination when different content are scheduled from different coordinated cells (III.14). In this case the coordination gain brought by the additional meaningful signal is larger and then the gains are higher compared to the previous case. Observe that performance degradation at high load is obtained starting from $\delta P = 15$ dB. Note that there is no throughput degradation at very low load in this case. This is due to the fact that the coordination

gain when using this transmission scheme is high enough in order to compensate the loss caused by the interference generated to other neighboring cells. The highest gains are also obtained at medium load. Maximum relative gains are around 25% and 35% when $K_{max} = 2$ and $K_{max} = 3$ respectively. They are obtained around $\delta P = 9$ dB.

Figure III.35 shows the case of inter-site coordination when using Scheme 1. The main conclusions are the same of the case of intra-site coordination. Observe that stability condition degradation is obtained for all evaluated values of δP. Figure III.36 shows the case of inter-site coordination when using Scheme 2. Maximum relative gains are around 40% and 50% when $K_{max} = 2$ and $K_{max} = 3$ respectively. They are obtained when $\delta P = 9$ dB.

ii. Dynamic clustering The case of dynamic clustering when using Scheme 1 is illustrated by Figure III.39. Obviously this clustering method provides the highest gains but this is at the cost of higher complexity. There is an excessive resource consumption illustrated by the stability condition degradation observed for all values of δP. Observe that the mean number of users increases very quickly at high load compared to the case without coordination especially when $K_{max} = 3$. The highest gains are around 40% and 50% when $K_{max} = 2$ and $K_{max} = 3$ respectively. Similarly to the case of static clustering, they are obtained for $\delta P = 6$.

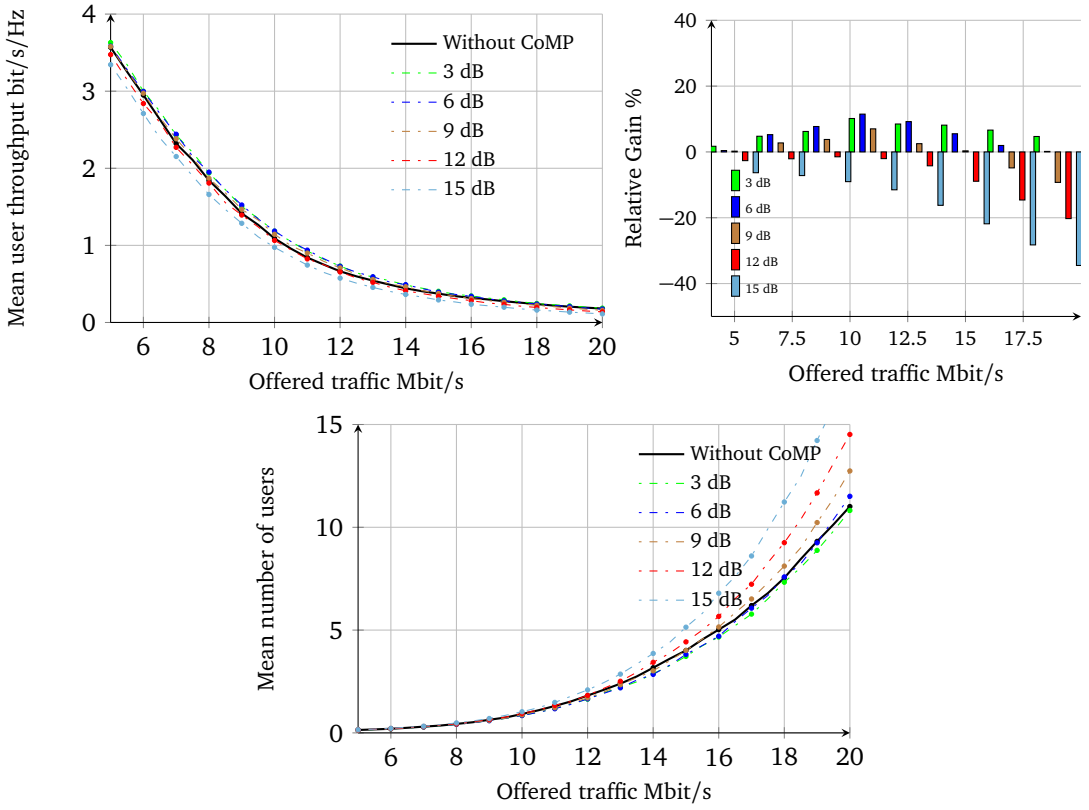
The case of dynamic clustering when using Scheme 2 is illustrated by Figure III.40. This case is theoretical and gives performance upper bound. Practically inter stream interference is not completely eliminated. Gains are impressive; the highest gains are around 150% when CoMP users are allowed to receive up to two different independent streams while they are around 300% when CoMP users can receive up to three streams. These highest gains are observed at medium load for $\delta P =$ 12-15 dB. Observe that there is no performance degradation neither at high load nor at low load. The coordination gain is very high so it overcomes the problem of extra resource consumption as well as the problem of interference generation that appears at low load.

iii. Impact of K_{max} In order to study the impact of the maximum number of coordinated cells K_{max} allowed for one user, we consider dynamic clustering with two values of δP: 3 dB and 12 dB under PF iterative scheduler. We vary the value of K_{max} from 2 cooperating cells to 5 cooperating cells. We evaluate in each case the mean user throughput as well as the mean number of users per eNodeB with respect to the offered traffic per eNodeB supposing that all users are static. Then, we deduce the relative throughput gain when applying JT CoMP compared to the case without coordination.

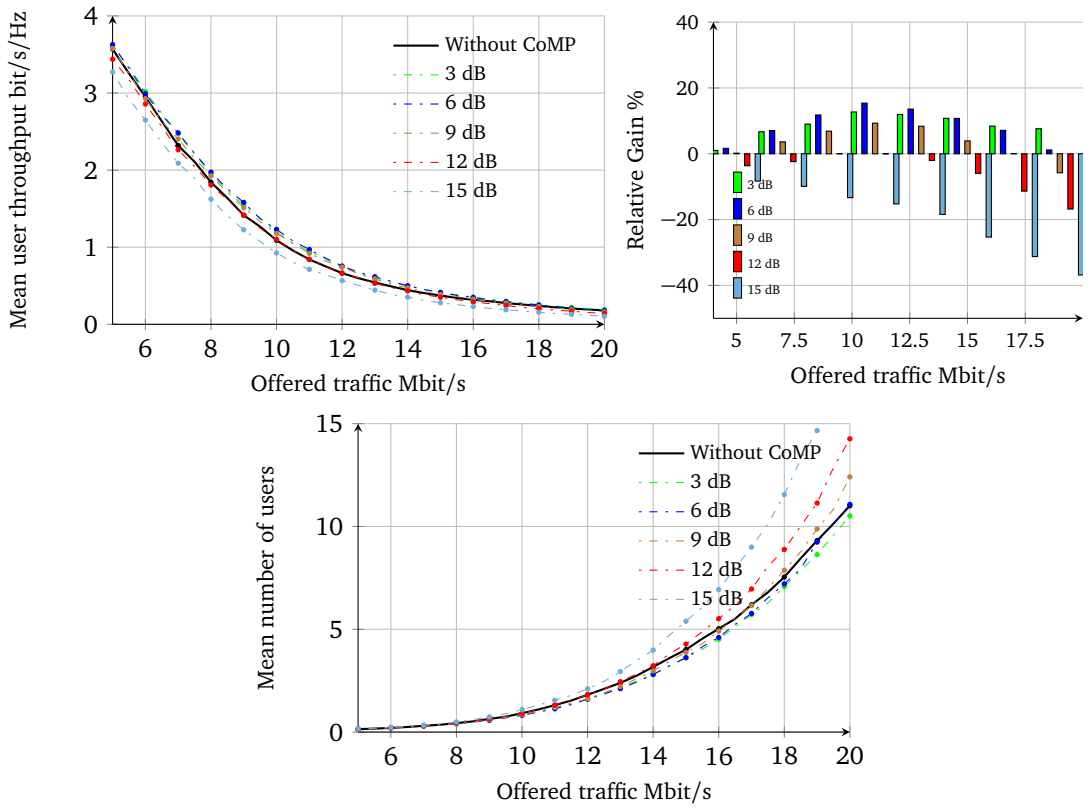
Figure III.41 shows the case of transmission Scheme 1 while Figure III.41 illustrates the case of transmission Scheme 2.

Observe that for $\delta P = 3$ dB the highest gains are achieved when $K_{max} = 3$. Further increasing the maximum number of cooperating cells for one user to more than three does not increase the gains. This is due to the fact that the probability that more than two cells satisfy the condition (I.1) is negligible in the considered network topology. Note that this conclusion is common for both transmission schemes.

However for a relatively high $\delta P = 12$ dB the results are different according to the transmission scheme. $\delta P = 12$ dB is considered as degrading performance when transmission Scheme 1 is applied under dynamic clustering according to the previous section, so the lower the number of cells involved in the transmission, the better the performance. However, $\delta P = 12$ dB is considered as leading to good performance when transmission Scheme 2 is applied. Moreover, the probability that more than two cells satisfy the condition (I.1) is not negligible when $\delta P = 12$ dB compared to the case when $\delta P = 3$ dB, so the higher the maximum number of coordinated cells the better the performance. However, the additional gain brought when $K_{max} > 3$ with respect to the gain when $K_{max} = 3$ is relatively small. So we can say that $K_{max} = 3$ is a good compromise between complexity and performance and enables to achieve most of the gains in a hexagonal network topology.

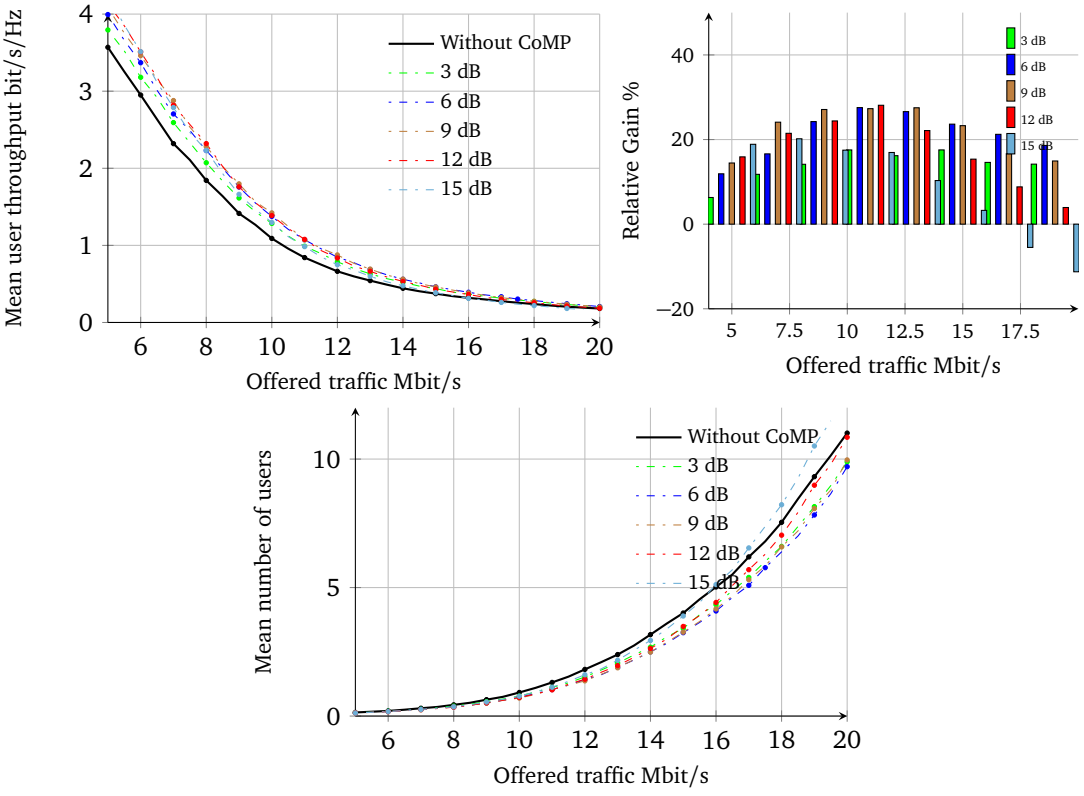


(a) $K_{max} = 2$ maximum coordinated cells

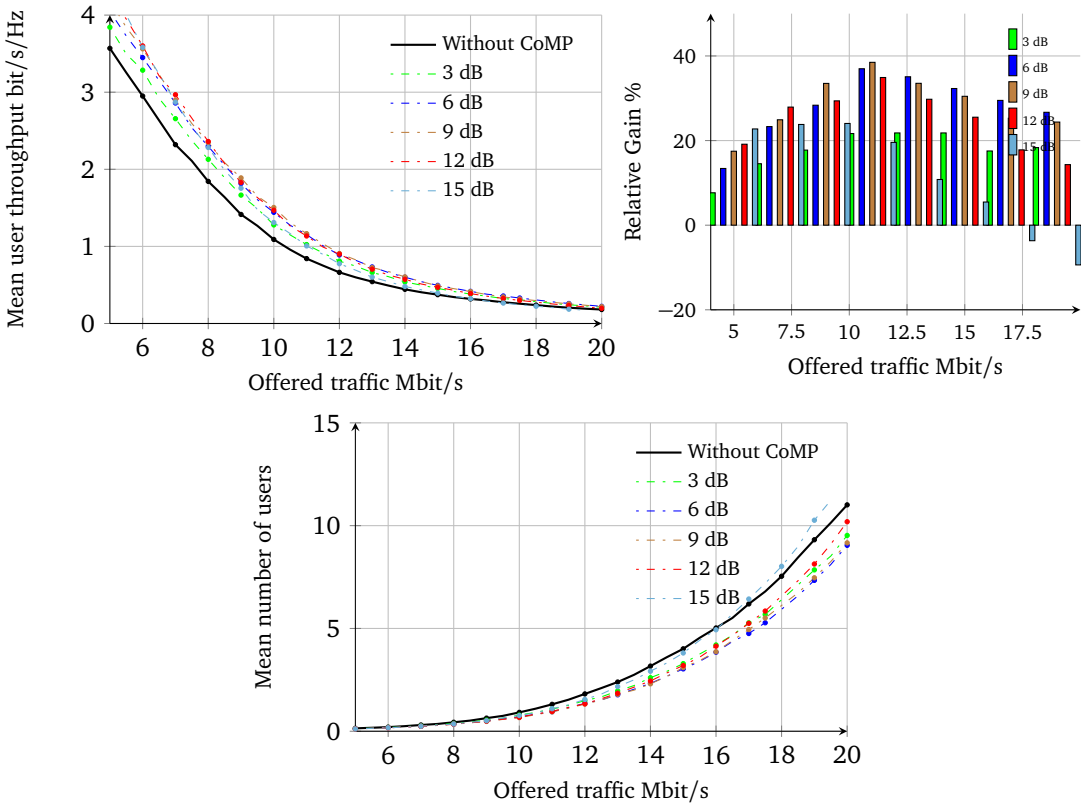


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.33: Throughput performance obtained by system-level simulation for Intra-site JT Scheme 1.

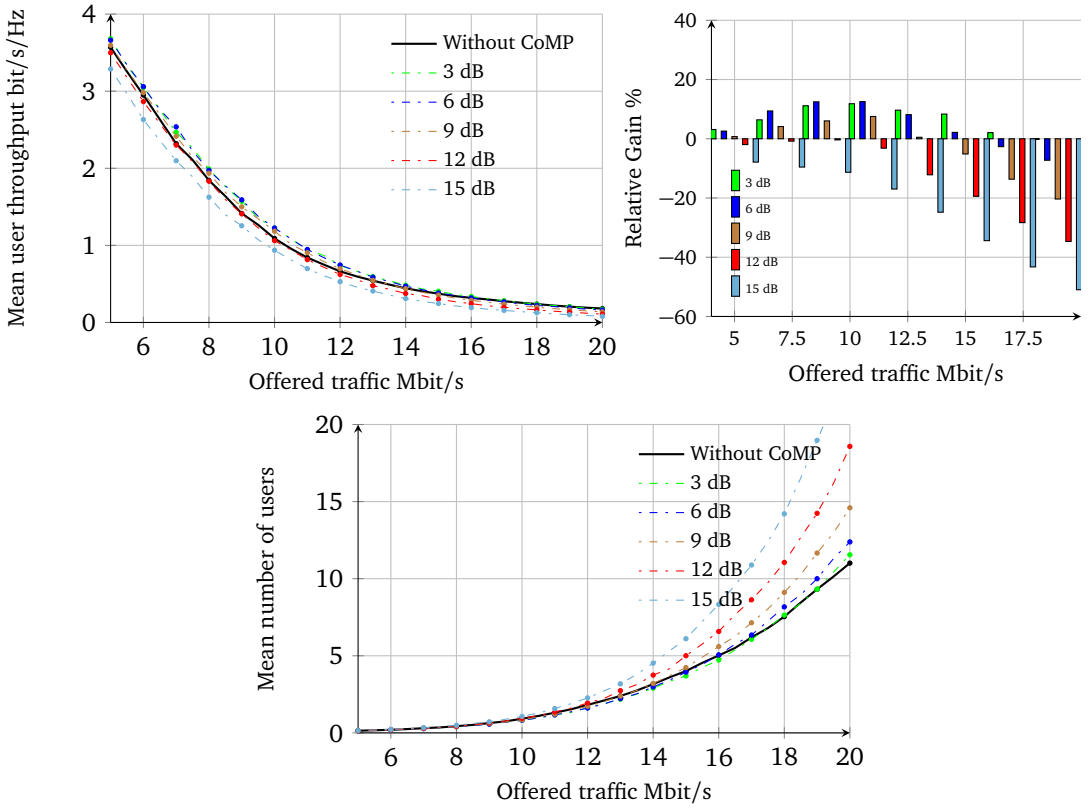


(a) $K_{max} = 2$ maximum coordinated cells

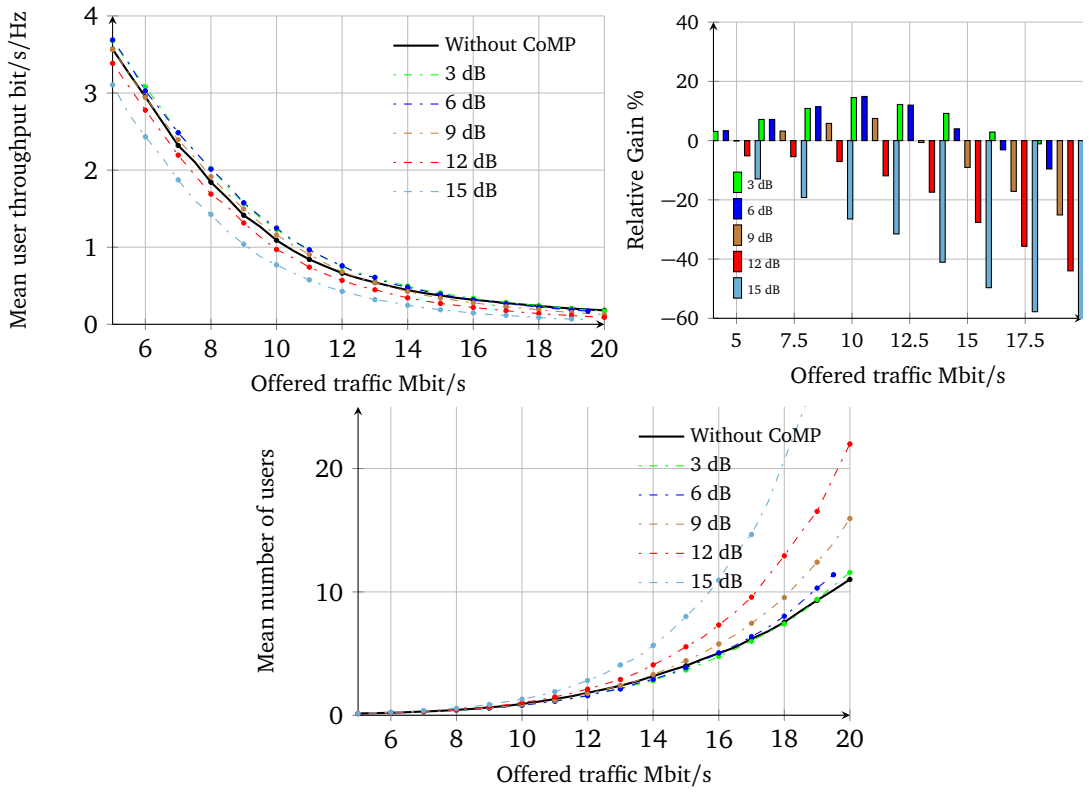


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.34: Throughput performance obtained by system-level simulation for Intra-site JT Scheme 2.

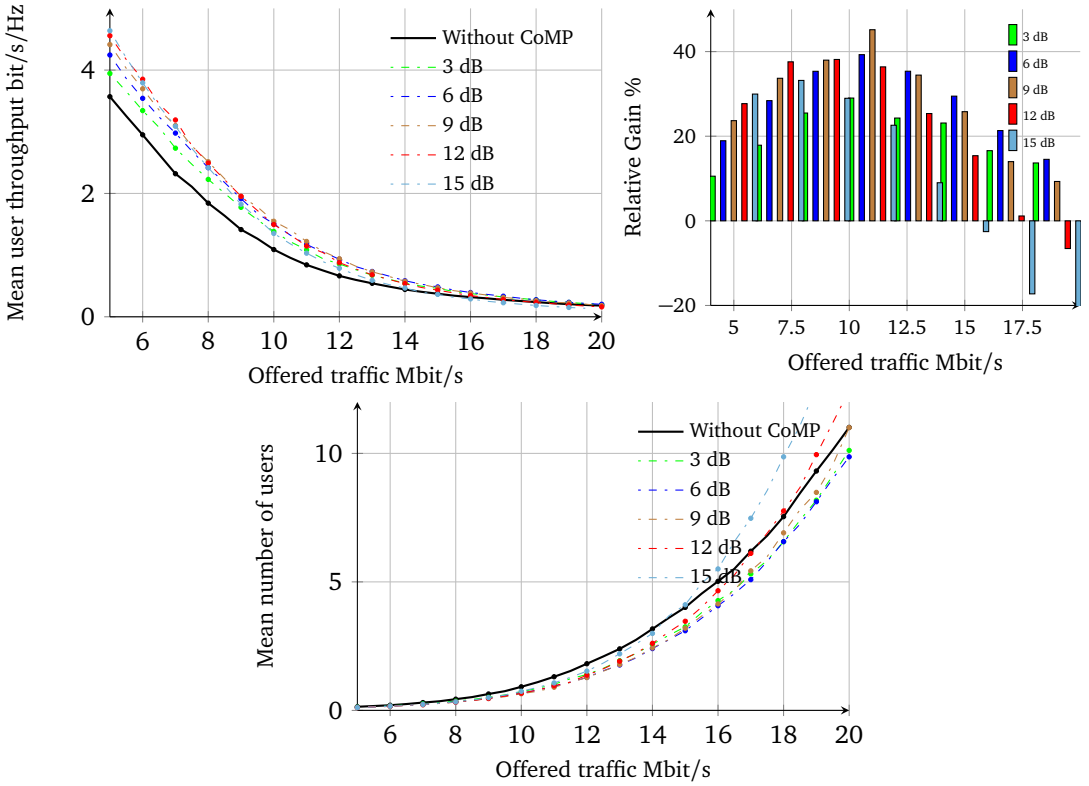


(a) $K_{max} = 2$ maximum coordinated cells

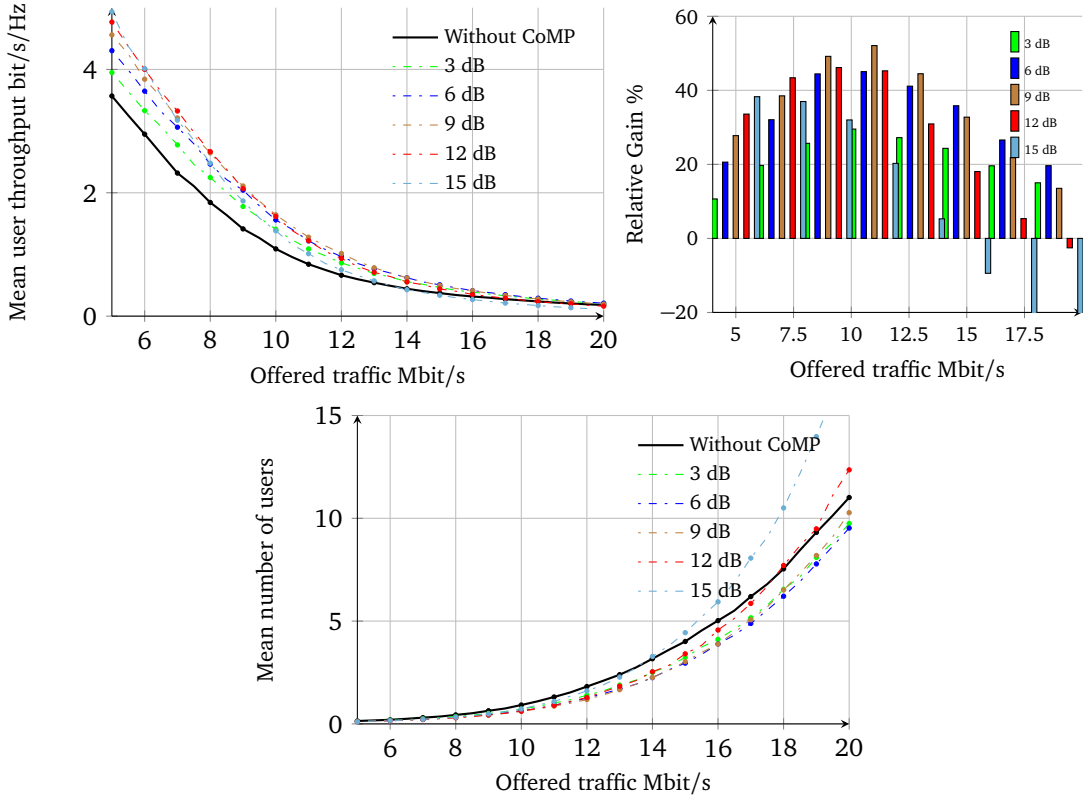


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.35: Throughput performance obtained by system-level simulation for Inter-site JT Scheme 1.

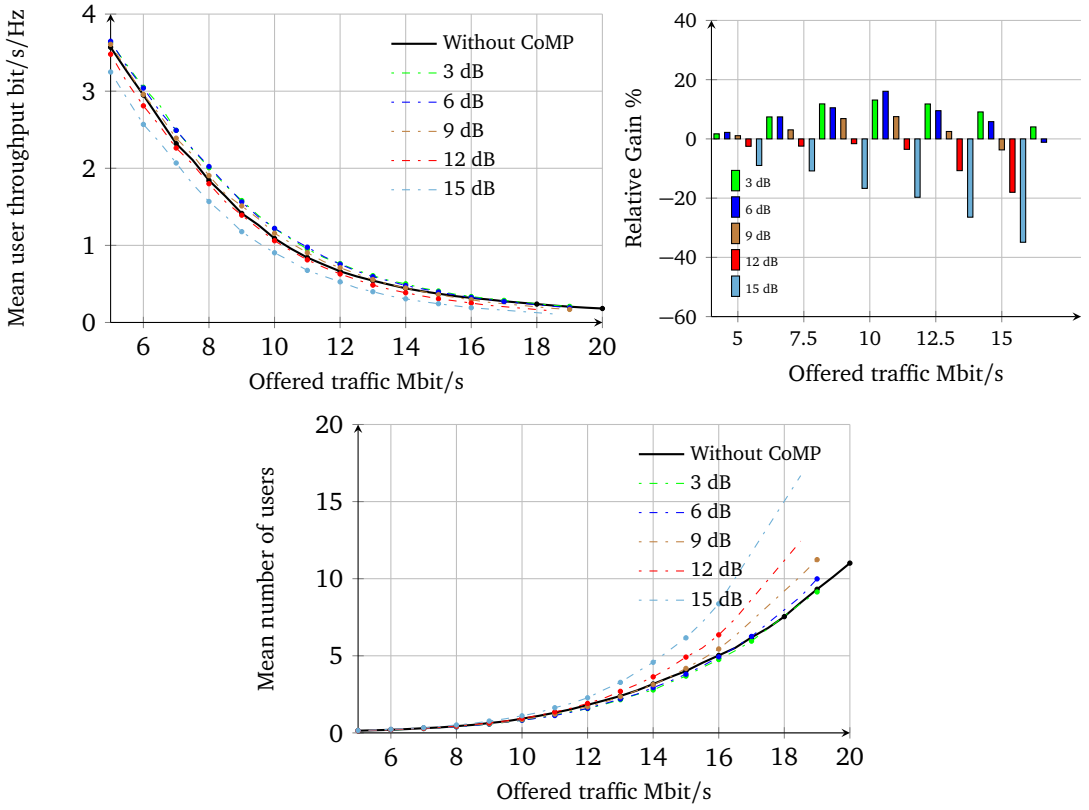


(a) $K_{max} = 2$ maximum coordinated cells

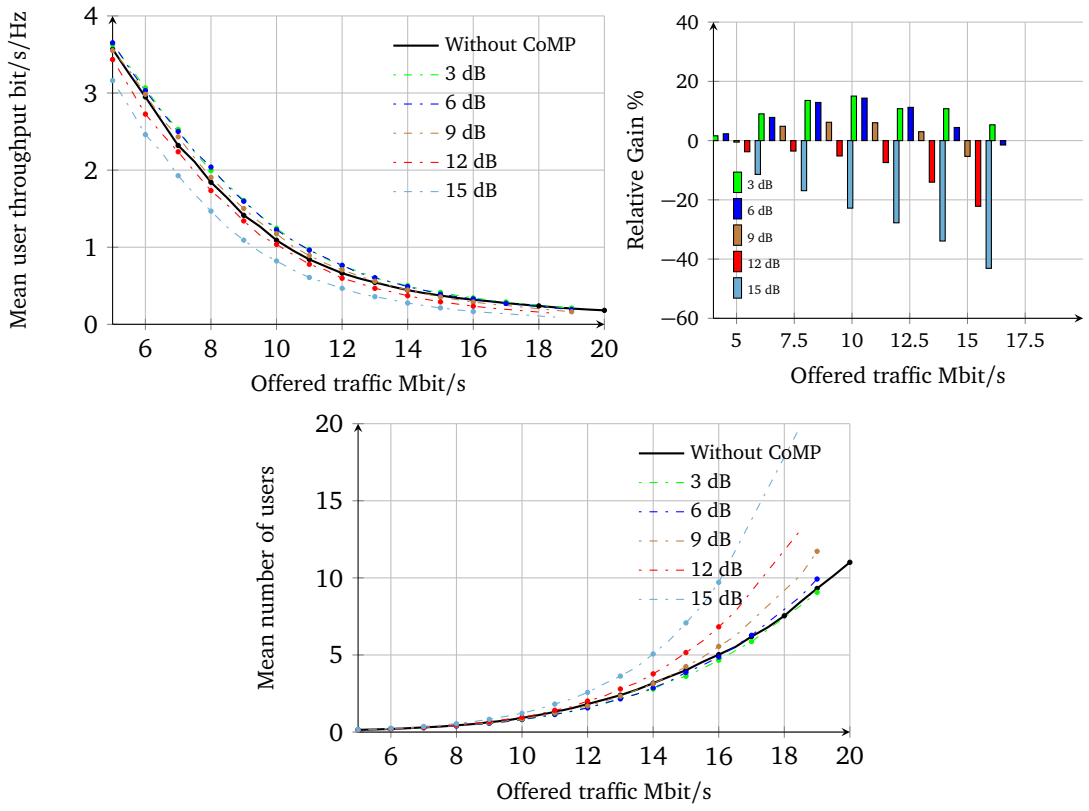


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.36: Throughput performance obtained by system-level simulation for Inter-site JT Scheme 2.

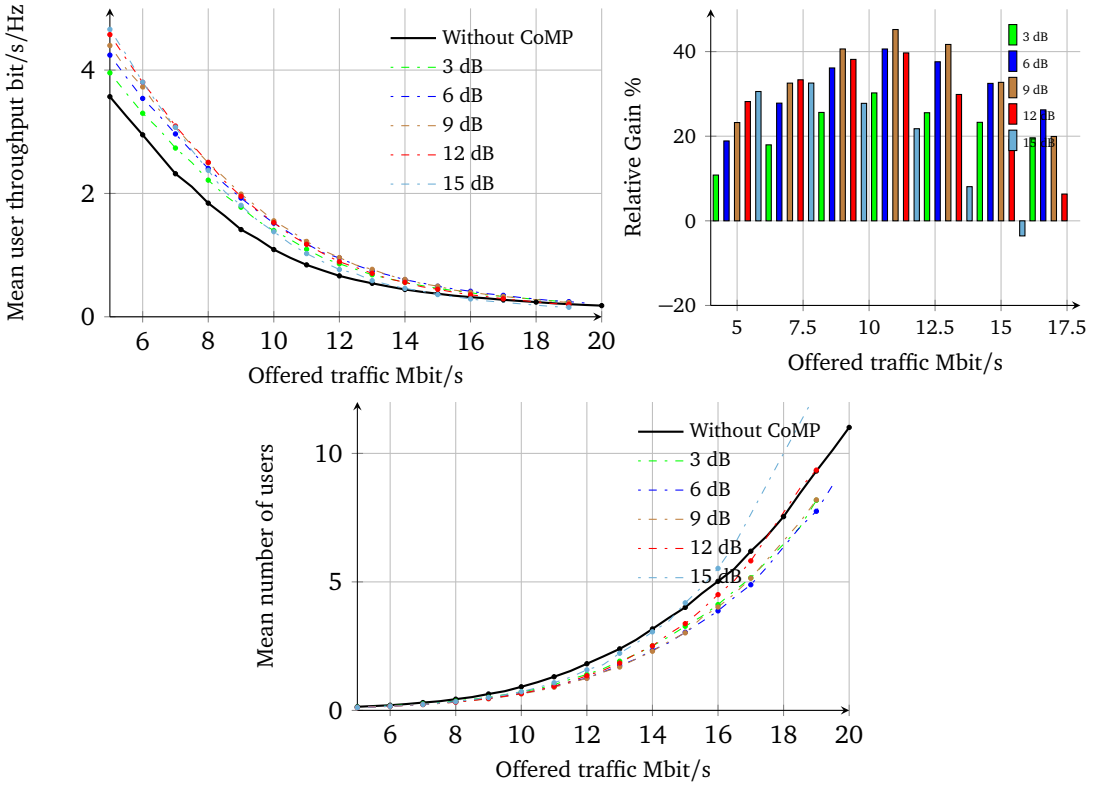


(a) $K_{max} = 2$ maximum coordinated cells

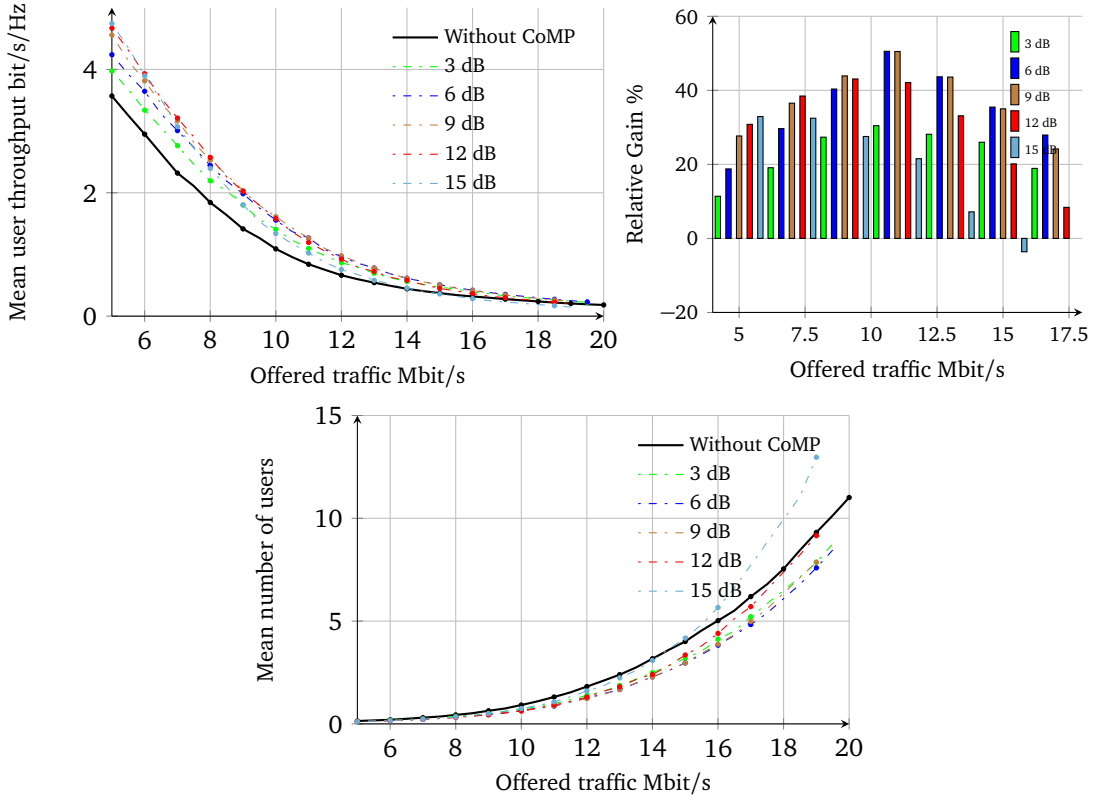


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.37: Throughput performance obtained by system-level simulation for Intra and Inter-site JT Scheme 1.

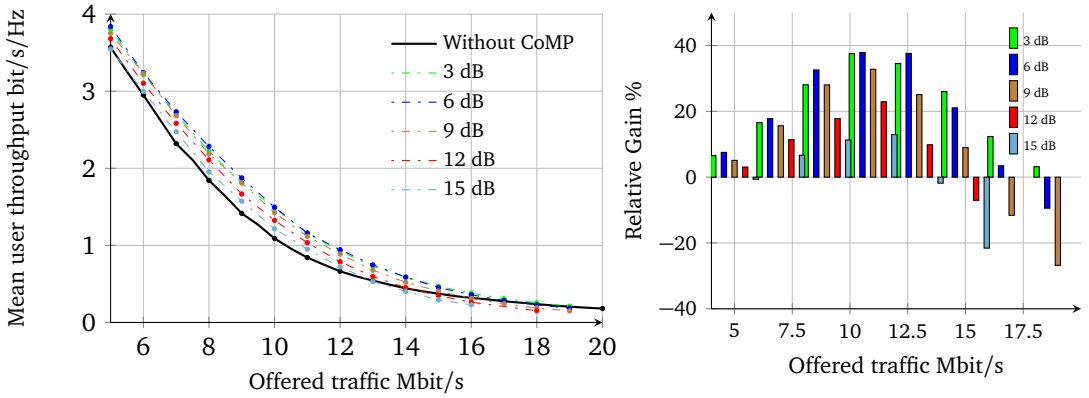


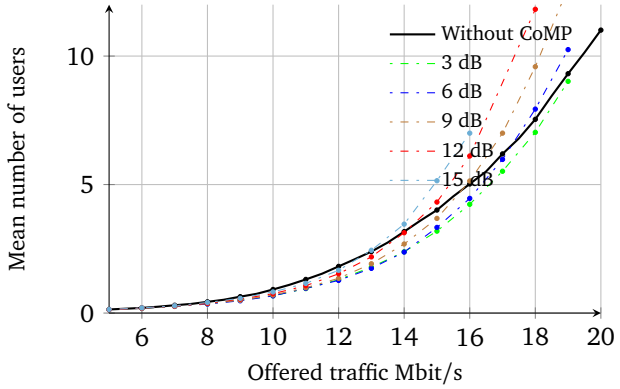
(a) $K_{max} = 2$ maximum coordinated cells



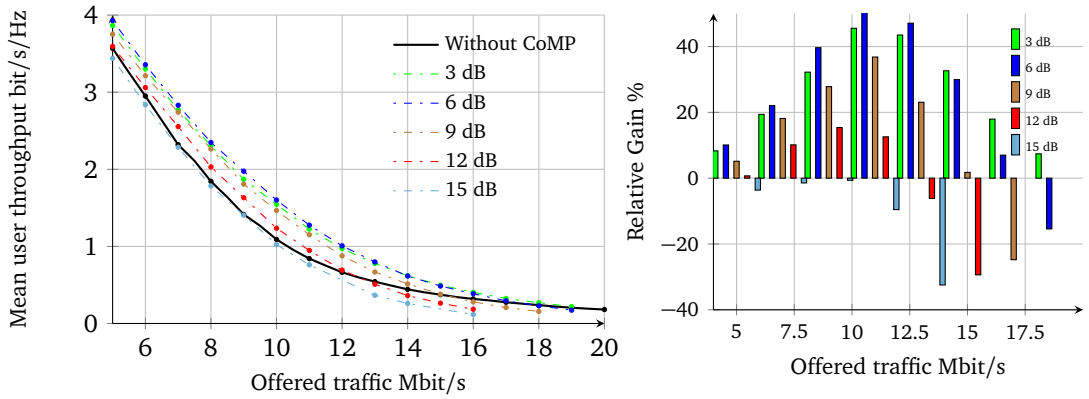
(b) $K_{max} = 3$ maximum coordinated cells

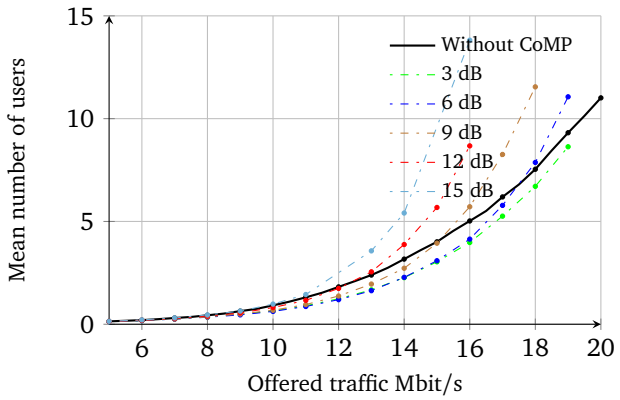
Figure III.38: Throughput performance obtained by system-level simulation for Intra and Inter-site JT Scheme 2.





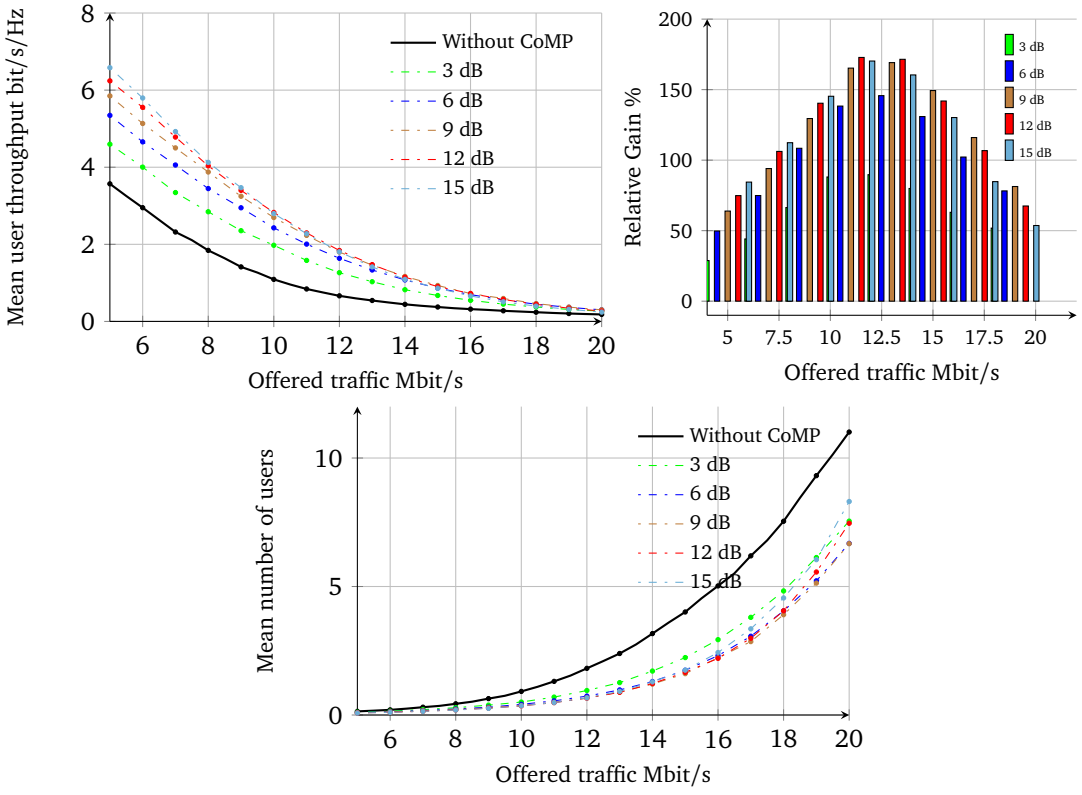
(a) $K_{max} = 2$ maximum coordinated cells



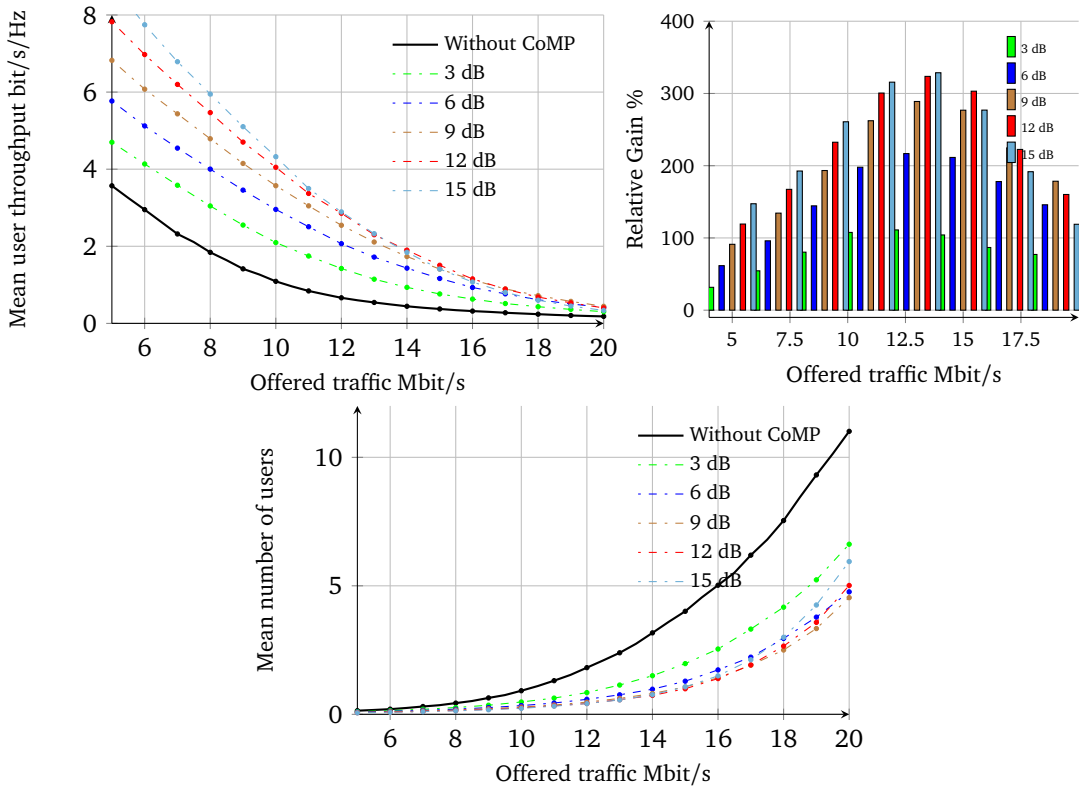


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.39: Throughput performance obtained by system-level simulation for Dynamic clustering JT Scheme 1.

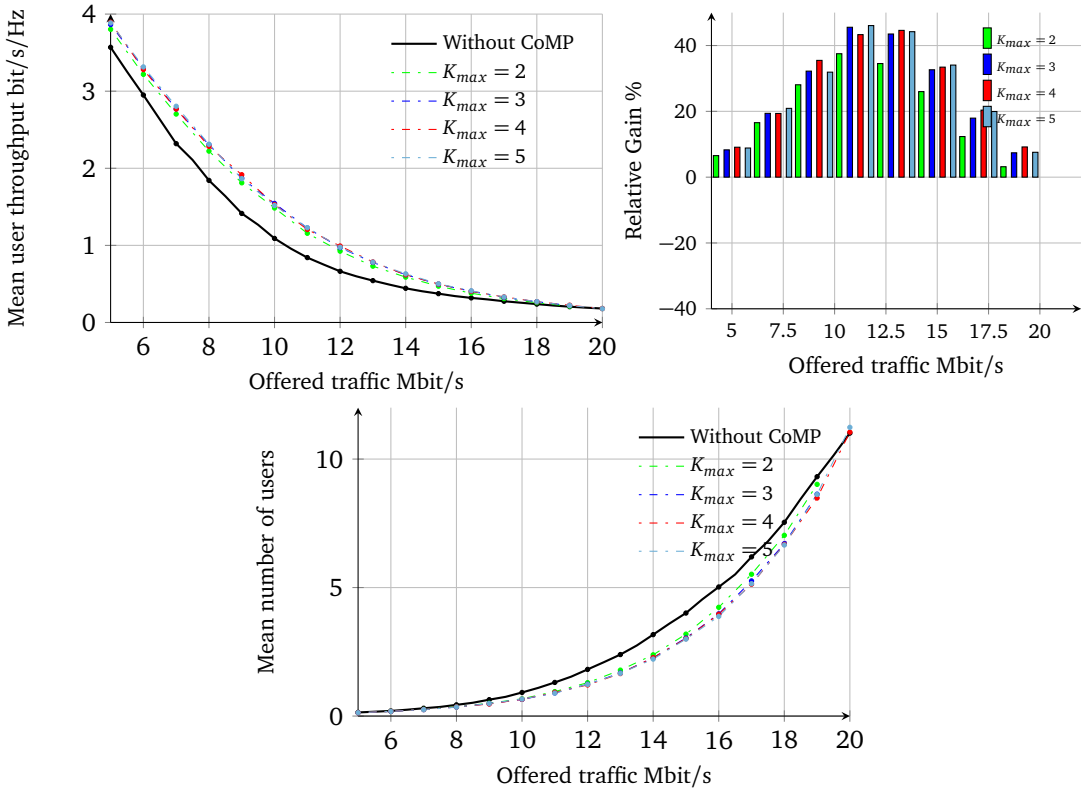


(a) $K_{max} = 2$ maximum coordinated cells

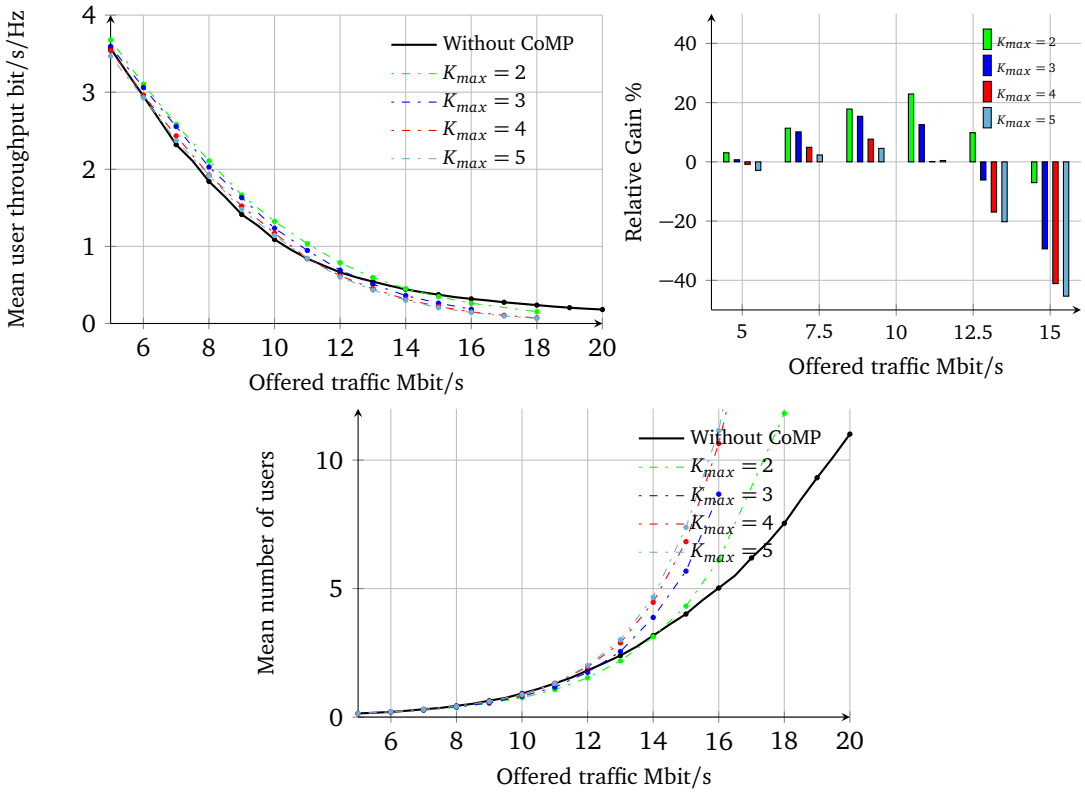


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.40: Throughput performance obtained by system-level simulation for Dynamic clustering JT Scheme 2.

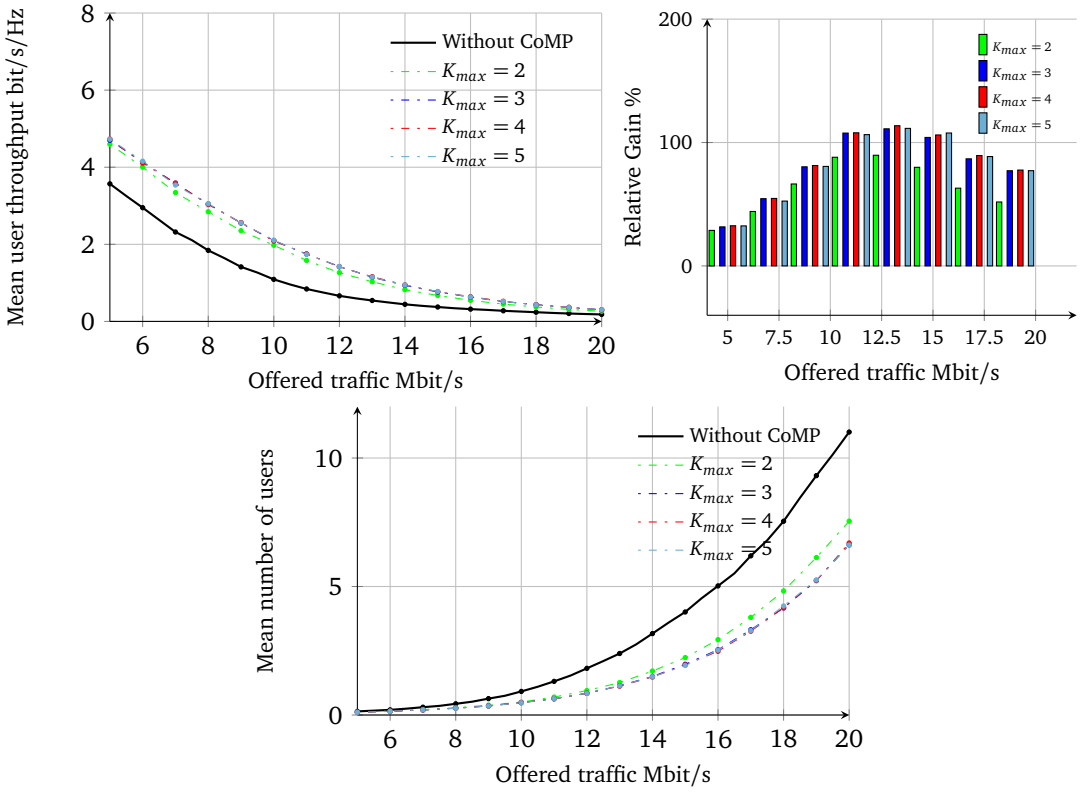


(a) $\delta P = 3dB$

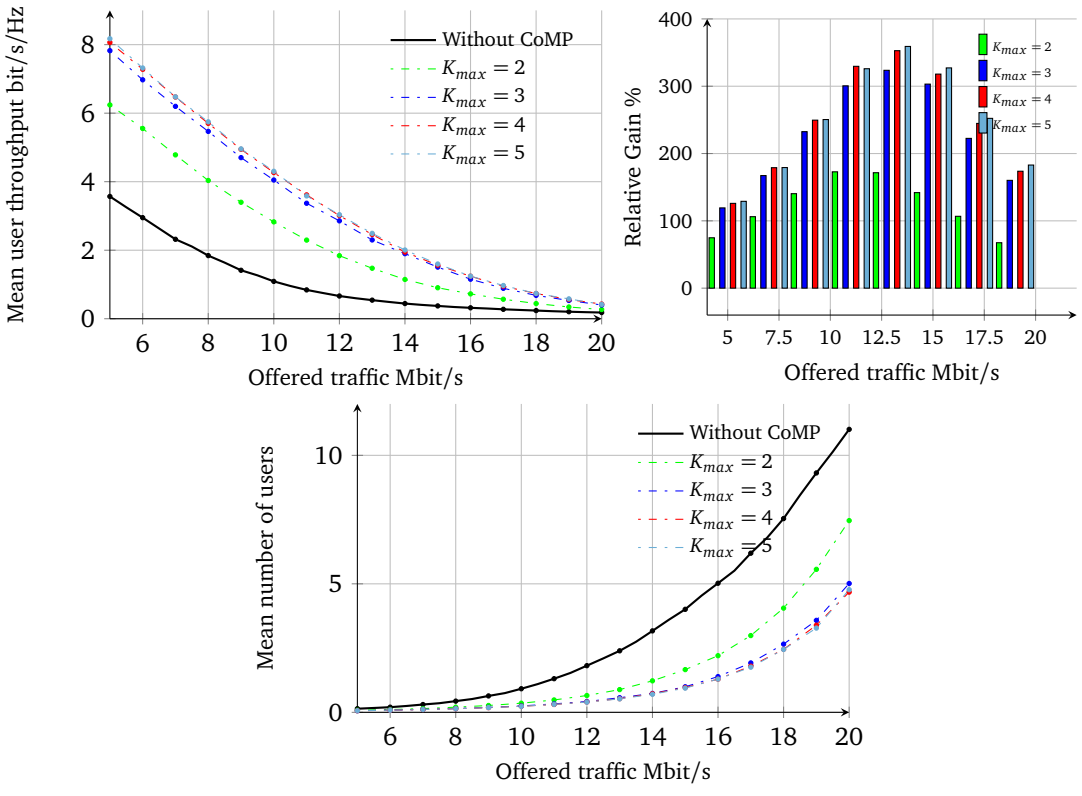


(b) $\delta P = 12dB$

Figure III.41: Throughput performance obtained by system-level simulation for Dynamic clustering JT Scheme 1 for different values of K_{max}.



(a) $\delta P = 3dB$



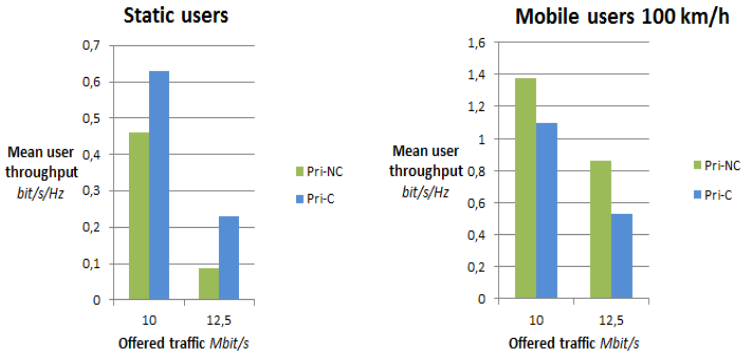
(b) $\delta P = 12dB$

Figure III.42: Throughput performance obtained by system-level simulation for Dynamic clustering JT Scheme 2 for different values of K_{max}.

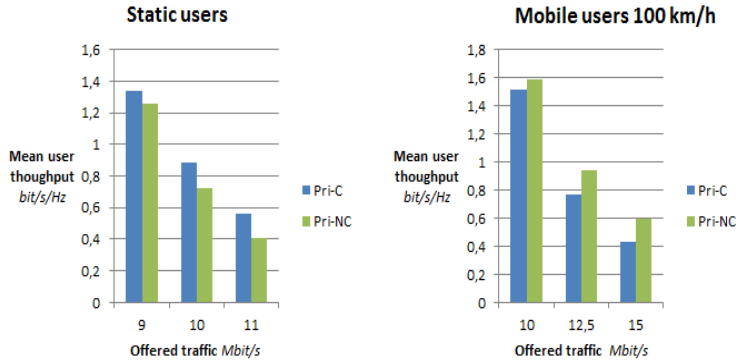
e) System level simulations

Prioritization strategies

According to Section III.3.5.1, the strategy of prioritizing CoMP users outperforms that of prioritizing non-CoMP users when users are static. However, Pri-C strategy leads to only limited performance compared to Pri-NC strategy when users are mobile. In order to validate these results obtained by analysis, we conduct system simulations, , using the LTE simulator described in Appendix A, in the case of intra-site JT with $\delta P = 12dB$ when limiting the maximum number of cooperating cells for one user to $K_{max} = 2$, for both transmission schemes: Scheme 1 and Scheme 2. We consider the two prioritization strategies: Pri-C and Pri-NC. Inside each category of users: CoMP users or non-CoMP users, we apply the classical PF metric in order to schedule users. We study two cases: in the first one all users are static while in the second one all users are moving at 100Km/h



(a) Scheme 1



(b) Scheme 2

Figure III.43: Throughput performance obtained by system-level simulation for Intra-site JT under prioritization scheduling strategies.

We evaluate performance in terms of mean user throughput with respect to the offered traffic per eNodeB as shown in Figure III.43. Observe that when users are static Pri-C strategy outperforms Pri-NC strategy. Results are reversed when users are mobile.

Figure III.44 illustrates the instantaneous variation of the number of users in the network (21 cells), under both scheduling strategies when users are static and when users are mobile for a particular offered traffic (11 Mbit/s when users are static and 15 Mbit/s when users are mobile). Observe

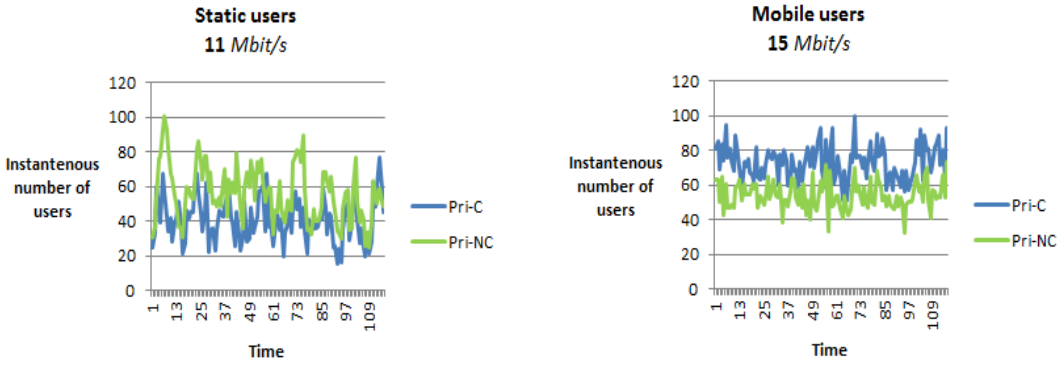


Figure III.44: Variation of the instantaneous number of users in the network obtained by system-level simulation for Intra-site JT Scheme 2 under prioritization scheduling strategies.

that under Pri-C strategy the networks can treat the incoming traffic faster that under Pri-NC scheduling strategy when users are static. However, when users are mobile traffic is treated more slowly under Pri-C strategy. This confirms further the previous results. The conclusion drawn from this evaluation is common for both interference scenarios: low interference case and high interference case. However, note that in the presence of fast fading the iterative scheduler that treats non-CoMP users and CoMP users equally without any prioritization policy, achieves better diversity gains compared to both considered prioritization strategies. Such prioritization strategies can be considered only in an environment where channel variations are negligible (like line of sight case). Otherwise, they can lead to limited performance if channel variations are considerable.

III.4 Gain-based joint transmission

III.4.1 Coordination scheme

In order to maintain stability of the system when applying JT as explained in §III.3.4.3a), a cell should cooperate only when it brings at least 100% mean throughput gain when cooperating. Based on this conclusion we propose a new cooperating cell definition method. Instead of defining a cooperating cell based on the received power which requires to predefine a threshold δP, see Section I.5.3, the definition is based on the throughput gain. Thus a neighboring cell which increases the mean throughput of a user by at least a factor of two, is involved in the transmission to this user.
We denote by $\mathbb{S} = \{s_1, s_2, \ldots s_{K_{max}-1}\}$ the set of candidate cooperating cells for a given user, that is the set of neighboring cells sorted from the highest to the lowest interfering cell. Note that the maximum number of cells belonging to this set can not exceed $K_{max}-1$. We denote by r'_n the transmission rate given that the first n cells in the set $\mathbb{S}$ candidate cooperating cells, that are $s_1, s_2, \ldots s_n$, are involved in the transmission; it is given by

$$r'_n = \boldsymbol{F}\left(1 + \frac{P_s + \sum_{c=s_1,s_2,\ldots s_n} P_c}{\sum_{i\neq s_1,s_2,\ldots s_n} P_i + \mathcal{N}}\right)$$

when Scheme 1 is applied, and by

$$r'_n = \boldsymbol{F}\left(1 + \frac{P_s}{\sum_{i\neq s_1,s_2,\ldots s_n} P_i + \mathcal{N}}\right) + \sum_{c=s_1,s_2,\ldots s_n} \boldsymbol{F}\left(1 + \frac{P_c}{\sum_{i\neq s_1,s_2,\ldots s_n} P_i + \mathcal{N}}\right)$$

when Scheme 2 is applied. The algorithm works as follows:

Data: Set $\mathbb{S}$ of candidate cooperating cells
Result: Set $\mathbb{C}$ of cooperating cells
$r'_0 \leftarrow r$
$i \leftarrow 1$
$j \leftarrow 0$
while $(i < K_{max})$ **do**
  if $r'_i/r - r'_j/r > (i-j) * \beta_T$ **then**
    $\mathbb{C} \leftarrow \mathbb{C} \cup \{s_{j+1}, \ldots s_i\}$
    $j \leftarrow i$
  end
  $i \leftarrow i+1$
end

Algorithm 7: Gain-based cooperating cell definition algorithm for JT.

We denote by β_T the coordination gain constraint, that should be 100% in order to maintain the same stability condition as in the case without coordination. We shall see that the gain brought by the cooperation of a cell depends to a great extent on the interference level. A user can be for instance strongly interfered by two neighboring cells. The cooperation of one of two cells does not bring 100% cooperation gain. However, the cooperation of both cells can bring 200% gain, satisfying the required condition. This may happen when the neighboring cells cause similar or close interference levels. In this case, it is not worth to involve just one cell in the transmission. Either both cells cooperate, or none of them cooperates.

Consider for instance the case of $K_{max} = 3$, the proposed algorithm checks if s_1 brings at least $\beta_T = 100\%$ gain, then if the case, it checks if s_2 brings at least additional 100% gain. In the opposite

case, where s_1 does not bring 100% gain, the algorithm checks if the cooperation of s_1 and s_2 can bring 200% cooperation gain.

Algorithm generalization

Consider now the case of a general non symmetric case (non symmetric network topology, non uniform traffic distribution...), as illustrated by Figure III.45. In this case the factor α_n (see §III.3.4.3a)) is no more equal to n. In order to avoid the degradation of the stability of the system, the condition:

$$\beta_n \geq \alpha_n \neq n$$

must be satisfied. Thus a more advanced coordination scheme combined with a SON (Self Organizing

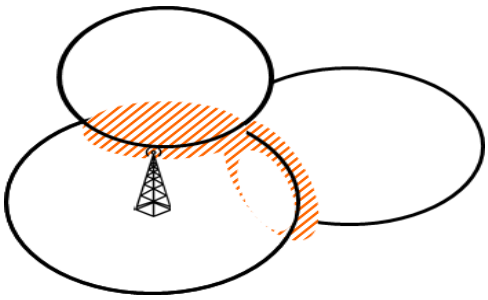


Figure III.45: A non symmetric network topology.

Network) algorithm can be used, so that the coordination gain constraint that fits best to the network characteristics, can be reconfigured each period of time. Thus, the traffic increase factor at the edge can be evaluated in each cell i through an exponentially smoothed average, and the coordination gain constraint, β_T^i, can be updated accordingly. In the general case, the algorithm works as follows:

Data: Set $\mathbb{S}$ of candidate cooperating cells
Result: Set $\mathbb{C}$ of cooperating cells
$r'_0 \leftarrow r$
$i \leftarrow 1$
$j \leftarrow 0$
while $(i < K_{max})$ **do**
  if $r'_i/r - r'_j/r > \sum_{k=j+1}^{i} \beta_T^k$ **then**
    $\mathbb{C} \leftarrow \mathbb{C} \cup \{s_{j+1}, \ldots s_i\}$
    $j \leftarrow i$
  end
  $i \leftarrow i + 1$
end

Algorithm 8: General gain-based cooperating cell definition algorithm for JT.

III.4.2 SINR map

Figure III.46 shows the coordination zone (black zone), that is the zone where users are in CoMP mode and require the cooperation of some neighboring cells to be served, under different clustering

strategies for the proposed algorithm. Figure III.47 shows the variation of the coordination area with respect to coordination gain threshold β_T. Observe that the higher the coordination gain threshold β_T, the lower the coordination area.

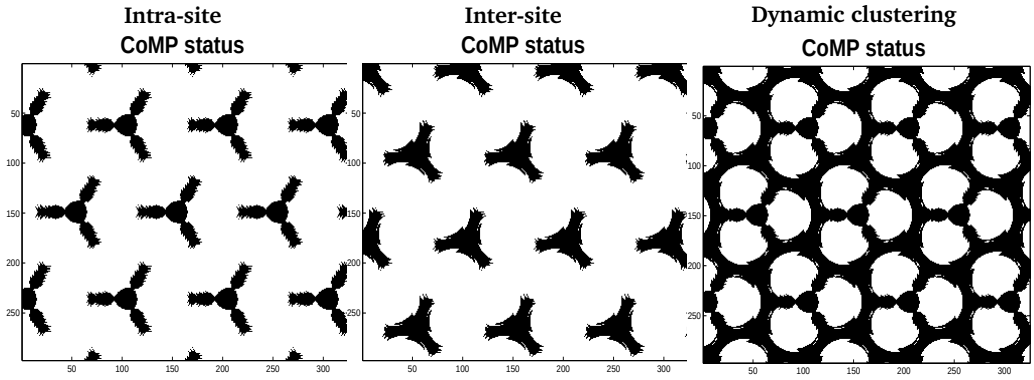


Figure III.46: Coordination area under different clustering methods with $K_{max} = 3$ (black area).

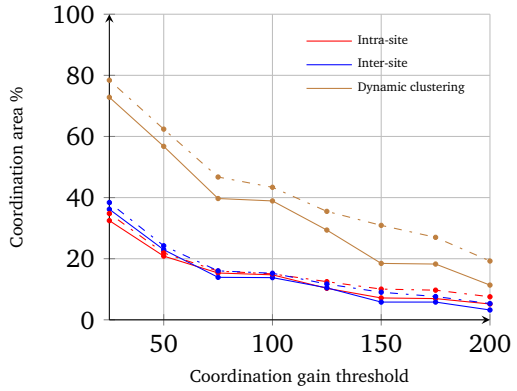


Figure III.47: Mean coordination area under different clustering methods with $K_{max} = 2$ (solid lines) and $K_{max} = 3$ (dashed lines).

Figure III.48 shows the SINR Maps (in dB) according to (III.12) and the service rate Maps (in bit/s/Hz) when applying the proposed algorithm for both transmission schemes as given by (III.13) and (III.14). The mapping between the SINR and the service rate is obtained according to LTE CQI table. Observe that radio conditions become more homogeneous under dynamic clustering.

Figure III.49 shows the mean coordination gain with respect to coordination gain threshold β_T for both transmission schemes, under different clustering strategies with $K_{max} = 2$ and $K_{max} = 3$. This is the mean throughput gain brought to a CoMP user evaluated as the ratio of the mean service rate of CoMP users to the mean service rate of CoMP candidate users when served without coordination. Scheme 2 provides higher mean coordination gain compared to Scheme 1. When $K_{max} = 2$, different clustering methods provide almost same coordination gain. However when $K_{max} = 3$, the highest coordination gain is obtained under dynamic clustering.

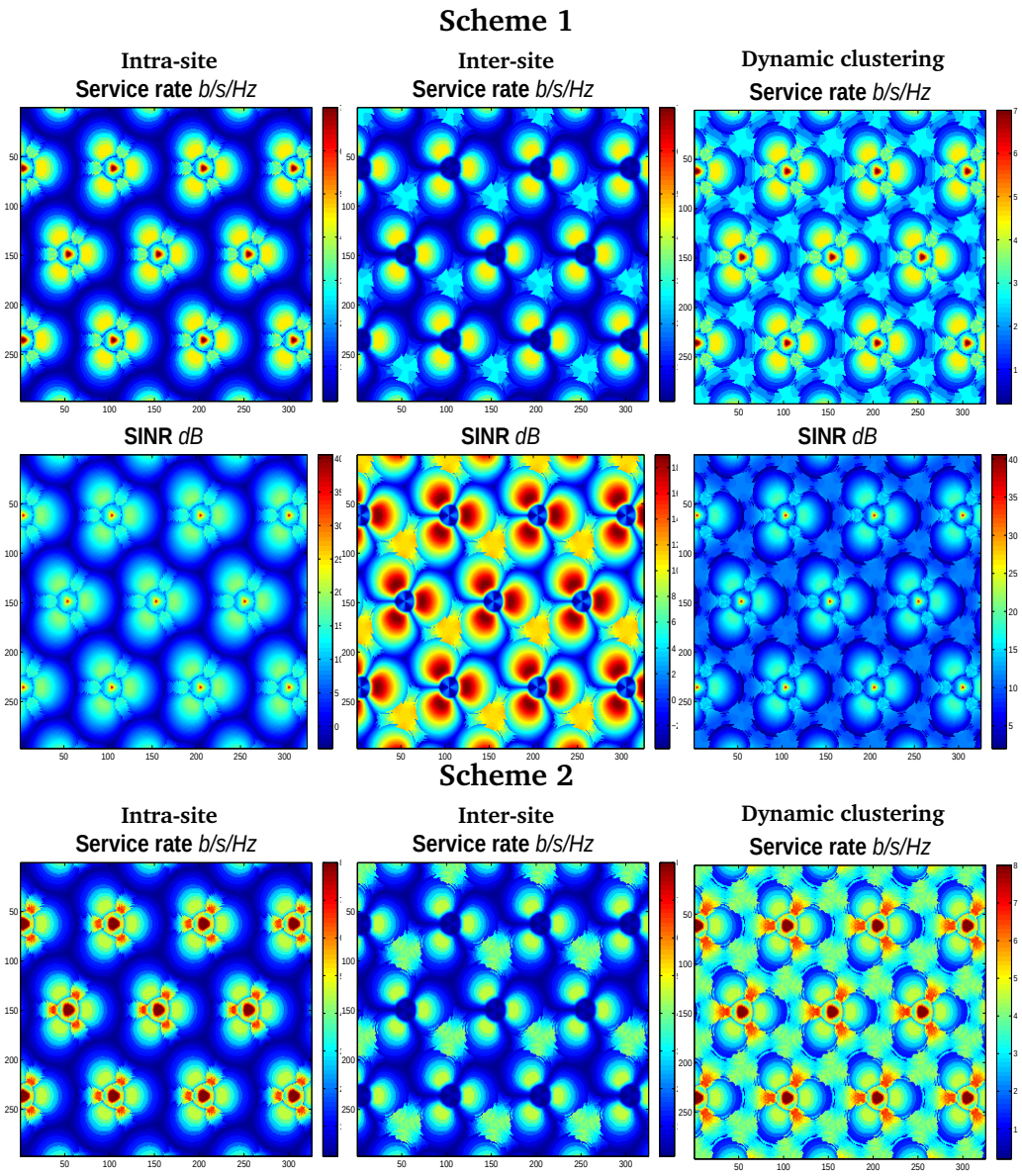


Figure III.48: Service rate and SINR maps for 100% coordination gain threshold.

III.4.2.1 SINR cumulative distribution function

Figure III.50 gives the cumulative distribution function of the SINR in the case without coordination and when activating JT with the proposed cooperating cell definition algorithm 7 under different clustering methods and for different values of maximum coordinated cells K_{max}. Observe that the mean SINR improvement in the case of intra-site coordination and in the case of inter-site coordination with $K_{max} = 3$ is around 1.7 dB. Under dynamic clustering, the mean SINR improves by 5 dB when $K_{max} = 3$.

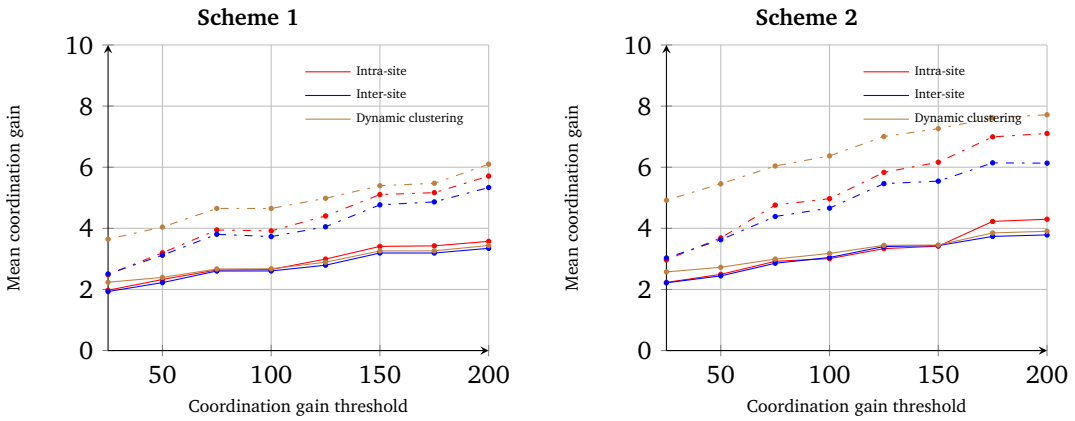


Figure III.49: Mean coordination gain under different clustering methods with $K_{max} = 2$ (solid lines) and $K_{max} = 3$ (dashed lines).

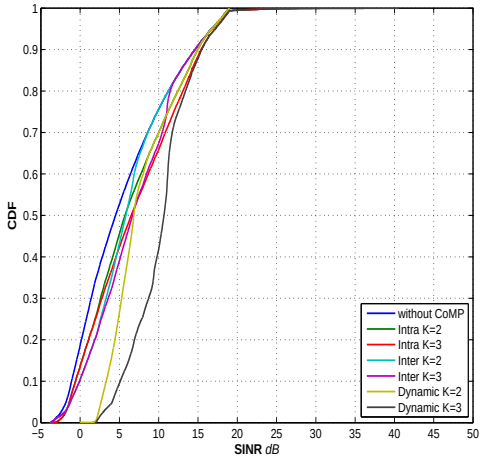


Figure III.50: SINR CDF for proposed JT CoMP

III.4.3 Flow level model

Note that all results obtained for the scenario of a single coordination zone see §III.3.4.1 and for the scenario of intra-site coordination see §III.3.4.2 are valid for the case of gain-based JT.

By applying the gain-based cooperating cell definition algorithm in a homogeneous symmetric network, we can guarantee that the mean rate gain of a CoMP user involving n cells in the transmission is higher than n which satisfies condition (III.29).

III.4.4 Numerical results

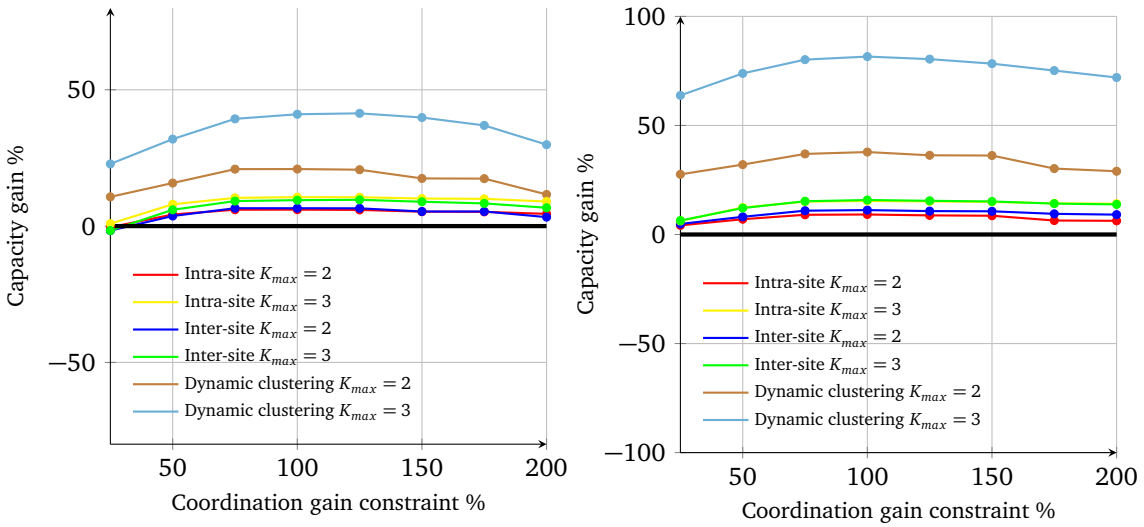
III.4.4.1 Capacity

We evaluate the capacity gain when applying the proposed cooperating cell definition algorithm 7 with different coordination gain constraint β_T, under different clustering methods and different maximum coordinated cells: $K_{max} = 2$, $K_{max} = 3$ for a regular hexagonal network (21 cells) as well as an

hexagonal network with eNodeB position and antenna direction dispersion. We study two transmission schemes and evaluate for each case the mean service rate in each cell at full interference, when all neighboring cells are transmitting (generating interference). We estimate the stability condition under a RR scheduler in the case without coordination (classical RR scheduler) as well as in the case when applying JT (RR iterative scheduler); it is nothing more than the ratio of the total traffic to the mean service rate. Then, we compare the stability condition in the presence of JT coordination to that in the absence of coordination techniques. Note that in the case when applying JT, the additional traffic induced by the cooperation and coming from neighboring cells is taken into account. The overall capacity gain in the network is evaluated as the arithmetic mean of capacity gains in different cells.

a) Regular hexagonal network

Figure III.51 shows the case of a regular hexagonal network. Observe that the highest capacity gains under RR scheduling strategy is obtained in the case of dynamic clustering. Obviously, Scheme 2 provides higher coordination gain and then higher capacity gains compared to Scheme 1. Observe that the maximum capacity gains are obtained around coordination gain constraint $\beta_T = 100\%$ and are given by Table III.3, for Scheme 1 and Scheme 2.



(a) Scheme 1

(b) Scheme 2

Figure III.51: Capacity gain of the proposed algorithm with respect to the coordination gain constraint in a regular hexagonal network.

b) Hexagonal network with position and antenna direction dispersion

Now, we add dispersion to the position of each eNodeB as well as the antenna direction of each sector in an hexagonal network. The new SINR map without any coordination between cells is given by Figure III.31. This is closer to a real network. The capacity gains for coordination gain constraint $\beta_T = 100\%$ are given by Table III.4, for Scheme 1 and Scheme 2. Note that these gains are of the same order of these obtained in a regular hexagonal network.

	Capacity Gain of the proposed algorithm with 100% coordination gain constraint
Intra-site $K_{max} = 2$	6.02%
Intra-site $K_{max} = 3$	10.6%
Inter-site $K_{max} = 2$	6.55%
Inter-site $K_{max} = 3$	9.52%
Dynamic $K_{max} = 2$	20.92%
Dynamic $K_{max} = 3$	41.02%

(a) Scheme 1

	Capacity Gain of the proposed algorithm with 100% coordination gain constraint
Intra-site $K_{max} = 2$	9.17 %
Intra-site $K_{max} = 3$	15.37 %
Inter-site $K_{max} = 2$	11.19 %
Inter-site $K_{max} = 3$	15.82 %
Dynamic $K_{max} = 2$	37.80 %
Dynamic $K_{max} = 3$	81.57 %

(b) Scheme 2

Table III.3: Capacity Gain of the proposed algorithm with 100 % coordination gain constraint in a regular hexagonal network.

	Capacity Gain of the proposed algorithm with 100% coordination constraint constraint
Intra-site $K_{max} = 2$	5.73%
Intra-site $K_{max} = 3$	9.89%
Inter-site $K_{max} = 2$	7.46%
Inter-site $K_{max} = 3$	9%
Dynamic $K_{max} = 2$	20.14%
Dynamic $K_{max} = 3$	33.58%

(a) Scheme 1

	Capacity Gain of the proposed algorithm with 100% coordination constraint
Intra-site $K_{max} = 2$	8.99 %
Intra-site $K_{max} = 3$	14.9 %
Inter-site $K_{max} = 2$	12.77 %
Inter-site $K_{max} = 3$	16.8 %
Dynamic $K_{max} = 2$	35.99 %
Dynamic $K_{max} = 3$	68.82 %

(b) Scheme 2

Table III.4: Capacity Gain of the proposed algorithm with 100 % coordination gain constraint in an hexagonal network with position and antenna direction dispersion.

III.4.4.2 System level simulations under iterative scheduler

a) Round robin

We first evaluate the performance of the proposed algorithm under iterative RR scheduler, using the LTE simulator described in Appendix A. We measure the performance in terms of mean user throughput as a function of the offered traffic, in the case of intra-site coordination, inter-site coordination as well as in the case of dynamic clustering. Figure III.52 shows the results for both transmission schemes, and for $K_{max} = 2$ and $K_{max} = 3$. Black line represents the case without coordination (under classical RR) and dashed lines represent the case when applying gain-based JT (under iterative RR). Results are very close to those obtained by analysis. Observe that capacity gains are close to those in Table III.4. There is no stability condition degradation when applying gain-based JT under an iterative RR scheduler.

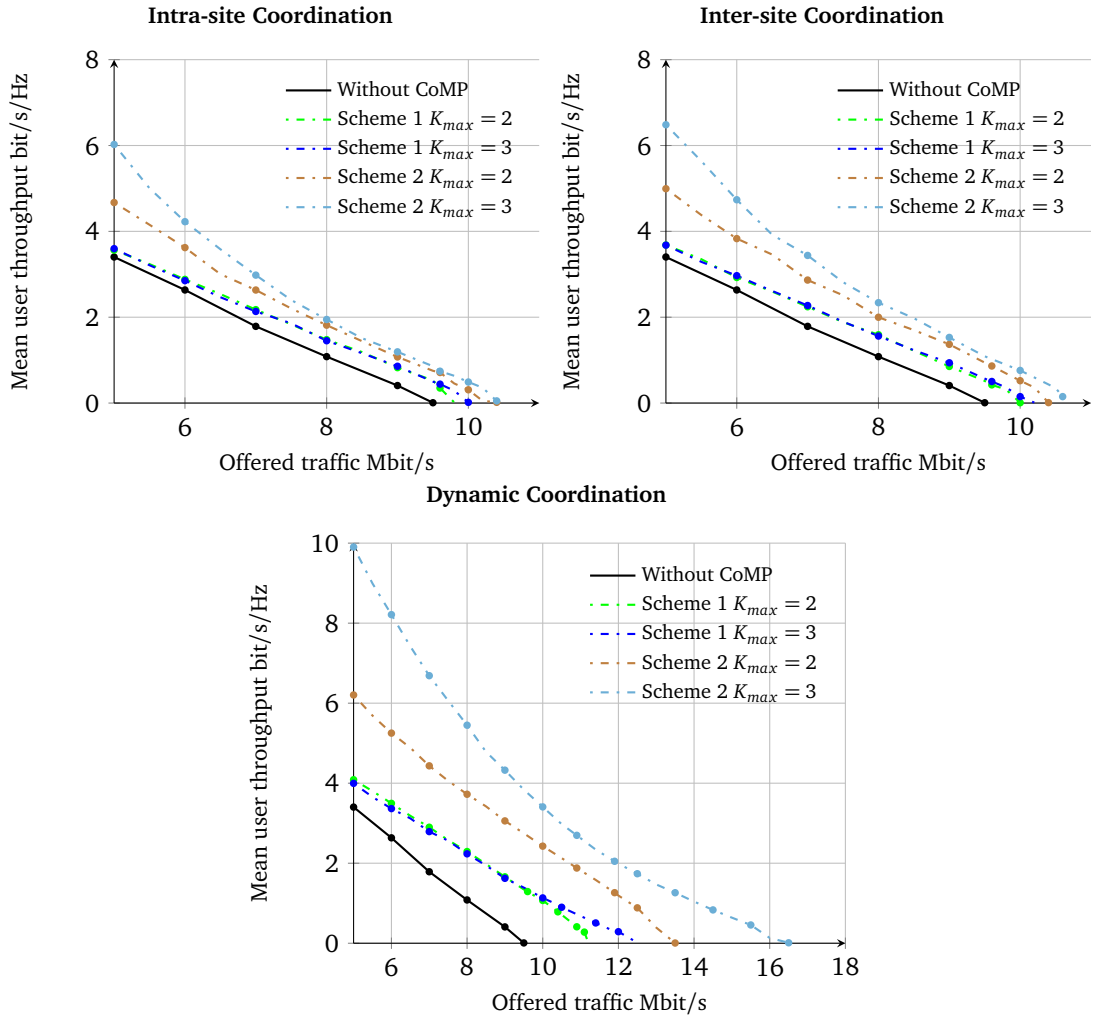


Figure III.52: Throughput performance obtained by system-level simulation for gain-based JT under round-robin scheduler.

b) Proportional fair

Advanced networks use opportunistic schedulers in order to exploit the opportunistic gain which is a relatively high gain. We evaluate now the proposed cooperating cell definition algorithm under an iterative PF scheduler. Figure III.53 and Figure III.54 show the performance under fixed clustering: intra-site coordination and inter-site coordination.

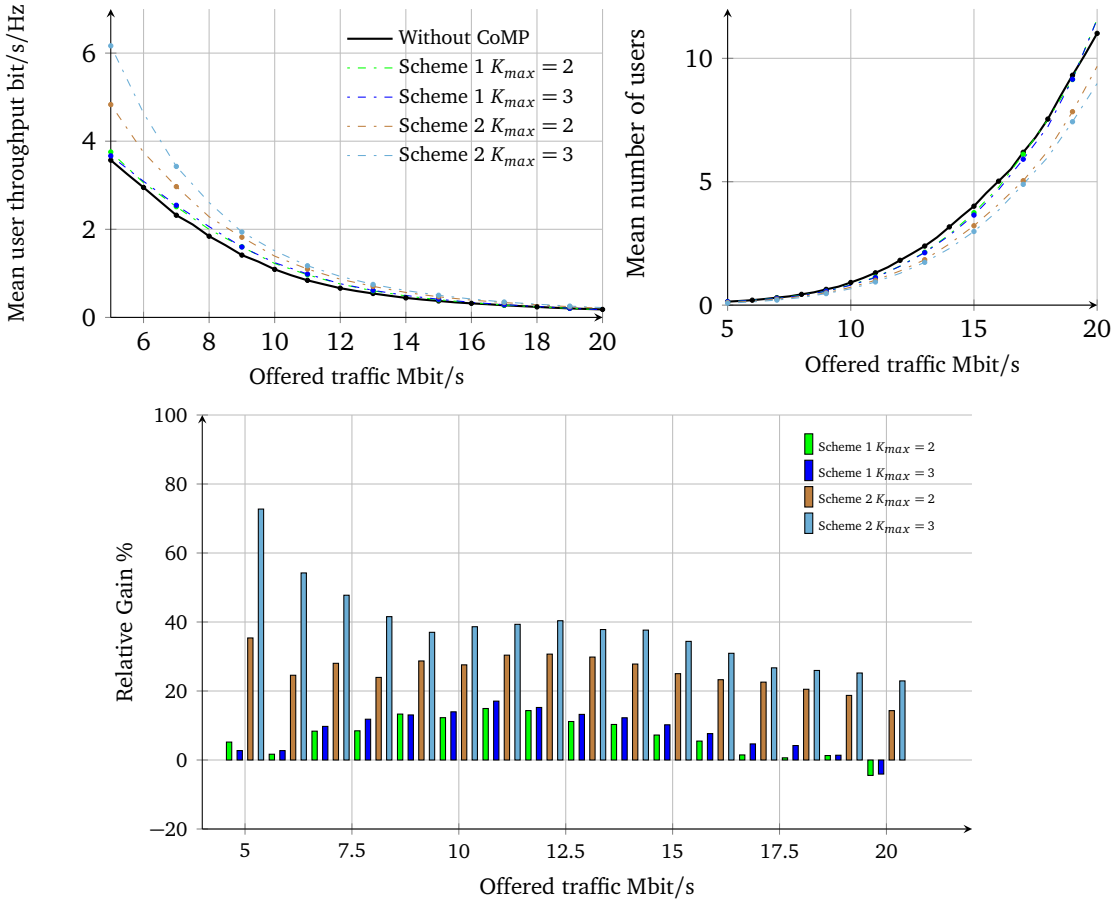


Figure III.53: Throughput performance obtained by system-level simulation for Intra-site gain-based JT under iterative scheduler.

When Scheme 1 is applied the highest gains are close to 20%. Note that these gains are higher than those obtained when applying classical cooperating cell definition, see Section I.5.3. However surprisingly, there is performance degradation at high load. Observe that at 20 Mb/s offered traffic, there is a loss of 5% in the case of intra-site coordination and 20% in the case of inter-site coordination.

Figure III.55 shows the performance under dynamic clustering. The highest gains obtained when transmission scheme 1 is applied are around 50%. Similarly, there is performance degradation at high load. The loss is around 45% at 20 Mbits/s offered traffic. However, there is no performance degradation when transmission scheme 2 is applied, under fixed and dynamic clustering.

Note that in the case of Scheme 1, gains at medium load are close to those obtained when considering previous cooperating cell definition algorithm with $\delta P = 6$ dB, see Section d). In the case of Scheme 2, performance at medium load are closer to those obtained when $\delta P = 9$ dB under fixed

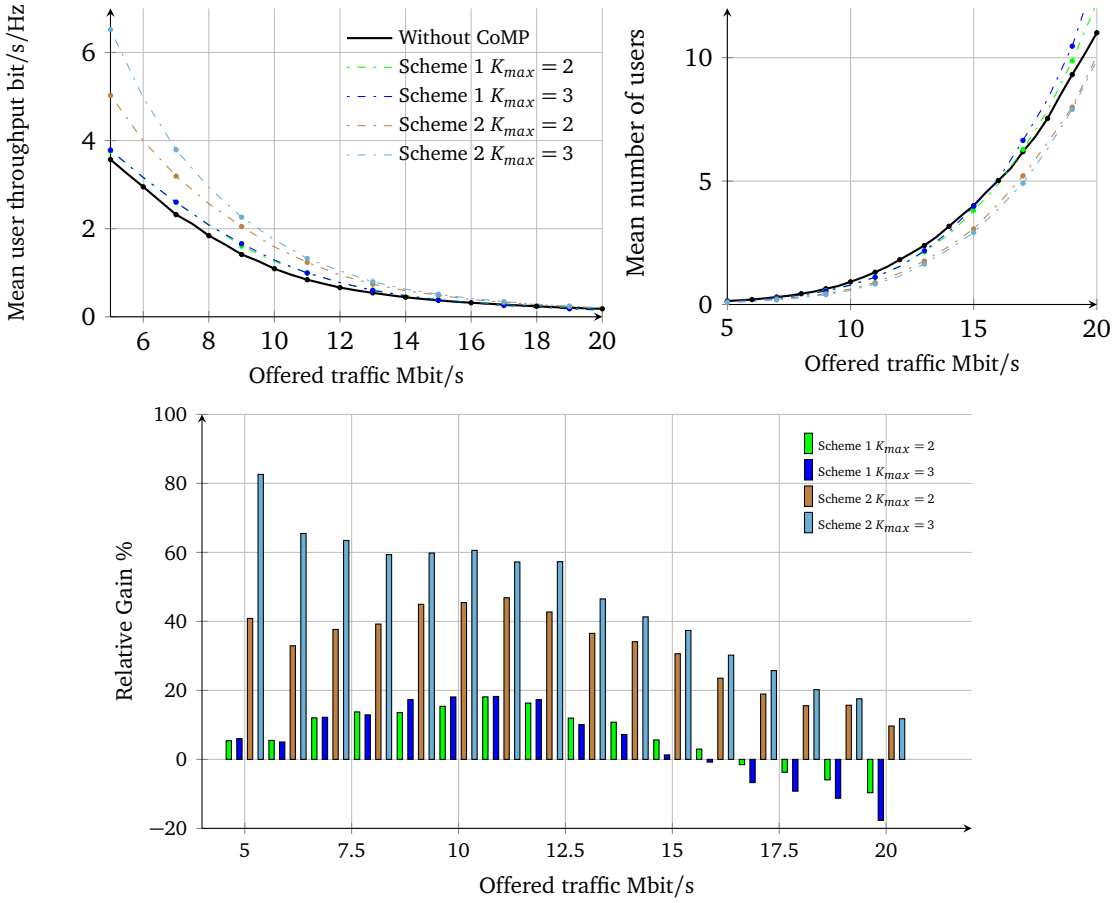


Figure III.54: Throughput performance obtained by system-level simulation for Inter-site gain-based JT under iterative scheduler.

clustering strategies, and to those obtained when $\delta P = 12$-15 dB under dynamic clustering strategy. Thus, there is not a fixed value of δP that corresponds to a coordination gain of 100%. The precise value of δP depends on the scenario, the clustering method and the transmission scheme.

III.4.5 Scheduling issues: Opportunistic gain vs Coordination gain

In the previous sections, we have seen that gain-based JT coordination brings performance and capacity gains under an iterative RR scheduler. However, performance degradation may occur at high load under an iterative PF scheduler. The scheduling strategy is a key point in the presence of coordination mechanisms. It strongly impacts the performance. Consider the case of a classical network without coordination operating under a classical scheduler (RR or PF). There is a significant throughput gain brought by the PF scheduler compared to the RR scheduler according to §II.2.3.1c) and Section II.2.3.2.

Now when JT coordination techniques and more precisely gain-based JT techniques are applied under an iterative RR scheduling strategy, there is a coordination gain brought by the coordination between cells which allows to enhance the network performance in terms of user throughput and

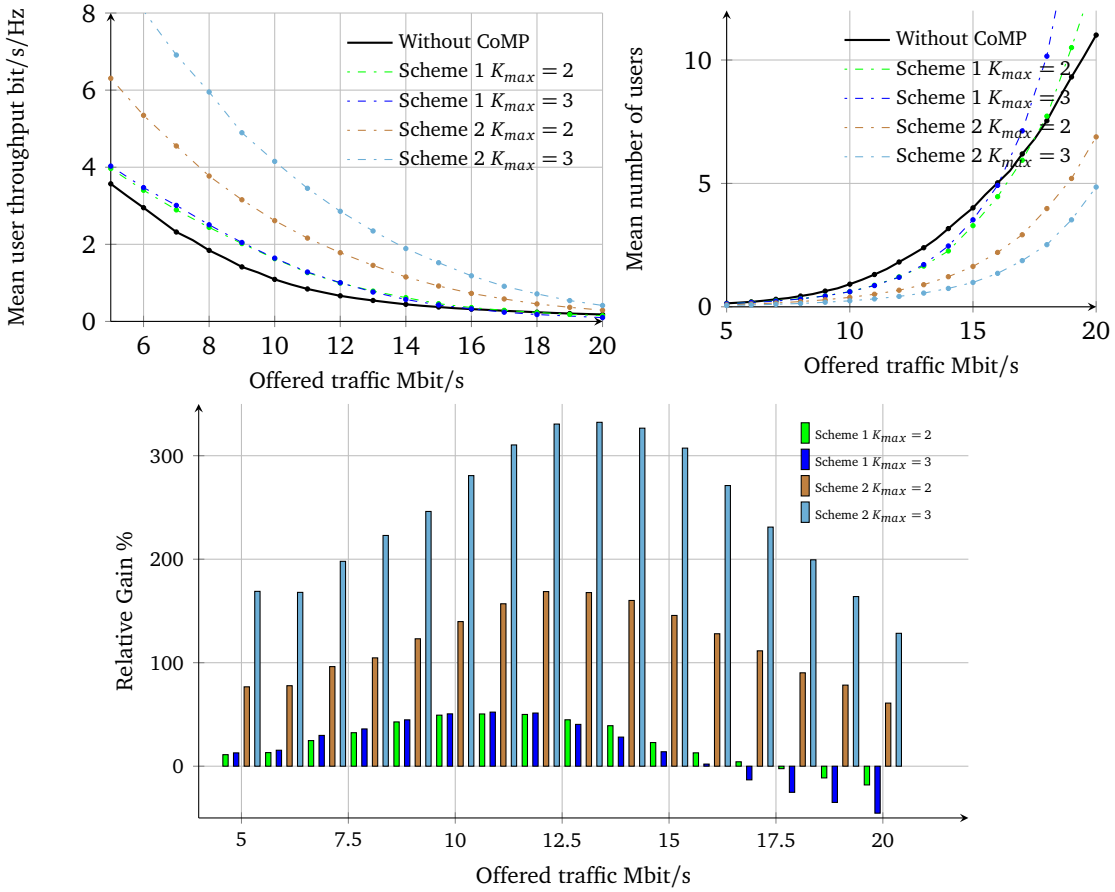


Figure III.55: Throughput performance obtained by system-level simulation for Dynamic gain-based JT.

capacity.

Under a PF scheduling strategy, the network is benefiting from a huge opportunistic gain. This gain may be reduced when applying an iterative scheduler. Consider the case of a cluster of 3 cells. We denote by n_1, n_2 and n_3 the number of users in the considered cells. In an iterative way, the first cell selects its best user (to be scheduled) among a set of n_1 users. The second cell must select its best user among a subset of $n'_2 \leq n_2$ users; all users requiring a transmission from the serving cell and all cooperating cells of the previous selected user are excluded from this subset, and so on... This scheduling strategy leads to the following rate gains (see Section II.2.2.4): $G(n_1)$ in the first cell, $G(n'_2 \leq n_2) \leq G(n_2)$ in the second cell and $G(n'_3 \leq n_3) \leq G(n_3)$ in the third cell. Observe that the total rate gain ($G(n_1)$, $G(n'_2)$, $G(n'_3)$) when applying an iterative PF strategy is less than the total rate gain ($G(n_1)$, $G(n_2)$, $G(n_3)$) obtained when each cell takes its scheduling decision independently without any coordination between cells. So, there is a loss due to the diminution of the opportunistic gain, together with a coordination gain. This results in performance degradation at high load when the coordination gain is not high enough in order to compensate the loss coming from a sub-optimal scheduling strategy, which is the case of Scheme 1. Thus, we should ensure that the introduction of a new scheduler, does not degrade existing performance.

III.4.6 Global scheduler with dynamic CoMP status

In order to cope with this problem, we consider in this section a global scheduler which selects each TTI the best combination of users that maximizes the sum of PF metrics. We assume that the status of a CoMP user is dynamic so that the coordination may be activated or deactivated according to the scheduling decision. The scheduling decisions in each cluster are taken at a central entity which has full knowledge of the channel radio conditions of each user associated with a cell taking part in the cooperative cluster. The following optimization problem is then formulated for each cluster of cells and for each sub-band of frequency f (subset of RBs):

$\forall f \in [1,2,\dots N_{RB}]$:

Scheme 1

$$\max \sum_{s=1}^{N} \sum_{u=1}^{n_j} \frac{1}{\bar{R}_u} \boldsymbol{F}\left(\frac{b_{u,s}^f P_{u,s}^f + \sum_{c \in \mathbb{C}_u} b_{u,c}^f P_{u,c}^f}{\sum_{i \neq c \in \mathbb{C}_u} P_{u,i}^f + \sum_{c \in \mathbb{C}_u} (1 - b_{u,c}^f) P_{u,c}^f + \mathcal{N}} \right)$$

Scheme 2

$$\max \sum_{s=1}^{N} \sum_{u=1}^{n_j} \frac{1}{\bar{R}_u} \boldsymbol{F}\left(\frac{b_{u,s}^f P_{u,s}^f}{\sum\limits_{i \neq c \in \mathbb{C}_u} P_{u,i}^f + \sum\limits_{c \in \mathbb{C}_u} (1 - b_{u,c}^f) P_{u,c}^f + \mathcal{N}} \right) + \sum_{c \in \mathbb{C}_u} \frac{1}{\bar{R}_u} \boldsymbol{F}\left(\frac{b_{u,c}^f P_{u,c}^f}{\sum\limits_{i \neq c \in \mathbb{C}_u} P_{u,i}^f + \sum\limits_{c \in \mathbb{C}_u} (1 - b_{u,c}^f) P_{u,c}^f + \mathcal{N}} \right)$$

$b_{u,x}^f \in \{0,1\}$ is a variable to indicate whether cell x is transmitting to user u on frequency sub-band f in the current TTI or not.

$$b_{u,x}^f = \begin{cases} 0 & \text{cell } x \text{ is not transmitting to user } u \text{ .} \\ \\ 1 & \text{cell } x \text{ is transmitting to user } u \text{ .} \end{cases}$$

We denote by $P_{u,x}^f$ the signal power received by user u from cell x over frequency sub-band f. $\bar{R}_u$ is the mean data rate of user u evaluated through an exponentially smoothed average over time.

In order to guarantee that a given subset of resources is allocated to only one user at a time, the following constraints should be fulfilled:

$\forall s \in [1,2,\dots N]$, $\forall f \in [1,2,\dots N_{RB}]$:

$$\sum_{u \in \mathbb{U}_s} b_{u,s}^f \leq 1 \tag{III.32}$$

where $\mathbb{U}_s$ is the set of users requiring transmission from cell s, including its own users (non-CoMP users and CoMP users) as well as CoMP users from coordinated cells. Cell s may transmit to one user on a given sub-band f at a time.

In order to reduce the complexity of the problem, we suppose that: $\forall c \in \mathbb{C}_u : b_{u,c}^f = b'_{u,f}$. Note that the set of cooperating cells $\mathbb{C}_u$ of user u should be predefined according to Algorithm 7. In other words, for each CoMP user, coordination is either activated for all cooperating cells or deactivated for all of them. When $b'_{u,f} = 0$, user u is served in non-CoMP mode (normal service mode) and when $b'_{u,f} = 1$ user u is served in CoMP mode. Note that in this case the optimization problem becomes:

$$\max \sum_{s=1}^{N} \left(\sum_{u \in \mathbb{U}_{\text{NCoMP},s}} b_{u,s}^f \frac{r_{u,f}}{\bar{R}_u} + \sum_{u \in \mathbb{U}_{\text{CoMP},s}} b_{u,s}^f \left((1 - b'_{u,f}) \frac{r_{u,f}}{\bar{R}_u} + b'_{u,f} \frac{r'_{u,f}}{\bar{R}_u} \right) \right)$$

where $\mathbb{U}_{\text{NCoMP,s}}$ is the set of non-CoMP users associated with serving cell s and $\mathbb{U}_{\text{CoMP,s}}$ is the set of CoMP users associated with serving cell s. $r_{u,f}$ is the instantaneous transmission rate of user u when served without any coordination on sub-band f as given by (III.11), and $r'_{u,f}$ is the instantaneous rate of user u when served by multiple cells at the same time on sub-band f as given by (III.13) and (III.14). Knowing $r_{u,f}$ of each user u (CoMP and non-CoMP users) and $r'_{u,f}$ of each CoMP user thanks to CQI feedback, we can determine the optimal values of $b^f_{u,s}$ (for all CoMP and non-CoMP users belonging to the considered cluster) and $b'_{u,f}$ (for all CoMP users belonging to the considered cluster) given by (III.32).

This scheduler provides better performance than the iterative scheduler. However, the complexity of such a scheduler increases significantly with the number of users and the size of the cluster. Moreover the fairness is no more achieved for each user alone; the scheduling of a given user depends also on the scheduling metrics of the users in the neighboring cells which can be scheduled together with the considered user.

III.4.7 Numerical results

III.4.7.1 System level simulations under global scheduler

Figure III.56 and Figure III.57 shows the performance under global scheduler, obtained using the LTE simulator described in Appendix A, for intra-site coordination and inter-site coordination respectively. For the sake of simplicity, we consider only one frequency sub-band in the case of fixed clustering where only a limited number of cells ($N = 3$) constitute a cooperative cluster. We measure performance in terms of mean user throughput and relative gain with respect to the case without coordination for Scheme 1 and Scheme 2. Observe that there is performance improvement at high load when the mean number of users strongly impacting the opportunistic gain becomes large and degradation is avoided when Scheme 1 is applied. Gains at high load are between 5% and 10% in this case (Scheme 1). Even though iterative scheduler provides comparable performance to global scheduler at low and medium load it is still insufficient in a highly loaded network where the opportunistic gain can not be well exploited under a sub-optimal iterative scheduler.

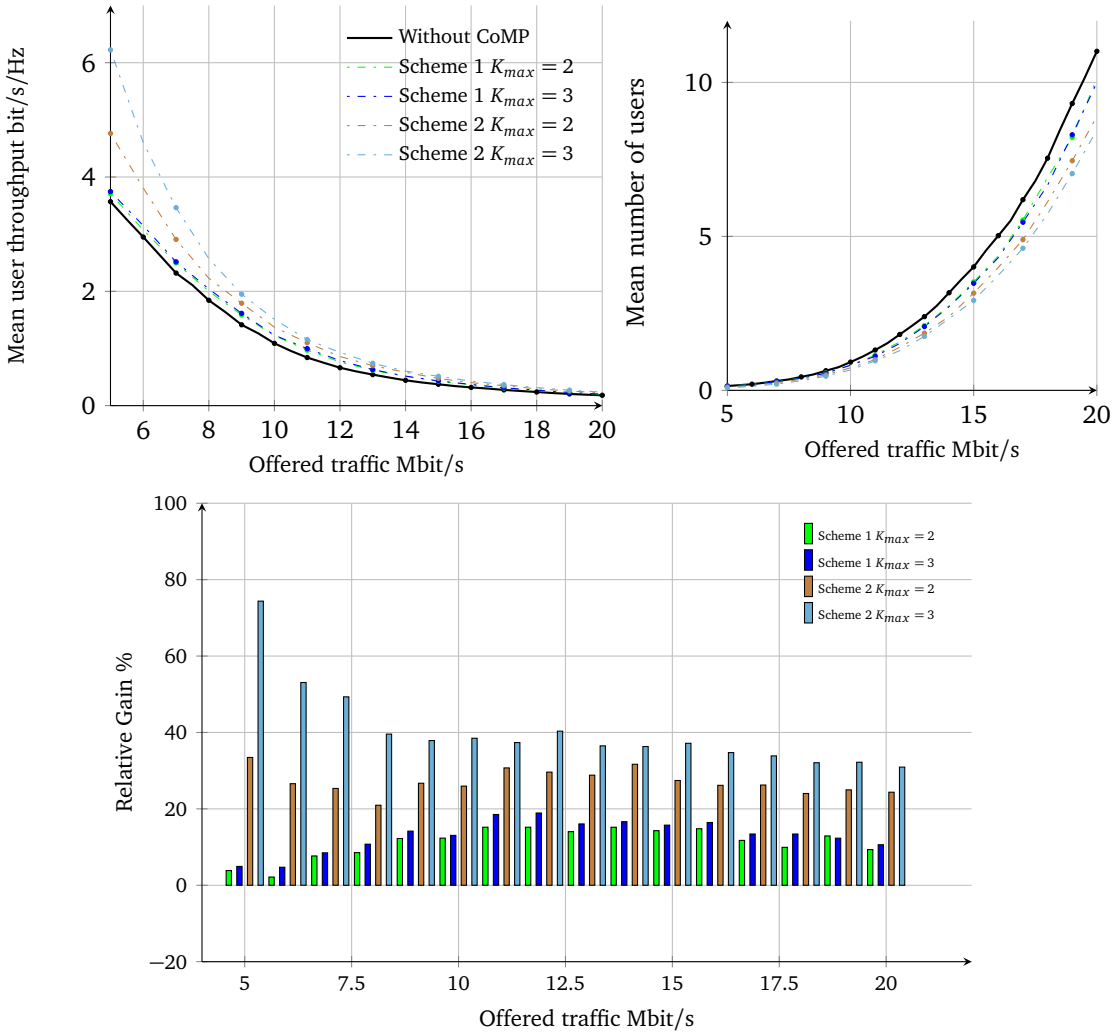


Figure III.56: Throughput performance obtained by system-level simulation for Intra-site gain-based JT under global scheduler with dynamic CoMP status.

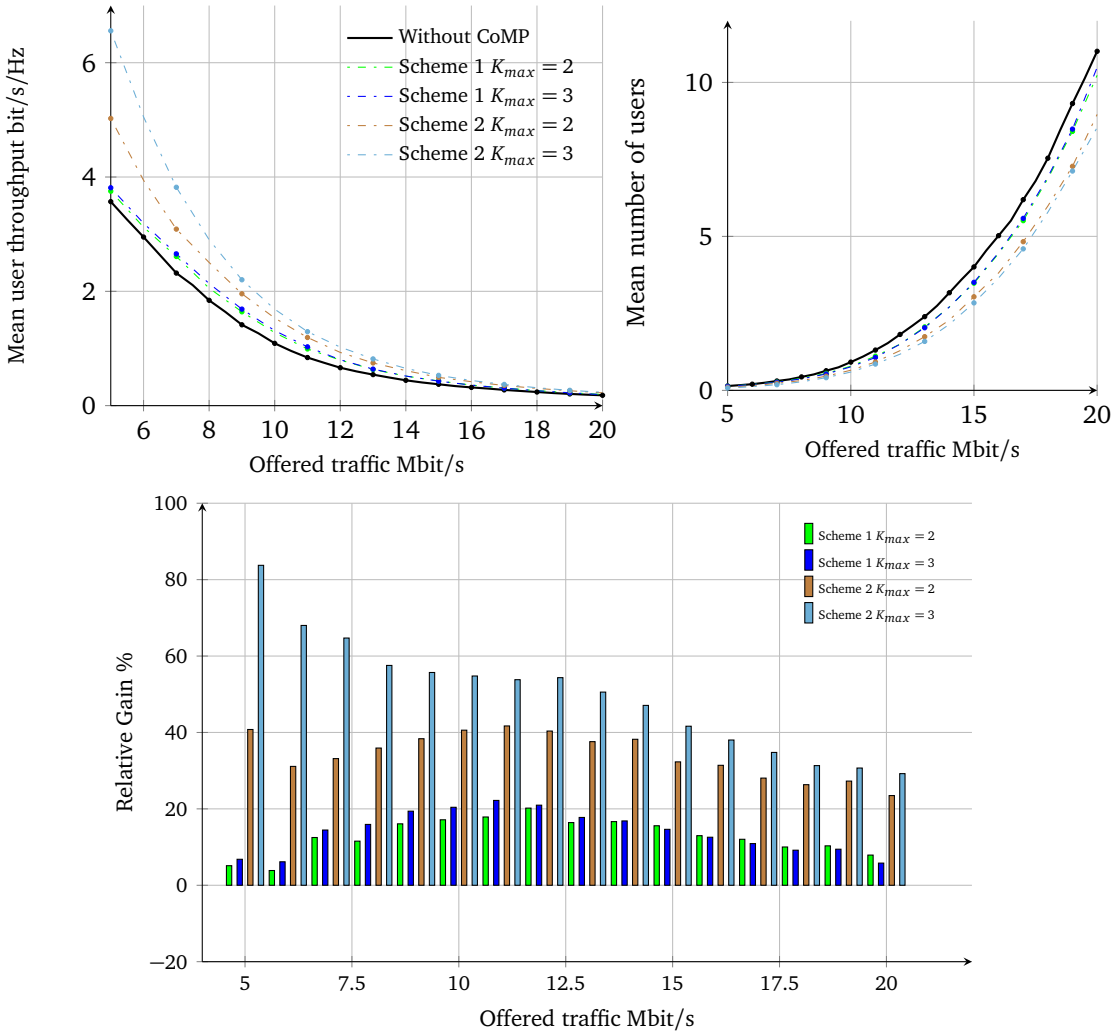


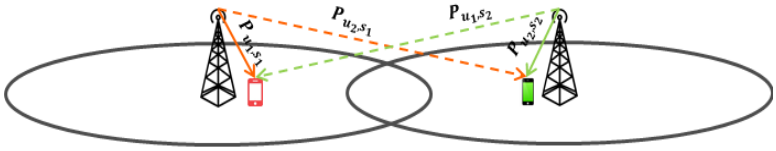
Figure III.57: Throughput performance obtained by system-level simulation for Inter-site gain-based JT under global scheduler with dynamic CoMP status.

III.5 Dynamic point blanking

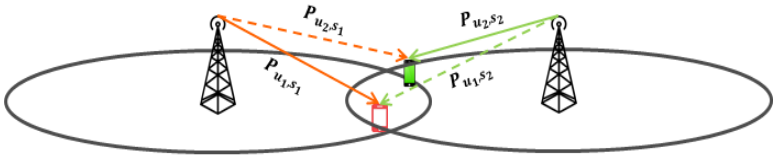
III.5.1 Description

LTE is designed to operate with a frequency reuse factor of one to maximize the spectral efficiency. However, data and control channels can experience a significant level of interference from neighboring cells, which reduces the achievable spectral efficiency, especially at the cell edge.

In order to demonstrate further the significance of interference we consider the simple case of a cellular system with two cells (s_1 and s_2) and one active user per cell (u_1 and u_2 respectively). Each user receives a signal from its serving cell and an interference from the other cell. In the first case, each user is close to its respective eNodeB as illustrated by Figure III.58a. The interference received from the neighboring cell is small compared to the power from the serving cell. In the second case, illustrated by Figure III.58b, users are located at the cell edge and the interference level seen by each user from the neighboring cell is similar to the power level received from the serving cell.



(a) Cell center users



(b) Cell edge users

Figure III.58: Different interference cases.

We denote by $P^f_{u,x}(t)$ the signal power received by user u from cell x over frequency sub-band f at time t. Note that $P^f_{u,x}(t) = P^x_{\text{tx}} G^f_{u,x}(t)$, where P^x_{tx} is the transmit power from cell x and $G^f_{u,x}(t)$ is the channel gain from cell x to user u including pathloss, see Appendix A. Consequently, the capacity of the system at time t can be written according to Shannon as follows:

$$R(t) = W \log_2\left(1 + \frac{P_{u_1,s_1}(t)}{P_{u_1,s_2}(t) + \mathcal{N}}\right) + W \log_2\left(1 + \frac{P_{u_2,s_2}(t)}{P_{u_2,s_1}(t) + \mathcal{N}}\right).$$

Actually, the optimal transmit powers that allow to achieve maximum capacity is different for two considered cases. In the first case the maximum throughput is achieved when both eNodeBs transmit at maximum power, while in the second the maximum capacity is reached by allowing only one eNodeB to transmit. It has been shown that the optimal power allocation for maximum capacity for this scenario with two base stations is binary in the general case; this means that either both base stations should be operating at maximum power in a given RB, or one of them should be turned off completely in that RB [33]. Hence the interest to mute some base stations when necessary.

DPB consists in identifying and dynamically muting the principal interferer(s) to the UEs in the coordination area. By muting the dominant interferer the SINR of the UE may be significantly im-

proved as dominant interferers may represent the majority of the whole interference. It seems unfair to UEs connected to the muted cell. However, since scheduling is dynamically performed, those UEs may also benefit from muting other cells in subsequent sub-frames.

We proceed by a similar way, as in Section III.3.1. Each cooperating cell $c \in \mathbb{C}$, the set of cooperating cells of the considered CoMP user, is muted. The perceived SINR by a given user can be written as follows:

$$\frac{P_s}{\sum_{i \neq c \in \mathbb{C}} P_i + \mathcal{N}}. \tag{III.33}$$

The resulting transmission rate is the given by:

$$F\left(\frac{P_s}{\sum_{i \neq c \in \mathbb{C}} P_i + \mathcal{N}}\right). \tag{III.34}$$

III.5.2 SINR cumulative distribution function

Figure III.59 and Figure III.60 give the cumulative distribution function of the SINR in the case without coordination (III.10) and when activating DPB with different values of δP as given by (III.33), under different clustering methods and for different values of maximum coordinated cells K_{max}. Figure III.59 shows the case of static clustering while Figure III.60 shows the case of dynamic clustering. These figures allow to evaluate the performance in terms of coverage improvement. Observe that most of the SINR gain comes from the elimination of the interference. Adding a new useful signal provides only limited additional SINR gain. In the case of intra-site coordination with $\delta P = 18$ dB, the mean SINR is improved by 1.5 dB when $K_{max} = 2$ and by 2.2 dB when $K_{max} = 3$. In the case of inter-site coordination with $\delta P = 18$ dB, the increase is around 1.5 dB when $K_{max} = 2$ and 3.3 dB when $K_{max} = 3$. Dynamic clustering provides the highest SINR improvement where the mean value increases by 6.8 dB when $K_{max} = 3$ and by 8.8 dB when there is no limit on the number of coordinated cells.

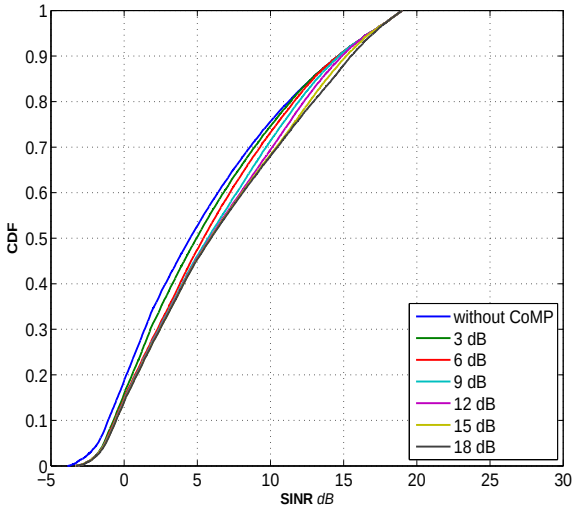
Observe that most of the gain comes from the elimination of the interference, where interference cancellation provides between 70% and 90% of the total gain brought by adding a new useful signal, see Table III.5. Note that the actual mean SINR can be even higher than the one represented in Figure III.59 and Figure III.60. Indeed, these cumulative distribution functions represent the worst case scenario of full interference (case of very high load), where SINR values are estimated in each point of the cell supposing that all neighboring non cooperating cells are transmitting. However, CoMP users as well as non CoMP users may benefit from the cooperation (muting) of some neighboring cells even though those cells do not take part in their cooperative cluster.

	$\delta SINR_{JT}$	$\delta SINR_{DPB}$	$\delta SINR_{DPB}/\delta SINR_{JT}$
Intra-site coordination $K_{max} = 2$	1.8	1.5	83 %
Intra-site coordination $K_{max} = 3$	2.8	2.2	78 %
Inter-site coordination $K_{max} = 2$	2.3	1.5	65 %
Inter-site coordination $K_{max} = 3$	4.5	3.3	73 %
Dynamic clustering $K_{max} = 3$	7.6	6.7	88 %
Dynamic clustering, unlimited number of coordinated cells	10.2	8.8	86 %

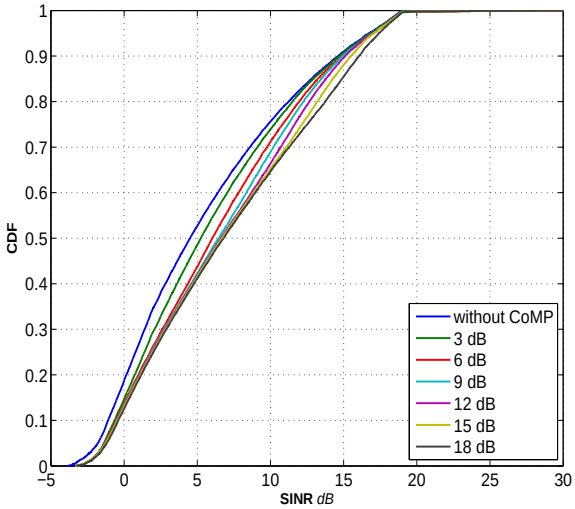
Table III.5: Mean SINR improvement when applying JT and DPB with $\delta P = 18$ dB.

Intra-site coordination

$K_{max} = 2$ **maximum coordinated cells**

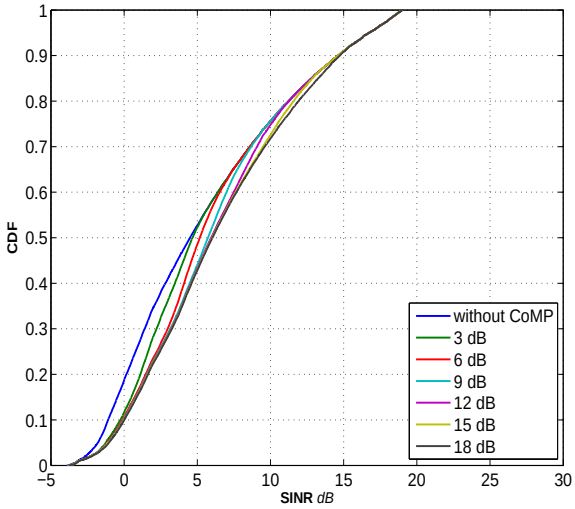


$K_{max} = 3$ **maximum coordinated cells**

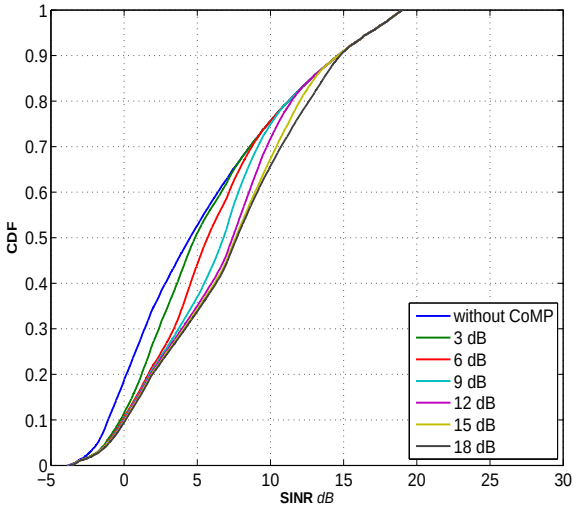


Inter-site coordination

$K_{max} = 2$ **maximum coordinated cells**

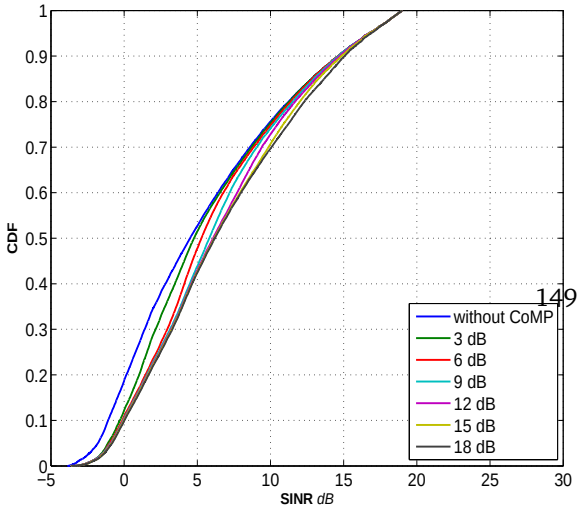


$K_{max} = 3$ **maximum coordinated cells**



Intra-site and Inter-site coordination

$K_{max} = 2$ **maximum coordinated cells**



$K_{max} = 3$ **maximum coordinated cells**

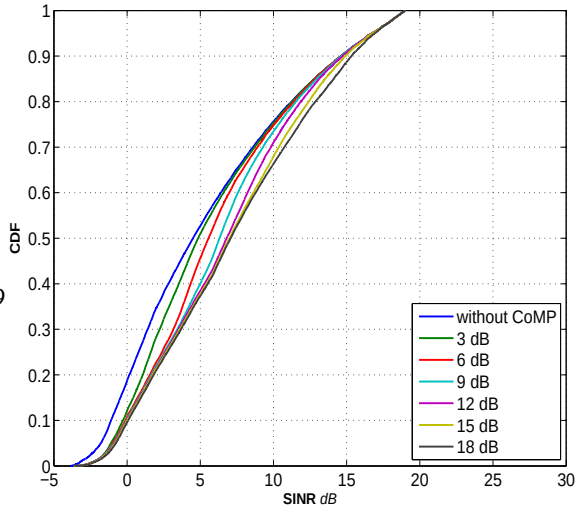
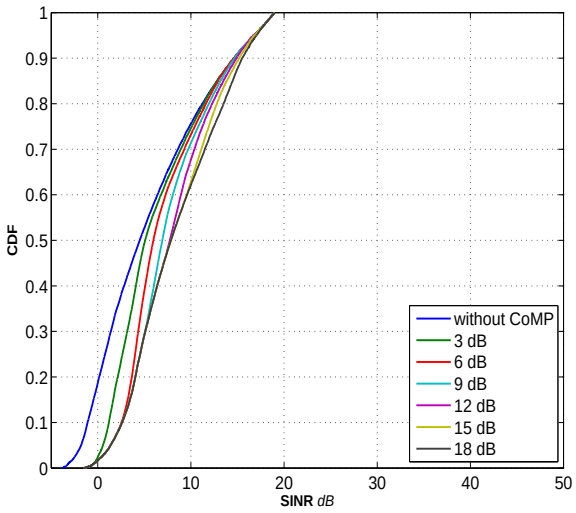


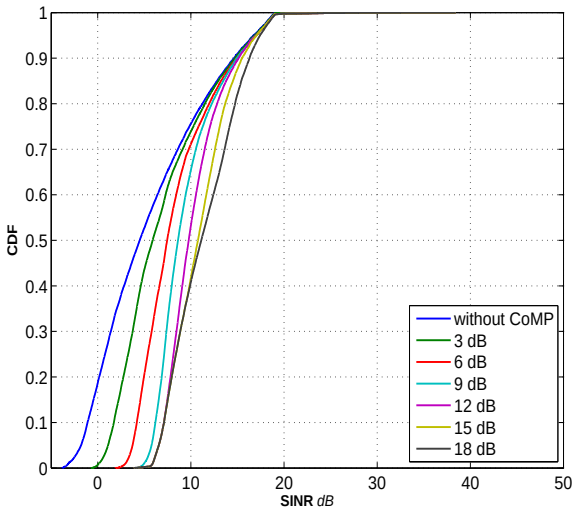
Figure III.59: SINR CDF for fixed clustering DPB CoMP

Dynamic clustering

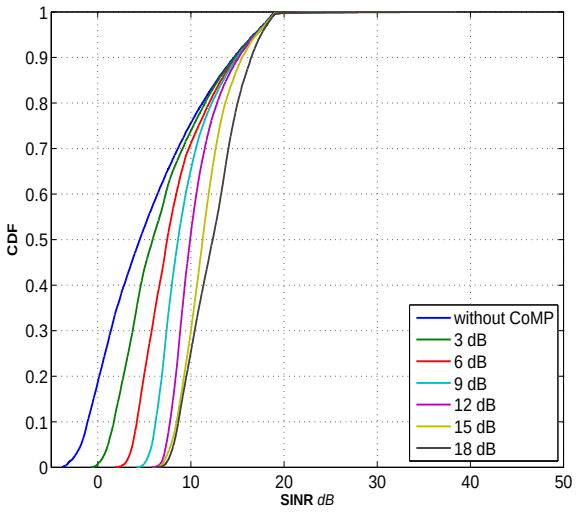
$K_{max} = 2$ **maximum coordinated cells**



$K_{max} = 3$ **maximum coordinated cells**



$K = 4$ **maximum coordinated cells**



unlimited coordinated cells number

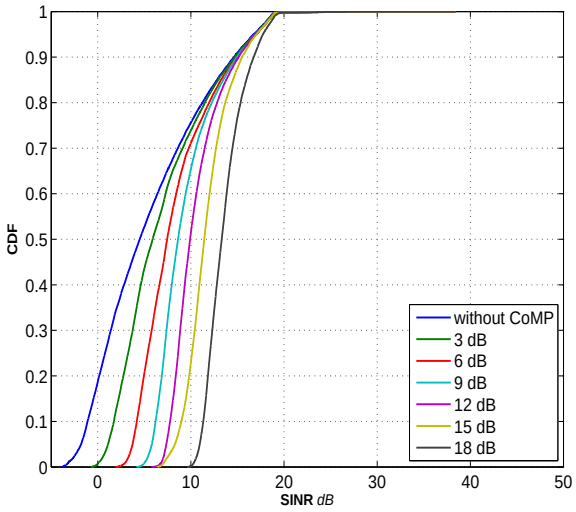


Figure III.60: SINR CDF for Dynamic clustering DPB

III.5.3 Scheduling schemes

III.5.3.1 Static allocation

This strategy consists in allocating a fixed fraction of resources to CoMP and non-CoMP users. All users selected to be served in CoMP mode use this pre-configured frequency sub-band. The simplicity of this strategy is at the cost of limited performance.

Non-CoMP users are scheduled according to classical scheduling strategies on their dedicated frequency sub-band while CoMP users are scheduled as follows:

```
Data: set A of coordinated cells with CoMP UEs to be served
Result: Set U_CoMP of UEs to schedule
for each cluster A do
    n ← |A| − 1
    while (A ≠ ∅) & (n > 0) do
        n ← n − 1
        max ← 0
        for each CoMP user u in cluster A do
            S ← {s} ∪ set of cooperating cells
            if (S ⊆ A)&(s is not muted) then
                Compute metric
                if max < metric then
                    max ← metric
                    selected user ← u
                end
            end
        end
        s ← serving cell of selected user
        A ← A \ {s}
        mute all cooperating cells
        U ← U ∪ {selected user}
    end
end
```

Algorithm 9: Centralized static allocation scheduling for DPB.

Note that at most $|A|-1$ CoMP users can be scheduled on the same resources in a given cluster of size $|A|$. This may happen when all CoMP scheduled users have the same cooperating cell.

III.5.3.2 Iterative scheduling

Consider in each cluster a central scheduler that has knowledge of the CSI/CQI information of all users being served within the cluster. In this case, the simplest scheduling strategy, treating CoMP users and non-CoMP users equally, is achieved through the iterative scheduler. This scheduler requires at most $|A|$ iterations in order to make scheduling decisions for all cells taking part in the cluster, as at most $|A|$ users can be scheduled in a given cluster of size $|A|$. This may happen when all scheduled users are non-CoMP users.

The first user is chosen either based on the scheduling metric (PF, max C/I...) or randomly (RR). This blocks only the serving cell of the scheduled user so that this cell can no more transmit to another user or cooperate with other cells by going blank, on the same RBs. Note that in the case where the first scheduled user is a CoMP user, all cooperating cells stay available for all their CoMP users attached to neighboring cells, in the sense that these cells can still cooperate (by going blank)

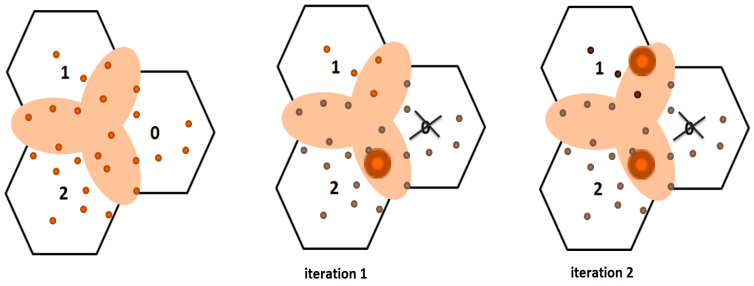


Figure III.61: Iterative scheduling for DPB.

as they are already muted. The next best user according to the same metric is chosen only if its serving cell and all its cooperating cells are still available after the first decision and so on till all cells in the cluster become unavailable... The serving cell is considered as available when it is neither transmitting nor muted. Cooperating cells are considered as available when they are not transmitting independently of the fact that they are muted or not. The algorithm works as follows:

Data: set A of coordinated cells with UEs to be served
Result: Set U of UEs to schedule
for each cluster A **do**
  $n \leftarrow |A|$
  while $(A \neq \emptyset)\,\&\,(n > 0)$ **do**
    $n \leftarrow n - 1$
    max $\leftarrow 0$
    for each user u in cluster A **do**
      $S \leftarrow \{s\} \cup$ set of cooperating cells
      if $(S \subseteq A)$&(s is not muted) **then**
        Compute metric
        if max $<$ metric **then**
          max $\leftarrow$ metric
          selected user $\leftarrow$ u
        end
      end
    end
    $s \leftarrow$ serving cell of selected user
    $A \leftarrow A \setminus \{s\}$
    mute all cooperating cells
    $U \leftarrow U \cup \{$selected user$\}$
  end
end

Algorithm 10: Centralized iterative scheduling for DPB.

Figure III.61 shows an example of a cluster of three cells, where the scheduling decisions in a given TTI are taken in just two iterations. In the first iteration a CoMP user associated with cell 2 is selected, this involves cell 2 in the transmission and forces cell 0 to go blank. Observe that all users requiring transmission or cooperation from cell 2 are then blocked. On the other hand, all users associated with cell 0 and requiring transmission from this cell (non-CoMP users as well as CoMP users) are blocked. As cell 0 can no more transmit, the last user to schedule in this TTI can be chosen within the subset of users associated with cell 1 (requiring transmission from cell 1), and not

requiring cooperation of cell 2.

III.5.3.3 Priority to non-CoMP users

Under this policy, non-CoMP users in each cell are scheduled first and are allocated all the radio resources whenever active; CoMP users wait untill serving cell becomes available to transmit and cooperating neighboring cells become capable to go blank. Scheduling is done in two steps, each TTI and for every subset of resource blocks. Non-CoMP users are selected first according to classical scheduling policies (RR, PF, max C/I); the scheduling decision in each cell is independent of the decisions in the neighboring cells in this case. In some cases, no users are scheduled within a cell due mainly to the fact that there is no non-CoMP users in this cell. The second step consists in assigning remaining resources to appropriate CoMP users. Users are chosen based on a particular scheduling metric (PF, max C/I...) or even randomly (for RR). The algorithm describing the second step works as follows:

Data: set A of coordinated cells with CoMP UEs to be served
Result: Set U_{CoMP} of CoMP UEs to schedule
for each cluster A **do**
  $n \leftarrow |A| - 1$
  while $(A \neq \emptyset)\,\&\,(n > 0)$ **do**
    $n \leftarrow n - 1$
    max $\leftarrow 0$
    for each CoMP user u in cluster A **do**
      $S \leftarrow \{s\} \cup$ set of cooperating cells
      if $(S \subseteq A)$&(s is not muted) **then**
        Compute metric
        if max < metric **then**
          max $\leftarrow$ metric
          selected user $\leftarrow$ u
        end
      end
    end
    $s \leftarrow$ serving cell of selected user
    $A \leftarrow A \setminus \{s\}$
    mute all cooperating cells
    $U \leftarrow U \cup \{$selected user$\}$
  end
end

Algorithm 11: Centralized Pri-NC scheduling for DPB.

Observe that the maximum number of iterations during the second scheduling phase is $|A| - 1$ due to the fact that at most $|A| - 1$ CoMP users can be scheduled at the same time in a given cluster. This may happen when $|A| - 1$ cells are transmitting and one cell is cooperating with the $|A| - 1$ cells by going blank.

III.5.3.4 Priority to CoMP users

The scheduling consists of two phases. In the first one, CoMP users are scheduled first. All serving cells and cooperating cells of the scheduled CoMP users are then blocked. In the second phase, Non-CoMP users are served only when there are no active CoMP users or when there are still available

cells neither transmitting nor muted after scheduling active CoMP users. The algorithm works as follows:

III.5.4 Flow level model

III.5.4.1 Single coordination zone and intra-site coordination scenarios

All results obtained for the scenario of a single coordination zone in Section III.3.4.1 are valid for the case of DPB. However, the results obtained for the scenario of intra-site coordination, in Section III.3.4.2, no more apply in the case of DPB. This is due to the fact that in the case of intra-site coordination, there are multiple coordination zones in the cluster of size K ($K = 3$ in this case). Unlike JT where at most $K/2$ ($= 1$) CoMP users can be scheduled together, in the case of DPB the number of simultaneously scheduled CoMP users can go till $K-1$ ($= 2$) as illustrated by Figure III.62; each scheduled CoMP user must belong to a different coordination zone. Thus the fractions of allocated resources, the stability conditions as well as the throughput performance depend on the coordination technique (JT or DPB).

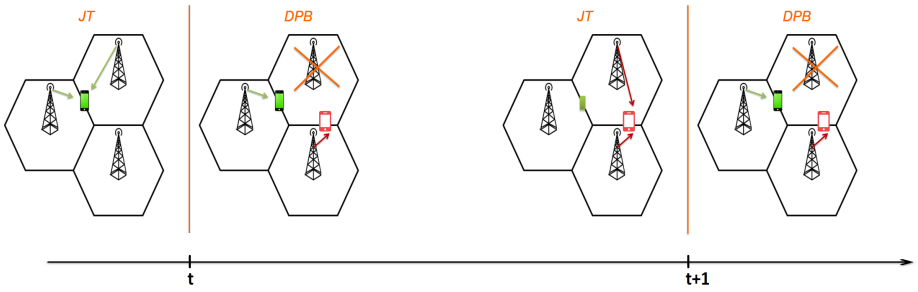


Figure III.62: JT vs DPB scheduling.

III.5.4.2 A single cell surrounded by its neighboring cells

a) Stability condition analysis

Note that when applying DPB, the mean service rates of cell center users and cell edge users are higher than the case without DPB. This is due to the fact that users benefit implicitly from the muting of some interfering cells which increases the global mean service rate in the network.

Supposing that DPB feature is activated in the network, we denote by $\mu'_{\text{center}} \geq \mu_{\text{center}}$ the mean service rate of a user in the cell center. $\mu'_{\text{edge,n}} \geq \mu_{\text{edge,n}}$ is the mean service rate of a user at the cell edge when served without any coordination, and that may involve n cells in the coordination when getting the CoMP status. We denote by β'_n the mean rate gain of a user located at the cell edge involving n cells in the coordination, one transmitting cell and $n-1$ muted cells. Consequently, the mean service rate of a CoMP user involving n cells in the coordination when DPB is activated is nothing more than $\beta'_n \mu'_{\text{edge,n}}$.

When activating DPB, the new cell load under RR iterative scheduler becomes as follows:

$$\rho' = \frac{p_{\text{center}}\lambda}{\mu'_{\text{center}}} + \sum_n \frac{\alpha_n}{\beta'_n} \frac{p_{\text{edge,n}}\lambda}{\mu'_{\text{edge,n}}} \tag{III.35}$$

α_n is the traffic increase factor in the particular cell edge zone involving n cells in the coordination. In fact, the condition that should be met regarding the coordination gain is less restrictive than in the

```
Data: set A of coordinated cells with UEs to be served
Result: Set U of UEs to schedule
for each cluster A do
    n ← |A|
    switch To NC ← 0
    while (A ≠ ∅) & (n > 0) do
        n ← n − 1
        max ← 0
        if !switch To NC then
            for each CoMP user u in cluster A do
                S ← {s} ∪ set of cooperating cells
                if (S ⊆ A)&(s is not muted) then
                    Compute metric
                    if max < metric then
                        max ← metric
                        selected user ← u
                    end
                end
            end
            if There is no selected user then
                switch To NC ← 1
            end
        end
        if switch To NC then
            for each non-CoMP user u in cluster A do
                S ← {s}
                if S ⊆ A then
                    Compute metric
                    if max < metric then
                        max ← metric
                        selected user ← u
                    end
                end
            end
        end
        s ← serving cell of selected user
        A ← A \ {s}
        mute all cooperating cells
        U ← U ∪ {selected user}
    end
end
```

Algorithm 12: Centralized Pri-C scheduling for DPB.

case of JT. This is not related only to the fact that $\mu'_{\text{center}} \geq \mu_{\text{center}}$ and $\mu'_{\text{center}} \geq \mu_{\text{center}}$ but also to the traffic increase factor α_n which is lower in this case. For the sake of simplicity, consider for instance the case of $K_{max} = 2$. Figure III.63 illustrates the CoMP users' traffic (orange zone) in the case of a hexagonal topology. In the worst case, only one CoMP user located in the orange zone is served and allocated given resources. In this case the traffic at the edge increases by a factor of 2. However, it is also very likely that 6 CoMP users each one associated with one of the 6 surrounding neighboring cells and requiring that the considered cell goes blank are scheduled together at the same time and allocated the same resources. Unlike JT, the cell can by going blank treat the additional traffic coming from neighboring cells in parallel. So, it is as if the traffic has increased by a factor of $7/6$. Thus the mean traffic increase factor in a symmetric hexagonal network is $7/6 < \alpha_2 < 2$. We can proceed in the same manner to show that $\alpha_n \leq n \ \forall n$.

Thus, one way to avoid the stability condition degradation when activating DPB is to guarantee that:

$$\beta'_n \geq n' \qquad \forall n,$$

where $n' \leq n$.

Note that $\beta'_n < \beta_n$. However according to Section III.5.2 most of the SINR gain comes from the elimination of the interference. Consequently, most of the rate gain is achieved through the elimination of the interference.

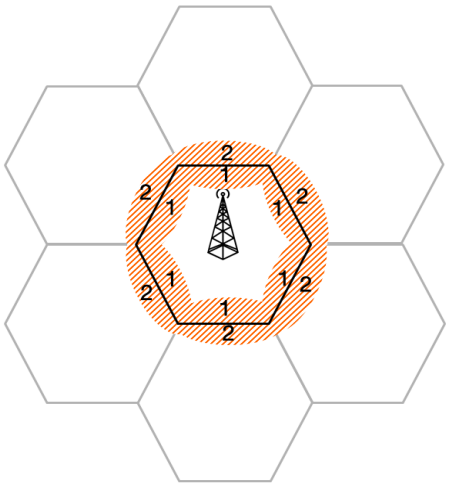


Figure III.63: Traffic increase at the edge.

b) Impact of interference on the coordination gain

Consider the case of only one cooperating cell, that is $\mathbb{C} = \{c\}$. In order to guarantee that there is no stability condition degradation, we consider the most restrictive condition in the case of DPB. Thus, the mean coordination gain should be higher than two.

We proceed in a similar way as in §III.3.4.3 b). The signal-to-noise ratio in the absence of coordi-

nation is.

$$SINR = \frac{P_s}{P_c + I + \mathcal{N}}$$

while that when c is cooperating (going blank) is:

$$SINR'' = \frac{P_s}{I + \mathcal{N}}$$

The SINR gain is then:

$$G = \frac{SINR''}{SINR}$$

The new service rate should be two times more the original service rate that is:

$$\frac{\log_2(1 + SINR'')}{\log_2(1 + SINR)} \geq 2$$

$$\frac{\log_2(1 + G \times SINR)}{\log_2(1 + SINR)} \geq 2$$

$\hookrightarrow$

$$G \geq SINR + 2$$

$$\frac{SINR''}{SINR} \geq SINR + 2$$

We get the following in-equation:

$$\frac{1}{\upsilon} \times (\zeta + \upsilon) \geq \frac{1}{\zeta + \upsilon} + 2,$$

which is nothing more than:

$$\upsilon^2 + \upsilon - \zeta^2 \leq 0$$

This in-equation is verified when:

$$\upsilon \leq \sqrt{\zeta^2 + 1/4} - 1/2. \tag{III.36}$$

Observe that when $P_s = P_c$, the residual interference $I + \mathcal{N}$ should be less than approximately $0.6P_s$ in order to satisfy the required condition.

III.5.5 Numerical results

III.5.5.1 System level simulations

We simulate 12×10^6 TTI, using the LTE simulator described in Appendix A, and we estimate the mean user throughput as the ratio of the mean flow size to mean flow duration for $\sigma = 1.25$ MBytes, under different clustering strategies: static clustering and dynamic clustering. We suppose that all users are static. We study the case without coordination under classical PF scheduler as well as the case with DPB for different power threshold values δP: 3 dB, 6 dB, 9 dB, 12 dB, 15 dB and 18 dB under PF iterative scheduler explained in Algorithm 10. We evaluate also in each case the mean number of users per eNodeB and the relative throughput gain when applying DPB compared to the case without coordination with respect to the offered traffic per eNodeB. For each clustering method, we consider two cases. In the first one, we limit the number of cooperating cells for one user to $K_{max} = 2$ while in the second one at most $K_{max} = 3$ cells can cooperate to serve one user.

a) Static clustering

Figure III.64 shows the case of intra-site coordination. Observe that the relative gain increases with load. At low load, the gain brought by the elimination of the interference due to the muting of interfering cells is relatively small. This is due to the fact that when neighboring cells are slightly loaded, they are naturally muted most of the time and thus cooperation is likely to be less triggered. When the network becomes more loaded, interference between different cells becomes more important. In this case, the blanking of interfering cells at the right moment, when the scheduled user is strongly affected by the generated interference, seems to bring interesting gains. Observe that the higher the power threshold δP, the higher the gains. Note that the relative gain can go up to 50% when only the most interfering cell in the site can be muted for a CoMP user ($K_{max} = 2$), while it goes up to 70% when $K_{max} = 3$. Unlike JT, applying DPB does not degrade performance at high load by reducing the stability condition. Quite the contrary, the mean number of users in each cell increases much more slowly compared to the case where DPB is not applied (without coordination). This technique seems to be promising as it is so simple to implement in the case of intra-site coordination where coordination and scheduling decisions can be taken centrally in each site.

Figure III.65 shows the case of inter-site coordination. The case of intra-site and inter-site coordination is illustrated by Figure III.66. The main conclusions are the same of the case of intra-site coordination except that relative gains seem to be higher than the previous case. Observe that maximum relative gains are around 80% when $K_{max} = 2$ and 100% when $K_{max} = 3$. However, from the point of view of complexity, intra-site coordination is much easier to implement.

b) Dynamic clustering

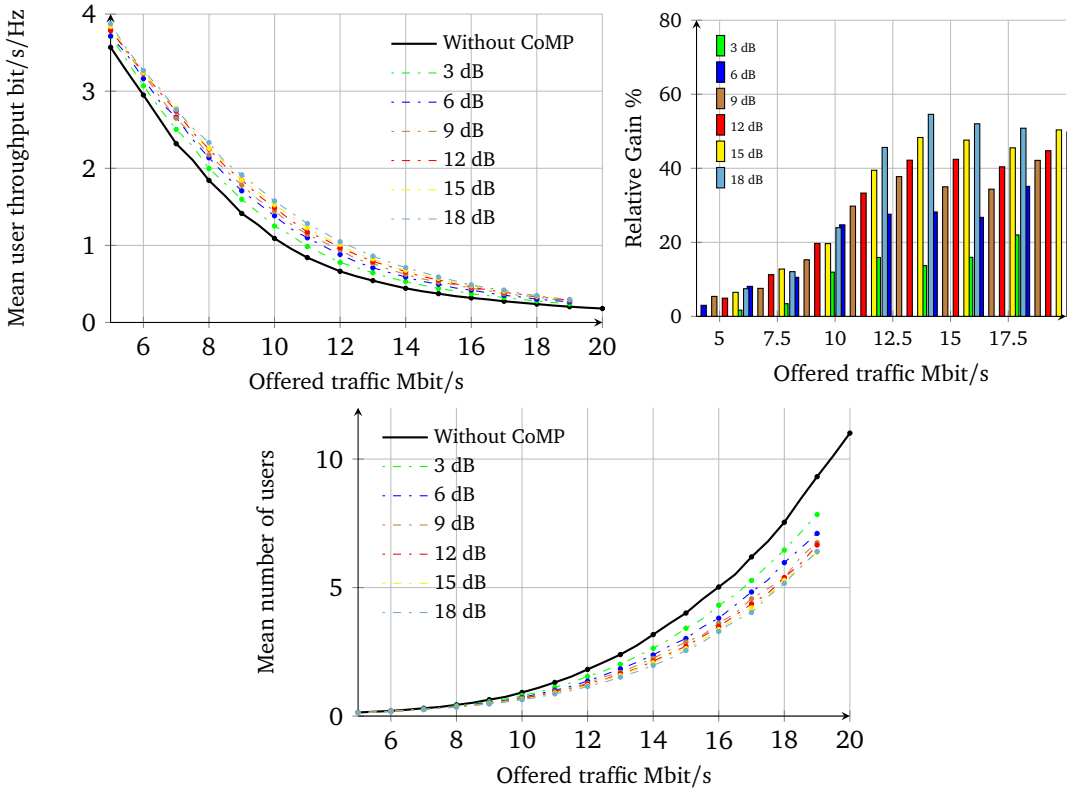
The case of dynamic clustering is illustrated by Figure III.67. The dynamic clustering method provides the highest gains but this is at the cost of higher complexity. Observe that gains go up to 150% when $K_{max} = 2$ and to 200% when $K_{max} = 3$. In the case of $K_{max} = 3$, gains increase with δP up to $\delta P = 12dB$ and then start to decrease.

Observe that under dynamic clustering the cluster size (21 cells) increases by a factor of 7 with respect to the previous cases where clusters are static of size $N = 3$. Even though, the maximum gains go from around 70%-100% (static clustering) till only 200% (dynamic clustering). The main reason is that in the case of static clustering some CoMP as well as non-CoMP users benefit implicitly from the muting of some interfering cells which are not necessarily defined as cooperating cells for the benefiting users. Thanks to the scheduler granularity that allows to take scheduling decisions each TTI, the measurement done by the UE and the CQI feedback, users are more likely to be scheduled at the best subframe where interfering cells belonging to other clusters (different from the cluster of the serving cell) go blank.

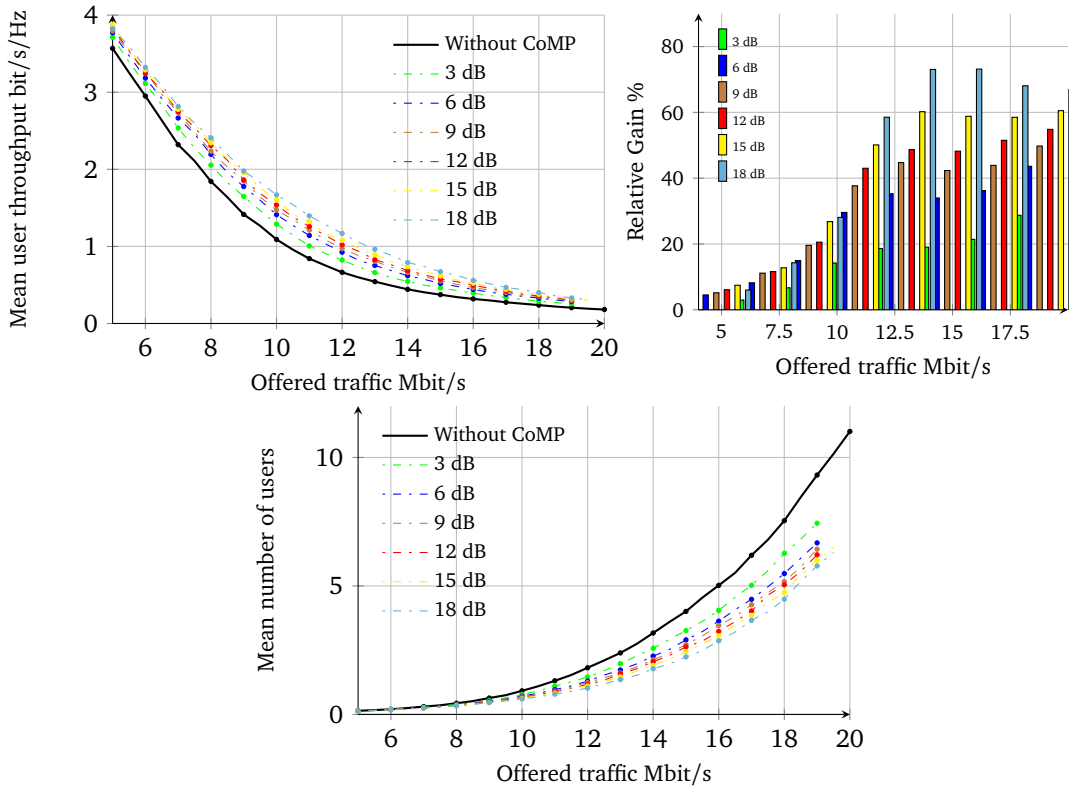
Note that the actual gains may be less then these gains. Indeed, users are scheduled based on measurements done in previous TTI. So, if the level of interference experienced by a scheduled user in the current TTI is higher then that measured in previous TTI, an error is more likely to happen in the decoding at the receiver side. This is due to the fact that the MCS (Modulation and Coding Scheme) assigned by the BS to the scheduled user, based on previous measurements, may be higher than the MCS that can be decoded by this user in the current TTI. To avoid this error, the scheduler should be aware of the channel state information of each user. An optimal scheduler allows to achieve the maximum gains. However, even suboptimal scheduling strategies allow to achieve high gains.

III.5.6 Global scheduler

The iterative scheduler corresponds to the simplest scheduling strategy that treats CoMP users and non-CoMP users equally without frequency band partitioning. It is a sub-optimal scheduling strategy.

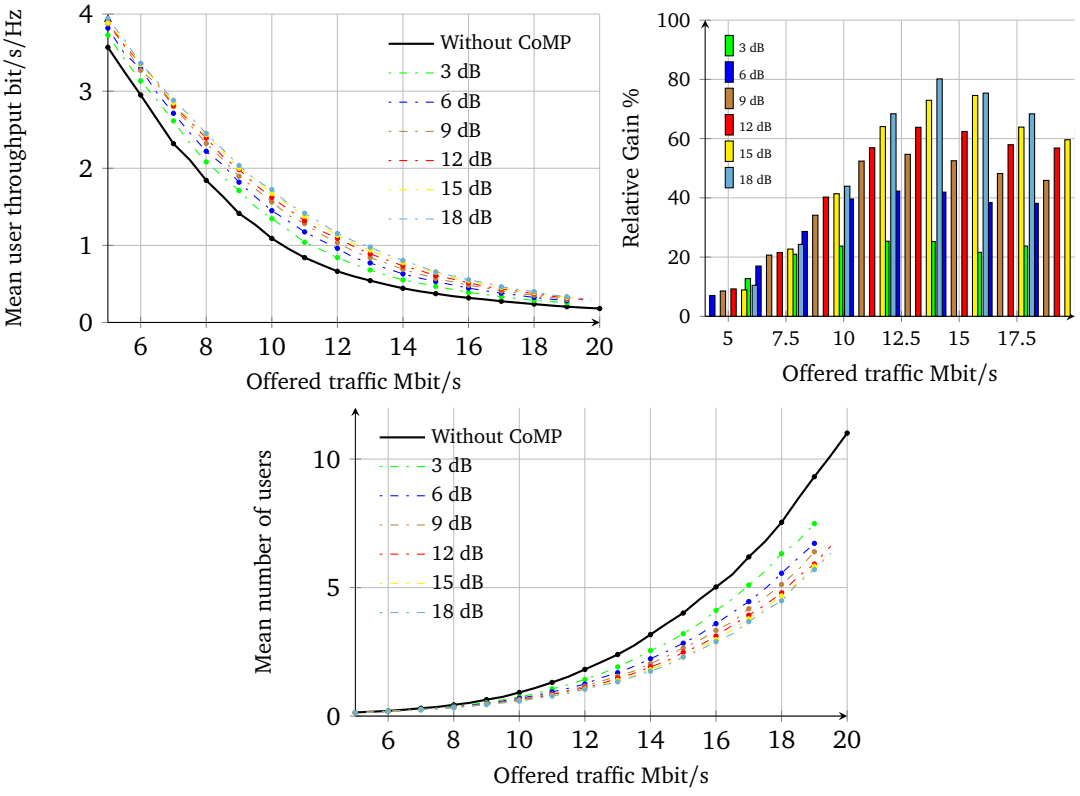


(a) $K_{max} = 2$ maximum coordinated cells

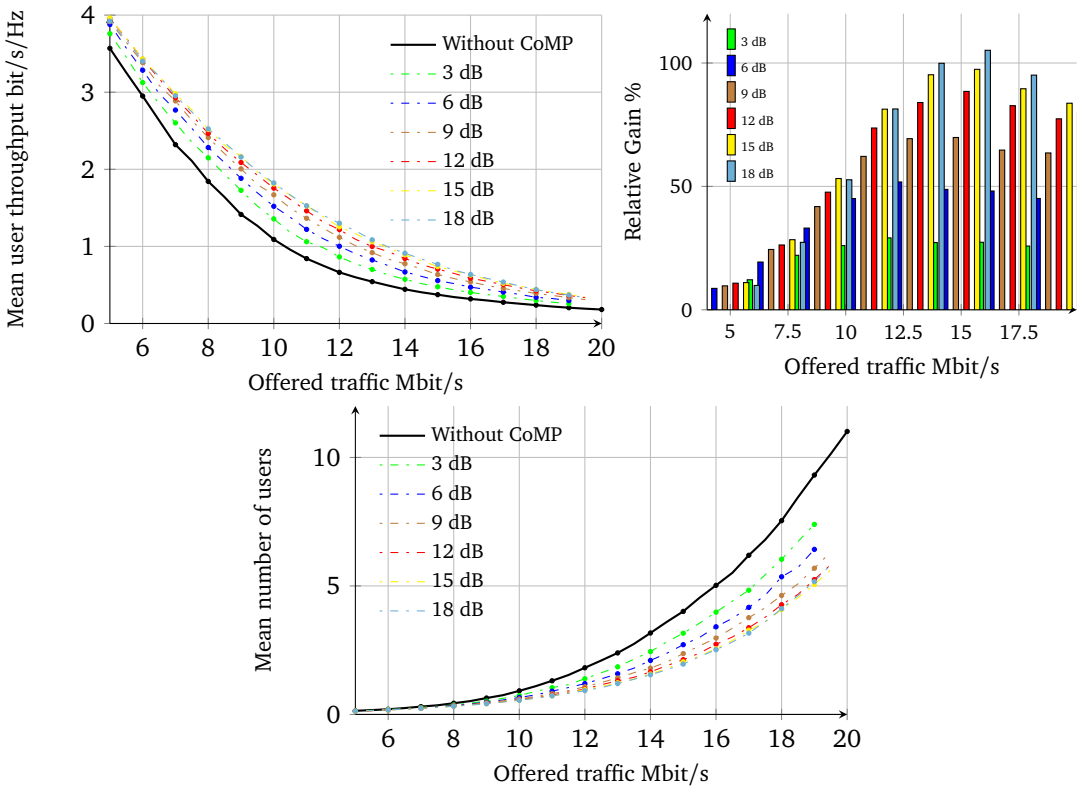


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.64: Throughput performance obtained by system-level simulation for Intra-site DPB.

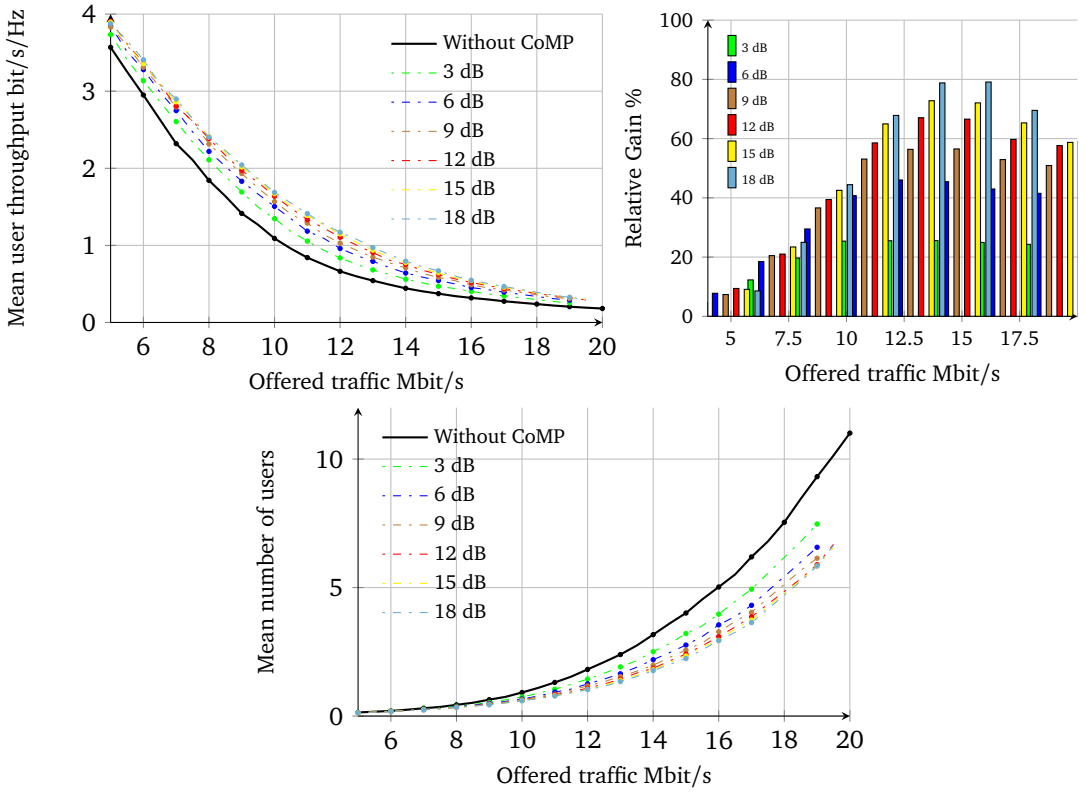


(a) $K_{max} = 2$ maximum coordinated cells

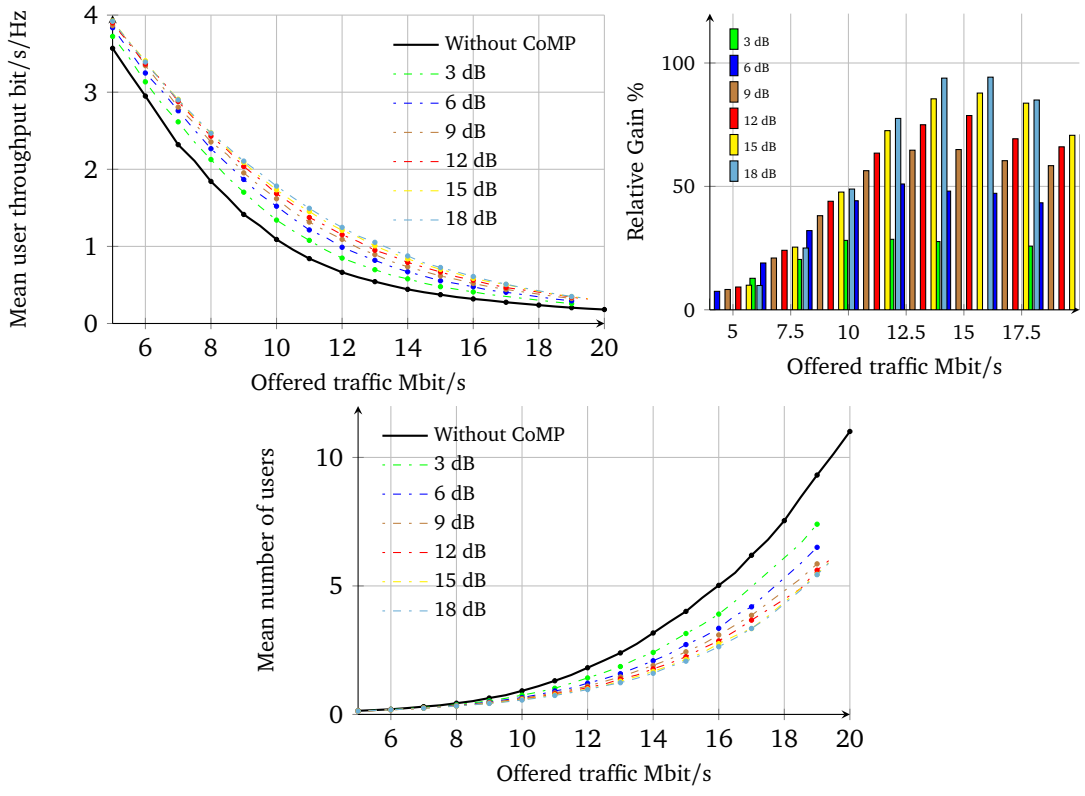


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.65: Throughput performance obtained by system-level simulation for Inter-site DPB.

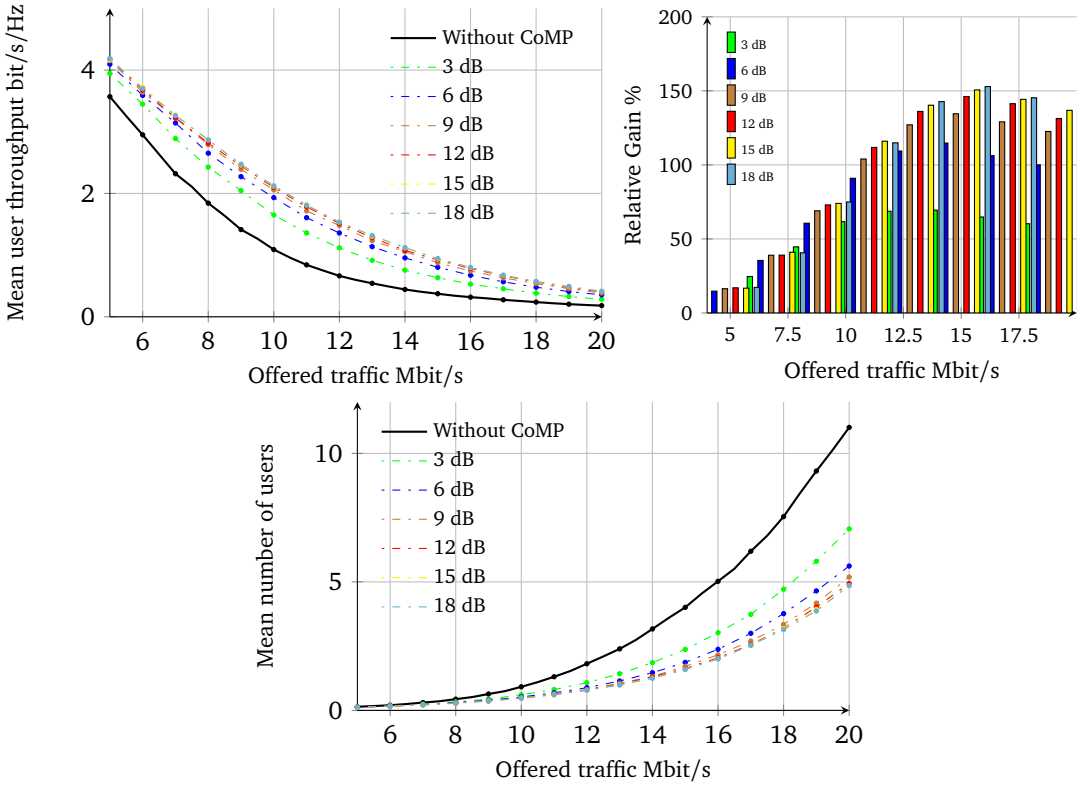


(a) $K_{max} = 2$ maximum coordinated cells

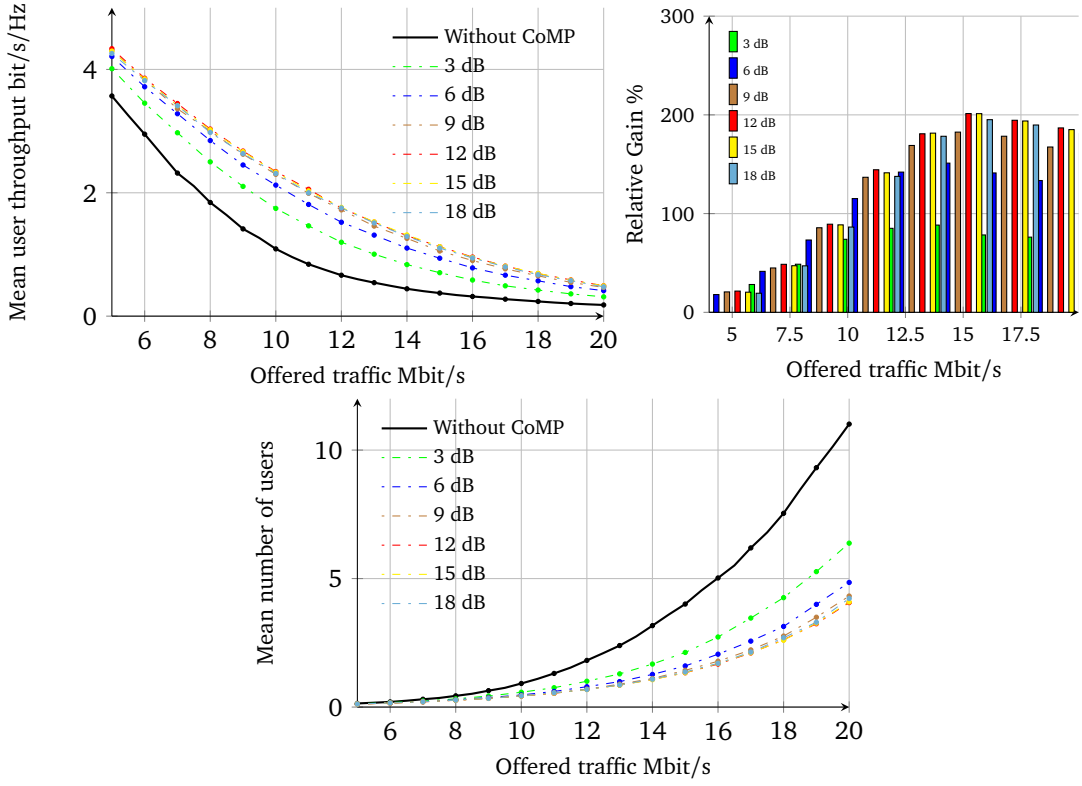


(b) $K_{max} = 3$ maximum coordinated cells

Figure III.66: Throughput performance obtained by system-level simulation for Intra and Inter-site DPB.



(a) $K_{max} = 2$ maximum coordinated cells



(b) $K_{max} = 3$ maximum coordinated cells

Figure III.67: Throughput performance obtained by system-level simulation for Dynamic clustering DPB.

However, a global scheduler may provide better performance. We assume in this section that the status of a CoMP user is dynamic so that the coordination may be activated or deactivated according to the scheduling decision. The following optimization problem is then formulated for each cluster of cells and for each sub-band of frequency f (subset of RBs):

$\forall f \in [1, 2, \dots N_{RB}]$:

$$\max \sum_{j=1}^{N} \sum_{u=1}^{n_j} \frac{1}{\bar{R}_u} \boldsymbol{F} \left(\frac{b_{u,s}^f P_{u,s}^f}{\sum_{i \neq c \in \mathbb{S}_s} P_{u,i}^f + \sum_{c \in \mathbb{S}_s} m_c^f P_{u,c}^f + \mathcal{N}} \right)$$

We denote by $\mathbb{S}_x$ the set of cells which take part in the cooperative cluster of cell x. This set does not include cell x.
$b_{u,x}^f \in \{0, 1\}$ is a variable to indicate whether cell x is transmitting to user u on frequency sub-band f in the current TTI or not.

$$b_{u,x}^f = \begin{cases} 0 & \text{cell } x \text{ is not transmitting to user } u \text{ .} \\ \\ 1 & \text{cell } x \text{ is transmitting to user } u \text{ .} \end{cases}$$

$m_x^f \in \{0, 1\}$ is a variable to indicate whether cell x is muted on frequency sub-band f or not.

$$m_x^f = \begin{cases} 0 & \text{cell } x \text{ is muted .} \\ \\ 1 & \text{cell } x \text{ is transmitting .} \end{cases}$$

Note that the following constraint should be fulfilled:

$\forall s \in [1, 2, \dots N]$, $\forall f \in [1, 2, \dots N_{RB}]$:

$$\sum_{u=1}^{n_s} b_{u,s}^f + 1 - m_s^f \leq 1 \tag{III.37}$$

Cell s should transmit to only one user on a given sub-band f at a time.

This scheduler may provide better performance than the iterative scheduler but this is at the cost of higher complexity. Indeed, this would require to exchange an enormous amount of signaling information. Observe that the power $P_{u,c}^f$ received from each cell $c \in \mathbb{S}_s$ should be reported by the UE, s being the serving cell of the UE. However, the residual interference, that is the interference caused by all cells which take part in the cooperative cluster of cell s can be measured by the UE. The higher the size of the cluster, the better is the performance and the higher is the complexity.

III.6 Conclusion

We have investigated in this chapter JP techniques, more precisely JT and DPB. We have evaluated the performance of both techniques through flow level models and system level simulations considering non-full buffer traffic models. A complete comparative study between these two techniques has been carried out through this chapter.

JT is a coordination technique which is helpful to improve cell edge performance by converting an interfering signal to a desired signal. However, the introduction of such a feature may create some problems due to the excessive resource consumption, especially in a highly loaded network. The performance when applying such a technique depends primarily on the coordination gain, that is the mean throughput gain brought by the cooperation of a cell. This coordination gain depends to a great extent on the scenario (high-low interference), the transmission scheme (one-multi stream) as well as the scheduling strategy. We have analyzed the capacity issues of this technique and we have proposed a new cooperating cell definition method where a cell cooperates only when it brings at least 100% mean throughput gain when cooperating. We have studied different scheduling strategies and we have shown that a global scheduler where the coordination can be activated and deactivated dynamically provides better performance compared to an iterative scheduler, but this is at the cost of higher complexity.

We have shown that JT could be interesting in an interference-dominated environment, particularly at medium load, when different streams are sent from different cooperating cells but this requires multiplexing and advanced receivers techniques. We have studied performance in multiple cases considering different clustering approaches and different clustering strategies. Gains depend mainly on the clustering strategy, the transmission scheme as well as the load of the network.

We have also considered the case of a low interference scenario where it is not recommended to activate JT feature. We have shown that, if we stick to performing coordination for static cell-edge users suffering from very degraded throughput, it is not worth to perform coordination for a mobile user who is able to move and to get better radio conditions. Consequently, we have proposed a mobility-aware scheduler which deprioritizes mobile CoMP users. We have shown that this scheduler improves performance by giving the chance to mobile cell-edge users to be served in better radio conditions where cell coordination is not required. This scheduler is suitable for elastic traffic where the delay is tolerable. However, it can no longer work when considering traffic with real-time constraints such as video streaming (Netflix, YouTube, ...).

The performance of DPB, which consists in prohibiting transmission from strongest interfering cell(s) on a set of physical resources, is also evaluated in several cases under different clustering methods assuming DPB scheduling granularity of one sub-frame. We have shown that most of the rate gain is achieved through the elimination of the interference and that DPB may present less restrictive stability condition compared to a JT scheme with a moderate coordination gain.

DPB seems to be more promising than JT especially in the case of intra-site coordination where coordination and scheduling decisions can be taken centrally in each site; it is a simpler and performant technique especially at higher loads, in a high interference scenario where only a limited number of antennas is deployed. From the technical point of view, intra-site coordination will be much easier to realize. However, in order to exhaust the full interference reduction potential of cell cooperation, dynamic cooperation is needed.

Synchronization presents a real challenge for both JP techniques, to this must of course be added the CSI feedback design as well as the backhaul characteristics in terms of latency and capacity especially when considering a distributed coordination approach.

Chapter IV

Coordinated Beamforming CoMP

IV.1 Introduction

One of the key requirements of LTE is to achieve enhanced performance in terms of cell throughput. Multiple Input Multiple Output (MIMO) is one of the most promising key technologies used in modern communication networks in order to achieve high data rates. Beamforming [65, 63], one of the downlink transmission modes defined in LTE, consists in using an array of antenna elements at the transmitter side, for the purpose of illuminating specific targeted areas. The radiation pattern is determined by the beamforming weights / phase shifts applied to different antenna elements as well as the geometric characteristics of the antenna array [70].

While two-dimensional (2D) MIMO beamforming techniques allow to adjust the beams in the horizontal dimension, three-dimensional (3D) MIMO beamforming techniques allow vertical beamforming combined with horizontal beamforming, see [48, 74]. Enhancements of LTE enable to support full dimension MIMO/Elevation Beamforming by the use of higher dimension MIMO of up to 64 antennas at the eNodeB. However, the array geometric characteristics are not specified; the number of antenna elements and the antenna architecture are left up to implementation. LTE does not even specify any methods for designing codebooks, as it has been done in [76].

Beamforming is one of the techniques that can be used to mitigate inter-cell interference (ICI), one of the main issues for current and next-generation wireless cellular networks. The performance in a system with beamforming depends on the scheduling strategy. While standard single cell scheduling is much more simple, coordinated scheduling/beamforming is needed, in order to further improve performance by avoiding interference between adjacent cells. The aim is to schedule, at a time, the set of users covered by beams which do not interfere between each other, see [75, 51, 31]. Finding the globally optimal downlink beamforming vector across all BSs in order to maximize a utility of interest, is very challenging due to the processing burden and the amount of feedback that is required in this case. This requires full channel state information (CSI) knowledge, which may not be feasible in practice. Some optimization problems are very complex as shown in [54] for a MISO channel. Thus, since low-complexity scheduling is needed, suboptimal scheduling algorithms, like iterative ones, are generally used.

Interference is significantly mitigated in a beamforming system, even in the absence of coordination. The tighter the beams, the less the probability that activated beams in neighboring cells interfere between each other. An interference free environment is guaranteed under coordinated schedulers. However, these schedulers, as shown in the literature, add significant processing com-

plexity, feedback and signaling overhead and their performance depends primarily on the accuracy of the feedback. But, is it really worth it ? In order to answer this question, we quantify, in this chapter, the additional gains brought by using coordinated schedulers with respect to a standard distributed scheduling case. We first study a system with blind beamforming in Section IV.3 , that is a system where beams activation is done in a distributed and non-coordinated manner in different cells. Then, we study in Section IV.4.1 some coordinated scheduling algorithms in a beamforming system. We show that channel-aware standard non-coordinated schedulers may provide similar performance as coordinated scheduler in the case of tight beams. Coordination provides only limited performance gains, compared to the case without coordination, while adding a lot of complexity to the network.

IV.2 Beamforming basics

Beamforming consists in using multiple antennas to control the direction of a wavefront by appropriately weighting the magnitude and phase of individual antenna signals (transmit beamforming), as illustrated in Figure IV.1. For example this makes it possible to provide better coverage to specific areas along the edges of cells. Because every single antenna in the array makes a contribution to the steered signal, an array gain (also called beamforming gain) is achieved. While two-dimensional antenna array structures allow to perform 3D beamforming, the focus of this work is only on one-dimensional linear array.

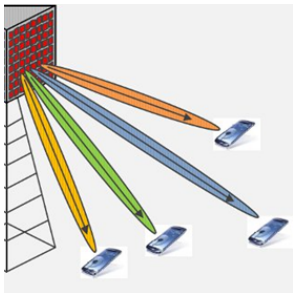


Figure IV.1: Beamforming.

IV.2.1 Array of two elements- Line of sight

First, consider the case of two similar isotropic antennas separated by a distance d, as shown in Figure IV.2. Observe that the wavefront of a signal must traverse the additional distance $d \sin\theta$ in order to reach the user (at point P), θ is the angle between the user and the main direction of the antenna.

The electric field radiated by the first antenna at point P is given by:

$$\vec{E}_1(P) = \alpha \frac{e^{-j\frac{2\pi f_c}{c}r}}{r}\vec{e}_P = \alpha \frac{e^{-j\frac{2\pi}{\lambda}r}}{r}\vec{e}_P,$$

where α is a constant, r is the distance from the considered antenna element to point P and $\vec{e}_P$ is the vector which carries the polarization at point P. f_c is the carrier frequency, c is the light speed and λ is the wavelength.

Similarly, the electric field radiated by the second antenna at point P is given by:

$$\vec{E}_2(P) = \alpha \frac{e^{-j\frac{2\pi}{\lambda}(r-d\sin\theta)}}{r - d\sin\theta}\vec{e}_P.$$

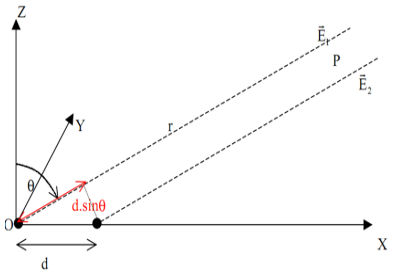


Figure IV.2: Two transmit antennas.

According to far-field approximations, $d\sin\theta$ can be neglected in the amplitude. However, it should be considered in the phase.

Then, the total electric field radiated at point P is nothing more than the sum:

$$\vec{E}(P) = \alpha\frac{e^{-j\frac{2\pi}{\lambda}r}}{r}\left(1+e^{j\frac{2\pi}{\lambda}d\sin\theta}\right) = 2\alpha\frac{e^{-j\frac{2\pi}{\lambda}r}}{r}e^{-j\frac{\pi}{\lambda}d\sin\theta}\cos\left(\frac{\pi}{\lambda}d\sin\theta\right),$$

$1+e^{j\frac{2\pi}{\lambda}d\sin\theta}$ is the array factor, which represents the modification induced by the second antenna element.

The corresponding radiation pattern (related to the electric field) is then given by:

$$E(\theta) = \left|\cos\frac{\pi d\sin\theta}{\lambda}\right|. \tag{IV.1}$$

Figure IV.3 illustrates the radiation pattern derived from (IV.1), for different values of d. The optimal value of d that allows to obtain maximum directivity is around 0.7λ. Generally, d is chosen between 0.5λ and λ. In this work, we consider the case where $d = 0.5\lambda$.

In the case of a non-isotropic antenna, the radiation pattern becomes:

$$E(\theta) = E_0(\theta)\left|\cos\frac{\pi d\sin\theta}{\lambda}\right|,$$

where $E_0(\theta)$ is the radiation pattern of one antenna element.

Suppose that a phase shift δ is applied to the second antenna element. In this case, the radiation pattern can be written as follows:

$$E(\theta) = E_0(\theta)\left|\cos\frac{\pi(d\sin\theta+\delta)}{\lambda}\right|.$$

Observe that in order to obtain maximum power in the direction θ, the phase shift should be set to:

$$\delta = -d\sin\theta.$$

IV.2.2 Array of N elements- Line of sight

Consider now the case of a uniform linear array as shown in Figure IV.4.

The array factor when N is even, can be written in this case, as follows:

$$e^{-j\frac{N-1}{2}\frac{2\pi}{\lambda}d\sin\theta} + e^{-j\frac{N-3}{2}\frac{2\pi}{\lambda}d\sin\theta} + \ldots + e^{j\frac{N-3}{2}\frac{2\pi}{\lambda}d\sin\theta} + e^{j\frac{N-1}{2}\frac{2\pi}{\lambda}d\sin\theta}.$$

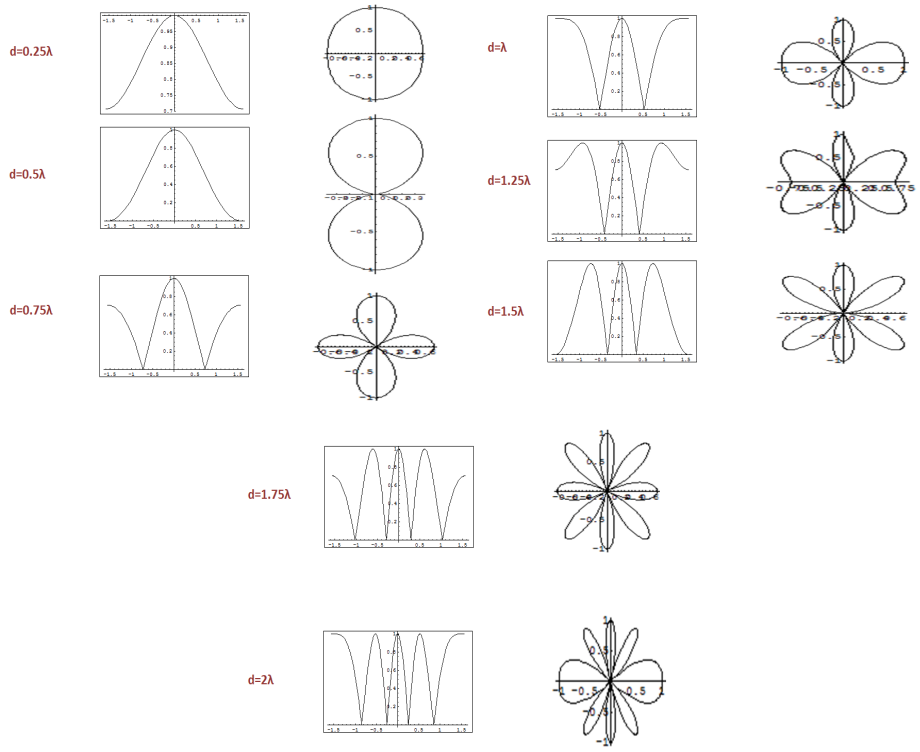


Figure IV.3: Radiation pattern for two antenna elements with different separating distance values.

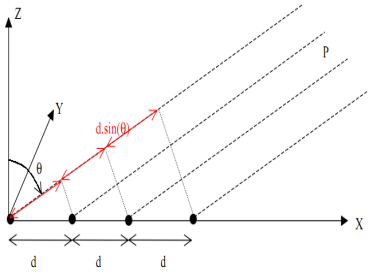


Figure IV.4: Uniform linear antenna array.

The phase shift between two adjacent antenna elements is $d \sin\theta$. We proceed in the same way in order to get the radiation pattern, that is for isotropic antenna elements given by:

$$E(\theta) = \left| \frac{\sin\left(\frac{N\pi d \sin\theta}{\lambda}\right)}{\sin\left(\frac{\pi d \sin\theta}{\lambda}\right)} \right|. \tag{IV.2}$$

Figure IV.5 gives the impact of the number of antenna elements on the radiation pattern given by equation (IV.2). Observe that the higher the antenna elements number N, the higher the directivity will be.

In the following, we suppose that $d = \lambda/2$ and that the signal is distributed over the N antennas.

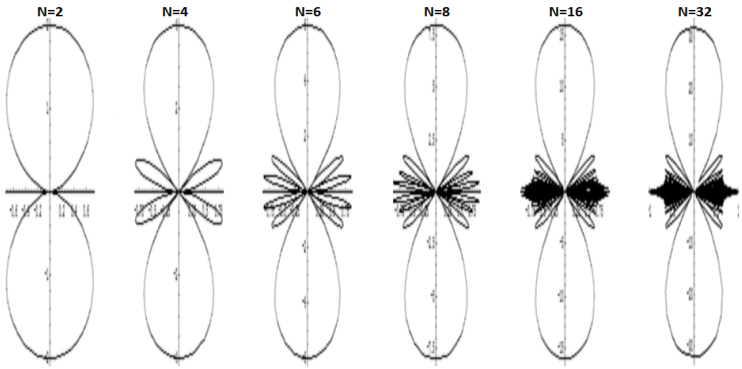


Figure IV.5: Impact of the number of antenna elements N, $d = \lambda/2$.

In order to steer the beam toward direction θ', the following precoding vector should be applied:

$$v = 1/\sqrt{N} \begin{pmatrix} e^{j\frac{N-1}{2}\pi\sin\theta'} \\ e^{j\frac{N-3}{2}\pi\sin\theta'} \\ \vdots \\ e^{-j\frac{N-3}{2}\pi\sin\theta'} \\ e^{-j\frac{N-1}{2}\pi\sin\theta'} \end{pmatrix}.$$

If the angle of the mobile from the main antenna direction is given by θ, the received power by the mobile at angle θ is given by:

$$p(\theta,\theta') = P_{tx}\left(\frac{2}{N}\sum_{k=1}^{N/2}\cos\left(\pi\frac{N-2k+1}{2}(\sin\theta-\sin\theta')\right)\right)^2, \tag{IV.3}$$

where P_{tx} is the total transmitted signal power. Observe that the maximum received power is obtained when $\theta = \theta'$.

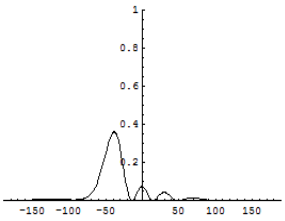
Adaptive beamforming consists in applying the appropriate phase shift that maximizes the power received by the served UE. The possible phase shifts can be chosen among predefined codebook. Thus, we define two beamforming types: Codebook-based Beamforming and Position-based Beamforming referred to by CB and PB respectively.

IV.2.3 Codebook-based beamforming CB

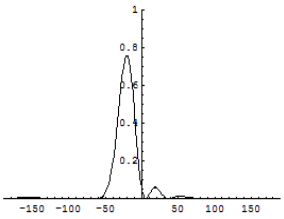
A codebook with a finite number of beams is defined. Each beam corresponds to a defined phase shift. Thus, a codebook of size N_B is defined, for instance, as follows:

$$\theta'_n = -\frac{\pi}{3} + \frac{\pi}{3N_B} + 2n\frac{\pi}{3N_B} \qquad n = 0\ldots N_B - 1.$$

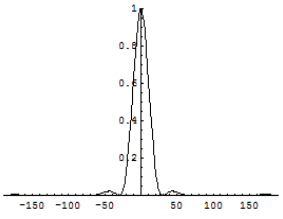
The user should return the index n of the beam which maximizes the received power. This beamforming type is required when the BS is not aware of the position of the user. A training period may be necessary in order to define the best beam of the user.



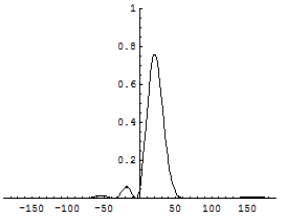
(a) $\theta' = -48°$



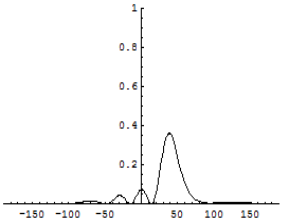
(b) $\theta' = -24°$



(c) $\theta' = 0$



(d) $\theta' = 24°$



(e) $\theta' = 48°$

Figure IV.6: Normalized received power with respect to the direction of the user for $N = 4$ considering a codebook of size $N_B = 5$.

For a static user, one training period may be sufficient. However, when the user is mobile the best beam should be updated continually. This requires rapid signal processing and powerful algorithms, especially when the number of beams is large.

Figure IV.6 shows an example of a predefined codebook of size $N_B = 5$, considering a uniform linear array of $N = 4$ antenna elements. This figure illustrates the normalized received power with respect to the direction of the user:

$$p(\theta, \theta')A(\theta)/P_{tx},$$

for different beams (phase shift), when taking into account the horizontal antenna pattern of a BS sector:

$$A(\theta) = -\min\left[12\left(\frac{\theta}{\theta_{3dB}}\right)^2, A_m\right],$$

$\theta_{3dB} = 70°$ is the 3dB beamwidth and $A_m = 20$dB is the maximum attenuation. Observe that, for each beam, the maximum power is obtained in the exact direction θ'.

Note that another similar vertical antenna pattern is used for elevation in simulations, see Appendix A.

IV.2.4 Position-based beamforming PB

This case considers an infinite codebook size. The BS is supposed to be aware of the location of each user so it can steer the beam in the exact direction of the user.

$$\theta' = \theta_{UE}.$$

This beamforming type corresponds to the optimal one. However, using this type comes at the cost of a higher complexity.

IV.3 Blind beamforming

IV.3.1 Blind scheduling

Blind scheduling refers to classical schedulers (RR, PF, MaxC/I, Mob-aware ..., see Chapter II), in the sense that each BS takes its scheduling decisions independently of the scheduling decisions taken by the neighboring BSs. This does not impact the performance in a system without beamforming since the interference experienced by the users remains the same regardless of the scheduled users in the neighboring BSs. However, in a system with beamforming, the performance of a user depends critically on the scheduled users (activated beams) in the neighboring cells.

This distributed approach benefits from low complexity since no information exchange is required between the different BSs. However, BSs choose the Modulation and Coding Scheme (MCS) of their scheduled users based on reported CQI evaluated through channel measurements done in the previous TTI, where a specific set of beams were activated in different cells. Note that this set of beams activated in the previous TTI may not remain the same in the current TTI, especially under RR strategy. So, if the level of interference experienced by a scheduled user in the current TTI is higher then that measured in previous TTI, an error may happen in the decoding at the receiver side.

In this work, we try to evaluate the error in decoding probability that may happen under blind scheduling. We shall see that this probability is significantly reduced under channel-aware scheduling strategies.

IV.3.2 SINR Map

We denote by u_i the scheduled user in cell i and by $\theta_{u,i}$ the angle of user u from the main antenna direction of BS i. $\theta'_{u,i}$ is the direction of the best beam that maximizes the received power of user u from BS i. $\theta'_{u,i}$ is chosen from a codebook in the case of CB beamforming while it is equal to $\theta_{u,i}$ in the case of PB beamforming.

Thus the SINR seen by user u_s from its serving BS s, when user u_i is scheduled by BS i ($\forall i \neq s$) is given by:

$$SINR(u_s) = \frac{p(\theta_{u_s,s}, \theta'_{u_s,s})P_s/P_{tx}}{\sum_{i\neq s} p(\theta_{u_s,i}, \theta'_{u_i,i})P_i/P_{tx} + \mathcal{N}}. \tag{IV.4}$$

Observe that the level of interference and the SINR depend primarily on the scheduled users in neighboring BSs.

The mean SINR of a user u_s located at position $(\theta_{u_s,s}, \dots \theta_{u_s,i}, \dots)$, under a RR scheduler, can be evaluated through the harmonic mean of different possible experienced SINR, evaluated for all possible combination of scheduled users in all neighboring BSs.

Figure IV.7-IV.13 show the SINR map (in dB) for $N = 4, 8, 16, 32, 64, 128, 256$ antenna elements respectively and for different codebook sizes (finite for CB beamforming and infinite for PB beamforming), considering a RR scheduler. The mean SINR experienced in each position is nothing more then the harmonic mean of different possible SINR as given by equation (IV.4). Observe that the higher the number of beams in the case of CB beamforming, the closer the performance to the case of PB beamforming.

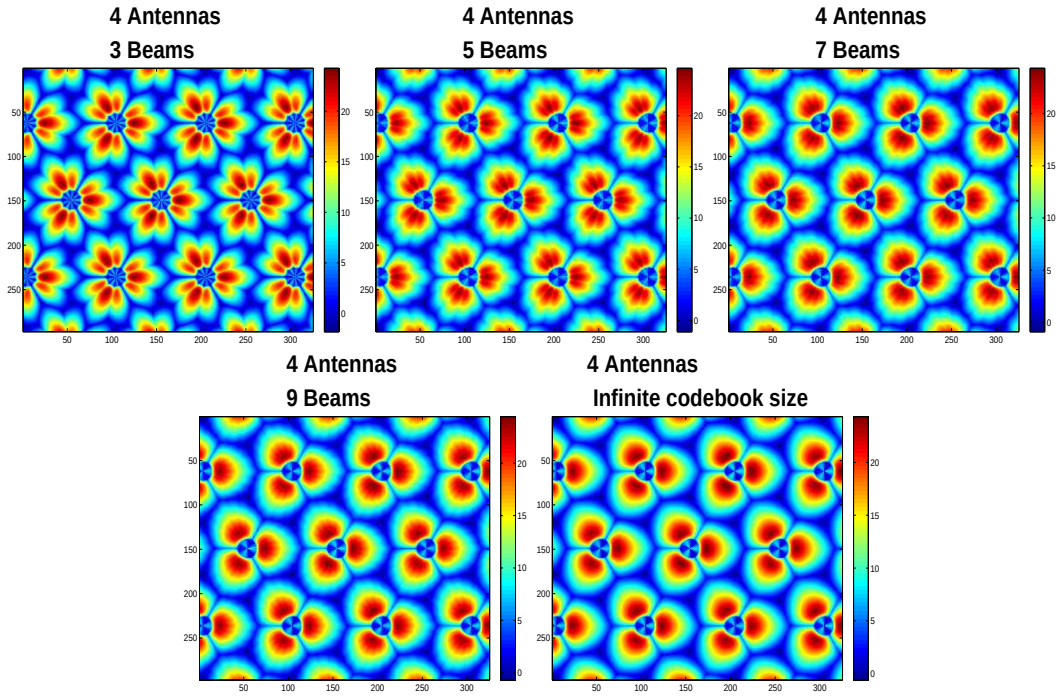


Figure IV.7: SINR maps for N=4 antennas and different codebook size

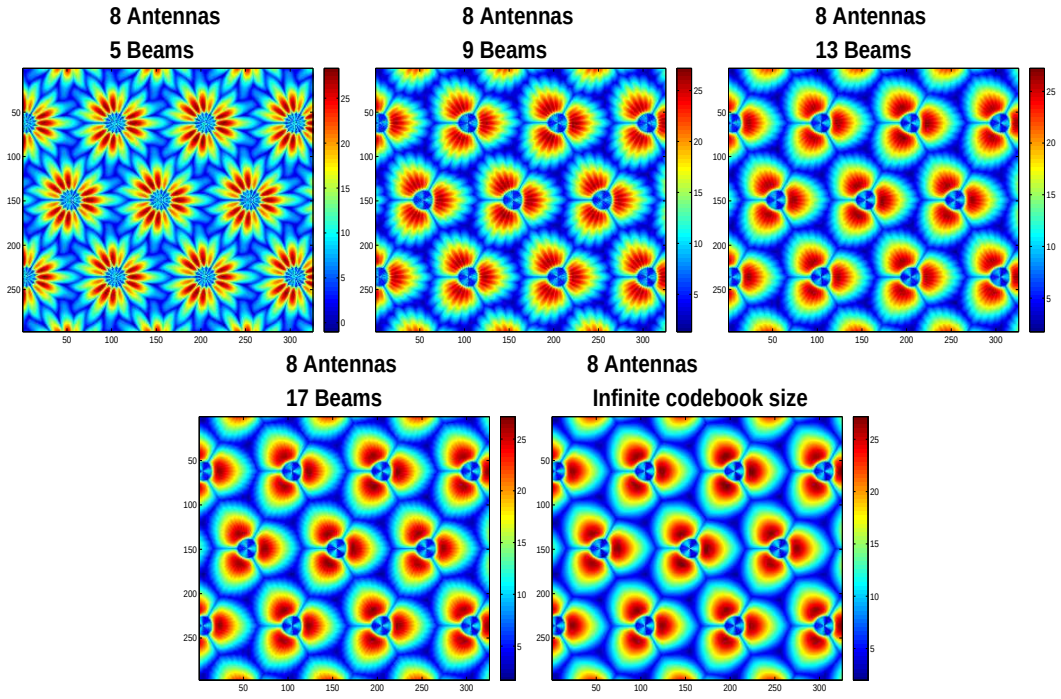


Figure IV.8: SINR maps for N=8 antennas and different codebook size

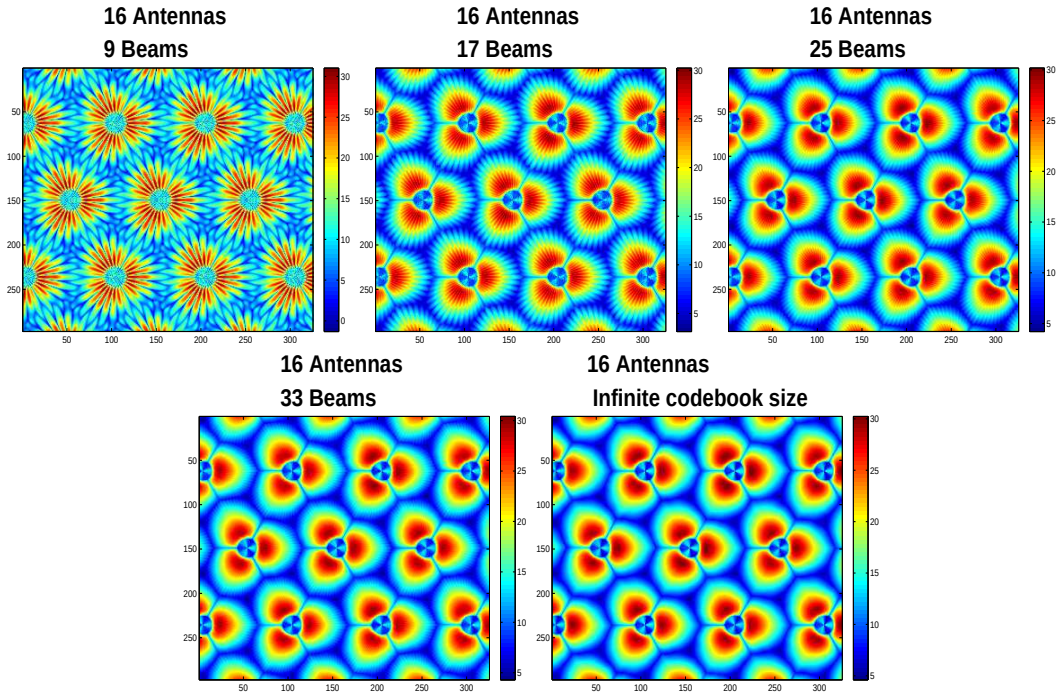


Figure IV.9: SINR maps for N=16 antennas and different codebook size

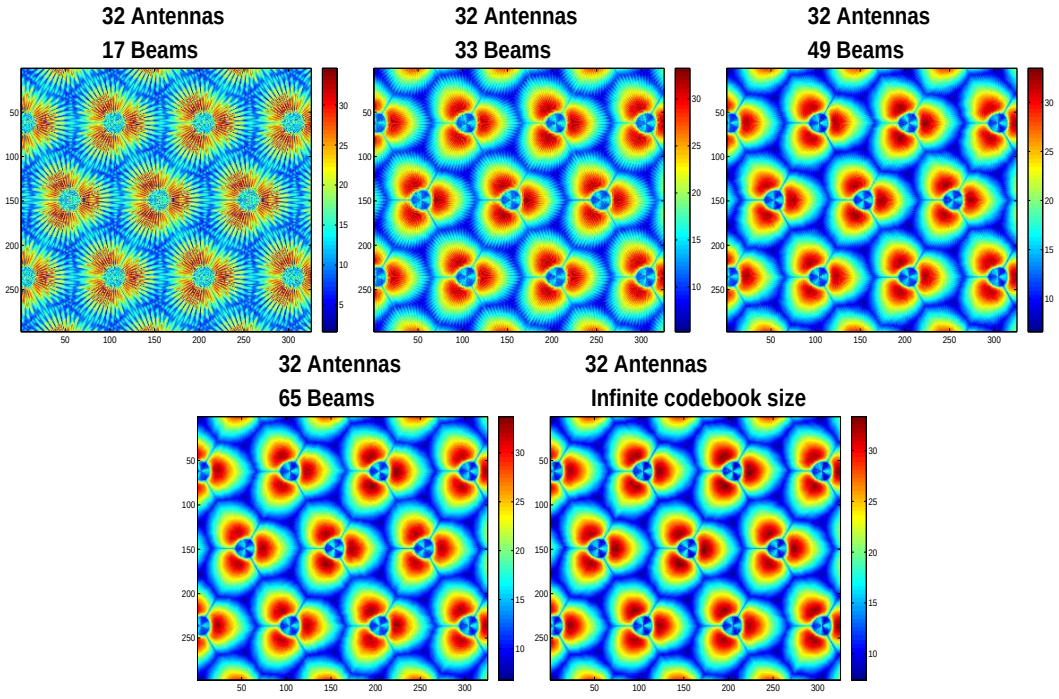


Figure IV.10: SINR maps for N=32 antennas and different codebook size

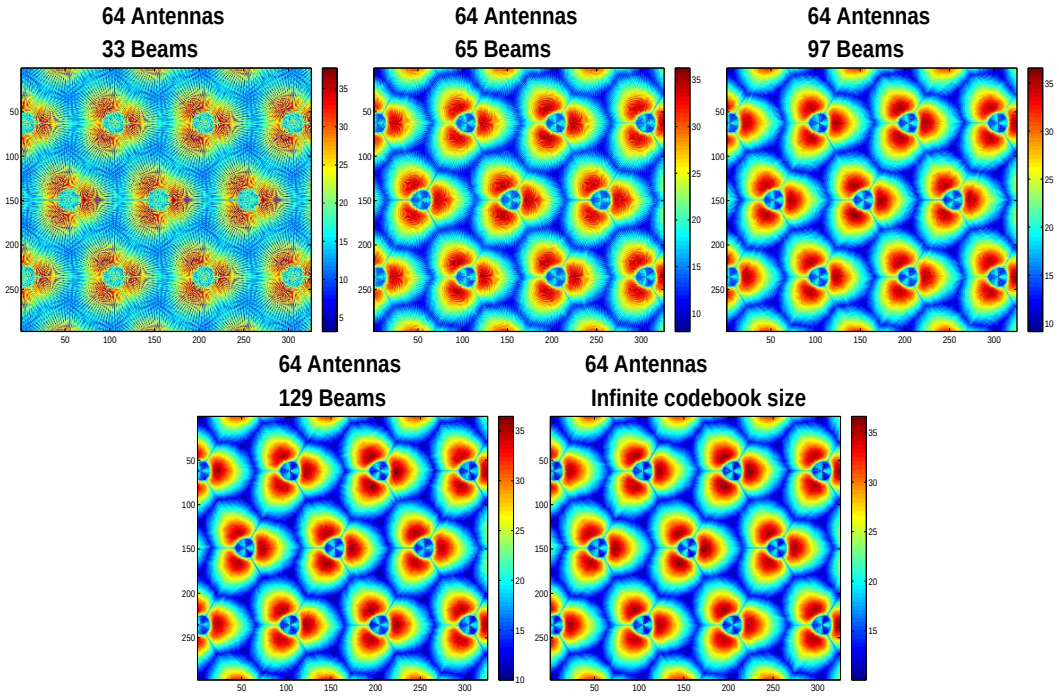


Figure IV.11: SINR maps for N=64 antennas and different codebook size

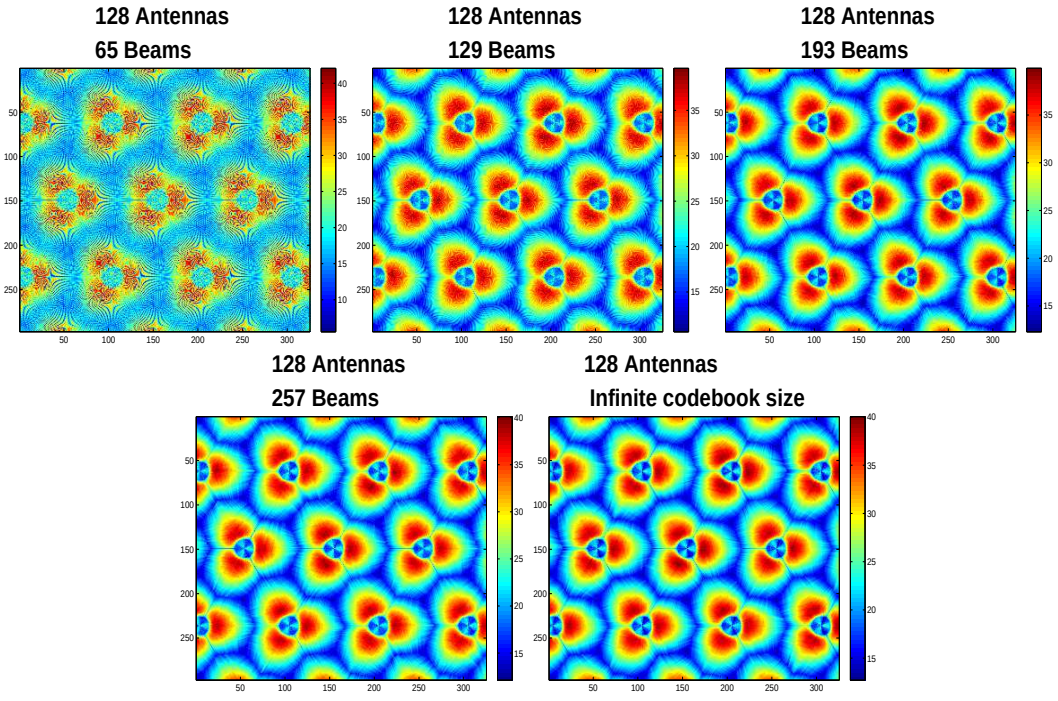


Figure IV.12: SINR maps for N=128 antennas and different codebook size

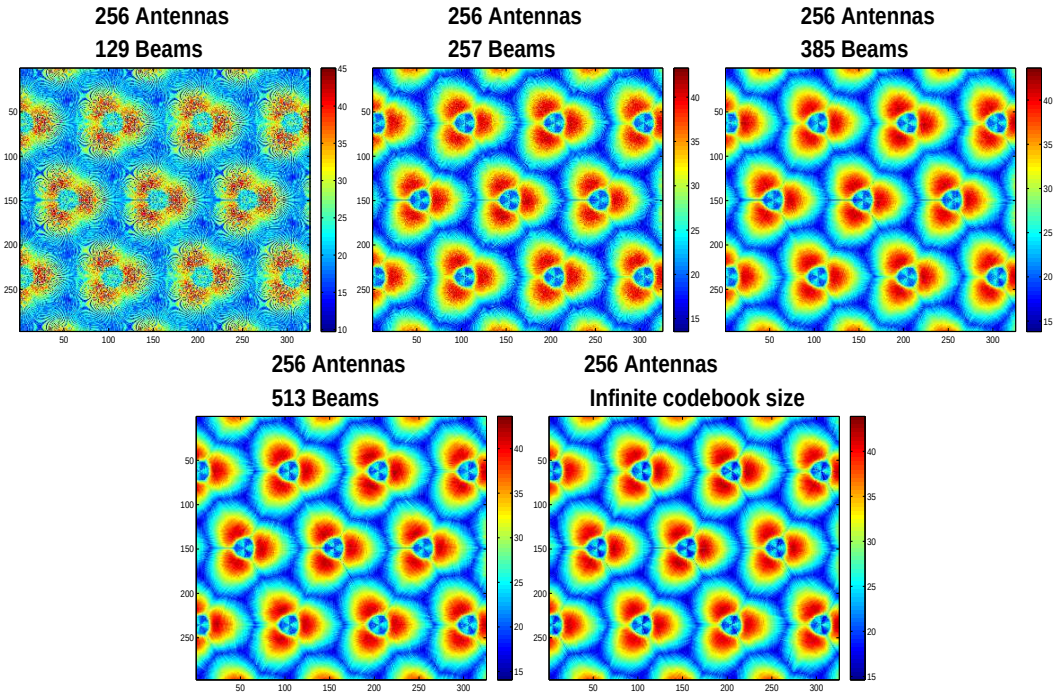


Figure IV.13: SINR maps for N=256 antennas and different codebook size

An insufficient number of beams may degrade the performance of some users, especially those who are located in a region similarly distanced from two beams' main direction (e.g., see Figure IV.7, case $N = 4$ and $N_B = 3$). Thus, the number of beams should be high enough in order to cover all zones of the cell. The higher the number of antennas, the tighter the beams and the lower the probability that a user located at a given position is strongly interfered by neighboring cells. The number of antennas as well as that of beams have also an impact on the minimum and the maximum SINR that can be experienced in the network. The maximum SINR is around 25 dB for $N = 4$ while it can reach 45 dB for $N = 256$. The minimum SINR is less than 0 dB for $N = 4$, while it is around 15 dB for $N = 256$. Note that this improvement is mainly due to the decrease in the level of interference, since the mean interference seen by a user in a system with beamforming is lower than the interference experienced in a system without beamforming.

Figure IV.14 gives the cumulative distribution function of the SINR for different number of antenna elements N, for both CB beamforming and PB beamforming and for different number of beams $N_B = N+1, 3N/2+1, 2N+1, \ldots \infty$.

Observe that the mean SINR increases significantly with the number of antennas. The mean SINR is around 5 dB in the case without beamforming. In the case with beamforming, the mean SINR increases to 8-9 dB when $N = 4$ and reaches 26-27 dB when $N = 256$. Observe that the mean SINR increases by around 3 dB when the number of antennas is doubled. The probability of experiencing an SINR<20 dB when $N = 256$ is lower than 10%.

Furthermore, increasing the size of the codebook increases slightly the mean SINR. However, observe that the cumulative distribution functions become very close when the number of beams is high enough to cover the whole zones of the cell ($3N/2+1$ in this case). Note that this number ($3N/2+1$) may be affected by the resolution of the map (taken equal to 5 m in this study) so it can be higher in real systems.

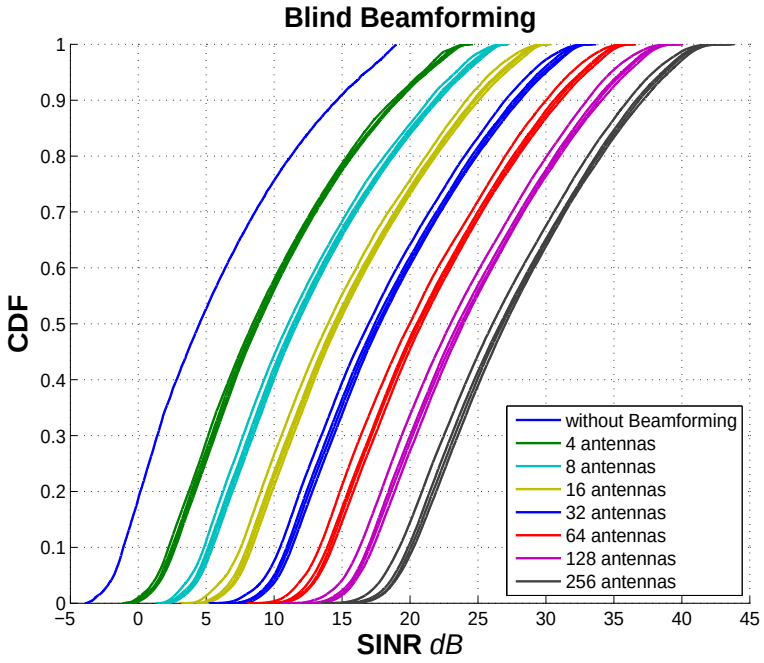


Figure IV.14: SINR CDF for blind beamforming for different antenna size, codebook size going from N+1 to infinite (from left to right)

IV.3.3 Capacity gain analysis

Given the SINR in equation (IV.4) of user u_s when user u_i is scheduled by BS i ($\forall i \neq s$), the bit rate of user u_s is written as follows:

$$r(u_s) = F\left(\frac{p(\theta_{u_s,s}, \theta'_{u_s,s})P_s/P_{tx}}{\sum_{i\neq s} p(\theta_{u_s,i}, \theta'_{u_i,i})P_i/P_{tx} + \mathcal{N}}\right).$$

The mean rate $\bar{r}(u_s)$ experienced by user u_s, under a RR scheduler, is the harmonic mean of all possible rates evaluated for all possible combinations of scheduled users in neighboring cells. Given the mean rate experienced by users in all locations of the cell, the mean user throughput, when allocated all the cell resources, is again evaluated through the harmonic mean.

Figure IV.15 shows the mean capacity gain, under RR scheduling strategy, of a system with beamforming compared to a system without beamforming, that is the mean rate gain of a user when allocated all the cell resources. Observe that the network capacity increases significantly with the number of antenna elements. It increases between 80% and 100% approximately when the number of antenna elements is doubled.

We consider both CB beamforming with different number of beams $N_B = N/2+1, N+1, 3N/2+1, 2N+1$, and PB beamforming (infinite codebook size). Increasing the size of the codebook, increases the mean capacity gain. However, the performance becomes very close when the number of beams is high enough to cover the whole zones of the cell. Note that the capacity gains, in the case of CB beamforming, do not take into account the signaling overhead, due to the amount of reference signals that should be sent to the user in this case.

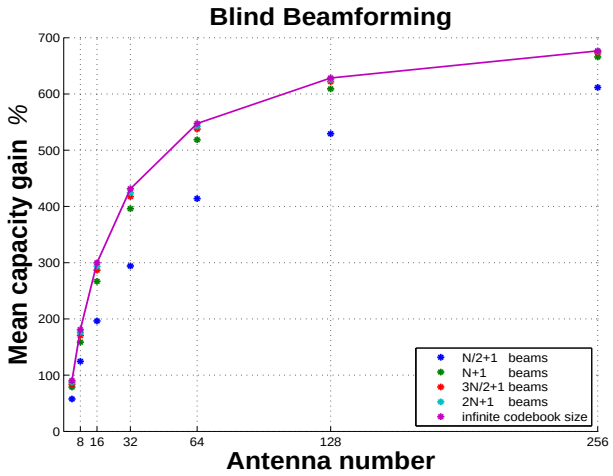


Figure IV.15: Mean capacity gain under RR blind scheduling with respect to the number of antennas, for different codebook size.

IV.3.4 Numerical results

Figure IV.16 gives the performance in terms of mean user throughput with respect to the offered traffic, obtained by system-level simulations, using the LTE simulator described in Appendix A, assuming that mobile devices are equipped with one omni-directional antenna. The figure shows results for a system without beamforming (similarly to Figure II.19) and for a system with blind beamforming

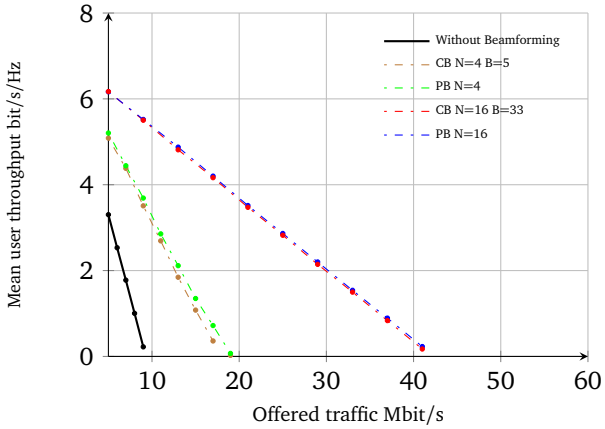
with $N = 4$ and $N = 16$ under different scheduling strategies, assuming static users with a single receive antenna. A user arriving in the network, gets connected to the cell providing the maximum mean received power. The index of the beam that maximizes the mean received power of each user is returned to the eNodeB at each TTI, in the case of codebook-based beamforming. However, when position-based beamforming is used, the eNodeB is supposed to be aware of the positions of the users, so that the beam is steered according to these positions. Since the number of antenna elements is not very high, we still consider fast fading and shadowing in a urban environment.

Figure IV.16a shows the performance under a round robin scheduler for both PB beamforming with $N = 4$ and $N = 16$, and CB beamforming with ($N = 4$, $N_B = 5$) and ($N = 16$, $N_B = 33$). For the sake of comparison, this figure shows also the performance under a round robin scheduler in a system without beamforming. Observe that when PB beamforming is used with $N = 4$, the capacity gain is around 100%, as predicted in the capacity analysis done in section IV.3.3. When $N = 16$ antenna elements are used on the transmitter side, the capacity gain is around 300%, as predicted in section IV.3.3. Performance in terms of mean user throughput of CB beamforming is very close to that of PB beamforming, when the number of beams is high enough to cover all the zones of the cell.

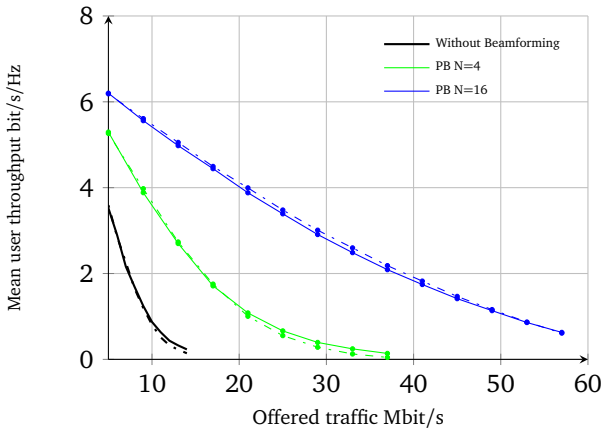
Note that this figure shows the ideal performance that can be obtained under a round robin scheduler. Actually, the MCS of each user is chosen based on the level of interference experienced in previous TTI which is generally not the same as the one experienced in the current TTI: the scheduled users and the associated activated beams in neighboring cells change each TTI under a round robin scheduler. So, if the level of interference experienced by a scheduled user in the current TTI is higher then that measured in previous TTI, an error is more likely to happen in the decoding at the receiver side. This error probability should be taken into account in the effective final transmitted volume. Figure IV.17b shows the error probability under a round robin scheduler in a system with blind beamforming. This error probability is evaluated as the percentage of transmitted data, when the MCS chosen by the BS (based on previous measurements) is higher than the MCS that can be decoded by the scheduled user in the current TTI. Observe that the error probability increases with the load of the system; it reaches 40% approximately at high load for both $N = 4$ and $N = 16$.

Consider now channel-aware scheduling, where scheduling decisions are taken based on the users' channel quality measured in previous TTI. Figure IV.16b gives the performance in terms of mean user throughput under PF scheduling strategy (solid lines) and max C/I strategy (dashed lines) for a system without and with beamforming, considering PB beamforming only, with $N = 4$ and $N = 16$. Compared to round robin scheduling strategy, there is an important opportunistic gain brought by these opportunistic schedulers. Observe that PF scheduler outperforms max C/I scheduler at high load when users are static, which is the same conclusion driven in Chapter II.

Similarly, under these schedulers, the MCS of each user is chosen based on the level of interference experienced in previous TTI. However, as scheduled users are chosen based on their channel quality, channel aware schedulers, more specifically max C/I schedulers, in different cells are more likely to chose a combination of users who do not interfere between each other. Generally, under max C/I scheduling strategy, the user experiencing the best radio conditions is scheduled a number of successive TTIs. The variation of the experienced interference is smoother in this case. Thus, the measured interference in previous TTI is more likely to be the same as the experienced interference in current TTI. The performance is primarily affected by granularity of the scheduler and the accuracy of the feedback. However, there might still be error but the probability of error is less than that under a round robin scheduler. Figure IV.17b shows the error probability in a system with blind beamforming under PF and max C/I scheduling strategies. Observe that the error probability is smaller under max



(a) Round-Robin

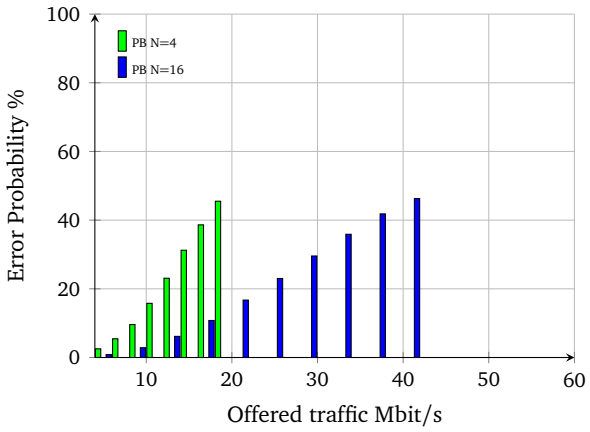


(b) Proportional Fair (solid lines) and max C/I (dashed lines)

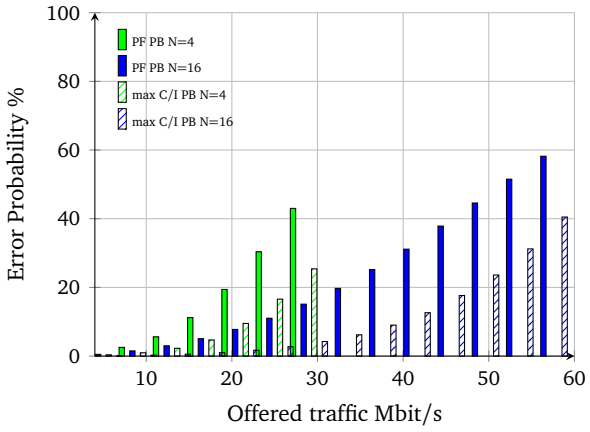
Figure IV.16: Throughput performance obtained by system-level simulation for blind beamforming under different scheduling strategies, for static users.

C/I scheduler compared to PF strategy. This is due to the fact that max C/I metric is only based on instantaneous channel quality while PF metric takes into account the amount of downloaded data, in addition to the channel quality which increases the error probability.

In order to quantify the gains brought by using more sophisticated coordinated schedulers, we consider in the next section, coordinated scheduling/ beamforming technique under different clustering strategies.



(a) Round-Robin scheduling



(b) Channel-aware scheduling

Figure IV.17: Error probability for blind beamforming under different scheduling strategies.

IV.4 Coordinated beamforming

IV.4.1 Coordinated scheduling

The scheduling is done by a coordination between different cells, so that the set of scheduled users in all coordinated cells (cluster) corresponds to the best set of users that can be scheduled together when performing beamforming, as shown in Figure IV.18. In other words, each cell activates the beam which less interferes with the scheduled users within the neighboring coordinated cells which reduces the interference.

For the sake of comparison and in order to quantify the maximum gains that can be brought by the coordination between different cells, we consider an optimal scheduling strategy within each cluster, where at each TTI the best combination of users that maximizes a given metric (PF sum metrics, max C/I sum metrics ...) is chosen. We consider a centralized scheduler in each cluster.

PF coordinated scheduling (PF-CS):

Under this strategy, the scheduled users are chosen based on the sum of PF metrics (the instan-

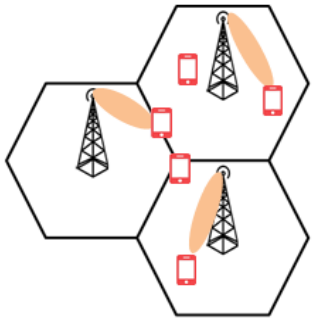


Figure IV.18: Coordinated scheduling / beamforming.

taneous data rate relative to the mean data rate), by solving the following problem:

$$\max \sum_{c\in\mathbb{C}} \sum_{u\in\mathbb{U}_c} \frac{1}{\bar{R}_u} \boldsymbol{F}\left(\frac{b^f_{u,c} p(\theta_{u,c},\theta'_{u,c}) P^f_{u,c}/P_{tx}}{\sum_{i\in\mathbb{C},i\neq c} \sum_{v\in\mathbb{U}_i} b^f_{v,i} p(\theta_{u,i},\theta'_{v,i}) P^f_{v,i}/P_{tx} + I_u + \mathcal{N}} \right)$$

$\forall c, \forall f \in [1,2,\dots N_{RB}]$:

$$\sum_{u\in\mathbb{U}_c} b^f_{u,c} \leq 1$$

We denote by $P^f_{u,x}$ the signal power received by user u from cell x over frequency sub-band f, including pathloss, shadowing, fast fading and taking into account the antenna pattern of one antenna element (See Appendix A). I_u is the residual interference caused to user u associated with cell c, by all neighboring cells not taking part in the cooperative cluster of cell c in previous TTI. $\bar{R}_u$ is the mean data rate of user u evaluated through an exponentially smoothed average over time.

$\mathbb{C}$ is the set of coordinated cells (cluster). $\mathbb{U}_x$ is the set of users associated with cell x (served by cell x). We consider only single user beamforming (SU beamforming); a cell can transmit to one user on a given sub-band f at a time.

$b^f_{u,x} \in \{0,1\}$ is a variable to indicate whether cell x is transmitting to user u on frequency sub-band f in the current TTI or not:

$$b^f_{u,x} = \begin{cases} 0 & \text{cell } x \text{ is not transmitting to user } u\ , \\ 1 & \text{cell } x \text{ is transmitting to user } u\ . \end{cases}$$

Max C/I coordinated scheduling (max C/I-CS):

Under this strategy, the scheduled users are chosen based on the sum of max C/I metrics, as follows:

$$\max \sum_{c\in\mathbb{C}} \sum_{u\in\mathbb{U}_c} \boldsymbol{F}\left(\frac{b^f_{u,c} p(\theta_{u,c},\theta'_{u,c}) P^f_{u,c}/P_{tx}}{\sum_{i\in\mathbb{C},i\neq c} \sum_{v\in\mathbb{U}_i} b^f_{v,i} p(\theta_{u,i},\theta'_{v,i}) P^f_{v,i}/P_{tx} + I_u + \mathcal{N}} \right)$$

$\forall c, \forall f \in [1,2,\dots N_{RB}]$:

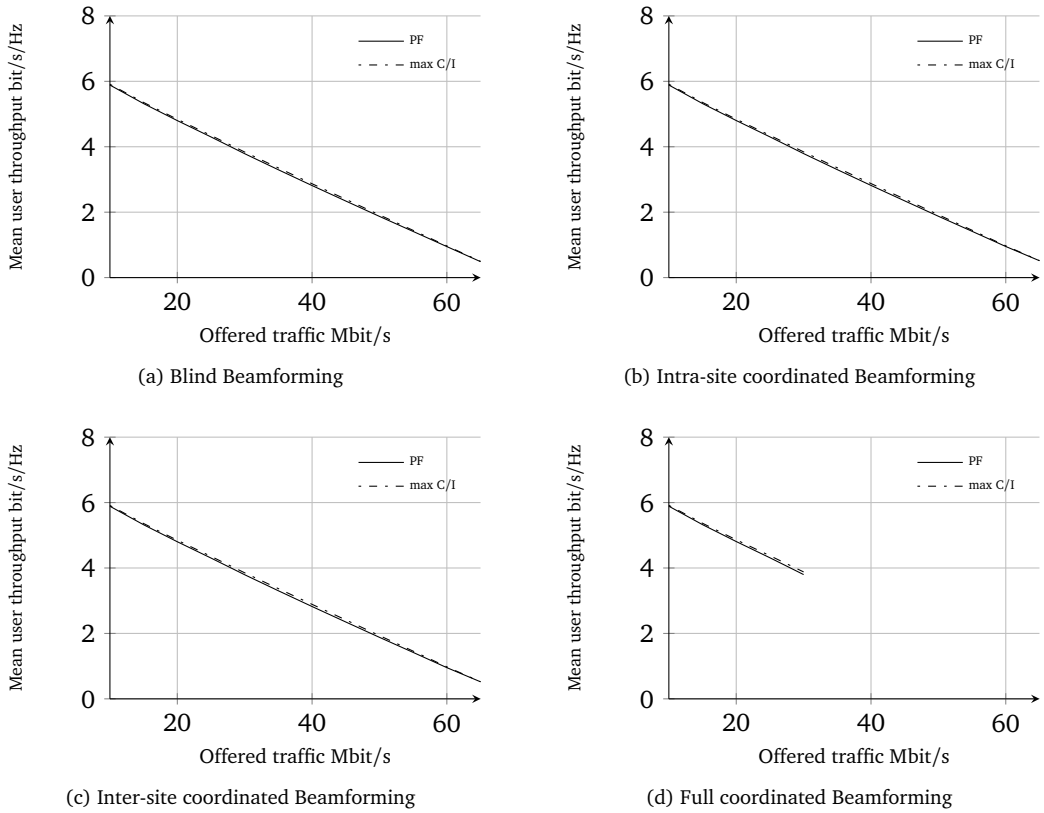
$$\sum_{u\in\mathbb{U}_c} b^f_{u,c} \leq 1$$

Observe that, unlike blind scheduling in Section IV.3.1 where standard CQI reports are sufficient, this centralized scheduler should be aware of the power received by all users from all cells taking part in the cooperative cluster, considering the different beams. This requires a large amount of feedback.

Small cooperative clusters, like in the case of intra-site coordination and inter-site coordination, does not guarantee zero error probability due to the residual interference I_u which is measured in previous TTI, and which can vary in the current TTI, according to the scheduled users and activated associated beams in non-cooperating cells. A fully coordinated network could guarantee zero error probability but this at the cost of higher complexity in terms of processing overhead and feedback.

IV.4.2 Numerical results

We consider that $N = 64$ antenna elements are deployed at each sector. In this case, the beams are tight enough allowing to neglect fast fading and shadowing effect. We consider PB beamforming only. We evaluate the performance of blind beamforming as well as coordinated beamforming under different clustering strategies. We consider intra-site coordination, inter-site coordination and full coordination. Standard PF and max C/I schedulers are considered in the case of blind beamforming while PF-CS and max C/I-CS strategies are considered in the case of coordinated beamforming.



(a) Blind Beamforming

(b) Intra-site coordinated Beamforming

(c) Inter-site coordinated Beamforming

(d) Full coordinated Beamforming

Figure IV.19: Throughput performance obtained by system-level simulation for blind beamforming and intra-site, inter-site and full coordinated beamforming under PF and max C/I scheduling strategies, for static users.

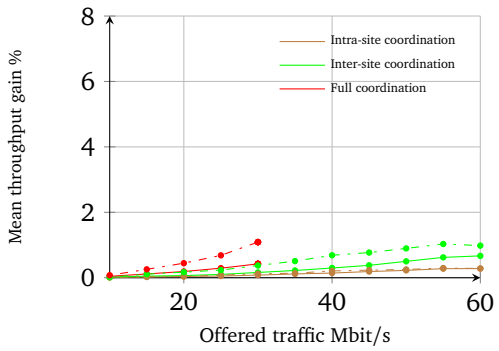
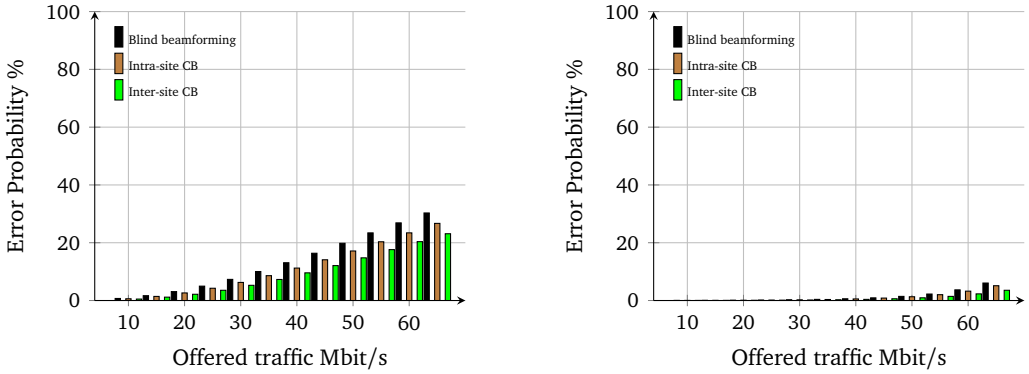


Figure IV.20: Throughput gain brought by intra-site, inter-site and full coordinated beamforming under PF-CS (solid lines) and max C/I-CS (dashed lines) strategies.



(a) PF scheduling

(b) max C/I scheduling

Figure IV.21: Error probability for blind beamforming, intra-site and inter-site coordinated beamforming under PF and max C/I scheduling strategies.

Figure IV.19 shows the mean user throughput performance obtained by system-level simulations under PF and max C/I scheduling strategies, assuming static users. In the absence of shadowing and fast fading, the performance in terms of mean user throughput under both scheduling strategies (PF and max C/I) must be close to the performance that can be obtained under round robin scheduling strategy. However, the error probability is less under channel-aware scheduling strategies, as seen in the previous section. Observe that the capacity gain is around 550% as predicted from Figure IV.15.

The performance obtained under intra-site, inter-site and full coordinated schedulers are very close to the performance obtained under standard scheduling strategies in a beamforming system, as shown in Figure IV.19. The gains brought by these sophisticated coordinated schedulers under different clustering strategies are illustrated in Figure IV.20. These gains are negligible in view of the added complexity, in terms of feedback requirements, backhaul capacity and speed, in addition to the complexity of the scheduling algorithm that increases significantly with the load of the system and the number of coordinated cells. The coordination brings only limited further gains.

A fully coordinated network guarantees zero error probability. However, we can see that the error probability, when $N = 64$ antenna elements are deployed at the transmitter side, is very low, especially under max C/I scheduling strategy, even in a system with blind beamforming, as shown in

Figure IV.21.

When using the PF metric, the maximum error probability is around 30% under blind beamforming. Intra-site coordination and inter-site coordination decrease slightly this probability. However, the error probability is lower in the case of inter-site coordination compared to the case of intra-site coordination.

When using the max C/I metric, the error probability is negligible under blind and coordinated beamforming (less than 6%). This leads to the following conclusion: using channel-aware standard non-coordinated schedulers, and more specifically the max C/I strategy, in a beamforming system where a large number of antenna is deployed, can guarantee negligible error probability without adding any complexity to the system.

Note that in the case where only a limited number of antennas is deployed at the BS side, coordination may be needed as the beams are larger in this case, thus the error probability becomes bigger. So, there is a trade-off between the number of antenna elements to be deployed and the complexity of the scheduler, the feedback requirements and the burden that may be placed on the network.

IV.5 Conclusion

We have studied in this chapter a beamforming system. We have defined two beamforming types: codebook-based beamforming and position-based beamforming. A codebook with a large number of beams which allow to cover all the zones in the cell, enables to achieve similar performance as position-based beamforming.

We have considered first a system with **blind beamforming**, where no coordination is performed between different cells. We have quantified the capacity gains with respect to a system without beamforming for different numbers of antenna elements. We have studied the performance under several scheduling strategies. We have shown that the probability of error (measured as the percentage of transmitted data when the MCS chosen by the BS is higher than the MCS that can be decoded by the scheduled user in the current TTI), is lower under channel aware schedulers than under round robin scheduler. Max C/I scheduler, which chooses users based only on their instantaneous channel quality, achieves lower error probability than PF scheduler. A high number of antenna elements enables to form tight beams and leads to lower error probability.

We have then considered a system with **coordinated beamforming** under different clustering strategies, in order to quantify the gains brought by using more sophisticated coordinated schedulers. We have shown that coordinated scheduling/beamforming brings only limited further gains compared to blind beamforming operated under standard non-coordinated channel-aware schedulers. In addition to the fact that only a small gain is brought by these schedulers, a high complexity is added to the network, in terms of feedback and signaling overhead, backhaul capacity and speed requirements, and processing burden which increases significantly with the load of the system and the number of coordinated cells.

Consequently, using channel-aware standard non-coordinated schedulers, and more specifically the max C/I strategy, in a beamforming system where a large number of antenna is deployed, can guarantee negligible error probability without any additional feedback requirements and without adding any complexity to the system.

However, more sophisticated coordinated schedulers may be needed in the case when only a limited number of antennas is deployed at the BS side in order to guarantee a low error probability, but this is at the detriment of higher processing overhead, a larger amount of feedback and signaling

placing an additional burden on the network. Thus there is a trade-off to be considered between the number of deployed antenna elements and the complexity of the scheduler, the feedback and signaling requirements.

The results obtained in this chapter offer many perspectives for future work. The scheduling strategy and the channel estimation are very interesting research topics to be considered in a system enabling beamforming. The results have been derived for the case of 2D beamforming. The real channel being 3D characterized, we aim to extend this study to the case of 3D beamforming by considering the vertical dimension. Other interesting topics can be studied as well and include the impact of mobility. Finally, in this study we have only considered the case of single-user (SU) beamforming. It would be also interesting to study the case of multi-user (MU) beamforming, see for instance [44, 69], where the scheduling seems to be very challenging.

Chapter V

Conclusion

In this thesis, we have evaluated the performance of several CoMP techniques and scheduling schemes, with and without mobility, through the analysis of flow level traffic models and system-level simulations. The thesis consists of two main parts. In the first part, we consider the case of a network without inter-cell coordination while in the second part, different cell coordination schemes are studied.

V.1 Network without inter-cell coordination

In the first part of the thesis, we have studied the impact of mobility in cellular networks, in the presence of elastic traffic or adaptive streaming traffic, without considering any coordination mechanism between different cells.

While advanced wireless systems exploit fast channel variations through opportunistic scheduling, we show that slow channel variations due to mobility can be exploited as well in the presence of elastic traffic. We have quantified the gains induced by mobility in cellular data networks under round-robin, max C/I and proportional fair policies. Note that we have not take into account the actual transmission rates reduction that may occur at high speeds due to the estimation and prediction problems, which is another problem that can be studied when dealing with mobility. The performance at flow level improves as the mobility of the users increases, under the different considered scheduling policies. However, the optimal choice of the fairness factor depends critically on mobility. In the presence of mobility, the more opportunistic the scheduler is the higher is the system capacity. In contrast, proportional fairness is the best strategy in the absence of mobility.

Thus we have proposed a **mobility-aware scheduler** that exploits the user mobility information to adapt to the mobility of users so it tends to be more opportunistic when all users are mobile and more fair when all users are static. Mobile users at the edge are deprioritized since they are likely to move and to be served in better radio conditions. We compare the performance of the proposed scheduler to that of other usual scheduling policies (round-robin, max C/I, proportional fair) through the analysis of flow level traffic models and we validate the results by system level simulations. It turns out that this scheduler is a good compromise in the presence of multiple mobility behaviors and improves the overall performance, especially at high load.

It is important to point out that the gain brought by this scheduling strategy is by exploiting only information related to the speed, without any trajectory information. The gains would certainly be higher with this information but this is at the cost of much higher complexity.

Although the focus of this work is on the downlink, we think that similar performance can be obtained on the uplink. An interesting problem for future research would include both aspects of mobility, performance at flow level as well as prediction and decoding errors related to mobility, a point that we let for future work. Impatience of users is also an important point to consider especially when dealing with prioritization scheduling strategies that have been ignored in this work. Other interesting problems for future research include opportunistic scheduling policies that prioritize short flows, as considered in [9] for instance, as well as heterogeneous networks consisting of both picocells and macro cells [40].

We have also examined in this thesis the performance under different scheduling strategies in the presence of adaptive streaming traffic. Two performance indicators, mean video bit rate and mean buffer surplus, which are calculated based on the flow-level dynamics, are considered in order to evaluate performance. We show that the max C/I policy pushes users to select higher video bit rates which may degrade the stability condition compared to that obtained under round robin scheduler, when users are static. However, when users are mobile, the max C/I policy provides better mean video bit rates, but may decrease the mean buffer surplus. Therefore, we propose a discriminatory scheduler that allows to balance both performance metrics, in the presence of mobility. Nevertheless, the case of mixed service, with both elastic traffic and adaptive streaming traffic, has not been considered in this work. Designing a performant scheduling strategy for mixed service types and mixed mobility behaviors is a challenging and interesting problem for future work.

V.2 Network with inter-cell coordination

The second part of this thesis deals with coordination schemes in the downlink mainly, **joint processing** and **coordinated scheduling/beamforming**.

We have first studied JP techniques, more precisely JT and DPB. We have evaluated the performance of both techniques through flow level models and system level simulations considering non-full buffer traffic models. The introduction of such a feature may create some problems due to the excessive resource consumption, especially in a highly loaded network. Indeed, the performance depends strongly on the interference level in the network through the coordination gain, that is the mean throughput gain brought by the cooperation of a cell to a cell-edge user.

We have investigated an interference dominated environment, where we have shown that JT, which consists in converting an interfering signal into a desired signal, could be interesting when applying a transmission scheme that enable a sufficiently high mean coordination gain that compensates the extra radio resource consumption incurred by this technique. This may require multiplexing and advanced receivers techniques.

We have analyzed the capacity issues in the presence of JT and have proposed a new coordination scheme where a cell cooperates only when it brings at least 100% mean throughput gain when cooperating, mainly in a symmetric network topology. In the case of a general network topology, a more advanced coordination scheme combined with SON algorithm can be used. The SON algorithm will be responsible of computing the exact coordination gain constraint that fits best to the network characteristics.

We have studied different scheduling strategies and we have shown that a global scheduler that maximizes the instantaneous sum rate and where the coordination can be activated and deactivated dynamically provides better performance compared to an iterative scheduler, but at the cost of higher complexity. We have shown that JT performs better at medium load. Gains depend mainly on the clustering strategy, the transmission scheme as well as the load of the network.

The performance of DPB, which consists in prohibiting transmission from strongest interfering cell(s), is also evaluated in several cases under different clustering methods assuming scheduling granularity of one TTI. We have shown that most of the rate gain is achieved through the elimination of the interference and that DPB may present less restrictive stability condition compared to a JT scheme with moderate coordination gain. The relative gain of DPB compared to the case without coordination increases with load. When the network is very highly loaded, interference between different cells becomes more important. In this case, the blanking of interfering cells at the right moment when the scheduled user is strongly affected by the generated interference seems to bring interesting gains. We show that DPB is more promising than JT since it is a simpler and performant technique, especially at high loads, where only a limited number of antennas is deployed. Even intra-site coordination, which is much easier to implement, brings important gains as all users, even those who are not defined as CoMP users, can benefit implicitly from the muting of some interfering cells which are not necessarily defined as cooperating cells for the benefiting users. However, in order to achieve potential full interference reduction, dynamic cooperation is essential. Thus, DPB seems to be simpler and more promising than JT especially in a high interference scenario where only a limited number of antennas is deployed.

We have also investigated the case of a low interference scenario where we have shown that it is not recommended to activate JP feature. However, if we stick to ensure uniformity of service by performing coordination, more precisely JT, for static cell-edge users suffering from very degraded throughput, it is not worth to perform JT for a mobile user since he is able to move and to get better radio conditions. Consequently, we have proposed a **mobility-aware scheduler**, which is the extension of the scheduler proposed in the case of a network without inter-cell coordination and which deprioritizes mobile CoMP users. This scheduling strategy improves the global performance by giving the chance to mobile cell-edge users to be served in better radio conditions where cell coordination is not required, which avoids the extra resource consumption for those mobile users. This scheduler is suitable for elastic traffic where the delay is tolerable. However, it can no longer work when considering traffic with some time constraints such as video streaming.

It is important to point out that this scheme, especially JT technique, presents many practical challenges in terms of backhaul latency and capacity, synchronization and feedback design.

We have then considered a system with beamforming, using uniform linear antenna array in order to study the gains expected from using CS/CB scheme. We have shown that, performing beamforming based on a codebook with a large number of beams which allow to cover all the zones in the cell, enables to achieve similar performance as position-based beamforming. Further increasing the number of beams does not bring important performance enhancement.

We have studied the case of **blind beamforming**, where no coordination is performed between different cells. The maximum expected capacity gains under a round robin scheduler, with respect to a system without beamforming have been quantified considering different numbers of antenna elements. We have also studied the performance in terms of mean user throughput under different scheduling strategies. We have evaluated the probability of error, measured as the percentage of transmitted data when the modulation order chosen by the BS is higher than the modulation order that can be decoded by the scheduled user at current time. It turns out that channel-aware schedulers, and more precisely max C/I scheduler, provides lower error probability than other scheduling strategies, and further increasing the number of antenna elements enables lower error probabilities.

We have then studied the performance under coordinated schedulers in order to quantify the further gains brought by using **coordinated beamforming**, compared to blind beamforming. We have shown that coordinated scheduling/beamforming brings only limited further gains compared to blind beamforming operated under standard non-coordinated channel-aware schedulers, especially

when a high number of antenna elements is deployed. Deploying such coordination schemes add high complexity to the network and optimal coordinated scheduling strategies cause high processing overhead.

While using channel-aware standard non-coordinated schedulers, and more specifically the max C/I strategy, is sufficient in a beamforming system where a large number of antenna elements is deployed at the BS, sophisticated coordinated schedulers may be needed in the case when only a limited number of antennas is deployed in order to guarantee a lower interference environment. However, the further gain brought by the coordination, is at the expense of higher processing overhead, a larger amount of feedback and signaling. Thus, there is a trade-off to be considered between the number of deployed antenna elements and the complexity of the scheduler, the feedback and signaling requirements.

The results have been derived only for the case of 2D beamforming. It would be also interesting to study the case of 3D beamforming which is better aligned with the channel characteristics. The scheduling strategy in the context of MU beamforming and the impact of mobility on the performance of a system enabling beamforming are also interesting problems for future research.

Appendices

Appendix A

This appendix describes the main characteristics of the LTE simulator used in this thesis. Note that this is a simplified C++ LTE simulator that I have developed for the purpose of evaluating different coordination schemes, as well as different scheduling strategies.

We consider the 21 hexagonal cells formed by 7 tri-sector sites (a reference site surrounded by 6 interfering sites), as illustrated by Figure 1. A wrap-around technique is employed in order to avoid border effects.

The simulator is based on the LTE technology. We assume a carrier frequency of 2GHz, a bandwidth of 10 MHz. We consider an urban environment, in an outdoor context with macro cells only and an inter-site distance of 500m.

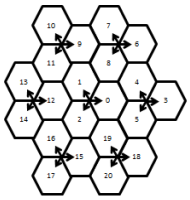


Figure 1: 7 tri-sector sites

Macroscopic pathloss

The macroscopic pathloss between an eNodeB sector and a UE is composed of both the propagation pathloss due to the distance as well as the antenna gain .

The macroscopic pathloss map of each sector is computed once and used during all the simulation time.

Distance-dependent pathloss

We consider the pathloss model of the ITU urban Macro cell (Uma) scenario, see [66]. Figure 2 shows the variation of this pathloss (red) with respect to the distance between the eNodeB and the user.

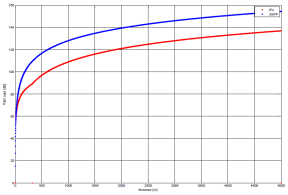


Figure 2: ITU Uma scenario pathloss.

BS antenna pattern

The horizontal antenna pattern, as illustrated by Figure 3, used for each BS, according to 3GPP, is given by:

$$A(\theta) = -\min\left[12\left(\frac{\theta}{\theta_{3dB}}\right)^2, A_m\right],$$

this is the relative antenna gain in dB, for $-180° \leq \theta \leq 180°$, $\theta_{3dB} = 70°$ is the 3dB beamwidth and $A_m = 20$dB is the maximum attenuation.

The vertical antenna pattern, used for the elevation is given by:

$$B(\phi) = -\min\left[12\left(\frac{\phi - \phi_{\text{tilt}}}{\phi_{3dB}}\right)^2, A_m\right],$$

$-90° \leq \phi \leq 90°$. The antenna tilt ϕ_{tilt} is assumed to be 12°, and the elevation 3dB angle ϕ_{3dB} is assumed to be 10°.

Consequently, the combined antenna is given by:

$$-\min[-(A(\theta) + B(\phi)), A_m]$$

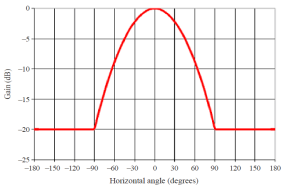


Figure 3: Antenna pattern for tri-sector cells

UE antenna pattern

The user is assumed to be equipped with omni directional antennas.

Shadowing

Large scale shadowing, or simply shadowing is fading that occurs on a large scale. It is caused by the geographical irregularities of the environment due to the presence of obstacles between the UE and the eNodeB, such as building...

Shadowing is generally interpreted as a log-normal deviation with respect to the macroscopic pathloss. In this simulator, shadowing is modeled as a zero-mean Gaussian random variable, with standard deviation 4dB.

A shadowing Map is generated for each site, that is 7 correlated shadowing maps, with a constant cross correlation $\rho_{\text{inter-site}} = 0.5$.

In order to simulate the mobility of users, spatial correlation between shadowing values at different points of the map is also taken into account. This correlation depends on the distance between the two considered points of the map. The farther away two positions, the lower the correlation, which is given by:

$$r(\delta d) = e^{\frac{-\delta d}{d_{corr}}},$$

assuming a shadowing correlation distance of $d_{corr} = 30$m; δd is the distance between the two considered points.

Thus the spatial correlation matrix for each shadowing map of n pixels is given by:

$$C_{\text{sh}} = \begin{bmatrix} 1 & r(\delta d_{1,2}) & r(\delta d_{1,3}) & \dots & r(\delta d_{1,n}) \\ r(\delta d_{2,1}) & 1 & r(\delta d_{2,3}) & \dots & r(\delta d_{2,n}) \\ r(\delta d_{3,1}) & r(\delta d_{3,2}) & 1 & \dots & r(\delta d_{3,n}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(\delta d_{n,1}) & r(\delta d_{n,2}) & r(\delta d_{n,3}) & \dots & 1 \end{bmatrix} \quad (1)$$

Given (1), correlated shadowing values are then generated from zero-mean complex independent identically distributed (i.i.d) as follows:

$$X = AY$$

Y is the vector of n independent complex log-normal random variables, X is the resulting vector of n correlated (1)) log-normal random variables, (according to the correlation matrix. Matrix A results from the Cholesky decomposition of the correlation matrix C_{sh}.

$$C_{\text{sh}} = AA^T$$

The shadowing maps are updated each 1s. Time correlation is also taken into account. Given the generated correlated shadowing vector of a given site X (which is correlated with the 6 shadowing vectors of other sites), the final shadowing vector at time $t+1$ is given by:

$$X_{t+1} = \rho X_t + \sqrt{1-\rho^2} X,$$

ρ is the 1s time correlation which is considered as 85% in this simulator. Figure 4 gives an example of shadowing maps generated as described in this section.

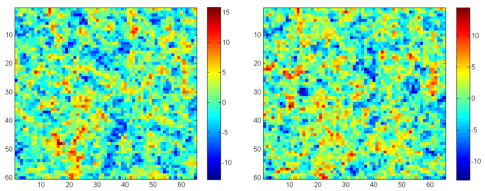


Figure 4: Shadowing map

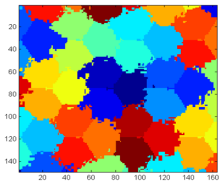


Figure 5: Serving cell.

Figure 5 shows the serving cell association of each point of the map, according to the received power including macroscopic pathloss and the shadowing.

Figure 6 shows the SINR maps without and with shadowing.

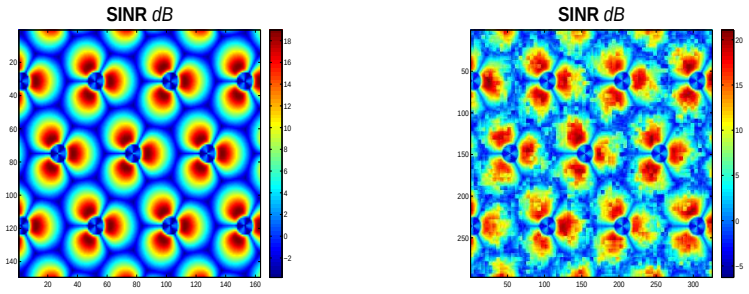


Figure 6: SINR map without and with shadowing (from left to right)

Fast fading

Small-scale channel variations in wireless communication environments are due to multipath propagation; a signal transmitted from the transmitter reaches the receiver through many different paths. Unlike macroscopic pathloss and shadowing, small-scale fading is updated each 1ms. We suppose, for instance, that eNodeBs and mobile devices are equipped with 2 antennas, that is the case of a 2x2 MIMO channel (in the case without beamforming). The channel matrix $H \in \mathbb{C}^{2x2}$ is then given by:

$$H = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}, \tag{2}$$

According to TS 36.101 [8], the spatial correlation matrix at the transmitter side is given by:

$$R_{\text{BS}} = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix},$$

while the correlation matrix at the receiver side is given by:

$$R_{\text{UE}} = \begin{bmatrix} 1 & \beta \\ \beta & 1 \end{bmatrix}$$

The spatial correlation matrix of the MIMO radio channel is then the Kronecker product, see [42], of the spatial correlation matrix at the transmitter and that at the receiver:

$$R_{\text{MIMO}} = R_{\text{BS}} \otimes R_{\text{UE}} = \begin{bmatrix} 1 & \beta & \alpha & \alpha\beta \\ \beta & 1 & \alpha\beta & \alpha \\ \alpha & \alpha\beta & 1 & \beta \\ \alpha\beta & \alpha & \beta & 1 \end{bmatrix}$$

The Cholesky decomposition of the correlation matrix gives:

$$A_{\text{MIMO}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \beta & \sqrt{1-\beta^2} & 0 & 0 \\ \alpha & 0 & \sqrt{1-\alpha^2} & \beta \\ \alpha\beta & \alpha\sqrt{1-\beta^2} & \beta\sqrt{1-\alpha^2} & \sqrt{1+\alpha^2\beta^2-\alpha^2-\beta^2} \end{bmatrix}$$

The correlated elements of the channel matrix are then generated as follows:

$$\begin{bmatrix} s_{11} \\ s_{12} \\ s_{21} \\ s_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \beta & \sqrt{1-\beta^2} & 0 & 0 \\ \alpha & 0 & \sqrt{1-\alpha^2} & \beta \\ \alpha\beta & \alpha\sqrt{1-\beta^2} & \beta\sqrt{1-\alpha^2} & \sqrt{1+\alpha^2\beta^2-\alpha^2-\beta^2} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{bmatrix}$$

x_{ik} are complex independent random variables:

$$x_{ik} = a_{ik} + jb_{ik},$$

where a_{ik} and b_{ik} are generated from a normal distribution $N(0, 1/\sqrt{2})$.
The MIMO correlation matrix coefficients has been defined in 3GPP, according to three different correlation types:

- Low correlation: $\alpha = 0$, $\beta = 0$
- Medium correlation: $\alpha = 0.3$, $\beta = 0.9$
- High correlation: $\alpha = 0.9$, $\beta = 0.9$

We consider in this simulator the medium correlation type. Time correlation is also taken into account. Thus 4 spatially correlated fading complex values are generated and updated each 1ms for each link between a sector and user.

LTE Codebook-based precoding

We assume that a single layer data stream is transmitted by two antennas, that is the case of transmit diversity LTE transmision mode. The codebook as defined in 3GPP in this case, is given by Figure 7. The precoding at the transmitter is done with the selected codebook V that maximizes the received power:

$$\arg\max_i |HV_i|^2$$

Codebook index	Precoding vector
0	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$
1	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$
2	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\j\end{bmatrix}$
3	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-j\end{bmatrix}$

Figure 7: Codebook for transmit diversity.

Traffic model

We consider a dynamic setting where users arrive over time and quit after being served. They arrive according to a Poisson process with uniform spatial distribution. Their Traffic consists of file transfers only. File sizes are generated from an exponential distribution with mean 4 Mbits, that is 1.25 MBytes. In order to evaluate the capacity of the network, we do not consider any admission control.

Acronyms

GSM Global System for Mobile Communications.

QoS Quality of Service.

EDGE Enhanced Datarates for GSM Evolution.

GPRS General Packet Radio Service.

UMTS Universal Mobile Telecommunications System.

HSPA High-Speed Packet Access.

3GPP Third Generation Partnership Project.

LTE Long-Term Evolution.

WCDMA Wideband Code Division Multiple Access.

OFDM Orthogonal Frequency Division Multiplexing.

OFDMA Orthogonal Frequency Division Multiple Access.

MIMO Multiple Input Multiple Output.

MU-MIMO multi-user MIMO.

FDD Frequency Division Duplex.

TDD Time Division Duplex.

UE User Equipment.

E-UTRAN Evolved UMTS Terrestrial Radio Access Network.

EPC Evolved Packet Core.

eNodeB Evolved Node B.

RRM Radio resource management.

RB Resource Block.

PRB Physical Resource Block.

TTI Transmission Time Interval.

MCS Modulation and Coding Scheme.

RE Resource Element.

CQI Channel Quality Indicator.

BLER Block Error Rate.

CoMP Coordinated Multipoint Transmission.

JT Joint Transmission.

JP Joint Processing.

DPB Dynamic Point Blanking.

DPS Dynamic Point Selection.

CS/CB Coordinated Scheduling/Coordinated Beamforming.

CSI Channel State nformation.

LTE-A LTE-Advanced.

SISO single input single output.

eICIC Enhanced Inter-Cell Interference Coordination.

CoMP Coordinated Multipoint .

SINR Signal-to-Interference-plus-Noise Ratio.

PDSCH Physical Downlink shared Channel.

BS Base Station.

SON Self Organizing Network.

Bibliography

[1] ISO/IEC DIS 23009-1. Information Technology - Dynamic Adaptive Streaming Over HTTP (DASH) - part 1: Media Presentation Description and Segment Formats. Technical report, 2011.

[2] 3GPP R1-083569. Further Discussion on Inter-Cell Interference Mitigation Through Limited Coordination . 2008.

[3] 3GPP R1-090613. Discussion on CoMP-SU-MIMO. 2009.

[4] 3GPP R1-091415. Further discussion of frequency plan scheme on CoMP-SU-MIMO. 2009.

[5] 3GPP TR 25.872. Technical specification group radio access network; high speed packet access (HSDPA) multipoint transmission . 2011.

[6] 3GPP TR 36.819. Coordinated Multi-point Operation for LTE Physical Layer Aspects. 2011-2012.

[7] 3GPP TR 36.913. Requirements for Further Advancements for E-UTRA (LTE-Advanced).

[8] 3GPP TS 36.101. Evolved Universal Terrestrial Radio Access, User Equipment radio transmission and reception.

[9] Samuli Aalto, Aleksi Penttinen, Pasi Lassila, and Prajwal Osti. On the optimal trade-off between srpt and opportunistic scheduling. In *Proceedings of the ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '11, pages 185–196, New York, NY, USA, 2011. ACM.

[10] N. Abbas, T. Bonald, and B. Sayrac. Opportunistic gains of mobility in cellular data networks. In *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2015 13th International Symposium on*, pages 315–322, May 2015.

[11] Nivine Abbas, Thomas Bonald, and Berna Sayrac. Opportunistic gains of mobility in cellular data networks. In *IEEE WiOpt, Mumbai India*, 2015.

[12] B.W. Abeysundara and A.E. Kamal. High-Speed Local Area Networks and their Performance. In *ACM Computing Surveys*.

[13] Eitan Altman, Chadi Barakat, Emmanuel Laborde, Patrick Brown, and Denis Collange. Fairness analysis of TCP/IP. In *Decision and Control, 2000. Proceedings of the 39th IEEE Conference on*, volume 1, pages 61–66. IEEE, 2000.

[14] K. Ashim, A. Koichi, S. Sumei, Z. Huiling, and W. Jiangzhou. Cooperative transmission strategy for downlink distributed antenna systems over time-varying channel. In *in Proc. IEEE GLOBECOM*, page 1798–1803, Dec. 2015.

[15] Urtzi Ayesta, Martin Erausquin, and Peter Jacko. A modeling framework for optimizing the flow-level scheduling with time-varying channels. *Perform. Eval.*, 67(11):1014–1029, 2010.

[16] Paolo Baracca, Federico Boccardi, and Nevio Benvenuto. A dynamic clustering algorithm for downlink comp systems with multiple antenna ues. *EURASIP J. Wireless Comm. and Networking*, 2014:125, 2014.

[17] Thorsten Biermann, Luca Scalia, Changsoon Choi, Holger Karl, and Wolfgang Kellerer. Comp clustering and backhaul limitations in cooperative cellular mobile access networks. *Pervasive and Mobile Computing*, 8(5):662–681, 2012.

[18] Thorsten Biermann, Luca Scalia, Changsoon Choi, Wolfgang Kellerer, and Holger Karl. How backhaul networks influence the feasibility of coordinated multipoint in cellular networks [accepted from open call]. *IEEE Communications Magazine*, 51(8), 2013.

[19] T. Bonald. A score-based opportunistic scheduler for fading radio channels. In *Proceedings of European Wireless*, 2004.

[20] T. Bonald and M. Feuillet. *Network Performance Analysis*. ISTE/Wiley, 2011.

[21] Thomas Bonald, Sem C. Borst, Nidhi Hegde, Matthieu Jonckheere, and Alexandre Proutiere. Flow-level performance and capacity of wireless networks with user mobility. *Queueing Syst.*, 63(1-4):131–164, 2009.

[22] Thomas Bonald, Sem C. Borst, and Alexandre Proutière. How mobility impacts the flow-level performance of wireless data systems. In *Proceedings IEEE INFOCOM 2004, Hong Kong, China, March 7-11*, 2004.

[23] Thomas Bonald, Salah-Eddine Elayoubi, and Yu-Ting Lin. A flow-level performance model for mobile networks carrying adaptive streaming traffic. In *IEEE Globecom*, Dec 2015.

[24] Thomas Bonald and Alexandre Proutière. On performance bounds for the integration of elastic and adaptive streaming flows. In *Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '04/Performance '04, pages 235–245, New York, NY, USA, 2004. ACM.

[25] S. Borst and N. Hegde. Integration of streaming and elastic traffic in wireless networks. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pages 1884–1892, May 2007.

[26] Sem C. Borst, Nidhi Hegde, and Alexandre Proutiere. Mobility-driven scheduling in wireless networks. In *INFOCOM. 19-25 April 2009, Rio de Janeiro, Brazil*, pages 1260–1268, 2009.

[27] Sem C. Borst, Alexandre Proutière, and Nidhi Hegde. Capacity of wireless data networks with intra- and inter-cell mobility. In *INFOCOM. 23-29 April 2006, Barcelona, Catalunya, Spain*, 2006.

[28] Stefan Brueck, Lu Zhao, Jochen Giese, and M. Awais Amin. Centralized Scheduling for Joint Transmission Coordinated Multi-Point in LTE-Advanced. In *Proc. ITG/IEEE Workshop on Smart Antennas, 23-24 Feb 2010, Bremen, Germany*, 2010.

[29] G. Chandrasekaran, T. Vu, A. Varshavsky, M. Gruteser, R.P. Martin, J. Yang, and Y. Chen. Vehicular speed estimation using received signal strength from mobile phones. In *Proc. of the 12th ACM international conference on Ubiquitous computing, Ubicomp*, volume 10, 2010.

[30] E.F. Chaponniere, P.J. Black, J.M. Holtzman, and D.N.C. Tse. Transmitter directed code division multiple access system using path diversity to equitably maximize throughput, September 10 2002. US Patent 6,449,490.

[31] Hayssam Dahrouj and Wei Yu. Coordinated beamforming for the multicell multi-antenna wireless system. *IEEE Trans. Wireless Communications*, 9(5):1748–1759, 2010.

[32] Minghai Feng, Xiaoming She, Lan Chen, and Yoshihisa Kishiyama. Enhanced dynamic cell selection with muting scheme for DL comp in LTE-A. In *Proceedings of the 71st IEEE Vehicular Technology Conference, VTC Spring 2010, 16-19 May 2010, Taipei, Taiwan*, pages 1–5, 2010.

[33] Anders Gjendemsjø, David Gesbert, Geir E. Øien, and Saad G. Kiani. Binary power control for sum rate maximization over multiple interfering links. *IEEE Trans. Wireless Communications*, 7(8):3164–3173, 2008.

[34] Majed Haddad, Dalia-Georgiana Herculea, Eitan Altman, Nidham Ben Rached, Veronique Capdevielle, Chung Shue Chen, and Frédéric Ratovelomanana. Mobility State Estimation in LTE. In *IEEE Wireless Communications and Networking Conference*, Doha, Qatar, June 2016. IEEE.

[35] Dalia-Georgiana Herculea, Majed Haddad, Veronique Capdevielle, and Chung Shue Chen. Network-based UE mobility estimation in mobile networks. In *ACM MobiCom, Poster Paper*, Paris, France, Sep 2015.

[36] Fan Huang, Yafeng Wang, Jian Geng, Mei Wu, and Dacheng Yang. Clustering approach in coordinated multi-point transmission/reception system. In *Proceedings of the 72nd IEEE Vehicular Technology Conference, VTC Fall 2010, 6-9 September 2010, Ottawa, Canada*, pages 1–5, 2010.

[37] Ralf Irmer, Heinz Droste, Patrick Marsch, Michael Grieger, Gerhard Fettweis, Stefan Brueck, Hans-Peter Mayer, Lars Thiele, and Volker Jungnickel. Coordinated multipoint: Concepts, performance, and field trial results. *IEEE Communications Magazine*, 49(2):102–111, 2011.

[38] V. Jungnickel and al. Coordinated multipoint trials in the downlink. In *EEE Broadband Wireless Access Workshop, co-located with GLOBECOM 09*, 2009.

[39] Mohamed Kadhem Karray. User's mobility effect on the performance of wireless cellular networks serving elastic traffic. *Wireless Networks*, 17(1):247–262, 2011.

[40] Veeraruna Kavitha, Sreenath Ramanath, and Eitan Altman. Spatial queueing for analysis, design and dimensioning of picocell networks with mobile users. *Performance Evaluation*, 68(8):710–727, 2011.

[41] F. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. In *Journal of the Operational Research Society*, volume 49, 1998.

[42] Jean-Philippe Kermoal, Laurent Schumacher, Klaus I. Pedersen, Preben E. Mogensen, and Frank Frederiksen. A stochastic MIMO radio channel model with experimental validation. *IEEE Journal on Selected Areas in Communications*, 20(6):1211–1226, 2002.

[43] Ahlem Khlass, Thomas Bonald, and Salah-Eddine Elayoubi. Flow-level performance of intra-site coordination in cellular networks. In *IEEE WiOpt, Tsukuba Japan, May 2013*, pages 216–223, 2013.

[44] Ahlem Khlass, Thomas Bonald, and Salah-Eddine Elayoubi. Flow-level modeling of multi-user beamforming in mobile networks. In *12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, WiOpt 2014, Hammamet, Tunisia, May 12-16, 2014*, pages 70–77, 2014.

[45] Jeongsim Kim, Bara Kim, Jerim Kim, and Yun Han Bae. Stability of flow-level scheduling with markovian time-varying channels. *Perform. Eval.*, 70(2):148–159, 2013.

[46] Leonard Kleinrock. *Queueing Systems*, volume I: Theory. Wiley Interscience, 1975.

[47] Troels E Kolding, Klaus Ingemann Pedersen, Jeroen Wigard, Frank Frederiksen, and P Elgaard Mogensen. High speed downlink packet access: Wcdma evolution. *IEEE Vehicular Technology Society News*, 50(1):4–10, 2003.

[48] M. Peng L. Song and Y. Li. Resource allocation for vertical sectorization in lte-advanced systems. 2013.

[49] Daewon Lee, Hanbyul Seo, Bruno Clerckx, Eric Hardouin, David Mazzarese, Satoshi Nagata, and Krishna Sayana. Coordinated multipoint transmission and reception in lte-advanced: deployment scenarios and operational challenges. *IEEE Communications Magazine*, 50(2):148–155, 2012.

[50] Juho Lee, Younsun Kim, Hyojin Lee, Boon Loong Ng, David Mazzarese, Jianghua Liu, Weimin Xiao, and Yongxing Zhou. Coordinated multipoint transmission and reception in lte-advanced systems. *IEEE Communications Magazine*, 50(11):44–50, 2012.

[51] Geoffrey Ye Li, Jinping Niu, Daewon Lee, Jiancun Fan, and Yusun Fu. Multi-cell coordinated scheduling and MIMO in LTE. *IEEE Communications Surveys and Tutorials*, 16(2):761–775, 2014.

[52] Mingju Li, Xiang Yun, Satoshi Nagata, and Lan Chen. Power allocation of dynamic point blanking for downlink comp in lte-advanced. In *International Conference on Wireless Communications and Signal Processing, WCSP 2013, Hangzhou, China, October 24-26, 2013*, pages 1–5, 2013.

[53] Jing Liu, Yongyu Chang, Qun Pan, Xin Zhang, and Dacheng Yang. A novel transmission scheme and scheduling algorithm for CoMP-SU-MIMO in LTE-A system. In *Proceedings of the 71st IEEE VTC , 16-19 May 2010, Taipei, Taiwan*, pages 1–5, 2010.

[54] Ya-Feng Liu, Yu-Hong Dai, and Zhi-Quan Luo. Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms. *IEEE Trans. Signal Processing*, 59(3):1142–1157, 2011.

[55] Helka-Liina Määttänen, Kari Hämäläinen, Juha Venäläinen, Karol Schober, Mihai Enescu, and Mikko Valkama. System-level performance of lte-advanced with joint transmission and dynamic point selection schemes. *EURASIP J. Adv. Sig. Proc.*, 2012:247, 2012.

[56] Patrick Marsch and Gerhard Fettweis. Static clustering for cooperative multi-point (comp) in mobile communications. In *Proceedings of IEEE International Conference on Communications, ICC 2011, Kyoto, Japan, 5-9 June, 2011*, pages 1–6, 2011.

[57] Zoltan Mayer, Jingya Li, Agisilaos Papadogiannis, and Tommy Svensson. On the impact of control channel reliability on coordinated multi-point transmission. *EURASIP J. Wireless Comm. and Networking*, 2014:28, 2014.

[58] Bishwarup Mondal, Eugene Visotsky, Timothy A. Thomas, Xiaoyi Wang, and Amitava Ghosh. Performance of downlink comp in LTE under practical constraints. In *23rd IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2012, Sydney, Australia, September 9-12, 2012*, pages 2049–2054, 2012.

[59] Gregory Morozov, Alexei Davydov, and Ilya Bolotin. Performance evaluation of dynamic point selection comp scheme in heterogeneous networks with FTP traffic model. In *4th International Congress on Ultra Modern Telecommunications and Control Systems, ICUMT 2012, St. Petersburg, Russia, October 3-5, 2012*, pages 922–926, 2012.

[60] D. Tse P. Viswanath and R. Laroia. *Opportunistic beamforming using dumb antennas*, page 1277–1294. 2002.

[61] J. Peroulas. Method and apparatus for estimating speed of a mobile terminal, September 4 2014. US Patent App. 14/356,155.

[62] D. Piazza and L B. Milstein. Multiuser Diversity-Mobility Tradeoff: Modeling and Performance Analysis of a Proportional Fair Scheduling. In *Proc. IEEE Globecom,vol.1 pp.906-910 Nov. 2002, Taipei, Taiwan*.

[63] Rohde and Schwarz. Lte beamforming measurements. *Application Note 1MA187*, September 2011.

[64] Mamoru Sawahashi, Yoshihisa Kishiyama, Akihito Morimoto, Daisuke Nishikawa, and Motohiro Tanno. Coordinated multipoint transmission/reception techniques for lte-advanced. *IEEE Wireless Commun.*, 17(3):26–34, 2010.

[65] B. Schulz. Lte transmission modes and beamforming. *Rhode and Schwarz White Paper*, 2011.

[66] M Series. Guidelines for evaluation of radio interface technologies for imt-advanced. In *Technical report, ITU-R M.2135-1, ITU*, 2009.

[67] I. Sodagar. The mpeg-dash standard for multimedia streaming over the internet. *MultiMedia, IEEE,* 18(4):62–67, April 2011.

[68] Shaohui Sun, Qiubin Gao, Ying Peng, Yingming Wang, and Lingyang Song. Interference management through comp in 3gpp lte-advanced networks. *IEEE Wireless Commun.*, 20(1):59–66, 2013.

[69] Alireza Tarighat, Mirette Sadek, and Ali H. Sayed. A multi user beamforming scheme for downlink MIMO channels based on maximizing signal-to-leakage ratios. In *2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '05, Philadelphia, Pennsylvania, USA, March 18-23, 2005*, pages 1129–1132, 2005.

[70] David Tse and Pramod Viswanath. Fundamentals of wireless communication. In *Cambridge University Press, New York, NY, USA*, 2005.

[71] David N. C. Tse. Multiuser diversity in wireless networks. In *Wireless Communication seminar, Stanford University April*, 2001.

[72] V. Vukadinovic and G. Karlsson. Video streaming performance under proportional fair scheduling. *IEEE Journal on Selected Areas in Communications*, 28(3):399–408, April 2010.

[73] Hui Wang, Xiaofeng Tao, and Ping Zhang. Adaptive modulation for dynamic point selection/-dynamic point blanking. *IEEE Communications Letters*, 19(3):343–346, 2015.

[74] D. Liang Y. Li, X. Ji and Y. Li. Dynamic beamforming for three-dimensional mimo technique in lte-advanced networks. 2013.

[75] Wei Yu, Taesoo Kwon, and Changyong Shin. Multicell coordination via joint scheduling, beamforming, and power spectrum adaptation. *IEEE Trans. Wireless Communications*, 12(7):1–14, 2013.

[76] Yuan Yuan, Ying Wang, Weidong Zhang, and Fei Peng. Separate horizontal and vertical codebook based 3d MIMO beamforming scheme in LTE-A networks. In *Proceedings of the 78th IEEE Vehicular Technology Conference, VTC Fall 2013, Las Vegas, NV, USA, September 2-5, 2013*, pages 1–5, 2013.

[77] Wenan Zhou, Yiyu Zhang, Pei Qin, Wei Chen, and Xu Li. Joint scheduling algorithms for LTE-A comp system. *JCP*, 8(11):2795–2801, 2013.

Conception et performance de schémas de coordination dans les réseaux cellulaires

Nivine Abbas

RESUME :

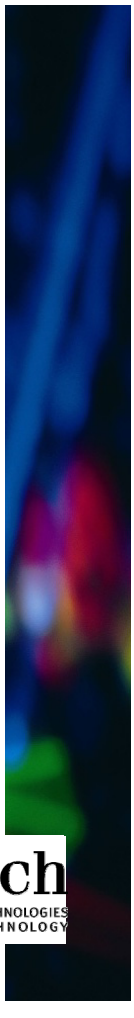
L'interférence entre stations de base est considérée comme le principal facteur limitant les performances des réseaux cellulaires. Nous nous intéressons aux différents schémas de coordination multi-point (CoMP) proposés dans la norme LTE-A pour y faire face, en tenant compte de l'aspect dynamique du trafic et de la mobilité des utilisateurs. Les résultats sont obtenus par l'analyse mathématique de modèles markoviens et par des simulations du système. Nous montrons l'importance de l'algorithme d'ordonnancement sur les performances en présence d'utilisateurs mobiles, pour des services de téléchargement de fichier et de streaming vidéo. Nous proposons un nouvel algorithme d'ordonnancement basé sur la dé-priorisation des utilisateurs mobiles se trouvant en bord de cellule, afin d'améliorer l'efficacité globale du système. Nous montrons ensuite qu'il est intéressant d'activer la technique dite Joint Processing uniquement dans un réseau à forte interférence, son activation dans un réseau à faible interférence pouvant conduire à une dégradation des performances. Nous proposons un nouveau mécanisme de coordination où une cellule ne coopère que lorsque sa coopération apporte un gain moyen de débit suffisant pour compenser les pertes de ressources engendrées. Nous considérons enfin la technique de formation de faisceaux coordonnée. Nous montrons notamment que la coordination n'est pas nécessaire lorsque l'on dispose d'un grand nombre d'antennes par station de base, un simple mécanisme d'ordonnancement opportuniste permettant d'obtenir des performances optimales.

MOTS-CLEFS: Réseaux cellulaires, mobilité, ordonnancement opportuniste, coordination multi-point CoMP, technique de formation de faisceaux, modélisation niveau flot du trafic, théorie des files d'attente, simulations systèmes.

ABSTRACT:

Interference is still the main limiting factor in cellular networks. We focus on the different coordinated multi-point schemes (CoMP) proposed in the LTE-A standard to cope with interference, taking into account the dynamic aspect of traffic and users' mobility. The results are obtained by the analysis of Markov models and system-level simulations. We show the important impact of the scheduling strategy on the network performance in the presence of mobile users considering elastic traffic and video streaming. We propose a new scheduler that deprioritizes mobile users at the cell edge, in order to improve the overall system efficiency. We show that it is interesting to activate Joint Processing technique only in a high-interference network, its activation in a low-interference network may lead to performance degradation. We propose a new coordination mechanism, where a cell cooperates only when its cooperation brings a sufficient mean throughput gain, which compensates the extra resource consumption. Finally, we show that the coordination of beams is not necessary when a large number of antennas is deployed at each base station; a simple opportunistic scheduling strategy provides optimal performance.

KEY-WORDS: Cellular data networks, mobility, opportunistic scheduling, coordinated multipoint CoMP, beamforming, flow-level modeling, queuing theory, system-level simulations.



TELECOM ParisTech

EDITE

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY