Optimisation de la Topologie des Réseaux Sans Fils

Philippe Ezran

N° national de thèse : 2018SACLC014

THESE DE DOCTORAT DE L'UNIVERSITÉ PARIS-SACLAY PREPAREE A CENTRALESUPÉLEC

ECOLE DOCTORALE N°580
Sciences et technologies de l'information et de la communication (STIC)

Spécialité : Réseaux, Information et Communications

Par

M. Philippe Ezran

Optimisation de la Topologie des Réseaux Sans Fils

(Topology Optimization of Wireless Networks)

Thèse présentée et soutenue à Gif-sur-Yvette, le 23 janvier 2018

Composition du jury :

M. Jean-Claude Belfiore,	Télécom ParisTech	Président du Jury
M. Eitan Altman,	INRIA	Rapporteur
M. Jean-Marie Gorce,	INSA-Lyon	Rapporteur
Mme Veronica Belmega,	ENSEA	Examinateur
M. Tijani Chahed,	Télécom Sud Paris	Examinateur
M. Paul Muhlethaler,	INRIA	Examinateur
M. Mérouane Debbah,	Chaire LANEAS - CentraleSupélec	Directeur de Thèse
M. Yoram Haddad,	Jerusalem College of Technology	Co-Directeur de Thèse

This thesis is dedicated to my mother Diane Ezran and my brother Alex Ezran.

Acknowledgments

I would like to express my deep gratitude to:

Mérouane Debbah, for giving me the chance to do this PhD and for his highly valuable advice, directing me towards promising and challenging topics.

Yoram Haddad, for initiating the project, welcoming me at the Jerusalem College of Technology and organizing the scientific, technical and financial support for this PhD.

Jean-Claude Belfiore, Eitan Altman, Jean-Marie Gorce, Veronica Belmega, Tijani Chahed and Paul Muhlethaler, for their participation to my PhD defense and for their constructive and helpful comments.

Hadar Sufiev, Yosef Pogrow, Guy Saadon, Aviram Zilberman, Michael Dreyfuss, Chana Touitou and Chana Klein, for the friendly atmosphere I enjoyed at the Jerusalem College of Technology.

Gil Katz, Loïg Godard, Ejder Baştuğ, Kenza Hamidouche, Anne Batalie and Eliza Dias for welcoming me during my fruitful and motivating visits at CentraleSupélec.

My final thanks go to my family and relatives for their constant support.

Thank you all.

Contents

Acronyms

BS	Base Station. 13-15, 19, 39-40, 45, 62, 112
Capex	Capital expenditure. 2, 21, 28, 46-47, 61
CDMA	Code Division Multiple Access. 11, 37
CFM	Composite Fade Margin. 6, 32
C-RAN	Cloud Radio Access Network. 12, 37-38
CSI	Channel State Information. 13, 38
DSSS	Direct Sequence - Spread Spectrum. 11, 37
EGC	Equal Gain Combining. 13, 39
FDD	Frequency division duplex. 9, 35
GSM	Global System for Mobile Communications. 3, 28
HetNet	Heterogeneous Network. 3, 12, 14, 29, 38-39
HP	Horizontal Polarization. 23, 49, 73, 89, 91
HRAN	High Radio Access Network. 10, 36
IDATE	Institut de l'audiovisuel et des télécommunications en Europe. 93
ILP	Integer Linear Programming. 69-70
ITU	International Telecommunication Union. 3, 17, 29, 42
LOS	line-of-sight. 4, 15, 23-24, 30, 42, 49, 73-75, 84
LRAN	Low Radio Access Network. 10, 36
LTE	Long-Term Evolution. 3, 21, 29, 47
MFSK	Multiple frequency-shift keying. 22, 48
MIMO	multiple input, multiple output. 3, 12-13, 15, 22-24, 29, 38-39, 41, 47-49, 73-74
MPWSP	maximum production $[0, 1)$ weighted k-set packing. 68
MQAM	M-ary Quadrature Amplitude Modulation. 22, 48
MRC	Maximal Ratio Combining. 13, 39
MSP	maximum k-set packing. 68
NFV	Network Functions Virtualization. 2, 28
NOS	Network Operating System. 28
NP	Non-deterministic polynomial-time. 20, 23-24, 46, 49-50, 68-70, 72, 115
OFDM	Orthogonal frequency-division multiplexing. 11, 37
OFDMA	Orthogonal frequency-division multiple access. 22, 48
Opex	Operating expenses. 2, 21, 28, 47
QAM	Quadrature Amplitude Modulation. 24, 49
QoS	Quality of Service. 19, 22, 27, 45, 48

SC	small cell. 14-15, 40
SDN	Software Defined Networking. 1-2, 19, 27-28, 45
SIMO	single input, multiple output. 22, 48
SINR	signal-to-interference-plus-noise ratio. 13-14, 19, 39-40, 45, 97, 113
SIR	signal-to-interference ratio. 20, 45
SLA	Service Level Agreement. 27
TCO	Total Cost of Ownership. 20, 46
TDD	Time division duplex. 9, 13, 15, 35, 38, 41
TDMA	Time Division Multiple Access. 12, 37
UMTS	Universal Mobile Telecommunications System. 3, 29
VP	Vertical Polarization. 23, 49, 73, 89, 91
XPIC	Cross-Polarization Interference Cancellation. 75, 84

List of Figures

List of Tables

Chapter 1

Resumé (French)

Les besoins des utilisateurs sont actuellement en constante augmentation en raison des nouvelles applications, services et usages. Ainsi, les services cloud actuels permettent aux utilisateurs de sauvegarder régulièrement leurs documents numériques dans des centres de données distants, ce qui peut provoquer une très haute consommation de bande passante sur une liaison sans fil. Plus généralement, la liaison sans fil est aujourd'hui de plus en plus utilisée, et les utilisateurs ressentent déjà des limitations de débit et des indisponibilités de service lorsqu'ils accèdent, par exemple, à leur réseau cellulaire. Alors que l'occupation du spectre est déjà élevée, il s'avère que nous approchons des limites des capacités des liaisons sans fils pour l'utilisateur. Cependant, de plus en plus de terminaux, comme les smartphones et les tablettes, sont équipés uniquement d'interface sans fils et les opérateurs doivent répondre à un besoin croissant en bande passante. Cela plaide en faveur d'un changement de paradigme dans la gestion des réseaux. Un réseau devrait pouvoir augmenter ou diminuer sa capacité sur demande, et dans la mesure du possible, fournir aux utilisateurs la qualité de service demandée et éviter de perturber les réseaux voisins. Cela requiert une entité de gestion de la bande passante capable d'examiner la demande et les accords de niveau de service de l'utilisateur, l'état des différents points d'accès et du canal et de décider en conséquence. Cette entité de gestion devra construire sa vision du réseau en recueillant les données de divers équipements de réseaux et réaliser une optimisation globale pour finalement proposer des options de connexion à chaque utilisateur pour chaque application. De plus, alors que nous nous approchons de la capacité limite de Shannon, la solution à un trafic mobile toujours croissant repose sur des cellules plus petites qui offriront aux utilisateurs un signal plus puissant et donc des débits plus élevés. Cependant, la gestion d'une multitude de petites cellules s'avère fort complexe si chacune d'entre elles doit se voir allouer un canal faiblement chargé, spécialement dans l'environnement actuel où les canaux radio deviennent une ressource rare et onéreuse. Il est nécessaire de gérer d'une nouvelle façon les appareils générant des interférences. Le nouveau paradigme qui doit répondre aux défis susmentionnés est dénommé réseau à définition logicielle ou Software Defined Networking (SDN). Il consiste à découpler le plan de contrôle du plan de données [2], [3], [4]. De nombreux travaux sont en cours dans ce domaine.

En parallèle, un sujet quelque peu négligé dans le domaine des réseaux cellulaires est la "liaison backhaul", qui est la liaison entre le réseau radio d'accès et le réseau central. Jusqu'à présent, il n'y avait pratiquement pas d'ingénierie dans le déploiement de ces liaisons car on supposait une topologie point à point. Aujourd'hui, avec la multiplication des petites cellules comme mentionné plus haut, une telle topologie ne sera plus possible. Les liaisons point-multipoint, la réutilisation des fréquences et le contrôle de puissance seront nécessaire dans le backhaul. Mais entretemps, le SDN permettra une vue centralisée de l'état du réseau.

Le SDN est défini dans [5] comme une architecture de réseau basée sur quatre principes:

- Les plans de contrôle et de données sont découplés.
- Les décisions relatives au transmission de données sont basés sur les flux et non pas sur la destination: tous les paquets d'un même flux reçoivent le même service des dispositifs de transmission.
- La logique de contrôle est déplacée vers une entité externe (Système d'exploitation réseau).
- Le réseau est programmable par des applications logicielles fonctionnant sur le système d'exploitation réseau qui interagit avec les dispositifs du plan de données.

Le SDN, associé à la virtualisation des fonctions de réseau ou Network Functions Virtualization (NFV), offre la possibilité de réduire le capex (coûts de matériel plus bas, équipements de réseau multi-fonctions , implémentation multi-utilisateurs) et l'opex (cycles de développement et de test plus courts, meilleure efficacité opérationnelle, consommation d'énergie plus basse) et de réduire le délai de commercialisation des nouveaux services (promotion de l'innovation, déploiement de service plus rapide, service ciblé par zone géographique) [6].

Parmi les objectifs d'intérêt dans ce contexte:

- Allocation de canal efficace: en tenant compte du minimum requis pour le trafic en temps réel et d'une allocation aussi flexible que possible.
- Résilience: éviter le déploiement de liaisons 1+1 par mise en commun des stratégies des différents utilisateurs (opérateurs).
- Optimisation de la disponibilité : critère des cinq neuf contre probabilité de panne.
- Allocation de ressource optimale, définie comme compromis entre efficacité du réseau et équité entre les utilisateurs.

1.1 Les réseaux sans fils

1.1.1 Introduction

L'industrie des télécommunications sans fil fait actuellement face à une croissance considérable pour des débits toujours plus hauts, stimulée par le développement des services mobiles de données. Afin de répondre à cette demande tout en conservant une qualité de service constante, les planificateurs de réseaux doivent augmenter les capacités du réseau tout en prenant en compte des considérations d'architecture et de topologie.

Depuis l'introduction de la téléphonie mobile en 1981, une nouvelle génération est apparue environ tous les dix ans: 2G (Global System for Mobile Communication, GSM) en 1991, 3G (Universal Mobile Telecommunication Service, UMTS) en 2001 et 4G (Long-Term Evolution, LTE) en 2012. Sur la base de ces observations, certains suggèrent que la 5G pourrait être introduite vers 2020 [7]. Cependant, il n'y a actuellement aucune spécification officielle sur la 5G et le contenu technologique exact du réseau mobile de prochaine génération fait l'objet d'un grand débat [8].

En raison des ressources limitées d'un réseau sans fils, son optimisation aborde un certain nombre de questions. Les fréquences du spectre radio présentent diverses caractéristiques de propagation et d'évanouissement. Leur efficacité spectrale peut être améliorée par une utilisation adéquate de la polarisation. Les techniques d'accès multiple visent à optimiser l'utilisation du spectre. Trois technologies, le MIMO (multiple input, multiple output) , les HetNets (Heterogeneous Networks) et le full duplex, devraient améliorer les efficacités spectrale et énergétique. Les topologies de réseaux présentent divers avantages et inconvénients en termes de disponibilité, coût et délai. La qualité de service et le coût sont des éléments clés qui doivent être pris en compte dans la planification de réseau. Ces questions sont présentées dans les sections suivantes de ce Chapitre.

1.1.2 Modèles de canaux

Le spectre

Le spectre radioélectrique est la partie du spectre électromagnétique comprise entre 3 Hz et 3 THz. La classification des fréquences de l'International Telecommunication Union (ITU) est donnée Table 1.1 ([9], p 27).

Band name	**Abbreviation**	**Frequency band**	**Wavelength**
Extrêmement basse fréquence	ELF	3 Hz - 30 Hz	100,000 km - 10,000 km
Super basse fréquence	SLF	30 Hz - 300 Hz	10,000 km - 1,000 km
Ultra basse fréquence	ULF	300 Hz - 3 kHz	1,000 km - 100 km
Très basse fréquence	VLF	3 kHz - 30 kHz	100 km - 10 km
Basse fréquence	LF	30 kHz - 300 kHz	10 km - 1 km
Moyenne fréquence	MF	300 kHz - 3 MHz	1 km - 100 m
Haute fréquence	HF	3 MHz - 30 MHz	100 m - 10 m
Très haute fréquence	VHF	30 MHz - 300 MHz	10 m - 1 m
Ultra haute fréquence	UHF	300 MHz - 3 GHz	1 m - 100 mm
Super haute fréquence	SHF	3 GHz - 30 GHz	100 mm - 10 mm
Extrêmement haute fréquence	EHF	30 GHz - 300 GHz	10 mm - 1 mm
Térahertz	THF	300 GHz - 3 THz	1 mm - 100 μm

Table 1.1: Le spectre radioélectrique.

Ces fréquences ont diverses caractéristiques. Les basses fréquences (30 kHz-300 kHz) peuvent être diffractées sur les obstacles et sont adaptées pour les transmissions longue distance. Les hautes fréquences (3 MHz-30 MHz) sont encore adéquates pour les communications intercontinentales grâce à la réflexion des ondes par l'ionosphère. Les fréquences supérieures à 30 MHz sont bloquées par les collines et les montagnes, et les fréquences supérieures à 300 MHz sont bloquées par les grands immeubles. Ces dernières fréquences sont adaptées pour les communications courte distance et permettent la réutilisation de fréquence. Les fréquences supérieures à 300 MHz se propagent en ligne de vue (Line-of-Sight, LOS) et réflexion sur le sol.

Les fréquences de 300 GHz à 300 THz sont absorbées par l'atmosphère terrestre, alors que les fréquences supérieures à 300 THz sont dans le spectre de la lumière visible.

Le développement des communications sans fils rend le spectre disponible de plus en plus rare et nécessite des solutions afin d'optimiser l'usage de ses ressources limitées.

Modèle de Propagation

Un modèle de propagation est proposé dans [10], Chapitre 3.

Affaiblissement en espace libre (Free-space loss) L'affaiblissement en espace libre est l'atténuation d'une onde électromagnétique dans un conduit non obstrué en ligne de vue à travers l'espace libre entre un émetteur et un récepteur. L'affaiblissement en espace libre est proportionnel au carré de la distance entre l'émetteur et le récepteur et proportionnel au carré de la fréquence du signal.

Equation de l'affaiblissement en espace libre:

$$FSPL = \left(\frac{4\pi d}{\lambda}\right)^2 = \left(\frac{4\pi df}{c}\right)^2 \tag{1.1}$$

où d est la distance entre l'émetteur et le récepteur, λ la longueur d'onde du signal, f la fréquence et c la vitesse de la lumière.

Cette equation peut s'écrire:

$$L_{FSL} = 92.45 + 20\log(f) + 20\log(d) \tag{1.2}$$

avec L_{FSL} en dB, f en GHz et d en km.

La dépendance envers la distance résulte de la dispersion de l'énergie électromagnétique dans l'espace libre. La dépendance envers la fréquence résulte du fait que la définition de l'affaiblissement prend en compte une antenne réceptrice théorique isotrope. L'ouverture A et le gain G d'une antenne sont liés par l'équation:

$$A = G\frac{\lambda^2}{4\pi} \tag{1.3}$$

Par conséquent, afin de maintenir le gain à 1, une plus haute fréquence nécessite une plus petite ouverture, et donc, une puissance reçue plus faible.

Atténuation Outre l'affaiblissement en espace libre, d'autres facteurs contribuent à l'atténuation du signal: la végétation, l'absorption atmosphérique, la pluie et les obstacles. L'atténuation est une fonction croissante de la fréquence et de la distance.

Un modèle pour l'atténuation par la végétation a été proposé par Weissberger [11].

Les principaux phénomènes d'absorption sont l'absorption par la pluie au-dessus de 10 GHz (l'absorption par les molécules d'eau a un maximum à 23 GHz) et l'absorption par l'oxygène dans la bande des 60 GHz [10], [7].

L'atténuation des ondes électromagnétiques en fonction de la fréquence est donnée Figure 1.1 [12].

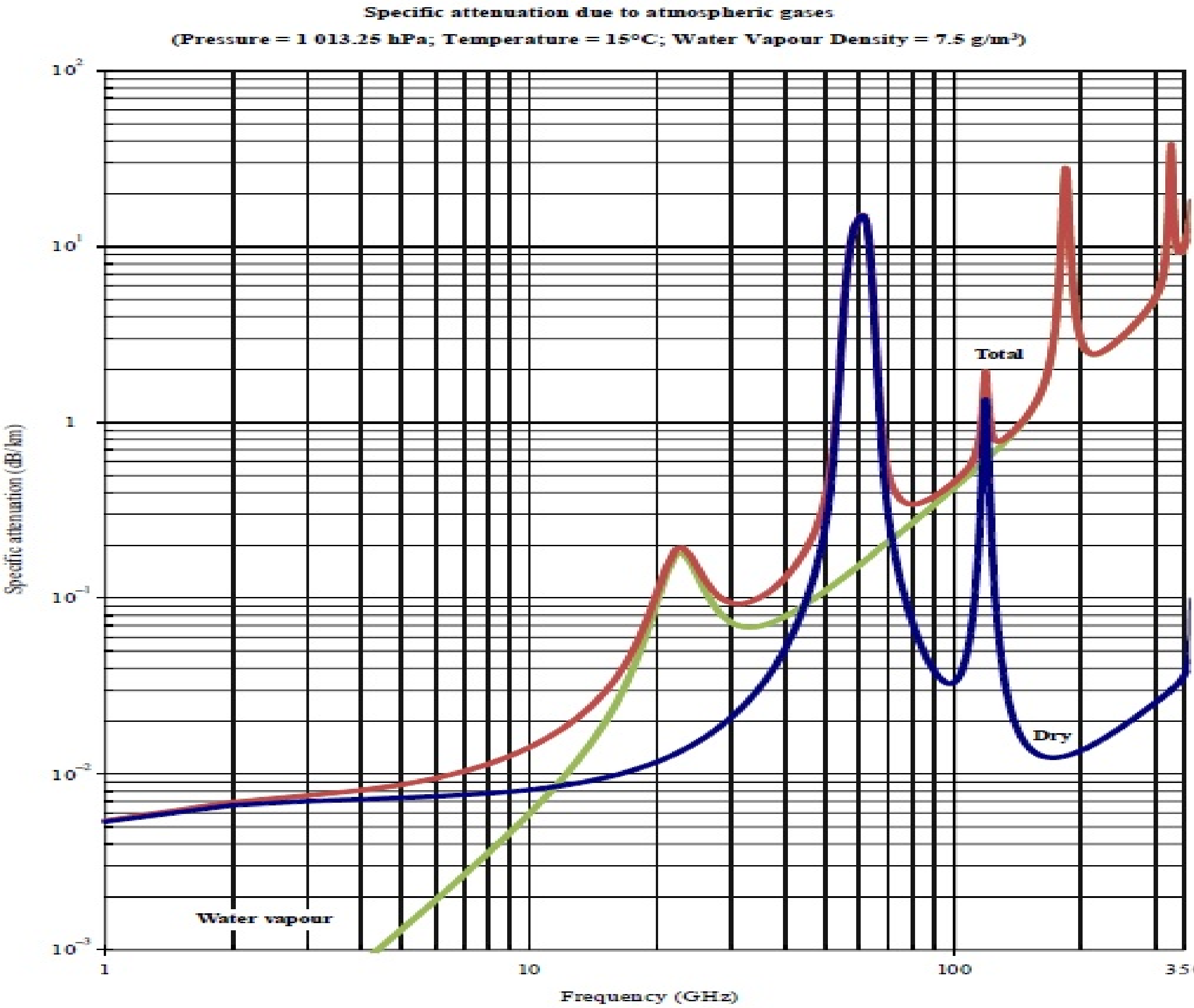

Figure 1.1: Absorption atmosphérique et moléculaire.

Évanouissement (Fading) L'évanouissement est l'atténuation d'un signal résultant des changements atmosphériques ou des réflexions dans le trajet de propagation. Le modèle le plus utilisé pour le fading est le modèle de Vigants, qui donne la probabilité de fading en fonction d'un facteur géographique, de la fréquence, de la distance et de la marge mixte d'évanouissement (Composite Fade Margin) [10], [13]:

$$P = 6.10^{-10} C f d^3 10^{-\frac{CFM}{10}} \tag{1.4}$$

avec

P: probabilité d'évanouissement comme fraction du temps

d: distance (km)

f: fréquence (MHz)

C: facteur lié aux conditions de propagation

CFM: Composite Fade Margin (dB)

Un plan donnant une indication du facteur C en fonction du lieu est donné dans [14].

Efficacité Spectrale

L'efficacité spectrale théorique d'un canal est donnée par la formule de Shannon-Hartley:

$$C = B \log_2(1 + SNR) \tag{1.5}$$

Comme le montre la Figure 1.2, l'efficacité spectrale des réseaux d'accès radio a été multipliée par environ 30 depuis 2000 [15].

Polarisation

Une onde électromagnétique comporte un champ électrique et un champ magnétique oscillant dans deux directions orthogonales à la direction de propagation et orthogonales l'une à l'autre. Si les champs tournent à la fréquence de l'onde, la polarisation est dite circulaire ou elliptique. Si les champs oscillent dans une seule direction, la polarisation est dite linéaire. Par convention, la direction d'une polarisation linéaire est la direction du champ électrique.

Lorsqu'une onde électromagnétique rencontre une surface séparant deux milieux d'indices de réfraction différents, une partie de son énergie est réfractée et une autre partie est réfléchie. Le plan contenant les ondes incidente, réfléchie et réfractée est appelé le plan d'incidence. Le rayon incident peut être considéré comme la somme de deux composantes:

- rayon de polarisation s, dont le champ électrique est orthogonal au plan d'incidence;
- rayon de polarisation p, dont le champ électrique est parallèle au plan d'incidence.

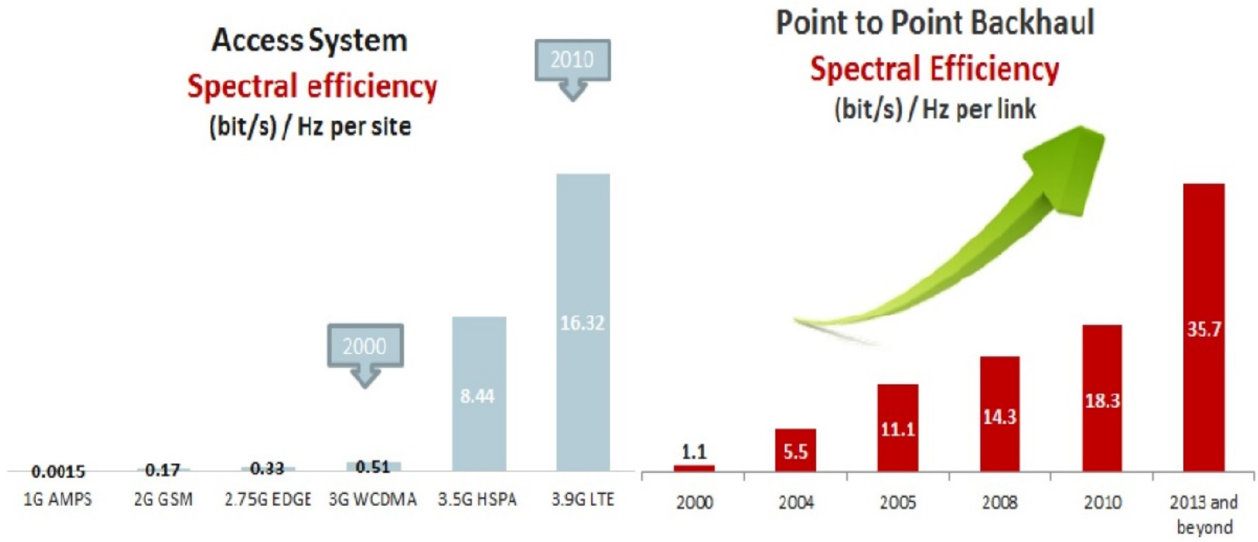

Figure 1.2: Les progrès de l'efficacité spectrale dans l'accès RAN et le backhaul hyperfréquences.

La relation entre l'angle incident θ_i et l'angle réfracté θ_t est donné par la loi de Snell-Descartes (voir Figure 1.3):

$$n_1 \sin \theta_i = n_2 \sin \theta_t \tag{1.6}$$

où n_1 and n_2 sont les indices de réfraction des deux milieux.

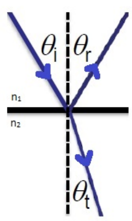

Figure 1.3: Réfraction.

La réflectance, qui est la fraction de la puissance incidente réfléchie par l'interface, est donnée par les équations de Fresnel, pour les rayons de polarisation s et les rayons de polarisation p, respectivement [16]:

$$R_s = \left(\frac{\sqrt{\frac{\mu_2}{\epsilon_2}} \cos\theta_i - \sqrt{\frac{\mu_1}{\epsilon_1}} \cos\theta_t}{\sqrt{\frac{\mu_2}{\epsilon_2}} \cos\theta_i + \sqrt{\frac{\mu_1}{\epsilon_1}} \cos\theta_t} \right)^2 \tag{1.7}$$

$$R_p = \left(\frac{\sqrt{\frac{\mu_2}{\epsilon_2}} \cos\theta_t - \sqrt{\frac{\mu_1}{\epsilon_1}} \cos\theta_i}{\sqrt{\frac{\mu_2}{\epsilon_2}} \cos\theta_t + \sqrt{\frac{\mu_1}{\epsilon_1}} \cos\theta_i} \right)^2 \tag{1.8}$$

où μ_1, μ_2, ϵ_1 et ϵ_2 sont les perméabilités magnétiques et les permittivités électriques des deux matériaux, respectivement.

Puisque $\sqrt{\frac{\mu_i}{\epsilon_i}} = \mu_i \frac{c}{n_i}$, ces équations peuvent s'écrire, pour des milieux non-magnétiques ($\mu_i = \mu_0$, où μ_0 est la perméabilité de l'espace libre):

$$R_s = \left(\frac{n_1 \cos\theta_i - n_2 \sqrt{1 - \left(\frac{n_1}{n_2} \sin\theta_i\right)^2}}{n_1 \cos\theta_i + n_2 \sqrt{1 - \left(\frac{n_1}{n_2} \sin\theta_i\right)^2}} \right)^2 \tag{1.9}$$

$$R_p = \left(\frac{n_1 \sqrt{1 - \left(\frac{n_1}{n_2} \sin\theta_i\right)^2} - n_2 \cos\theta_i}{n_1 \sqrt{1 - \left(\frac{n_1}{n_2} \sin\theta_i\right)^2} + n_2 \cos\theta_i} \right)^2 \tag{1.10}$$

La transmittance, qui est la fraction de la puissance incidente réfractée par l'interface, se déduit de la loi de conservation de l'énergie :

$$T_s = 1 - R_s \tag{1.11}$$

et

$$T_p = 1 - T_p \tag{1.12}$$

Pour un angle particulier, $R_p = 0$. Cela se produit lorsque $n_1 \cos\theta_t = n_2 \cos\theta_i$. De cette équation et de la loi de Snell-Descartes, on obtient:

$$\sin 2\theta_i = \sin 2\theta_t \tag{1.13}$$

ce qui implique $\theta_t = \frac{\pi}{2} - \theta_i$ et $\tan\theta_i = \frac{n_2}{n_1}$. Cet angle est connu sous le nom d'angle de Brewster.

La figure 1.4 représente R_s et R_p en fonction de l'angle d'incidence pour $n_1 = 1$ et $n_2 = 1.5$ (indice de réfraction du verre). Dans ce cas, l'angle de Brewster vaut environ 56°.

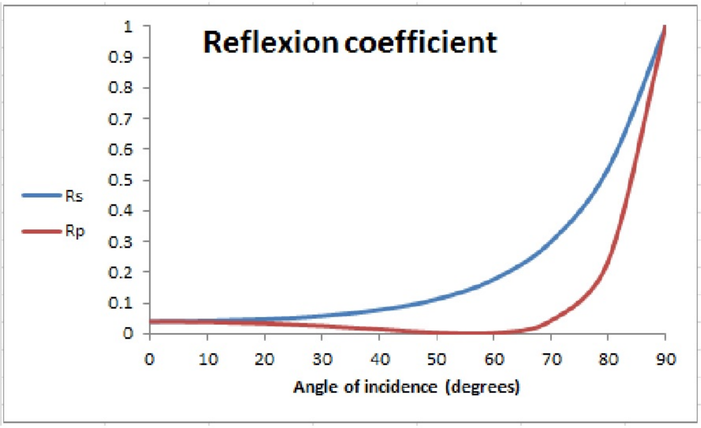

Figure 1.4: Équations de Fresnel.

Une conséquence importante des équations de Fresnel est qu'il n'existe pas de principe de retour inverse pour la polarisation. Si une polarisation $\mathbf{E}_1$ est transformée en $\mathbf{E}_2$ après réflexion, le rayon inverse ayant la polarisation $\mathbf{E}_2$ et se réfléchissant au même point aura en général une polarisation différente de $\mathbf{E}_1$ après réflexion.

La diversité de polarisation permet la réutilisation des fréquences dans un réseau. Dans l'état de l'art actuel, la polarisation est presque exclusivement verticale ou horizontale, bien que les polarisations inclinées à ±45° et les polarisations circulaires soient aussi considérées. Cela est principalement dû à deux raisons:

- Les antennes sont éloignées les unes des autres, et par conséquent, une différence d'altitude peut être négligée par rapport à la distance; tous les nœuds d'un réseau peuvent être considérés comme coplanaires.
- La plupart des réseaux en anneau utilisent le frequency-division duplex (FDD) et par conséquent comportent un nombre pair de nœuds.

Il convient de noter qu'avec le déploiement des small cells, du time division duplex (TDD) et du full duplex, ces raisons ne s'appliqueront plus, et d'autres solutions basée sur la polarisation devront être considérées à l'avenir, notamment dans les réseaux en anneau.

1.1.3 Accès multiple

Réseaux Backhaul

L'environnement backhaul est communément décrit comme étant la partie d'un réseau mobile qui raccorde les stations de base au réseau central. Il est souvent divisé en deux parties (voir Figure 1.5 [1]):

- Low Radio Access Network (LRAN), qui est la partie du backhaul RAN servant d'accès au site cellulaire.
- High Radio Access Network (HRAN), qui est la partie du réseau qui collecte, regroupe, et concentre le trafic du LRAN et le connecte aux contrôleurs radio.

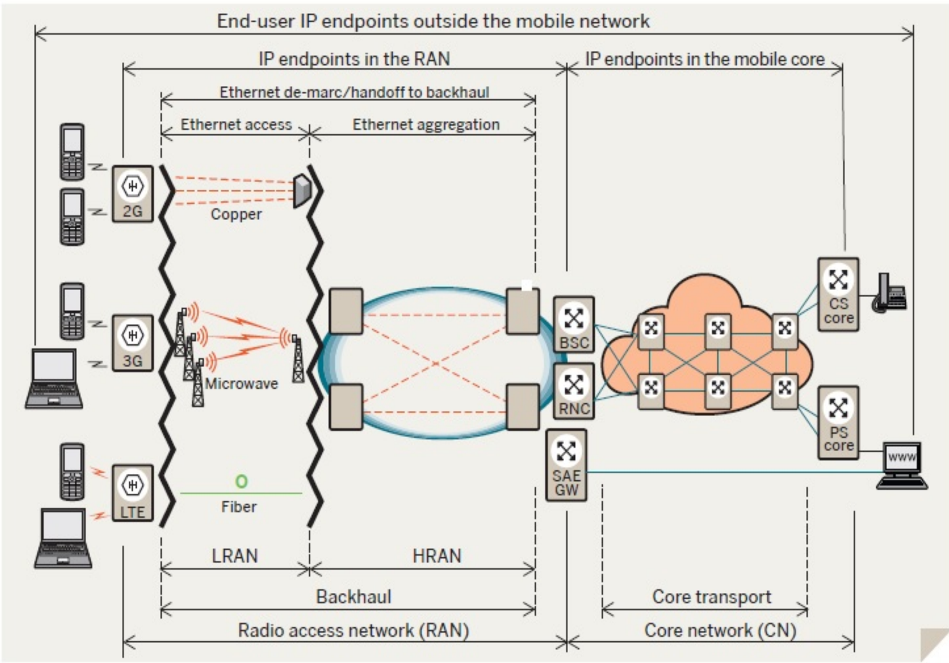

Figure 1.5: Limites du réseau central et du backhaul [1].

Chaque génération de technologie de réseau a ses propres spécifications concernant le backhaul en termes de délai, de synchronisation, de disponibilité, de taux d'erreur et de taux de transmission.

Les avantages et les inconvénients de la fibre par rapport aux hyperfréquences ont été largement étudiés. La fibre a une capacité quasiment illimitée. Cependant, elle est également capitalistique (le coût est fonction de la distance), est souvent fournie par les concurrents (dans le cas de liaisons louées), et n'est pas toujours disponible, en raison de contraintes géographiques ou administratives. D'un autre côté, les hyperfréquences répondent aux besoins présents et futurs en capacité, sont flexibles et très économiques, et peuvent être déployés rapidement. Le consensus général est que les segments du backhaul liés au regroupage ou au réseau central, qui requièrent une capacité élevée, doivent être déployées sur la fibre, alors que les sites d'extrémité et les segments de pré-regroupage peuvent utiliser les hyperfréquences [15].

Les futurs réseaux backhaul devraient utiliser des fréquences de transmission plus élevées (ondes millimétriques) et par conséquent nécessiteront davantage d'antennes par station de base [17].

Étalement du spectre

Dans les techniques à étalement de spectre, un signal généré dans une bande de fréquence déterminée est étalé en fréquence. Les techniques à étalement de spectre ont démarré dans le domaine militaire pour des raisons de sécurité et sont maintenant largement utilisées en raison des avantages qu'elles offrent sur les plans de la résistance aux interférences, au bruit et à l'évanouissement.

Sauts de fréquence Dans la technique des sauts de fréquence, la fréquence porteuse change plusieurs fois pendant la transmission. L'ensemble de la communication est acheminé sur de nombreux canaux, d'après un motif connu de l'émetteur et du receveur. Les signaux sont difficiles à intercepter et fortement résistants au brouillage intentionnel. Cette technique est donc particulièrement adaptée aux applications militaires. Entre autres avantages, les sauts de fréquence répartissent sur l'ensemble du spectre les caractéristiques du signal reçu et réduisent la probabilité de scénario catastrophe résultant du fading [18].

Direct sequence - Spread spectrum (DSSS) Dans le DSSS, le signal transmis est multiplié par un signal à large bande. Le spectre du signal obtenu étant le produit de convolution des deux spectres, sa largeur de bande est légèrement supérieure à celle du signal à large bande. Le DSSS réduit donc significativement la densité spectrale de puissance. Le signal est plus résistant aux attaques malveillantes et au fading [18].

Code Division Multiple Access (CDMA) Le CDMA est utilisé en complément du DSSS. Chaque utilisateur a son propre code d'étalement de spectre. Afin d'éviter les interférences entre les signaux, les codes doivent être orthogonaux deux à deux.

Orthogonal Frequency Division Multiplexing (OFDM)

L'OFDM convertit un train de données à haut débit en plusieurs trains de données à bas débit, transmis sur plusieurs canaux parallèles à bande étroite. Cette technique permet d'augmenter la durée du signal, ce qui peut être nécessaire à cause de la dispersion. L'OFDM offre une meilleure résistance à l'interférence à bande étroite et au fading sélectif en fréquence [10]. Les erreurs résultant de canaux de basse qualité peuvent être corrigées par des codes correcteurs d'erreurs.

L'OFDM est destiné à un seul utilisateur, mais peut être adapté à un schéma multi-utilisateur: dans ce cas, chaque utilisateur se voit allouer plusieurs fréquences non adjacentes, ce qui assure une diversité en fréquence. Cette solution est souvent utilisée avec le Time Division Multiple Access (TDMA) [18].

Cloud Radio Access Network (C-RAN)

Le C-RAN est une nouvelle architecture, destinée à répondre aux défis soulevés par les futurs réseaux cellulaires et en particulier le coût résultant du nombre élevé de stations de base dans les réseaux hétérogènes (Heterogeneous Networks, HetNets). Le traitement en bande de base est centralisé et partagé entre plusieurs sites dans une unité virtuelle [19]. Le C-RAN nécessite un nombre d'unités en bande de base plus faible et réduit la consommation d'énergie.

1.1.4 Technologies

L'augmentation des capacités du réseaux nécessite une densification massive du réseau et des outils de traitement du signal. Il existe trois approches: augmenter le nombre d'antennes sur les sites existants (massive MIMO), les HetNets et le full duplex [20].

Massive MIMO

Un système MIMO point à point avec M_t antennes d'émission et M_r antennes de réception est présenté Figure 1.6 [21].

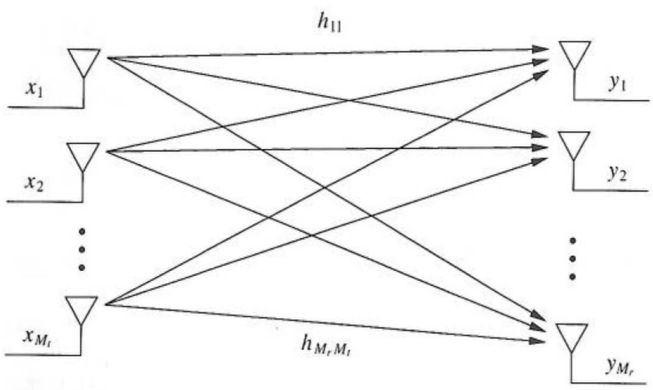

Figure 1.6: Systèmes MIMO.

Le symbole émis $\mathbf{x}$ à M_t dimensions et le symbole reçu $\mathbf{y}$ à M_r dimensions sont liés par l'équation:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1.14}$$

où $\mathbf{H}$ est la matrice $M_r \times M_t$ des gains de canal h_{ij} de l'antenne d'émission j à l'antenne de réception i et $\mathbf{n}$ le vecteur bruit à M_r-dimensions.

La technologie MIMO nécessite une information d'état de canal (channel state information, CSI) au niveau de l'émetteur. La performance d'un système MIMO multi-utilisateurs dépend fortement de la corrélation entre l'état réel du canal et son estimation au niveau de l'émetteur [22]. C'est pourquoi cette technologie est adaptée au TDD où la réciprocité de canal peut être exploitée [23]. Aujourd'hui, dans la plus part des mises en œuvre des systèmes MIMO, chaque station de base (BS) ne fait fonctionner qu'un faible nombre d'antennes (moins de 10) [24].

Les ondes millimétriques constituent la solution la plus pertinente pour le massive MIMO, en raison de la grande bande passante disponible [25] et de la possibilité de placer de nombreuses antennes dans un faible volume [26]

Trois techniques peuvent exploiter la diversité pour le récepteur [27]:

- La diversité de sélection choisit le conduit offrant le meilleur rapport signal sur bruit plus interférence (signal-to-interference-plus-noise ratio, SINR).
- La combinaison à rapport maximal (Maximal Ratio Combining, MRC) prend ses décisions sur la base d'une combinaison linéaire optimale entre les signaux.
- La combinaison à égalité de gain (Equal Gain Combining, EGC) ajoute les signaux après que ceux-ci aient été cophasés.

En massive MIMO, une station de base peut être équipée de centaines ou milliers d'antennes. Une analyse asymptotique montre que les effets du bruit non corrélé et du fast fading disparaissent, le nombre de terminaux par cellule et le débit par terminal sont indépendants de la taille des cellules, l'efficacité spectrale est indépendante de la largeur de bande, et l'énergie nécessaire par bit diminue [28].

Étant donné que le nombre d'antennes excède largement le nombre de terminaux, le massive MIMO offre un grand nombre de degrés de liberté, qui peuvent être utilisés pour réduire la puissance d'émission ou pour pour augmenter la tolérance à de défauts matériels, ce qui contribue à la réduction des coûts dans le déploiement du réseau [29].

Le massive MIMO est donc considéré comme étant une technologie permettant un multiplexage spatial dynamique, améliorant la fiabilité des liaisons et réduisant la puissance émise [30].

Réseaux hétérogènes

Un réseau hétérogène est typiquement composé d'une multitude de technologies d'accès radio, d'architectures, de solutions de transmission, et de stations de base ayant des puissances de transmission variables [31]. Les questions clés dans les réseaux hétérogènes sont le regroupement des stations de base, les small cells et le captage d'énergie.

Regroupement des stations de base L'introduction des HetNets remet en question certaines hypothèses bien établies telles que la modélisation de l'emplacement des BS par un réseau hexagonal uniforme ou l'affectation d'un utilisateur à la BS fournissant le plus fort signal [32]. Divers modèles théoriques, dont la caractérisation de la distribution des SINR, l'association intelligente de cellules et le réutilisation fractionnaire de fréquences ont été étudiées dans [33], montrant que l'ajout d'éléments hétérogènes augmente significativement la performance du réseau. La mobilité de l'utilisateur est généralement modélisée dans le cadre de la géométrie stochastique [34].

Small cells (SC) Les small cells permettent la réutilisation de fréquence en différents sites cellulaires. Elles sont habituellement classifiées d'après leur zone de couverture: les microcells ont des zones de couverture de l'ordre de 2 km, les picocells ont des zones de couverture de l'ordre de 200 m et les femtocells ont des zones de couverture de l'ordre de 10 m.

Dans le contexte du backhaul sans fil, les small cells doivent remplir de nombreuses spécifications: bas coût, petite taille, haute capacité, basse puissance, installation rapide et facile, tolérance au mouvement, fonctionnement en zone précaire, résistance à des changements imprévus et incontrôlés de l'environnement, scalabilité et faible temps de latence [35].

L'introduction des small cells augmente l'hétérogénéité dans les réseaux cellulaires. La question du backhaul est généralement considérée dans le cadre d'un modèle hiérarchique multi-tier. Les performances en termes de délai ont été étudiées dans [36] et [37].

Les small cells ont démontré leur efficacité à améliorer la capacité du réseau sans augmenter la puissance émise. Cependant, elles posent des problèmes pratiques d'interférence entre cellules, d'acquisition de site cellulaire et de déploiement [38].

Principaux avantages des femtocells [39]:

- Meilleure couverture et capacité: grâce aux courtes distances et aux interférences réduites, les femtocells permettent de réduire la puissance émise, d'obtenir un meilleur SINR et de prolonger la durée de vie des batteries.
- Réductions de coûts: le déploiement des femtocells réduira le besoin en macro-BS.
- Réduction des résiliations d'abonnement: la meilleure couverture en intérieur offerte par les femtocells réduira l'insatisfaction des clients.

Captage d'énergie Une BS auto-alimentée est réaliste pour plusieurs raisons [40]:

- Les small cells couvrent des zones très petites demandent donc des puissances d'émission beaucoup plus faibles que celles des macrocellules conventionnelles.
- En raison de la part croissante du trafic en rafale, les charges des BSs varieront fortement dans le temps et dans l'espace. Dans les zones denses, de nombreuses BSs pourront être éteintes la plupart du temps et ne fonctionneront que de façon intermittente en fonction de la demande.
- Les techniques de captage d'énergie deviennent économiques.
- Le backhaul sans fil à haute vitesse devient une réalité pour les small cells, ce qui élimine le besoin d'autres connections filaires.

Full duplex

Lorsqu'un nœud émet et reçoit simultanément à la même fréquence, le signal entrant est dominé par le signal sortant [41]. Afin d'émettre et de recevoir des signaux à la même fréquence et en même temps, le full duplex utilise l'annulation d'interférence. Celle-ci peut être réalisée par deux antennes émettrices placées aux distances d et $d + \frac{\lambda}{2}$ (λ étant la longueur d'onde) de l'antenne réceptrice [42].

Une architecture de réseau hétérogène basée sur le TDD et comprenant des stations de base MIMO et des Small Cells est proposée dans [23] et semble donner des résultats prometteurs en termes de débit par unité de surface.

En mode TDD complet, des intervalles de temps séparés sont alloués pour les communications entre la station de base et les small cells et pour les échanges de données entre les small cells et les équipements utilisateurs. Le mode TDD complet et le full duplexing sont comparés dans [20]. Le mode TDD complet est la solution la plus simple, mais le full duplexing peut obtenir de meilleurs résultats suivant les conditions: d'après les simulations, le full duplexing a les meilleures performances si le nombre de SC est faible et les liaisons sont assez solides.

1.1.5 Topologie

Les principaux paramètres qui entrent en ligne de compte pour la topologie de réseaux sont les conditions de propagation (pluie, ligne de vue) et les considérations d'infrastructure.

Les topologies les plus utilisés sont présentées Figure 1.7.

La topologie maillée fournit une connectivité importante entre les nœuds. Elle est généralement utilisée dans les data centers et les réseaux centraux.

La topologie en étoile requiert des liaisons radio plus longues et le LOS pour chaque liaison, et ne fait que peu de réutilisation de fréquence. Elle est optimale en termes de

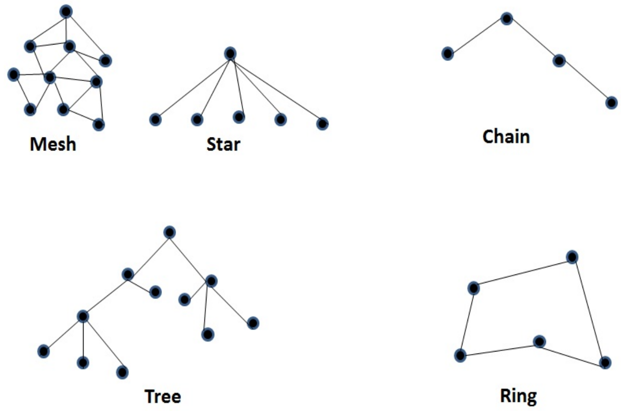

Figure 1.7: Topologies de réseau.

délai. La topologie en étoile est généralement inefficace pour les hyperfréquences, mais peut être utilisée aux fréquences plus basses.

La topologie en chaîne est la plus économique en termes de nombre d'antennes: $2n-2$, n étant le nombre de nœuds. Cependant, elle est vulnérable en cas de panne d'une liaison. Dans la plupart des cas, assurer une redondance sur chaque liaison annulerait l'avantage économique offert par la topologie en chaîne. De plus, la topologie en chaîne a des performances médiocres en termes de délais.

Dans la topologie en arbre, chaque liaison dont la panne affecterait plus d'un site est généralement protégée.

La topologie en anneau peut supporter toute panne unique de liaison. Cependant, ses performances en termes de délai sont médiocres.

Les topologies les plus utilisées sont les arbres et les anneaux, ou une combinaison des deux [10], [43].

1.1.6 Qualité de Service

Disponibilité

L'indisponibilité est généralement définie comme étant la somme des indisponibilités des nœuds du réseau [10]. D'après cette définition, une indisponibilité d'un grand nombre de nœuds dans le réseau a un poids plus grand qu'une indisponibilité d'un petit nombre de nœuds. Cependant, cette définition peut conduire à des indisponibilités supérieures à 100%. Pour cette raison, nous définissons l'indisponibilité comme étant le pourcentage du temps où le réseau est totalement ou partiellement défaillant.

Des définitions des recommandations G.826, G.827, G.828 et G.8201 de l'ITU sont rapportées ci-dessous:

Bloc : un bloc est un ensemble de bits consécutifs associés au conduit, chaque bit appartenant à un bloc et un seul. Des bits consécutifs peuvent ne pas être contigus dans le temps.

Défauts : les trois catégories suivantes de défaut relatives au signal entrant sont définies:

- d1 perte du signal;
- d2 signal d'indication d'alarme;
- d3 perte de verrouillage de trames.

Bloc erroné : bloc dont un ou plusieurs bits sont erronés

Seconde erronée : période d'une seconde comportant un ou plusieurs blocs erronés ou au moins un défaut.

Seconde gravement erronée : période d'une seconde comportant un taux de blocs erronés égal ou supérieur à 30% ou au moins un défaut. L'ensemble des secondes gravement erronées est un sous-ensemble des secondes erronées.

Etat indisponible :

- **Critères applicables à un sens**: une période d'indisponibilité commence au début de dix secondes gravement erronées consécutives. Ces dix secondes font partie du temps d'indisponibilité. Une nouvelle période de disponibilité commence au début de dix secondes consécutives ne comportant pas de seconde gravement erronée. Ces dix secondes font partie du temps de disponibilité.
- **Critère applicable à un conduit ou à une connexion bidirectionnel**: un conduit ou une connexion bidirectionnel est en état d'indisponibilité si un sens ou les deux sont en état d'indisponibilité.

Disponibilité dans diverses topologies de réseaux

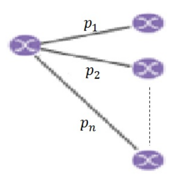

Figure 1.8: Étoile.

Étoile Si les probabilités de panne sont indépendantes, la disponibilité totale est:

$$A = \prod_{i=1}^{n} (1 - p_i) \tag{1.15}$$

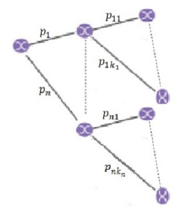

Figure 1.9: Arbre.

Arbre Il résulte de notre définition de la disponibilité qu'un réseau en arbre est disponible si et seulement si toutes ses liaisons sont disponibles.

Si les probabilités de panne sont indépendantes, la disponibilité totale est:

$$A = \prod_{i=1}^{n} (1 - p_{i_1 \ldots i_l}) \tag{1.16}$$

Il convient de noter que la redondance sur une liaison réduit la probabilité de panne, mais ne modifie pas la formule générale ci-dessus.

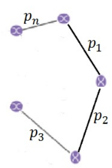

Figure 1.10: Anneau.

Anneau La disponibilité d'un anneau est la probabilité qu'il n'y ait pas plus d'une panne sur cet anneau.

Si les probabilités de panne sont indépendantes, la disponibilité totale est:

$$A = \prod_{i=1}^{n} (1 - p_k) + \sum_{i=1}^{n} p_i \prod_{\substack{1 \leq j \leq n \\ j \neq i}} (1 - p_j) \tag{1.17}$$

Résilience

La résilience est la capacité du réseau à fournir et maintenir un niveau de service acceptable malgré les erreurs et les procédures exceptionnelles.

Alors que la disponibilité d'une connexion à fibre optique est tout ou rien, les techniques de modulation et codage adaptatifs permettent à une liaison hyperfréquence d'utiliser des modulations de bas débit en cas de conditions dégradées.

La topologie en anneau a la résilience la plus élevée: en cas de panne d'une liaison, chaque nœud reste connecté au réseau. Seule une double panne provoque une interruption de service.

La topologie en arbre avec redondance offre une protection contre les pannes d'équipements, mais pas contre les défaillances d'un site et n'offre pas de redondance de conduit.

La SDN sans fil améliore la connectivité de l'utilisateur final et la QoS, permet la planification multi-réseau, améliore la sécurité et permet la localisation de l'utilisateur [3].

Les spécifications concernant la qualité de service peuvent être définies selon plusieurs paramètres [7]:

- Le débit agrégé est la quantité totale de données que le réseau peut transmettre, exprimée en $bit/s/m^2$.
- Le débit limite, ou débit à 5%, est le plus mauvais débit qu'un utilisateur peut raisonnablement recevoir lorsqu'il est à portée du réseau.
- Le débit maximum est le meilleur débit qu'un utilisateur peut espérer obtenir dans des conditions de configuration plausible du réseau.
- La latence est la durée entre la stimulation et la réponse.

Les spécifications dépendent de l'application: la vidéo haute définition nécessite des débits élevés, alors que les applications dans le domaine de la sécurité ou les applications ludiques interactives donnent la priorité à une faible latence.

Affectation à une Station de Base

L'affectation à une Station de Base est une question essentielle dans les réseaux hétérogènes multi-tier, et elle devrait prendre encore plus avec l'introduction des réseaux 5G, où la SDN permettra des stratégies d'affectation élaborées, dont le but sera l'optimisation de la QoS pour l'utilisateur final.

De nombreuses approches proposent une affectation basée sur un seul critère. Dans [44], l'utilisateur se connecte à la BS qui offre le SINR le plus élevé. En connectant l'utilisateur à la BS qui offre la puissance moyenne reçue maximum sur le long terme, où les files d'attente des BS sont toutes pleines, il est prouvé dans [45] que le nombre de niveaux et la densité des BS n'affecte que peu la probabilité de panne et le débit ergodique moyen. En connectant l'utilisateur à la BS qui offre le plus fort signal, il est prouvé dans [33] que la probabilité de couverture est indépendante du nombre de niveaux, des puissances émises, de la densité des BS et de la distribution de fading.

Cependant, l'affectation à une Station de Base ne doit pas être basée uniquement sur la puissance ou le SINR. La charge peut être un paramètre plus important [46].

La symétrie entre les liaisons descendante et montante n'est pas indispensable. L'affectation à une Station de Base avec découplage des deux liaisons a été étudiée dans [47]. Pour une liaison montante sous condition de partage du spectre, si tous les niveaux ont le même rapport signal sur interférence (SIR) cible, l'affectation basée sur la distance est optimale dans de nombreuses configurations d'atténuation et de contrôle de puissance. Pour les deux sens, en cas de partage orthogonal du spectre, l'allocation optimale consiste à partager le spectre entre les niveaux en proportion du nombre d'utilisateurs associés à chaque niveau.

L'optimisation de l'affectation à une Station de Base sous le critère de proportional fairness a été étudiée dans [48] et [49].

L'affectation à une Station de Base conjuguée à l'allocation de ressources a été étudiée dans [50] dans le cadre d'un algorithme distribué basé sur une tarification dynamique et dans [51], où il a été prouvé qu'un protocole efficace et non manipulable (ie l'intérêt de l'utilisateur est de déclarer honnêtement ses informations) pour une combinaison optimale de l'affectation à une Station de Base et des stratégies d'allocation de ressources est nécessairement NP-hard.

1.1.7 Coût

La Figure 1.11 indique le coût total de possession (Total Cost of Ownership, TCO) pour les liaisons à fibre optique et backhaul hyperfréquence. ([Source: Analysys Mason, 2011]).

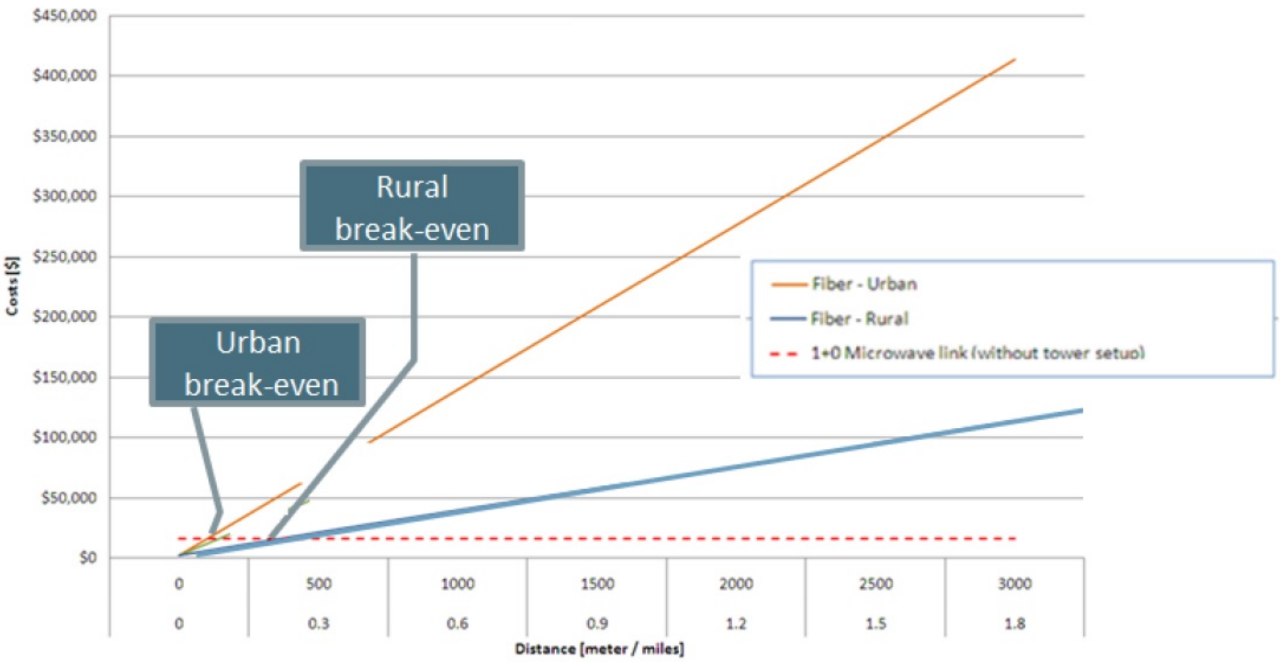

Figure 1.11: TCO des liaisons à fibre optique et backhaul hyperfréquence à 150 Mbit/s en zone urbaine.

Alors que le coût d'une liaison hyperfréquence est indépendant de la distance, le coût d'une fibre est une fonction linéaire de la distance. De plus, la fibre est plus onéreuse en zone urbaine. La fibre en zone urbaine ne peut être économique que pour les très courtes distances. Cependant, cette comparaison doit être interprétée avec prudence: les coûts de génie civil représentent une grande partie des dépenses de capex pour la fibre. Ils peuvent représenter jusqu'à 80% du coût total, surtout en zone urbaine. Dans de nombreux cas et notamment dans les zones de développement, ces dépenses sont de toute façon indispensables pour l'investissement d'infrastructure (routes, chemins de fer, canalisation, électricité). C'est pourquoi le paramètre pertinent pour la comparaison fibre/sans fil est le coût additionnel généré par la fibre.

Le Backhaul peut représenter plus de la moitié des coûts opérationnels d'un opérateur sans fil. En Amérique du Nord et en Europe, il est souvent plus opportun d'utiliser des liaisons louées; en d'autres lieux, le réseau est souvent auto-construit, avec des options de partage d'infrastructure et de concentration du dernier kilomètre par un pôle [52].

Une comparaison des capex et opex à 5 ans pour une liaison point à point et une fibre louée est présentée Figure 1.12 [53].

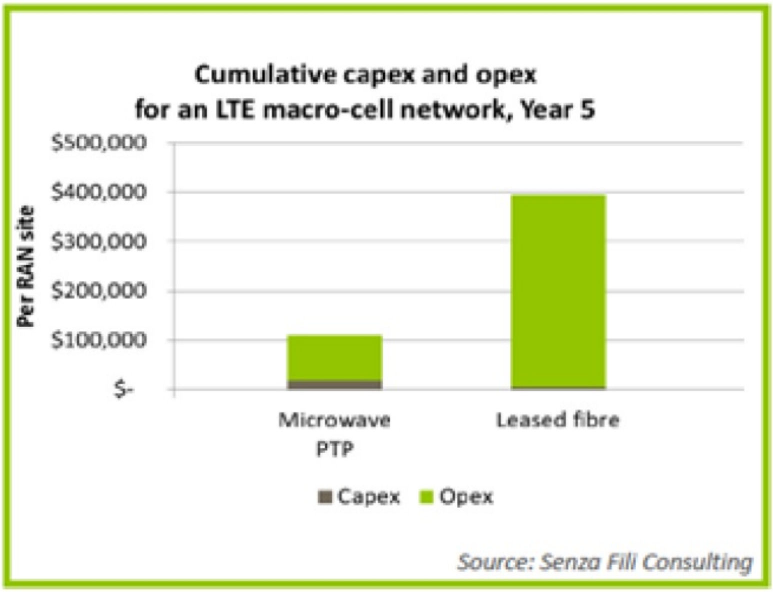

Figure 1.12: Capex et opex cumulés pour un réseau LTE macro-cell, Année 5.

La fibre louée a un capex plus faible, mais s'avère beaucoup plus coûteuse après quelques années.

Un des principaux défis pour les réseaux radio 5G est que la consommation d'énergie et le coût par liaison n'augmentent pas. Étant donné que l'on prévoit une multiplication du débit par lien par un facteur d'environ 100, il est nécessaire de réduire la consommation d'énergie par bit et le coût par bit par un facteur d'environ 100 [7]. Deux approches permettent de réduire la consommation d'énergie d'une liaison radio: les small cells réduisent la distance au terminal; en massive MIMO, les pertes d'énergie sont plus faibles car les faisceaux sont plus directifs [54].

Efficacité énergétique

Les questions d'efficacité énergétique ont été longuement analysées dans [55] et des moyens d'économie d'énergie ont été proposés:

- La gestion des ressources radio peut permettre d'économiser l'énergie par identification des faibles charges de trafic et différenciation des services. La commutation en fonction du trafic peut permettre d'économiser beaucoup d'énergie lorsque le trafic est faible. La tolérance au retard d'applications tels que l'email ou le chargement de fichier peut être exploitée afin de réduire les variations de la charge de trafic. Plusieurs questions importantes, la collaboration entre cellules voisines, l'équilibre efficacité énergétique et le QoS, l'allocation de bande passante, les MIMO et les antennes directionnelles nécessitent une étude plus approfondie.
- Les stratégies de déploiement de réseau permettent d'économiser l'énergie. Les picocells et femtocells d'intérieur réduisent l'atténuation. Les réseaux relais réduisent l'atténuation et l'interférence. Les réseaux coopératif, où les nœuds relais codent les paquets reçus de différents flux avant de les transmettre permettent de diminuer le nombre de transmissions.
- A courte distance, le SIMO (single input, multiple output) peur être plus économe en énergie que le MIMO à cause de la consommation des circuits. En OFDMA (orthogonal frequency-division multiple access), une stratégie pertinente d'allocation de canal permet d'économiser l'énergie. Ces domaines nécessitent une étude plus approfondie.

L'approche répandue "bits-par-Joule" ne prend pas en compte tous les paramètres pertinents. Une mesure de l'efficacité énergétique, prenant en compte la puissance émise et la consommation des circuits, est impérative. En particulier, l'idée communément admise que le Multiple Frequency Shift Keying (MFSK) est plus économe en énergie que le M-ary Quadrature Amplitude Modulation (MQAM) peut être réfutée lorsque la consommation du circuit est prise en compte, notamment pour les applications à courte distance.

Des solutions pour le backhaul cellulaire ont été analysées dans [52]:

- Les hyperfréquences, qui nécessitent un investissement initial élevé, mais des coûts opérationnels bas.

- Les solutions basées sur Ethernet, qui soulèvent le problème de la synchronisation.
- Les ondes millimétriques, inadaptées aux longues distances.
- Le backhaul satellite, cher, mais adapté aux zones éloignées.
- Les communications optiques en espace libre, qui fonctionnent sans câble et sans licence, mais nécessitent une haute stabilité des installations et sont exposées aux obstructions et aux atténuations.

1.2 Travaux connexes

1.2.1 Optimisation de la disponibilité

L'optimisation de la disponibilité des réseaux de télécommunication est étroitement liée à la théorie des graphes [56]. Le problème du matching maximal a été traité dans [57], [58], [59], [60]. Le meilleur algorithme connu a une complexité de $O(|V|^3)$, où V est le nombre de sommets du graphe. Le problème du matching de dimension k est NP-hard pour $k \geq 3$ [61].

1.2.2 Polarisation

La diversité de polarisation a longtemps été considérée comme un moyen supplémentaire pour répondre à la demande croissante en bande passante [10], [18]. Cette solution n'est adaptée que pour les liaisons en visibilité directe, en raison des interférences de polarisations croisées résultant des réflexions. Étant donné que les réseaux conventionnels n'utilisaient que la polarisation verticale (VP), la diversité de polarisation était réalisée en ajoutant une polarisation horizontale (HP). En raison de ce contexte historique, la diversité de polarisation était, dans une large majorité des cas, limitée aux polarisations horizontale/verticale.

Depuis, d'autres options ont été envisagées. Les performances respectives de HP/VP et des polarisations inclinées à $\pm 45°$ ont été comparées, afin de déterminer le couple optimal de polarisations orthogonales [62], [63], [64]. Ces travaux s'ajoutent à l'intérêt croissant que la diversité de polarisation a acquis ces dernières années.

Le multiplexage en polarisation dans les canaux non-LOS a été étudié dans [65]. Cependant, la solution proposée n'est pertinente que pour les signaux à un seul sens de transmission, car il n'y a pas de principe de retour inverse pour la polarisation. Pour cette raison, l'hypothèse LOS est nécessaire pour le multiplexage en polarisation dans les liaisons full-duplex. Cette hypothèse peut sembler très restrictive, spécialement dans le contexte de la technologie MIMO: le pire des cas se produit pour les réseaux d'antennes compacts dans des conditions de propagation LOS, où le rang de la matrice du canal est égale à 1 [66], [24]. Cependant, de récents travaux ont souligné le fait que les ondes millimétriques constituent la solution la plus pertinente pour le massive MIMO [25], [26]. En raison du caractère hautement directionnel et quasi-optique de la propagation des ondes millimétriques, les conditions de propagation LOS devraient être majoritaires dans les systèmes MIMO utilisant des ondes millimétriques [67], [68], [69].

Le multiplexage en polarisation est actuellement considéré comme une méthode permettant de doubler l'efficacité spectrale sur un seul canal [70]. Un exemple de diversité de polarisation a été proposé dans [71], avec six canaux de 500 MHz, chacun d'entre eux délivrant un débit de 1 Gbit/s, utilisant une modulation en quadrature 16-QAM Orthogonal Frequency Division Multiplexing.

1.2.3 Allocation des ressources

La question de l'allocation des ressources se pose chaque fois que la demande excède l'offre. Cette problématique dépasse largement le domaine des télécommunications et fait l'objet d'études approfondies en économie et en sociologie. Dans la gestion des réseaux de télécommunication, la question se pose notamment pour l'allocation des débits et l'affectation à une station de base dans les réseaux multi-tier [51], [47].

Certaines approches donnent la priorité à un équilibre entre la consommation d'énergie et les revenus [72], [55]. De plus, de nombreuses méthodes ont été proposées afin d'améliorer l'efficacité énergétique [73], [74].

D'autres approches introduisent une tarification afin d'optimiser à la fois la consommation d'énergie et les revenus [75], [50].

La caractéristique commune de toutes ces approches est que la question de l'allocation des ressources résulte du fait que l'offre est limitée. L'objectif est d'optimiser en un certain sens l'efficacité du réseau tout en maintenant une certaine équité entre les utilisateurs.

1.3 Plan de la Thèse et Contributions

La présente étude explore trois domaines différents liés à l'utilisation optimale des ressources du réseau: optimisation de la disponibilité, polarisation et allocation des ressources.

1.3.1 Optimisation de la disponibilité

Le problème de l'optimisation de la disponibilité d'un réseau en anneau est formulée au Chapitre 4. Dans cette formulation, le nombre de stations de base et le nombre d'anneaux sont donnés. Il est prouvé que si chaque anneau est constitué de 3 nœuds, le problème peut être résolu en temps polynomial, alors que pour les anneaux plus grands, le problème est NP-hard (non-deterministic polynomial-time hard). Dans ce dernier cas, des méthodes d'approximation, basées sur la programmation linéaire, sont proposées, afin de converger vers la solution.

1.3.2 Polarisation

Un nouveau paradigme sur la diversité de polarisation est proposé au Chapitre 5, dans lequel un seul canal est utilisé dans un réseau en anneau. La condition est que la polarisation de toute liaison doit être orthogonale aux polarisations des deux liaisons adjacentes. La solution proposée peut améliorer l'efficacité spectrale jusqu'à 50% par comparaison avec la solution répandue de multiplexage de polarisation. De plus, elle a des implications sur la topologie des réseaux et l'allocation de canal.

1.3.3 Allocation des ressources

Alors que les stratégies d'allocation de ressources utilisent diverses définitions pour l'efficacité du réseau et l'équité, la plupart d'entre elles sont basées sur la maximisation d'une fonction d'utilité. Le formalisme mathématique sous-jacent de ces approches est le même que celui utilisé dans le modèle de Bernoulli en finance, où un agent économique est supposé maximiser sa fonction d'utilité espérée. Ce modèle est réfuté par les paradoxes d'Allais, qui donnent des exemples de comportements rationnels qui ne peuvent pas être décrits par la maximisation d'une quelconque fonction d'utilité. En transposant ces paradoxes aux réseaux de télécommunication dans le cadre de l'allocation de ressources, il est possible de construire des exemples d'opérateurs rationnels dont le choix optimal ne peut pas être décrit par la maximisation d'une quelconque fonction d'utilité. Un modèle similaire à l'optimisation du compromis risque-rentabilité en finance est proposé au Chapitre 6, où le but est d'optimiser le compromis entre efficacité du réseau et équité.

1.4 Publications

Les travaux publiés au cours de cette thèse sont listés ci-dessous.

1.4.1 Journaux

[76] P. Ezran, Y. Haddad, and M. Debbah, "Availability optimization in a ring-based network topology", Computer Networks, vol. 124, pp. 27-32, 2017.

[77] P. Ezran, Y. Haddad, and M. Debbah, "Polarization Planning for Wireless Networks", Wireless Personal Communications, pp. 1-20, 2017.

[78] P. Ezran, Y. Haddad, and M. Debbah, "Allais' Paradox and Resource Allocation in Telecommunication Networks" *soumis à* Telecommunication Systems, 2018.

1.4.2 Conférence

[79] P. Ezran, Y. Haddad, and M. Debbah, ”Polarization Diversity in Ring Topology Networks”, IEEE VTC 2016-Fall, 2016.

1.4.3 Brevet

[80] P. Ezran and Y. Haddad, ”Wireless electromagnetic communication network using polarization diversity”, May 26 2016, European Patent WO2016079748 (A1).

Chapter 2

Introduction

Nowadays, user needs are constantly increasing with the new applications, services and usages. Today's cloud services, for example, allow users to fully and constantly backup their digital documents to remote data-centers, which can result in a very high bandwidth consumption on a wireless link. More generally, the wireless link appears to be more and more utilized today and users already experience data rate limitations or services unavailability when using, for example, their cellular network access. As spectrum occupancy is already high, it appears that we are approaching the limit of the user's wireless experience. But more and more terminals, such as smartphones and tablets, are equipped with wireless interface only and operators have to respond to an increasing need for wireless bandwidth. This advocates for a paradigm shift in the way wireless networks are managed. A network should be able to increase or decrease its capacity on demand, providing the desired QoS to users when possible and avoiding generating disturbances to close networks when possible. This requires a bandwidth management entity capable of examining the demand and the user's service level agreements (SLAs), but also the state of the different access points, of the wireless channel, etc. and to take appropriate decisions. This management entity shall build its vision of the network by gathering data from various network's equipments and perform global optimization to ultimately suggest connection options to each user and for each application. In addition, as we approach every day to the Shannon capacity limit, the solution to the ever increasing mobile traffic to/from the internet relies on smaller cells which will offer the users better signal strength and thus higher data rates. However, management of multitude of small cells turns out to be quite complex, if we are required to allocate a light loaded channel to each one, especially in the current environment where radio channel becomes a very scarce and expensive resource. New way to manage the interfering wireless devices is required. The new paradigm that is expected to provide answers to the aforementioned challenges is referred to as Software Defined Networking (SDN), which consists in decoupling the control plane from the data plane [2], [3], [4]. Lots of research is now ongoing on this field.

In parallel, a somewhat forgotten area in the field of cellular networks is the "backhaul link" which is the link between the radio access network and the core network. Up to now there was almost no engineering in the deployment of those links since a point to point topology was assumed. Today with the rise of multitude of small cells as mentioned before, this is no longer possible. Point to Multipoint and frequency reuse together with power control will be required at the backhaul. But meanwhile the SDN will enable centralized view of the network state.

SDN is defined in [5] as a network architecture with four tenets:

- The control and data planes are decoupled.
- Forwarding decisions are flow-based, instead of destination-based: all packets of a flow receive identical service policies at the forwarding devices.
- Control logic is moved to an external entity (Network Operating System, NOS).
- The network is programmable through software applications running on top of the NOS which interacts with the underlying data plane devices.

SDN, coupled with Network Functions Virtualization (NFV), offers potential to reduce capex (lower hardware costs, multi-function network equipments, multi-user implementation) and opex (shorter development and test cycles, better operational efficiency, lower energy consumption) and to speed up the time-to-market for new services (boosting innovation, faster service deployment, targeted service by geography) [6].

Some objectives of interest in this context are:

- Efficient channel allocation: with consideration of minimum required for real time traffic and flexible allocation for best effort.
- Resiliency: avoid 1+1 link deployment by sharing strategies between links from different users (operators).
- Optimization on the availability : the five "9's" against increased outage probability.
- Optimal resource allocation as a trade-off between network efficiency and fairness.

2.1 Wireless Networks

2.1.1 Introduction

The wireless industry is presently facing a tremendous growth of demand for higher data rates, driven by the development of mobile data wireless services. In order to meet this demand while keeping a consistent quality of service, network planners need to increase network capacity while taking into account architecture and topology considerations.

Since the introduction of mobile telephony in 1981, a new generation has appeared approximately every ten years: 2G (Global System for Mobile Communication, GSM) in

1991, 3G (Universal Mobile Telecommunication Service, UMTS) in 2001 and 4G (Long-Term Evolution, LTE) in 2012. Based on these observations, some suggest that 5G could be introduced around 2020 [7]. However, there is not presently any official specification on 5G and the exact technological content of the next generation mobile network is a major topic of debate [8].

Due to the limited resources of a wireless network, its optimization addresses a number of issues. The radio spectrum frequencies offer various features of propagation and fading. Their spectral efficiency can be enhanced by a proper use of polarization. Multiple Access techniques aim to optimize the use of the radio spectrum. Three technologies, MIMO (multiple input, multiple output), HetNets (Heterogeneous Networks) and full duplexing, are expected to improve spectral and energy efficiency. Network topologies offer various pros and cons regarding availability, cost and delay. Quality of Service and cost are key considerations which must be taken into account in network planning. These issues are presented in the following sections of this Chapter.

2.1.2 Channel models

Spectrum

The radio spectrum is the part of the electromagnetic spectrum from 3 Hz to 3 THz. International Telecommunication Union (ITU) frequency classification is provided in Table 2.1 ([9], p 27).

Band name	**Abbreviation**	**Frequency band**	**Wavelength**
Extremely low frequency	ELF	3 Hz - 30 Hz	100,000 km - 10,000 km
Super low frequency	SLF	30 Hz - 300 Hz	10,000 km - 1,000 km
Ultra low frequency	ULF	300 Hz - 3 kHz	1,000 km - 100 km
Very low frequency	VLF	3 kHz - 30 kHz	100 km - 10 km
Low frequency	LF	30 kHz - 300 kHz	10 km - 1 km
Medium frequency	MF	300 kHz - 3 MHz	1 km - 100 m
High frequency	HF	3 MHz - 30 MHz	100 m - 10 m
Very high frequency	VHF	30 MHz - 300 MHz	10 m - 1 m
Ultra high frequency	UHF	300 MHz - 3 GHz	1 m - 100 mm
Super high frequency	SHF	3 GHz - 30 GHz	100 mm - 10 mm
Extremely high frequency	EHF	30 GHz - 300 GHz	10 mm - 1 mm
Tremendously high frequency	THF	300 GHz - 3 THz	1 mm - 100 μm

Table 2.1: Radio spectrum.

These frequencies offer various features. Low frequencies (30 kHz-300 kHz) can diffract over obstacles and are suitable for long distances. High frequencies (3 MHz-30 MHz) are still suitable for intercontinental communications thanks to the reflection of waves by the ionosphere. Frequencies higher than 30 MHz are blocked by hills and mountains, and frequencies higher than 300 MHz are blocked by large buildings. These frequencies are suitable for short-distance communications and allow frequency reuse. Frequencies higher than 300 MHz travel by Line-of-Sight (LOS) and ground reflection.

Frequencies from 300 GHz to 300 THz are absorbed by Earth's atmosphere, while frequencies over 300 THz are in the spectrum of visible light.

The development of wireless communications makes the available spectrum scarcer and scarcer and requires solutions in order to optimize the use of its limited resources.

Propagation Model

A propagation model is proposed in [10], Chapter 3.

Free-space loss Free-space loss is the attenuation of an electromagnetic wave which results from an unobstructed line-of-sight path through free space between an transmitter and a receiver. Free-space loss is proportional to the square of the distance between the transmitter and the receiver and proportional to the square of the frequency of the signal.

The equation for free-space loss is:

$$FSPL = \left(\frac{4\pi d}{\lambda}\right)^2 = \left(\frac{4\pi df}{c}\right)^2 \tag{2.1}$$

where d is the distance between the transmitter and the receiver, λ is the signal wavelength, f is the signal frequency and c is the speed of light in vacuum.

This equation can be written:

$$L_{FSL} = 92.45 + 20\log(f) + 20\log(d) \tag{2.2}$$

with L_{FSL} in dB, f in GHz and d in km.

The distance dependency of the free-space loss is caused by the spreading out of electromagnetic energy in free space. The frequency dependency results from the fact that path loss is defined on the use of a theoretical isotropic receiving antenna. The aperture A and the gain G of an antenna are related by the equation:

$$A = G\frac{\lambda^2}{4\pi} \tag{2.3}$$

Therefore, in order to keep the gain equal to 1, a higher frequency requires a smaller aperture, and thus, a lower received power.

Attenuation In addition to the free-space loss, other factors contribute to the signal attenuation: vegetation, gas absorption, rain and obstacles. Attenuation is a growing function of frequency and a growing function of distance.

A model for vegetation attenuation has been proposed by Weissberger [11].

The most significant absorption phenomena are rain absorption above 10 GHz (absorption by water molecules has a peak at 23 GHz) and oxygen absorption within the 60 GHz band [10], [7].

Figure 2.1 shows the attenuation of electromagnetic waves as a function of frequency [12].

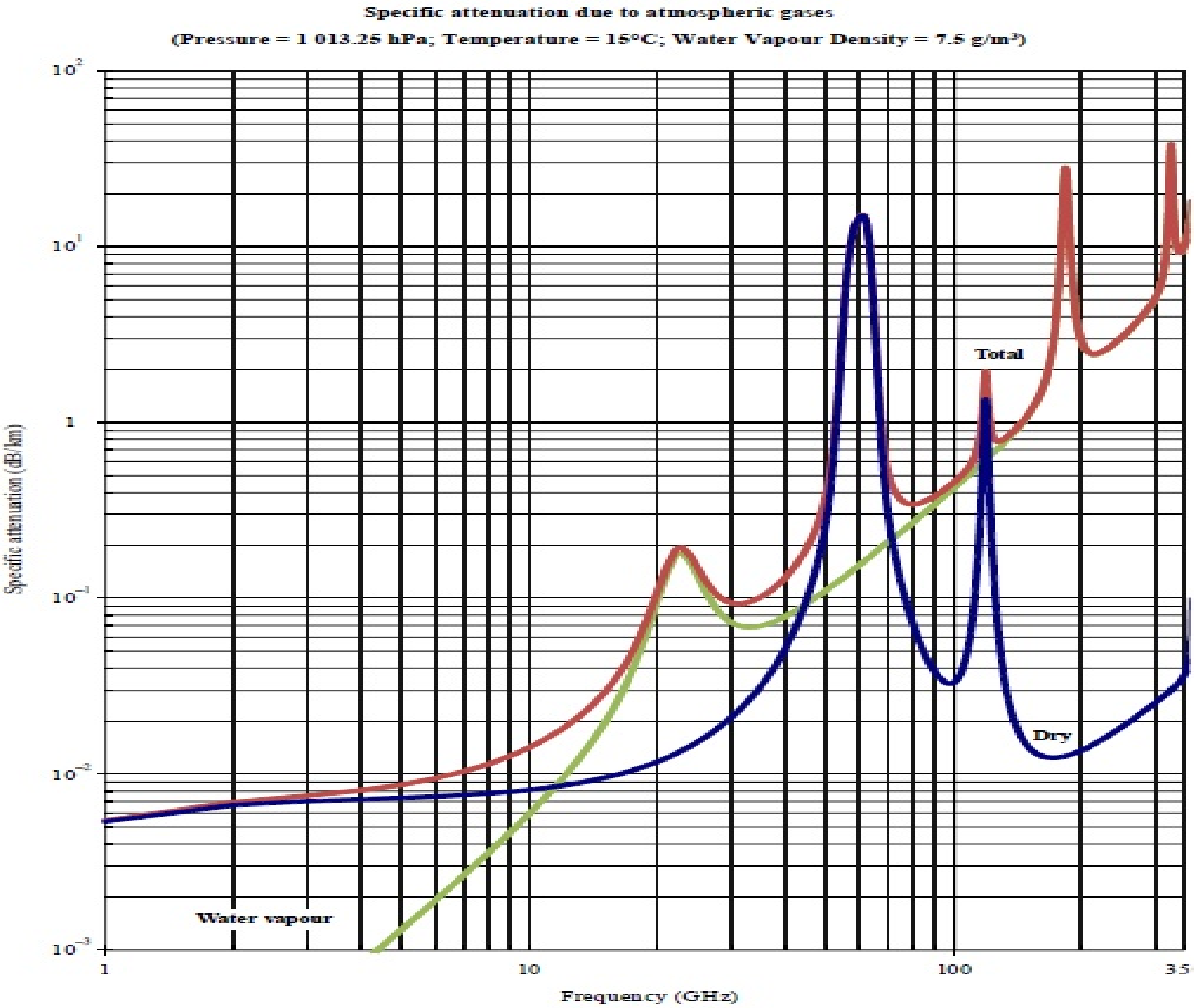

Figure 2.1: Atmospheric and molecular absorption.

Fading Fading is the attenuation of a signal due to atmospheric changes or reflections in the propagation path. The most widely used model for fading is Vigants model, which gives the fading probability as a function of climate/terrain factor, frequency, distance and Composite Fade Margin [10], [13]:

$$P = 6.10^{-10} C f d^3 10^{-\frac{CFM}{10}} \tag{2.4}$$

where

P: probability of a fade as a fraction of time

d: path length (km)

f: frequency (MHz)

C: propagation conditions factor

CFM: Composite Fade Margin (dB)

A map providing an indication of the C factor according to the location is given in [14].

Spectral Efficiency

The theoretical channel spectral efficiency is given by Shannon-Hartley capacity formula:

$$C = B \ \log_2(1 + SNR) \tag{2.5}$$

As shown on Figure 2.2, spectral efficiency of radio access networks has increased some 30 times since 2000 [15].

Polarization

In an electromagnetic wave, electric field and magnetic field are oscillating in two directions orthogonal to the propagation direction and orthogonal to each other. If the fields rotate at the wave frequency, the polarization is circular or elliptical. If the fields oscillate in one single direction, the polarization is linear. By convention, the direction of a linear polarization is the direction of the electric field.

When an electromagnetic wave encounters a boundary between two media with different refractive indices, a part of its energy is refracted and another part is reflected. The plane containing the incident, reflected and refracted rays is called plane of incidence. The incident ray can be considered as the sum of two components:

- s-polarized ray, which electric field is orthogonal to the plane of incidence;
- p-polarized ray, which electric field is parallel to the plane of incidence.

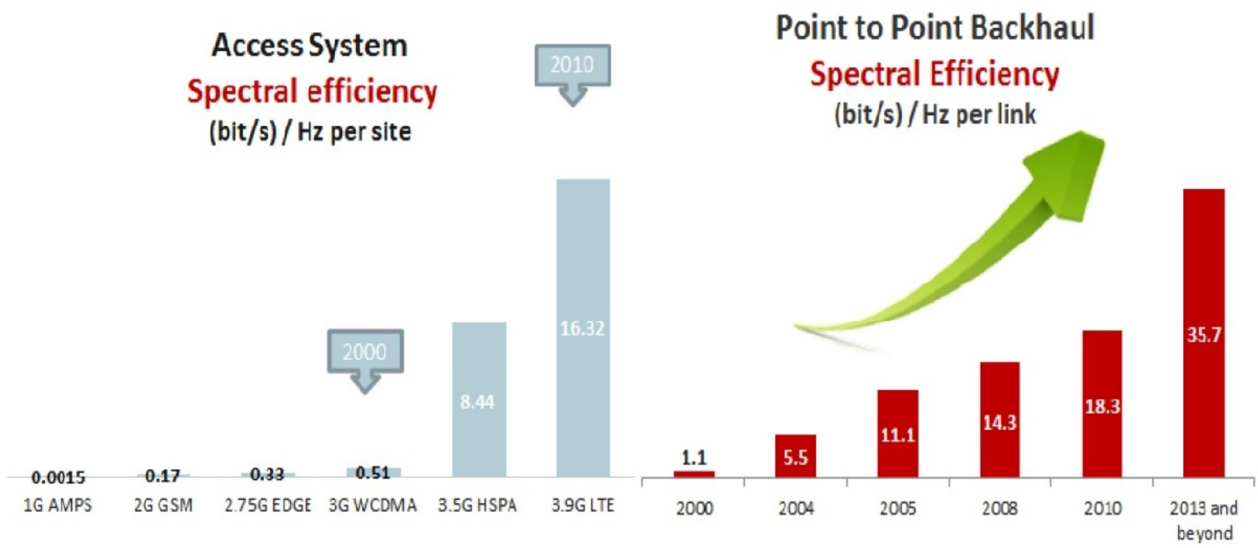

Figure 2.2: Spectral efficiency advances in RAN access and Microwave backhauling.

The relationship between the incident angle θ_i and the refracted angle θ_t is given by Snell-Descartes law (see Figure 2.3):

$$n_1 \sin\theta_i = n_2 \sin\theta_t \tag{2.6}$$

where n_1 and n_2 are the refractive indices of the two media.

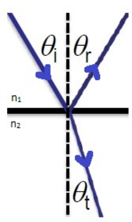

Figure 2.3: Refraction.

The reflectance, which is the fraction of the incident power which is reflected from the interface, is given by Fresnel equations, for s-polarized rays and p-polarized rays, respectively [16]:

$$R_s = \left(\frac{\sqrt{\frac{\mu_2}{\epsilon_2}} \cos\theta_i - \sqrt{\frac{\mu_1}{\epsilon_1}} \cos\theta_t}{\sqrt{\frac{\mu_2}{\epsilon_2}} \cos\theta_i + \sqrt{\frac{\mu_1}{\epsilon_1}} \cos\theta_t} \right)^2 \tag{2.7}$$

$$R_p = \left(\frac{\sqrt{\frac{\mu_2}{\epsilon_2}} \cos\theta_t - \sqrt{\frac{\mu_1}{\epsilon_1}} \cos\theta_i}{\sqrt{\frac{\mu_2}{\epsilon_2}} \cos\theta_t + \sqrt{\frac{\mu_1}{\epsilon_1}} \cos\theta_i} \right)^2 \tag{2.8}$$

where μ_1, μ_2, ϵ_1 and ϵ_2 are the magnetic permeabilities and the electric permittivities of the two materials, respectively.

Since $\sqrt{\frac{\mu_i}{\epsilon_i}} = \mu_i \frac{c}{n_i}$, these equations can be written, for non-magnetic media ($\mu_i = \mu_0$, where μ_0 is the permeability of free space):

$$R_s = \left(\frac{n_1 \cos\theta_i - n_2 \sqrt{1 - \left(\frac{n_1}{n_2} \sin\theta_i\right)^2}}{n_1 \cos\theta_i + n_2 \sqrt{1 - \left(\frac{n_1}{n_2} \sin\theta_i\right)^2}} \right)^2 \tag{2.9}$$

$$R_p = \left(\frac{n_1 \sqrt{1 - \left(\frac{n_1}{n_2} \sin\theta_i\right)^2} - n_2 \cos\theta_i}{n_1 \sqrt{1 - \left(\frac{n_1}{n_2} \sin\theta_i\right)^2} + n_2 \cos\theta_i} \right)^2 \tag{2.10}$$

The transmittance, which is the fraction of the incident power which is refracted, is deduced from the law of conservation of energy :

$$T_s = 1 - R_s \tag{2.11}$$

and

$$T_p = 1 - T_p \tag{2.12}$$

For one particular angle, $R_p = 0$. This occurs when $n_1 \cos\theta_t = n_2 \cos\theta_i$. Using this equation together with Snell-Descartes law, we obtain:

$$\sin 2\theta_i = \sin 2\theta_t \tag{2.13}$$

which means $\theta_t = \frac{\pi}{2} - \theta_i$ and $\tan\theta_i = \frac{n_2}{n_1}$. This angle is known as Brewster's angle.

Figure 2.4 represents R_s and R_p as a function of the angle of incidence for $n_1 = 1$ and $n_2 = 1.5$ (refractive index of glass). In this case, Brewster's angle is about 56°.

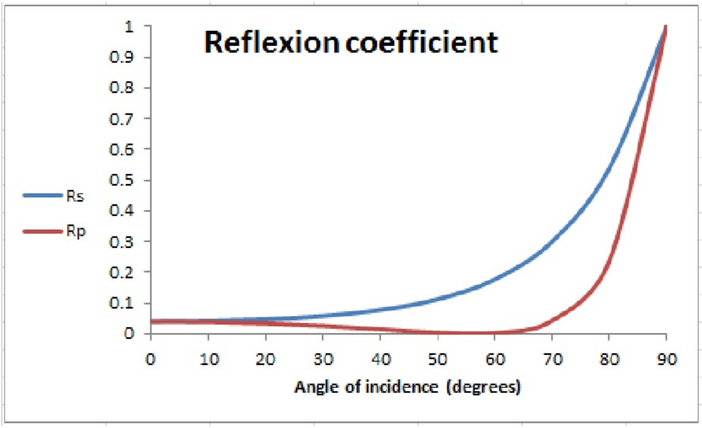

Figure 2.4: Fresnel Equations.

A major consequence of Fresnel equations is that there is no reversibility principle for polarization. If a polarization $\mathbf{E}_1$ is changed to $\mathbf{E}_2$ after reflection, the reverse ray carrying polarization $\mathbf{E}_2$ reflecting at the same point will generally have a polarization different from $\mathbf{E}_1$ after reflection.

Polarization diversity enables frequency reuse in a network. According to the present state of art, polarization is almost exclusively vertical or horizontal, though $\pm 45°$ slanted polarization and circular polarization are also considered. This is mainly due to two reasons:

- Antennas are located far away one from each other, and therefore, a difference in altitude can be neglected relative to the distance; all the nodes of a network can be considered to be co-planar.
- Most of the ring topology networks use frequency-division duplex (FDD) and therefore include an even number of nodes.

It should be noted that with the deployment of small cells, time division duplex (TDD) and full duplexing, these reasons will not stand anymore and other polarization-based solutions shall be considered in the future, especially in ring-topology networks.

2.1.3 Multiple access

Backhaul networks

The backhaul environment is commonly referred to as the part of a mobile network which connects the base stations to the core network. It is often segmented into two parts (see Figure 2.5 [1]):

- Low Radio Access Network (LRAN), which is the cell site access part of the RAN backhaul.
- High Radio Access Network (HRAN), which is the part of the network which collects, aggregates, and concentrates traffic from LRAN for connecting into the radio controllers.

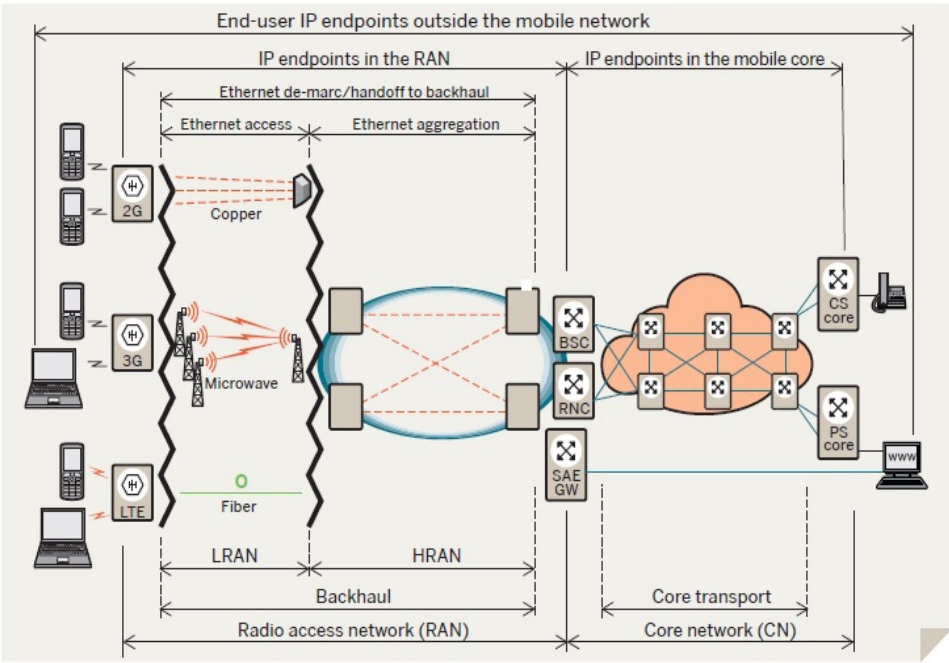

Figure 2.5: Radio, core and backhaul network boundaries [1].

Each mobile network technology generation puts its specific set of requirements on backhaul, in terms of delay, synchronization, availability, error rate and transmission rate.

Pros and cons of fiber vs microwave have been largely studied. Fiber has virtually unlimited capacity. However, it is also capital intensive (cost function of distance), often sourced from competitors (in case of leased lines), and is not always available, due to geographic or administrative constraints. On the other hand, microwave meets current and future capacity demands, is flexible and highly cost effective, and can be rapidly deployed. The general consensus is that backhaul's large aggregation and core segments, which require high capacity, should be deployed over fiber, while tail sites and pre-aggregation segments can use microwave [15].

Future backhaul networks are expected to use higher transmission frequencies (mm-waves) and consequently require more transmit antennas per base station [17].

Spread spectrum

In spread spectrum techniques, a signal generated in a given bandwidth is spread in frequency. Spread spectrum techniques started in the military area for security purposes and are presently widely used thanks to the advantages they offer regarding resistance to interferences, noise and fading.

Frequency hopping In frequency hopping, the carrier frequency is changes several times during transmission. The whole communication is carried over many frequency channels, according to a pattern known by both the transmitter and the receiver. Spread spectrum signals are difficult to intercept and highly resistant to deliberate jamming. Therefore, this technique is especially relevant for military purposes. Among other advantages, frequency hopping averages over the spectrum the received signal characteristics and reduces the probability of disastrous scenarios resulting from fading [18].

Direct sequence - Spread spectrum (DSSS) In DSSS, the transmitted signal is multiplied by a large-bandwidth signal. Since the spectrum of the resulting signal is the convolution of the two spectra, the bandwidth of the resulting signal is somewhat larger than that of the large-bandwidth signal. As a result, DSSS reduces significantly the power spectral density. The signal is more resistant to malicious attacks and fading [18].

Code Division Multiple Access (CDMA) CDMA is used on top of DSSS. Each user has a different spreading code. In order to avoid interference between signals, spreading codes of any two users must be orthogonal.

Orthogonal Frequency Division Multiplexing (OFDM)

OFDM converts one high-rate datastream into several low-rate streams, transmitted over parallel narrowband channels. This technique allows a higher signal duration, which can be required because of delay dispersion. OFDM offers higher resistance to narrowband interference and frequency-selective fading [10]. Errors resulting from low-quality channels can be corrected by error correction coding.

OFDM is aimed for one single user, but can be adapted to a multi-user scheme: in this case, each user is assigned several subcarriers which are not adjacent to each other, providing frequency diversity. This solution is often used together with Time Division Multiple Access (TDMA) [18].

Cloud Radio Access Network (C-RAN)

C-RAN is a proposed novel architecture, intended to answer the challenges raised by future cellular networks and in particular the cost resulting from the high number of Base

Stations in heterogeneous networks. In C-RAN, baseband processing is centralized and shared among sites in a virtualized baseband unit pool [19]. Therefore, C-RAN requires fewer baseband units and reduces power consumption.

2.1.4 Technologies

Increasing network capacity requires massive network densification and signal processing tools. There are three traditional approaches: using more antennas at existing cell sites (massive MIMO), HetNets and full duplexing [20].

Massive MIMO

A point-to-point MIMO system with M_t transmit and M_r receive antennas is shown on Figure 2.6 [21].

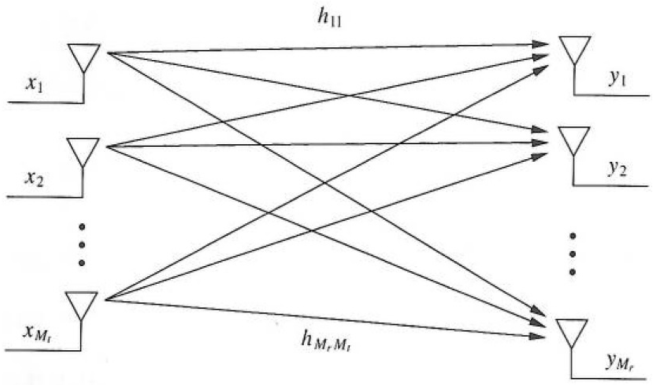

Figure 2.6: MIMO systems.

In this system, the M_t-dimensional transmitted symbol $\mathbf{x}$ and the M_r-dimensional received symbol $\mathbf{y}$ are related by the following equation:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2.14}$$

where $\mathbf{H}$ is the $M_r \times M_t$ matrix of channel gains h_{ij} from transmit antenna j to receive antenna i and $\mathbf{n}$ is the M_r-dimensional noise vector.

MIMO technology requires channel state information (CSI) at the transmitter. The performance of the multi-user MIMO system strongly depends on the correlation between the real channel and the estimated channel state information at the transmitter [22]. This is why this technology is adapted to TDD where channel reciprocity can be exploited [23].

Presently, in most MIMO implementations, each base station (BS) operates only a small number (less than 10) of antennas [24].

Millimeter wave is the most relevant solution for massive MIMO, due to the large available bandwidth [25] and the possibility to pack many antenna elements within a limited volume [26]

Three techniques can exploit diversity for the receiver [27]:

- Selection diversity chooses the path with the highest signal-to-interference-plus-noise ratio (SINR).
- Maximal Ratio Combining (MRC) makes decisions based on an optimal linear combination of the path signals.
- Equal Gain Combining (EGC) adds the path signals after they have been cophased.

In massive MIMO, a cellular base station can be equipped with hundreds or thousands of antennas. Asymptotic analysis shows that the effects of uncorrelated noise and fast fading vanish, the number of terminals per cell and the throughput per terminal are independent of the size of the cells, the spectral efficiency is independent of bandwidth, and the required transmitted energy per bit vanishes [28].

Since the number of antennas largely exceeds the number of terminals, massive MIMO offers a great number of degrees of freedom, which can be used to reduce the transmit power or to tolerate larger hardware impairments, which contributes to cost savings in the network deployment [29].

Therefore, massive MIMO is considered as a technology allowing aggressive spatial multiplexing, improving link reliability and reducing the radiated power [30].

Heterogeneous networks

A heterogeneous network is typically composed of multiple radio access technologies, architectures, transmission solutions, and base stations of varying transmission power [31]. Key issues in heterogeneous networks are BS clustering, small cells and energy harvesting.

BS clustering The introduction of HetNets challenges well established assumptions such as modeling the BS locations by a uniform hexagonal grid or assigning a user to the BS providing the strongest signal [32]. Various theoretical models including characterization of SINR distribution, intelligent cell association and fractional frequency reuse have been studied in [33], showing that adding heterogeneous elements improves significantly the network performance. User mobility modeling is usually based upon stochastic geometric framework [34].

Small cells (SC) Small cells enable frequency reuse across various cell sites. They are commonly classified according to their coverage areas: microcells have coverage areas in the range of 2 km, picocells have coverage areas in the range of 200 m and femtocells have coverage areas in the range of 10 m.

In the wireless backhaul context, small cells have to meet multiple requirements: low cost, small form, high capacity, low power, fast and easy installation, tolerance of sway, ability to operate from precarious locations, ability to cope with unforeseen and uncontrolled changes in the environment, scalability and low latency [35].

The introduction of small cells increases heterogeneity in cellular networks. The backhaul issue is usually considered in the scope of a hierarchical multi-tier model. Delay performance has been studied in [36] and [37].

Small cells have proved to be an effective way to improve network capacity without increasing the radiated power. However, this solution faces practical considerations of inter-cell interference, cell site acquisition and deployment [38].

The key benefits of femtocells are [39]:

- Better coverage and capacity: thanks to short distances and reduced interferences, femtocells enable to lower transmit power, achieve higher SINR and prolong handset battery life.
- Costs benefits: the deployment of femtocells will reduce the need for adding macro-BS towers.
- Reduced subscriber turnover: the enhanced in-building coverage provided by femtocells will reduce customer dissatisfaction.

Energy harvesting Various factors make self-powered BS realistic [40]:

- Small cells cover much smaller areas and hence require significantly smaller transmit powers compared to the conventional macrocells.
- Due to the increasingly bursty nature of traffic, the loads on the BSs will experience massive variation in space and time. In dense deployments, this means that many BSs can, in principle, be turned off most of the time and only be requested to wake up intermittently based on the traffic demand.
- Energy harvesting techniques are becoming cost-effective.
- High-speed wireless backhaul is rapidly becoming a reality for small cells, which eliminates the need for other wired connections.

Full duplexing

When a node transmits and receives simultaneously at the same frequency, the useful signal at the receiver of the incoming link is overwhelmed by the transmitted signal of the node itself [41]. For this reason, full duplexing uses self-interference cancellation in order to transmit and receive signals on the same frequency at the same time. Self-interference cancellation may be achieved by using two transmit antennas, placed at distances d and $d + \frac{\lambda}{2}$ (λ being the wavelength) from the received antenna [42].

A Time Division Duplex based heterogeneous network architecture consisting of massive MIMO Base Stations and Small Cells is proposed in [23] and seems to give promising results in terms of throughput per area.

In Complete TDD, separate time slots are allocated for communications between the base station and small cells and data exchange between small cells and their user equipments. Complete TDD and full duplexing are compared in [20]. Complete TDD is the simplest solution, but full duplexing can achieve better results, depending on conditions: according to the simulations, full duplexing can outperform complete TDD if the number of SC is small and the links between base stations and user equipments are strong enough.

2.1.5 Topology

The main parameters to consider for network topology are propagation factors (rain, Line of Sight issues) and infrastructural considerations.

The most commonly used network topologies are shown on Figure 2.7.

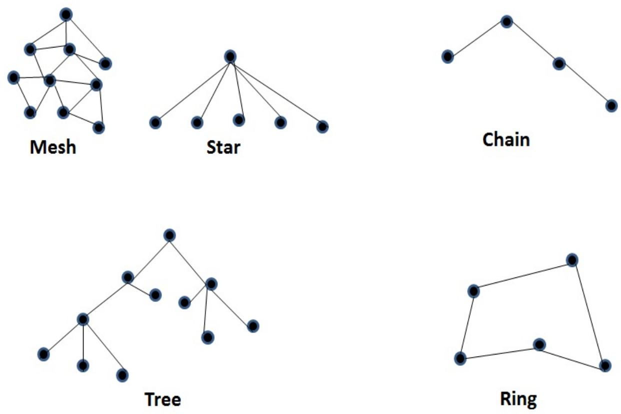

Figure 2.7: Network Topologies.

Mesh topology is designed to provide rich connectivity between nodes. It is generally found in data centers and network cores.

Star topology requires longer radio links and LOS for each link, and makes very poor frequency reuse. It is optimal in terms of delay. Star topology is usually inefficient for microwave systems, but can be used at lower frequencies.

Chain topology is the most cost-saving topology in terms of number of antennas: $2n-2$, n being the number of nodes. However, it is sensitive to any link failure. In most cases, ensuring the required protection for each link would cancel the economic benefit offered by chain topology. In addition, chain topology offers poor performance in terms of delay.

In a tree topology, every link which supports more than one site is generally protected.

Ring topology can support any single link failure event. However, delay performance is poor.

The most common topologies are trees and rings or a combination of both [10], [43].

2.1.6 Quality of Service

Availability

Unavailability is generally defined as the sum of the unavailabilities of the network nodes [10]. According to this definition, an unavailability which impacts a great number of nodes in the network gets a higher weight than an unavailability which impacts a small number of nodes. However, this definition may lead to unavailabilities greater than 100%. For this reason, we define the unavailability as the percentage of time for which all or part of the network is down.

We report hereafter some definitions from the ITU recommendations G.826, G.827, G.828 and G.8201:

Block : A block is a set of consecutive bits associated with the path; each bit belongs to one and only one block. Consecutive bits may not be contiguous in time.

Defects : the three following categories of defects related to the incoming signal are defined:

- d1 loss of signal;
- d2 alarm indication signal;
- d3 loss of frame alignment.

Errored block (EB) : A block in which one or more bits are in error.

Errored second (ES) : A one-second period with one or more errored blocks or at least one defect.

Severely errored second (SES) : A one-second period which contains at least 30% of errored blocks or at least one defect. SES is a subset of ES.

Unavailable state :

- **For a single direction**: a period of unavailable time begins at the onset of ten consecutive SES events. These ten seconds are considered to be part of unavailable time. A new period of available time begins at the onset of ten consecutive non-SES events. These ten seconds are considered to be part of available time.
- **For a bidirectional path or connection**: A bidirectional path or connection is in the unavailable state if either one or both directions are in the unavailable state.

Availability in various topology networks

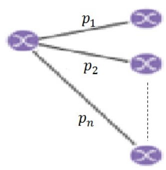

Figure 2.8: Star.

Star If the failure probabilities are independent, then the total availability is:

$$A = \prod_{i=1}^{n}(1 - p_i) \tag{2.15}$$

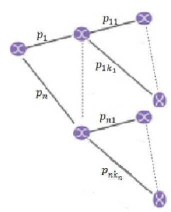

Figure 2.9: Tree.

Tree As a consequence of our definition of availability, a tree topology network is available if and only if all links are available.

If the failure probabilities are independent, then the total availability is:

$$A = \prod_{i=1}^{n} (1 - p_{i_1 \dots i_l}) \tag{2.16}$$

It should be noted that protecting a link reduces the failure probability of this link, but does not modify the general formula above.

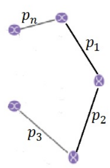

Figure 2.10: Ring.

Ring The availability of a ring topology network is the probability of no more than one failure event.

If the failure probabilities are independent, then the total availability is:

$$A = \prod_{i=1}^{n} (1 - p_k) + \sum_{i=1}^{n} p_i \prod_{\substack{1 \le j \le n \\ j \ne i}} (1 - p_j) \tag{2.17}$$

Resiliency

Resiliency is the ability of the network to provide and maintain an acceptable level of service in the face of faults and challenges to normal operations.

While the availability of an optical fiber connection is all or nothing, Automatic Coding Modulation enables a microwave link to use lower modulations in degraded conditions.

Ring topology has the highest resiliency: when a link fails, each node of the ring remains connected to the network. Only a double link failure generates an interruption of service.

The protected tree is protected against equipment failures, but not against site failure and does not provide path redundancy.

Wireless SDN improves end-user connectivity and QoS, enables multi-network planning, improves security and enables user localization [3].

Requirements regarding Quality of Service can be defined in accordance to several parameters [7]:

- Aggregate data rate refers to the total amount of data the network can serve, expressed in bit/s/area.
- Edge rate, or 5% data rate, is the worst data rate that a user can reasonably expect to receive when in range of the network.
- Peak rate is the best-case data rate that a user can hope to achieve under any conceivable network configuration.
- Latency is the time length between stimulation and response.

Requirements will be dependent upon the application: high definition video will require high data rates while safety applications or two-way gaming will give priority to low latency.

Base Station Assignment

Base Station assignment is a crucial issue in multi-tier heterogeneous networks and its importance is expected to grow even more with the introduction of 5G networks, where the SDN will enable sophisticated assignment strategies, whose end goal is QoS optimization for the end user.

Many approaches suggest a single-criteria Base Station assignment. In [44], the user connects to the BS with the maximum SINR. By connecting the user to the BS which offers the maximum long-term average received power, where all BS have full queues, it is proved in [45] that the number of tiers and density of base stations at most weakly affect the outage probability and the average ergodic rate. By connecting the user to the BS with the strongest signal, [33] provides a simplified HetNet model, where the coverage probability is independent of the number of tiers, the transmitted powers, the density of BS and the fading distribution.

However, Base Station assignment should not be based only upon signal power or SINR. Load can be a more important parameter [46].

Downlink and uplink user associations do not have to be symmetric. Base Station assignment with downlink-uplink decoupling has been studied in [47]. For uplink under spectrum sharing, if all tiers have the same target Signal-to-interference ratio (SIR), distance-based user association is optimal under a variety of path loss and power control settings. For both downlink and uplink, under orthogonal spectrum partition, the optimal proportion of spectrum allocated to each tier should match the proportion of users associated with that tier.

Optimization of base-station assignment under proportional fairness criterion has been studied in [48] and [49].

Joint Base Station assignment and resource allocation has been studied in [50] in the scope of a distributed algorithm based on dynamic pricing and in [51], where it has been proved that any efficient and strategy-proof (ie the user's best interest is to truthfully report his private information) protocol for optimal combination of assignment of users to base stations and resource allocation strategies is NP-hard.

2.1.7 Cost

Figure 2.11 shows the Total Cost of Ownership (TCO) for fiber and microwave backhaul links ([Source: Analysys Mason, 2011]).

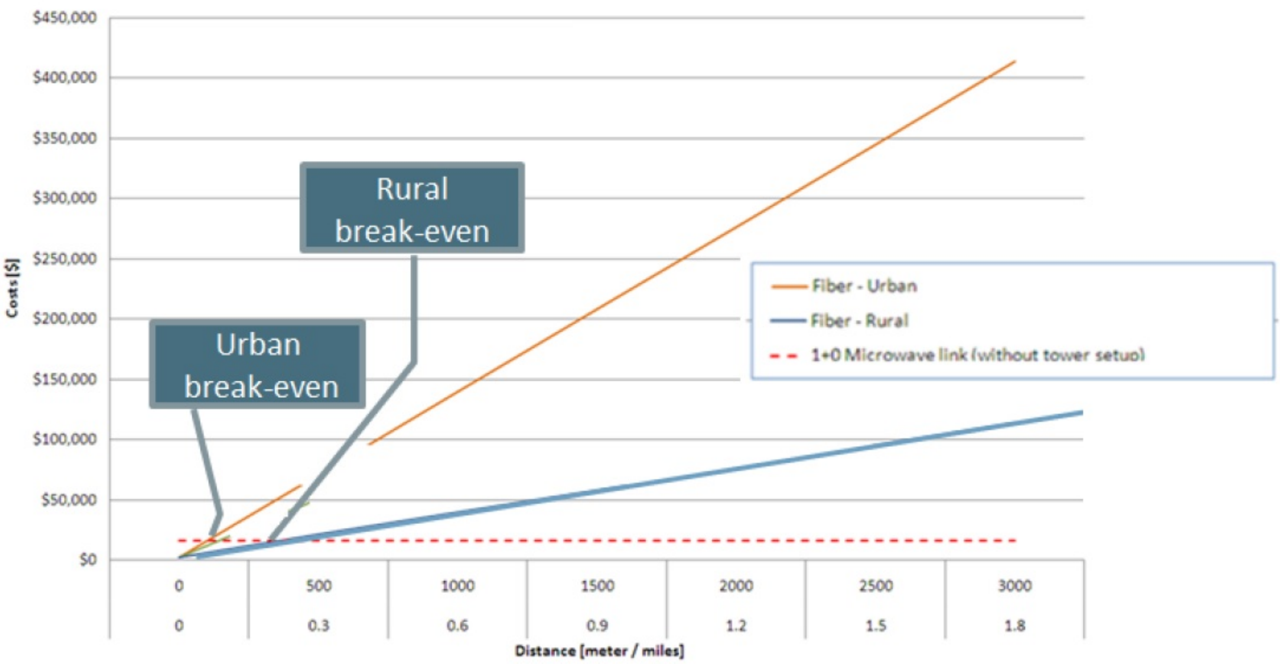

Figure 2.11: TCO of fiber and 150 Mbit/s microwave backhaul links in urban areas.

While the cost of a microwave link is independent of distance, the cost of fiber is a linear function of distance. Besides, fiber is more expensive in urban areas. The choice of fiber in urban areas may be cost-effective for very short distances only. However, this comparison must be considered cautiously: a large part of capex expenses for fiber is civil engineering. Civil engineering costs can represent up to 80% of total cost, especially in urban areas. In many cases and particularly in developing areas, these civil engineering expenses must be done anyway for infrastructure investment (road, rail, pipeline, electricity). Therefore, the relevant parameter which has to be taken into consideration for the fiber/wireless comparison is the additional cost generated by the fiber.

Backhaul can amount to over a half of a wireless carrier's recurring operating cost. In North America and Europe, it is generally more expedient to use leased lines, while in other places, the network is usually self-built, with the possibility of infrastructure sharing and concentration of multiple last-mile links via a hub [52].

Figure 2.12 compares 5-year capex and opex for microwave point-to-point and leased fiber [53].

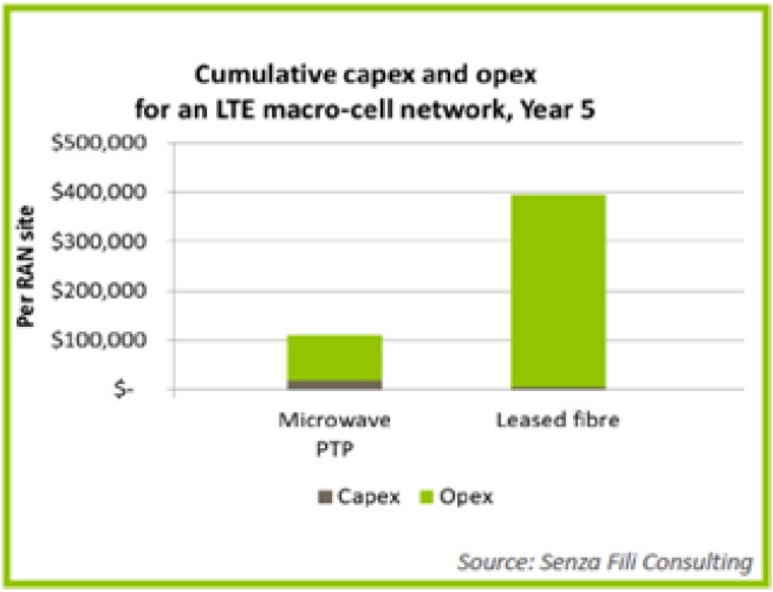

Figure 2.12: Cumulative capex and opex for an LTE macro-cell network, Year 5.

Leased fiber has a lower capex, but proves to be much more expensive after a few years.

One main challenge for 5G radio networks is that energy consumption and cost shall not increase on a per-link basis. Since the per-link data rate is expected to increase by about 100x, energy consumption per bit and cost per bit will need to fall by about 100x [7]. Two approaches enable reduction of energy consumption on the radio link: first, small cells reduce the distance to the terminal; second, in massive MIMO, energy waste is reduced because beams are more directive [54].

Energy efficiency

Energy efficiency issues have been deeply analyzed in [55] and energy saving means have been proposed:

- Radio resource management can save energy from low traffic loads and service differentiation. Traffic-aware mode switching can save a lot of energy which is otherwise wasted when the traffic load is low. Delay tolerance of applications such as email and file downloading can be exploited in order to reduce the variations in traffic load. Several important issues, such as collaboration between neighboring cells, balance between energy efficiency and QoS requirements, bandwidth allocation and spatial-domain solutions require further study.

- Network deployment strategies can save energy. Indoor picocells and femtocells reduce penetration loss and path loss. Relay networks reduce path loss and interference. Cooperative communications, where relay nodes encode the received packets from different traffic flows before forwarding the encoded packet leads to fewer transmissions.

- At short distance, SIMO (single input, multiple output) may be more energy efficient than MIMO because of circuit power. In OFDMA (orthogonal frequency-division multiple access), a relevant channel allocation strategy can save energy. Energy-efficient schemes in both MIMO and OFDMA need further investigation.

The popular ”bits-per-Joule” approach does not take into account all the relevant parameters. An energy efficiency metric is imperative. It has to take into account transmit power and circuit power. In particular, the commonly held idea that Multiple Frequency Shift Keying (MFSK) is more energy-efficient than M-ary Quadrature Amplitude Modulation (MQAM) may be disproved when circuit power is considered, especially in short-range applications.

Solutions for cellular backhaul have been analyzed in [52]:

- Microwave radio, which requires high initial investment, but low operating costs.

- Carrier Ethernet-based solutions, which raises the challenge of synchronization.

- Millimeter-wave radio, which is unsuitable for long-distance.

- Satellite backhaul, which is expensive, but the most viable for remote locations.

- Free-space optical which operates with no cable and no spectrum license, but requires high-stability mounting and faces problems due to obstructions and fog attenuation.

2.2 Related works

2.2.1 Availability Optimization

Availability optimization in telecommunication networks is closely related to graph theory [56]. The maximum 2-ring division problem has been treated in [57], [58], [59], [60]. The best known algorithm has a complexity of $O(|V|^3)$, where V is the number of the vertices in the graph. The k-dimensional matching problem is known to be NP-hard for $k \geq 3$ [61].

2.2.2 Polarization

Polarization diversity has long been regarded as another mean to meet the growing demand for wireless spectrum [10], [18]. This solution is suitable for line-of-sight links only, because of the cross-polarization interferences resulting from reflections. Since conventional networks used only vertical polarization (VP), polarization diversity was achieved by adding a horizontal polarization (HP). Because of this historical context, polarization diversity was, in a large majority of cases, restricted to horizontal/vertical.

Since then, other options have been considered. The respective performances of HP/VP versus $\pm 45°$ slanted polarization have been compared, in order to determine the optimal pair of orthogonal polarizations [62], [63], [64]. These works came along with the increased interest polarization diversity has gained these recent years in frequency planning.

Polarization multiplexing in non-LOS channels has been studied in [65]. However, the proposed solution is relevant only for one-way transmission signals, since there is no reversibility principle for polarization. For this reason, the LOS assumption is necessary for polarization multiplexing in full-duplex links. This assumption could be considered as very restrictive, especially in the context of the MIMO technology: the worst case scenario occurs for compact arrays under LOS propagation conditions, where the rank of the channel matrix equals to 1 [66], [24]. However, recent works have highlighted that millimeter wave is the most relevant solution for massive MIMO [25], [26]. Due to the highly directional and quasi-optical nature of millimeter wave propagation, LOS propagation condition should be dominant in MIMO systems with millimeter wave communications [67], [68], [69].

Polarization multiplexing is presently seen as a method for doubling spectral efficiency on a single channel [70]. An example of polarization diversity has been proposed in [71], with six 500 MHz channels, each one delivering a throughput of 1 Gbit/s, using 16-Quadrature Amplitude Modulation (QAM) Orthogonal Frequency Division Multiplexing.

2.2.3 Resource Allocation

The question of resource allocation arises whenever demand exceeds supply. This issue goes far beyond the field of telecommunications and is extensively studied in economics and sociology. In telecommunication network management, the question arises especially in throughput allocation and base-station assignment in multi-tier networks [51], [47].

Some approaches favor the search of balance between energy consumption and throughput revenue [72], [55]. In addition, many methods have been proposed in order to improve energy efficiency [73], [74].

Other approaches introduce a pricing in order to optimize both energy consumption and throughput revenue [75], [50].

The common feature of all these approaches is that the question of resource allocation results from the fact that supply is limited. The objective is to optimize in some way the network efficiency while maintaining a certain fairness among the users.

2.3 Thesis Outline and Contributions

The present work investigates three different fields related to the optimal use of the network resources: availability optimization, polarization and resource allocation.

2.3.1 Availability optimization

In Chapter 4, the problem of availability optimization of a ring topology network is formulated. In this formulation, the number of cellular sites and the number of rings are given. It is proved that if each ring includes 3 nodes, the problem can be solved in a polynomial time, while for bigger rings, the problem is NP-hard (non-deterministic polynomial-time hard). In this latter case, approximation methods, based on linear programming, are proposed, in order to converge to the solution.

2.3.2 Polarization

In Chapter 5, a new paradigm regarding polarization diversity is proposed, where one single channel is used in a ring topology network. The condition is that the polarization of any link is orthogonal to the polarizations of the two adjacent links. The proposed solution can improve spectrum efficiency by up to 50% in comparison with the widespread polarization multiplexing solution. Furthermore, it has implications on network topology and channel allocation.

2.3.3 Resource allocation

While resource allocation policies use various definitions for network efficiency and fairness, most of them are based on maximization of a utility function. The mathematical formalism underlying these approaches is the same as the mathematical formalism used in the Bernoulli model in finance, where a player is supposed to maximize his expected utility function. This model is disproved by Allais' paradoxes, which provide examples of rational behaviors which cannot be described by the maximization of any utility function. By transposing theses paradoxes to telecommunication networks for the purpose of resource allocation, it is possible to build examples of rational operators whose optimal choice cannot be described by the maximization of any utility function. A model similar to the risk-return trade-off optimization in finance is proposed in Chapter 6, where the purpose is to optimize the trade-off between network efficiency and fairness.

2.4 Publications

The publications made during the course of this PhD are listed below.

2.4.1 Journal articles

[76] P. Ezran, Y. Haddad, and M. Debbah, "Availability optimization in a ring-based network topology", Computer Networks, vol. 124, pp. 27-32, 2017.

[77] P. Ezran, Y. Haddad, and M. Debbah, "Polarization Planning for Wireless Networks", Wireless Personal Communications, pp. 1-20, 2017.

[78] P. Ezran, Y. Haddad, and M. Debbah, "Allais' Paradox and Resource Allocation in Telecommunication Networks" *submitted to* Telecommunication Systems, 2018.

2.4.2 Conference paper

[79] P. Ezran, Y. Haddad, and M. Debbah, "Polarization Diversity in Ring Topology Networks", IEEE VTC 2016-Fall, 2016.

2.4.3 Patent

[80] P. Ezran and Y. Haddad, "Wireless electromagnetic communication network using polarization diversity", May 26 2016, European Patent WO2016079748 (A1).

Chapter 3

Mathematical Tools

In this chapter, we provide some background and theoretical results which are used in the following chapters. Graph theory results are used in Chapter 4. Results on economic behavior are used in Chapter 6.

3.1 Graph theory

3.1.1 Definitions

Undirected graph: an undirected graph is an ordered pair $G = (V, E)$, where V is a set of vertices (or nodes) and E is a set of edges, which are 2-element subsets of V.

Directed graph: a directed graph is an ordered pair $G = (V, E)$, where V is a set of vertices (or nodes) and E is a set of edges, which are ordered pairs of element of V.

Weighted graph: a weighted graph is a graph in which a number is assigned to each edge.

Matching: given an undirected graph $G = (V, E)$, a matching M is a subset of E where no two edges of M share a common vertex.

3.1.2 Maximum weighted matching

Weights in graphs may represent various parameters: cost, distance, time, capacity, availability. Many telecommunication networks planning problems are related to maximization or minimization of weighted graphs or weighted matchings.

Given an undirected weighted graph $G = (V, E)$, a matching M is a maximum weighted matching if it fulfills the condition:

For each matching $\mathcal{M}'$ in G, $\sum_{(M_i,M_j)\in\mathcal{M}} w(M_i, M_j) \geq \sum_{(M_i,M_j)\in\mathcal{M}'} w(M_i, M_j)$.

Algorithms for maximum weighted matching are provided in [57], with a running time complexity of $O(|V|^2|E|)$, and in [58], [59], [60] with a running time complexity of $O(|V|^3)$

3.2 Economic behavior

The behavior of economic players (individuals or institutions) are the result of rational and irrational factors. For the purpose of our study, which draws a parallel between economic behavior and network operator strategy in resource allocation, we will focus on the rational factors, which are the expected return and the risk. In order to understand the impact of these two factors upon the behavior of economic players, we present below various kinds of lotteries and the theories which have been developed in order to explain what could be a plausible decision of a player who would be proposed to participate to these lotteries.

3.2.1 Saint Petersburg paradox

Consider a game where a coin is tossed at each stage until it comes up tails. The number of tosses, n determines the prize, which equals 2^n. What would be a fair price to enter the game?

The expected payoff is:

$$\sum_{n=1}^{+\infty} \left(\frac{1}{2}\right)^n 2^n = +\infty \tag{3.1}$$

Considering only the expected gain would lead to the conclusion that a player would agree to enter the game at any price. This conclusion is in conflict with observations: most people are not ready to pay a huge amount of money to enter this game.

3.2.2 Expected utility theory

The expected utility theory was introduced by Daniel Bernoulli. According to Bernoulli, the determination of the value of an item must not be based on the price, but rather on the utility it yields. The marginal utility of money is a decreasing function of the hold amount, which means the the utility function must be concave. A common utility function, suggested by Bernoulli, is the logarithm function.

The expected utility of a game is:

$$\sum_{x \in O} p(x)U(x) \tag{3.2}$$

where O is the set of the outcomes, $p(x)$ and $U(x)$ are the probability and the utility of outcome x, respectively.

Bernoulli's expected utility theory does not completely solve Saint-Petersburg paradox, since it is possible to change the lottery in such a way that the paradox reappears. For example, if $U(x) = \ln x$, the paradox reappears if the prize is e^{2^n} instead of 2^n. More generally, if the user's utility function U is unbounded, then it is possible to generate a variant of the paradox by choosing a prize x_n fulfilling the condition $U(x_n) \geq 2^n$, n being the first toss coming up tails.

3.2.3 The Von Neumann-Morgenstern formulation

The expected utility model was formally developed by John Von Neumann and Oscar Morgenstern [81]. In this model, a rational player is able to define his preference between two lotteries: if $\varphi(A) \geq \varphi(B)$, the player prefers lottery A to lottery B ($A \succeq B$).

Von Neumann and Morgenstern stated four axioms which define a rational decision maker:

- Completeness: for every A and B, $A \succeq B$ or $B \succeq A$.
- Transitivity: for every A, B and C, if $A \succeq B$ and $B \succeq C$, then $A \succeq C$.
- Independence: for any $t \in [0, 1]$ and for every A, B and C, if $A \succeq B$, then $tA + (1-t)C \succeq tB + (1-t)C$.
- Continuity: for every A, B and C, if $A \succeq B \succeq C$, there exist some $t \in [0, 1]$ such that B is equally good as $tA + (1-t)C$.

3.2.4 Allais' paradoxes

Allais' first paradox

Allais' first paradox is based on the four lotteries described in Table 3.1[1].

Allais claims that a rational player can prefer lottery $1A$ to lottery $1B$ and lottery $2B$ to lottery $2A$. The reason is risk aversion: if the risk is low, the player will prefer the more secure choice. If the risk is high anyway, the player will try to maximize the risk premium. However, these preferences contradict the expected utility model:

If the player prefers Lottery $1A$ to Lottery $1B$, then:

$$f(1M\$) > 10\% f(5M\$) + 89\% f(1M\$) + 1\% f(0) \tag{3.3}$$

[1]French francs were originally used in all the paradoxes

Lottery	Chance	Winnings
1A	100%	1 M$
1B	10% 89% 1%	5 M$ 1 M$ 0
2A	11% 89%	1 M$ 0
2B	10% 90%	5 M$ 0

Table 3.1: Allais paradox

Therefore,

$$11\%f(1M\$) > 10\%f(5M\$) + 1\%f(0) \tag{3.4}$$

But if he prefers Lottery $2B$ to Lottery $2A$, then:

$$10\%f(5M\$) + 90\%f(0) > 11\%f(1M\$) + 89\%f(0) \tag{3.5}$$

So,

$$11\%f(1M\$) < 10\%f(5M\$) + 1\%f(0) \tag{3.6}$$

Inequalities (3.4) and (3.6) clearly contradict each other.

Allais' second paradox

Allais' second paradox is aimed to disprove the independence assumption formulated by Von Neumann and Morgenstern. The independence assumption is based on the following argument: the lottery $tA + (1-t)C$ (resp. $tB + (1-t)C$) can be performed in two steps: first we draw lots between lotteries A (resp. B) and C with probabilities t and $(1-t)$, and then we play the chosen lottery. If the probability t event occurs, then we play lottery A (resp. lottery B). If the probability $(1-t)$ event occurs, then we play lottery C anyway. Since lottery A is assumed to be preferable to lottery B, the player shall prefer lottery $tA + (1-t)C$ to lottery $tB + (1-t)C$.

However, Allais points out that the choice must be *ex ante* and not *ex post*, and provides an example based upon the three lotteries described in Table 3.2.

With $t = 1\%$, the combined lotteries are described in Table 3.3.

Rational, but cautious, players may prefer lottery A to lottery B and lottery $tB + (1-t)C$ to lottery $tA + (1-t)C$. Once again, as in the first paradox, the reason is risk aversion.

Lottery	Chance	Winnings
A	100%	100 M$
	98%	500 M$
B	2%	0
C	100%	0

Table 3.2: Independence assumption - Counter-example

Lottery	Chance	Winnings
	1%	100 M$
tA+(1-t)C	99%	0
	0.98%	500 M$
tB+(1-t)C	99.02%	0

Table 3.3: Combined lotteries

Allais' main conclusion is that all the properties of the probability distribution must be taken into account for any rational choice involving risk.

3.2.5 Risk-return trade-off in portfolio management

In many financial models, the considered probability distributions are Gaussian. In this case, the mean and the volatility (standard deviation) provide all the information and the player can define his risk aversion profile according to these two parameters. Otherwise, the definition of the player's risk aversion profile based on mean and standard deviation is only an approximation.

The portfolio management model was developed by Harry Markowitz in 1952 [82]. Given a set of individual assets with their respective expected returns, volatilities and correlations, it can be proved that the set of feasible portfolio is delimited by a curve, which is called the minimum-variance frontier ([83], Chapter 7). The upper part of the minimum-variance frontier is called the efficient frontier of risky assets. Any rational investor shall define his risky portfolio on the efficient frontier (Fig. 3.1).

The choice of a risky portfolio on the efficient frontier is a matter of risk aversion. A risk-averse investor will choose a portfolio close to the global minimum-variance portfolio, in order to reduce risk. A non-risk-averse investor will choose a more risky portfolio, offering a higher expected return.

3.2.6 Prospect theory

Prospect theory [84] proposes a more accurate description of decision making under risk, compared to expected utility theory. The model takes into account psychological factors

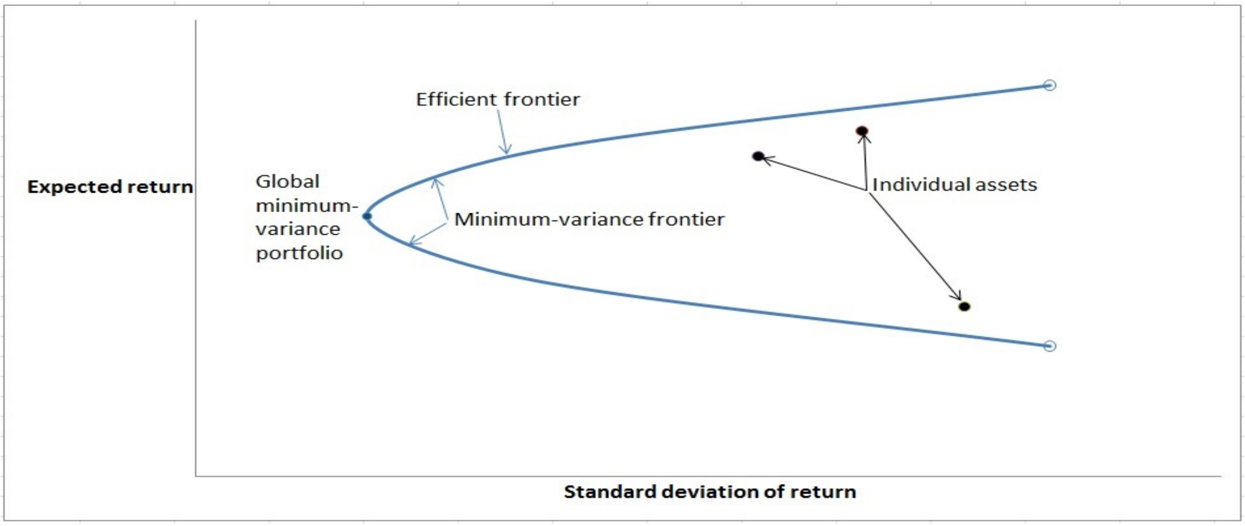

Figure 3.1: The efficient frontier of risky assets.

which take part in the decision process, such as risk aversion or underweighting low-probability outcomes. Prospect theory addresses Allais paradoxes and points out that besides the final outcomes and probabilities, the way the problem is presented has an impact on the payer's decision. For example, in Allais' second paradox, the player will express different preferences whether the game is presented as a one-stage game or a two-stage game. Mathematically, the value of a prospect $(x_1, p_1; ...; x_n, p_n)$ which yields outcome x_i with probability p_i is:

$$V = \sum_{i=1}^{n} \pi(p_i)v(x_i) \tag{3.7}$$

where the decision weight $\pi(p_i)$ reflects the impact of p_i on the over-all value of the prospect, while the value function $v(x_i)$ reflects the subjective value of the outcome.

The value function v has similar properties to utility function in the expected utility theory. It is therefore a concave function for gains ($x_i \geq 0$) and a convex function for losses $(x_i \leq 0)$, with $v(0) = 0$.

The weighting function π must fulfill the following properties:

- $\pi(0) = 0$
- $\pi(1) = 1$
- $\pi(p) > p$ for low probabilities
- $\pi(p) + \pi(1 - p) < 1$ for $0 < p < 1$

These conditions imply discontinuity of function π on 0 and 1. In addition, they imply violations of dominance: let us consider a game where a player wins $1M\$ - \epsilon$ with probability $\frac{1}{2}$ and $1M\$ + \epsilon$ with probability $\frac{1}{2}$. When ϵ tends to 0, the game tends to be equivalent to a certain win of $1M\$$. However, the value function of the game will be:

$V = (v(1M\$ - \epsilon) + v(1M\$ + \epsilon))\pi\left(\frac{1}{2}\right)$, which tends to $2v(1M\$)\pi\left(\frac{1}{2}\right) < v(1M\$)$ when ϵ tends to 0.

In order to answer Allais' paradoxes, the prospect theory introduced a non-linear weighting function π. This key property of non-linearity brings in turn another drawback: the discontinuity of the prospect value.

Chapter 4

Availability Optimization in Ring-Based Network Topology

4.1 Overview

The choice of a network technology is mainly a matter of cost, availability and resiliency. For a given technology, these parameters will also impact the choice of the topology.

While fiber is capital intensive (cost function of distance) and offers limited availability, wireless is highly cost effective and flexible [15]. Besides, fiber is more expensive in urban areas. The choice of fiber in urban areas may be cost-effective for very short distances only. However, this comparison must be considered cautiously: a large part of capex expenses for fiber is civil engineering. Civil engineering costs can represent up to 80% of total cost, especially in urban areas. In many cases and particularly in developing areas, these civil engineering expenses must be done anyway for infrastructure investment (road, rail, pipeline, electricity). Therefore, the relevant parameter which has to be taken into consideration for the fiber/wireless cost comparison is the additional cost generated by the fiber.

Availability is a key parameter which quantifies network performance. This parameter is closely related to reliability. The difference between these two concepts is that reliability refers to failure-free operation during an interval, while availability refers to failure-free operation at a given instant of time [85].

As explained in Chapter 2, unavailability is the percentage of time for which all or part of the network is down.

While the availability of an optical fiber connection is all or nothing, line-of-sight and propagation considerations must be taken into account in wireless links. Automatic Coding Modulation enables a microwave link to use lower modulations in degraded conditions. This difference has an impact on the network resiliency, which is the ability of the network to provide and maintain an acceptable level of service in the face of faults and challenges to normal operations [86].

In the case of a radio network, the main parameters which come into account are propagation factors and infrastructural considerations. The most common topologies used in radio backhaul networks are trees and rings or a combination of both. Since tree topology generally offers shorter paths and lower costs, while ring topology generally ensures a better availability, a ring-tree combination can be an efficient solution to cumulate the advantages of both technologies [43].

Various causes of a network failure are identified in [86]: unusual traffic load, accidents and human mistakes, large-scale disasters, malicious attacks, environmental challenges and failures at a lower layer. A relevant network availability strategy must reduce as much as possible the failure probability of any link in the network and add redundancy in order to minimize the impact of a single link failure on the availability of the network nodes.

Statistical approaches have been proposed in order to optimize availability [87], [88] for systems subject to random failures. These approaches are based upon maintenance considerations for a partially observable system.

Backhaul can be made of fiber or microwave radio. In both cases, the goal is to connect the base stations to the core network. In some cases, when the gateway to the core is not far, this can be performed in one hop. But in rural areas or in Ultra Dense Networks, where there are a huge number of small BS to connect, this can require multiple hops. In this Chapter, we consider the Microwave Radio technology as the medium to perform the backhaul. We assume that we have a large number of BS to be connected to a single aggregation node which itself will be connected to the core network (Figure 4.1). This latter connection is assumed to be wired and therefore out of the scope of our study. Making a single large ring raises serious delay issues since the Backhaul for a BS might require several hops. In addition, it might raise serious availability issues since the disconnection of two links can affect a large number of BS. Therefore, it could be preferable to split the network into several rings.

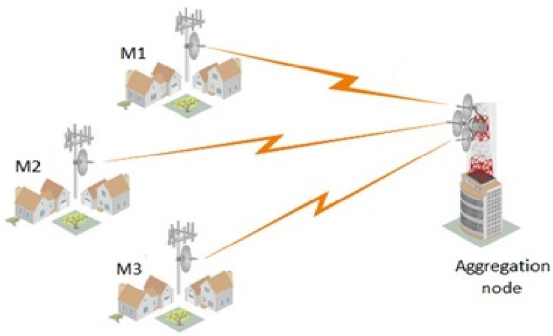

Figure 4.1: Backhaul network. The aggregation node handles all the traffic produced by and to nodes M1, M2, M3.

In this Chapter, we will study the question of topology optimization from the point of view of availability maximization. Given an aggregation node and n cellular sites, what is the best topology based on rings, each one of them including the aggregation node, which maximizes availability?

4.2 Simplified model

In a first step, we build a simplified model, based on the following five assumptions. Though the last two assumptions of this model are not realistic, this simplified approach will enable us to draw basic conclusions regarding backhaul network topologies.

Assumptions:

- the network includes n cellular sites (in addition to the aggregation node);
- the network topology is made of h rings; each cellular site belongs to one single ring;
- for $1 \leq i \leq h$ ring i includes n_i cellular sites and the aggregation node; $n_1 \geq n_2 \geq \cdots \geq n_h \geq 2$;
- same failure probability for all links: p;
- failure events are uncorrelated.

n and n_i are related by the following equation:

$$n = \sum_{i=1}^{h} n_i \tag{4.1}$$

Availability: the condition for availability is that all cellular sites are connected to the aggregation node. This condition is fulfilled if there is no more than one failure in each ring.

$$A = \prod_{i=1}^{h} \left((1-p)^{n_i+1} + (n_i+1)p(1-p)^{n_i}\right) \tag{4.2}$$

If $p \ll 1$, this expression can be approximated by its second-order Taylor development:

$$A = 1 - n\frac{p^2}{2} - \frac{p^2}{2}\sum_{i=1}^{h} {n_i}^2 + o(p^2) \tag{4.3}$$

Therefore,

$$A = 1 - n\frac{p^2}{2} - \frac{p^2}{2}\left(hV(n_i) + \frac{n^2}{h}\right) + o(p^2) \tag{4.4}$$

where $V(n_i)$ is the empirical variance of the n_i distribution:

$$V(n_i) = \frac{1}{h}\sum_{i=1}^{h}\left(n_i - \frac{n}{h}\right)^2 = \frac{1}{h}\sum_{i=1}^{h}{n_i}^2 - \left(\frac{n}{h}\right)^2 \tag{4.5}$$

Therefore, increasing the number of rings reduces the maximum path length and unavailability. On the other hand, it requires more antennas. In any case, given the number of rings, it is preferable that the empirical variance of the ring size distribution be as small as possible.

For a given number of rings h, the maximum availability is obtained when the empirical variance is minimized, which means when the numbers of cellular sites in the rings are as close as possible to $\frac{n}{h}$. Let q and r be the quotient and the remainder of the Euclidean division of n by h:

$$n = qh + r; 0 \leq r \leq h - 1 \tag{4.6}$$

Then, $n_1 = \cdots = n_r = q + 1$ and $n_{r+1} = \cdots = n_h = q$.

Therefore, the best availability is:

$$A_{h,p} = \left((1-p)^{q+2} + (q+2)p(1-p)^{q+1}\right)^r\left((1-p)^{q+1} + (q+1)p(1-p)^q\right)^{h-r} \tag{4.7}$$

$$A_{h,p} = 1 - n\frac{p^2}{2} - \frac{p^2}{2}\left(r(q+1)^2 + (h-r)q^2\right) + o(p^2) \tag{4.8}$$

$$A_{h,p} = 1 - \frac{p^2}{2}(n + hq^2 + 2rq + r) + o(p^2) \tag{4.9}$$

$A_{h,p}$ is an increasing function of h and a decreasing function of p. Of course, since the total number of antennas is $2n + 2h$, increasing h increases the cost.

The results above are illustrated with the following numerical application:

$p = 0.01$

$n = 100$

$2 \leq h \leq 50$

The cost and the availability are growing functions of the number of rings h. This defines a curve of feasible. According to the price the operator is ready to pay for a given level of availability, it is possible to define an acceptable set. Any point of the feasible curve which is inside the acceptable set is a relevant choice for the operator (Figure 4.2).

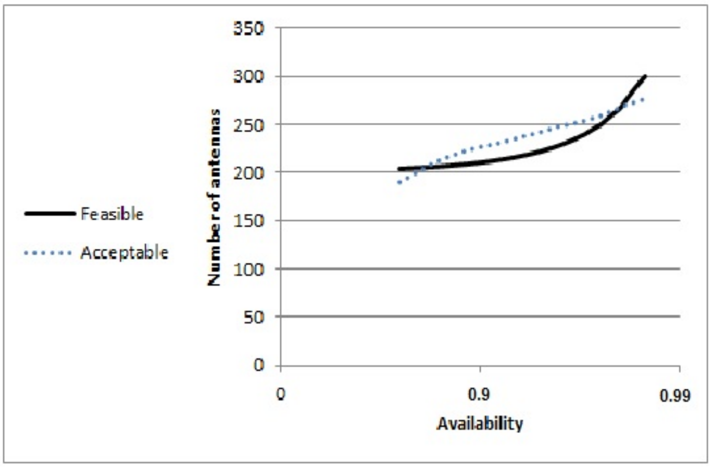

Figure 4.2: Cost-Availability balance.

It should be noted that this conclusion cannot be generalized to n rings, each one of them including one cellular node: in this case, the network topology would be a star topology and a cellular site would be unavailable in case of single failure. The availability of a star topology network including n cellular sites is:

$$A = (1-p)^n = 1 - np + o(p) \tag{4.10}$$

As a consequence, availability is maximized when all the rings include 2 nodes, in addition to the aggregation node. Assuming that n is even, then $h = \frac{n}{2}$ and

$$A = \left((1-p)^3 + 3p(1-p)^2\right)^{\frac{n}{2}} = (1 - 3p^2 + 2p^3)^{\frac{n}{2}} = 1 - \frac{3}{2}np^2 + o(p^2) \tag{4.11}$$

4.3 General model

We now assume that links may have various failure probabilities. The network includes one aggregation node (O) and n cellular sites M_1,M_2,...,M_n.

Let p_i be the failure probability of the link OM_i and p_{ij} the failure probability of the link M_iM_j.

Assuming that each ring includes the aggregation node, a ring may be defined by an ordered sequence of cellular sites.

Let us define:

$V = \{M_1, M_2, ..., M_n\}$.

R: the set of rings including the aggregation node and cellular sites of V.

A ring r of R can be defined with an m-tuple $(M_{i_1}, M_{i_2}, ..., M_{i_m})$, made from the ordered list of the nodes of r starting from the aggregation node but not including it.

The availability of r is:

$$A(r) = (1-p_{i_1})(1-p_{i_m}) \prod_{l=1}^{m-1} (1-p_{i_l i_{l+1}}) + p_{i_1}(1-p_{i_m}) \prod_{l=1}^{m-1} (1-p_{i_l i_{l+1}}) + p_{i_m}(1-p_{i_1}) \prod_{l=1}^{m-1} (1-p_{i_l i_{l+1}})$$
$$+ (1-p_{i_1})(1-p_{i_m}) \prod_{l=1}^{m-1} (1-p_{i_l i_{l+1}}) \sum_{j=1}^{m-1} \frac{p_{i_j i_{j+1}}}{1-p_{i_j i_{j+1}}} \quad (4.12)$$

For a given $s \leq n$, we try to maximize the following expression:

$$\max_{\substack{r_1 \cup ... \cup r_s = V \\ r_i \cap r_j = \varnothing}} \prod_{i=1}^{h} A(r_i)$$

Particular case $\frac{n}{2}$ rings, each one including the aggregation node and two cellular sites

In this section, we assume the following:

- n is an even number;
- the network includes $\frac{n}{2}$ rings, each one including the aggregation node and 2 cellular sites;
- for $1 \leq i \leq n$, p_i is the failure probability of the link OM_i;
- for $1 \leq i, j \leq n$, p_{ij} is the failure probability of the link M_iM_j;
- failure events are uncorrelated.

We can calculate the availability of the ring OM_iM_j.

$$A_{ij} = 1 - p_ip_j - p_ip_{ij} - p_jp_{ij} + 2p_ip_jp_{ij} \quad (4.13)$$

We try to maximize the expression:

$$\prod_{(O,M_i,M_j) \in R} A_{ij} \quad (4.14)$$

which is equivalent to maximize the expression:

$$\sum_{(O,M_i,M_j) \in R} \log A_{ij} \quad (4.15)$$

The problem can be regarded as a search of a perfect matching in a weighted graph: given $G = (V, E, w)$ an undirected weighted graph, the goal is to compute a perfect matching (ie a subset of edges $E' \subseteq E$ such that each node in V has exactly one incident edge in E') for a maximum total weight $w(E')$.

The maximum 2-ring division problem can be solved efficiently (in polynomial time) as followed: given a network as described above, an undirected weighted graph $G = (V, E)$ should be constructed where $V = \{M_1, M_2, \dots, M_n\}$ and $E = (M_i, M_j)|1 \leq i < j \leq n$ (a full graph). The weight function $w : E \to \mathbb{R}$ is defined as followed: $\forall i, j, 1 \leq i < j \leq n, w(M_i, M_j) = \log A_{ij}$. Then, due to [56], finding a maximum 2-ring division in the original network is equivalent to finding a matching $\mathcal{M}$ in G such that for each matching $\mathcal{M}'$ in G, $\sum_{(M_i,M_j)\in\mathcal{M}} w(M_i, M_j) \geq \sum_{(M_i,M_j)\in\mathcal{M}'} w(M_i, M_j)$.

This problem is a well-known problem called maximum weighted matching. In 1964, Jack Edmonds was the first to develop a polynomial time algorithm to solve this problem [57]. A straight forward implementation of Edmonds' algorithm will have a running time complexity of $O(|V|^2|E|)$, and hence in our problem $O(|V|^4)$ (because the constructed graph is fully meshed. i.e. $E = \Theta(|V|^2)$). Over the years, several variants, implementations and improvements of Edmonds' idea where suggested, some of them in [56], [58], [59]. Overall, the best know algorithm for a full graph has a running time complexity of $O(|V|^3)$ [58], [59], [60].

Now, solving the maximum 2-ring division problem is done in 2 phases:

1. Computing $logA_{ij}$ for each i,j, $1 \leq i < j \leq n$ and constructing an undirected weighted full graph G, as described earlier.

2. Solving the weighted maximum matching problem on G.

The running time complexity of phase 1 is $\binom{n}{2}\Theta(1)+\Theta(n^2) = \Theta(n^2)$. The running time complexity of phase 2 is $O(n^3)$. Therefore, the running time complexity of the proposed algorithm for solving the maximum 2-ring division problem is $O(n^3)$. Hence, the decision problem corresponding to the maximum 2-ring division problem is in P.

Conclusion: it is possible to connect an even number n of cellular sites with $n/2$ rings, each of one including 2 cellular sites and the aggregation node. The running time is $O(n^3)$.

General case $\frac{n}{k}$ rings, each one including the aggregation node and k cellular sites; $k \geq 3$

In this section, we assume the following:

- n is a multiple of k;

- the network includes $\frac{n}{k}$ rings, each one including the aggregation node and k cellular sites;

- for $1 \leq i \leq n$, p_i is the failure probability of the link OM_i;
- for $1 \leq i, j \leq n$, p_{ij} is the failure probability of the link M_iM_j;
- failure events are uncorrelated.

At first, we investigate the relation between the general maximum k-ring division problem and an NP-Complete problem.

Let $\mathcal{P}_k(n)$ be the set of k-combinations of $\{1, 2, \ldots, n\}$. Given a family of sets $F \subseteq \mathcal{P}_k(n)$ for $k \geq 3$, a k-set packing of $\{1, 2, \ldots, n\}$ is a set $S \subseteq F$ such that $\forall s_1, s_2 \in S, s_1 \cap s_2 = \varnothing$. The maximum k-set packing problem (MSP) is to find a k-set packing S of $\{1, 2, \ldots, n\}$ such that for each k-set packing S' of $\{1, 2, \ldots, n\}$, $|S| \geq |S'|$. The corresponding decision problem $(d - MSP)$ is a well-known NP-Complete problem [89], [90].

We define the maximum production $[0, 1)$ weighted k-set packing (MPWSP) as followed: given a family $F = \mathcal{P}_k(n)$ where $n = mk$ for some $m \in \mathbb{N}$ and a weight function $w : F \to [0, 1)$, the MPWSP problem is to find a k-set packing S of $\{1, 2, \ldots, n\}$ such that for each k-set packing S' of F, $\prod_{u \in S} w(u) \geq \prod_{u \in S'} w(u)$.

Let $d - MPWSP$ denote the corresponding decision problem to MPWSP. $d - MSP$ is a particular case of $d - MPWSP$ with:

- $w(X) = 1$ for $X \in F$
- $w(X) = 0$ for $X \notin F$

Therefore, $d - MPWSP$ is as least as hard as $d - MSP$. Thus, d-MPWSP is NP-Hard (and in fact, d-MPWSP is NP-Complete).

Given an algorithm to solve the MPWSP problem, it can be used to solve the general maximum k-ring division problem as followed: let $A^{max}_{i_1 i_2 \ldots i_k}$ be the highest availability of all the rings including the k nodes i_1, i_2,..., i_k and the aggregation node:

$$A^{max}_{i_1 i_2 \ldots i_k} = max(A_{j_1 j_2 \ldots j_k} | j_1 j_2 \ldots j_k \text{ is a permutation of } i_1 i_2 \ldots i_k) \quad (4.16)$$

and

$$c_{i_1 i_2 \ldots i_k} = log(A^{max}_{i_1 i_2 \ldots i_k}) \quad (4.17)$$

An instance of MPWSP could be constructed by defining a family of sets $F = \mathcal{P}_k(n)$ and a weight function $w(i_1, i_2, \ldots, i_k) = A^{max}_{i_1 i_2 \ldots i_k}$. Clearly, a solution to the constructed MPWSP instance yields a solution to the original maximum k-ring division problem.

Reciprocally, let us consider the following particular case:

$$
\begin{aligned}
& V = U_1 \cup ... \cup U_k \\
& |U_1| = ... = |U_k| = \frac{n}{k} \\
& \forall u \in U_1, p_u = 0 \\
& \forall u \in U_2 \cup ... \cup U_k, p_u = 1 \\
& \forall u_i \in U_i, \forall u_j \in U_j, |j - i| \neq 1 \rightarrow p_{ij} = 1 \\
& \forall u_i \in U_i, \forall u_j \in U_j, |j - i| = 1 \rightarrow p_{ij} \in [0, 1]
\end{aligned}
$$

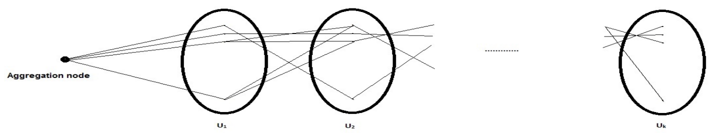

Figure 4.3: Particular case of k-ring division.

In this particular case, the aggregation node is connected to all the nodes of U_1 and no other node. Every connection between the aggregation node and anyone of the nodes of U_1 is assumed to be free of failure-risk. A node in a given subset U_i can be connected only to the nodes belonging to the adjacent sets U_{i-1} and U_{i+1} (Figure 4.3).

Then, the maximization of $c_{i_1 i_2 ... i_k}$ is a k-dimensional matching problem, which is known to be NP-hard [61] for $k \geq 3$. Therefore, the general problem is NP-hard for $k \geq 3$.

4.4 Approximation methods

Since the general maximum k-ring division problem is NP-hard for $k \geq 3$, we propose hereafter approximation methods in order to converge to the solution.

4.4.1 Formalization as an Integer Linear Programming (ILP) Problem

We can present the k-ring division problem as an ILP: the idea is to define binary variables which correspond to a k-ring.

$$P = \max \sum_{\{i_1, i_2, ..., i_k\} \in \mathcal{P}_k(n)} c_{i_1 i_2 ... i_k} x_{i_1 i_2 ... i_k} \tag{4.18}$$

subject to

$$\sum_{\substack{\{i_1,i_2,\ldots,i_k\}\in\mathcal{P}_k(n)\\ j\in\{i_1,i_2,\ldots,i_k\}}} x_{i_1 i_2\ldots i_k} = 1, \forall j \in \{1,2,\ldots,n\} \tag{4.19}$$

$$x_{i_1 i_2\ldots i_k} \in \{0,1\}, \forall\{i_1,i_2,\ldots,i_k\} \in \mathcal{P}_k(n) \tag{4.20}$$

The purpose of this method is to characterize the network topology by binary values: $x_{i_1 i_2\ldots i_k} = 1$ if the nodes $i_1, i_2, \ldots, i_k$, together with the aggregation node, form a ring, and $x_{i_1 i_2\ldots i_k} = 0$ else.

Constraints (4.19) and (4.20) forces each node j to be in exactly one k-ring.

General ILP is known to be NP-Hard [60]. However, linear programming can be solved in polynomial time. By replacing constraint (4.20) in (4.18) with the constraint:

$$x_{i_1 i_2\ldots i_k} \geq 0, \forall\{i_1,i_2,\ldots,i_k\} \in \mathcal{P}_k(n) \tag{4.21}$$

(also known as LP relaxation) we get a polynomial-time solvable linear program.

Without loss of generality, we can assume that $c_{i_1 i_2\ldots i_k} > 0$ for each $\{i_1, i_2, \ldots, i_k\} \in \mathcal{P}_k(n)$, since we can always add any constant to all the coefficients $c_{i_1 i_2\ldots i_k}$. Doing that does not change the set of vectors that maximizes the problem, because due to the constraints, each feasible vector contains exactly $\frac{n}{k}$ ones and $\binom{n}{k} - \frac{n}{k}$ zeros. Therefore, adding K to all the coefficients $c_{i_1 i_2\ldots i_k}$ is equivalent to adding the constant $K\frac{n}{k}$ to the original objective function.

Lemma 1. *Assuming $c_{i_1 i_2\ldots i_k} > 0$ for each $\{i_1, i_2, \ldots, i_k\} \in \mathcal{P}_k(n)$, a vector x which maximizes the system*

$$P' = \max \sum_{\{i_1,i_2,\ldots,i_k\}\in\mathcal{P}_k(n)} c_{i_1 i_2\ldots i_k} x_{i_1 i_2\ldots i_k} \tag{4.22}$$

subject to

$$\sum_{\substack{\{i_1,i_2,\ldots,i_k\}\in\mathcal{P}_k(n)\\ j\in\{i_1,i_2,\ldots,i_k\}}} x_{i_1 i_2\ldots i_k} \leq 1, \forall j \in \{1,2,\ldots,n\} \tag{4.23}$$

$$x_{i_1 i_2\ldots i_k} \in \{0,1\}, \forall\{i_1,i_2,\ldots,i_k\} \in \mathcal{P}_k(n) \tag{4.24}$$

is feasible to (4.18).

Proof. All we need to show is that $\sum_{\substack{\{i_1,i_2,\ldots,i_k\}\in\mathcal{P}_k(n)\\ j\in\{i_1,i_2,\ldots,i_k\}}} x_{i_1 i_2\ldots i_k} = 1, \forall j \in \{1,2,\ldots,n\}$. Assume by contradiction that there is a j for which $\sum_{\substack{\{i_1,i_2,\ldots,i_k\}\in\mathcal{P}_k(n)\\ j\in\{i_1,i_2,\ldots,i_k\}}} x_{i_1 i_2\ldots i_k} = 0$ (there is no other possibility since x satisfies constraint (4.24); this means that there is a node j that is not in any k-ring). Since n is a multiplier of k, there are $k-1$ other nodes that are not in any k-ring, therefore, a new ring can be added to the sum contradicting the fact that x maximizes P'. $\square$

Lemma 2. *Assuming $c_{i_1 i_2 \dots i_k} > 0$ for each $\{i_1, i_2, \dots, i_k\} \in \mathcal{P}_k(n)$, a vector x which maximizes (4.22) maximizes (4.18).*

Proof. Straight from Lemma (1) and from the fact that any feasible vector in (4.18) is a feasible vector in (4.22). $\square$

However, not each solution to the relaxation yields a solution to the original problem. Consider the following example:

$$n = 6; k = 3 \tag{4.25}$$

$$c_{124} = c_{135} = c_{236} = c_{456} = 1;\ \text{all other } c_{ijl} = 0 \tag{4.26}$$

Then, $\max \sum c_{ijl} x_{ijl}$ subject to (4.19), (4.21) is obtained only for:

$$x_{124} = x_{135} = x_{236} = x_{456} = \frac{1}{2};\ \text{all other } x_{ijl} = 0 \tag{4.27}$$

4.4.2 Power method

In order to favour the emergence of a maximum which has exclusively integer coordinates, we introduce an exponent α. The purpose of this exponent is to penalize potential non-integer solutions.

For each $\alpha > 0$ we define:

$$P_\alpha = \max \sum_{\{i_1, i_2, \dots, i_k\} \in \mathcal{P}_k(n)} c_{i_1 i_2 \dots i_k} x^\alpha_{i_1 i_2 \dots i_k} \tag{4.28}$$

subject to

$$\sum_{\substack{\{i_1, i_2, \dots, i_k\} \in \mathcal{P}_k(n) \\ j \in \{i_1, i_2, \dots, i_k\}}} x_{i_1 i_2 \dots i_k} \leq 1, \forall j \in \{1, 2, \dots, n\} \tag{4.29}$$

$$x_{i_1 i_2 \dots i_k} \geq 0, \forall \{i_1, i_2, \dots, i_k\} \in \mathcal{P}_k(n) \tag{4.30}$$

We also define x_α a vector in $\mathbb{R}^{\binom{n}{k}}$ which maximizes (4.28) (can be one of many if there is more than one vector which maximizes P_α).

Lemma 3. *If $x_\alpha \in \{0,1\}^{\binom{n}{k}}$, then x_α maximizes (4.22).*

Proof. x_α is obviously feasible to (4.22). Assume by contradiction that there is y which satisfies (4.23) and (4.24) for which $P'(y) > P'(x_\alpha)$. Since all coordinates of x_α and y are 0 or 1, $P'(x_\alpha) = P_\alpha(x_\alpha)$ and $P'(y) = P_\alpha(y)$. Therefore, $P_\alpha(y) > P_\alpha(x_\alpha)$, contradicting the fact that x_α maximizes (4.28). □

Lemma 4. $\lim_{\alpha\to\infty} P_\alpha(x_\alpha) = \lim_{\alpha\to\infty} P_\alpha(\lfloor x_\alpha \rfloor)$

Proof.

$$\begin{aligned}\lim_{\alpha\to\infty} P_\alpha(x_\alpha) &= \lim_{\alpha\to\infty} \sum_{x_{i_1 i_2 \ldots i_k} \text{ coordinates of } x_\alpha} c_{i_1 i_2 \ldots i_k} x^\alpha_{i_1 i_2 \ldots i_k} \\ &= \lim_{\alpha\to\infty} \sum_{x_{i_1 i_2 \ldots i_k} \text{ coordinates of } x_\alpha} c_{i_1 i_2 \ldots i_k} \lfloor x_{i_1 i_2 \ldots i_k} \rfloor^\alpha = \lim_{\alpha\to\infty} P_\alpha(\lfloor x_\alpha \rfloor) \end{aligned} \quad (4.31)$$

The second equality derives from the fact that constraints (4.29) and (4.30) force that $0 \le x_{i_1 i_2 \ldots i_k} \le 1$. Hence:

- if $x_{i_1 i_2 \ldots i_k} = 1$, then $c_{i_1 i_2 \ldots i_k} x^\alpha_{i_1 i_2 \ldots i_k} = c_{i_1 i_2 \ldots i_k} \lfloor x_{i_1 i_2 \ldots i_k} \rfloor^\alpha = c_{i_1 i_2 \ldots i_k}$
- if $x_{i_1 i_2 \ldots i_k} < 1$, then $\lim_{\alpha\to\infty} c_{i_1 i_2 \ldots i_k} {x_{i_1 i_2 \ldots i_k}}^\alpha = 0$ and $c_{i_1 i_2 \ldots i_k} \lfloor x_{i_1 i_2 \ldots i_k} \rfloor^\alpha = 0$

□

Theorem 5. *Assuming $c_{i_1 i_2 \ldots i_k} > 0$ for each $\{i_1, i_2, \ldots, i_k\} \in \mathcal{P}_k(n)$, as $\alpha \to \infty$, $\lfloor x_\alpha \rfloor$ maximizes (4.18).*

Proof. From lemma (4), we conclude that $\lfloor x_\alpha \rfloor$ maximizes (4.28). Then, from lemma (3), we conclude that $\lfloor x_\alpha \rfloor$ maximizes (4.22). Finally, from lemma (2), we conclude that $\lfloor x_\alpha \rfloor$ maximizes (4.18). □

Theorem (5) offers an alternative way to solve (4.18) for a large α and select the floor values of the elements in x_α.

4.5 Closing Remarks

Availability is maximized when the number of rings is high and the ring size distribution is regular. In this Chapter, we show that the partition of a network including an aggregation node and n cellular sites into $\frac{n}{2}$ rings, each one including the aggregation node and 2 cellular sites, can be solved in a time of $O(n^3)$. Regarding a partition with larger rings, the problem is similar to a k-set partition problem, which is NP-hard for $k \ge 3$. We propose an approximation method, based on linear programming and use of an exponent aimed to accelerate the convergence.

Chapter 5

Polarization Planning for Wireless Networks

5.1 Overview

The wireless industry is presently facing a tremendous growth of demand for higher data rates, driven by the development of mobile data services. In order to meet this demand, a number of solutions are widely considered: network densification (including deployment of small cells [91], [92], [23]), use of wider spectrum [91], [92], increase of spectral efficiency [91], [23].

Polarization diversity has long been regarded as another mean to meet the growing demand for wireless spectrum [10], [18]. This solution is suitable for line-of-sight links only, because of the cross-polarization interferences resulting from reflections. Since conventional networks used only vertical polarization (VP), polarization diversity was achieved by adding a horizontal polarization (HP). Because of this historical context, polarization diversity was, in a large majority of cases, restricted to horizontal/vertical.

Since then, other options have been considered. The respective performances of HP/VP versus $\pm 45°$ slanted polarization have been compared, in order to determine the optimal pair of orthogonal polarizations [62], [63], [64]. These works came along with the increased interest polarization diversity has gained these recent years in frequency planning.

Polarization multiplexing in non-LOS channels has been studied in [65]. However, the proposed solution is relevant only for one-way transmission signals, since there is no reversibility principle for polarization. For this reason, the LOS assumption is necessary for polarization multiplexing in full-duplex links. This assumption could be considered as very restrictive, especially in the context of the MIMO technology: the worst case scenario occurs for compact arrays under LOS propagation conditions, where the rank of the channel matrix equals to 1 [66], [24]. However, recent works have highlighted that millimeter wave is the most relevant solution for massive MIMO [25], [26]. Due to the highly directional and quasi-optical nature of millimeter wave propagation, LOS propagation condition should be dominant in MIMO systems with millimeter wave communications [67], [68], [69].

Polarization multiplexing is presently seen as a method for doubling spectral efficiency on a single channel [70]. To the best of our knowledge, network coverage with one single channel by just optimizing the polarization directions has not been studied yet. In this Chapter, we show how an appropriate use of polarization diversity in ring topology networks improves spectrum efficiency.

While the current approach focuses on optimizing a point-to-point link, we propose a paradigm shift for polarization diversity: instead of restricting the choice to two predefined polarizations such as horizontal/vertical or $\pm 45°$, our approach, which will be called hereafter "inclined polarization", is based on polarizations fulfilling the condition that the polarization of any link is orthogonal to the polarization of the two adjacent links. This approach is suitable for all kinds of ring topology wireless networks whether the base stations are fixed or mobile and can use any kind of polarized antenna.

As it will be shown, the polarizations will be determined by the base stations' locations. Therefore, in a network with fixed base stations, our solution will not require any extra cost compared to networks using horizontal/vertical polarizations. A network with mobile base stations will require a centralized control system and polarization-agile antennas. Such antennas, which polarization state can be changed dynamically, have been studied in [93], [94].

Optimal synthesis of beam pattern having any state of polarization using an array of antennas has been addressed in [95], [96]. These polarization synthesis techniques can be used in our model for polarization planning with mobile base stations.

5.2 Polarization diversity

In an electromagnetic wave, electric field and magnetic field are oscillating in two directions orthogonal to the propagation direction and orthogonal to each other. If the fields rotate at the wave frequency, the polarization is circular or elliptical. If the fields oscillate in one single direction, the polarization is linear. By convention, the direction of a linear polarization is the direction of the electric field.

It should be noted that vertical polarization, in the strict sense of the word, is possible only if both base stations are at the same altitude. When antennas were located far away one from each other, a difference in altitude could be neglected relative to the distance and all the nodes of a network could roughly be considered to be co-planar. However, with the deployment of small cells, this difference in altitude will become more and more significant and vertical polarization will often reveal to be impossible.

Polarization diversity enables to reuse the same frequency in a network. In order to avoid interferences, the two polarizations received by a base station at the same frequency must be orthogonal to each other. In practice, due to rotational misalignment of antennas or weather conditions, the signal in one polarization may interfere with the other. However, it is possible to cancel this interference by using the XPIC (Cross-Polarization Interference Cancellation) algorithm [70].

In a chain comprising the nodes A, B and C, the polarization $\mathbf{E}_{AB}$ between A and B may be chosen arbitrarily (upon the condition it is orthogonal to the line AB), but then, the reuse of the same frequency requires that the polarization $\mathbf{E}_{BC}$ between B and C shall be orthogonal to $\mathbf{E}_{AB}$. Since $\mathbf{E}_{BC}$ shall also be orthogonal to the line BC, the direction of $\mathbf{E}_{AB}$ generally determines that of $\mathbf{E}_{BC}$.

This solution can be extended to any number of nodes, provided that all links are LOS: connecting all the nodes of a chain with one single channel is always possible by choosing each polarization orthogonal to the previous one and to the propagation line.

5.3 Polarization in ring networks

Notations: we use bold letters to denote matrices and vectors. $\|\mathbf{u}\|$ denotes the euclidean norm of vector $\mathbf{u}$. $\mathbf{u}||\mathbf{v}$ means that vectors $\mathbf{u}$ and $\mathbf{v}$ are parallel. The symbol $\times$ denotes the vector product. If $\triangle ABC$ is a triangle, we denote $\angle BAC$ the angle in A.

We assume in the following that all the transmissions are LOS and that the nodes are located far away enough one from each other, so that the signal received by a node does not cause interference to other nodes.

Since the polarization of a link generally determines the polarization of the next one, the use of one single frequency by polarization diversity in a ring topology network faces the question of closing the loop: the first polarization will determine the second one, which will determine the third one and so forth until the last one. The last polarization, which is determined by a sequence of constraints, is not necessarily orthogonal to the first one. In this section, we will study under what conditions it is possible to choose the first polarization in such a way that this orthogonality property is fulfilled.

5.3.1 Formulating the problem

Let us consider n nodes A_1, A_2,..., A_n. We define the following unitary vectors:

$$\mathbf{u_1} = \frac{\mathbf{A_1A_2}}{\|\mathbf{A_1A_2}\|}, \mathbf{u_2} = \frac{\mathbf{A_2A_3}}{\|\mathbf{A_2A_3}\|}, \ldots, \mathbf{u_n} = \frac{\mathbf{A_nA_1}}{\|\mathbf{A_nA_1}\|}$$

and the polarizations: $\mathbf{E_1}$ between A_1 and A_2, $\mathbf{E_2}$ between A_2 and A_3,..., $\mathbf{E_n}$ between A_n and A_1 (see Figure 5.1).

Since each polarization shall be orthogonal to the previous one and to the propagation direction, it must be parallel to the vector product of these two vectors:

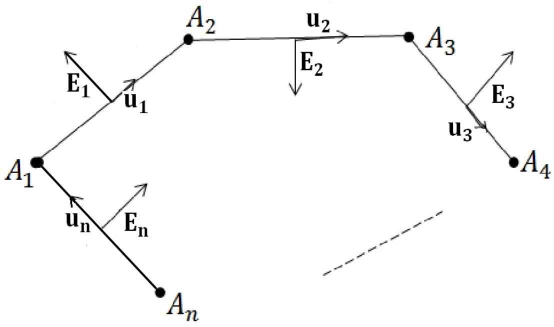

Figure 5.1: Ring network.

$\mathbf{E_1}||\mathbf{E_n} \times \mathbf{u_1}$

$\mathbf{E_2}||\mathbf{E_1} \times \mathbf{u_2}$

$\vdots$

$\mathbf{E_n}||\mathbf{E_{n-1}} \times \mathbf{u_n}$

Therefore, $\mathbf{E_n}$ must fulfill the following condition:

$$\mathbf{E_n}||((\mathbf{E_n} \times \mathbf{u_1}) \times \cdots \times \mathbf{u_{n-1}}) \times \mathbf{u_n} \tag{5.1}$$

Let us define n vector planes: P_1 orthogonal to $\mathbf{u_1}$, P_2 orthogonal to $\mathbf{u_2}$,..., P_n orthogonal to $\mathbf{u_n}$. We can now define the following n homomorphisms (for simplicity of notation, we put $P_0 = P_n$):

$$\phi_i : P_{i-1} \to P_i$$
$$\mathbf{x} \mapsto \mathbf{x} \times \mathbf{u_i}$$

Then, $\phi = \phi_n \phi_{n-1} \dots \phi_1$ is an endomorphism in P_n.

The above condition (5.1) can be written:

$$\mathbf{E_n}||\phi(\mathbf{E_n}) \tag{5.2}$$

Therefore, the problem can be expressed as the search of an eigenvector for the endomorphism ϕ.

5.3.2 Matrix expression

Condition for the existence of an eigenvector

It is well known that the eigenvalues of an endomorphism ϕ in a vector plane are the solutions of the equation:

$$X^2 - tr(\phi)X + det(\phi) = 0 \tag{5.3}$$

where $tr(\phi)$ and $det(\phi)$ are the trace and the determinant of ϕ, respectively.

This equation has solutions if and only if:

$$tr(\phi)^2 - 4det(\phi) \geq 0 \tag{5.4}$$

Choosing an appropriate basis for each plane

For each vector plane P_i, we define two vectors $\mathbf{v_i}$ and $\mathbf{w_i}$ fulfilling the following conditions:

- $\mathbf{v_i}$ is in the plane containing A_i, A_{i+1} and A_{i+2} (for simplicity of notation, we put $A_{n+1} = A_1$ and $A_{n+2} = A_2$).
- ($\mathbf{u_i}$,$\mathbf{v_i}$,$\mathbf{w_i}$) is a direct orthonormal basis of the space.

Note that the choice of $\mathbf{v_i}$ is not unique: two opposite vectors fulfill the condition. Anyone of both can be chosen arbitrarily.

We also define in P_i two vectors $\mathbf{v'_i}$ and $\mathbf{w'_i}$ fulfilling the following conditions:

- $\mathbf{w'_i} = \mathbf{w_{i-1}}$ (for simplicity of notation, we put $\mathbf{w_0} = \mathbf{w_n}$).
- ($\mathbf{u_i}$,$\mathbf{v'_i}$,$\mathbf{w'_i}$) is a direct orthonormal basis of the space.

  The plane containing A_{i-1}, A_i and A_{i+1} is represented in Figure 5.2 with the bases ($\mathbf{u_{i-1}}$,$\mathbf{v_{i-1}}$,$\mathbf{w_{i-1}}$) and ($\mathbf{u_i}$,$\mathbf{v'_i}$,$\mathbf{w'_i}$).

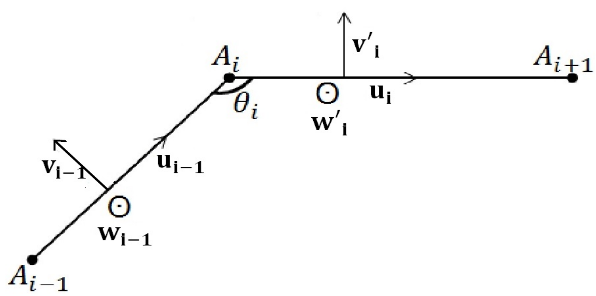

Figure 5.2: Vector basis.

Calculation of the matrix

B_i=($\mathbf{v_i}$,$\mathbf{w_i}$) and B'_i=($\mathbf{v'_i}$,$\mathbf{w'_i}$) are two orthonormal bases of P_i. Since ($\mathbf{u_i}$,$\mathbf{v_i}$,$\mathbf{w_i}$) and ($\mathbf{u_i}$,$\mathbf{v'_i}$,$\mathbf{w'_i}$) are both direct orthonormal bases of the space, the change of basis matrix from B_i to B'_i is a rotation matrix:

$$\mathbf{R_i} = \begin{pmatrix} \cos\alpha_i & -\sin\alpha_i \\ \sin\alpha_i & \cos\alpha_i \end{pmatrix} \tag{5.5}$$

where α_i is the rotation angle of matrix $\mathbf{R_i}$.

We define the angle $\theta_i = \angle A_{i-1}A_iA_{i+1}$.

Since $\mathbf{v_{i-1}} \times \mathbf{u_i} = \sin(\theta_i + \frac{\pi}{2})\mathbf{w'_i} = \cos\theta_i \mathbf{w'_i}$ and $\mathbf{w_{i-1}} \times \mathbf{u_i} = \mathbf{v'_i}$, the matrix of ϕ_i with respect to the bases B_{i-1} and B'_i is:

$$\mathbf{M_i} = \begin{pmatrix} 0 & 1 \\ \cos\theta_i & 0 \end{pmatrix} \tag{5.6}$$

Therefore, the matrix of ϕ in B_n is:

$$\mathbf{M} = \mathbf{R_n M_n R_{n-1} M_{n-1} \ldots R_2 M_2 R_1 M_1} \tag{5.7}$$

5.3.3 Solution of the problem

All the nodes are in the same plane

If all the nodes are in the same plane, it is possible to choose B_i and B'_i so that $B_i = B'_i$ for all i. Then, all the rotation matrices equal the identity matrix. Equation (5.7) becomes:

$$\mathbf{M} = \mathbf{M_n M_{n-1} \ldots M_2 M_1} \tag{5.8}$$

First case: the number of nodes is even
If n is an even number, equations (5.6) and (5.8) give:

$$\mathbf{M} = \begin{pmatrix} \cos\theta_1\cos\theta_3\ldots\cos\theta_{n-1} & 0 \\ 0 & \cos\theta_2\cos\theta_4\ldots\cos\theta_n \end{pmatrix} \tag{5.9}$$

In this case, the eigenvectors are obviously $\mathbf{v_n}$ and $\mathbf{w_n}$. This matches the intuitive solution of choosing polarization in the plane containing all the nodes and orthogonal to this plane, alternately.

In addition, in the particular case where $\cos\theta_1\cos\theta_3\ldots\cos\theta_{n-1} = \cos\theta_2\cos\theta_4\ldots\cos\theta_n$, then ϕ is a homothety. Any vector is an eigenvector and the first polarization can be chosen arbitrarily.

Second case: the number of nodes is odd
If n is an odd number, equations (5.6) and (5.8) give:

$$\mathbf{M} = \begin{pmatrix} 0 & \cos\theta_2 \cos\theta_4 \ldots \cos\theta_{n-1} \\ \cos\theta_1 \cos\theta_3 \ldots \cos\theta_n & 0 \end{pmatrix} \tag{5.10}$$

Then, $tr(\phi) = tr(M) = 0$,

and $det(\phi) = det(\mathbf{M}) = -\cos\theta_1 \cos\theta_2 \ldots \cos\theta_n$.

Equation (5.3) becomes:

$$X^2 - \cos\theta_1 \cos\theta_2 \ldots \cos\theta_n = 0 \tag{5.11}$$

This equation has a solution if and only if $\cos\theta_1 \cos\theta_2 \ldots \cos\theta_n \geq 0$, which means if and only if the number or obtuse angles in the polygon is even.

Since the total number of angles in the polygon is odd, we can express a necessary and sufficient condition:

Equation (5.3) has a solution if and only if the number of acute angles in the polygon is odd.

The eigenvectors coordinates in B_n can be easily calculated:

$$\mathbf{E_n} = \lambda \begin{pmatrix} \sqrt{|\cos\theta_2 \cos\theta_4 \ldots \cos\theta_{n-1}|} \\ \pm\sqrt{|\cos\theta_1 \cos\theta_3 \ldots \cos\theta_n|} \end{pmatrix}, \lambda \in \mathbb{R} \tag{5.12}$$

General case

If we do not assume that all the nodes are in the same plane, the calculation of $\mathbf{M}$ from equation (5.7) is much more complicated and $tr(\phi)$ cannot be calculated easily. However, as a consequence of equation (5.4), if $det(\phi) \leq 0$, the endomorphism ϕ has eigenvalues. Since the determinants of rotation matrices $\mathbf{R_i}$ equal to 1, we can still calculate $det(\phi)$:

$$det(\phi) = det(\mathbf{M}) = (-1)^n \cos\theta_1 \cos\theta_2 \ldots \cos\theta_n \tag{5.13}$$

It is therefore possible to express a sufficient, though not necessary, condition for the existence of a solution to equation (5.3):

If the number of acute angles in the polygon is odd, then $det(\phi) \leq 0$ and equation (5.3) has a solution.

Particular case of 4 nodes

Theorem 6. *If $n = 4$, equation (5.3) has always a solution.*

Proof. Without loss of generality, we may assume that the nodes have the following coordinates:

$$A_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, A_2 \begin{pmatrix} a \\ b \\ 0 \end{pmatrix}, A_3 \begin{pmatrix} c \\ d \\ e \end{pmatrix}, A_4 \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

According to the notations above,

$$\mathbf{u_4} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v_4} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{w_4} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

We obtain, after calculation:

$$(((\mathbf{v_4} \times \mathbf{A_1A_2}) \times \mathbf{A_2A_3}) \times \mathbf{A_3A_4}) \times \mathbf{A_4A_1} = (a-1) \begin{pmatrix} 0 \\ d(b-d) + c(a-c) \\ e(b-d) \end{pmatrix} \tag{5.14}$$

and

$$(((\mathbf{w_4} \times \mathbf{A_1A_2}) \times \mathbf{A_2A_3}) \times \mathbf{A_3A_4}) \times \mathbf{A_4A_1} = \begin{pmatrix} 0 \\ (1-a)ed + bec \\ (1-a)(e^2 + c(c-a)) + bc(b-d) \end{pmatrix} \tag{5.15}$$

Therefore, $M = \frac{1}{\|\mathbf{A_1A_2}\|\|\mathbf{A_2A_3}\|\|\mathbf{A_3A_4}\|\|\mathbf{A_4A_1}\|} A$, with

$$A = \begin{pmatrix} (a-1)(d(b-d) + c(a-c)) & (1-a)ed + bec \\ (a-1)e(b-d) & (1-a)(e^2 + c(c-a)) + bc(b-d) \end{pmatrix} \tag{5.16}$$

$$tr(A)^2 - 4det(A) = ((a-1)(d(b-d)+e^2) - bc(b-d))^2 + 4e^2(a-1)(b-d)((1-a)d+bc) \tag{5.17}$$

$$tr(A)^2 - 4det(A) = \big((a-1)(d(b-d) - e^2) - bc(b-d)\big)^2 \geq 0 \tag{5.18}$$

Therefore, A has eigenvectors and ϕ has eigenvectors. This proves that it is possible to use one single frequency in any ring topology network including four nodes.

□

Summarized results

The results obtained above are summarized in Table 5.1.

Nodes	**In the same plane**	**Not in the same plane**
Odd number	Solution if and only if there is an odd number of acute angles.	If the number of acute angles is odd, then there is a solution.
Even number≥ 6	There is a solution.	
4 nodes	There is a solution.	

Table 5.1: One single frequency in a ring network.

5.3.4 Special cases

Below are three application cases for the results we obtained.

Ring topology with a right angle

If there is a right angle in the ring, e.g. in A_i, then $det(M_i) = 0$ and therefore $det(\phi) = 0$. Equation (5.3) has a solution.

One can build a solution by choosing $\mathbf{E_i} = \mathbf{u_{i-1}}$ and then applying successively ϕ_{i+1},..., ϕ_n, ϕ_1,...,ϕ_{i-1}. The last vector will be in P_{i-1} and therefore will be orthogonal to $\mathbf{E_i}$.

Triangle topology

As a result of the "odd number of acute angles" condition, equation (5.3) has a solution if and only if the three angles of the triangle are acute.

The case of the triangle is the best illustration of the advantages of the solution proposed in this Chapter over the present state of the art: if B is the available bandwidth, the use of horizontal/vertical polarization enables to double the bandwidth, and since the same frequency cannot be used on two adjacent links with the same type of polarization, it is possible to allocate up to $\frac{2}{3}B$ to each link. By defining three polarizations orthogonal to each other, our solution enables to allocate B to each link.

As an example, the polarizations in the case of an equilateral triangle are given below (see Figure 5.3).

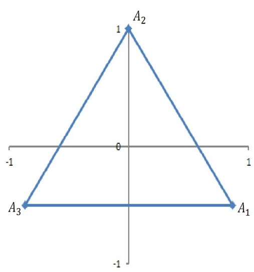

Figure 5.3: Equilateral triangle.

With the notations above, we obtain the following polarizations:

$$\mathbf{E_1} = \begin{pmatrix} -1/\sqrt{2} \\ -1/\sqrt{6} \\ \pm 1/\sqrt{3} \end{pmatrix} \mathbf{E_2} = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{6} \\ \pm 1/\sqrt{3} \end{pmatrix} \mathbf{E_3} = \begin{pmatrix} 0 \\ \sqrt{2/3} \\ \pm 1/\sqrt{3} \end{pmatrix}$$

Regular pentagon topology

In a convex regular pentagon, all angles are obtuse. The ”odd number of acute angles” condition is not fulfilled and polarization diversity does not enable the use of one single frequency over the ring. However, if the pentagon vertices are connected according to a sheriff star, then all the angles are acute and the use of one single frequency is possible (see Figure 5.4).

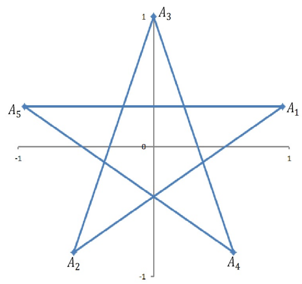

Figure 5.4: Sheriff star.

The method described above enables to calculate the polarizations:

$$\mathbf{E_1} = \begin{pmatrix} \pm 0.437 \\ \mp 0.602 \\ 0.669 \end{pmatrix} \mathbf{E_2} = \begin{pmatrix} -0.707 \\ 0.230 \\ \pm 0.669 \end{pmatrix} \mathbf{E_3} = \begin{pmatrix} \pm 0.707 \\ \pm 0.230 \\ 0.669 \end{pmatrix}$$

$$\mathbf{E_4} = \begin{pmatrix} -0.437 \\ -0.602 \\ \pm 0.669 \end{pmatrix} \mathbf{E_5} = \begin{pmatrix} 0 \\ 0.743 \\ \pm 0.669 \end{pmatrix}$$

Another solution also enables the use of one single frequency: a pentagon with three long sides and two short sides (see Figure 5.5).

The method described above enables to calculate the polarizations:

$$\mathbf{E_1} = \begin{pmatrix} \pm 0.309 \\ \mp 0.1 \\ 0.292 \end{pmatrix} \mathbf{E_2} = \begin{pmatrix} -0.278 \\ -0.09 \\ \pm 0.263 \end{pmatrix} \mathbf{E_3} = \begin{pmatrix} \pm 0.25 \\ \mp 0.081 \\ 0.236 \end{pmatrix}$$

$$\mathbf{E_4} = \begin{pmatrix} -0.225 \\ -0.073 \\ \pm 0.213 \end{pmatrix} \mathbf{E_5} = \begin{pmatrix} 0 \\ 0.946 \\ \pm 0.325 \end{pmatrix}$$

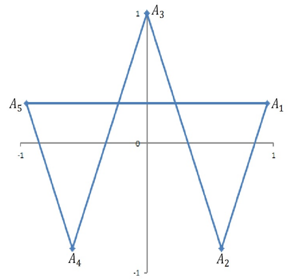

Figure 5.5: Pentagon with 3 long sides and 2 short sides.

In a pentagon, the use of horizontal/vertical polarization enables to allocate at best 80% of the available bandwidth to each link. An example of optimal allocation can be obtained by dividing the available band into 5 equal bands B_1, B_2, B_3, B_4, B_5 and performing the allocation shown on Table 5.2.

Link	Frequency bands	Polarizations
A_1A_2	B_1, B_4	Horizontal/Vertical
A_2A_3	B_2, B_5	Horizontal/Vertical
A_3A_4	B_3, B_1	Horizontal/Vertical
A_4A_5	B_4, B_2	Horizontal/Vertical
A_5A_1	B_5, B_3	Horizontal/Vertical

Table 5.2: Frequency and polarization allocation in a pentagon.

In comparison, our solution enables to allocate the whole bandwidth to each link.

More generally, for any odd number n, in an n-sided polygon, the use of horizontal/vertical polarization enables to allocate at best $\frac{n-1}{n}B$ to each link, B being the available bandwidth. When our solution is applicable, it enables to allocate the whole bandwidth to each link.

In an even-number sided polygon, the use of horizontal/vertical polarization also enables to allocate the whole bandwidth to each link by sharing the band into two halves and allocating alternately to each link the lower part and the upper part, always with two polarizations. However, since the frequencies in the neighborhood of the band bounds are often disturbed in practice by interferences from the contiguous band, our solution should be preferable though the advantage is less evident than that of the odd-number sided polygon case.

5.4 Spectral efficiency

The benefits of inclined polarization can be highlighted by its consequences in terms of spectral efficiency. We will hereafter compare the performances of polarization multiplexing and inclined polarization in the two ring topologies described above: triangle and regular pentagon.

According to the Shannon-Hartley theorem, the maximum bit rate which can be transmitted over a channel is given by the expression:

$$C = B\ \log_2(1 + \frac{S}{N}) \tag{5.19}$$

where

C is the channel capacity in bit/s;

B is the channel's bandwidth in Hertz;

S is the average signal power over the bandwidth, in Watt;

N is the average noise or interference power over the bandwidth, in Watt.

Classical polarization multiplexing, such as horizontal/vertical, enables frequency reuse. On the other hand, it creates some cross-polarization interference which will not be totally suppressed by the XPIC algorithm. The ultra-high-performance antenna discrimination between two orthogonal polarizations is typically close to $40\,dB$ [10]. We will retain in the following a cross-polarization interference of $-30\,dB$. Therefore, an interference noise of λS will be added to the thermal noise N, with $\lambda = 10^{-3}$.

By comparison to classical polarization multiplexing, inclined polarization enhances the frequency reuse possibilities for the same cross-polarization interference. However, if a change in the network topology is required, inclined polarization will cause higher distances and therefore lower received power. Therefore, the spectral efficiency comparison for the regular pentagon depends on the propagation model.

In the following, we will assume that the path attenuation is proportional to d^α, d being the distance and α a real number greater or equal to 2. Since all the transmissions are assumed to be LOS, attenuation due to vegetation and obstacles can be ignored. However, attenuation due to gas absorption and precipitation can be added to the free-space loss. Therefore, in order to evaluate the impact of the propagation model on the relevance of our solution, we will perform the comparisons in the scope of two different assumptions:

- $\alpha = 2$ (free-space loss model);
- $\alpha = 3$.

5.4.1 Triangle topology

No polarization

Since the same frequency cannot be used in two adjacent links, only one third of the bandwidth is available for each link. Therefore, the spectral efficiency is:

$$\frac{C}{B} = \frac{1}{3}\log_2(1 + \frac{S}{N}) \tag{5.20}$$

Polarization multiplexing

Polarization multiplexing doubles the bandwidth available for each link. Therefore:

$$\frac{C}{B} = \frac{2}{3}\log_2(1 + \frac{S}{N + \lambda S}) = \frac{2}{3}\log_2(1 + \frac{S/N}{1 + \lambda S/N}) \tag{5.21}$$

Inclined polarization

Inclined polarization enables the use of all the available bandwidth:

$$\frac{C}{B} = \log_2(1 + \frac{S}{N + \lambda S}) = \log_2(1 + \frac{S/N}{1 + \lambda S/N}) \tag{5.22}$$

As a result of Equations (5.20), (5.21) and (5.22), for a given signal-to-noise ratio, the spectral efficiency is almost doubled with horizontal/vertical polarization as long as the dominant noise is the thermal noise. When the signal-to-noise ratio tends to infinity, the use of polarization is not relevant anymore.

As shown on Figure 5.6, inclined polarization improves by 50% the spectral efficiency offered by horizontal/vertical polarization. This improvement does not depend upon the propagation model.

5.4.2 Regular pentagon topology

In order to compare the spectral efficiencies of the various strategies, we will assume that the power emitted by each node is such that the received power at the adjacent node of the convex pentagon is S.

In a pentagon, the best frequency allocation strategy is given in Table 5.2. It enables to allocate two fifths of the total bandwidth without polarization and four fifths of the total bandwidth with polarization multiplexing.

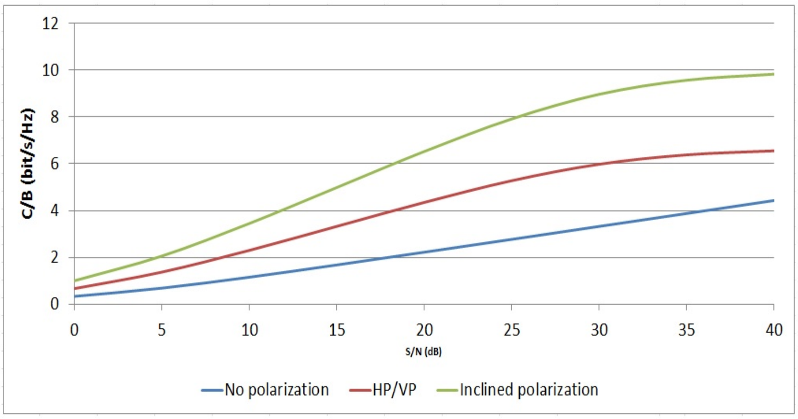

Figure 5.6: Spectral efficiency in a triangle.

No polarization

$$\frac{C}{B} = \frac{2}{5}\log_2(1+\frac{S}{N}) \tag{5.23}$$

Polarization multiplexing

$$\frac{C}{B} = \frac{4}{5}\log_2(1+\frac{S}{N+\lambda S}) = \frac{4}{5}\log_2(1+\frac{S/N}{1+\lambda S/N}) \tag{5.24}$$

Inclined polarization - Sheriff star

Taking the notations of Figure 5.4, we first calculate the ratio between the side of a sheriff star and the side of a regular pentagon:

$$\frac{A_1A_2}{A_1A_4} = \frac{\sin\frac{3\pi}{5}}{\sin\frac{\pi}{5}} = 1 + 2\cos\frac{2\pi}{5} = \frac{\sqrt{5}+1}{2} \tag{5.25}$$

Therefore, for the same emitting power, the ratio between the received power in a regular pentagon and the received power in a sheriff star is:

$$\frac{A_1{A_2}^{\alpha}}{A_1{A_4}^{\alpha}} = \left(\frac{\sqrt{5}+1}{2}\right)^{\alpha} \tag{5.26}$$

As a result, the spectral efficiency is:

$$\frac{C}{B} = \log_2\left(1 + \frac{\left(\frac{2}{\sqrt{5}+1}\right)^{\alpha} S/N}{1 + \lambda\left(\frac{2}{\sqrt{5}+1}\right)^{\alpha} S/N}\right) \tag{5.27}$$

- If $\alpha = 2$ and $\lambda = 10^{-3}$, this equation becomes:

$$\frac{C}{B} = \log_2\left(1 + \frac{\frac{2}{\sqrt{5}+3} S/N}{1 + 10^{-3}\frac{2}{\sqrt{5}+3} S/N}\right) \tag{5.28}$$

- If $\alpha = 3$ and $\lambda = 10^{-3}$, this equation becomes:

$$\frac{C}{B} = \log_2\left(1 + \frac{\frac{1}{\sqrt{5}+2} S/N}{1 + 10^{-3}\frac{1}{\sqrt{5}+2} S/N}\right) \tag{5.29}$$

Inclined polarization - Pentagon with 3 long sides and 2 short sides

In order to obtain a given spectral efficiency $\frac{C}{B}$, the required signal-to-noise ratio is different for a long side and for a short side (see Figure 5.5):

Long side:

$$\left(\frac{S}{N}\right)_{long} = \left(\frac{\sqrt{5}+1}{2}\right)^{\alpha} \frac{2^{\frac{C}{B}} - 1}{1 + \lambda - \lambda 2^{\frac{C}{B}}} \tag{5.30}$$

Short side:

$$\left(\frac{S}{N}\right)_{short} = \frac{2^{\frac{C}{B}} - 1}{1 + \lambda - \lambda 2^{\frac{C}{B}}} \tag{5.31}$$

Therefore, the average signal-to-noise ratio is:

$$\left(\frac{S}{N}\right)_{average} = \frac{1}{5}\left(3\left(\frac{S}{N}\right)_{long} + 2\left(\frac{S}{N}\right)_{short}\right) \tag{5.32}$$

$$\left(\frac{S}{N}\right)_{average} = \frac{1}{5}\left(3\left(\frac{\sqrt{5}+1}{2}\right)^{\alpha} + 2\right) \frac{2^{\frac{C}{B}} - 1}{1 + \lambda - \lambda 2^{\frac{C}{B}}} \tag{5.33}$$

The spectral efficiency can now be expressed as a function of the average signal-to-noise ratio:

$$\frac{C}{B} = \log_2\left(1 + \frac{\frac{5}{3\left(\frac{\sqrt{5}+1}{2}\right)^{\alpha}+2}\left(\frac{S}{N}\right)_{average}}{1 + \lambda\frac{5}{3\left(\frac{\sqrt{5}+1}{2}\right)^{\alpha}+2}\left(\frac{S}{N}\right)_{average}}\right) \tag{5.34}$$

- If $\alpha = 2$ and $\lambda = 10^{-3}$, this equation becomes:

$$\frac{C}{B} = \log_2 \left(1 + \frac{\frac{10}{3\sqrt{5}+13} \left(\frac{S}{N}\right)_{average}}{1 + 10^{-3} \frac{10}{3\sqrt{5}+13} \left(\frac{S}{N}\right)_{average}} \right) \tag{5.35}$$

- If $\alpha = 3$ and $\lambda = 10^{-3}$, this equation becomes:

$$\frac{C}{B} = \log_2 \left(1 + \frac{\frac{5}{3\sqrt{5}+8} \left(\frac{S}{N}\right)_{average}}{1 + 10^{-3} \frac{5}{3\sqrt{5}+8} \left(\frac{S}{N}\right)_{average}} \right) \tag{5.36}$$

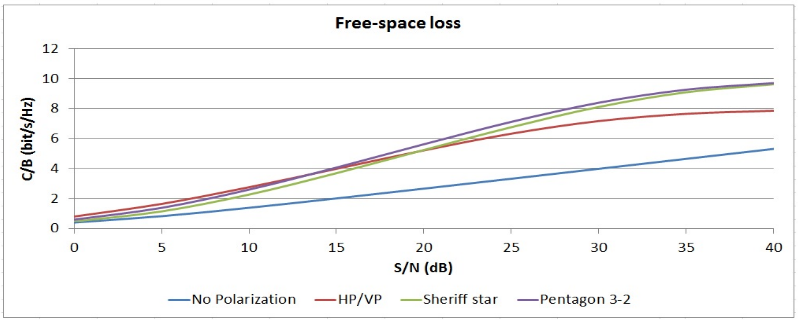

Figure 5.7: Spectral efficiency in a pentagon - Free-Space Loss.

As shown on Figure 5.7, for $\alpha = 2$, the sheriff star is a more relevant solution than the horizontal/vertical polarization if the signal-to-noise ratio is greater than $19.7\,dB$. The pentagon with three long sides and two short sides is the best solution if the signal-to-noise ratio is greater than $13.5\,dB$.

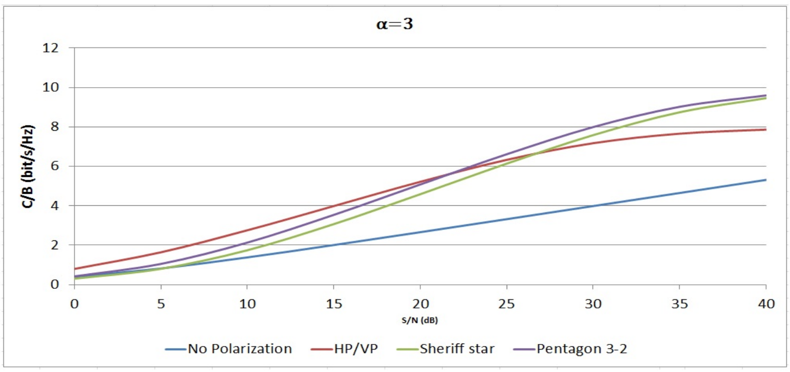

Figure 5.8: Spectral efficiency in a pentagon - $\alpha = 3$.

As shown on Figure 5.8, for $\alpha = 3$, the sheriff star is a more relevant solution than the horizontal/vertical polarization if the signal-to-noise ratio is greater than $26.8\,dB$. The pentagon with three long sides and two short sides is the best solution if the signal-to-noise ratio is greater than $21.8\,dB$.

In any case, the pentagon with three long sides and two short sides always offers a better spectral efficiency than the sheriff star.

5.5 Channel Allocation

Channel allocation in a wireless network is usually considered as a matter of graph coloring. To each channel corresponds a color. Allocating channels in a network while avoiding channel interferences is equivalent to coloring the edges of a graph while ensuring that two edges having a common vertex are not of the same color. As long as the use of polarization is restricted to HP/VP, this paradigm is not affected: polarization remains a black box which doubles the number of channels and channel allocation is still a matter of graph coloring.

However, the use of polarization we propose in this Chapter has implications on channel allocation. These implications will be illustrated by the following example.

Let us consider nine nodes A, B, C, A', B', C', A'', B'' and C''. We assume that the triangles ABC, $A'B'C'$, $A''B''C''$, $AA'A''$, $BB'B''$ and $CC'C''$ are connected and that each one of these triangles has three acute angles.

We also assume that all nodes are located far enough away one from each other to avoid any interference between two links which do not share a common node (see Figure 5.9).

For the sake of clarity, triangles ABC, $A'B'C'$ and $A''B''C''$ are represented in black, triangle $AA'A''$ is represented in green, triangle $BB'B''$ is represented in purple and triangle $CC'C''$ is represented in orange. These colors are not related to the concept of graph coloring.

Theorem 7. *Covering the network of Figure 5.9 by a graph coloring approach requires at least 5 channels.*

Proof. Let us assume that the network is covered by 4 channels, a, b, c and d. Since each one of the 9 vertices of the network belongs to 4 edges, each one of the 4 channels, a, b, c and d, must be used by exactly one of these 4 edges.

Triangle ABC requires three different channels. Without any loss of generality, we can assume that a is the channel used for the edge BC, b the channel used for the edge AC, and c the channel used for the edge AB.

Then, the two edges AA' and AA'' must use the channels a and d. As well, the two edges BB' and BB'' must use the channels b and d and the two edges CC' and CC'' must use the channels c and d.

Therefore, channel d must be used at least twice between triangle ABC and one of the

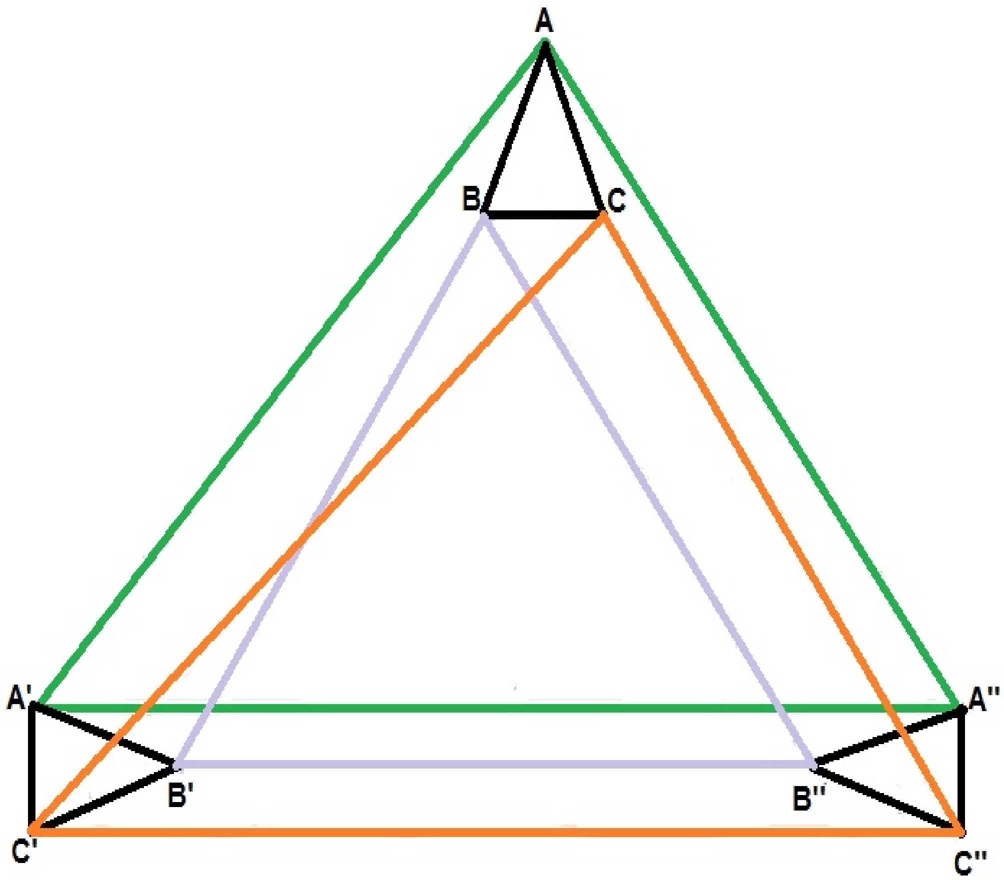

Figure 5.9: Example of wireless network.

triangles $A'B'C'$ or $A''B''C''$. Without any loss of generality, we can assume that channel d is used for the edges AA'' and BB''. As a result, channel a is used for the edge AA' and channel b is used for the edge BB'.

If channel d is allocated to the edge CC', then no edge containing C'' can use channel d: $C''C$ and $C''C'$ because channel d is already used for $C'C''$, and $C''A''$ (resp. $C''B''$) because channel d is already used for AA'' (resp. BB''). This contradicts the fact that each one of the four channels must be used by exactly one of the four edges containing C''. Therefore, it is necessary to allocate channel c to CC' and channel d to CC''.

Then, since channel d is already allocated to AA'', it cannot be allocated to any of the edges AA' or $A'A''$. Therefore, it must be allocated to one of the two edges $A'B'$ or $A'C'$.

Likewise, channel d must be allocated to one of the two edges $B'A'$ or $B'C'$, and one of the two edges $C'A'$ or $C'B'$. This implies it must be used twice in the triangle $A'B'C'$. This is impossible.

□

Therefore at least 5 channels (and therefore at least 3 frequencies if polarization diversity is restricted to HP/VP) are required in order to cover the network. Figure 5.10 provides an example of such a covering.

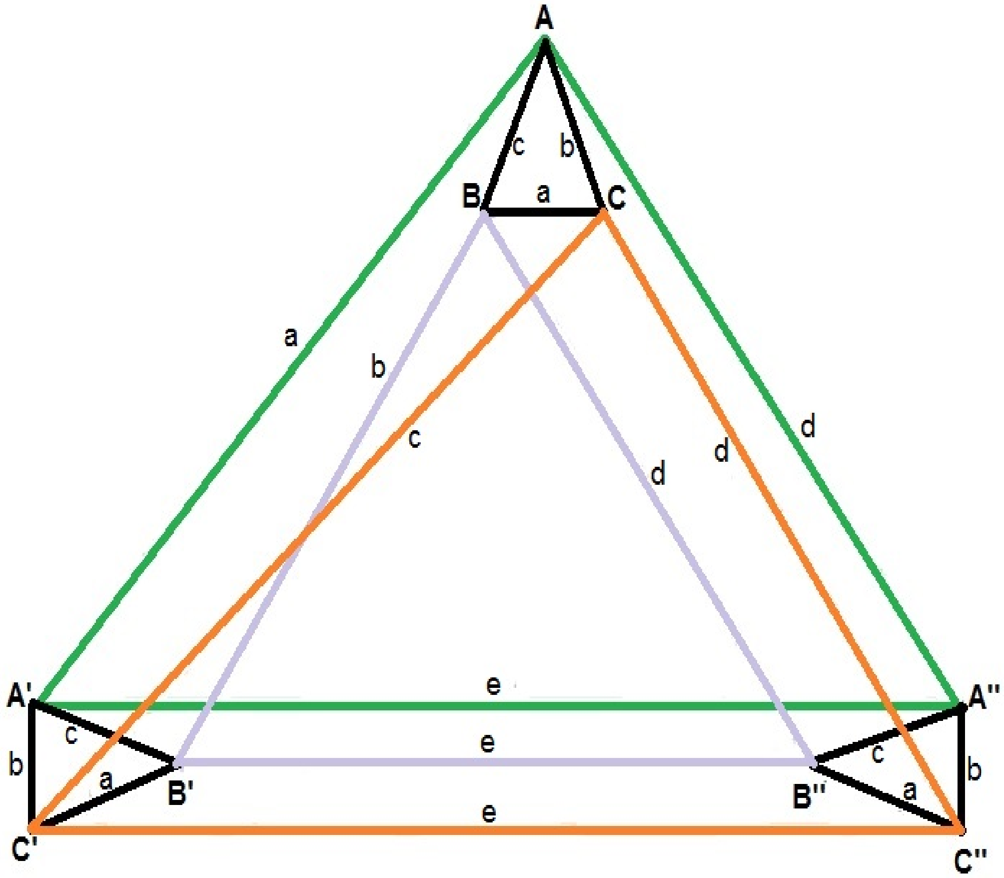

Figure 5.10: Channel allocation in a wireless network.

By comparison, the use of inclined polarization enables network coverage with 2 frequencies only: one frequency for the triangles ABC, $A'B'C'$ and $A''B''C''$, and another frequency for the triangles $AA'A''$, $BB'B''$ and $CC'C''$.

5.6 Closing remarks

Polarization diversity enables significant improvements regarding frequency allocation in various ring topology networks. In this Chapter, we show that the appropriate polarizations are the eigenvectors of an endomorphism. We give a necessary and sufficient condition for the existence of such a solution in the case all the base stations are on the same plane and a sufficient condition in the general case. We show that the appropriate polarization on a given link of a ring depends upon the position of all the nodes of the ring. The choice of polarization is not restricted to vertical and horizontal: inclined polarization can be an efficient mean to improve spectrum efficiency, even when all the base stations of a network are on a horizontal plane. It is always the best option when compatible with the network topology and can be considered as an alternative even if a change in the network topology is required. As shown in the examples above, inclined polarization is particularly efficient in triangles, where it can improve the spectrum efficiency by 50% in comparison with horizontal/vertical polarization and it is still fairly efficient in a number of other configurations. Furthermore, polarization diversity brings a new paradigm regarding channel allocation in a wireless network. Beyond being a black box which doubles the number of available channels, polarization diversity enables much more significant improvements on resource use. Further work still has to be done regarding topology optimization and channel allocation taking into account the opportunities inclined polarization can offer as described in the present Chapter.

Chapter 6

Fairness-Efficiency Trade-off in Network Resource Allocation

6.1 Overview

The mobile telecommunication sector is presently enjoying a tremendous growth, and this trend is expected to continue in the foreseen future. According to an IDATE report [97], global mobile subscription should grow from 5,328 million in 2010 to 9,684 million in 2020 (+81.8%), while total worldwide mobile traffic should grow from 3.8 Ebyte in 2010 to 127 Ebyte in 2020 (x33).

This tremendous growth will inevitably cause saturation and congestion problems. In spite of the numerous solutions which are considered in order to meet the growing demand, there is no doubt that the question of resource allocation will become a key issue in the next years, especially in the scope of the introduction of 5G networks.

In a typical wireless network, a number of Base Stations are connected to an aggregation node, which is wirily connected to the main network. As long as spectral resources were abundant, the main constraint of the wireless network was the total rate, which cannot exceed the maximal rate supported by the wired connection of the aggregation node to the network. However, with the introduction of 5G networks and massive use of small cells, spectral resources will become scarcer and scarcer. As a result, the interferences can become in many cases the limiting factor for resource allocation.

In the following, we present the existing tools for measuring fairness and network efficiency, and propose a new approach, based on fairness-efficiency trade-off, for resource allocation.

6.1.1 Fairness

Various approaches have been proposed for fairness in allocation of a single type of resource. Let $x = (x_1, ..., x_n)$ be a resource allocation vector. A variety of metrics, such as the ratio between the smallest and the largest allocations (min-max ratio), Jain's index, or proportional fairness have been proposed:

Min-max ratio

The min-max ratio is the ratio between the lowest allocation and the highest allocation.

$$\min - \max(x_1, ...x_n) = \frac{\min_i x_i}{\max_i x_i} \tag{6.1}$$

Jain's index

$$J(x_1, x_2, ..., x_n) = \frac{(\sum_{i=1}^{n} x_i)^2}{n \sum_{i=1}^{n} x_i^2} \tag{6.2}$$

Jain's index ranges from $\frac{1}{n}$ in the worst case, when all x_i but one equal to 0, to 1 in the best case, when all x_i are equal.

By noting $\hat{x}$ and $\hat{\sigma_x}^2$ the empirical mean and the empirical squared error of x, respectively, Jain's index can be expressed as a function of these two parameters:

$$J(x_1, x_2, ..., x_n) = \frac{1}{1 + \frac{\hat{\sigma_x}^2}{\hat{x}^2}} \tag{6.3}$$

Proportional fairness

Proportional fairness is based upon the assumption that the users' utility functions are logarithmic. Resource allocations x_i^* are proportionally fair if they maximize $\sum_{i=1}^{n} \log x_i$.

In weighted proportional fairness, the optimal allocation is obtained by maximizing $\sum_{i=1}^{n} p_i \log x_i$.

The concept of proportional fairness has been generalized in [98], which introduces the definition of (p, α)-proportional fairness: a feasible resource vector x^* is (p, α)-proportionally fair if for any feasible vector x, $\sum_{i=1}^{n} p_i \frac{x_i - x_i^*}{{x_i^*}^{\alpha}} \leq 0$.

It is shown in [98] that (p, α)-proportional fair resource allocation is equivalent to the maximization of the function:

$$f_\alpha(x) = \begin{cases} \ln x, & \text{if } \alpha = 1 \\ \frac{x^{1-\alpha}}{1-\alpha}, & \text{otherwise} \end{cases} \tag{6.4}$$

This definition can be slightly modified in order to have a family of functions continuous in α:

$$f_\alpha(x) = \int_1^x \frac{dt}{t^\alpha} = \begin{cases} \ln x, & \text{if } \alpha = 1 \\ \frac{x^{1-\alpha}-1}{1-\alpha}, & \text{otherwise} \end{cases} \tag{6.5}$$

Entropy

Entropy was introduced by Shannon [99] in information theory in order to measure the expected value of the information contained in a message. Assuming that a random variable can take n distinct values, with probabilities p_1, p_2,...,p_n, Shannon entropy is defined as:

$$H(p_1, ..., p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{6.6}$$

Shannon entropy equals to 0 when one probability equals to 1 and all other probabilities equal to 0. Shannon entropy reaches its maximum value, $\log_2 n$ when all probabilities equal to $\frac{1}{n}$. More generally, the more homogeneous the probability distribution is, the greater Shannon entropy is. For this reason, Shannon entropy can be used as a fairness measure. By defining:

$$p_i = \frac{x_i}{\sum_{j=1}^{n} x_j} \tag{6.7}$$

$H(p_1, ..., p_n)$ is a fairness measure of the resource allocation vector $x = (x_1, ..., x_n)$.

In order to unify these various theories, [100] and [101] developed an axiomatic approach which is summarized below. A fairness measure is a function

$$f : \mathbb{R}^n \to \mathbb{R}$$

where $x \in \mathbb{R}^n$ is an allocation vector representing the resource allocated to each user, fulfilling the following axioms:

- Continuity: f is continuous for any $n \geq 1$.
- Homogeneity: $f(x) = f(t.x), \forall t > 0$
- Saturation: As the number of users tends to infinity, fairness value of equal resource allocation becomes independent of the number of users: $\lim_{n\to\infty} \frac{1_n}{1_{n+1}}$.
- Partition:

The partition axiom uses the concept of mean function:

$$h : \mathbb{R}^2 \to \mathbb{R}$$

is a mean function iff there exist a continuous and strictly monotonic function g and two positive weights s_1 and s_2 fulfilling $s_1 + s_2 = 1$ such that: $\forall u, v \in \mathbb{R}, g(h(u, v)) = s_1 g(u) + s_2 g(v)$.

Considering an arbitrary partition of the system into two subsystems, the partition axiom states that there exists a mean function

$$h : \mathbb{R}^2 \to \mathbb{R}$$

such that the fairness ratio of two resource allocation vectors $x = [x^1 x^2]$ and $y = [y^1 y^2]$ equals the mean of the fairness ratios of the two suballocations: $\frac{f(x)}{f(y)} = h\left(\frac{f(x^1)}{f(y^1)}, \frac{f(x^2)}{f(y^2)}\right)$.

- Starvation: In a two-user system, an equal allocation is more fair than starving one user: $f([11]) \geq f([10])$.

Starting with these five axioms, [100] generates a family of fairness measures from a generator function:

$$f_\beta(x) = \text{sign}(1-\beta) \left(\sum_{i=1}^{n} \left(\frac{x_i}{\sum_{j=1}^{n} x_j} \right)^{1-\beta} \right)^{\frac{1}{\beta}} \tag{6.8}$$

where sign(.) is the sign function.

Fairness in network resource allocation encounters the question of fair balance of utility functions, rather than fair balance of throughputs: fair balance of throughputs can lead to very unsatisfied users. For example, if two users have an all-or-nothing utility function for the total available throughput in the network, the best solution is to choose arbitrarily one of the users and to attribute him the whole throughput.

6.1.2 Utility function

On the other hand, maximization of utility function is also considered. These two approaches are clearly different, since scale-invariant metrics are unaffected by the total amount of allocated resources, while maximization of utility functions leads to Pareto optimal resource allocations.

In [102], users' satisfaction is defined taking into account the available resource. As a consequence, a user receiving a given allocation will have a greater satisfaction if all or part of the allocation he receives has been taken at the expense of other users. This definition leads to a trade-off between fairness and efficiency which favors compromise from the part of the users. This is fairly adapted to a resource sharing problem where users know all the allocations or at least the total allocation. However, the basic need of a user connecting to a telecommunication network is to transmit his data, without taking into account other users' allocations. Therefore, in our model, the utility function of a user only depends upon the throughput he receives.

We assume in the following that each user requests a given throughput R and has a utility function f fulfilling the following conditions:

- $\forall x \leq 0, f(x) = 0$
- $\forall x \geq R, f(x) = 1$
- $\forall x < R, f(x) < 1$
- f is a growing function

It has been shown in [103] that if the utility function is concave, then the general utility is maximized when resources are equally allocated. However, we will not make any assumption on convexity since all kinds of convexity properties can occur in practice for utility functions.

We assume that each user's utility function is known by the network, with no possibility of cheating.

Since the question is sharing one resource between several users, we will deal with cardinal utility rather than ordinal utility. The utility functions are considered according to their absolute values and treated as additive and multiplicative. Therefore, unlike most utility functions in use in economic modeling, the utility functions we consider are affected by composition with a growing function.

Many well-known resource allocation optimization policies are particular cases of utility functions. For example:

- a constraint requiring the $SINR$ to be greater than a threshold [104], [105] is equivalent to an all-or-nothing utility function;
- optimizing the total throughput [106] is equivalent to a linear utility function.
- proportional fairness [107], [108] is equivalent to a logarithmic utility function.

Taking into account the fact that each player tries to optimize his utility function, the fairness measure of resource allocation will be defined as Jain's index of utility functions:

$$F(x_1, x_2, \ldots, x_n) = J\left(f_1(x_1), f_2(x_2), ..., f_n(x_n)\right) = \frac{\left(\sum_{i=1}^{n} f_i(x_i)\right)^2}{n \sum_{i=1}^{n} f_i(x_i)^2}$$

In the case where there are several classes of users, Jain's index may be weighted: $F(x_1, x_2, \ldots, x_n) = J\left(f_1(x_1), f_2(x_2), ..., f_n(x_n)\right)$ $F(x_1, x_2, \ldots, x_n) = \frac{\left(\sum_{i=1}^{n} \beta_i f_i(x_i)\right)^2}{n \sum_{i=1}^{n} (\beta_i f_i(x_i))^2}$

A generalized definition of utility function has been proposed in [107] and [109] in the case of a multi-resource allocation: let n_k be the number of users using service k, $1 \leq k \leq N$. The objective function to maximize will be:

$$\sum_{k=1}^{N} \lambda_k U \left(\sum_{i=1}^{n_k} \mu_{k,i} R_{k,i} \right) \tag{6.9}$$

where U is a utility function and λ_k and $\mu_{k,i}$ are the weight of resource k and the weight of user i in resource k, respectively.

This model has two particular cases:

- if $N = 1$ (one resource), the resource allocation will be the maximization of the weighted rates;
- if $n_k = 1$ for all k (one user per resource), the resource allocation will be the maximization of the weighted sum of utility functions (for example proportional fairness if function U is a logarithm).

6.2 A parallel between finance and resource allocation

The expected utility model was introduced by Daniel Bernoulli in the 18$^{\text{th}}$ century. In order to solve the well-known Saint Petersburg paradox, Bernoulli introduced the utility function and suggested that a rational player behaves as if he tries to maximize his expected utility function. This model was formally developed by John Von Neumann and Oscar Morgenstern [81], who stated four axioms which define a rational decision maker. In 1953, Maurice Allais published simple examples disproving Bernoulli's utility function and Von Neumann and Morgenstern's independence assumption [110] (see Chapter 3).

Allais' paradoxes can be transposed to resource allocation in wireless networks.

6.2.1 Utility function

Let us consider the four following networks, proposing various rates for equally-weighted users (see Table 6.1).

While it is generally assumed that Network $1A$ is more efficient than Network $1B$, it is also assumed that Network $2B$ is more efficient than Network $2A$. However, as in the example above, these statements contradict the definition of network efficiency.

Network	Users	Rate
1A	100%	1 Mbit/s
1B	10% 89% 1%	5 Mbit/s 1 Mbit/s 0
2A	11% 89%	1 Mbit/s 0
2B	10% 90%	5 Mbit/s 0

Table 6.1: Allais paradox - Transposition to resource allocation

6.2.2 Independence assumption

Let us consider the three following networks, proposing various rates for equally-weighted users (see Table 6.2).

Network	Users	Rate
A	100%	100 Mbit/s
B	98% 2%	500 Mbit/s 0
C	100%	0

Table 6.2: Independence assumption - Transposition to resource allocation

With $t = 1\%$, the resulting networks are described in Table 6.3.

Network	Users	Rate
tA+(1-t)C	1% 99%	100 Mbit/s 0
tB+(1-t)C	0.98% 99.02%	500 Mbit/s 0

Table 6.3: Combined networks

An operator may prefer Network A to Network B, and Network $tB + (1 - t)C$ to Network $tA + (1 - t)C$.

In finance, the most widely accepted explanation for Allais' paradox is risk aversion: if the risk is low, the player will prefer the more secure choice. If the risk is high anyway, the player will try to maximize the risk premium.

A similar explanation can be adopted for network efficiency: in Network $1B$, 10% fully satisfied users will be negatively overwhelmed by 1% of fully unsatisfied users. Network $1A$ will be preferred because it does not let any user fully unsatisfied.

On the other hand, the difference of fully unsatisfied users between Networks $2A$ and $2B$ is not significant (89% vs 90%). For 1% more unsatisfied users, it seems worthy to enhance the satisfaction level of other users.

Therefore, as in the example of utility function in finance, network efficiency shall be considered as a prescriptive theory, rather than a descriptive theory.

A major consequence on this paradox is that resource allocation in a network must be made by considering the network as a whole. Sharing the network into subnetworks and optimizing resource allocation in each one of them will lead to suboptimal allocation.

6.3 Fairness-efficiency trade-off

A survey of fair optimization methods and models for resource allocation in telecommunication networks is provided in [111]. The concept of trade-off between fairness and utility is mentioned in [112]. A metric to evaluate the price of fairness in terms of efficiency loss and the price of efficiency in terms of fairness loss is provided in [113]. A method of management of the efficiency-fairness trade-off by controlling the system fairness index is proposed in [114]. An α-fair dynamically adapted scheduling strategy optimizing coverage and capacity in self-organizing networks is proposed in [115]. In order to introduce our own definition of fairness-efficiency trade-off, we will first define the concepts of network efficiency and unfairness aversion.

6.3.1 Network efficiency

A simple measure for network efficiency can be the total binary rate of the network. However, we prefer to relate network efficiency to the user's utility functions. The reason is that a network can be inefficient though the binary rate may be high. For example, if two users have an all-or-nothing utility function for the total available throughput in the network, choosing arbitrarily one of the users and attributing him the whole throughput is a more efficient solution than sharing the throughput into two equal parts.

Optimizing the total binary rate is a particular case of our model, since it is equivalent to rule that the users' utility functions are linear.

Therefore, we define the network efficiency as the sum, optionally weighted, of the user's utility functions:

$$U(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} \alpha_i f_i(x_i) \tag{6.10}$$

6.3.2 Unfairness aversion

The definition we adopt for unfairness aversion is similar to that of risk aversion in finance.

Unfairness aversion is the reluctance of a network to accept a given unfairness level for a given efficiency level rather than a lower unfairness level for a lower efficiency level.

The concept of unfairness aversion gives a measure of the price the network is ready to pay in terms of efficiency in order to get a better fairness.

Therefore, the network's indifference curves are the set of points in the efficiency-fairness map which are equally satisfying from the network's point of view.

Thus, the network's characteristics can be represented by a point in a two-axis map. Among all the feasible points, the fairness-efficiency trade-off is the point located on the highest indifference curve.

If the fairness is defined by Jain's index, which can be, as mentioned above, expressed as a function of empirical mean and empirical squared error, then the fairness-efficiency trade-off is equivalent to the risk-return trade-off in finance. Though mean and standard deviation do not provide all the information about probability distribution, these two parameters may provide a fairly good approximation in order to describe the operator's preferences.

Unfairness aversion can be characterized by an objective function $\varphi\left(U,J\right)$ where the curves $\varphi\left(U,J\right) = C$ are the indifference curves, C denoting a constant.

6.3.3 Trade-off between fairness and efficiency

The optimal trade-off between fairness and efficiency is given by the constrained optimization:

$$\max \varphi\left(U(x_1, x_2, ..., x_n), J(x_1, x_2, ..., x_n)\right) \tag{6.11}$$

subject to $\sum_{i=1}^{n} x_i = X$, X being the total allocated resource.

The approach described above can be illustrated by the following examples. In these examples, two kinds of utility functions will be used:

- linear utility function: the user's satisfaction is proportional to the rate he gets;
- all-or-nothing utility function: the user is satisfied if and only if he gets the rate he wants.

In practice, linear utility functions can be encountered for example for data transmission, where the time transmission will be inversely proportional to the rate, while all-or-nothing utility functions can be met for real-time video applications.

6.4 Case studies

The approach presented above will be illustrated by two case studies involving a small number of players. The purpose is to show how the fairness-efficiency trade-off can define the optimal resource allocation and to highlight the effect of discontinuous utility functions.

Case study 1: two users with linear utility functions Let us assume that the total resource of the network is $1Mbit/s$ to be shared between two users: one gold user A and one silver user B. We assume that if the rate x (in Mbit/s) is allocated to A, it is possible to allocate up to $1-x$ to B.

We will assume that both users have the same linear utility function: if x is allocated to A and $1-x$ allocated to B, then $f_A(x) = x$ and $f_B(1-x) = 1-x$.

In order to give more weight to user A, we will define the network efficiency as follows:

$$U(x; 1-x) = 3f_A(x) + f_B(1-x) = 3x + 1 - x = 2x + 1 \tag{6.12}$$

Fairness will be evaluated by weighted Jain's index:

$$J(x; 1-x) = \frac{(f_A(x) + 3f_B(1-x))^2}{2\left(f_A(x)^2 + (3f_B(1-x))^2\right)} = \frac{(x + 3(1-x))^2}{2\left(x^2 + (3(1-x))^2\right)} = \frac{(3-2x)^2}{2(10x^2 - 18x + 9)} \tag{6.13}$$

We assume the objective function is:

$$\varphi\left(U, J\right) = J(x; 1-x)U(x; 1-x) \tag{6.14}$$

On Figure 6.1, the blue curve is the feasible set and the red curves are indifference curves. The optimal trade-off between fairness and network efficiency is given by the tangent point of the feasible set and the indifference curve. It is obtained for $x = 0.8$; $1-x = 0.2$; $U(0.8; 0.2) = 2.6$; $J(0.8; 0.2) = 0.98$.

Case study 2: two users with linear utility functions (services like regular data) and two users with all-or-nothing utility functions (services like real time video or audio) Let us assume that the total resource of the network is $1Mbit/s$ to be shared between four users:

- A is a gold user whose utility function is $f_A(x_A) = 1$ if $x_A \geq \frac{1}{2}$; $f_A(x_A) = 0$ else.
- B is a silver user whose utility function is $f_B(x_B) = 1$ if $x_B \geq \frac{1}{2}$; $f_B(x_B) = 0$ else.
- C is a silver user whose utility function is $f_C(x_C) = x_C$.

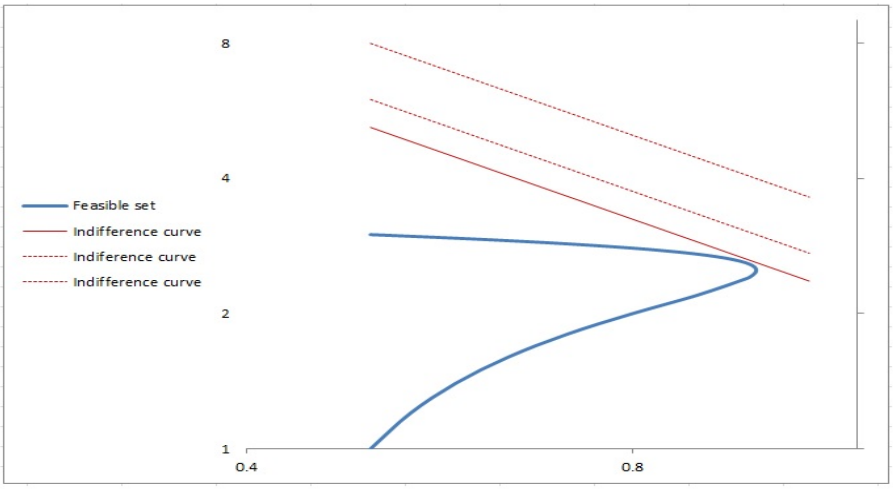

Figure 6.1: Fairness-efficiency trade-off - Case 1.

- D is a bronze user whose utility function is $f_D(x_D) = x_D$.

In order to take into account the preference between users, we will define the network efficiency as follows:

$$U(x_A; x_B; x_C; x_D) = 9f_A(x_A) + 3f_B(x_B) + 3f_C(x_C) + f_D(x_D) \tag{6.15}$$

Fairness will be evaluated by weighted Jain's index:

$$J(x_A; x_B; x_C; x_D) = \frac{(f_A(x_A) + 3f_B(x_B) + 3f_C(x_C) + 9f_D(x_D))^2}{4\left(f_A(x_A)^2 + (3f_B(x_B))^2 + (3f_C(x_C))^2 + (9f_D(x_D))^2\right)} \tag{6.16}$$

As shown on Figure 6.2, the frontier of the feasible set includes two separate subsets:

One point corresponding to the allocation $x_A = x_B = \frac{1}{2}$; $x_C = x_D = 0$ (green point)

One curve corresponding to the allocation $x_A = \frac{1}{2}$; $x_B = 0$; $0 \leq x_C \leq \frac{1}{2}$; $x_D = \frac{1}{2} - x_C$ (blue curve)

We assume the objective function is:

$$\varphi\left(U, J\right)) = J(x_A; x_B; x_C; x_D)U(x_A; x_B; x_C; x_D)^{\alpha} \tag{6.17}$$

where α denotes a constant.

Then:

- for $\alpha < 4.026$, the optimal trade-off between fairness and network efficiency will be on the blue curve;
- for $\alpha = 4.026$, the green point $(\frac{1}{2}; \frac{1}{2}; 0; 0)$ and the black point $(\frac{1}{2}; 0; 0.395; 0.105)$ will be equally satisfying from the network's point of view;
- for $\alpha > 4.026$, the green point will be the optimal trade-off between fairness and network efficiency.

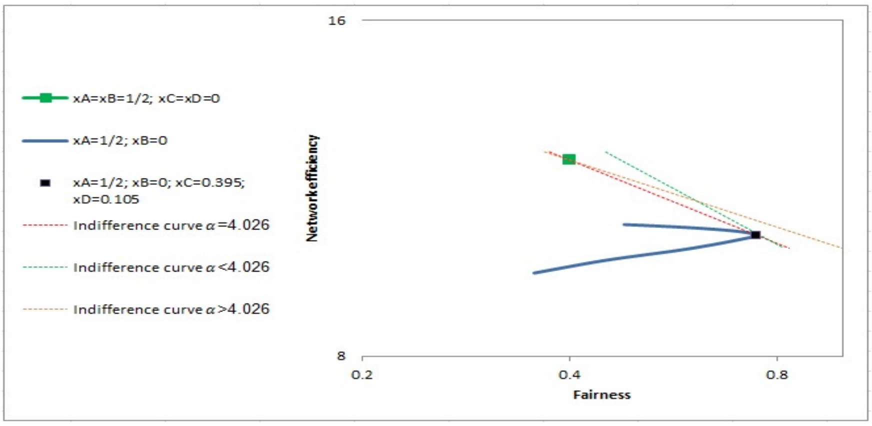

Figure 6.2: Fairness-efficiency trade-off - Case 2.

As illustrated in this case study, discontinuous utility functions, such as all-or-nothing functions, can generate high sensitivity to indifference curves characterizing unfairness aversion.

6.5 Simulations

We ran simulations of the approach presented above in a network involving a large number of users.

6.5.1 System description

Users arrive into the network according to a Poisson process. Each user i is characterized by 4 parameters:

- Utility function (linear or all-or-nothing)

- Requested rate: R_i for an all-or-nothing user, R'_i for a linear user. The requested rate follows a lognormal law
- Volume of data which needs to be transmitted: Q_i for an all-or-nothing user, Q'_i for a linear user. The volume of data follows a lognormal law.
- User class (gold, silver, bronze)

We denote r_i (resp. r'_i) the rate allocated to all-or-nothing (resp. linear) user i and q_i (resp. q'_i) the remaining data to be transmitted by all-or-nothing (resp. linear) user i. At each instant, the network includes n all-or-nothing users and m linear users. $N = n + m$ is the total number of users.

An event is the entry of a newcomer or the exit of a user when his transmission is achieved.

The system will process the resource allocation which optimizes the fairness-efficiency trade-off, under the constraint that the total rate is R_{total}.

The system will provide statistical outputs regarding the rate of satisfaction according to the type of users (class, demand or utility function).

- If the utility function of user i is all-or nothing, it is defined as:
  - $f_i(r_i) = 1_{[R_i,+\infty[}$
- If the utility function of user i is linear, it is defined as:
  - $f'_i(r'_i) = r'_i/R'_i$ for $0 \leq x_i \leq R'_i$
  - $f'_i(r'_i) = 1$ for $r'_i \geq R'_i$

The network efficiency is:

$U(r_1, \ldots, r_n, r'_1, \ldots r'_m) = \sum_{i=1}^{n} \alpha_i f_i(r_i) + \sum_{i=1}^{m} \alpha'_i f'_i(r'_i)$, where:

- $\alpha_i = 9$ or $\alpha'_i = 9$ if user i is a gold user;
- $\alpha_i = 3$ or $\alpha'_i = 3$ if user i is a silver user;
- $\alpha_i = 1$ or $\alpha'_i = 1$ if user i is a bronze user.

Jain's index is:

$F(r_1, \ldots, r_n, r'_1, \ldots r'_m) = J\left(f_1(r_1), \ldots, f_n(r_n), f'_1(r'_1), \ldots, f'_m(r'_m)\right) = \frac{\left(\sum_{i=1}^{n} \beta_i f_i(r_i) + \sum_{i=1}^{m} \beta'_i f'_i(r'_i)\right)^2}{N\left(\sum_{i=1}^{n} (\beta_i f_i(r_i))^2 + \sum_{i=1}^{m} (\beta'_i f'_i(r'_i))^2\right)}$, where:

- $\beta_i = 1$ or $\beta'_i = 1$ if user i is a gold user

- $\beta_i = 3$ or $\beta'_i = 3$ if user i is a silver user
- $\beta_i = 9$ or $\beta'_i = 9$ if user i is a bronze user

We deal with three kinds of indifference curves:

- $\phi(U, J) = U(r_1, \ldots, r_n, r'_1, \ldots r'_m)$;
- $\phi(U, J) = U(r_1, \ldots, r_n, r'_1, \ldots r'_m) J\left(f_1(r_1), \ldots, f_n(r_n), f'_1(r'_1), \ldots, f'_m(r'_m)\right)^{\eta}$, where η is a strictly positive number;
- $\phi(U, J) = J\left(f_1(r_1), \ldots, f_n(r_n), f'_1(r'_1), \ldots, f'_m(r'_m)\right)$.

For each event, the system will process the optimal allocation:

$$\max_{\sum_i r_i \leq R_{total}} \phi(U, J)$$

6.5.2 Algorithm

Any all-or-nothing user shall be either fully served or not served at all. As a result, all-or-nothing users bring discontinuities in the network efficiency and Jain's index, making impossible the use of classical algorithms for optimization.

Taking into account all the serve/not serve possibilities for each all-or-nothing user would lead to an exponentially complex problem.

Therefore, we present hereafter an approximation which enables to find a sub-optimal solution with a polynomial complexity. Under reasonable assumptions, this sub-optimal solution will tend to the optimal solution when the number of users tends to infinity.

Ordering the all-or-nothing users

The all-or-nothing users will be ordered according to the $\frac{\alpha_k}{R_k}$ ratio. The purpose of comparing these ratios is to evaluate the average impact of each unit of rate on the network efficiency.

After this ordering, we have n ordered all-or-nothing users and m linear users, with $n + m = N$.

$$X_1 = \begin{cases} R_1 \\ Q_1 \\ \alpha_1 \end{cases}, X_2 = \begin{cases} R_2 \\ Q_2 \\ \alpha_2 \end{cases}, \ldots, X_n = \begin{cases} R_n \\ Q_n \\ \alpha_n \end{cases}, \text{ with } \frac{\alpha_1}{R_1} \geq \frac{\alpha_2}{R_2} \geq \ldots \geq \frac{\alpha_n}{R_n}$$

$$X_{n+1} = \begin{cases} R_{n+1} \\ Q_{n+1} \\ \alpha_{n+1} \end{cases}, X_{n+2} = \begin{cases} R_{n+2} \\ Q_{n+2} \\ \alpha_{n+2} \end{cases}, \ldots, X_{n+m} = \begin{cases} R_{n+m} \\ Q_{n+m} \\ \alpha_{n+m} \end{cases}$$

Basic assumption: priority among all-or-nothing users

We now assume that if an all-or-nothing user X_k, $1 \leq k \leq n$ is served, then all the all-or-nothing users who are better rated than him (ie X_i, $1 \leq i \leq k$) are also served. This assumption is an approximation and, since rate demands are not breakable, may lead to a sub-optimal solution. The priority that we define among all-or-nothing users is based on the following considerations:

- **Lemma 8.** *Let X_j and X_k be two all-or-nothing users. If $\alpha_j = \alpha_k$ and $R_j \leq R_k$, then:*

$$\max_{\substack{\sum_i r_i \leq R_{total} \\ r_j = R_j \\ r_k = 0}} \phi(U, J) \geq \max_{\substack{\sum_i r_i \leq R_{total} \\ r_j = 0 \\ r_k = R_k}} \phi(U, J)$$

Proof.

$$\max_{\substack{\sum_i r_i \leq R_{total} \\ r_j = R_j \\ r_k = 0}} \phi(U, J) = \max_{\substack{\sum_{i \notin \{j,k\}} r_i \leq R_{total} - R_j \\ r_j = R_j \\ r_k = 0}} \phi(U, J)$$

$$\max_{\substack{\sum_i r_i \leq R_{total} \\ r_j = 0 \\ r_k = R_k}} \phi(U, J) = \max_{\substack{\sum_{i \notin \{j,k\}} r_i \leq R_{total} - R_k \\ r_j = 0 \\ r_k = R_k}} \phi(U, J)$$

Since $U(r_1, ..., r_j = R_j, ..., r_k = 0, ..., r_n, r'_1, ...r'_m) = U(r_1, ..., r_j = 0, ..., r_k = R_k, ..., r_n, r'_1, ...r'_m)$ and $F(r_1, ..., r_j = R_j, ..., r_k = 0, ..., r_n, r'_1, ...r'_m) = F(r_1, ..., r_j = 0, ..., r_k = R_k, ..., r_n, r'_1, ...r'_m)$, the theorem results from the fact that $R_{total} - R_j \geq R_{total} - R_k$.

□

- **Lemma 9.** *Let X_{j_1},..., X_{j_q} be q all-or-nothing users with $\alpha_{j_1} = ... = \alpha_{j_q} = \alpha_j$ and X_{k_1},..., X_{k_p} p all-or-nothing users with $\alpha_{k_1} = ... = \alpha_{k_p} = \alpha_k$. If $\frac{\alpha_j}{\alpha_k} = \frac{p}{q}$, then $U(r_1, ..., r_{j_1} = R_{j_1}, ..., r_{j_q} = R_{j_q}, r_{k_1} = 0, ..., , r_{k_p} = 0, ..., r_n, r'_1, ...r'_m) = U(r_1, ..., r_{j_1} = 0, ..., r_{j_q} = 0, r_{k_1} = R_{k_1}, ..., , r_{k_p} = R_{k_p}, ..., r_n, r'_1, ...r'_m)$*

Proof. $U(r_1, ..., r_{j_1} = R_{j_1}, ..., r_{j_q} = R_{j_q}, r_{k_1} = 0, ..., , r_{k_p} = 0, ..., r_n, r'_1, ...r'_m) = q\alpha_j + \sum_{i=\notin\{j_1,...j_q,k_1,...k_p\}} \alpha_i f_i(r_i) + \sum_{i=1}^m \alpha'_i f'_i(r'_i)$
$U(r_1, ..., r_{j_1} = 0, ..., r_{j_q} = 0, r_{k_1} = R_{k_1}, ..., , r_{k_p} = R_{k_p}, ..., r_n, r'_1, ...r'_m) = p\alpha_k + \sum_{i=\notin\{j_1,...j_q,k_1,...k_p\}} \alpha_i f_i(r_i) + \sum_{i=1}^m \alpha'_i f'_i(r'_i)$
Since $q\alpha_j = p\alpha_k$, the equality is obtained.

□

Lemma 9 shows that there are two groups of users which are equivalent with regard to the network efficiency. Of course, these two groups may have a different impact on Jain's index. However, the expression of Jain's index shows that if the number of other users is much larger than the size of these two groups, serving either one of these two groups should have little effect on Jain's index. This gives legitimacy to ordering the all-or-nothing users according to the $\frac{\alpha_k}{R_k}$ ratio, since this ratio represents the gain in network efficiency relative to the allocated rate.

This approximation enables to reduce the number of optimizations from 2^n to $n+1$.

The algorithm

The simulation algorithm is described in Figure 6.3

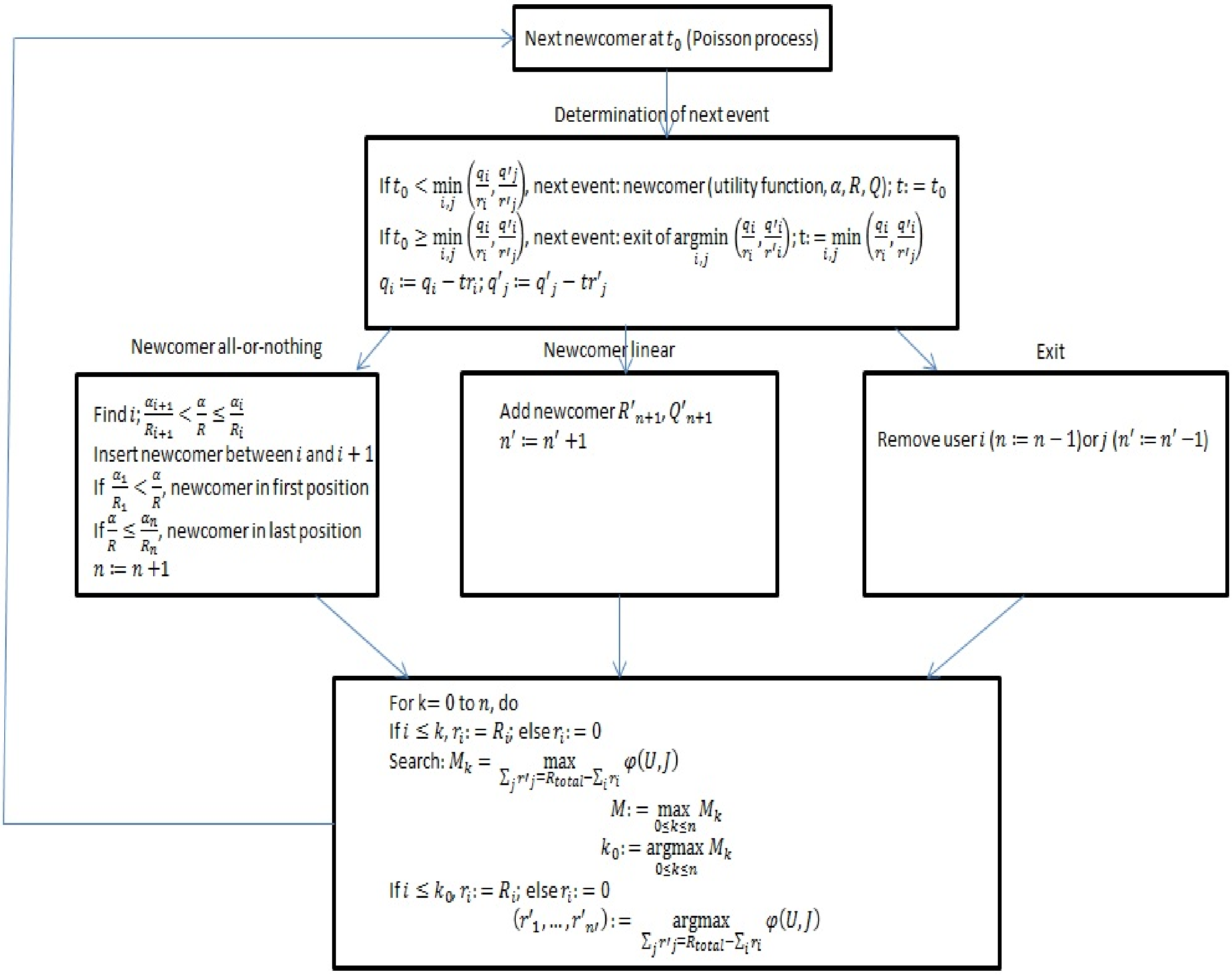

Figure 6.3: Simulation Algorithm.

Simulation parameters

The simulations were run with the following parameters:

- $R_{total} = 1,000$ Mbit/s
- Number of events: 4,000.
- Requested rate $R_i = R_0 2^{u_i}$ or $R'_i = R_0 2^{u'_i}$, with $R_0 = 100$ Mbit/s , and u_i and u'_i following a standard normal distribution.
- Volume of data which needs to be transmitted: $Q_i = Q_0 2^{v_i}$ or $Q'_i = Q_0 2^{v'_i}$, with $Q_0 = 1,000$ Mbit, and v_i and v'_i following a standard normal distribution.
- Newcomers arrive according to a Poisson process. The Poisson parameter λ is adjusted in order to have a stable long-term number of users in the network.

  If n users, each one willing to transmit a volume of data $Q_i = Q_0 2^{v_i}$, $1 \leq i \leq n$ are in the network, the average requested time to exit all of them is:

$$T = \overline{\sum_{i=1}^{n} \frac{Q_i}{R_{total}}} = \frac{Q_0}{R_{total}} \overline{\sum_{i=1}^{n} 2^{v_i}} = \frac{Q_0}{R_{total}} \overline{\sum_{i=1}^{n} e^{v_i \ln 2}} = \frac{nQ_0}{R_{total}} exp(\ln 2^2/2) \qquad (6.18)$$

  Therefore, in order that the average number of incoming users equals the average number of exiting users, $\lambda = \frac{R_{total}}{Q_0 exp(\ln 2^2/2)} \approx 0.78645$.
- All random processes are independent.

Fairness-efficiency trade-off :

In order to evaluate the impact of the fairness-efficiency trade-off on the network performance, we ran the simulations with three different objective functions:

- $\phi(U, J) = U$; maximization of network efficiency;
- $\phi(U, J) = UJ$; trade-off between fairness and efficiency, with $\eta = 1$;
- $\phi(U, J) = J$; maximization of fairness.

It should be noted that the resource which is allocated is not storable. For this reason, the efficiency and the fairness are measured at any time for the users who are connected to the network at this time. There is no fairness between users connected at different times to the network.

Results

The minimum time a user i stays in the network is $\frac{Q_i}{R_i}$ or $\frac{Q'_i}{R'_i}$. Therefore, a relevant parameter to evaluate the network performance, from the point of view of user i, is $x_i = \frac{Q_i}{t_i R_i}$ or $x'_i = \frac{Q'_i}{t'_i R'_i}$, where t_i (resp. t'_i) is the delay between the arrival of the all-or-nothing user (resp. linear user) i to the network and the end of his transaction.

$\phi(\mathbf{U}, \mathbf{J}) = \mathbf{U}$

Figures 6.4, 6.5 and 6.6 provide the cumulative distribution functions of x_i resulting from the simulations for each category of users. Means and standard deviations are summarized in Table 6.4.

For each class of users, cumulative distribution functions, means and standard deviation are similar for all-or-nothing utility function users and linear utility function users.

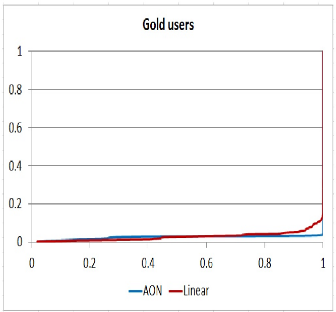

Figure 6.4: Gold users.

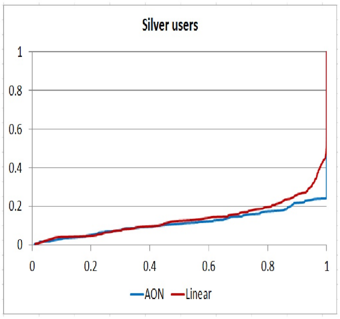

Figure 6.5: Silver users.

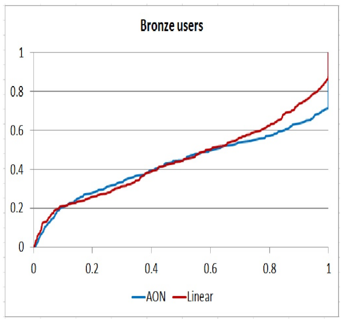

Figure 6.6: Bronze users.

AON Users	Mean	Standard deviation	Linear Users	Mean	Standard deviation
Gold	0.98	0.14	**Gold**	0.97	0.12
Silver	0.89	0.25	**Silver**	0.87	0.26
Bronze	0.57	0.39	**Bronze**	0.55	0.37

Table 6.4: Simulations results $\phi(U, J) = U$.

$\phi(\mathbf{U}, \mathbf{J}) = \mathbf{UJ}$

Figures 6.7, 6.8 and 6.9 provide the cumulative distribution functions of x_i resulting from the simulations for each category of users. Means and standard deviations are summarized in Table 6.5.

For each class of users, the number of fully-satisfied users ($x_i = 1$) is much greater for all-or-nothing utility function users. The standard deviations and the numbers of presumably disappointed customers (low values of x_i) are much greater for all-or-nothing utility function users.

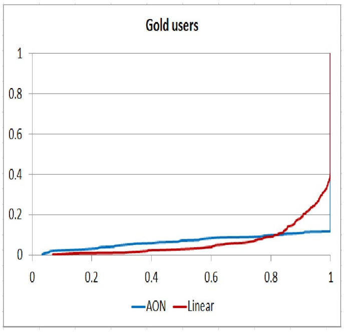

Figure 6.7: Gold users.

Figure 6.8: Silver users.

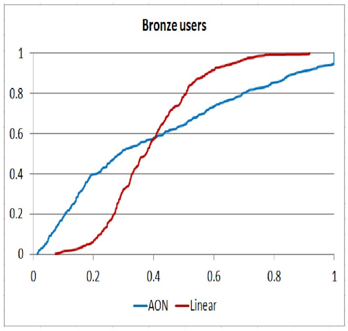

Figure 6.9: Bronze users.

AON Users	Mean	Standard deviation	Linear Users	Mean	Standard deviation
Gold	0.93	0.21	**Gold**	0.94	0.14
Silver	0.78	0.32	**Silver**	0.76	0.24
Bronze	0.39	0.30	**Bronze**	0.39	0.15

Table 6.5: Simulations results $\phi(U, J) = UJ$.

$\boldsymbol{\phi(\mathbf{U}, \mathbf{J}) = \mathbf{J}}$

Figures 6.10, 6.11 and 6.12 provide the cumulative distribution functions of x_i resulting from the simulations for each category of users. Means and standard deviations are summarized in Table 6.6.

For each class of users, the number of fully-satisfied users ($x_i = 1$) is much greater for all-or-nothing utility function users. The standard deviations and the numbers of presumably disappointed customers (low values of x_i) are much greater for all-or-nothing utility function users.

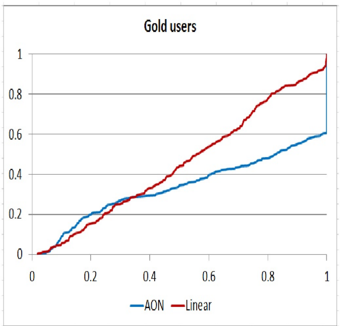

Figure 6.10: Gold users.

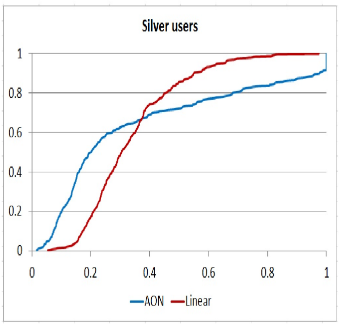

Figure 6.11: Silver users.

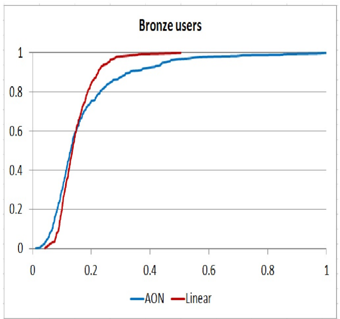

Figure 6.12: Bronze users.

The performance is strongly degraded when the network maximizes the fairness ($\phi(U, J) = J$). When comparing the performances between maximizing the network efficiency ($\phi(U, J) = U$) and maximizing a trade-off between fairness and efficiency ($\phi(U, J) = UJ$), it turns out that the means of x_i are always higher for $\phi(U, J) = U$. However, as the comparison

AON Users	Mean	Standard deviation	Linear Users	Mean	Standard deviation
Gold	0.67	0.36	**Gold**	0.55	0.28
Silver	0.35	0.32	**Silver**	0.34	0.15
Bronze	0.17	0.14	**Bronze**	0.15	0.06

Table 6.6: Simulations results $\phi(U, J) = J$.

between the cumulative distribution functions of x_i shows, the number of presumably disappointed customers (low values of x_i) is also significantly greater.

This example highlights the fact that, by defining its fairness-efficiency trade-off function and adjusting the parameter η in the objective function $\phi(U, J) = UJ^\eta$, an operator can define its priorities and find the proper balance between maximizing the network efficiency and minimizing the number of disappointed customers.

6.6 Closing Remarks

Tools already existing in finance can be successfully transposed to resource allocation in telecommunication networks. By introducing the concepts of network efficiency and unfairness aversion, the fairness-efficiency trade-off can be treated similarly to the well-known risk-return trade-off in finance. This approach enables to define an optimal resource allocation without conflicting with paradoxes resulting from the maximization of a utility function.

The resource allocation depicted above is based above the assumption that the users do not encounter any interference limitation and that the only limiting factor is the total rate allocated to the users. However, another case may occur, where the spectral resources offered by the network are more restricting than the link between the wireless network and the core network. In that case, the interferences generated by the users will be the limiting factor and the constraint on the sum of the rates allocated to the users will not hold anymore.

Statistical analysis shows that users located near the edge of coverage areas are over-represented by comparison with other users. The reason is that these users tend to have higher communication durations in order to perform their operations.

Since these users are located far away from the BS, they usually generate more interferences than other users. Therefore, in order to optimize the network efficiency, it is generally preferable to eliminate them. The fairness-efficiency trade-off should favor users located close to the BS and compensate the statistical bias mentioned above.

The power received by a mobile u_i $(1 \leq i \leq m)$ from Base Station B_j $(1 \leq j \leq n)$ is:

$$P_{ij} = KP_j h_{ji} d(B_j, u_i)^{-\alpha} \tag{6.19}$$

where K is a constant, P_j is the power transmitted by Base Station B_j, h_{ji} is the random channel gain and α is the path-loss exponent. If u_i is served by Base Station B_k, then, the interference at user i is:

$$SINR_i = \frac{KP_k h_{ki} d(B_k, u_i)^{-\alpha}}{\sum_{j \neq k} KP_j h_{ji} d(B_j, u_i)^{-\alpha} + N_0} \tag{6.20}$$

According to the Shannon-Hartley theorem, the maximum bit rate which can be transmitted over the channel is given by:

$$C_i = B \log_2(1 + SINR_i) \tag{6.21}$$

Thus, the network efficiency is:

$$U(P_1, \ldots, P_n) = \sum_{i=1}^{m} \alpha_i f_i(C_i) \tag{6.22}$$

and Jain's index:

$$J(P_1, \ldots, P_n) = \frac{\left(\sum_{i=1}^{n} \beta_i f_i(C_i)\right)^2}{n \sum_{i=1}^{n} \left(\beta_i f_i(C_i)\right)^2} \tag{6.23}$$

Therefore, knowing the network's parameters, the channel model, the position of each user and the power allocated to each Base Station, it is possible to calculate the network efficiency and Jain's index.

From this point, as in the case where the limiting factor is the total rate, it is possible to determine the optimal fairness-efficiency trade-off by maximizing:

$\phi(U, J) = U(P_1, \ldots, P_n) J\left(P_1, \ldots, P_n\right)^{\eta}$, where η is a strictly positive number.

6.6.1 Network without noise ($N_0 = 0$)

If $N_0 = 0$, the $SINR$ is not changed by multiplying all the powers P_j, $1 \leq j \leq n$ by a constant. The network efficiency and Jain's index will not be changed either.

Therefore, it is possible to add an additional constraint, for example forcing the total power to a fixed value: $\sum_{i=1}^{m} P_i = P$.

6.6.2 Network with noise ($N_0 \neq 0$)

If $N_0 \neq 0$, as a result of the definition of the $SINR$, multiplying all the powers P_j, $1 \leq j \leq n$ by a constant λ is equivalent to dividing N_0 by λ . Therefore, if the allocated powers are not constrained, the solution of the optimization problem will be powers which tend to infinity, and the network efficiency will be the network efficiency of a network without noise. In order to avoid powers tending to infinity, it is necessary to add a constraint, for example forcing the total power to a fixed value: $\sum_{i=1}^{m} P_i = P$.

Chapter 7

Conclusions and Outlook

Three different network optimization problems have been considered in this thesis. The proposed solutions are aimed to enable resource limited networks to deal with the increasing demand for bandwidth. For each one of the considered problems, the road is open for further research.

7.1 Availability

One basic assumption in our model is that failure events are uncorrelated. This assumption is plausible in wired links, but unlikely in wireless links. A more plausible model for wireless rings could include correlations between failure events. Introducing these correlations will add more complexity.

We reached the conclusion in our model that availability optimization is an NP-hard problem if the rings include 4 nodes or more. For this reason, we propose approximation methods based on Linear Programming. We introduce an exponent, aimed to reduce the probability of non-integer solutions. Optimization of these methods requires further investigation.

Moreover, NP-hard is not the last word. Quantum computers will be able to solve NP-hard problems in the future and it may be possible to write an algorithm for quantum computer which could solve availability optimization problems in a polynomial time.

Direct extension of the approach used for ring-topology networks to other topologies will often lead to trivial problems: according to our definition, the network is unavailable when at least one of its nodes is unavailable. As a result, for most common topologies, excluding rings (chain, star, tree), the networks is unavailable when at least one link is down. However, non-trivial problems based on trade-off between availability, cost and delay can be formulated for these topologies.

For example, if each link has a cost and a failure probability, each feasible solution is characterized by a total cost and an availability. These two parameters correspond to a point in the availability-cost plane. In a similar manner as risk aversion is defined in finance, or as we defined unfairness aversion in Chapter 6, it is possible to define unavailability aversion, which is the cost that the network is ready to pay in order to improve its availability. Thus, the network will be characterized by indifference curves in the availability-cost plane and the problem will be to find which feasible solution optimizes the availability-cost trade-off.

7.2 Polarization

A new paradigm has been proposed. As it has been highlighted, it has aftermath in topology and channel allocation, as illustrated by the examples given in this thesis. In each one of these two fields, a systematic optimization research could be led:

- Topology

  Given n nodes A_1, A_2,..., A_n in the space, is there a ring connecting these n nodes and enabling network coverage by one single channel with the use of polarization?

  Examples given in Chapter 5 show that when a ring is not adapted to the inclined polarization solution, a change in the topology can solve the problem. However, a criterion characterizing these rings still needs to be found.

- Channel allocation

  What are the networks in which polarization enables a more efficient channel allocation than the common approach based upon graph coloring?

  An example of such a network has been provided in Chapter 5. A systematic identification of these rings is still to be done.

In addition to these academic studies, the proposed paradigm is aimed to serve industrial applications. The concept still needs to be validated by laboratory experiments, under real operating conditions. The tolerance to misplacement or misalignment of antennas, to imperfect orthogonality of polarizations, must be measured in order to confirm the industrial applicability of the solution.

The orthogonality condition on polarizations is much more challenging for two signals coming from two different sources than it is in the classical point-to-point case. A dual-polarized antenna can be designed to ensure that the two transmitted polarizations are orthogonal to each other, even though each one of them may have some inaccuracy. On the other hand, two different antennas are subject to different conditions. Ensuring polarization orthogonality under real operating conditions, including wind or ground motion, requires an adaptive centralized control of the polarizations.

In a further step, the implementation of the proposed solution in a network with mobile base stations could be considered. As stated in Chapter 5, this network will require a centralized control system and polarization-agile antennas.

7.3 Resource allocation

The parallelism with Allais' paradoxes in finance shows that present approaches, based on a function maximization, do not always describe properly the strategy of a rational operator. The prospect theory, intended to answer Allais' paradoxes, faces problems of discontinuity, and therefore does not seem to be applicable to telecommunication networks. We proposed therefore another approach, based on a trade-off between fairness and efficiency. The main concepts of risk aversion and risk-return trade-off can be translated in the telecommunication context, taking into account the specific characteristics of this field.

In finance, the debate is far from over. A similar debate needs to be held in resource allocation. In any case, the distribution of allocated resources must be taken into account, and not only an objective function.

The approach presented in this thesis can be extended in more complex situations, for example interconnected networks of rival operators. In such a context, each operator would define its own fairness-efficiency trade-off. Existing tools used in game theory could be used to optimize the strategies of operators and customers.

Bibliography

[1] R. Chundury, “Mobile broadband backhaul: Addressing the challenge,” *Planning Backhaul Networks, Ericsson Review*, no. 3, pp. 4–9, 2008.

[2] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, “Openflow: enabling innovation in campus networks,” *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.

[3] C. Chaudet and Y. Haddad, “Wireless software defined networks: Challenges and opportunities,” in *Microwaves, Communications, Antennas and Electronics Systems (COMCAS), 2013 IEEE International Conference on.* IEEE, 2013, pp. 1–5.

[4] K.-K. Yap, M. Kobayashi, D. Underhill, S. Seetharaman, P. Kazemian, and N. McKeown, “The stanford openroads deployment,” in *Proceedings of the 4th ACM international workshop on Experimental evaluation and characterization.* ACM, 2009, pp. 59–66.

[5] D. Kreutz, F. M. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, “Software-defined networking: A comprehensive survey,” *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14–76, 2015.

[6] J. DiGiglio and D. Ricci, “High performance, open standard virtualization with nfv and sdn,” *White paper, Intel Corporation and Wind River*, 2013.

[7] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, “What will 5g be?” *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[8] R. Chávez-Santiago, M. Szydełko, A. Kliks, F. Foukalas, Y. Haddad, K. E. Nolan, M. Y. Kelly, M. T. Masonta, and I. Balasingham, “5g: The convergence of wireless communications,” *Wireless Personal Communications*, vol. 83, no. 3, pp. 1617–1642, 2015.

[9] L. A. Belov, S. M. Smolskiy, and V. N. Kochemasov, *Handbook of RF, microwave, and millimeter-wave components.* Artech house, 2012.

[10] H. Lehpamer, *Microwave Transmission Networks.* McGraw-Hill, 2010.

[11] M. A. Weissberger, "An initial critical summary of models for predicting the attenuation of radio waves by trees," ELECTROMAGNETIC COMPATIBILITY ANALYSIS CENTER ANNAPOLIS MD, Tech. Rep., 1982.

[12] I. R. Sector, "Recommendation itu-r p. 676–10, attenuation by atmospheric gases," *International Telecommunications Union*, 2013.

[13] N. Nepomuceno, "Network optimization for wireless microwave backhaul," Ph.D. dissertation, Université Nice Sophia Antipolis, 2010.

[14] H. R. Anderson, *Fixed broadband wireless system design.* John Wiley & Sons, 2003.

[15] "Microwave Backhauling - Meeting the Capacity Challenges of 4G/LTE and Beyond," *Ceragon Technical Brief*, pp. 1–10, 2012.

[16] E. Hecht, "Optics, 4th," *International edition, Addison-Wesley, San Francisco*, vol. 3, 2002.

[17] H. S. Dhillon and G. Caire, "Wireless backhaul networks: Capacity bound, scalability analysis and design guidelines," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6043–6056, 2015.

[18] A. F. Molisch, *Wireless Communications.* Wiley, 2010.

[19] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud ran for mobile networks—a technology overview," *IEEE Communications surveys & tutorials*, vol. 17, no. 1, pp. 405–426, 2015.

[20] B. Li, D. Zhu, and P. Liang, "Small cell in-band wireless backhaul in massive mimo systems: A cooperation of next-generation techniques," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 7057–7069, 2015.

[21] A. Goldsmith, *Wireless communications.* Cambridge university press, 2005.

[22] A. Khadka, K. Adachi, S. Sun, H. Zhu, and J. Wang, "Cooperative transmission strategy for downlink distributed antenna systems over time-varying channel," in *Global Communications Conference (GLOBECOM), 2015 IEEE.* IEEE, 2015, pp. 1–6.

[23] J. Hoydis, K. Hosseini, S. ten Brink, and M. Debbah, "Making Smart Use of Excess Antennas: Massive MIMO, Small Cells, and TDD," *Bell Labs Technical Journal*, vol. 18, no. 2, pp. 5–21, Sep. 2013.

[24] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive mimo: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[25] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive mimo: the next wireless revolution?" *IEEE Communications Magazine*, vol. 52, no. 9, pp. 56–62, 2014.

[26] Q. Zhang, S. Jin, K.-K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive mimo systems with arbitrary-rank channel means," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 966–981, 2014.

[27] S. Sanayei and A. Nosratinia, "Antenna selection in mimo systems," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 68–73, 2004.

[28] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[29] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive mimo systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7112–7139, 2014.

[30] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser mimo systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.

[31] S. Landstrom, A. Furuskãr, K. Johansson, L. Falconetti, and F. Kronestedt, "Heterogeneous networks–increasing cellular capacity," *Jour. Ericson Review*, vol. 89, pp. 4–9, 2011.

[32] Q. Zhang, C. He, and L. Jiang, "Base station clustering in heterogeneous network with finite backhaul capacity," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on.* IEEE, 2015, pp. 2899–2903.

[33] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. A. Thomas, J. G. Andrews, P. Xia, H. S. Jo *et al.*, "Heterogeneous cellular networks: From theory to practice," *IEEE Communications Magazine*, vol. 50, no. 6, 2012.

[34] W. Bao and B. Liang, "Stochastic geometric analysis of user mobility in heterogeneous wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2212–2225, 2015.

[35] M. Paolini, L. Hiley, F. Rayal, and S. Fili, "Small-cell backhaul: Industry trends and market overview."

[36] D. C. Chen, T. Q. Quek, and M. Kountouris, "Backhauling in heterogeneous cellular networks: Modeling and tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3194–3206, 2015.

[37] ——, "Wireless backhaul in small cell networks: Modelling and analysis," in *Vehicular Technology Conference (VTC Spring), 2014 IEEE 79th.* IEEE, 2014, pp. 1–6.

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

[39] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communications magazine*, vol. 46, no. 9, 2008.

[40] H. S. Dhillon, Y. Li, P. Nuggehalli, Z. Pi, and J. G. Andrews, "Fundamentals of heterogeneous cellular networks with energy harvesting," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2782–2797, 2014.

[41] P. C. Weeraddana, M. Codreanu, M. Latva-aho, and A. Ephremides, "On the effect of self-interference cancelation in multihop wireless networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, no. 1, p. 513952, 2010.

[42] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking.* ACM, 2010, pp. 1–12.

[43] R. Nadiv and T. Naveh, "Wireless Backhaul Topologies: Analysing Backhaul Topology Strategies," *Ceragon White Paper*, pp. 1–15, 2010.

[44] P. Madhusudhanan, J. G. Restrepo, Y. Liu, and T. X. Brown, "Downlink coverage analysis in a heterogeneous cellular network," in *Global communications conference (GLOBECOM), 2012 IEEE.* IEEE, 2012, pp. 4170–4175.

[45] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink sinr analysis," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3484–3495, 2012.

[46] J. G. Andrews, "Seven ways that hetnets are a cellular paradigm shift," *IEEE Communications Magazine*, vol. 51, no. 3, pp. 136–144, 2013.

[47] Y. Lin, W. Bao, W. Yu, and B. Liang, "Optimizing user association and spectrum allocation in HetNets: A utility perspective," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1025–1039, 2015.

[48] K. Shen and W. Yu, "Distributed pricing-based user association for downlink heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1100–1113, 2014.

[49] N. Wang, E. Hossain, and V. K. Bhargava, "Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier hetnets with large-scale antenna arrays," *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3251–3268, 2016.

[50] J.-W. Lee, R. R. Mazumdar, and N. B. Shroff, "Joint resource allocation and base-station assignment for the downlink in CDMA networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 1, pp. 1–14, 2006.

[51] M. Hong and A. Garcia, "Mechanism design for base station association and resource allocation in downlink OFDMA network," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 11, pp. 2238–2250, 2012.

[52] S. Chia, M. Gasparroni, and P. Brick, "The next challenge for cellular networks: Backhaul," *IEEE Microwave Magazine*, vol. 10, no. 5, 2009.

[53] M. Paolini, "Crucial economics for mobile data backhaul," *Senza Fili Consulting*, 2011.

[54] R. Cluster, "5g radio network architecture," *Tech. Rep., Radio Access and Spectrum-FP 7 Future Internet Cluster*, 2014.

[55] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 167–178, 2013.

[56] H. N. Gabow, "Implementation of algorithms for maximum matching on nonbipartite graphs," Ph.D. dissertation, 1974.

[57] J. Edmonds, "Maximum matching and a polyhedron with 0,1 vertices," *J. Res. Nat. Bur. Standards B 69.1965*, vol. 69, no. 125-130, pp. 55–56, 1965.

[58] H. N. Gabow, "An efficient implementation of Edmonds' algorithm for maximum matching on graphs," *Journal of the ACM (JACM) 23.2*, pp. 221–234, 1976.

[59] ——, *Data structures for weighted matching and nearest common ancestors with linking.* University of Colorado, Boulder, Department of Computer Science, 1990.

[60] A. Schrijver, *Theory of linear and integer programming.* Wiley, 1998.

[61] S. Polyakovskiy, F. C. Spieksma, and G. J. Woeginger, "The three-dimensional matching problem in Kalmanson matrices," *Journal of Combinatorial Optimization 26.1*, no. 1, pp. 1–9, 2013.

[62] A. Turkmani, A. Arowojolu, P. Jefford, and C. Kellett, "An Experimental Evaluation of the Performance of Two-Branch Space and Polarization Diversity Schemes at 1800 MHz," *IEEE-J-VT*, vol. 44, no. 2, pp. 318–326, May 1995.

[63] J. J. Lempiăinen and J. K. Laiho-Steffens, "The Performance of Polarization Diversity Schemes at a Base Station in Small/Micro Cells at 1800 MHz," *IEEE-J-VT*, vol. 47, no. 3, pp. 1087–1092, Aug. 1998.

[64] B. S. Collins, "Polarization Diversity Antennas for Compact Base Stations," *Microwave Journal*, vol. 43, no. 1, pp. 76–88, Jan. 2000.

[65] S.-C. Kwon and G. L. Stüber, "Polarization division multiple access on NLoS wideband wireless fading channels," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3726–3737, 2014.

[66] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[67] A. M. Sayeed and N. Behdad, "Continuous aperture phased mimo: A new architecture for optimum line-of-sight links," in *Antennas and Propagation (APSURSI), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 293–296.

[68] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace mimo for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 7, pp. 3814–3827, 2013.

[69] A. Sayeed and J. Brady, "Beamspace mimo for high-dimensional multiuser communication at millimeter-wave frequencies," in *Global Communications Conference (GLOBECOM), 2013 IEEE*. IEEE, 2013, pp. 3679–3684.

[70] J. Hansryd and J. Edstam, "Microwave Capacity Evolution," *Ericsson Review*, no. 1, pp. 22–27, 2011.

[71] K. S. Kim, B.-S. Kim, M.-S. Kang, W.-J. Byun, and H. C. Park, "16-qam ofdm-based w-band polarization-division duplex communication system with multi-gigabit performance," *ETRI Journal*, vol. 36, no. 2, pp. 206–213, 2014.

[72] I. Maric, B. Bostjancic, and A. Goldsmith, "Resource allocation for constrained backhaul in picocell networks," in *Information Theory and Applications Workshop (ITA), 2011*. IEEE, 2011, pp. 1–6.

[73] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with hybrid energy harvesting base station," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3412–3427, 2013.

[74] C. Li, S. Song, J. Zhang, and K. B. Letaief, "Maximizing energy efficiency in wireless networks with a minimum average throughput requirement," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2012, pp. 1130–1134.

[75] X. Kang, R. Zhang, and M. Motani, "Price-based resource allocation for spectrum-sharing femtocell networks: A stackelberg game approach," *IEEE Journal on Selected areas in Communications*, vol. 30, no. 3, pp. 538–549, 2012.

[76] P. Ezran, Y. Haddad, and M. Debbah, "Availability optimization in a ring-based network topology," *Computer Networks*, vol. 124, pp. 27–32, 2017.

[77] ——, "Polarization planning for wireless networks," *Wireless Personal Communications*, pp. 1–20, 2017.

[78] ——, "Allais' paradox and resource allocation in telecommunication networks," *submitted to Telecommunication Systems*, 2018.

[79] ——, "Polarization Diversity in Ring Topology Networks," *IEEE VTC 2016-Fall*, 2016.

[80] P. Ezran and Y. Haddad, "Wireless electromagnetic communication network using polarization diversity," May 26 2016, European Patent WO2016079748 (A1).

[81] J. Von Neumann and O. Morgenstern, *Theory of games and economic behavior*. Princeton university press, 2007.

[82] H. Markowitz, "Portfolio selection," *The journal of finance*, vol. 7, no. 1, pp. 77–91, 1952.

[83] Z. Bodie, A. Kane, and A. J. Marcus, *Investment and portfolio management*. McGraw-Hill Irwin, 2011.

[84] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica: Journal of the econometric society*, pp. 263–291, 1979.

[85] M. Al-Kuwaiti, N. Kyriakopoulos, and S. Hussein, "A comparative analysis of network dependability, fault-tolerance, reliability, security, and survivability," *IEEE Communications Surveys & Tutorials 11.2*, pp. 106–124, 2009.

[86] J. P. Sterbenz, D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, and P. Smith, "Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines," *Computer Networks 54.8*, pp. 1245–1265, 2010.

[87] I. B. Gertsbakh, *Statistical Reliability Theory*. CRC Press, 1988.

[88] R. Jiang, M. J. Kim, and V. Makis, "Availability maximization under partial observations," *OR spectrum*, vol. 35, no. 3, pp. 691–710, 2013.

[89] V. Kolmogorov, "Blossom V: a new implementation of a minimum cost perfect matching algorithm," *Mathematical Programming Computation 1.1*, pp. 43–47, 2009.

[90] E. Hazan, S. Safra, and O. Schwartz, "On the complexity of approximating k-set packing," *computational complexity 15.1*, pp. 20–39, 2006.

[91] Qualcomm, "The 1000x Mobile Data Challenge," Nov. 2013. [Online]. Available: https://www.qualcomm.com/1000x/

[92] D. Bojic *et al.*, "Advanced Wireless and Optical Technologies for Small-Cell Mobile Backhaul with Dynamic Software-Defined Management," *IEEE-M-COM*, vol. 51, no. 9, pp. 86–93, Sep. 2013.

[93] S.-C. Kwon and A. F. Molisch, "Capacity maximization with polarization-agile antennas in the mimo communication system," in *Global Communications Conference (GLOBECOM), 2015 IEEE*. IEEE, 2015, pp. 1–6.

[94] O. H. Karabey, S. Bildik, S. Bausch, S. Strunck, A. Gaebler, and R. Jakoby, "Continuously polarization agile antenna by using liquid crystal-based tunable variable delay lines," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 1, pp. 70–76, 2013.

[95] J.-J. Xiao and A. Nehorai, "Optimal polarized beampattern synthesis using a vector antenna array," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 576–587, 2009.

[96] B. Fuchs and J. J. Fuchs, "Optimal polarization synthesis of arbitrary arrays with focused power pattern," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 12, pp. 4512–4519, 2011.

[97] F. Pujol, "Mobile traffic forecasts 2010-2020 & offloading solutions," *IDATE Consulting and Research, May*, vol. 15, 2011.

[98] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking (ToN)*, vol. 8, no. 5, pp. 556–567, 2000.

[99] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[100] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, *An axiomatic theory of fairness in network resource allocation.* IEEE, 2010.

[101] C. Joe-Wong, S. Sen, T. Lan, and M. Chiang, "Multiresource allocation: Fairness-efficiency tradeoffs in a unifying framework," *IEEE/ACM Transactions on Networking (TON)*, vol. 21, no. 6, pp. 1785–1798, 2013.

[102] F. Fossati, S. Moretti, and S. Secci, "A mood value for fair resource allocations," *arXiv preprint arXiv:1701.07991*, 2017.

[103] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.

[104] L. Le and E. Hossain, "Multihop cellular networks: Potential gains, research challenges, and a resource allocation framework," *IEEE Communications Magazine*, vol. 45, no. 9, pp. 66–73, 2007.

[105] I. Koutsopoulos and L. Tassiulas, "Adaptive resource allocation in SDMA-based wireless broadband networks with OFDM signaling," in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3. IEEE, 2002, pp. 1376–1385.

[106] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *2006 IEEE International Symposium on Information Theory.* IEEE, 2006, pp. 1394–1398.

[107] F. Kelly, "Charging and rate control for elastic traffic," *European transactions on Telecommunications*, vol. 8, no. 1, pp. 33–37, 1997.

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

[109] B. Rouzbehani, L. M. Correia, and L. Caeito, "A Model for Virtual Radio Resource Management in C-RAN," *EURO-COST CA15104 TD(16)02012*, 2016.

[110] M. Allais, "The so-called Allais paradox and rational decisions under uncertainty," in *Expected utility hypotheses and the Allais paradox*. Springer, 1979, pp. 437–681.

[111] W. Ogryczak, H. Luss, M. Pióro, D. Nace, and A. Tomaszewski, "Fair optimization and networks: A survey," *Journal of Applied Mathematics*, vol. 2014, 2014.

[112] S. Huaizhou, R. V. Prasad, E. Onur, and I. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 5–24, 2014.

[113] D. Bertsimas, V. F. Farias, and N. Trichakis, "On the efficiency-fairness trade-off," *Management Science*, vol. 58, no. 12, pp. 2234–2250, 2012.

[114] E. B. Rodrigues and F. Casadevall, "Control of the trade-off between resource efficiency and user fairness in wireless networks using utility-based adaptive resource allocation," *IEEE Communications Magazine*, vol. 49, no. 9, 2011.

[115] R. Combes, Z. Altman, and E. Altman, "On the use of packet scheduling in self-optimization processes: application to coverage-capacity optimization," in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2010 Proceedings of the 8th International Symposium on*. IEEE, 2010, pp. 98–107.

ÉCOLE DOCTORALE

Sciences et technologies de l'information et de la communication (STIC)

Titre : Optimisation de la Topologie des Réseaux Sans Fils

Mots clés : Réseaux cellulaires, topologie, anneaux, disponibilité, résilience, polarisation, réutilisation des fréquences, allocation de ressources, équité, efficacité du réseau

Résumé : L'industrie des télécommunications sans fil fait actuellement face à une croissance considérable pour des débits toujours plus hauts, stimulée par le développement des services mobiles de données. Ce développement rend le spectre disponible de plus en plus rare et nécessite des solutions afin d'optimiser l'usage de ses ressources limitées.

Le principal défi auquel les réseaux sans fils font face est de maximiser la disponibilité, la résilience et la qualité de service, tout en minimisant les coûts et en assurant entre les utilisateurs une allocation de ressources équitable.

Cette thèse tente de présenter des solutions à ces problèmes et se focalise sur trois thèmes.

Sur le premier thème, le but est de trouver la topologie en anneau qui optimise la disponibilité. Il est montré que les algorithmes développés dans le cadre de la théorie des graphes peuvent être utilisés de manière efficace pour définir en temps polynomial la topologie en anneau optimale si les anneaux sont petits (deux nœuds en plus du nœud d'agrégation). Pour les anneaux plus grands, le problème est NP-hard.

Le deuxième thème concerne la polarisation. Nous proposons une solution innovante qui peut améliorer l'efficacité spectrale jusqu'à 50% par comparaison avec l'état de l'art. Le paradigme proposé introduit de nouvelles perspectives au sujet de l'optimisation de la topologie et de l'allocation de canal.

Le troisième thème concerne l'allocation de ressources. Nous remettons en question l'approche présente, basée sur l'optimisation de l'efficacité du réseau. Nous montrons que cette approche est similaire au modèle d'utilité espérée de Bernoulli, qui a été réfuté par les paradoxes d'Allais. C'est pourquoi nous introduisons le concept d'aversion au manque d'équité et considérons la question d'allocation de ressources comme un compromis entre efficacité du réseau et équité.

Title : Topology Optimization of Wireless Networks

Keywords : Cellular networks, topology, ring networks, availability, resiliency, polarization, frequency reuse, resource allocation, fairness, network efficiency

Abstract: The wireless telecommunication sector is presently facing a tremendous growth of demand for higher data rates, driven by the development of mobile data services. This development makes the available spectrum scarcer and scarcer and requires solutions in order to optimize the use of its limited resources.

The main challenge wireless networks are facing is to maximize availability, resiliency and Quality of Service, while minimizing costs and ensuring fair resource allocation among users.

The present thesis will try to present solutions to these issues and will focus on three topics.

On the first topic, the purpose is to find the ring-based topology which optimizes availability. It will be shown that algorithms which have been developed in the field of graph theory can be used efficiently to define in polynomial time the optimal ring network topology if the rings are small (two nodes in addition to the aggregation node). For bigger rings, the problem will be NP-hard.

The second topic deals with polarization. We propose an innovative solution which can improve spectral efficiency in wireless ring networks by up to 50% in comparison with the state of the art. The proposed paradigm brings new perspectives regarding topology optimization and channel allocation.

The third topic deals with resource allocation. We question the present approach based on optimization of network efficiency. We show that this approach is similar to Bernoulli's expected utility model, which has been disproved by Allais' paradoxes. For this reason, we introduce the concept of unfairness aversion and consider the question of resource allocation as a trade-off between network efficiency and fairness.