جامعة الحسن الأول
UNIVERSITÉ HASSAN 1ER

كلية العلوم والتقنيات سطات
FACULTÉ DES SCIENCES ET TECHNIQUES SETTAT

Université Hassan 1er
Centre d'Études Doctorales

Faculté des Sciences et Techniques
Settat

THÈSE DE DOCTORAT

Pour l'obtention de grade de Docteur en Sciences et Techniques

Formation Doctorale : Mathématiques, Informatique et Applications

Spécialité : **Mathématiques Appliquées**

Sous le thème

A Novel Blind Handover Technique for Heterogeneous Mobile Networks based on Analytical Approaches

Présentée par :

Adnane EL HANJRI

Soutenue le : **15/06/2022**

A la Faculté des Sciences et Techniques de Settat devant le jury composé de :

Pr. Mohammed KHALIS	**PES**	**FST, Settat**	**Président**
Pr. Mounir RIFI	**PES**	**EST, Casablanca**	**Rapporteur**
Pr. Driss EL OUADGHIRI	**PES**	**FS, Meknès**	**Rapporteur**
Pr. Mohamed HANINI	**PH**	**FST, Settat**	**Rapporteur**
Pr. Abdelghani BEN TAHAR	**PES**	**FST, Settat**	**Examinateur**
Pr. Said EL KAFHALI	**PH**	**FST, Settat**	**Examinateur**
Pr. Aawatif HAYAR	**PES**	**ENSEM, Casablanca**	**Co-encadrante de thèse**
Pr. Abdelkrim HAQIQ	**PES**	**FST, Settat**	**Directeur de thèse**

Année Universitaire : 2021/2022

Acknowledgments

This dissertation summarizes the research works conducted in the Computer, Networks, Mobility and Modeling laboratory (IR2M) of the Department of Mathematics and Computer Science at Hassan First University of Settat.

First of all, special appreciation goes to my supervisor, Dr. Abdelkrim HAQIQ, Professor at the Faculty of Sciences and Techniques, Hassan First University, who supervised me throughout this thesis and who shared his brilliant intuitions, ideas and advice. I am grateful to him for his kindness, permanent availability, support, and encouragements. I would also like to thank him for his trust in me, which allowed me to develop my ideas and gave me clear insights.

I would like to thank Dr. Aawatif HAYAR, Professor at the National School of Electricity and Mechanics, Hassan Second University and Minister of Solidarity, Social Inclusion and Family. This thesis is the fruit of a collaboration of more than five years with her. Thanks to her, I understood what rigor and precision mean. Her help and her availability are really appreciated.

I extend my thanks to Dr. Mohammed KHALIS, Professor at the Faculty of Sciences and Techniques of Settat, for giving me the honor of being the President of the jury of my thesis defense.

I would also like to thank Dr. Mounir RIFI, Dr. Driss EL OUADGHIRI, and Dr. Mohamed HANINI, for having accepted to devote their precious time to review and examine this report. The high quality and relevance of their comments will allow me to improve it.

I would also like to thank Dr. Abdelghani BEN TAHAR and Dr. Said EL KAFHALI, who kindly acted as examiners.

My deep and sincere gratitude goes to all the members of my family, and to all my friends, who have contributed, in one way or another, to the success of this work.

Abstract

Information and Communication technologies established a very quick evaluation and deployment. Last decades has been characterized by developmental changes in mobile communication networks. The fifth-generation networks (5G) has recently emerged to improve the quality of service to users with very high throughput.

Among the challenges of this new technology (5G), we find the problem of Handover Management, since it suggests to consider the deployment of Small Cells over Macro cells layer which introduces Heterogeneous networks. This problem must be solved in order to maintain the connectivity every time and everywhere for all mobile users.

This thesis is interested in this problem and therefore proposes a new approach, to have an efficient, blind and rapid handover just by analyzing the received signal density function instead of demodulating and analyzing the received signal itself as in classical handover. The proposed approach exploits some mathematical tools like, Kullback Leibler Distance, Akaike Information Criterion and Akaike Weight.

We first modelled the wireless communication signal received using the Rayleigh and Rice distributions, and then applied our approach to detect the best handover. After that, we have proposed another approach based on a mathematical tool called "Compressive Sampling", to select instantly and at low energy cost the appropriate cell without degradation of the primary signal sparsity to keep the linearity and the properties of the original signal in order to be able to apply the distribution analysis detector on the compressed measurements.

Finally, we estimated the performance of the networks in terms of channel capacity and outage probability based on Nakagami distribution, which has modeled a wide class of transmission channel conditions. The proposed approaches were evaluated and numerically validated.

Keywords: 5G Networks, Heterogeneous Networks, Small Cells, Macro Cells, Handovers, Information Theory, Performance Evaluation, Nakagami Distribution, Rayleigh and Rice Distribution, Kullback Leibler Distance.

Résumé

Les technologies de l'information et de la communication connaissent une évolution et un déploiement très rapides. Ces dernières décennies ont été caractérisées par des développements considérables dans le domaine des réseaux de communications mobiles. La cinquième génération de ces réseaux (5G) a récemment émergé pour améliorer davantage la qualité de service offerte aux utilisateurs avec des débits très élevés.

Parmi les défis de cette nouvelle technologie (5G), on trouve le problème de gestion de la mobilité inter-cellulaires (problème de Handover), car elle propose l'utilisation des petites cellules avec des macro cellules, ce qui rend le réseau hétérogène. Il faut donc bien résoudre ce problème pour maintenir la connectivité à chaque instant et partout pour tous les utilisateurs mobiles.

Cette thèse s'intéresse à ce problème et donc propose une nouvelle approche, pour avoir un Handover efficace, aveugle et rapide en analysant juste la fonction de densité de probabilité du signal reçu au lieu de démoduler et d'analyser le signal reçu lui-même comme dans le Handover classique. Notre approche est basée sur des outils mathématiques tels que, la distance de Kullback Leibler, le critère d'information d'Akaike et le poids d'Akaike.

Nous avons d'abord modélisé le signal de communication sans fil reçu en utilisant les distributions de Rayleigh et Rice, puis nous avons appliqué cette approche pour détecter le meilleur Handover. Ensuite, nous avons proposé une autre approche basée sur un outil mathématique appelé "Compressive Sampling", pour sélectionner instantanément et à moindre coût énergétique, la cellule adéquate sans dégradation de la qualité de communication, en profitant de la rareté du signal primaire pour conserver la linéarité et les propriétés du signal d'origine, et ceci afin de pouvoir appliquer le détecteur d'analyse de distribution sur les mesures compressées.

A la fin, nous avons estimé les performances du réseau en termes de capacité du canal et de la probabilité de coupure en se basant sur la distribution de Nakagami, qui a la capacité de modéliser les différents types de canaux de transmissions. Les approches proposées ont été évaluées et validées numériquement.

Mots-clés: Réseaux 5G, Réseaux Mobiles Hétérogènes, Petites Cellules, Macro Cellules, Mobilité Inter-Cellulaires, Théorie de l'Information, Évaluation des Performances, Distribution de Nakagami, Distribution de Rayleigh et Rice, Distance de Kullback Leibler.

Contents

List of Abreviations

- **3GPP** Third Generation Partnership Project
- **AI** Artificial Intelligence
- **AIC** Akaike Information Criterion
- **AICc** corrected AIC
- **AMPS** Advanced Mobile Phone Service
- **APs** Access Points
- **BDMA** Beam Division Multiple Access
- **BER** Bit Error Ratio
- **BIC** Bayesian Information Criterion
- **BS** Base Station
- **CDMA** Code Division Multiple Access
- **ComP** Coordinated Multiple Point
- **CS** Compressive Sampling
- **DAD** Distribution Analysis Detector
- **DVB** Digital Video Broadcasting
- **EMP** Entropy Maximization Principle
- **eNB** evolved NodeB
- **EVDO** Evolution-Data Optimized
- **FBMC** Filter Bank Multi Carrier
- **GSM** Global Systems for Mobile communications
- **HetNets** Heterogeneous Networks
- **HSUPA** High Speed Uplink Packet Access

- **HSDPA** High Speed Downlink Packet Access
- **ICT** Information and Communications Technologies
- **IoT** Internet of Things
- **IoV** Internet of Vehicles
- **IP** Internet Protocol
- **IT** Information Theory
- **ITS** Intelligent Transportation Systems
- **KLD** Kullback Leibler Distance
- **Li-Fi** Light Fidelity
- **LLN** Law of Large Numbers
- **LoS** Line-of-Sight
- **LTE** Long Term Evolution
- **MAHO** Mobile Assisted Handover
- **MAICE** Minimum AIC Estimate
- **MCHD** Mobile Controlled Handover Decision
- **MLE** Maximum Likelihood Estimator
- **MMS** Multimedia Messaging Service
- **MNs** Mobile nodes
- **MS** Mobile Station
- **MSE** Mean Square Error
- **MVNO** Mobile Virtual Network Operator
- **NAHO** Network Assisted Handover
- **NCHD** Network Controlled Handover Decision
- **NFV** Network Function Virtualization
- **NLoS** non-Line-of-Sight
- **NMT** Nordic Mobile Telephon
- **PCI** Physical Cell Id
- **QoE** Quality of Experience

- **QoS** Quality of Service
- **RAN** Radio Access Network
- **R&D** Reaserch and Developement
- **RF** Radio Frequency
- **RMSE** Root Mean Square Error
- **RNC** Radio Network Controller
- **RRC** Radio Resource Connection
- **RSRP** Reference Signal Received Power
- **RSRQ** Reference Signal Received Quality
- **RSSI** Received Signal Strength Indicator
- **SCs** Small Cells
- **SDN** Software-Defined Networking
- **SINR** Signal-to-Interference-plus-Noise Ratio
- **SMS** Short Message Service
- **SNR** Signal-to-Noise Ratio
- **SP** Stochastic Process
- **TACS** Total Access Communication System
- **TDMA** Time Division Multiple Access
- **UE** User Equipment
- **UDN** Ultra-Dense Network
- **UMTS** Universal Mobile Telecommunications Systems
- **VXLAN** Virtual eXtensible LAN
- **WCDMA** Wideband Code Division Multiple Access
- **Wi-Fi** Wireless Fidelity
- **WIMAX** Worldwide Interoperability for Microwave Access

List of figures

List of tables

General Introduction

Contents

1 Thesis context

According to the United Nations about 55 % of the world's population lives in cities [1], and it is expected that the population will increase 1.5 times by 2050 [2]. The concentration of the population in urban areas causes problems such as an increase in traffic congestion, air pollution, crime, and environmental deterioration.

The Internet of Things (IoT) is a recent digital communication paradigm in which everyday life objects can communicate with each other and with the users using the Internet. Hence, the IoT aims to expand the Internet concept, making it more immerse, by enabling easy interaction with a wide variety of devices such as home appliances, surveillance cameras, industrial actuators, traffic lights, and vehicles. In this context, data are being generated and gathered from a vast number of connected devices. The integration of Cloud Computing and Big Data technologies plays a significant part in handling different types of data, according to the requirements, creating more valuable services. Such technologies are crucial to ensure the IoT paradigm in urban scenarios, which is known as Smart City.

Fueled by this unprecedented growth in the connected devices number, mobile data traffic, and the limitations of the 4G technologies to address this enormous data demand, industry and academia efforts are focused on defining the specifications for 5G services, signaling the dawn of the 5G era.

Fifth Generation (5G) has recently emerged to satisfy the increasing demand for high data bit rates. A device with 5G will be able to maintain network connectivity every moment and everywhere, opening the possibility to connect all the devices in the network. To this end, the basic 5G system design is expected to provide support for up to a million simultaneous connections per square kilometer, enabling the introduction of a variety of emerging concepts within IoT services.

A key piece of this shift is the deployment of Small Cells over the Macrocells layer which introduces a new type of network called Heterogeneous Networks (HetNets). But the high number of small cells and their low coverage imply more Handovers to provide continuous connectivity, and the selection, quickly and at low energy cost, of the appropriate one in the vicinity of thousands is also a key problem.

The work done in this thesis falls in the Handover Management. The goal is to investigate a new approach to manage handovers based on some Information Theory tools and estimate the performance of the network.

2 Thesis objectives

The main goal of this thesis is to propose a new approach in the Management of Handovers, to have an efficient, blind, and rapid handover. For this reason, we opted for Mathematical Modeling approach exploiting Information Theory concepts, in order to achieve the following objectives:

- Analyze received signal density function instead of demodulating and analyzing received signal itself as in classical handover;
- Estimate different models of signal distribution;
- Use a Mathematical technique to reduce computation complexity and energy consumption;
- Estimate the performance of the network in terms of Channel capacity and Outage probability;
- Compare between different models of signal distributions.

3 Contributions

The main contributions of this thesis are made for modeling the Handover Management approach using information theory concepts, they can be divided into three contributions:

- The first contribution consists of presenting a new approach, to have an efficient, blind, and rapid handover just by analyzing the received signal density function instead of demodulating and analyzing the received signal itself as in classical handover. The proposed approach exploits Kullback Leibler Distance (KLD), Akaike Information Criterion (AIC), and akaike weights, in order to decide blindly the best handover i.e. the best base station (BS) for each user.

  This contribution has been published in the following reference :

  Adnane EL HANJRI, Aawatif HAYAR and Abdelkrim HAQIQ, "Features detection based blind handover using kullback leibler distance for 5G HetNets systems", IAES International Journal of Artificial Intelligence (IJ-AI), Vol. 9, No. 2, pp. 193-202, 2020. (Scopus) [3]

- In the second contribution, we combine Compressive Sampling and the approach presented in the first contribution. The Compressive Sampling algorithm is designed to take advantage of the sparsity of the primary signal and to keep the linearity and properties of the original signal in order to be able to apply Distribution Analysis Detector on the compressed measurements, to reduce computation complexity.

  This contribution has been published in the following reference :

  Adnane EL HANJRI, Aawatif HAYAR and Abdelkrim HAQIQ, "Combined Compressive Sampling Techniques and Features Detection using Kullback Leibler Distance to Manage Handovers", Proceeding of the 5th IEEE International Smart Cities Conference (ISC2), Casablanca, Morocco, October 14-17, 2019. (IEEE) [4]

- The third contribution estimates the performance of networks in terms of Channel Capacity and Outage Probability based on Rice and Nakagami Distribution Models by using concepts from Information Theory, then a comparison between Rice distribution Model and Nakagami distribution Model is made.

  This contribution has been published in the following reference :

Adnane EL HANJRI, Aawatif HAYAR and Abdelkrim HAQIQ, "Blind Handover detection based on KLD and Channel Capacity, Outage Probability Estimation for Rice and Nakagami Models", IAENG International Journal of Computer Science, vol. 48, no. 4, pp 1087-1094, 2021. (Scopus) [5]

4 Thesis organization

The manuscript is composed of an abstract, un résumé, a general introduction, five chapters, a conclusion and future work, a list of publications, a bibliography and an appendix. After the General introduction which introduces the thesis context, objectives and also highlights the main contributions in this report, the remainder of this report is organized as follows:

1. Chapter 1 presents the background notions related to handovers. It starts by presenting the history of Information and Communication Technology. Then, introduces the Smart City concept and 5G networks. Finally, describes the handovers in details.

2. Chapter 2 presents in details the Mathematical tools used in this thesis report. First, describes random variables and stochastic processes. After that, describes Information Theory and finally, presents Kullback Leibler Distance, and Akaike Information Criterion.

3. In chapter 3, we present the first contribution, Features Detection based Blind Handover using Kullback Leibler Distance for 5G HetNets Systems, where a new approach, to have an efficient, blind and rapid handover just by analyzing Received Signal probability density function instead of demodulating and analyzing Received Signal itself as in classical handover, is presented. The proposed method exploits KL Distance, Akaike Information Criterion (AIC), and Akaike weights, in order to decide blindly the best handover and the best Base Station (BS) for each user.

4. In chapter 4, the second contribution is presented, Combined Compressive Sampling Techniques and Features Detection using Kullback Leibler Distance to Manage Handovers, we present a new Handover technique that combines Distribution Analysis Detector and Compressive Sampling Techniques. The Compressive Sampling algorithm is designed to take advantage of the primary signal's sparsity and to keep the linearity and properties of the original signal in order to be able to apply Distribution Analysis Detector on the compressed measurements.

5. In Chapter 5, the third contribution, Blind Handover detection based on KLD and Channel Capacity, Outage Probability Estimation for Rice and Nakagami Models, where we estimate the performance of networks in terms of Channel Capacity and Outage Probability based on Rice and Nakagami Distribution Models. We obtain Channel Capacity and Outage Probability by using concepts from Information Theory. Through Numerical evaluations, we show that Nakagami distribution Model is more efficient than Rice distribution Model.

This manuscript ends with a conclusion that summarizes the contributions and the principal results of the thesis, and gives some perspectives and possible future work.

Chapter 1

Survey on 5G Networks and Handovers

Contents

The purpose of this chapter is to provide an overview of 5G networks and also briefly discuss the problems of handovers.

1.1 Introduction

We have been witnessing exponential growth in the amount of traffic carried through mobile networks [6]. According to the Cisco visual networking index (Figure 1.1) [7], mobile data traffic has doubled during the last years, extrapolating this trend for the rest of the decade shows that global mobile traffic will increase 1000x from 2010 to 2030.

Figure 1.1 – Mobile Network growth

Apart from 1000x traffic growth, the increasing number of connected devices imposes another challenge on the future mobile network. It is envisaged that in the future connected society, having a profound socio-economic impact, and enriching our daily lives with a lot of services from media entertainment (e. g video) to more sensitive safety-critical applications (e.g. e-commerce, e-health..)

5G Technology [8] [9] has recently emerged to satisfy the increasing demand for high data bit rates and to change the world by connecting anything to anything from anywhere to anywhere. Faced with global exponential mobile data traffic, the deployment of 5G systems will encounter new challenges in terms of data rate, mobility support, and QoE (Quality of Experience).

In addition to utilizing more spectrum, the most powerful technique to address this data demand is through network densification, i.e. deploying more Small Cells (SCs) [10] to serve a geographical area and thereby achieving cell splitting gains.

Small cells by strict definition are low-power wireless access points that operate in the licensed spectrum to provide improved cellular coverage, capacity, and applications for homes and enterprises as well as metropolitan and rural public spaces.

Small cells are used to extend the coverage of mobile networks to indoor areas where outdoor signals do not penetrate well or to add network capacity in areas with very dense phone usage. They make the best use of the available spectrum by re-using the same frequencies many times within a geographical area.

It is clear why there was early and broad industry agreement that small cells will be a crucial component of 5G networks because they have the ability to significantly increase network capacity, density, and coverage, especially indoors. They are a relatively low-cost deployment option and, because they are low power devices, are relatively cheap and efficient to run to give a low total cost of ownership.

The deployment of SCs over the Macrocells layer introduces a new type of network called Heterogeneous Networks (HetNets) [11]. However, with the densification of mobile Heterogeneous Networks (HetNets) through the introduction of Femtocells at a large scale, the management of networking processes such as configuration, optimization, and maintenance is becoming a real burden for mobile operators. And, while it may sound like a highly technical term, a HetNet is simply the banding together of different sized cells to provide ultra-dense coverage in defined geographic areas.

Because of the low coverage of SCs, it is essential to support seamless handovers [12][13][14] to provide continuous connectivity and user-perceived Quality of Service (QoS), within any wide area network [15]. In addition, due to the high number of SCs, handovers increase, and the selection, quickly and at low energy cost, of the appropriate one in the vicinity of thousands is also a key problem.

Handover is the mechanism that transfers an ongoing call from one cell to another as a user moves through the coverage area of a cellular system. Every handover process contains three phases logically. The first step concerns the measurement or information gathering phase, where the User Equipment (UE) measures the signal strength of every potential neighbor BS and the current serving station. The second phase is about the handover decision, where the current serving BS decides about initializing the handover based on the measured data from the first stage. And the last one is the cell exchange when the UE releases the serving evolved NodeB (eNB) and connects to the new one.

There is Classical Handovers and Blind Handovers, but in our work, we are interested in the Blind Handover. An existing feature in which the network node, may initiate a handover procedure for a terminal without doing conventional measurement configuration and without considering measurement reports is Blind Handover. Actually, the efficiency of the handover process is a proportional relationship with the performance of the cellular network.

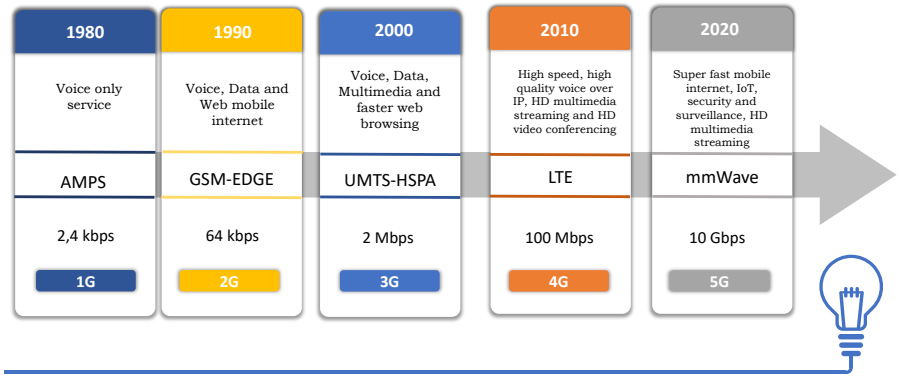

Figure 1.2 – Cellular Generations

1.2 History

Information and Communication technologies [16] established a very quick evolution and deployment, each decade has been characterized by developmental changes in mobile communications (Figure 1.2) [17], since 1^{st} generation cellular telephone. This generation was announced in the initial 1980s. It has a data rate of up to 2.4 kbps. Major subscribers were Advanced Mobile Phone Service (AMPS) [18], Nordic Mobile Telephone (NMT) [19], and Total Access Communication System (TACS) [20]. It has a lot of disadvantages like below par capacity, reckless handover, inferior voice associations, and no security since voice calls were stored and played in radio towers due to which vulnerability of these calls from unwanted eavesdropping by third party increases. This generation of cellular services notices an exponential growth rate in their subscriptions, and by the late 1980s capacity limits were already reached in the largest markets with (1G) cellular systems.

In reply to such massive demand, 2^{nd} Generation (2G) cellular systems were presented in the early 1990s. Digital technology is used in 2^{nd} generation mobile telephones. Global Systems for Mobile communications (GSM) [21] was the first 2^{nd} generation system, chiefly used for voice communication and having a data rate up to 64 kbps. 2G mobile handset battery lasts longer because of the radio signals having low power. It also provides services like Short Message Service (SMS) and e-mail. Vital eminent technologies were GSM, Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA).

Third Generation (3G) cellular systems were presented after the year 2000 [11]. This technology allows simultaneous use of speech and data services and with a higher data rate and giving rise to geolocation information. It imparts a transmission rate of up to 2 Mbps. Third-generation (3G) systems merge high-speed mobile access to services based on Internet Protocol

(IP). Aside from transmission rate, unconventional improvement was made for maintaining QoS. Additional amenities like global roaming and improved voice quality made 3G a remarkable generation. The major disadvantage for 3G handsets is that they require more power than most 2G models. Along with this 3G network, plans are more expensive than 2G. Since 3G involves the introduction and utilization of Wideband Code Division Multiple Access (WCDMA), Universal Mobile Telecommunications Systems (UMTS), and Code Division Multiple Access (CDMA) 2000 technologies, the evolving technologies like High-Speed Uplink/Downlink Packet Access (HSUPA/HSDPA) and Evolution-Data Optimized (EVDO) has made an intermediate wireless generation between 3G and 4G named 3.5 G with an improved data rate of 5-30 Mbps.

Recently, Fourth Generation (4G) [22] [23]cellular systems use voice over Internet Protocol (VoIP) and Multimedia applications with broadband access. These 4G systems are based on Multicarrier Modulation Multiplexing techniques or Advanced Simple Carrier Modulation Multiplexing techniques. 4G is generally referred to as the descendant of the 3G and 2G standards. 3^{th} Generation Partnership Project (3GPP) is presently standardizing Long Term Evolution (LTE) Advanced as forthcoming 4G standard along with Mobile Worldwide Interoperability for Microwave Access (WiMAX). A 4G system improves the prevailing communication networks by imparting a complete and reliable solution based on IP. Amenities like voice, data, and multimedia will be imparted to subscribers on every time and everywhere basis and at quite higher data rates as related to earlier generations. Applications that are being made to use a 4G network are Multimedia Messaging Service (MMS), Digital Video Broadcasting (DVB), and video chat, High Definition TV content, and mobile TV.

With an exponential increase in the demand of the users, 4G will now be easily replaced with 5G [8] [9] [24] with an advanced access technology named Beam Division Multiple Access (BDMA) and Non- and quasi-orthogonal or Filter Bank multi-carrier (FBMC) multiple access. The concept behind the BDMA technique is explained by considering the case of the base station communicating with the mobile stations. In this communication, an orthogonal beam is allocated to each mobile station and the BDMA technique will divide that antenna beam according to locations of the mobile stations for giving multiple accesses to the mobile stations, which correspondingly increase the capacity of the system. An idea to shift towards 5G [25] is based on current drifts, it is commonly assumed that 5G cellular networks must address six challenges that are not effectively addressed by 4G i.e. higher capacity, higher data rate, lower End to End latency, massive device connectivity, reduced cost and consistent Quality of Experience provisioning.

Currently, Information technology has become a part of our society, having a deep socio-economic impact and make our life easier with a lot of services from media entertainment to

commerce and health care. Scientists predict that in the future every physical object can be connected to the network (Internet of Things). Otherwise, in today's network, with the growth in Mobile traffic, energy consumption represents a key source of expenditure for operators.

A summary of comparison between 1G, 2G, 3G, 4G and 5G is shown in table 1.1

Table 1.1 – Comparison between 1G, 2G, 3G, 4G and 5G

Generations	1G	2G	3G	4G	5G
Start	1970–1984	1980–1999	1990–2002	2000–2010	2010–2020
Data Bandwidth	2 kbps	14–64 kbps	2 Mbps	100 Mbps	1–10 Gbps
Technology	AMPS NMI TACS	TDMA CDMA GPS	HSPA	LTE	LTE Advanced
Services	Voice only services	Voice data services web mobile internet	Browsing multimedia tv streaming video calling	High speed	Super Fast

Such technologies respond to the need of most national governments to adopt Information and Communications Technologies (ICT) solutions in the management of public affairs and they are crucial to ensure the IoT paradigm in urban scenarios, which is known as Smart City.

1.3 Smart City

Urbanization is one of the most important social-economic phenomena in today's world. As of now, according to the United Nations 55 % of the global population lives in cities [1]. As urbanization picks up speed, it is expected to exceed 70 % by 2050 [2]. As more people flock to cities, new mega-cities and city clusters begin to take form. At the same time, problems such as traffic congestion, pollution, resource scarcity, and lowered quality of life become more prominent, making sustainability a shared concern of city managers. The concept of a smart city [26] was conceived in such context. It advocates for the use of advanced technologies, particularly IT technologies, to achieve sustainable city development, and promises to be the best solution for sustainability for city managers.

Meanwhile, information and communication technologies (ICT) are burgeoning around the world. Key technologies like 5G network, IoT, cloud computing, big data analytics, and next-generation geoinformation system, once novel concepts on paper, are being experimented on and implemented in real applications. This has spawned new usage scenarios and innovative management models, bringing more possibilities to smart cities. As information technologies

mature, the condition becomes ripe for digitalization and smart management of the city, which can effectively solve urbanization-related problems, laying a solid foundation for smart city development.

1.3.1 Definition

The European Commission defines a smart city as "a place where traditional networks and services are made more efficient with the use of digital and telecommunication technologies for the benefit of its inhabitants and business" [27]. Diffusion and availability of new technologies are required to transform a city into a smart city, contributing to reach a high level of urban sustainable development and improved quality of life for its citizens, table 1.2.

Table 1.2 – 5G smart city envisioned

5G smart city envisioned	
Application	Typical Scenario
Smart Government	Major public emergency response Online one-stop government services Identification by facial recognition
Smart Env. Protection	Environment monitoring Smart garbage bin
Smart Security	UHD real-time monitoring Robot patrol Drone patrol
Smart Transportation	Remote/self-driving AR-assisted Navigation Smart traffic planning
Smart Power	Real-time grid monitoring Smart allocation of energy Remote grid maintenance
Smart Logistics	Autonomous driving transportation Drone delivery Real-time tracking of goods
Smart Healthcare	Auto collection of health data Remote surgery Remote diagnosis & treatment
Smart Education	Immersive teaching & learning Remote interactive learning
Smart Household	Furniture IoT, remote control Immersive entertainment

Smart cities use the Internet of Things (IoT) to collect real-time data to better understand how demand patterns are changing and respond with faster and lower-cost solutions. Broadly

speaking, digital city ecosystems are designed to run on ICT frameworks that connect several dedicated networks of mobile devices, sensors, connected cars, home appliances, communication gateways and data centers.

By 2025 the IoT trends suggest the number of connected devices worldwide will rise to 75 billion [28]. The increasing number of objects that interconnect generates an unprecedented volume of data that the city can analyze locally in order to make more informed decisions about what changes or new projects will most benefit residents. The term "Massive IoT" is a description of the enormous number of IoT sensors and devices that will be communicating with each other.

To achieve a vision in which millions of devices are connected, the IoT standard must ensure both scalability and versatility, offering enough capacity and network efficiency to connect millions of devices while also providing advanced features—such as longer battery life and a wider coverage area—to facilitate the expansion of new use cases. The existing 4G network was designed primarily to enhance mobile data services, however, it still suffers from numerous limitations. These limitations include poor support for simultaneous connections, high power consumption, and too high a price per bit. 5G is expected to unlock the potential of the IoT and be a driving force for the smart city by addressing and overcoming these issues.

1.3.2 Smart City & Socio economic satisfaction

Smart cities are aimed at improving the use of public resources, increasing the quality of the services with a focus on comfort, maintenance, and sustainability, while the operational costs of the public utilities are reduced, within an IoT framework [29]. In general, IoT-based smart city applications can be grouped into four categories, as can be seen in [30]. Personal and Home Applications is the first category and includes home appliances connected and ubiquitous e-healthcare services which help doctors monitor patients remotely [25]. Utility Applications is the second category and includes smart water network monitoring, air quality, video-based surveillance, public safety, and emergency services. The third category is Industrial Applications which usually consists of a network of industrial machinery within a production environment. The last category is centered in Intelligent Transportation Systems (ITS) or in general, Mobility Applications. The latter category includes emerging concepts such as autonomous vehicles, vehicle networks, traffic management, congestion control, among others.

5G technology has a number of features that will positively impact digital experiences and smart cities, (Figure 1.3). In addition to a higher speed to upload and download data, it ensures very short latency times and the ability to connect multiple devices at the same time.

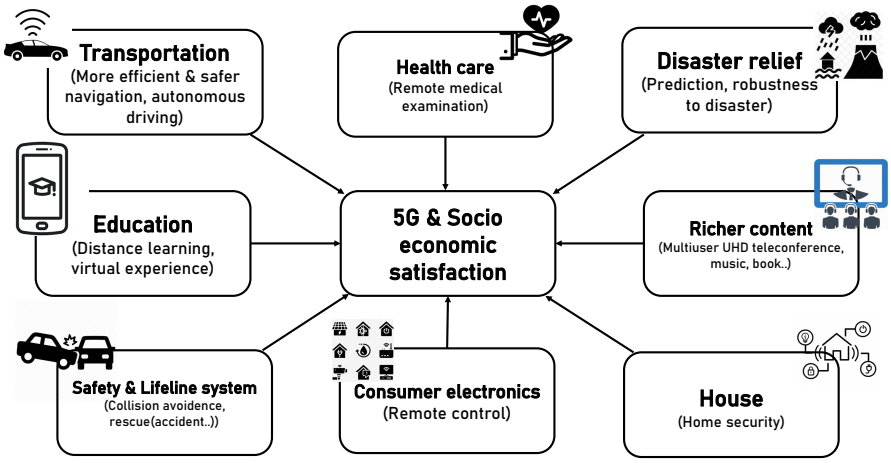

Figure 1.3 – 5G and socio economic satisfaction

Less latency means compressing the time between sending and receiving the signal. 5G brings the range to at least under 10 milliseconds (that is, half the most advanced 4G could achieve) and in best cases around 1-millisecond delays, meaning data will be transferred about in real-time.

Moreover, with new networks, speed and latency don't get worse even with tens of thousands of connected devices, 5G, therefore, offers more device density.

The combination of high density and low latency will deeply transform our cities. Today, in crowded vacation spots or at stadiums, the connection can sometimes get worse. With 5G it will no longer be like this, it will be possible to have a huge number (up to one million) of connections simultaneously for each square kilometer. This means, in addition to personal devices such as smartphones, tablets, smart speakers, and PCs, also many other devices, objects and sensors will be capable of capturing information and dialogue with each other. The focus will be on extreme simplicity, low-power consumption to ensure longer operation time and pervasive coverage for reaching challenging locations, as well as increased connection density so that networks can handle the massive number of devices deployed for IoT applications.

Therefore, 5G essentially removes one of the brakes on the development of the Internet of Things, (Figure 1.4), which will thus be able to express its potential not only in the home environment but also in industrial plants, in public buildings, or on the streets.

Ultimately 5G, moves the construction of smart cities from the theoretical to the practical and paves the way for the development and deployment of new applications ranging from monitoring air quality, energy use and traffic patterns to street lighting, smart parking, crowd

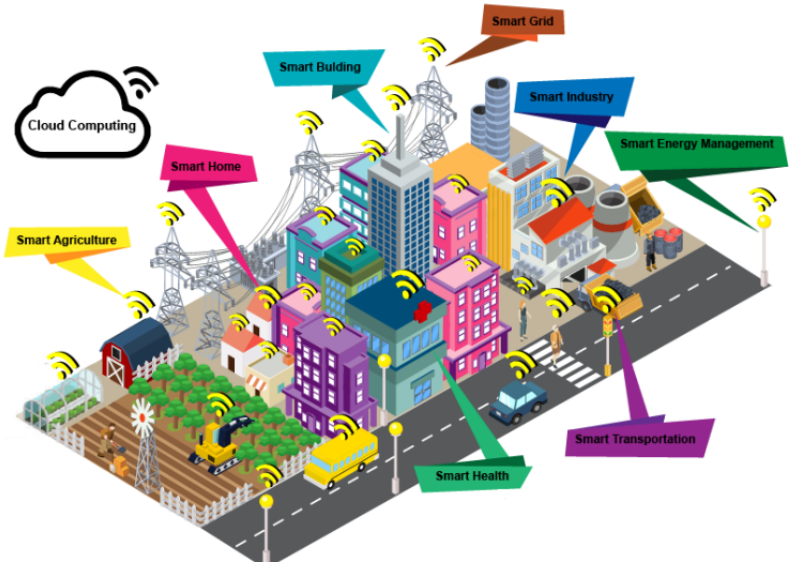

Figure 1.4 – Illustration of an IoT-based Smart City

management, and emergency response. The smart city uses digital solutions, technology, and data to improve significantly several key quality of life indicators. This leads to improved traffic, and commute time, accelerated emergency response time, reduced healthcare costs, decreased water consumption, unrecycled waste and harmful emissions, and ultimately in a huge saving potential.

Some of the most relevant approaches within each application category focusing on mobility applications are presented below:

A. Personal and Home Applications

Presently, a small percentage of people have a fitness device also known as a tag device, but the opportunities are vast and with 5G, smart tag devices are expected to become more prevalent. Unlike today's devices, future 5G devices will be fully connected since there will not be a need to be tied to a smartphone for internet access. Companies such as Samsung are developing health care and fitness devices that not only record exercise performance and make recommendations about exercise routines but also send to the user vital health information to an expert in real-time to prevent or monitor medical emergencies [31].

Moreover, with 5G, homes are expected to continuously become smarter through security (remote video security monitoring and control and wireless-controlled door locks), and comfort (command by voice, remote control using smartphones, and thermostat regulation).

B. Utilities Applications

Using 5G for urban IoT might provide monitor service of the whole energy consumption in the city, thus enabling authorities to access detailed and valuable information about the energy required by the different public services (e.g., public lighting, traffic lights, surveillance cameras, heating/cooling of public buildings, among others). This will allow identifying the main energy consumption sources and then planning in order to optimize city energy management.

In addition to the economic benefit of optimizing energy resources, 5G is expected to help public safety by saving lives through disaster and emergency response or improving crime detection and monitoring. Suspicious baggage in airports, vandalism, and criminal identification can be combated by using surveillance cameras and computer vision techniques. When the threat is detected, using a 5G fast connection, this will be informed to public safety personnel of the status of threats and might help to coordinate response actions. Moreover, in [32] is presented a security system which when detecting the face of a known criminal even if the crime has not yet been committed, the system capture live photos and actual location for sent to the nearby police stations.

C. Industrial Applications

The exponential rise of recent technologies (including big data, cloud computing, artificial intelligence, and 5G) has attracted great interest from industry to integrate ICT in the production environment. The melding of industrial machinery with ICT opens up opportunities to accelerate productivity, reduce waste, increase efficiency and improve the working experience in the production environment. Agriculture is a specific area where IoT has enormous potential. Using sensors with wireless connectivity for crop fields can help to optimize growing and minimize the use of water and fertilizers. Livestock, tanks, and other farm equipment can be monitored remotely, making farming more efficient by reducing production costs.

D. Mobility Applications

Increasingly, urban vehicles are becoming a moving sensor platform that provides environmental information to drivers, and soon such information could be uploaded to the cloud. The sensors data will be available to a network of autonomous vehicles that exchange their information with each other in order to optimize a well-defined function. Thus, vehicles would become another device connected to the Internet.

Ideally, when the human control is removed, the autonomous vehicles should cooperate to allow handling traffic more efficiently, with lower delays, less pollution, and better driver and passenger comfort. For instance, for disaster management, the vehicular network should be able to coordinate the evacuation of dangerous areas in a quick and orderly manner. This requires

being able to communicate with each other also have access to resources as ambulances, police vehicles, or information about escape routes, as shown in [33].

Nevertheless, due to the complexity of simultaneous control of hundreds of thousands of vehicles, current 4G technologies are not able to support such a large device density. Some other critical features as latency and quality of service are necessary to achieve it. For instance, it would take about 1.5 m for a vehicle with 4G to apply its brakes. While a vehicle with 5G would only require 2.5 cm to do so, helping avoid accidents. In the same way, if a vehicle enters an area with low coverage or is very populated, a 4G connection fails. However, a 5G connection theoretically will always have coverage, allowing keeping stable the connection anywhere and anytime.

Therefore, within the objectives of IoT and smart cities, vehicles play an important role that leads to the Internet of Vehicles (IoV) which is not only centred on the interaction between vehicles, but also on humans, cities or even countries.

1.3.3 Smart City & technology

Smart City makes use of Artificial Intelligence, cloud based services and Internet of Things (IoT) devices such as connected sensors, lights, and meters to collect and analyze data. The cities then use this data to improve infrastructure, public utilities, services and more.

A. Artificial Intelligence (AI)

As a city upgrades its production, lifestyle, and city management with smart technologies, strong demand will be generated for a new generation of AI technologies, products, services, and solutions. The 5G era fulfills the demand of smart cities for concurrent access to numerous intelligent devices and realizes millisecond-level response in device interaction, facilitating diverse AI applications to achieve the Intelligence of Things. In smart city applications, deep learning, as a type of AI, is gaining fast traction. As AI technologies, such as robotics, language recognition, image recognition, and natural language processing, become woven into the ubiquitous connections of smart cities, city managers will be in a better position to make informed decisions and provide intelligent public services.

B. The Internet of Things (IoT)

As smart city development deepens, in addition to computers, smartphones, and smart cameras, a wide variety of smart terminals are being deployed on a large scale, including smart robots, smart electricity meters, smart manhole covers, and smart industrial modules. As a new-generation network infrastructure and the cornerstone of the Intelligence of Everything,

5G and the wide connectivity it enables will facilitate the deployment of intelligent terminals and achieve ubiquitous connection among people and things. Through perception devices and connected things, data and information will be captured to form a massive peripheral nervous system of the city, providing solid support for digital twin city and giving city managers access to timely and accurate information.

C. Big data analysis

Data represents a strategic resource of the future and will be generated in large amounts by the smart terminal sensors deployed across the city. With high bandwidth and mass connectivity, 5G powers the entire process of big data analysis, including data acquisition, data fusion, data modeling and data mining, to extract value from mass city data and provide effective and timely support to city managers in city management and decision-making.

D. Cloud computing

Cloud computing offers flexible computing and a usage-based fee model, allowing information and resources to be coordinated and shared to the fullest extent on the "cloud". With cloud computing technology, physically dispersed computing power can be integrated and used for data storage and processing at the lowest possible costs, with the highest possible returns. With the high bandwidth of 5G, more data can be stored on the cloud. With its low latency, data uploading takes less time. With its enhanced load capacity, more IoT devices can connect to the cloud. Such a cloud-edge collaboration will improve the efficiency of business operation.

E. Blockchain

5G connectivity brings massive amounts of end-to-end information exchange, especially in large-scale business applications that pose higher requirements for security. The integration of 5G and the distributed ledger technology that underpins blockchain can be applied to information authentication, location and identification management, as well as spectrum sharing, etc. It will change the business model and architecture of future networks and drive a transformation from an information network to value network, extracting values inside the network and information assets.

1.4 5G Networks

Mobile networks, which have a 40-year history that parallels the Internet's, have undergone significant change. The first two generations supported voice and then text, with 3G defining the transition to broadband access, supporting data rates measured in hundreds of kilobits-per-second. Today, the industry is at 4G (supporting data rates typically measured in the few

megabits-per-second) and transitioning to 5G, with the promise of a tenfold increase in data rates.

From an architectural point of view, 5G defines an infrastructure for integrating a variety of access technologies, including both existing technologies such as the LTE network, new fixed or mobile access network technologies, while maintaining compatibility with existing technologies. In addition, 5G will integrate into a completely virtualized, flexible, and programmable environment, emerging paradigms of networking such as the software-defined network or Software-Defined Networking (SDN) and the virtualization of the functional components of the network or Network Function Virtualization (NFV).

From a spectrum perspective, 5G will be allocated new frequency bands to meet the growing demand for connection of connected devices, highly densified networks, and the many instances of mobile connectivity.

From the perspective of users and customers, 5G will define an ecosystem of access to ubiquitous services, very high data rates, very low latency, and greatly improved QoS and QoE.

The mobile networks of the future, including those of the 5G, are intended to combine within them telecommunications infrastructures as well as advanced management mechanisms based on the IP protocol. A wide variety of access networks or wireless communication technologies will allow mobile users or nodes (MNs) to access the services offered through several network attachment points or access points (APs).

Within communication networks, the term mobility refers to the possibility of accessing services, regardless of the location and movement of the user. There are three types of mobility: terminal mobility, service mobility, and personal mobility. We are interested in this work on terminal mobility. Mobility management has two components: location management (location management) and succession management (handover management).

1.4.1 Requirements of 5G

It is necessary to collect and accept the requirements of 5G to get a clear idea of the characteristics of systems, users and operations (Figure 1.5).

- High speed: It is defined as the theoretically achievable maximum data rate that can be assigned to a single mobile station assuming conditions without error when all available radio resources are used for the link corresponding (i.e., excluding radio resources used for

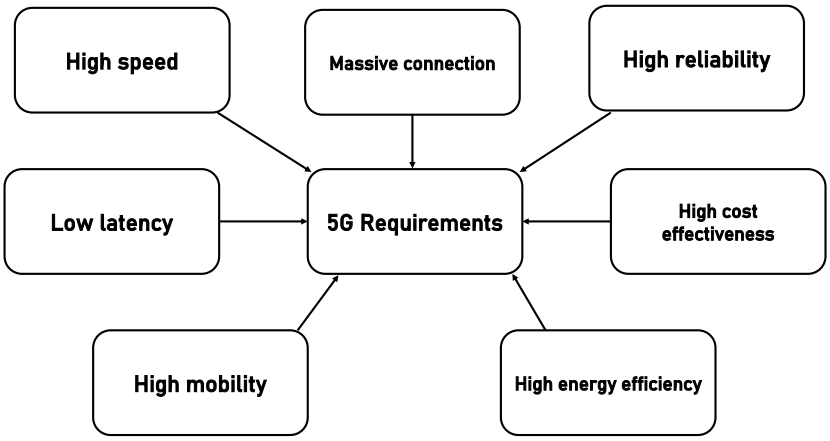

Figure 1.5 – 5G requirements

physical layer synchronization, reference signals or drivers, guard bands and warranties) 5G needs to:

  - Provide tens of Gbps peak data rate.
  - Provide up to 1 Gbps user experienced data rate.
  - Provide areal capacity of 10 Mbps per square meter.

- Energetic efficiency: Energy efficiency should be significantly improved compared to 4G in order to increase the battery life of connected objects (batteries that can be up to 100 times less energy intensive). Every effort must be made on energy optimization in order to have a good energy gain without degrading the performance of the network and devices. But the technology should allow a native flexibility for the operator to configure a compromise between energy efficiency and performance where warranted.

- Massive Connection:

  - Identify all devices over the world
  - Provide services to a million terminals per square kilometer.

- High reliability: Provide 99.99 % service availability even in an extrem situation.

- Mobility: Maintaining a very good level of service quality in the context of high-speed mobility is an essential requirement for 5G and above all for the development of autonomous cars.

- Cost effectiveness: Since the network functions always come with separate proprietary hardware entities, the deployment of new network services means a high cost for energy, capital investment challenges and rarity of skills necessary to design, integration and

operation of complex hardware-based appliances. Moreover, hardware-based appliances rapidly reach end of life, requiring much of the procure-design integrate-deploy cycle to be repeated with little or no revenue benefit. Worse, hardware life cycles are becoming shorter as technology and services innovation accelerates, inhibiting the roll out of new revenue earning network services and constraining innovation in an increasingly network-centric connected world. Therefore, it is necessary to decouple the software from hardware to architect operator's network towards deploying network services onto virtualized industry server, switch and storage. Besides, in order to cut down the investment of Mobile Virtual Network Operator (MVNO), future network should support network virtualization, which reduces the deployment of base station equipment and energy to run wireless network through sharing the network infrastructure of network operator. It is equivalent to say that the network resources can be utilized efficiently through network virtualization. Therefore, both decoupling the software from hardware and network virtualization are necessary for 5G network to reduce the CAPEX and OPEX, through reducing the equipment cost and improving the utilization of network resources. So 5G networks need to:

  - Improve cost effectiveness in network side even handling huge volume of traffic.
  - Reduction of devices' cost especially in sensors.

- Latency: This requirement is defined as the time required for a data packet to pass from source to destination through a network. The arrival of this new standard should allow an ultra-short latency time, equivalent to less than one tenth of the latency time of current communication systems. A very low latency will be motivated by the need to support new applications. Some cases of 5G use envisaged, such as connected and autonomous cars, emergency services, connected drones, remote medical operations, etc, may require a much lower latency than is possible with today's mobile communication systems.

Faced with these requirements, different technologies have aroused a certain enthusiasm in the scientific community. The main technologies that have been selected as the ideal candidate on the physical layer to meet the requirements of 5G are: Full duplex, Massive MIMO, Small cells in millimeter frequency bands and Heterogeneous network.

1.4.2 Key components of 5G

To achieve the objectives in terms of high throughput, reduction of energy consumption, latency, etc. The combination of various technologies will be necessary.

Massive MIMO: One of the solutions is the introduction of Massive MIMO for the next generation, ie the use of massive scale multiple antennas. The purpose of this technology is to increase the advantages of traditional MIMO systems. It refers to a scenario where specific technological components enable the cost effective deployment of cellular systems using hundreds of antennas in cellular base stations. This technique is one approach to increase channel capacity and to provide high multiplexing and diversity gains in the uplink and downlink directions. This performance will depend heavily on the number of antennas in the base station compared to the number of users.

Full duplex: It is currently at the heart of discussions on the definition of 5G and recent results obtained by academics have proven the technical feasibility. The advantage of this technology would also be recognition of the channel used in both directions. This allows the next generation to meet the latency and security requirements imposed especially in the case of stand-alone communications.

Small Cell: The idea of network approximation of reducing the distance between the last network access points and users is currently considered the only solution to meet coverage requirements. The capacity per user is reduced in the case of macrocells, due to the large number of users. Solving this type of problem amounts to creating small cells in these macrocells, in which the deployment of 5G radio communications equipment will take place. These cells are called "Small cells". Advanced techniques such as the cohabitation of different cell types ie heterogeneous networks, including network densification and backhaul, have emerged as key technologies applied in small cells to meet the requirements of 5G. The use of millimeter bands in these small cells will also be one of the most important technologies to provide high data rate services for the next generation.

Heterogeneous network: In telecommunications systems, a homogeneous network refers to the use of a standard RAN (Radio Access Network) made up only of base stations of the macrocell type. However, if the RAN is made up of different types of base stations (BSs), such as macrocells and different types of small cells, this is referred to as a heterogeneous network. Recent work on the deployment of small cells considers heterogeneous networks as a crucial means in the face of the growing demand for data traffic. The concept of this technology is based on a mixture of cells of different sizes (macro, micro, pico and / or femto) in order to increase the proximity between cells and mobile terminals. Considered an attractive solution for 5G networks, the heterogeneous network has been extensively detailed in Release 12 of the 3GPP standardization. Implementation will require a combination of a variety of small relay cells with different power levels under the cover of a macrocell. In order to produce a better service of quality and energy efficiency, the management of the change of cell will lead to constraints in real time, which is not tolerable with the arrival of autonomous cars. Operators

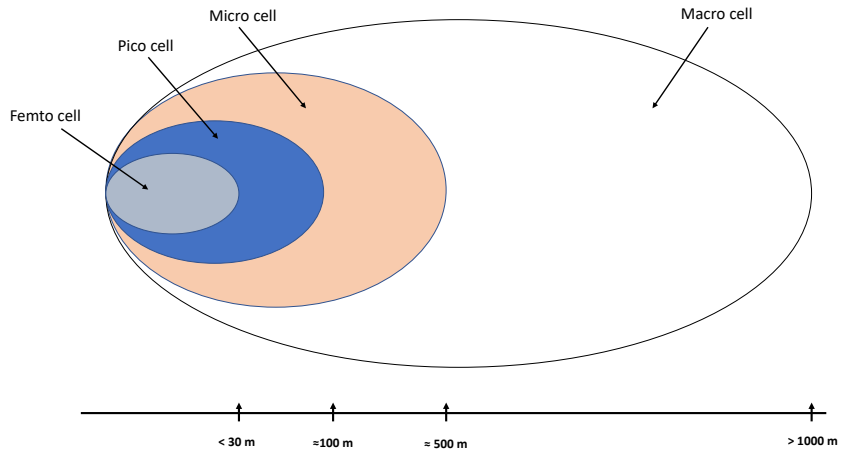

Figure 1.6 – Different types of Cells

will therefore have to find new methods of predicting mobility in order to solve this Handover problem.

1.4.3 Heterogeneous Network and Small cells

A heterogeneous 5G network incorporates two different kinds of cells namely macrocells and small cells. The primary operation of heterogeneous networks is that numerous smaller cells operate under the umbrella of macrocells to increase the coverage, promote frequency reuse, and enhance the ability to support high traffic rates for all areas.

Small cells are classified into microcells, picocells, and femtocells. The important features of macrocell and small cells for heterogeneous networks are as follows (Figure 1.6):

A. Macrocell

The cell diameter for a macrocell may range from 1 km (about) within the city to more than 40 km within a non-urban environment. The base station of a macrocell requires the installation of powerful antennas on the top of a tall building or a tower at a height of at least 30 m or more. With respect to the transmission power of the antennas, the value also can be set by 46 dBm. Moreover, the transceiver equipment of the base station is located within the same tower or the building, maintaining the temperature of the surrounding area. More precisely, even these transceiver devices are placed in the proximity of the antennas based on the reduced sizes as a result of the recent advancements.

In broadband wireless communication systems, a macrocell is regarded as the first element used for facilitation, and blanket coverage of an inhabited area. As a result, macrocells serve two

main objectives with the capability of blanket coverage whilst minimizing initial cost. Moreover, growth aspects of the network with respect to the inclusion of traffic and subscribers are also facilitated. Consequently, the radius of the macrocell is dependent on the following factors,

- Peak time facilitation regarding the management of the amount of the anticipated traffic.
- Maximization of the coverage distance.

Peak time facilitation of traffic is associated with the anticipated density of the subscriber, where increasing subscribers require enhanced traffic capacity. Accordingly, the service capacity of each base station is limited to facilitating a set number of users at a particular time. In addition to this, the macrocell coverage area can be small for a densely inhabited area. With respect to the relatively small population density of rural and suburban areas, base stations are capable of facilitating the coverage of a larger area. Therefore, the range of the cell radius entails several "kilometres" for suburban areas, several "hundred meters" for densely populated urban areas, and several “tens of kilometres” for rural areas.

B. Microcell

Compared to the macrocell, a microcell is smaller and possesses considerable differences in terms of requiring a smaller antenna with reduced power and power settings for transmission. These cells typically provide coverages up to 500 m and can be deployed indoors or outside to fill gaps in terms of both capacity and coverage. For instance, developed urban areas often/may encounter problems of congestion in terms of weakened signals. Moreover, microcell system designs could also be used in highly dense areas such as stadiums or concert arenas, for increasing the capacity of the system.

The base station of a microcell is generally directly interconnected with the central network. This connection is established using the link of an optical fibre. However, at certain times, increased traffic is routed by means of the base station of a macrocell, using a point-to-point wireless.

C. Picocell

A picocell serves as a wireless base station that is designed with very low power output for coverage of a very small area, like a single floor of an office building. The cellular network of picocells tends provide of coverage extensions for indoor areas that face weakened receiving of the signals. Moreover, network capacity is also increased in the areas with dense usage of phones, like airports, train stations or the shopping malls. With respect to the effectiveness of picocells, extensive research has been carried out for enhancing their cost-related and coverage-related performance in serving unapproachable places in the world.

D. Femtocell

Femtocells are smaller in coverage size and transmit power compared to picocells, microcells, and macrocells. Femtocells can be used in a home or small office buildings. The transmit power of femtocells is often less than 100 mw. Normally it can be seen that the coverage range of femtocells is less than 30 meters. Through Ethernet, femtocells are capable of using the cable modem or even a DSL (digital subscriber line) for backhauling data and voice calls within the range of the operator's network and the internet connection of the consumer. Hence, the mobile network of the operator is extended by means of using the internet connection of the consumer.

Mobility is undoubtedly a core feature of wireless communication systems. Seamless mobility across multiple cell coverage regions are achieved through handover mechanism in mobile networks. In the next section we will present the Handover.

1.5 Handovers

In a wireless network the need of managing the receivers mobility and ensuring continuous service led to the develop of procedures for a good network performance. This is achieved by supporting the handover algorithm, which is a procedure that can be applied to every kind of wireless communication network and it is also possible within the same radio system or between heterogeneous systems which are standardized by protocols developed by different standardization bodies.

The handover is the key to enabling the function of mobility and service continuity among a variety of wireless access technologies. It is the process of changing channels (frequency, time slot, spreading code, or combination of them) associated with the current connection while communication is in progress. In cellular telecommunications, the term handover refers to the procedure of transferring a call or a data section while the mobile station is moving away from a coverage area, called a cell, to another cell. This process is carried out to avoid the interruption of an in-progress call when the mobile gets outside the range of the cell.

In fact, a handover initiation is a process by which a handover is started as a consequence of the fact that the current link is unacceptably degraded and/or another base station can provide a better communication link.

1.5.1 Handover Process

Whatever the reasons for a mobile node to leave its current network to go to a new network (Handover), this process must be imperceptible to the user. Handover latency (time between

disconnection and reconnection) should not exceed a certain limit threshold, otherwise, this leads to a deterioration in the quality of service, especially for real-time applications. To achieve this goal, the Handover process is done through 3 phases:

A. Handover initiation and information gathering

A Handover process must begin when a mobile node needs to leave its point of attachment to the current network to go to connect to another network where the quality of service will be better. Usually, the reason may be low signal strength or a value of one or more quality of service parameters that fall below a certain threshold. During this phase, the mobile node continuously scans the networks in its surroundings, collecting the necessary information from each. This information is essential for the network selection phase. Among this information, we find those that are related to the network such as network coverage radius, packet loss rate, bandwidth, Bit Error Ratio (BER), Signal to Interference Ratio (SINR), etc. Other informations are more related to the mobile, such as signal strength, battery life, mobile speed.

As part of the scanning procedure, the UE measures the received signal strength from the selected base stations and reports the measurement results back to the serving BS using different methods:

Relative signal strength: This method works by selecting the BS from which the strongest signal is received. The decision is based on the computation of the average of the measurements of the received signal. It has been observed that this method generates many unnecessary handovers, even when the signal of the current BS is still at an acceptable level. Some variations have been proposed, as we see in the following.

Relative signal strength with threshold: This method allows a UE to hand off only if the current signal is sufficiently weak (less than threshold) and the other is the stronger of the two. The effect of the threshold depends on its relative value as compared to the signal strengths of the two BSs at the point at which they are equal. A threshold is not used alone in actual practice because its effectiveness depends on prior knowledge of the crossover signal strength between the current and candidate BSs.

Relative signal strength with hysteresis: This scheme allows a user to hand off only if the new BS is sufficiently stronger (by a hysteresis margin) than the current one. This technique prevents the so-called ping-pong effect, the repeated handover between two BSs caused by rapid fluctuations in the received signal strengths from both BSs. The first handover, however, may be unnecessary if the serving BS is sufficiently strong.

B. Handover Decision

To enhance the capacity and QoS of cellular systems, efficient handover algorithms are cost effective ways.

Several parameters have been proposed in the research literature for use in the handover decision algorithms. We briefly explain each of them below:

Handover delay: is the duration between the initiation and completion of the handover process, and is related to the complexity of the Handover Decision process. Reduction of the handover delay is especially important for delay sensitive voice or multimedia applications.

Number of handovers: Reducing the number of handovers is usually preferred as frequent handovers would cause wastage of network resources. A handover is considered to be superfluous when a handover back to the original point of attachment is needed within certain time duration, and the number of such handovers should be minimized.

Handover failure probability: A handover failure occurs when the handover is initiated but the target network does not have sufficient resources to complete it, or when the mobile terminal moves out of the coverage of the target network before the process is finalized. In the former case, the handover failure probability is related to the channel availability of the target network, while in the latter case it is related to the mobility of the user.

Throughput: The throughput refers to the data rate delivered to the mobile terminals on the network. Handover to a network candidate with higher throughput is usually desirable.

C. Handover Execution

In the previous phase, we have chosen the network to which the mobile node should connect. During this phase, the execution of the cut links with the old network and the connection with the new network is made. This execution can be done according to one of the following 4 cases: (1) Decision controlled by the network (Network Controlled Handover Decision, NCHD) usually used by operators to distribute network loads. (2) Mobile Controlled Handover Decision (MCHD). (3) Handover initiated by the network and assisted by the mobile (network-initiated but Mobile Assisted Handover, MAHO) and, (4) Handover initiated by the mobile and assisted by the network (mobile-initiated but Network Assisted Handover NAHO).

1.5.2 Types of Handover

The Handover is classified as horizontal (intra-system) and vertical (inter-system). Vertical refers to the overlapping of wireless networks (Figure 1.7).

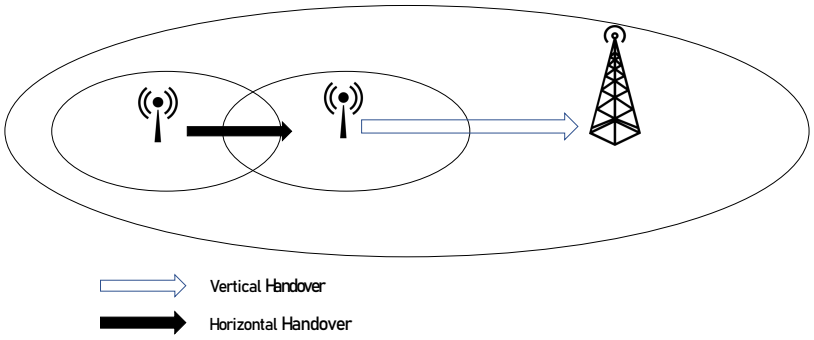

Figure 1.7 – Horizontal Handover vs Vertical Handover

A mobile node moves with the single technology network from one access point to the other one is a horizontal handover. For example, if a mobile user moves from one base station to the base station of another area, then the mobile user of the GSM network makes a horizontal handover. A horizontal handover is a traditional handover.

The transformation of an ongoing session or call from one cell to another cell having different access technologies is called Vertical Handover. For example when a mobile user is moving from a GSM-based network to the UMTS network, here the access technologies are changed so the handover in this case is the Vertical Handover.

In addition to the above classification of inter-cell and intra-cell classification of handovers, they also can be divided into hard and soft handovers (Figure 1.8):

Hard Handover is one in which the channel in the source cell is released and only then the channel in the target cell is engaged. Thus the connection to the source is broken before or 'as' the connection to the target is made—for this reason, such handovers are also known as break-before-make. Hard handovers are intended to be instantaneous in order to minimize the disruption to the call. A hard handover is perceived by network engineers as an event during the call. It requires the least processing by the network providing service. When the mobile is between base stations, then the mobile can switch with any of the base stations, so the base stations bounce the link with the mobile back and forth. This is called "ping-pong".

Soft Handover is one in which the channel in the source cell is retained and used for a while in parallel with the channel in the target cell. In this case, the connection to the target is established before the connection to the source is broken, hence this handover is called

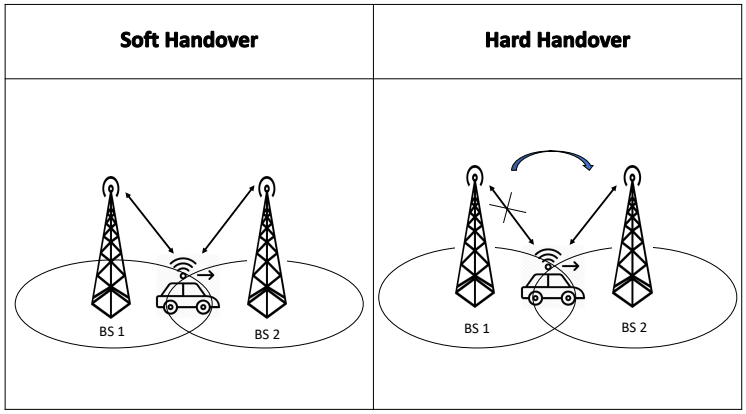

Figure 1.8 – Soft Handover vs Hard Handover

make-before-break. The interval, during which the two connections are used in parallel, maybe brief or substantial. For this reason, the soft handover is perceived by network engineers as a state of the call, rather than a brief event. Soft handovers may involve using connections to more than two cells: connections to three, four or more cells can be maintained by one phone at the same time. When a call is in a state of soft handover, the signal of the best of all used channels can be used for the call at a given moment or all the signals can be combined to produce a clearer copy of the signal. The latter is more advantageous, and when such combining is performed both in the downlink (forward link) and the uplink (reverse link) the handover is termed softer. Softer handovers are possible when the cells involved in the handovers have a single cell site.

1.5.3 Handover related challenges

While there are many open challenges in designing Ultra-Dense Networks (UDN), (Figure 1.9), our focus here is on identifying a solution to the problem of handover management during mobility. Understandably, the high number of small cell and mobile node deployments on the UDN increases the handover count[34]. This situation can result in a large accumulation of unnecessary and frequent handovers. Specifically, if these frequent handovers occur among the target and presently serving cells continuously, a back*and-forth signaling storm (the so-called ping-pong handover problem) is observed. Thus, network resources and energy get consumed at more than the usual rate because of the control traffic spike, which can also increase the risk of handover failure. Moreover, the mobility-related signaling overhead between the mobile node, and serving and target eNBs is increased [24].

To solve these problems, different mobility management algorithms are proposed in the current literature. In [35] handover mechanism with a modified signaling procedure is proposed to solve the unnecessary handover problem. Also, [36] proposes a state-dependent handover decision

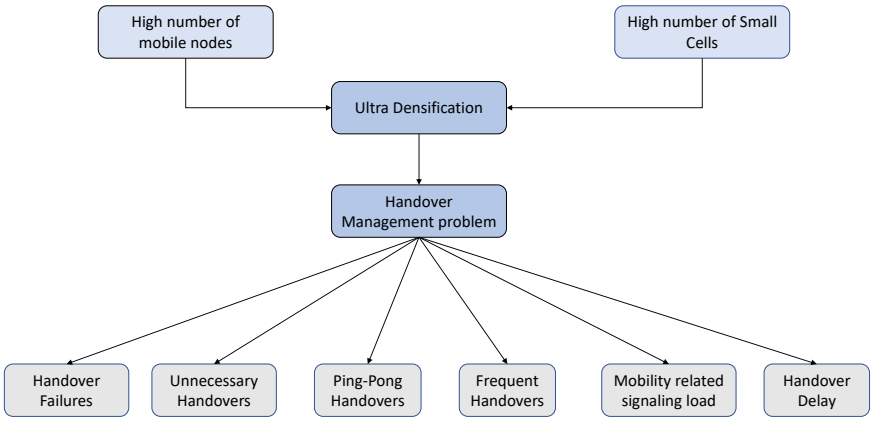

Figure 1.9 – Handover Problem Tree

algorithm to reduce the handover failure rate and improve the small cell utilization. Moreover, [37] proposes a cooperation-based cell clustering scheme to decrease the frequent handovers in dense small cell networks. Additionally, [38] investigates the relation between handover failure and ping-pong rates during the handover process. These works solve only some specific handover problems.

Also, practical delays observed during the handover procedure, and the unique scenarios emerging from the data/control channel separated architecture of 5G networks are not considered in these works. On the other hand, in the special 5G architecture control, data channels are managed by the different macro cells and small cells. Accordingly, during mobility management, these two different cell connections should be handled at the same time. Moreover, the handover delay is cumulative if the same device undergoes multiple handovers, resulting in severe impairment to the end-user experience. Therefore, we believe that minimizing the handover delay is a key issue in the design of future ultra-dense 5G networks. Additionally, any delay management scheme for handovers in 5G networks must be executed in two tiers, for both control and data channels, which has not been investigated so far.

In the Third Generation Partnership Project (3GPP) LTE handover standards [39] (based on the X2 interface), in the handover preparation phase, mobile nodes measure the RRM parameters, such as reference signal received power (RSRP) and reference signal received quality (RSRQ), of a high number of evolved node BS (eNBs) to choose the eNBs that provide the triggering condition (e.g., RSRP higher than a threshold) (so-called searching process). Then the mobile node transfers these measurement reports to the serving eNB. The serving eNB decides the handover by using these results, and the handover request is sent to the target eNB. According to the admission control results of the target eNB, the

handover acknowledgment message is sent to the serving eNB. The searching process and resulting mobility-related signaling overhead increase the handover delay. More specifically, this handover delay observed in the handover preparation phase to access the best target eNB becomes cumulative in UDNs.

Each small cell receives a large number of handover requests, followed by the local execution of the admission control algorithm for each accepted request. If these incoming handover request arrival rates are greater than the admission control rate, a high number of the requests wait in the queue of the target eNB. Also, the excessively long time to empty its queue means that the requests wait idly in the queue, and this situation further contributes to delays.

To alleviate the above issues, specific 5G architecture requirements arising from the densification of mobile nodes and small cells should be considered during the mobility management in 5G UDN architecture. Unlike the conventional mechanisms, different approaches based on software-defined networks (SDNs) and stochastic geometry concepts are proposed for solving the handover delay problem [35] [36] [37] [38]. However, these works do not consider an admission control mechanism to estimate available resources in the target eNB.

1.6 Conclusion

In this chapter, an introduction to 5G networks is presented in the context of the Smart City environment. Thereafter, we have given the historical evolution of Information and Communication technologies. In this chapter, we also presented the Handover process, types, and related challenges. In the following chapter, the mathematical tools for modeling our problem and presenting our approaches will be detailed.

Chapter 2

Mathematical Tools

Contents

In this chapter, the mathematical tools are introduced.

2.1 Introduction

The continuing evolution of computer systems and telecommunication networks emphasizes the growing need for tools that facilitate the study of their behavior, in order to have a reliable, efficient, and effective system. This is why there is a need for models and tools to analyze the system in all these phases in order to optimize its resources and improve its performance.

Indeed, the development of a complex system requires not only qualitative modeling to verify its Logical correctness but also a priori validation of the performance of the system during the design phase. Performance metrics (such as Probability of Detection, Channel Capacity, Outage probability, etc.) are generally very difficult to calculate.

In this sense that analytical modeling, based on Information theory and the theory of stochastic processes, appears to be a powerful approach due to their capacity and flexibility to model large systems. They are powerful tools for evaluating the performance of telecommunication networks.

The purpose of this chapter is to present the Mathematical tools used for the numerical application of our contributions, Information theory, Stochastic process, Kullback Leibler Distance, Akaike Information Criterion. As the resulting statements make extensive use of probabilities, we will begin the presentation by recalling some elements of probability theory.

2.2 Random Variable

In probability and statistics, a random variable, random quantity, aleatory variable, or stochastic variable is described informally as a variable whose values depend on outcomes of a random phenomenon. The formal mathematical treatment of random variables is a topic in probability theory. In that context, a random variable is understood as a measurable function defined on a probability space that maps from the sample space to the real numbers.

Random variable is a function from all possible outcomes to real values, used for defining probability mass functions. A random variable's possible values might represent the possible outcomes of a yet-to-be-performed experiment, or the possible outcomes of a past experiment whose already-existing value is uncertain (for example, because of imprecise measurements or quantum uncertainty). They may also conceptually represent either the results of an "objectively" random process (such as rolling a die) or the "subjective" randomness that results from incomplete knowledge of a quantity. The meaning of the probabilities assigned to the potential values of a random variable is not part of probability theory itself, but is instead related to philosophical arguments over the interpretation of probability. There are two types of random variables, discrete and continuous (Figure 2.1).

In this part, We collect some elementary concepts and properties in connection with random variables [40][41][42].

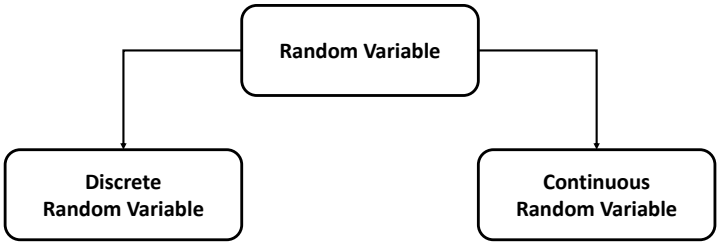

Figure 2.1 – Types of Random Variable

2.2.1 Probability Space

We denote the possible set of outcomes of a random experiment by Ω. Subsets A, $A \subseteq \Omega$, are called events. These events are assigned probabilities. The probability is a mapping $A \mapsto P(A) \in [0, 1], A \subseteq \Omega$, which fulfills the axioms of probability,

- $P(A) \geq 0$,
- P(Ω)=1,
- P($\bigcup A_i$)=$\sum_i P(A_i)$ for $A_i \cap A_j = \varnothing$ with $i \neq j$

where $\{A_i\}$ may be a possibly infinite sequence of pairwise disjoint events. For a well-defined mapping, we do not consider every possible event but in particular only those being contained in σ-algebras. A **σ-algebra** $\mathbb{F}$ of Ω is defined as a system of subsets containing

- The empty set $\varnothing$,
- The complement A^c of every subset A $\in \mathbb{F}$,
- The union $\bigcup A_i$ of a possibly infinite sequence of elements $A_i \in \mathbb{F}$.

Of course, a σ-algebra is not unique but can be constructed according to problems of interest. The interrelated triple of set of outcomes, σ-algebra and probability measure, (Ω, $\mathbb{F}$, P), is also called a **probability space**.

2.2.2 Discrete Random Variables

A real-valued one-dimensional **random variable** X maps the set of outcomes Ω of the space (Ω, $\mathbb{F}$, P) to the real numbers

$$X : \Omega \mapsto \mathbb{R}$$
$$\omega \mapsto X(\omega)$$

Again, however, not all such possible mappings can be considered. In particular, a random variable is required to have the property of **measurability**. This implies the following: A subset $B \subseteq \mathbb{R}$ defines an event of Ω in such a way that:

$$X^{-1}(B) := \{\omega \in \Omega | X(\omega) \in B\}$$

This so-called inverse image $X^{-1}(B) \subseteq \Omega$ of B contains exactly the very elements of Ω which are mapped by X to B. Let $\mathbb{B}$ be a family of sets consisting of subsets of $\mathbb{R}$. Then as measurability it is required from a random variable X that for all $B \in \mathbb{B}$ all inverse images are contained in the σ-algebra $\mathbb{F} : X^{-1}(\mathbb{B}) \in \mathbb{F}$. Thereby the probability measure P on $\mathbb{F}$ is conveyed to $\mathbb{B}$, i.e. the probability function P_x assigning values to X is induced as follows:

$$P_x(B) = P(X^{-1}(B)), B \in \mathbb{B}$$

Thus, strictly speaking, X does not map from Ω to $\mathbb{R}$ but from one probability space to another:

$$X : (\Omega, \mathbb{F}, P) \mapsto (\mathbb{R}, \mathbb{B}, P_x)$$

where $\mathbb{B}$ now denotes a σ-algebra. In particular, for $x \in \mathbb{R}$ the event $X \leq x$ has an induced probability leading to **the distribution function** of X defined as follows:

$$F_x(x) := P(X \leq x) = P(X \in]-\infty, x]) = P(X^{-1}(]-\infty, x])), x \in \mathbb{R}$$

2.2.3 Continuous Random Variables

For most of all problems in practice we do not explicitly construct a random experiment with probability P in order to derive probabilities P_x of a random variable X. Typically we start directly with the quantity of interest X modeling a probability distribution without inducing it. In particular, this is the case for so called continuous variables. For a continuous random variable every value taken from a real interval is a possible realization. As a continuous random variable can therefore take uncountably many values it is not possible to calculate a probability $P(x_1 < X \leq x_2)$ by summing up the individual probabilities. Instead, probabilities are calculated by integrating a probability density. We assume the function $f(x)$ to be continuous (or at least Riemann-integrable) and to be nonnegative for all $x \in \mathbb{R}$. Then f is called (probability) density (or **density function**) of X if it holds for arbitrary numbers $x_1 < x_2$ that

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$$

The area beneath the density function therefore measures the probability with which the continuous random variable takes on values of the interval considered. In general, a density is defined by two properties:

1. f(x)$\geq$ 0,
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$

Thus, the distribution function $F(x) = P(X \leq x)$ of a continuous random variable X is calculated as follows:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

If there is the danger of a confusion, we sometimes subscript the distribution function, e.g. $F_x(0) = P(x \leq 0)$.

2.2.4 Expected value and variance

As is well known, the **expected value** E(X) (also called **expectation**) of a continuous random variable X with continuous density f is defined as follows:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

For (measurable) mappings g, transformations g(X) are again random variables, and the expected value is given by:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

In particular, for each power of X so-called **moments** are defined for $k = 1, 2, \ldots$

$$\mu_k = E[X^k]$$

Note that this term represents integrals which are not necessarily finite (then one says: the respective moments do not exist). There are even random variables whose density f allows for very large observations in absolute value with such a high probability that even the expected value μ_1 is not finite. If nothing else is suggested, we will always assume random variables with finite moments without pointing out explicitly.

Often we consider so-called centered moments where $g(X)$ is chosen as $(X - E(X))^k$. For $k = 2$ the **variance** is obtained (often denoted by σ^2):

$$\sigma^2 = Var(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)dx$$

Elementarily, the following additive decomposition is shown:

$$Var(X) = E(X^2) - (E(X))^2 = \mu_2 - \mu_1^2$$

Since, $Var(X) \geq 0$ by construction, this gives rise to the following inequality:

$$(E(X))^2 \leq E(X^2)$$

In addition to centering, for higher moments a standardization is typically considered. The following measures of **skewness** and **kurtosis** with $k = 3$ and $k = 4$, respectively, are widely used:

$$\gamma_1 = \frac{E[(X - \mu_1)^3]}{\sigma^3}$$

$$\gamma_2 = \frac{E[(X - \mu_1)^4]}{\sigma^4}$$

The skewness coefficient is used to measure deviations from symmetry. If X exhibits a density f which is symmetric around the expected value, it obviously follows that $\gamma_1 = 0$. The interpretation of the kurtosis coefficient is more difficult. Generally, γ_2 is taken as a measure for a distribution's "peakedness", or alternatively, for how probable extreme observations ("outliers") are. Frequently, the normal distribution is taken as a reference. For every normal distribution (also called Gaussian distribution) it holds that the kurtosis takes the value 3. Furthermore, it can be shown that it holds always true that,

$$\gamma_2 \geq 1$$

2.3 Joint and Conditional Distributions

2.3.1 Joint Distribution and Independence

In order to restrict the notational burden, we only consider the three-dimensional case of continuous random variables X, Y and Z with the joint density function $f_{x,y,z}$ mapping from $\mathbb{R}^3$ to $\mathbb{R}$. For arbitrary real numbers a, b and c, probabilities are defined as multiple (or iterated) integrals:

$$P(X \leq a, Y \leq b, Z \leq c) = \int_{-\infty}^{c} \int_{-\infty}^{b} \int_{-\infty}^{a} f_{x,y,z}(x, y, z) dx dy dz$$

As long as f is a continuous function, the order of integration does not matter. The variables are called **stochastically independent** if, for arbitrary arguments, the joint distribution is given as the product of the marginal densities:

$$f_{x,y,z}(x, y, z) = f_x(x) f_y(y) f_z(z)$$

which implies pairwise independence:

$$f_{x,y}(x, y) = f_x(x) f_y(y)$$

The joint probability

$$P(X \leq a, Y \leq b, Z \leq c) = \int_{-\infty}^{c} \int_{-\infty}^{b} \int_{-\infty}^{a} f_x(x) f_y(y) f_z(z) dx dy dz$$

is, under independence, factorized to

$$P(X \leq a, Y \leq b, Z \leq c) = P(X \leq a) P(Y \leq b) P(Z \leq c)$$

2.3.2 Conditional Distributions

Conditional distributions and densities, respectively, are defined as the ratio of the joint density and the "conditioning density", i.e. they are defined by the following density functions (where positive denominators are assumed):

$$f_{x|y}(x) = \frac{f_{x,y}(x, y)}{f_y(y)}$$

$$f_{x|y,z}(x) = \frac{f_{x,y,z}(x, y, z)}{f_{y,z}(y, z)}$$

$$f_{x,y|z}(x, y) = \frac{f_{x,y,z}(x, y, z)}{f_z(z)}$$

It should be clear that these conditional densities are in fact density functions. In case of independence it holds by definition that the conditional and the unconditional densities are equal, e.g.

$$f_{x|y}(x) = f_x(x)$$

This is very intuitive: In case of two independent random variables, one does not have any influence on the probability with which the other takes on values.

2.3.3 Conditional Expectation

If the random variables X and Y are not independent and if the realization of Y is known, $Y = y$, then the expectation of X will be affected:

$$E(X|Y = y) = \int_{-\infty}^{+\infty} x f_{x|y}(x) dx$$

AnaLogously, we define the conditional expectation of a random variable Z, $Z = h(X, Y)$, $h : \mathbb{R}^2 \to \mathbb{R}$, given $Y = y$ as:

$$E(Z|Y = y) = E(h(X, Y)|Y = y)$$

In particular, for $h(X, Y) = Xg(Y)$, with $g : \mathbb{R} \to \mathbb{R}$ one therefore obtains,

$$E(Xg(Y)|Y = y) = g(y)E(X|Y = y)$$

Here, the marginal density of X is replaced by the conditional density conditioned on the value $Y = y$.

Technically, we can calculate the density conditioned on the random variable Y instead of conditioned on a value $Y = y$,

$$f_{x|Y}(x) = \frac{f_{x,y}(x, Y)}{f_y(Y)}$$

By $f_{x|Y}(x)$ a transformation of the random variable Y and consequently a new random variable is obtained.

With the notation introduced above, it holds that:

$$E_y[E_x(X|Y)] = E_x[X]$$

$$E_h[g(Y)X|Y] = g(Y)E_x[X|Y] \; for \; h(X, Y) = Xg(Y)$$

2.4 Stochastic processes

For the last five decades, Stochastic processes have been increasingly realized as an important branch of study not only in statistics, but also in other disciplines such as mathematics, engineering and computer sciences. The stochastic processes is preferred mainly because of the fact that it is probabilistic rather than deterministic in any real life situation. Moreover, stochastic processes deal with a group of variables, which depend on time, thus it could give better representation than a single random variable.

We talk about a stochastic process when the evolution of a variable over time is unpredictable. That is, it is impossible, to know the position of the variable at an instant, to accurately predict its position at the next instant. The unpredictability of stochastic processes does not, however, mean that they are completely random (i.e. that the value over a given interval is strictly independent of past values). The analysis of stochastic processes aims to show that apparently chaotic series include order, an order less obvious than in a series determined by a simple equation, but nevertheless mathematically formalizable.

A univariate **stochastic process (SP)** [40] is a family of (real-valued) random variables, $\{X(t,\omega)\}_{t\in\mathbb{T}}$, for a given index set $\mathbb{T}$:

$$\begin{aligned} X : \mathbb{T} * \Omega &\mapsto \mathbb{R} \\ (t,\omega) &\mapsto X(t,\omega) \end{aligned}$$

The subscript $t \in \mathbb{T}$ is always to be interpreted as "time". At a fixed point in time t_0 the stochastic process is therefore simply a random variable,

$$\begin{aligned} X : \Omega &\mapsto \mathbb{R} \\ \omega &\mapsto X(t_0,\omega) \end{aligned}$$

A fixed ω_0, however, results in a path, a trajectory or a realization of a process which is also often referred to as **time series**,

$$\begin{aligned} X : \mathbb{T} &\mapsto \mathbb{R} \\ t &\mapsto X(t,\omega_0) \end{aligned}$$

In fact, a stochastic process is a rather complex object. In order to characterize it mathematically, random vectors of arbitrary, finite length n at arbitrary points in time $t_1 < \ldots < t_n$ have to be considered:

$$X_n(t_i) := (X(t_1,\omega),\ldots,X(t_n,\omega))',\ \ t_1 < .. < t_n$$

The multivariate distribution of such an arbitrary random vector characterizes a stochastic process. In particular, certain minimal requirements for the finite dimensional distribution of $X_n(t_i)$ guarantee that a stochastic process exists at all.

Depending on the countability or non-countability of the index set $\mathbb{T}$, discrete time and continuous-time SPs are distinguished. In the case of sequences of random variables, we talk about **discrete-time processes**, where the index set consists of integers , $\mathbb{T} \subseteq \mathbb{N}$ or $\mathbb{T} \subseteq \mathbb{Z}$. For discrete-time processes we agree upon lower case letters as an abbreviation without explicitly denoting the dependence on ω,

$$\{x_t\}_{t\in\mathbb{T}} \ for\ \{X(t,\omega)\}_{t\in\mathbb{T}}$$

For so-called **continuous-time processes** the index set $\mathbb{T}$ is a real interval, $\mathbb{T} = [a,b] \subseteq \mathbb{R}$, frequently $\mathbb{T} = [0,T]$ or $\mathbb{T} = [0,1]$, however, open intervals are also admitted. For continuous-time processes we also suppress the dependence on ω notationally and write in a shorter way

$$X(t), t \in \mathbb{T}\ for\ \{X(t,\omega)\}_{t\in\mathbb{T}}$$

Most important types of stochastic processes are: Brownian Motion, Markov Chain, Gaussian processes and Poisson processes.

2.4.1 Gaussian Process

In probability theory and statistics, a Gaussian process is a stochastic process (a collection of random variables indexed by time or space), such that every finite collection of those random variables has a multivariate normal distribution, i.e. every finite linear combination of them is normally distributed. The distribution of a Gaussian process is the joint distribution of all those (infinitely many) random variables, and as such, it is a distribution over functions with a continuous domain, e.g. time or space.

The concept of Gaussian processes is named after Carl Friedrich Gauss because it is based on the notion of the Gaussian distribution (normal distribution). Gaussian processes can be seen as an infinite-dimensional generalization of multivariate normal distributions.

Definition

A time continuous stochastic process $\{X_t, t \in T\}$ is Gaussian if and only if for every finite set of indices $\{t_1, t_2, .., t_k\}$ in the index set T, $X_{t_1,..,t_k} = (X_{t_1}, .., X_{t_k})$ is a multivariate Gaussian random variable. That is the same as saying every linear combination of $(X_{t_1}, .., X_{t_k})$ has a univariate normal (or Gaussian) distribution.

2.4.2 Poisson Process

The Poisson process is one of the most widely-used counting processes. It is usually used in scenarios where we are counting the occurrences of certain events that appear to happen at a certain rate, but completely at random (without a certain structure).

Poisson random variable

Here, we briefly review some properties of the Poisson random variable. A discrete random variable X is said to be a Poisson random variable with parameter μ, shown as $X \sim Poisson(\mu)$, if its range is $R_X = \{0, 1, 2, ..\}$ and its distribution is given by

$$P_X(k) = \begin{cases} \frac{e^{-\mu}\mu^k}{k!}, & \text{for k} \in R_X \\ 0, & \text{Otherwise} \end{cases}$$

Here are some useful facts:

1. If $X \sim Poisson(\mu)$, then $E(X) = \mu$, and $Var(X) = \mu$.
2. If $X_i \sim Poisson(\mu_i)$, for $i = 1, 2, .., n$, and the $X_i's$ are independent, then

$$X_1 + X_2 + .. + X_n \sim Poisson(\mu_1 + \mu_2 + .. + \mu_n)$$

Definition

Let $\lambda > 0$ be fixed. The counting process $\{N(t), t \in [0, \infty)\}$ is called a Poisson process with rates λ if all the following conditions hold:

1. N(0)=0.
2. N(t) has independent increments.

3. the number of arrivals in any interval of length $\tau > 0$ has Poisson($\lambda\tau$) distribution.

Note that from the above definition, we conclude that in a Poisson process, the distribution of the number of arrivals in any interval depends only on the length of the interval, and not on the exact location of the interval on the real line. Therefore the Poisson process has stationary increments.

2.5 Maximum Likelihood

The most common method for estimating parameters in a parametric model is the maximum likelihood method [43]. Let $X_1, \ldots, X_n$ be iid with density function $f(x, \theta)$.

Definition
The **likelihood function** is defined by

$$\mathbb{L}_n(\theta) = \Pi_{i=1}^{n} f(X_i, \theta)$$

The **Log-likelihood function** is defined by $l_n(\theta) = Log\mathbb{L}_n(\theta)$

The likelihood function is just the joint density of the data, except that we treat it as a function of the parameter θ. Thus, $\mathbb{L}_n : \Theta \mapsto [0, +\infty[$. The likelihood function is not a density function: in general, it is not true that $\mathbb{L}_n(\theta)$ integrates to 1 (with respect to θ).

Definition
The **maximum likelihood estimator** MLE, denoted by $\widehat{\theta}_n$, is the value of θ that maximizes $\mathbb{L}_n(\theta)$.

The maximum of $l_n(\theta)$ occurs at the same place as the maximum of $L_n(\theta)$, so maximizing the Log-likelihood leads to the same answer as maximizing the likelihood. Often, it is easier to work with the Log-likelihood.

Remark
If we multiply $\mathbb{L}_n(\theta)$ by any positive constant c (not depending on θ) then this will not change the MLE. Hence, we shall often drop constants in the likelihood function.

Properties of Maximum Likelihood Estimators

Under certain conditions on the model, the maximum likelihood estimator $\widehat{\theta}_n$ possesses many properties that make it an appealing choice of estimator. The main properties of the MLE are:

1. The MLE is **consistent**: $\widehat{\theta}_n \mapsto \theta_*$ where θ_* denotes the true value of the parameter θ.
2. The MLE is **equivariant**: if $\widehat{\theta}_n$ is the MLE of θ then for any function g, $g(\widehat{\theta}_n)$ is the MLE of $g(\theta)$.
3. The MLE is **asymptotically Normal**: $(\widehat{\theta}_n) - \theta_*)/\widehat{se} \mapsto N(0, 1)$ also, the estimated standard error $\widehat{se}$ can often be computed analytically.

4. The MLE is **asymptotically optimal** or **efficient**: roughly, this means that among all well-behaved estimators, the MLE has the smallest variance, at least for large samples.

5. The MLE is approximately the Bayes estimator.

2.6 Information Theory

Information theory is a probabilistic theory to study the quantification, storage and communication of information content of a set of messages, where the computer coding satisfies a precise statistical distribution [44][45].

This theory is originally proposed by Claude Shannon in 1948 in a landmark paper titled "A Mathematical Theory of Communication" [46]. It was restricted to analyze the means to be implemented in telecommunication techniques to transmit information as quickly as possible and with maximum security. After that, promote methods that would diminish the probability of Error in the recognition of the message. A central concept to develop these methods is the measurement of information, in the mathematical sense of term.

In information theory there is some important quantities of information like, entropy, Mutual information and Kullback Leibler Distance.

2.6.1 Uncertainty and information

A. Qualitative description of the information

Before finding a way to quantitatively measure information, we will try to clarify the concept of information. As we have seen, the most appropriate way to describe a communication system is to give it a probabilistic model. Qualitatively, providing information consists of removing some of the uncertainty about the outcome of a random experiment.

In general, consider a pair of random variable (X, Y) and the two events $X = x$ and $Y = y$. The conditional probability $p(x|y)$ can be interpreted as the modification made to the probability $p(x)$ of the event x when we receive the information that the event y occurs. The information "y is realized" changes the probability of x, i.e. the uncertainty on the realization of x, from $p(x)$ to $p(x|y)$. More precisely,

- If $p(x|y) \leq p(x)$, the uncertainty on x increases,
- If $p(x|y) \geq p(x)$, the uncertainty on x decreases.

A decrease in uncertainty on x should be understood as a gain of information on x and conversely, an increase in uncertainty on x should be understood as a loss of information on x. This quantity is called mutual information of x and y, and can be positive or negative. We will see later that its average is always positive.

The maximum information gain on x will be obtained when $p(x|y) = 1$, that is to say essentially when $x = y$. This quantity, a function of the probability, will be called the proper information of x noted $I(x) = f(p(x))$. To quantify the information, we will need to specify and define this function $f()$.

B. Quantitative description of the information

The proper information of x must be a function of its probability, $I(x) = f(p(x))$. To define f () we admit,

- The proper information of x is a decreasing function of $p(x)$: indeed a certain event brings no information, whereas an improbable event will bring a lot.
- The proper information is an additive quantity: if the events x and y are statistically independent then the total information that they can provide is the sum of the proper information, $f(p(x, y)) = f(p(x)p(y)) = f(p(x)) + f(p(y))$

We must therefore choose a function of the form $I(x) = \lambda Log(p(x)), with\ \lambda < 0$ to ensure the decrease with respect to $p(x)$. The choice of λ will depend on the information unit we choose. In this chapter we use the bit.

C. Definition: Proper information

The specific information of the event $x \in \mathcal{X}$ is defined by

$$I(x) = -Log(p(x))$$

The specific information is interpreted as the "Quantity of information provided by the realization of an event".

Note that the specific information is always positive or zero, and that the more improbable an event, the greater its specific information. Conversely, when $p(x) = 1$, we have $I(x) = 0$, that is to say that the realization of a certain event does not provide any information, which seems consistent with the intuition.

One can also define a pair of random variable (X, Y), the conditional proper information which is equal to the quantity of information supplied by an event x knowing that the event y has occurred.

Definition

The conditional proper information of the event $x \in \mathcal{X}$, knowing $y \in \mathcal{Y}$ is defined by

$$I(x|y) = -Log(p(x|y)$$

This last definition allows us to give a new interpretation of the mutual information between two events. Indeed according to the relation

$$I(x, y) = I(x) - I(x|y)$$

The mutual information between x and y is equal to the quantity of information provided by x minus the quantity of information that x would provide if y were to occur.

2.6.2 Mutual Information

The mutual information is a measure defined over two or more random variables, which measures the amount of information that can be obtained about one random variable by observing another. It is a crucial quantity in communication where it can be used to maximize the amount of information shared between sent and received signals.

The mathematical definition of mutual information of X relative to Y is given by,

$$I(X, Y) = \int_{\mathbb{R}} \int_{\mathbb{R}} p(x, y) Log \left(\frac{p(x, y)}{p(x)p(y)} \right) dxdy$$

In other words, the mutual information is the difference between the "mess" inherent in X, and the "mess" left in X after knowing Y. We consider a pair of random variable (X, Y). To give a quantitative measure of what brings us the realization of an event $y \in \mathcal{Y}$ on the possibility of realization of another event $x \in \mathcal{X}$, the occurrence of the event y transforms the a priori probability $p(x)$ of the event x into the posterior probability $p(x|y)$.

Definition

Mutual information between events $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ is defined by

$$I(x, y) = Log(\frac{p(x|y)}{p(x)})$$

Note that this definition is symmetric, indeed, we have by definition of the conditional probability, $p(x, y) = p(x|y)p(y) = p(y|x)p(x)$, so

$$I(x, y) = I(y, x) = Log(\frac{p(x, y)}{p(x)p(y)})$$

A. Conditional Mutual information

Conditional mutual information satisfies the same set of relations given for mutual information except that all the terms are now conditioned on a random variable Z. We state these relations in the next two propositions.

Proposition 1

The mutual information between a random variable X and itself conditioning on a random variable Z is equal to the conditional entropy of X given Z, i.e., $I(X, X|Z) = H(X|Z)$.

Proposition 2

$$I(X, Y|Z) = H(X|Z) - H(X|Y, Z)$$
$$I(X, Y|Z) = H(Y|Z) - H(Y|X, Z)$$

and

$$I(X, Y|Z) = H(X|Z) + H(Y|Z) - H(X, Y|Z)$$

provided that all the conditional entropies are finite

2.6.3 Entropy

A fundamental measure in information theory is "Entropy" which is a measure of the uncertainty of a random variable involved in the value of a random variable, based on the probability density function of each source symbol to be communicated [47]. The corresponding formula for a continuous random variable with probability density function $f(x)$ with finite or infinite support X

The entropy $H(X)$ of a random variable X is defined by

$$H(X) = -\int_X f(x)Log(f(x))dx$$

The Log is to the base 2 and entropy is expressed in bits. If the base of the Logarithm is b, we denote the entropy as $H_b(X)$. If the base of the Logarithm is e, the entropy is measured in nats. Unless otherwise specified, we will take all Logarithms to base 2, and hence all the entropies will be measured in bits. Note that entropy is a functional of the distribution of X. It does not depend on the actual values taken by the random variable X, but only on the probabilities.

Prepositions

We have the following inequalities:

- $H \geq 0$
- $H \leq Log(card(X))$ where X is a finite set of random variables.

We now extend the definition to a pair of random variables. There is nothing really new in this definition because (X, Y) can be considered to be a single vector-valued random variable.

Definition

The joint entropy $H(X, Y)$ of a pair of continuous random variables (X, Y) with a joint distribution $f(x, y)$ is defined as

$$H(X, Y) = -\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} f(x, y)Log(f(x, y))dxdy$$

which can also be expressed as

$$H(X, Y) = -E[Log(f(X, Y))]$$

We also define the conditional entropy of a random variable given another as the expected value of the entropies of the conditional distributions, averaged over the conditioning random variable.

Definition (Conditional Entropy)

If $(X, Y) \sim f(x, y)$, the conditional entropy $H(Y|X)$ is defined as

$$\begin{aligned} H(Y|X) &= \int_{x\in\mathcal{X}} f(x)H(Y|X=x)dx \\ &= -\int_{x\in\mathcal{X}} f(x)\int_{y\in\mathcal{Y}} f(y|x)Log(f(y|x))dxdy \\ &= -\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} f(x,y)Log(f(y|x))dxdy \\ &= -E[Log(f(Y|X))] \end{aligned} \tag{2.1}$$

The naturalness of the definition of joint entropy and conditional entropy is exhibited by the fact that the entropy of a pair of random variables is the entropy of one plus the conditional entropy of the other. This is proved in the following theorem.

Theorem (Chain Rule)

$$H(X, Y) = H(X) + H(Y|X)$$

Proof

$$\begin{aligned} H(X,Y) &= -\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} f(x,y)Log(f(x,y))dxdy \\ &= -\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} f(x,y)Logf(x)f(y|x)dxdy \\ &= -\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} f(x,y)Logf(x)dxdy - \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} f(x,y)Logf(y|x)dxdy \\ &= -\int_{x\in\mathcal{X}} f(x)Logf(x)dx - \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} f(x,y)Logf(y|x)dxdy \\ &= H(X) + H(Y|X) \end{aligned} \tag{2.2}$$

Equivalently, we can write

$$Logf(X, Y) = Logf(X) + Logf(Y|X)$$

and take the expectation of both sides of the equation to obtain the theorem.

It can be seen that in fact it gives us insight about how far are X and Y from being independent from each other. Through some simple algebra, we can prove that (Figure 2.2)

$$I(X, Y) = H(X) - H(X|Y)$$

2.6.4 Channel Capacity

A communicates with B means that the physical acts of A have induced a desired physical state in B. This transfer of information is a physical process and therefore is subject to the uncontrollable ambient noise and imperfections of the physical signaling process itself. The communication is successful if the receiver B and the transmitter A agree on what was sent.

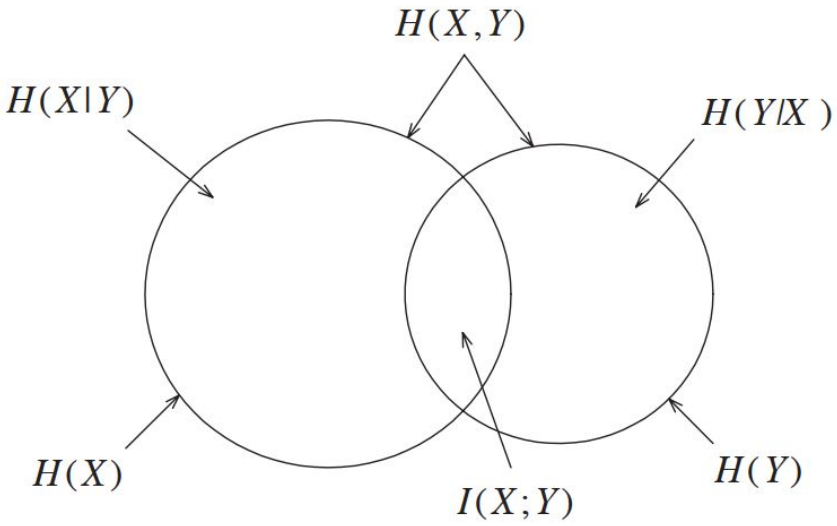

Figure 2.2 – Information Diagram

In this section we find the maximum number of distinguishable signals for n uses of a communication channel. This number grows exponentially with n, and the exponent is known as the channel capacity. The characterization of the channel capacity (the Logarithm of the number of distinguishable signals) as the maximum mutual information is the central and most famous success of information theory.

Source symbols from some finite alphabet are mapped into some sequence of channel symbols, which then produces the output sequence of the channel. The output sequence is random but has a distribution that depends on the input sequence. From the output sequence, we attempt to recover the transmitted message.

Each of the possible input sequences induces a probability distribution on the output sequences. Since two different input sequences may give rise to the same output sequence, the inputs are confusable. We can choose a "non confusable" subset of input sequences so that with high probability there is only one highly likely input that could have caused the particular output. We can then reconstruct the input sequences at the output with a negligible probability of error. By mapping the source into the appropriate "widely spaced" input sequences to the channel, we can transmit a message with very low probability of error and reconstruct the source message at the output. The maximum rate at which this can be done is called **the capacity** of the channel[48].

Definition

We define a discrete channel to be a system consisting of an input alphabet $\mathbb{X}$ and output alphabet $\mathbb{Y}$ and a probability transition matrix $p(y|x)$ that expresses the probability of observing the output symbol y given that we send the symbol x. The channel is said to be memoryless if the probability distribution of the output depends only on the input at that time and is conditionally independent of previous channel inputs or outputs.

Definition

We define the "information" channel capacity of a discrete memoryless channel as

$$C = \max_{p(X)} I(X, Y)$$

where the maximum is taken over all possible input distributions $p(x)$. The units of C are bits per input symbol into the channel.

An operational definition of channel capacity as the highest rate in bits per channel use at which information can be sent with arbitrarily low probability of error. Shannon's second theorem establishes that the information channel capacity is equal to the operational channel capacity. Thus, we drop the word information in most discussions of channel capacity.

Properties of Channel Capacity

1. $C \geq 0$ since $I(X, Y) \geq 0$.
2. $C \leq Log(X)$ since $C = maxI(X, Y) \leq maxH(X) = Log(X)$.
3. $C \leq Log(Y)$ for the same reason.
4. $I(X, Y)$ is a continuous function of $p(x)$.
5. $I(X, Y)$ is a concave function of $p(x)$.

In the next section we will introduce a new concept: Kullback Leibler Distance which is the relative entropy.

2.7 Kullback Leibler Distance

In any statistical problem we are given a set of observations. These observations are the values of some random variables whose probability distribution is usually unknown to us, or we have some knowledge of it. From the information provided by the data, we draw inferences about the unknown aspects of the underlying distribution, such as the unknown "true" parameter values of the distribution which govern the generation of the observed data and also govern the generation of any future observations if we adopt the predictive point of view.

We shall express a model in the form of a probability distribution and regard fitting a model to the data as estimating the true probability distribution from the data and treat the estimation and the evaluation of a model together as one entity rather than separating them. In the statistical literature during the past fifty years, there has been a meaningless separation of estimation and testing which did not help the development of a practical and successful statistical model selection and evaluation procedure.

If we had an objective measure (a metric) of the distance between the model and the true distribution, a good inference procedure ought to make this distance as small as possible. One of the most common measures to do that is Kullback Leibler Distance (KLD)[49][50].

KLD also known as Relative Entropy in information theory, is an important statistical measure used to quantify the dissimilarity between two distribution functions. The KLD is an extension of Shannon's concept of information.

Because of its simplicity in both theory and applications, KLD is used in different problems like image processing, compression, classification...

Definition
Let us consider f and g to be two distribution functions. The KLD of $f(x)$ relative to $g(x)$, which is a measure of the information lost when $g(x)$ is used to approximate $f(x)$, is defined as,

$$D_{KL}(f(x)||g(x)) = \int_{-\infty}^{+\infty} f(x) Log\left(\frac{f(x)}{g(x)}\right) dx$$

Typically, f represents data, observations, or a precisely calculated distribution function. And, g represents a theory, model, description or approximation of f.

Properties

1. $D_{KL}(f||g) \geq 0$.
2. $f = g$ ssi $D_{KL}(f||g) = 0$ almost everywhere.
3. $D_{KL}(f||g) \neq D_{KL}(g||f)$.

Distance property
Although the Kullback-Leibler information is used to measure a difference between two density functions, it is not a distance in the topoLogical sense. Indeed, the triangular inequality is not verified as well as the property of symmetry. This also explains the importance of the choice between the functions f and g. By convention, f is chosen as the density function associated with the experimental measurements and g as the approximation density function. However, there are modifications for symmetrical KL information, for example by using,

$$d(p, q) = I(p, q) + I(q, p) \tag{2.3}$$

this is called Kullback-Leibler distance.

Formulation in the form of expectation
By developing the sum, we get

$$I(p, q) = E_p[Log(p(x))] - E_p[Log(q(x))]$$

Why KLD

KLD is preferred over Mean Square Error (MSE), Root Mean Square Error (RMSE) and L1/L2 Regularization. KLD is a measure on probability distribution, it essentially captures the information loss between ground truth distribution and predicted.

L1/L2 Regularization, MSE and RMSE doesn't do well with probabilities, because of the power operations involved in the calculation of loss. Probabilities, being fractions under 1, are significantly effected by any power operations (square or root), and considering we are calculating the squares of differences of probabilities, the values that are summed are abnormally small, essentially barely learning anything as the random initialization itself starts with an abnormally small loss, almost always staying constant.

And in comparison with Euclidean Distance, KLD has a probabilistic/ statistical meaning while Euclidean distance has not.

On the other hand, KLD has better properties with respect to real data, and it is easier to implement in software. The KLD between models is a fundamental quantity in science and information theory and is the Logical basis for model selection in conjunction with likelihood inference.

Mean Log Likelihood as an Estimate of KLD

In this section we introduce the concept of mean Log likelihood as a measure for the goodness of fit of a model and state entropy maximization principle (EMP) according to Akaike.

Suppose that the generation of data is described by a model given by a probability density function $f(x|\theta)$. Given n independent observations from the same distribution regarded as a function of a vector-valued parameter, $\theta = (\theta_1, \theta_2, .., \theta_K)$, $k = 1, 2, .., K$, the likelihood function for the set of data is

$$L(\theta) = f(x_1, \ldots, x_n|\theta) = \Pi_{i=1}^{n} f(x_i|\theta)$$

The Log likelihood function, $l(\theta)$ (often called the support), is the natural Logarithm of $L(\theta)$ and is defined by

$$l(\theta) \equiv LogL(\theta) = \sum_{i=1}^{n} Logf(x_i|\theta)$$

regarded as a random variable, is the sum of i.i.d, random variables $Logf(x_i|\theta)$, $i = 1, 2, ..n$.

We define the average or mean Log likelihood of the sample by

$$\frac{1}{n} l(\theta) \equiv \frac{1}{n} LogL(\theta) = \frac{1}{n} \sum_{i=1}^{n} Logf(x_i|\theta) = l_n(\theta)$$

which can be interpreted as an estimator of the "distance" between the true probability density and the model.

As we discussed in the previous section, the KLD is not observable. However, it can be consistently estimated from the observed data and operationalized.

Certainly, one approach to measure how well the maximum likelihood model $f(x_i|\theta)$ "matches" the data would be to test the hypothesis of the KLD.

Since our estimation of KLD is based on the mean Log likelihood (which is also an estimate of the expected Log likelihood), and since the maximum likelihood estimates are biased, then there is the inevitable risk of error of estimation of the KLD when the maximum likelihood estimators of the parameters of the model is used.

Indeed, in defining AIC, Akaike has exactly this consideration of the bias by penalizing extra parameters when the maximum likelihood estimates are used in estimating the expected Log likelihood by the mean Log likelihood.

Next, we derive AIC in detail as a natural sample estimate of the expected Log likelihood.

2.8 Model Selection Techniques

A key ingredient in data analysis for reliable and reproducible statistical inference or prediction, and thus it is central to scientific studies in such fields as ecoLogy, economics, engineering, finance, political science, bioLogy, epidemioLogy and network science is Model Selection. There has been a long history of model selection techniques that arise from researches in statistics, information theory, and signal processing. **Model selection** is the task of selecting a statistical model from a set of candidate models, given data. Model selection may also refer to the problem of selecting a few representative models from a large set of computational models for the purpose of decision making or optimization under uncertainty.

There have been many overview papers on model selection scattered in the communities of signal processing, statistics, machine learning, epidemioLogy, chemometrics, and ecoLogy and evolution.

There are two main objectives in learning from data. One is for scientific discovery, understanding of the data-generation process, and interpretation of the nature of the data. A scientist, e.g., may use the data to support a physical model or identify genes that clearly promote early onset of a disease. Another objective of learning from data is for prediction, i.e., to quantitatively describe future observations. Here the data scientist does not necessarily care about obtaining an accurate probabilistic description of the data. Of course, one may also be interested in both directions.

In tune with the two different objectives, model selection can also have two directions: model selection for inference and model selection for prediction. The first one is intended to identify the best model for the data, which hopefully provides a reliable characterization of the sources of uncertainty for scientific insight and interpretation. And the second is to choose a model as a vehicle to arrive at a model or method that offers top performance. For the former goal, it is crucially important that the selected model is not too sensitive to the sample size. For the latter, however, the selected model may simply be the lucky winner among a few close competitors, yet the predictive performance can still be (nearly) the best possible. If so, the model selection is perfectly fine for the second goal (prediction), but the use of the selected model for insight and interpretation may be severely unreliable and misleading. Associated with the first goal of model selection for inference or identifying the best candidate is the following concept of selection consistency.

A wide variety of model selection methods have been proposed in the past few decades, motivated by different viewpoints and justified under various circumstances. We review some of the representative approaches in these contexts in this section.

2.8.1 Akaike Information Criterion

Information criteria generally refer to model selection methods that are based on likelihood functions and applicable to parametric model-based problems. Here we introduce some information criteria whose asymptotic performances are well understood.

The process of evaluating candidate models is termed model selection or model evaluation. It is clearly not desirable to always deem the most complex model the best, and it is generally accepted that the best model is the one that provides an adequate account of the data while using a minimum number of parameters.

One of the more popular methods of comparing multiple models, taking both descriptive accuracy and parsimony into account, is the Akaike information criterion (AIC)[51][52]. The objective of AIC model selection is to estimate the information loss when the probability distribution f associated with the true (generating) model is approximated by probability distribution g, associated with the model that is to be evaluated.

Suppose that the data is generated by some anonym process P. We consider two candidate models to represent P: Q_1 and Q_2. If we knew P, then we could found the information lost from using Q_1 to represent P by calculating the KLD, $D_{KL}(P||Q_1)$, similarly, the information lost from using Q_2 to represent P by calculating $D_{KL}(P||Q_2)$. We would then choose the candidate model that minimized the information lost.

In the case where we do not know P, we cannot do that. Akaike showed, however, that we can estimate how much more (or less) information is lost by Q_1 then by Q_2. The estimate through, is only valid asymptotically, if the number of data points is small, then some correction is often necessary.

The process of evaluating candidate models is termed model selection or model evaluation.

Akaike has shown that choosing the model with the lowest expected information loss (i.e., the model that minimizes the expected Kullback– Leibler discrepancy) is asymptotically equivalent to choosing a model K that has the lowest AIC value. Akaike defined "an Information Criterion" (AIC),

$$AIC = -2Log(L(\theta|data)) + 2K$$

where L is the likelihood function, θ is the maximum likelihood estimate of θ, and K the number of estimate parameters.

Thus, we should select the model that yields the smallest value of AIC because this model is estimated to be "closest" to the unknown reality that generated the data, from among the candidate models considered.

Here Log denotes the natural Logarithm. The simple procedure which selects a model with the minimum AIC among a set of models defines the minimum AIC estimate (MAICE). The introduction of AIC helped the recognition of the importance of modeling in statistics and many practically useful statistical procedures have been developed as minimum AIC procedures; see, for example, Akaike [51][52]. In spite of the accumulation of successful results in practical applications the Logical foundation of MAICE has been continuously questioned by theoretically minded statisticians.

This seems a very simple concept, select the fitted approximation model that is estimated, on average, to be closest to the unknown truth P. If all the models in the set are poor, AIC attempts to select the best approximating model of those in the candidate set and ranks the rest.

The AIC rewards descriptive accuracy via the maximum likelihood, and penalizes lack of parsimony according to the number of free parameters (note that models with smaller AIC values are to be preferred).

An extension of AIC is the **Takeuchi's information criterion** [53], derived in a way that allows model misspecification, but it is rarely used in practice due to its computational complexity. In the context of generalized estimating equations for correlated response data, a variant of AIC based on quasilikelihood is derived in [54]. Takeuchi's Information Criterion is useful in cases where the model is not particular close to truth. Attractive in theory, rarely used in practice because we need a very large simple size to obtain good estimates.

Finite-sample corrected AIC (AICc) [55] was proposed as a corrected version of the AIC for small-sample study. It selects the model that minimizes

$$AICc_m = AIC_m + \frac{2(d_m + 1)(d_m + 2)}{n - d_m - 2}$$

Unless the sample size n is small compared with model dimension d_m, there is little difference between AIC_c and AIC. Another modified AIC that replaces the constant two with a different positive number has also been studied in [56].

There is also **Bayesian Information Criterion (BIC)** [57], which is similar to the formula of AIC, but with different penalty for the number of parameters. The simulation study demonstrates, in particular that AIC sometimes selects a much better model than BIC even when the "true model" is in the candidate set. The reason is that, for finite n, BIC can have a substantial risk of selecting a very bad model from the candidate set. This reason can arise even when n is much larger than K^2. With AIC, the risk of selecting a very bad model is minimized.

Comparison Between AIC and BIC

In this subsection, we review some research advances in the understanding of AIC, BIC, and related criteria. The choice of AIC and BIC to focus on here because they represent two cornerstones of model selection principles and theories. We are only concerned with the settings where the sample size is larger than the model dimension.

Despite the widespread use of the AIC, some believe that it is too liberal and tends to select overly complex models. It has been pointed out that the AIC neglects the sampling variability of the estimated parameters. When the likelihood values for these parameters are not highly concentrated around their maximum value, this can lead to overly optimistic assessments. Furthermore, the AIC is not consistent. That is, as the number of observations n grows very large, the probability that the AIC recovers at rue low-dimensional model does not approach unity. A popular alternative model selection criterion is the Bayesian information criterion or BIC [58]. The BIC for model i is defined as

$$BIC_i = -2LogL_i + K_iLog(n)$$

where n is the number of observations that enter into the likelihood calculation. A comparison of BIC and AIC shows that the BIC penalty term is larger than the AIC penalty term when $n \leq e^2$. Although the equations of AIC and BIC look very similar, they originate from quite different frameworks. The BIC assumes that the true generation model is in the set of candidate models, and it measures the degree of belief that a certain model is the true data-generating model. The AIC does not assume that any of the candidate models is necessarily true, but rather calculates for each model the Kullback– Leibler discrepancy, which is a measure of distance between the probability density generated by the model and reality. A formal comparison in terms of performance between AIC and BIC is very difficult, particularly because AIC and BIC address different questions. Most simulations that show BIC to perform better than AIC assume that the true model is in the candidate set and that it is relatively low dimensional. In contrast, most simulations that favor AIC over BIC assume that reality is infinitely dimensional,and hence the true model is not in the candidate set.

Recall that AIC is asymptotically efficient for the nonparametric framework and is also minimax optimal. In contrast, BIC is consistent and asymptotically efficient for the parametric framework. Despite the good properties of AIC and BIC, they have their own drawbacks. AIC is known to be inconsistent in a parametric framework where there are at least two correct candidate models. As a result, AIC is not asymptotically efficient in such a framework. Why do AIC and BIC work in those ways? Theoretical arguments in those aspects are highly nontrivial and have motivated a vast literature since the formulations of AIC and BIC.

To briefly summarize, for asymptotic efficiency, AIC (respectively, BIC) is only suitable in non parametric (respectively, parametric) settings. There has been a debate between AIC and BIC in model selection practice, centering on whether the data-generating process is in a parametric framework or not. The same debate was sometimes raised under other terminoLogy. In a parametric (respectively, non parametric) framework, the true data-generating model is often said to be well specified (respectively, misspecified) or finite (respectively, infinite) dimensional. (To see a reason for such terminoLogy, consider, e.g., the regression analysis using polynomial basis function as covariates. If the true regression function is indeed a polynomial, then it can be parameterized with a finite number of parameters; if it is an exponential function, then it cannot be parameterized with any finite dimensional parameter.) Without prior knowledge on how the observations were generated, determining which method to use becomes very challenging.

2.8.2 Akaike Weight

Because AIC contains various constants and is a function of complex size, we routinely recommend computing the AIC differences: $\Delta_i = AIC_i - AIC_{min}$, where AIC_{min} is the smallest AIC value in the set, so the larger Δ_i is the less valid fitted model.

For the next step, we note that the AIC is an unbiased estimator of minus twice the expected Log likelihood of the model.

From the differences in AIC, we can then obtain an estimate of the relative likelihood L of model i by the simple transform:

$$L(M_i|data) \propto exp{-\frac{1}{2}\Delta_i(AIC)}$$

where $\propto$ stands for "is proportional to".

To better interprete the relative likelihood of models given the data and the set of N models, we normalize then to be a set of positive "Akaike Weight" [59]adding to 1,

$$w_i = \frac{exp(\frac{-1}{2}\Delta_i)}{\sum_{n=1}^{N} exp(\frac{-1}{2}\Delta_n)}$$

A given Weight w_i can be interpreted as the probability that the model i is the best model (in the AIC sense, that it minimizes the Kullback–Leibler discrepancy), given the data and the set of candidate models. Thus, the strength of evidence in favor of one model over the other is obtained by dividing their Akaike weights. Note that the Akaike weights are subject to sampling variability, and that a different sample will most likely generate a different set of weights for the models in the candidate set.

Akaike weights are easy to compute from the raw AIC values and provide a straightforward interpretation as the probabilities of each model's being the best model in an AIC sense (i.e., the model that has the smallest Kullback– Leibler distance, given the data and the set of candidate models). The use of Akaike weights gives the reader greater insight into the relative merits of the competing models. In addition, Akaike weights quantify conclusions based on AIC analyses by specifying the amount of statistical confidence for the model with the lowest AIC value. Given these considerable advantages,we believe that it is in many circumstances very useful to supplement the standard results of AIC model comparison analysis with presentation of Akaike weights.

2.9 Complexity study

When executing an algorithm, the computer performs a series of very simple operations such as comparing small numbers, for example. We then measure the time complexity of an algorithm as the number of these elementary operations. For example, considering the addition of 2 digits elementary, setting the addition of two numbers of n digits will cause us to perform n additions

to 1 digit, the complexity will therefore be n. On the other hand, posing the multiplication of these two same numbers (by the method learned in school) will have a complexity of the order of n2. The point of counting only these operations (and not real time) is to free yourself from the power itself of the computer. A newer computer will definitely be much faster than the computer left in the closet since the early 2000s, but the number of basic operations will remain about the same. In this way, we can speak of the intrinsic complexity of a problem: it is the smallest number of elementary operations necessary for an algorithm in order to solve this problem.

After the implementation of the approach, in the complexity study, we will study the complexity required to derive the algorithm. It will also provide simulation results assessing the performance in terms of execution time for the proposed algorithm in comparison with the reference algorithms described in Related works.

The complexity of the algorithm is measured through the number of complex multiplications that the algorithm has to perform for the calculation of the test statistic. It is difficult to say anything exact about the computational complexity of the proposed algorithm since this depends on the implementation of the sub functions.

Complexity of an algorithm is dominated by the computation of the parameters. The running time depends on the implementation, but can in general be done in 2N time since it only requires 2N multiplications.

2.10 Conclusion

In this chapter, the mathematical tools on which we have relied to deal with the problems addressed in this thesis, are introduced in detail. They allow the modeling and performance evaluation of the approaches that we will present in the following chapters which represent the subject of the contributions of this thesis. In the following chapters, the contributions will be developed, starting with presenting our new approach to manage handovers.

Chapter 3

Features Detection based Blind Handover using Kullback Leibler Distance for 5G HetNets Systems

Contents

In this chapter, we study a new approach to manage handovers using KLD.

3.1 Introduction

Mobile cellular communication has become increasingly one of the most interesting research area over the past few years. The exponentially increasing demand for wireless data services require a massive network densification that is neither economically nor ecoLogically viable with current cellular system architectures.

Fifth Generation (5G) [8] [9] have recently emerged to satisfy the increasing demand for high data bit rates. A crucial requirement for 5G networks is the deployment of Small Cells (SCs) [10] over Macrocells layer which introduces a new type of networks called Heterogeneous Networks (HetNets) [11]. A HetNet is simply the banding together of different sized cells to provide ultra dense coverage in defined geographic areas.

Small Cells (SCs) are low-powered cellular radio access nodes that operate in licensed and unlicensed spectrum that have a range of 10 meters to a few kilometers. They are a crucial component of 5G networks, because they have the ability to significantly increase network capacity, density and coverage, especially indoors. They are a relatively low cost deployment option and, because they are low power devices, are relatively cheap and efficient to run to give a low total cost of ownership.

Like every other technoLogy, SCs have some drawbacks that give rise to some major concern on part of the end users.

In this contribution, we are going to study the problem of the management of handovers. Handover is the practice of retaining a user's active connection when a mobile terminal changes its connection point to the access network (called "point of attachment") [12][13][14].

Because of the low coverage of SCs, it is essential to support seamless handovers to provide continuous connectivity within any wide area network. In addition, due to the high number of SCs, handovers increase, and the selection, quickly and at low energy cost, of the appropriate one in the vicinity of thousands is also a key problem. Hence, we propose a new approach to operate and manage a blind handover which is an existing feature in which the network node may initiate a handover procedure for a terminal without doing conventional measurement configuration and without considering measurement reports.

For 5G Networks, Artificial Intelligence can be broadly applied in the Blind Handover techniques. The usage of Artificial Intelligence techniques in the Handover decision process will reduce the computation complexity which already exists in the conventional approachs.

The main idea is to operate efficient, blind and rapid handover just by analyzing Received Signal Density Function(df) instead of demodulating and analysing Received Signal itself as in classical handover. The goal within our contribution is to exploit KL Distance, Akaike

Information Criterion (AIC) and Akaike weights [51][52] [59] in order to decide blindly the best handover and the best BS for each user [3].

The remainder of the chapter is organized as follows. We begin by introducing then Related work in section 2. In section 3, we present KL Distance and the formulation of our problem. In section 4 we give a brief review of model selection using AIC: the AIC is presented and the Akaike weights are derived. The approach based on model selection is developed in Section 5. The evaluation of the result is in section 6. The last section will be devoted to the conclusion.

3.2 Related work

In this section, we explain some related work to the different handover techniques available and we introduce the literature review on user association and mobility load balancing for 5G small cell networks.

The main objective of mobility management in 5G small cell Networks is to assure the continuity of the service during the handover process. Handover is the most sensitive point in the convergence of any two adjacent cells.

In the classical Handovers, before initiating the handover procedure, the User Equipment (UE) has to perform cell measurement and report, Radio Resource Connection (RRC) reconfiguration random access as soon as it enters a crossed cell. In [60] it was proposed that the users boarding a high-speed vehicle are handed over as a whole group. Therefore, system performance is enhanced in terms of mobility and resource management.

A novel technique to avoid handover problems in LTE has been proposed by authors in [61], which is based on collaboration among macro-femtocells grouped according to nearby base stations. Each group pre-fetches higher layer packets to reduce the latency in handover process.

Another study [62] introduced a new handover algorithm for mobile relay stations to improve the handover success rate. The algorithm is based on the relative speeds of the UEs to the serving eNB.

Many Handover schemes have been proposed to solve the problem associated with the frequent handover of mobile equipment in a high mobility circumstance. One of the optimized handover processes for LTE network is designed based on the coordinated multiple point (ComP) transmission technoLogy and dual-vehicle station coordination mechanisms [63]. The underlying technique of this optimized scheme makes use of the diversity gain in the overlapping area of two adjacent eNBs as the UE moves from one to another.

Because of the high level of diversity gain, the probability of handover failure decreases and the Quality of Service (QoS) improves because of the high level of reliability [64]. Such a technique can be useful for fast-moving UEs because it reduces the probability of service interruption.

In fact, several ways of research have been developed only for the static scenario (no mobility) with a focus on maximizing throughput across both networks while minimizing the number of handovers and eliminating the ping-pong effect.

In [65], authors proposed an algorithm for HO decision using the metric of Received Signal Strength (RSS), however, using RSS in heterogeneous networks does not give good results.

Furthermore, articles [66] [67] combined other metrics with RSS, such as distance between UE and eNB antennas, and the service cost. But the algorithm becomes more complex as well as excessive delay and high power [68].

Similarly, in [69] cost and/or speed of movement of mobile users have been used as the main indicators and RSS algorithm as a secondary metric. This approach brings better results in terms of rates, cost and blocking probability.

Work in [70] used the signal-to-noise ratio (SNR) and traffic type as the metrics for the HO decision. Its goal was to maximize the throughput of the network and minimize the ping-pong effect. Lin, H et al. proposed a QoS-Based Vertical HO. In addition, reference [68] uses the combined effects mentioned above including signal-to-interference-noise Ratio (SINR) to make HO decisions for multi-attribute QoS considerations. Still, all the above mentioned proposed techniques were studied from the core network point of view, however, integrating Wi-Fi in RAN makes it a different issue that needs to be investigated in terms of mobility and resource management.

In [71] presents Software Defined Networks (SDNs) as a solution to be adopted by service providers to offer new services to their customers while ensuring a better quality of service (QoS), and optimizing the use of services. An implementation of the SDN approach in wireless networks improves their service quality and interoperability. Indeed, during a Handover operation several problems can occur and can have a direct impact on end-to-end transmission delay, packet loss during Handover. For this, authors in [71] propose the Software Defined Mobile Networks approach and Software Defined Wireless Networks to improve the performance of cellular networks, by deploying several algorithms at the SDN controller to improve execution time of the vertical and horizontal Handover, also we deployed an access point, which integrates the OpenFlow protocol and communicates with the SDN controller.

An existing feature in which the network node, a node, may initiate a handover procedure for a terminal without doing conventional measurement configuration and without considering measurement reports is Blind Handover. This feature may be beneficial when a fast handover is needed and candidate cell measurements are unavailable, or would impose an unwanted delay. Using the blind handover in such case removes the time and signaling needed to conduct handover measurements, hence giving the desired fast handover.

Blind Handover Techniques [72]: A beacon pilot blind handover technique has been proposed in which the target network, which normally operates at a frequency f, broadcasts a beacon pilot at the same frequency f, as the frequency of the primary network. This beacon pilot consists of

a pilot channel and a synchronisation channel and enables the mobile terminal to evaluate the propagation loss between itself and the target network. One disadvantage of the beacon pilot approach is that it requires deployment of pilot antennas, increasing the cost of the system infrastructure. Another disadvantage arises in the case of an intra-system, inter-frequency handover between primary and target networks which are UMTS FDD networks operating at adjacent frequencies. In this case the pilot transmission can generate interference on the target network, making its capacity decrease.

Another known blind handover consists in a direct blind handover in which a look-up table is held, for example, in the Radio Network Controller (RNC) of the primary network (assuming an UMTS FDD primary network). This look-up table (or planning table') indicates, for each primary cell, which target cell should be used in a handover. If the handover is between systems having co-located cells then this blind handover method works reasonably well. However, in the case where the transfer is an inter-system transfer there is no guarantee that the boundaries of the cells of the two systems will be defined in the same locations. If the primary and target cells are not co-located then the quality of the connection available in the target cell will vary depending upon the geographic location of the mobile terminal within the primary cell. Thus, for mobile terminals at certain locations within the primary cell, the target cell specified in the planning table will not be the best one to use.

3.3 Description and formulation of the problem

The main idea in this contribution is to detect the best BS for each user (Best Handover) by exploiting model selection techniques and especially the Akaike Information Criterion (AIC). In its most basic forms, model selection is one of the fundamental tasks of scientific inquiry. Determining the principle that explains a series of observations is often linked directly to a mathematical model predicting those observations. Model selection is the task of selecting a statistical model from a set of candidate models, given data.
It was shown in [73] that, when signal demodulation cannot be performed, the received wireless communication signal can be, roughly, modeled using Rayleigh and Rician distribution. Therefore, we propose to calculate in blindly process the Received Signal for each BS and Analyze AIC in order to determine the best handover.

Figure 3.1 presents an illustrated model of Small Cells Network.

The notation used in this chapter are given in the Table 3.1 below.

In this section, we will give a short review of the basic ideas. In fact, it is assumed that the samples of the Received Signal for each BS are distributed according to an original density function f_k, called the operating model, where $k \in \{1, 2, 3, 4, 5, 6\}$ is the index of BS. Since only a finite number of observations is available, the operating model is usually unknown. Therefore, approximating model (i.e candidate model) must be specified using the observed data, in order to estimate the operating model. The candidate model is denoted as g_θ^k , where θ indicates the dimensional parameter vector, which specifies the density function.

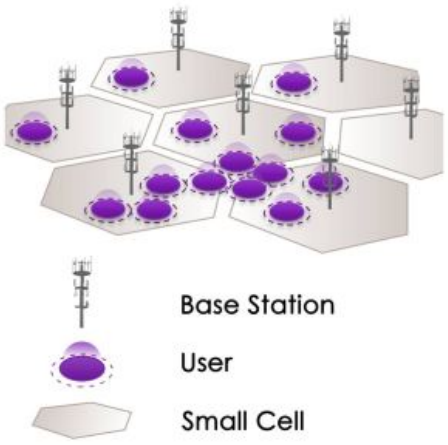

Figure 3.1 – Model of Small Cells Network

Notation	Description
k	index of the Base Station
f_k	original density function of the received signal
g_θ^k	approximating model
θ	the dimensional parameter vector
D_{KL}	Kullback Leibler Distance
h(.)	differential entropy
N	number of observations
$x_1, \ldots, x_N$	independent observations
AIC_k	Akaike Information Criterion of Base Station k
U	the dimension of the parameter vector θ
L_k	Likelihood function
ϕ_k	AIC differences
w_k	Akaike weight
$I_0(.)$	modified Bessel function of the first kind with order zero
μ_k	the mean or expectation of the distribution
σ_k	the standard deviation
λ	threshold
P_{FA}	probability of False Alarm

Table 3.1 – Notation and TerminoLogy

In information theory, the Kullback-Leibler distance describes the discrepancy between the two density functions f_k and g_θ^k and is given by [51],

$$D_{KL}(f_k(x)||g_\theta^k(x)) = \int_X f_k(x) Log\left(\frac{f_k(x)}{g_\theta^k(x)}\right) dx$$

$$D_{KL}(f_k \;\; g_\theta^k) = E(Log(f_k(x))) - E(Log(g_\theta^k(x)))$$

$$D_{KL}(f_k \;\; g_\theta^k) = \int_X f_k(x) Log(f_k(x)) dx - \int_X f_k(x) Log(g_\theta^k(x)) dx$$

$$D_{KL}(f_k \;\; g_\theta^k) = -h_i(x) - \int_X f_k(x) Log(g_\theta^k(x)) dx \tag{3.1}$$

where the random variable X is distributed according to the original but unknown density function f and $h(.)$ denotes differential entropy. Since, the original density function f_k is not known, this distance measure is not directly applicable.

It is known, however, that the Kullback-Leibler distance is nonnegative, this implies that,

$$-\int f_k(x) Log(g_\theta^k(x)) dx = h_i(x) + D(f_k \;\; g_\theta^k) \tag{3.2}$$

approaches the differential entropy of X from above for increasing quality of the model g_θ^k .

The differential entropy of X is reached if and only if $f = g_\theta$.

In probability theory, the law of large numbers (LLN) is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value and will tend to become closer to the expected value as more trials are performed.

Applying the weak law of large numbers, this expression (3.2) can be approximated by averaging the Log-likelihood values given the model over N independent observations $x_1, x_2, \ldots, x_N$ according to:

$$-\int_X f_k(x) Log\left(g_\theta^k(x)\right) dx \approx -\frac{1}{N}\sum_{n=1}^{N} g_\theta^k(x_n) \tag{3.3}$$

The Log-likelihood depends on the estimated vector θ, which itself is a function of the actual observations $x_1, x_2, \ldots, x_N$. If another set of observations $\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_N$ is used, a different Kullback Leibler distance would be obtained.

The expected Kullback-Leibler discrepancy is given by

$$-E_\theta\left(\int_X f_k(x) Log(g_\theta^k(x)) dx\right) \tag{3.4}$$

where the expectation is taken with respect to the distribution of the estimated parameter vector θ. This expression (3.4) cannot be computed, but estimated.

3.4 Model selection

The information theoretic criteria was first introduced by Akaike in [51] for model selection. Assuming a candidate model, the idea is to decide if the distribution of the observed signal fits the candidate model. The AIC criterion is an approximately unbiased estimator for (3.4) and is given by:

$$AIC_k = -2\sum_{n=1}^{N} Log(g_{\widehat{\theta}}^{k}(x_n)) + 2U \tag{3.5}$$

where U indicates the dimension of the parameter vector θ.

One should select the model that yields the smallest value of AIC because this model is estimated to be the closest to the unknown reality that generated the data, from among the candidate models considered.

The parameter vector θ for each family should be estimated using the minimum discrepancy estimator $\widehat{\theta}$, which minimizes the empirical discrepancy. This is the discrepancy between the approximating model and the model obtained by regarding the observations as the whole population.

In statistics, maximum likelihood estimation (MLE) is a method of estimating the parameters of a probability distribution by maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable. The point in the parameter space that maximizes the likelihood function is called the maximum likelihood estimate. The maximum likelihood estimator is the minimum discrepancy estimator for the Kullback–Leibler discrepancy [51].

Consider a density function distribution parameterized by an unknown parameter θ, associated with either a known density function or a known probability mass function, denoted as f_{θ}^{k}. As a function of θ with $x_1, x_2, \ldots, x_N$ fixed, the likelihood function is:

$$L_k(\theta) = f_{\theta}^{k}(x_1, x_2, \ldots, x_N) = \prod_{n=1}^{N} f_{\theta}(x_n) \tag{3.6}$$

The method of maximum likelihood estimates θ by finding the value of θ that maximizes $L_k(\theta)$. The maximum likelihood estimator (MLE) of θ is given by:

$$\widehat{\theta} = \arg_{\theta} max L_k(\theta) \tag{3.7}$$

Commonly, one assumes that the data drawn from a particular distribution are i.i.d. with unknown parameters. This considerably simplifies the problem because the likelihood can then be written as a product of N univariate density functions,

$$L_k(\theta) = \prod_{n=1}^{N} f_k(x_n|\theta) \tag{3.8}$$

and since maxima are unaffected by monotone transformations, one can take the Logarithm of this expression to turn it into a sum:

$$L_k^*(\theta) = \sum_{n=1}^{N} Logf_k(x_n|\theta) \tag{3.9}$$

Consequently, the expression of the maximum likelihood in our case is [73]:

$$\widehat{\theta} = \arg_\theta max \frac{1}{N} \sum_{n=1}^{N} Log(g_\theta^k(x_n)) \tag{3.10}$$

The maximum of this expression can then be found numerically using various optimization algorithms [74]. This contrasts with seeking an unbiased estimator of θ, which may not necessarily yield the MLE but which will yield a value that (on average) will neither tend to over-estimate nor under-estimate the true value of θ. The maximum likelihood estimator may not be unique, or indeed may not even exist.

Because AIC contains various constants and is a function of sample size, we routinely recommend computing the AIC differences (in addition to the actual AIC values),

$$\phi_k = AIC_k - AIC_{min} \tag{3.11}$$

where AIC_{min} denotes the minimum AIC value over all BSs.

Akaike weights can be computed using (3.5), in order to decide if the distribution of the Received Signal fits the candidate distribution or not. The Akaike weights can be interpreted as estimate for the probabilities that the corresponding candidate distribution show the best modeling fit. It provides another measure of the strength of evidence for this model, and is given by:

$$W_k = \frac{e^{-1/2\phi_k}}{\sum_{i=1}^{6} e^{-1/2\phi_i}} \quad where \quad k \in \{1,2,3,4,5,6\} \tag{3.12}$$

The Akaike weights allow us not only to decide if the distribution of the Received Signal fits the Gaussian distribution, but also provide information about the relative approximation quality of this distribution.

The maximum Likelihood estimator is the minimum discrepancy estimator for the KL discrepancy [51].

In our problem, we want Light Of Sight (LOS) signal between the BS and the users, which is a characteristic of electromagnetic radiation or acoustic wave propagation which means waves travel in a direct path from the source to the receiver. Consequently, we are going to use the Rice distribution. So the density function for the Received Signal for each BS is given by,

$$g_\theta^k(x|\mu_k, \sigma_k) = \frac{x}{\sigma_k^2} exp\left(-\frac{(x^2+\mu_k^2)}{2\sigma_k^2}\right) I_0\left(\frac{x\mu_k}{\sigma_k^2}\right) \tag{3.13}$$

Where $I_0(\frac{x\mu_k}{\sigma_k^2})$ is the modified Bessel function of the first kind with order zero , μ_k is the mean or expectation of the distribution (and also its median and mode) and σ_k is the standard deviation.

The approximated density function leads to the following Log-likelihood function,

$$L_k^*(\mu_k, \sigma_k) = Log\left(\frac{\prod_{i=1}^{N} x_i}{\sigma_k^{2N}} exp\left(\frac{-\sum_{i=1}^{N}(x_i^2+\mu_k^2)}{2\sigma_k^2}\right)\prod_{i=1}^{N} I_0\left(\frac{x_i\mu_k}{\sigma_k^2}\right)\right) \tag{3.14}$$

Parameters μ_k and σ_k are given by the solution of the following set of equations,

$$\begin{cases} \mu_k - \frac{1}{N}\sum_{i=1}^{N} x_i \frac{I_1\left(\frac{x_i\mu_k}{\sigma_k^2}\right)}{I_0\left(\frac{x_i\mu_k}{\sigma_k^2}\right)} = 0 \\ 2\sigma_k + \mu_k^2 - \frac{1}{N}\sum_{i=1}^{N} x_i^2 = 0 \end{cases} \tag{3.15}$$

Where $I_1\left(\frac{x_i\mu_k}{\sigma_k^2}\right) = -I_0\left(\frac{x_i\mu_k}{\sigma_k^2}\right) + \frac{\sigma_k^2}{2x\mu_k} I_0\left(\frac{x_i\mu_k}{\sigma_k^2}\right)$ is the modified Bessel function with order one. When $\frac{x_i\mu_k}{\sigma_k^2} >> 0.25$ and $I_0(\frac{x_i\mu_k}{\sigma_k^2}) = \frac{exp\left(\frac{x_i\mu_k}{\sigma_k^2}\right)}{\sqrt{2\pi\frac{x_i\mu_k}{\sigma_k^2}}}$, (3.15) can be expressed as:

$$\begin{cases} \mu_k^2 + \frac{1}{N}\sum_{i=1}^{N} x_i\mu_k - \frac{\sigma_k^2}{2} = 0 \\ \mu_k^2 - \frac{1}{N}\sum_{i=1}^{N} x_i^2 + 2\sigma_k^2 = 0 \end{cases} \tag{3.16}$$

Resolving (3.16), the MLE for the parameters $\widehat{\mu_k}$, $\widehat{\sigma_k}$ can be expressed as:

$$\begin{cases} \widehat{\mu_k} = \frac{-2\sum_{i=1}^{N} x_i + \sqrt{(4(\sum_{i=1}^{N} x_i)^2 + 5N\sum_{i=1}^{N} x_i^2)}}{5N} \\ \widehat{\sigma_k}^2 = \frac{1}{2}\widehat{\mu_k}^2 + \frac{1}{2N}\sum_{i=1}^{N} x_i^2 \end{cases} \tag{3.17}$$

And the parameter vector $\theta = (\sigma_k, \mu_k)$

In the Numerical Application section, we will compare the Rice Distribution Model and the Rayleigh Distribution Model, and the density function for the Rayleigh distribution is given by,

$$g(x|\sigma) = \frac{x}{\sigma^2} exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{3.18}$$

which leads to a Log-likelihood function

$$L^*(\sigma) = \sum_{i=1}^{p} Log(x_i) - pLog(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{p} x_i^2 \tag{3.19}$$

where the parameter $\theta = (\sigma)$. The MLE of the parameter σ is given by

$$\widehat{\sigma}^2 = \frac{1}{2p}\sum_{i=1}^{p} x_i^2 \tag{3.20}$$

3.5 Distribution Analysis Detector of Handovers

In this section, we present our new approach called Distribution Analysis Detector (DAD) to detect the best handover based on exploiting model selection techniques and especially AIC introduced by Akaike in [51].

We consider that the initial signal can be modeled using Gaussian distribution and its norm can be modeled using Rician distribution.

The flow chart of the proposed algorithm is shown in Figure 3.2, which can be implemented in four steps:

After the input of the values of the Received Signal for each BS (observations), in the first step we compute the parameters $\widehat{\mu_k}$ and $\widehat{\sigma_k}$ (MLE parameters), then g_{θ}^{k} the df for the Received Signal for each BS k. Once we get g_{θ}^{k}, we calculate AIC_k and W_k for each BS.

The Akaike weights allow us not only to decide if the distribution of the Received Signal fits the suitable distribution, but also provide information about the best signal (best BS) for each user.

If the Akaike weight of Rician distribution of the BS_k is higher than the Akaike weights of other BSs, then there is no Handover, and if the Akaike weight of BS_k is lower than the Akaike weight of BS_i where $i \in \{1, 2, 3, 4, 5, 6\}$ then there is Handover from BS_k to BS_i.

$$\lambda_{threshold}(x_n) = \begin{cases} W_k - W_i < \lambda_{threshold} & Handover \quad (H_0) \\ W_k - W_i > \lambda_{threshold} & No \quad Handover \quad (H_1) \end{cases} \tag{3.21}$$

The decision threshold is determined by using the probability of false alarm P_{FA} [75]. The threshold $\lambda_{threshold}$ for a given false alarm probability [75] is determined by solving the equation

$$P_{FA} = P(\lambda_{threshold}(x) < \lambda_{threshold} | H_1) \tag{3.22}$$

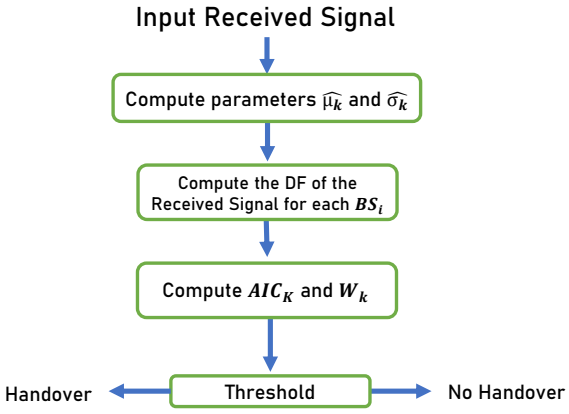

Figure 3.2 – Flowchart of Algorithm of Blind Handover based on distribution analysis

3.6 Study Case

The proposed Blind Detection approach is evaluated using the software package Matlab R2018a.

We summarize the Estimated parameters for each BS_i and their Akaike Weights in the table 3.3.

We apply the approach in Figure 3.2 and we compute the Akaike Weights for the BSs in terms to choose the best BS for the user.

Figure 3.4 depicts the Akaike Weights with Gaussian distribution obtained from the 6 BSs. It is clearly shown that the BS which has the Maximum Akaike weight is the first BS, so the

Index of BS_i	$\widehat{\mu_i}$	$\widehat{\sigma_i}$	Akaike Weight
1	6.1233	1.8749	0.69
2	2.3310	2.6959	0.01
3	2.3720	6.8872	0
4	3.2764	7.6820	0
5	1.3220	4.4329	0
6	1.4843	4.9318	0.3

Figure 3.3 – Estimated parameters for each BS_i and their Akaike Weights

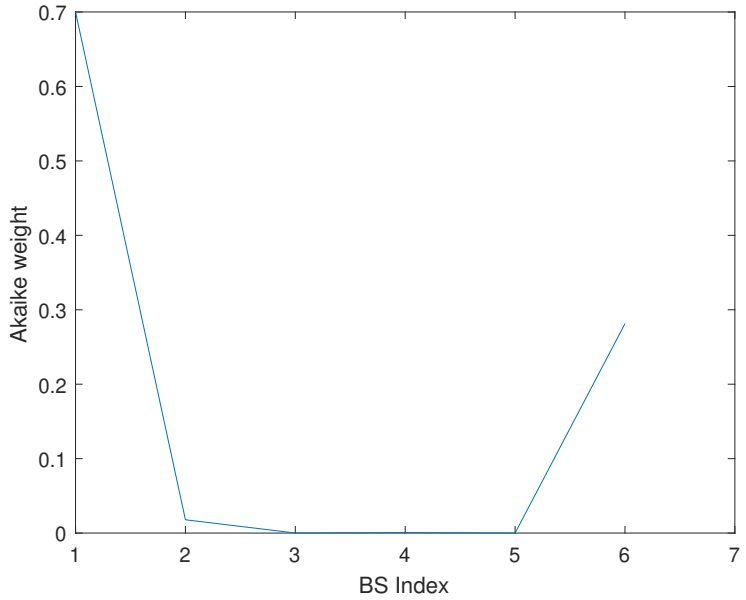

Figure 3.4 – Akaike Weights of the six BSs at time t

best BS for the user is the BS_1.

In Figure 3.5, we can see the difference between Rice and Rayleigh Distribution of the Received Signal. When the Signal between the BS and the UE is suffering from shadowing by a high building over the sensing channel, it definitely can decrease the Received Signal due to the low received SNR. When the SNR is low, the noise distribution will dominate in the convolution and the resulting distribution will tend to become close to Gaussian even if the signal has an arbitrary non Gaussian distribution, and the envelope (norm) distribution of the signal is close to Rayleigh distribution .

Another important property is the contribution of the dominant propagation paths on the distribution of the Received Signal. The envelope distribution of the Received Signal tend to become close to Rician even if the input has a non Rician distribution.

The Akaike weight of Rician distribution is higher than Akaike weight of Rayleigh distribution that mean that BS with Rician Distribution is the best for the UE.

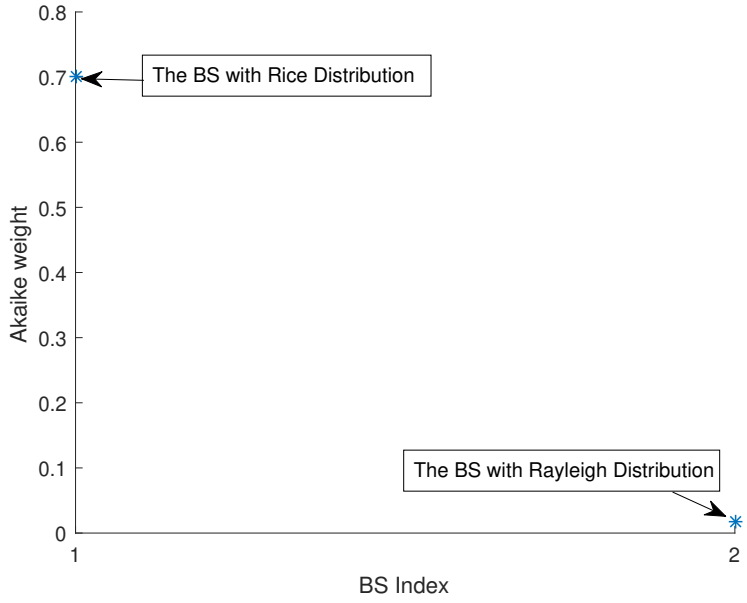

Figure 3.5 – The comparison of Akaike weight of two BSs with different distribution

3.7 Conclusion

In this chapter, we studied a new approach to manage the handovers between a number of users and Base Stations of Small Cells. Our idea has been based on analyzing the density function of the Received Signal for each BS, to provide an indication of the intensity of the Received Signal, and exploit KL Distance, Akaike Information Criterion and Akaike Weight in order to decide the best handover and the best BS for each user. The proposed Blind Detection Approach is evaluated using the software package Matlab R2018a. Another major concern in the Handover Management is the computation complexity.

Chapter 4

Combined Compressive Sampling Techniques and Features Detection using Kullback Leibler Distance to Manage Handovers

Contents

In this chapter, we introduce the concept of Compressive Sampling and we combine it with DAD.

4.1 Introduction

Over the previous decades, we have endorsed a growing demand for Mobile Cellular Communications Systems as a result of user demand for more flexible, wireless, smaller and practical devices. All this need engender more massive network densification.

The concept of cell densification has evolved from 4-5 macrocell base stations $(MBSs)/km^2$ in the 3G wireless networks to about 8–10 microcell BSs/km^2 in 4G. The primary aim of cell densification is to address the problem of capacity and coverage by spatial frequency reuse and offloading the data traffic to the SCs. The macrocells in 4G have a smaller area than those in 3G. Further, SCs bring users closer to the BS, which reduces the access distance, and consequently the path loss between them. In 5G, the SC area would be reduced further to support low-power transmissions, and hence the cell density is expected to increase to 40–50 small cell BSs $(SBSs)/km^2$.

A vital component that have the ability to significantly enhance network capacity, density and coverage are Small Cells (SCs)[10]. Like every other technology SCs have some shortcomings, and one of them is the problem of the Management of Handovers [12][13][14]. However, the Ultra Dense Networks increase the interference power levels, and consequently, the overall network performance might degrade due to inter-cell co-channel interference (CCI). Hence, interference mitigation is paramount for a network to achieve high spectral efficiency.

In this contribution our idea is to operate efficient and rapid Handover by Analyzing Received Signal Density Function instead of demodulating and Analyzing Received Signal itself, and shorten computational complexity and Analyze less required measurements with Compressive Sampling (CS) [76][77] which has been considered as a promising technique to reconstruct the original signal.

Compressive sensing makes it possible to reconstruct a sparse signal by taking less samples than Nyquist sampling, and thus Analyse the received signal is doable by CS. A Sparse Signal or Compressible Signal is a signal that is actually reliant on a number of degrees of freedom which is smaller than the dimension of the signal sampled at Nyquist rate. In general, signal of useful interest can be particularly nearly sparse [76].

In CS, in some basis, a sparse representation may be recovered from a small set of non-adaptive linear measurements. A sensing matrix catch less measurements of the signal, and the original signal can be reconstructed from partial observations by solving a straightforward convex optimization problem [76][78]. In [79] and [80] conditions on this sensing matrix are introduced which are sufficient in order to recover the original signal stably. And remarkably, a random matrix fulfills the conditions with high probability and performs an effective sensing.

Generally, for detection purposes it is not necessary to reconstruct the original signal, but only an estimate of the relevant sufficient statistics for the problem at hand is enough. This leads to less required measurements and lower computational complexity.

The remainder of the chapter is organized as follows. We begin by an introduction of our work. A brief overview about Compressive Sampling (CS) and the selection of the sensing matrix in section 2. In section 3, we revisit blind Handover technique based on Distribution Analysis presented in chapitre 3. In section 4 we will see some numerical applications. The last section will be devoted to the conclusion.

4.2 Compressive Sampling

Today's world is data driven. In many emerging applications such as medical imaging, video, data analysis, spectroscopy etc., the amount of data generated is too high. The resulting Nyquist rate is so high that we end up with far many samples. This will pose a tremendous challenge, as it is extremely difficult to build such devices that are capable of acquiring samples at the necessary rate. We can overcome this computational challenge especially in dealing with high-dimensional data by "Compression" techniques. In many practical problems of science and technoLogy, one encounters the task of inferring quantities of interest from measured information. The most popular technique used for signal compression is transform coding (which finds a basis that gives the sparser representation of signal).

This technique selects the most important data from the data set i.e., only the most important data is considered for the signal reconstruction process, thereby omitting the zero values (which are obtained by sparse representation of an image in some basis) . This is similar to data modelling/analysis where we extract the most important data from a huge data set.

Definition

Compressive Sensing (CS) also known as Sparse Signal Sampling is a new framework to reconstruct signals accurately and efficiently with less number of samples i.e., less than the Nyquist rate. According to the Shannon Nyquist sampling theorem, a signal can be reconstructed at a rate of twice the highest frequency of the signal.

Generally, compressive sensing works with sparse signals. In many applications the signal of interest is sparse i.e., the signal has a sparse representation in some predetermined basis where most of the coefficients are zero. Traditional measurement techniques oversample the signal heavily. Compressive sensing technique avoids excessive oversampling by linear sampling operators.

Fundamentals of compressive sensing

Compressive sensing, also referred to as compressed sensing or compressive sampling, is an emerging area in signal processing and information theory which has attracted a lot of attention recently. The motivation behind compressive sensing is to do "sampling" and "compression" at the same time. In conventional wisdom, in order to fully recover a signal, one has to sample the signal at a sampling rate equal or greater to the Nyquist sampling rate. However, in many applications such as imaging, sensor networks, astronomy, high-speed anaLog-to-digital

compression and bioLogical systems, the signals we are interested in are often "sparse" over a certain basis.

For instance, in signal and image processing, one would like to reconstruct a signal from measured data. When the information acquisition process is linear, the problem reduces to solving a linear system of equations. The measurement acquisition process is described by a matrix multiplication with a fat sensing-matrix and thus the reconstruction problem is an under-determined system of linear equations. The challenge in cs is two-fold: firstly, how to produce the measurement vector in practice and secondly, based on knowing the measurement vector and measurement matrix, how to find the correct underlying sparse signal-vector. The theory behind cs is based on the observation that many natural signals, such as sound or images, can be well approximated with a sparse representation in some domain. In sparse signal processing, one is usually interested in solving some detection or estimation problem, based on a measurement-vector whose size is not necessarily smaller than the size of the sparse signal-vector.

In mathematical terms, the observed data $y \in \mathbb{C}^m$ is connected to the signal $x \in \mathbb{C}^N$ of interest via

$$Ax = y$$

The matrix $A \in \mathbb{C}^{m*N}$ models the linear measurement (information) process. Then one tries to recover the vector $x \in \mathbb{C}^N$ by solving the above linear system. Traditional wisdom suggests that the number m of measurements, i.e., the amount of measured data, must be at least as large as the signal length N (the number of components of x). This principle is the basis for most devices used in current technoLogy, such as anaLog-to-digital conversion, medical imaging, radar, and mobile communication. Indeed, if $m < N$, then classical linear algebra indicates that the linear system is underdetermined and that there are infinitely many solutions (provided, of course, that there exists at least one). In other words, without additional information, it is impossible to recover x from y in the case $m < N$. This fact also relates to the Shannon sampling theorem, which states that the sampling rate of a continuous-time signal must be twice its highest frequency in order to ensure reconstruction.

Thus, it came as a surprise that under certain assumptions it is actually possible to reconstruct signals when the number m of available measurements is smaller than the signal length N. Even more surprisingly, efficient algorithms do exist for the reconstruction. The underlying assumption which makes all this possible is sparsity. The research area associated to this phenomenon has become known as compressive sensing, compressed sensing, compressive sampling, or sparse recovery [76][77].

Many natural signals are sparse in some basis and are reconstructed efficiently using Compressive Sensing framework. The CS framework has two major steps The first is Signal Acquisition – it is the process of acquiring compressed measurements and it is known as sensing. The second is Reconstruction – that is recovering of the original sparse signal from compressed measurements and is known as reconstruction.

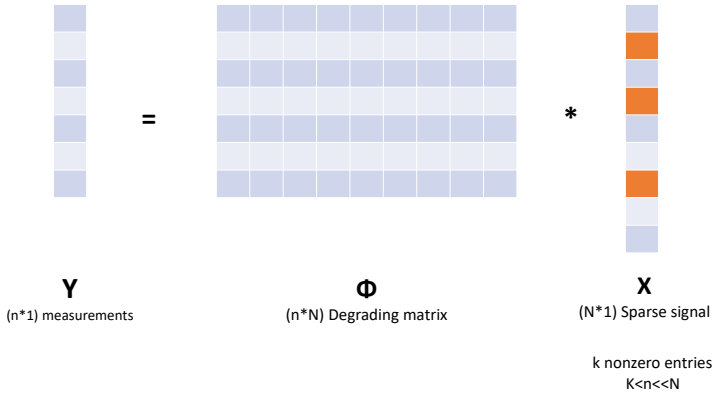

Figure 4.1 – Compressive Sampling

A signal can be presented as a $N * 1$ column vector in $\mathbb{R}^N$. A basis is a $N * N$ matrix (discrete cosine transform, wavelet, canonical etc.) which makes a signal sparse. In this thesis, we have considered wavelet basis for sparse representation of a signal. A signal x can be expressed as

$$x = \psi s$$

where s is a sparse representation of a signal $N * 1$ and ψ is a $N * N$ wavelet basis. The main compressive signal framework is expressed by the following equation,

$$y = \phi x = \phi \psi s$$

where ϕ is a degrading matrix of size $n * N$, which under-samples the sparse signal less than the Nyquist rate. The resultant y will be a low dimensional matrix of size $n * 1$. Since there are more unknowns than measurements, the system is classified as an undetermined system. It is clear that we cannot obtain an accurate input signal using the conventional inverse transform. Instead, we can obtain an input signal using compressive signal algorithms from fewer measurements.

The goal of compressive sensing is to design the matrix ϕ and a reconstruction algorithm so that for k-sparse signals we require only a "small" number of measurements, i.e. $m = k$ or slightly larger.

There are two fundamental properties underlying compressive signal: Sparsity and Incoherence.

Sparsity

Sparsity is a relevent term. In terms of signal theory, a sparse structure may mean having few large coefficients and many small coefficients. So the signal can be approximated by these large coefficients. In terms of anaLog signals, it may mean that the signal is constituted by

few of its basis functions at a unit interval of time, out of a large dictionary of possible basis functions. An other way of putting it is that we can define a signal in terms of such basis functions which give it a sparse structure but still we need to know the basis functions to reconstruct the signal.

A signal is called sparse if most of its components are zero. As empirically observed, many real-world signals are compressible in the sense that they are well approximated by sparse signals—often after an appropriate change of basis, Sparsity leads to dimensionality reduction and efficient modeling.

Mathematically, a vector x is expanded in basis $\psi = [\psi_1, \psi_2, ...\psi_N]$ and s is expressed as,

$$s = \psi^T x$$

in which s is the weighing coefficients of a signal.

When a signal has a sparse expansion in the wavelet basis, most of its coefficients are small and relatively few larger coefficients capture most of the information. Therefore, one can discard the smaller coefficients without any perceptual loss. A vector is sparse if it has fewer numbers of non-zeros than the number of zeros. By putting all together, we can say that the signal is k-sparse when it satisfies this condition $k < n << N$, (Figure 4.1).

Incoherence:

We have a pair of orthonormal bases (ϕ, ψ). The symbol ϕ is for sensing the object x and ψ is the representation of the object x. The coherence between the the pair of the basis (ϕ, ψ) is expressed by the following equation

$$\mu(\phi, \psi) = \sqrt{N}. \max_{1 \geq i,j \geq N} |\langle \phi_i, \psi_j \rangle|$$

Compressive signal is mostly concerned with low incoherence pairs. If the basis ϕ and ψ contain correlated elements, it has high coherence. Otherwise, they are less coherent. From linear algebra, the coherence between the bases are bounded with $\mu(\phi, \psi) \in [1, \sqrt{N}]$. If we consider noiselets for ϕ and wavelets for ψ hen the coherence between them is $\sqrt{2}$.

Looking closer at the standard compressive sensing problem consisting in the reconstruction of a sparse vector $x \in \mathbb{C}^N$ from underdetermined measurements $y = Ax \in \mathbb{C}^m, m < N$

In fact, compressive sensing is not fitted for arbitrary matrices $A \in \mathbb{C}^{N*m}$. For instance, if A is made of rows of the identity matrix, then $y = Ax$ simply picks some entries of x, and hence, it contains mostly zero entries. In particular, no information is obtained about the nonzero entries of x not caught in y, and the reconstruction appears impossible for such a matrix A. Therefore, compressive sensing is not only concerned with the recovery algorithm—the first question on the design of the measurement matrix is equally important and delicate. We also emphasize that the matrix A should ideally be designed for all signals x simultaneously, with a measurement process which is nonadaptive in the sense that the type of measurements for the datum y_j (i.e., the j_{th} row of A) does not depend on the previously observed data $y_1, \ldots, y_{j-1}$. As it turns out, adaptive measurements do not provide better theoretical performance in general.

minimisation Problem

One major approach, Basis Pursuit, relaxes the l_0-minimization problem to an l_1 minimization problem:

$$min\ ||x||_1 \qquad subject\ to \qquad Ax = y$$

Simply put, instead of trying to find the solution with the smallest l_0-norm, l_1 minimization tries to find the solution with the minimum l_1 norm. Surprisingly, this relaxation often recovers x exactly when x is sparse or accurately when x is an approximately sparse signal or compressible signal. Please note that the measurement matrix A is given and fixed in advance, and does not depend on the signal, but as long as the signals are sufficiently sparse and the measurement matrix satisfies some conditions independent of the signals, the l_1 minimization will succeed. That is, even though l_1-norm is different from the quasi-norm l_0, the solution of l_1 often comes as the sparsest solution.

Algorithms

For practical purposes, the availability of reasonably fast reconstruction algorithms is essential. This feature is arguably the one which brought so much attention to compressive sensing. The first algorithmic approach coming to mind is probably l_0-minimization. Introducing the notation $||x||_0$ for the number of nonzero entries of a vector x, it is natural to try to reconstruct x as a solution of the combinatorial optimization problem

$$minimize\ ||z||_0 \qquad subject\ to \qquad Az = y$$

In words, we search for the sparsest vector consistent with the measured data $y = Ax$. Unfortunately, l_0-minimization is NP-hard in general. Thus, it may seem quite surprising that fast and provably effective reconstruction algorithms do exist. A very popular and by now well-understood method is basis pursuit or l_1-minimization, which consists in finding the minimizer of the problem

$$minimize\ ||z||_1 \qquad subject\ to \qquad Az = y$$

Since the l_1-norm $||.||_1$ is a convex function, this optimization problem can be solved with efficient methods from convex optimization. Basis pursuit can be interpreted as the convex relaxation of l_0-minimization. Alternative reconstruction methods include greedy-type methods such as orthogonal matching pursuit, as well as thresholding-based methods including iterative hard thresholding.

Random Matrices

We first explain how to design the matrix ϕ. The ultimate goal is to have some matrix ϕ which does not destroy any information contained in the original signal x. However since $\phi \in \mathbb{R}^{n*N}$ and $m < n$ this is not possible in general as the equation of y is under-determined, making the problem of solving for s or x ill-posed.

By restricting ourselves to k-sparse signals we can do significantly better. If the position of the k non-zero entries of s were known a priori, i.e. we could form the $n*k$ matrix where $n > k$ and solve the least squares problem restricted to the non-zero positions of x. A necessary and sufficient condition for this $n * k$ system of equations to be well conditioned is that for any k-sparse vector $v \in \mathbb{R}^{\mathbb{N}}$ we have

$$1 - \epsilon \leq \frac{||\phi\psi v||_2}{||v||_2} \leq 1 + \epsilon$$

for some $\epsilon > 0$, that is the matrix $\phi\psi$ must almost preserve the lengths of these k-sparse vectors.

Definition For each integer $k = 1, 2, \ldots$ define the isometry constant $\sigma_k > 0$ of a matrix ϕ as the smallest number such that

$$1 - \sigma_k \leq \frac{||\phi x||_2^2}{||x||_2^2} \leq 1 + \sigma_k$$

holds for all k-sparse vectors x. We say that a matrix ϕ has the restricted isometry property (RIP) of order k if σ_k is not too close to 1.

The restricted isometry constants give a measure of how much they can change the length of a k-sparse vector. They also relate to the kernel of the matrix: suppose for instance that $\sigma_k < 1$ for some integer k. This implies that that there are no k-sparse vectors in the kernel of ϕ.

One approach to obtaining a matrix F with the RIP of high order is to use random matrices. The $n * N$ matrices generated according to the following rules:

- form ϕ by sampling n column vectors uniformly on the unit sphere in $\mathbb{R}^n$.
- let the entries of ϕ be i.i.d. normal with mean 0 and variance $\frac{1}{n}$, or
- let the entries ϕ by i.i.d. symmetric Bernoulli distributed, e.g. $\phi_{ij} = ś\frac{1}{\sqrt{n}}$ with equal probability, or any other subgaussian distribution.

all obey the restricted isometry property of order k provided that

$$m \geq C.kLog(\frac{n}{k})$$

for some constant C.

Stability

Compressive sensing features another crucial aspect, namely, its reconstruction algorithms are stable. This means that the reconstruction error stays under control when the vectors are not exactly sparse and when the measurements y are slightly inaccurate. In this situation, one may, for instance, solve the quadratically constrained l_1-minimization problem

$$minimize \ ||z||_1 \qquad subject \ to \qquad ||Az - y||_2 \leq \eta$$

Without the stability requirement, the compressive sensing problem would be swiftly resolved and would not present much interest since most practical applications involve noise and compressibility rather than sparsity.

From a general viewpoint, sparsity and, more generally, compressibility has played and continues to play a fundamental role in many fields of science. Sparsity leads to efficient estimations, efficient compression, dimensionality reduction and efficient modeling. A signal having a sparse representation in some basis can be reconstructed from a small set of nonadaptive, linear measurements. Briefly, this is accomplished by generalizing the notion of a measurement or sample to mean the application of a linear functional to the data.

Let $x \in \mathbb{R}^N$ be a signal with expansion in an orthonormal basis Ψ as

$$x(t) = \sum_{j=0}^{N-1} \alpha_j \psi_j(t), \quad t = 0, 1, .., N-1 \tag{4.1}$$

where Ψ is the $N * N$ matrix with the waveforms ψ_j as rows. To use convenient matrix notations we can write the decomposition as $x = \Psi\alpha$ or equivalently, $\alpha = \Psi^* x$ where Ψ^* denotes conjugate transpose of Ψ.

A Sparse Signal, is a signal which contains only a limited number of non-zero elements compared to its dimension, the implication of sparsity is now clear: when a signal has a sparse expansion, one can discard the small coefficients without much perceptual loss, in our conditions, in the Ψ basis if the coefficient sequence α is supported on a small set [76]. We say that a vector α is S-sparse if its support $j : \alpha_j \neq 0$ is of cardinality less or equal to S.

Hence, to recover all the N coefficients of x, vector α, from measurements y about x of the form

$$y_m = \langle x, \phi \rangle = \sum_{n=0}^{N-1} \phi_{mn} x[n], \quad m = 0, ..M-1 \tag{4.2}$$

or

$$y = \Phi x = \Phi\Psi\alpha = \Theta\alpha \tag{4.3}$$

where Φ is $M * M$ matrix, called the Sensing Matrix. We are concerned in the case that $M << N$, and the rows of Φ are incoherent with the columns of Ψ. Then it is shown that the recovered signal x^* is given by $x^* = \Psi\alpha^*$, and α^* is the solution to the convex optimization program

$$min_{\widetilde{\alpha} \in R^n} \ \widetilde{\alpha}_{\ l_1} \ subject \ to \ \ \Phi\Psi\widetilde{\alpha} = \Theta\widetilde{\alpha} = y \tag{4.4}$$

where $\widetilde{\alpha}_{\ l_1} := \sum_{j=1}^{N} |\widetilde{\alpha}|$.

The compressed sensing (CS) theory affirms that there exists a counting factor $c > 1$ such that only $M := cS$ incoherent measurements y are needed to recover x with high probability. We additionally have to notice that except l_1-minimization solution other methods like greedy algorithms in [78] exist for recovering the sparse signal [76] [78][81].

In case of noisy measurements, i.e., $y = \Phi x + e$, where e is the noise with $e_{l_2} \leq \varepsilon$, [6] shows that solution to

$$min_{\widetilde{\alpha} \in R^n} \ \widetilde{\alpha}_{\ l_1} \ subject \ to \ \ \Theta\widetilde{\alpha} - y_l \leq \varepsilon \tag{4.5}$$

recovers the sparse signal with an error at most proportional to the noise level.

Our intention is to apply the Handover Detection using DAD on the Compressed measurements of the observed signal, so we have to maintain the linearity and properties of the original signal. For this reason we must identify the suitable sensing matrix according to the detection technique.

To identify the sensing matrix we start by examining the Fourier transform of the signal $x \in \mathbb{R}^N$ [82].

$$X_l = \sum_{n=0}^{N-1} x[n]exp(-wln), \quad l = 0, ..., N-1 \tag{4.6}$$

where, $w = \frac{2\pi i}{n}$ and i is the imaginary unit.

The Fourrier transform of the measured signal is:

$$Y_k = \sum_{m=0}^{M-1} y[m]exp(-wkm), \quad k = 0, ..., M-1 \tag{4.7}$$

We replace $y[m]$ and we have

$$Y_k = \sum_{m=0}^{M-1} (\sum_{n=0}^{N-1} \phi_{mn} x[n])exp(-wkm), \quad k = 0, ..., M-1. \tag{4.8}$$

where ϕ_{mn} denotes the element of Φ at the cross of row m and column n.

Then, by linearity properties we have:

$$Y_k = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \phi_n[m]exp(-wkm)x[n], \quad k = 0, ..., M-1. \tag{4.9}$$

where: $\phi_n[m]$ denotes the m^{th} element of the n^{th} column vector of Φ, and we see that,

$$\sum_{m=0}^{M-1} \phi_n[m]exp(-wkm) = \hat{\Phi}_{n_k}, \quad k = 0, ..., M-1. \tag{4.10}$$

that is the Fourrier transform of the n^{th} column vector of Φ,

$$Y_k = \sum_{n=0}^{N-1} \hat{\Phi}_{n_k} x[n], k = 0, ..., M-1. \tag{4.11}$$

And, to feed the detection algorithm directly by the compressed measurements we observe that,

$$Y_k(w) = aX_l(w), k \in \{0, ..., M-1\}, \quad l \in \{0, ..., N-1\}. \tag{4.12}$$

where $a > 0$ is a constant, and $\hat{\Phi}_{n_k}$ is described as

$$\sum_{m=0}^{M-1} \phi_n[m] exp(-wkm) = \hat{\Phi}_{n_k}, \quad k = 0, ..., M-1. \tag{4.13}$$

So finally, we obtain that,

$$\hat{\Phi}_{n_k} = aexp(-wzn), z \in \{1, ..., N\}, \quad k \in \{0, ..., M-1\}. \tag{4.14}$$

And accordingly from inverse Fourrier transform we have

$$\phi_n = a\delta(n-z), \quad z \in \{1, ..., N\}. \tag{4.15}$$

which means that any row vector of the sensing matrix is a Dirac Function, that is solely one column of each row is nonzero.

To generate the sensing matrix we may begin by generating Φ^T matrix by randomly choosing M columns of an identity matrix I_N. The sensing matrix, Φ, is given by transpose of Φ^T, where the columns of the sensing matrix are unit-normed. So the sensing matrix Φ that we utilize has a form like this

$$\Phi \sim \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \end{bmatrix}_{M*N} \tag{4.16}$$

This form of sensing matrix permit to us to employ the compressed measurements from each Base Station directly as input to the DAD algorithm and accordingly avoiding the computation complexity of reconstructing the original signal. Then, the description of the DAD is interpreted.

4.3 Blind Handover Technique based on Distribution Analysis

In this section, we present the major ideas of the Blind Handover technique based on Distribution Analysis. The intention of our Blind Handover is to choose between the ensuing two hypothesis:

$$\lambda_{threshold}(x_n) = \begin{cases} W_k - W_i < \lambda_{threshold} & Handover \quad (H_0) \\ W_k - W_i > \lambda_{threshold} & No\,Handover \quad (H_1) \end{cases} \tag{4.17}$$

where $\lambda_{threshold}(x_n)$ is the estimated decision, and $\lambda_{threshold}$ is the decision threshold and is determinated by using the Probability of False Alarm P_{FA}.

The threshold $\lambda_{threshold}$ for a specified False Alarm Probability is determinated by solving the equation

$$P_{FA} = P(H_0|H_1) = P(\lambda_{threshold}(x) < \lambda_{threshold}|H_1) \tag{4.18}$$

That is the probability of detecting a Handover under the hypothesis H_1, and the Probability of Detection P_D expressed as:

$$P_D = 1 - P_M = 1 - P(H_1|H_0) \tag{4.19}$$

where P_M indicates the probability of a missed detection of Handover.

The principal idea is to detect the best BS for each user (Best Handover) by exploiting Model Selection techniques and especially the AIC. It was shown in [73] that, when signal demodulation cannot be performed, the received wireless communication signal can be, particularly, modeled using Rayleigh and Rician distribution. Consequently, we suggest to calculate in Blindly process the Received Signal for each BS and Analyze AIC in order to determine the Best Handover.

Actually, it is given that the samples of the Received Signal for each BS are distributed according to an original probability density function f_k where $k \in \{1, 2, 3, 4, 5, 6\}$ is the index of BS, called the operating model. Considering just a limited number of observations is available, the operating model is generally unknown. Therefore, approximating model (i.e candidate model) must be specified using the observed data, in order to estimate the operating model. The candidate model is denoted as g_θ^k , where θ indicates the U-dimensional parameter vector, which specifies the probability density function.

In information theory, the Kullback-Leibler distance characterizes the discrepancy between f_k and g_θ^k and is given by:

$$D(f_k \ \ g_\theta^k) = E(Log(f_k(x))) - E(Log(g_\theta^k(x)))$$

This distance measure is not directly applicable, because the original probability density function f_k is not known.

And the expected Kullback-Leibler discrepancy is given by

$$-E_\theta(\int f_k(x) Log(g_\theta^k(x)) dx) \tag{4.20}$$

This expression cannot be computed, but estimated.

An approximately unbiased estimator for (15) is the AIC criterion initiated by Akaike for model selection and is given by

$$AIC_k = -2\sum_{n=1}^{N} Log(g_\theta^k(x_n)) + 2U \tag{4.21}$$

where U expresses the dimension of the parameter vector θ.

Akaike's proposed was to select the model which gives the minimum AIC.

The parameter vector θ for each family have to be estimated applying the minimum discrepancy estimator $\widehat{\theta}$, which minimizes the empirical discrepancy. This is the discrepancy between the approximating model and the model obtained by regarding the observations as the whole population.

Because AIC contains various constants and is a function of sample size, we routinely recommend computing the AIC differences:

$$\phi_k = AIC_k - AIC_{min} \tag{4.22}$$

where AIC_{min} denotes the minimum AIC value over all BSs.

Akaike weights can be calculated using (4.21), so that we can decide if the distribution of the Received Signal is conform to the candidate distribution or not. The Akaike weights may be elucidated as estimate for the probabilities that the corresponding candidate distribution show the best modeling fit. It determines additional measure of the strength of evidence for this model, and is given by:

$$W_k = \frac{e^{-1/2\phi_k}}{\sum_{i=1}^{6} e^{-1/2\phi_i}} \quad where \quad k \in \{1,2,3,4,5,6\} \tag{4.23}$$

The Akaike weights permit us not only to decide if the distribution of the Received Signal fits the Gaussian distribution, but also provide information about the relative approximation quality of this distribution.

The maximum Likelihood estimator is the minimum discrepancy estimator for the KL discrepancy.

In our problem, we want Light Of Sight (LOS) signal between the BS and the users. Consequently, we are going to use the Rice distribution. So the probability density function for the Received Signal for each BS is given by:

$$g_\theta^k(x|\mu_k, \sigma_k) = \frac{x}{\sigma_k^2} exp(-\frac{(x^2+\mu_k^2)}{2\sigma_k^2}) I_0(\frac{x\mu_k}{\sigma_k^2}) \tag{4.24}$$

Where $I_0(\frac{x\mu_k}{\sigma_k^2})$ is the modified Bessel function of the first kind with order zero , μ_k is the mean or expectation of the distribution (and also its median and mode) and σ_k is the standard deviation.

The approximated probability density function leads to the following Log-likelihood function :

$$L_k^*(\mu_k, \sigma_k) = Log(\frac{\prod_{i=1}^{N} x_i}{\sigma_k^{2N}} exp(\frac{-\sum_{i=1}^{N}(x_i^2+\mu_k^2)}{2\sigma_k^2}) \prod_{i=1}^{N} I_0(\frac{x_i\mu_k}{\sigma_k^2})) \tag{4.25}$$

Parameters μ_k and σ_k are given by the solution of the following set of equations:

$$\begin{cases} \mu_k - \frac{1}{N}\sum_{i=1}^{N} x_i \frac{I_1(\frac{x_i\mu_k}{\sigma_k^2})}{I_0(\frac{x_i\mu_k}{\sigma_k^2})} = 0 \\ 2\sigma_k + \mu_k^2 - \frac{1}{N}\sum_{i=1}^{N} x_i^2 = 0 \end{cases} \tag{4.26}$$

Where $I_1(\frac{x_i\mu_k}{\sigma_k^2}) = -I_0(\frac{x_i\mu_k}{\sigma_k^2}) + \frac{\sigma_k^2}{2x\mu_k} I_0(\frac{x_i\mu_k}{\sigma_k^2})$ is the modified Bessel function with order one. When $\frac{x_i\mu_k}{\sigma_k^2} >> 0.25$ and $I_0(\frac{x_i\mu_k}{\sigma_k^2}) = \frac{exp(\frac{x_i\mu_k}{\sigma_k^2})}{\sqrt{2\pi\frac{x_i\mu_k}{\sigma_k^2}}}$, (21) can be expressed as:

$$\begin{cases} \mu_k^2 + \frac{1}{N}\sum_{i=1}^{N} x_i\mu_k - \frac{\sigma_k^2}{2} = 0 \\ \mu_k^2 - \frac{1}{N}\sum_{i=1}^{N} x_i^2 + 2\sigma_k^2 = 0 \end{cases} \tag{4.27}$$

Resolving (4.27), the MLE for the parameters $\widehat{\mu_k}$, $\widehat{\sigma_k}$ can be expressed as:

$$\begin{cases} \widehat{\mu_k} = \frac{-2\sum_{i=1}^{N} x_i + \sqrt{(4(\sum_{i=1}^{N} x_i)^2 + 5N\sum_{i=1}^{N} x_i^2)}}{5N} \\ \widehat{\sigma_k}^2 = \frac{1}{2}\widehat{\mu_k}^2 + \frac{1}{2N}\sum_{i=1}^{N} x_i^2 \end{cases} \tag{4.28}$$

And the parameter vector $\theta = (\sigma_k, \mu_k)$

4.4 Study Case

The proposed Blind Detection approach is evaluated using the software package Matlab R2018a.

In this section we investigate the performance of the proposed algorithm in comparison with Distribution Analysis Detector (DAD).

We apply the approach presented in 4.2 and we compute the Akaike Weights for the BSs in terms to choose the best BS for the user.

We summarize the Estimated parameters for each BS_i and their Akaike Weights in the table 4.3.

The complexity of Handover Detection is a major concern in Handover Management. Using the implementation steps of the two detectors, we will study the complexity required for each detector to derive their Handover Algorithm.

The complexity of the algorithm is measured through the number of complex multiplications that the algorithms has to perform for the calculation of the test statistics. We summarize the number of multiplications required for each technique in table 4.4.

Note that N refers to the number of samples of the received signal and M is the number of samples after Compress Sensing. For the computer performance note that the Laptop Processor 11th Gen Intel(R) Core(TM) i7-1165G7 2.80GHz 2.80 GHz.

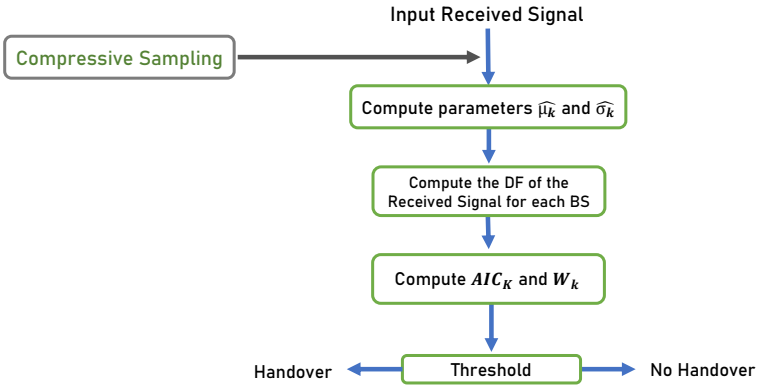

Figure 4.2 – Flowchart of Algorithm of Blind Handover based on distribution analysis with Compressive Sampling

Index of BS_i	$\widehat{\mu_i}$	$\widehat{\sigma_i}$	Akaike Weight
1	6.3030	1.9866	0
2	4.7226	1.1153	0
3	5.5804	1.9369	0
4	6.5761	2.5385	0.3
5	6.2326	2.2945	0.1
6	7.1991	2.7916	0.6

Figure 4.3 – Estimated parameters for each BS_i and their Akaike Weights

Handover Detection Technique	Complexity	Elapsed Time (s)
Distribution Analysis Detector	2N	0.202697
Distribution Analysis Detector + Compressive Sampling	2M	0.065626

Figure 4.4 – Complexity Comparison of the two Handovers Detection Techniques: DAD and DAD+ Compress Sensing

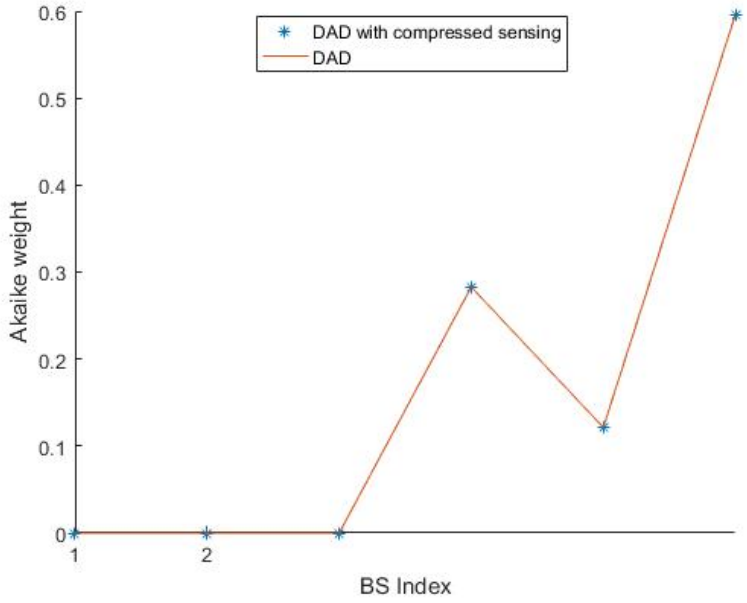

Figure 4.5 – Akaike Weight vs BS index for two detectors: DAD and DAD with Compress Sensing

From these results, we find that Compress Sensing decreases the complexity and the elapsed time as compared to the simple Distribution Analysis Detector.

Figure 4.5 depicts the comparison between two detectors. This figure shows the Akaike Weight for six Base Stations calculated with two techniques DAD and DAD with Compress Sensing. From the Numerical application results, we show that the two detectors give the same Akaike weight which means that the use of Compress Sensing doesn't change the final result. It is clearly shown that the Base Station which has the Maximum Akaike Weight is the six Base Station, so the best Base Station for the user is the Base Station 6.

4.5 Conclusion

We presented in this chapter a new sensing technique which combines compressive sampling and DAD to detect Handovers. In the first step, we designed a sensing matrix which keeps the linear properties of the sampled primary signal. Then, we applied the compressed measurements to the DAD in order to make the detection possible with less number of samples or smaller sampling rate. The analysis of the complexity of the proposed technique shows that it can be reduced. The Numerical comparison at different sampling rates shows that the new designed scheme achieves the same results as the DAD while preserving a low computational complexity. The Handover between base stations and User effect the capacity of the network and there is a correlation between the cell-quality and network performance.

Chapter 5

Blind Handover detection based on KLD and Channel Capacity, Outage Probability Estimation for Rice and Nakagami Models

Contents

In this chapter, we propose our blind Handover detection based on KLD and Channel Capacity, Outage Probability Estimation for Rice and Nakagami Models.

5.1 Introduction

From 2010 to 2020, we can observe an exponential growth in the amount of traffic carried through mobile networks [6]. According to Cisco, Mobile data traffic has doubled during 2010–2011, and most of the devices work in a wireless manner, which rely them with power battery and that can limit the amount of time they can operate, so, one of the biggest obstacles to this technoLogy is the limited battery lifetime. To confront this problem, Fifth Generation (5G) [83] [8] [9] recommend to consider the implementation of small cells [10], to decrease the transmit power and optimize energy consumption.

A wireless network designates a type of network which has the particularity of being wireless, capable of establishing a connection between several channels [44] sharing space and frequency band, and to provide continuous connectivity we need more handovers [14] [12][13] between those channels. To establish a Handover there is lot of metrics: Received Signal Strength Indicator (RSSI), distance, load, battery and Physical Cell Id (PCI).

Today's world is data driven. In many emerging applications such as medical imaging, video, data analysis, spectroscopy etc., the amount of data generated is too high. The resulting Nyquist rate is so high that we end up with far many samples. This will pose a tremendous challenge, as it is extremely difficult to build such devices that are capable of acquiring samples at the necessary rate. We can overcome this computational challenge especially in dealing with high-dimensional data by "Compression" techniques.

For future wireless communication systems, the capacity is increasingly required to provide some new technoLogies at high data transmission rate throughput on a limited bandwidth and power. The notion of channel capacity has been central to the development of wireless communication systems, with the advent of novel error correction coding mechanisms that have resulted in achieving performance very close to the limits promised by channel capacity.

The Handover between base stations and User effect the capacity of the network and there is a correlation between the cell-quality and network performance. The Shannon capacity of a channel defines its theoretical upper bound for the maximum rate of data transmission at an arbitrarily small bit error rate (BER), without any delay or complexity constraints. Therefore, the Shannon capacity represents an optimistic bound for practical communication schemes, and also serves as a bench-mark against which to compare the spectral efficiency of all practical adaptive transmission schemes.

The variability of channel capacity in a communication channel causes the change in the probability that a given information rate is not supported, called the Outage probability, which

is defined as the probability that information rate is less than the required threshold information rate. It is the probability that an outage will occur within a specified time period.

In [4] the authors presented a new approach to perform efficient handover through the analysis of received signal density function based on some information theory tools, called, Kullback Leiber Distance (KLD), Akaike Information Criterion (AIC) and Akaike weights.

The main objectives of this chapter is to apply the Compressive Sampling on the proposed approach presented in [4] for Handover management on the Nakagami distribution model [84] which is the typical channel condition often considered in the transmission of communication systems, to feed directly with the compressed measurements and determine blindly the best handover and the best BS for each user and also estimate the performance of the network in terms of channel capacity and outage probability and compare between the two models of signal distribution Rice and Nakagami [5].

The remainder of the chapter is structured as follows. The following presents some related work. A brief overview about Compressive Sampling (CS) in section 3. The Nakagami distribution model is described in section 4. In section 5, we represent KL Distance, the Maximum Likelihood Estimator and the approach. After that, the Channel Capacity and the Outage probability for Rice and Nakagami distribution are estimated in section 6. Numerical evaluations are presented in Section 7 in order to compare our models. We end our paper by a conclusion.

5.2 Related Work

With the expeditious advancement in R&D of wireless technoLogies, the integration of various technoLogies offers multiple services anytime and anywhere. The major goal of this networks is a seamless connection for the user. This process of change in the communication channel is called the handover, which is an important component in wireless network mobility management.

Diverse appropriate Handover management techniques are available in the literature. A crucial requirement is designed for the switch is based on one parameter measurement. Received Signal Strength Indicator (RSSI) protocol consider received signal strength as a criteria to select the appropriate channel for the user.

The proposed strategy in [85] considers hybrid RF small-cell networks, which create frequent unnecessary handover because the measurements are based only on RSSI. Furthermore, user mobility decrease the system throughput.

Otherwise, Chang et al. [86] evolved a handover decision method in two phases: RSSI prediction and Markov decision process. The strategy uses the traditional measure of RSSI and compares the RSSI values of the serving point of attachment. This increase the computation complexity and unneeded handovers.

A mathematical model to skip unneeded handover to the WLAN network in [87]. This strategy measures the RSSI available for the mobile user. Mobility of the mobile node in this scenario seems to be unnoticed.

In brief, for those strategies the RSSI is the only criterion in the handover decision strategy. However, the signal fluctuations resulting from the fading effect cause the undesirable so-called ping pong effect, which rise the probability of call loss during the handover process.

There is also, computation handover decision in the literature, which fail to consider a network load, results in over-utilization of the network, so the network capacity becomes full, this causes call blocking and call dropping. The traditional strategies are the simplest ones because the handover relies on one parameter. Multiple parameters are adopted in [88] where the complexity still exists and the mobile terminal based decision makes the handover policy unreliable in a heterogeneous environment.

There is a special feature, where the network node is allowed to start a handover process for a terminal without considering traditional measurements configuration is called the Blind Handover [89] like beacon pilot technique which have some drawbacks like the increase of the cost of the system infrastructure and the decrease on the network capacity because of the generation of interference on the target network.

5.3 Compressive Sampling

Compressive sensing is an emerging research field that has applications in signal processing, error correction, medical imaging, seismoLogy, and many more other areas. It promises to efficiently recover a sparse signal vector via a much smaller number of linear measurements than its dimension.

Before we go into greater technical details, we will first give the general signal models discussed in this paper [?]. Let $x \in \mathbf{R}^N$ be a signal with expansion in an orthonormal basis Ψ as

$$x = \Psi s \tag{5.1}$$

where Ψ is a $N * N$ matrix and s is the sparse representation wavelet basis. A sparse signal, is a signal which contains only a limited number of non-zero elements compared to its dimension, in our conditions, in the Ψ basis if the coefficient sequence s is supported on a small set.

Hence, to recover all the N coefficients of x, vector s, from measurements y about x of the form

$$y = \Phi x = \Phi \Psi s = \Theta s \tag{5.2}$$

where Φ is $M * M$ matrix, called the Sensing Matrix. We are concerned in the case that $M << N$, and the rows of Φ are incoherent with the columns of Ψ.

The aim of compressive sensing is to design the matrix Φ and a reconstruction algorithm so that for k-sparse signals we require only a "small" number of measurements, i.e. $m \approx k$.

In compressive sensing, random measurement matrices are generally used and l_1 minimization algorithms often use linear programming or other optimization methods to recover the sparse signal vectors. But explicitly constructible measurement matrices providing performance guarantees were elusive and l_1 minimization algorithms are often very demanding in computational complexity for applications involving very large problem dimensions.

The popular and powerful l_1 minimization algorithms generally give better sparsity recovery performances than known greedy decoding algorithms.

Then it is shown that the recovered signal x^* is given by $x^* = \Psi s^*$, and s^* is the solution to the convex optimization program

$$\min_{\widetilde{s}\in \mathbf{R}^N} ||\widetilde{s}|_{l_1} subject\ to\ \Phi\Psi\widetilde{s}\ = \Theta\widetilde{s}\ = y \tag{5.3}$$

The compressed sensing (CS) theory confirms that there exists a counting factor $c > 1$ such that only $M := cS$ incoherent measurements y are needed to recover x with high probability.

Furthermore, we have to observe that except l_1-minimization solution other methods like greedy algorithms in [78] exist for recovering the sparse signal [76] [78][81].

Our intention is to apply the Handover detection using Distribution Analysis Detector (DAD) on the compressed measurements of the observed signal, so we have to maintain the linearity and properties of the original signal. For this consideration we have to recognize the suitable sensing matrix conform to the detection technique.

To identify the sensing matrix we start by examining the Fourier transform of the signal $x \in \mathbf{R}^N$ [82].

$$X_l = \sum_{n=0}^{N-1} x[n]exp(-wln),\ l = 0, 1, ..., N-1 \tag{5.4}$$

where $w = \frac{2\pi i}{n}$ and i is the imaginary unit. The Fourrier transform of the measured signal is

$$Y_k = \sum_{m=0}^{M-1} y[m]exp(-wkm),\ k = 0, 1, ...M-1 \tag{5.5}$$

And, to satisfy the detection algorithm directly by the compressed measurements we observe that,

$$Y_k(w) = aX_l(x),\ k \in 0, .., M-1,\ l \in 0, .., N-1 \tag{5.6}$$

where $a > 0$ is a constant, and $\hat{\Phi}_{n_k}$ is described as

$$\sum_{m=0}^{M-1} \phi_n[m]exp(-wkm) = \hat{\Phi}_{n_k},\ k = 0, .., M-1 \tag{5.7}$$

So finally, we obtain that

$$\hat{\Phi}_{n_k} = aexp(-wzn),\ z \in 0,..,N,\ k = 0,..,M-1 \tag{5.8}$$

And accordingly from inverse Fourrier transform we have

$$\phi_n = a\delta(n-z),\ z \in 0,..,N \tag{5.9}$$

which means that any row vector of the sensing matrix is a Dirac Function, that is solely one column of each row is nonzero.

To make the sensing matrix we may begin by generating Φ^T matrix by randomly choosing M columns of an identity matrix N. The sensing matrix, Φ, is given by transpose of Φ^T , where the columns of the sensing matrix are unit-normed. So the sensing matrix Φ that we apply has a form like this

$$\Phi \sim \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \end{bmatrix}_{M*N} \tag{5.10}$$

This form of sensing matrix permit us to apply the compressed measurements from each BS directly as input to the DAD algorithm and accordingly avoiding the computation complexity of reconstructing the original signal.

5.4 Nakagami distribution

In the early 1940's, the Nakagami distribution was introduced by Nakagami in [84]. The Nakagami distribution has gained a lot of attention lately, because of its capability to model a vast class of fading channel conditions, and present a closer match to empirical data than the Rayleigh and Rice distributions.

The Nakagami distribution, like the Rayleigh and Ricean distributions, is based on the Normal, or Gaussian, distribution. The Nakagami density function is described by,

$$f(x) = \frac{2m^m}{\Gamma(m)\Omega^m} x^{2m-1} exp(-\frac{m}{\Omega}x^2)\ , \quad x \geq 0 \quad m \geq 1/2 \tag{5.11}$$

where m is a shape parameter, and Ω is the mean signal power, and the function Gamma is defined by, $\Gamma(m) = \int_0^\infty t^{m-1}exp(-t)dt$. For integer values of m, the distribution describes the summation of m orthogonal independent Rayleigh distributed random variables. That is, for N Rayleigh distributed random variables X_i, the density function of the random variable Y,

defined as,

$$Y = \sqrt{\sum_{i=1}^{N} X_i^2} \tag{5.12}$$

is given by a Nakagami distribution with $m = N$.

As special cases, Nakagami-m includes Rayleigh distribution when $m = 1$, and one sided Gaussian distribution for $m = 1/2$, when $m \to \infty$ the distribution becomes an impulse (no fading). This basically means that if $m < 1$, the Nakagami distributed fading is more severe than Rayleigh fading, and for values of $m > 1$ the fading circumstances are less severe. For values of $m > 1$, the Nakagami distribution closely approximates the Ricean distribution and the parameters m and the Ricean factor K which determines the severity of the Ricean fading, can be mapped via the equations [84]

$$K = \frac{\sqrt{m^2 - m}}{m - \sqrt{m^2 - m}} \quad , \qquad m > 1 \tag{5.13}$$

$$m = \frac{(K+1)^2}{(2K+1)} \tag{5.14}$$

$$\Omega = \mu^2 + 2\sigma^2 \tag{5.15}$$

So,

$$\begin{cases} \mu^2 = & \frac{K\Omega}{K+1} \\ 2\sigma^2 = & \frac{\Omega}{K+1} \end{cases} \tag{5.16}$$

Nakagami distribution is more appropriate to use in analytical expression than Rice distribution, because the Rice distribution equation consist of a Bessel function which is unreachable.

It is found in divers samples, that to describe the distribution of the measured fading of radio channels, the Nakagami distribution model is more convenient than Rice and Nakagami distributions models.

The channel models which have the signal amplitude resulting from a large number of multi path components, each with approximately the same amplitude, adding together, are frequently modeled by the Rayleigh distribution. It is described by the following density function,

$$p(x) = \frac{x}{\sigma^2} exp(-\frac{x^2}{2\sigma^2}) \ , \qquad x \geq 0 \tag{5.17}$$

where σ is the variance of the distribution. And the signal amplitude distribution which have multi path components with one dominant component emergent from a line of sight path between

the transmitter and the receiver, is described by the Rice distribution. It is described by the two parameter density function,

$$p(x) = \frac{x}{\sigma^2} exp(-\frac{(x^2+\mu^2)}{2\sigma^2}) I_0(\frac{x\mu}{\sigma^2}) \ , \qquad x \geq 0 \tag{5.18}$$

where μ is the mean of the distribution, σ its variance, and $I_0(.)$ the modified Bessel function of the first kind, order zero.

One additional examination is that the Nakagami distribution models the majority of measured and simulated results more closely than a Ricean distribution and the Rayleigh distribution. This involves that the hypothesis used to derive a Rayleigh or Ricean distributed channel are not accurate for indoor communications.

5.5 KL Techniques for Nakagami distribution

The main idea of this section is to exploit the distribution analysis techniques based on KLD and AIC [4] to calculate in blindly process the Received Signal density function modeled using Nakagami distribution for each BS and Analyze Akaike weight in order to determine the appropriate handover.

The notation used in this chapter are given in the Table 5.1 below.

Actually, the Received Signal for each BS is distributed according to an original density function f_k where $k \in \{1, 2, 3, 4, 5\}$ is the index of BS. Since only a finite number of observations is available, this function is usually unknown. To estimate the original density function, we use some observed data and an approximating model. We denote the approximating model as g_θ^k , where θ indicates the U-dimensional parameter vector, which specifies the density function.

In information theory [90], to compute the distinction between the two density functions f_k and g_θ^k we use KLD, given by [51],

$$D(f_k \ \ g_\theta^k) = E(Log(f_k(x))) - E(Log(g_\theta^k(x)))$$

$$D(f_k \ \ g_\theta^k) = -h_i(x) - \int_X f_k(x) Log(g_\theta^k(x)) dx \tag{5.19}$$

where f_k and g_θ^k are density functions defined over a set X, and f_k is absolutely continuous with respect to g_θ^k and h(.) denotes differential entropy. This distance measure is not directly applicable, because the original density function f_k is unidentified. It is known, however, that

Notation	Description
k	index of the Base Station
L	number of Base Stations
f_k	original density function of the received signal
g_θ^k	approximating model
θ	the dimensional parameter vector
D_{KL}	Kullback Leibler Distance
h(.)	differential entropy
N	number of observations
M	number of samples after compressive sampling
$x_1, \ldots, x_N$	independent observations
AIC_k	Akaike Information Criterion of Base Station k
U	the dimension of the parameter vector θ
L_k	Likelihood function
ϕ_k	AIC differences
w_k	Akaike weight
m	shape parameter
Ω	mean signal power
$\Gamma(m)$	function Gamma
K	Rician factor
$I_0(.)$	modified Bessel function of the first kind with order zero
μ	the mean or expectation of the distribution
σ	the standard deviation
C	channel capacity
H	entropy
P	power constraint
N	noise variance
γ	SNR
P_{out}	outage probability
P_{FA}	probability of False Alarm

Table 5.1 – Notation and TerminoLogy

the KLD is positive, this implies that,

$$-\int_X f_k(x) Log(g_\theta^k(x))dx = h_i(x) + D(f_k \;\; g_\theta^k) \tag{5.20}$$

approaches the differential entropy of X from above for increasing quality of the model g_θ^k .

This expression (5.20) can be approximated by applying the weak law of large numbers and averaging the Log-likelihood values given the model over N independent observations $x_1, x_2, \ldots, x_N$ according to,

$$-\int_X f_k(x) Log(g_\theta^k(x))dx \approx -\frac{1}{N}\sum_{n=1}^{N} g_\theta^k(x_n) \tag{5.21}$$

The expected KLD is given by,

$$-E_\theta\left(\int_X f_k(x) Log(g_\theta^k(x))dx\right) \tag{5.22}$$

This expression (5.22) cannot be computed, but estimated.

The model selection in information theoretic criteria was presented by Akaike. Considering a candidate model, the concept is to decide if the distribution of the observed signal match with the candidate model. The AIC criterion is an unbiased approximation estimator for (5.22), defined as,

$$AIC_k = -2\sum_{n=1}^{N} Log(g_\theta^k(x_n)) + 2U \tag{5.23}$$

where U indicates the dimension of the parameter vector θ.

In our approach, we have to select the closest estimation to the unknown reality that generated the data, which have the smallest AIC, from among the candidate models considered.

The parameter vector θ for each family need an estimation using the minimum discrepancy estimator $\widehat{\theta}$, which minimizes the empirical discrepancy. This is the difference between the approximating model and the original model. The maximum likelihood estimator [91] is the minimum discrepancy estimator for the KLD.

Because AIC contains various constants, it is recommend to compute the AIC differences,

$$\phi_k = AIC_k - AIC_{min} \tag{5.24}$$

where AIC_{min} denotes the minimum AIC value over all BSs.

Akaike weights can be computed using (5.23), with the intention of providing another measure of the strength of evidence for this model, and is given by,

$$W_k = \frac{e^{-1/2\phi_k}}{\sum_{i=1}^{5} e^{-1/2\phi_i}} \ , \text{ where } k \in \{1, 2, 3, 4, 5\} \tag{5.25}$$

The Akaike weights allow us not only to decide if the distribution of the Received Signal fits the Gaussian distribution, but also provide information about the relative approximation quality of this distribution.

In our problem, we model our signal between the BS and the users by a Nakagami distribution. So the density function for the Received Signal for each BS is given by the equation,

$$g(x) = \frac{2m^m}{\Gamma(m)\Omega^m} x^{2m-1} exp(-\frac{m}{\Omega}x^2) \ , \quad x \geq 0 \quad m \geq 1/2 \tag{5.26}$$

5.5.1 Maximum Likelihood Estimator of the parameters

Consider a density function with unidentified parameter θ, associated with either a known density function, denoted as f_θ^k. As a function of θ with $x_1, x_2, \ldots, x_N$ fixed, the likelihood function is,

$$L_k(\theta) = f_\theta^k(x_1, x_2, \ldots, x_N) \tag{5.27}$$

The method of maximum likelihood estimates θ by calculating the value of θ that maximizes $L_k(\theta)$. The maximum likelihood estimator (MLE) [91] of θ is given by,

$$\widehat{\theta} = \underset{\theta}{\text{argmax}}\ L_k(\theta) \tag{5.28}$$

Generally, one assumes that the data drawn from a particular distribution are Independent and identically distributed (iid) with unknown parameters. This considerably simplifies the problem because the likelihood can then be written as a product of N univariate densities function,

$$L_k(\theta) = \prod_{n=1}^{N} f_k(x_n|\theta) \tag{5.29}$$

and since maxima are unaffected by monotone transformations, one can take the Logarithm of this expression to turn it into a sum,

$$L_k^*(\theta) = \sum_{n=1}^{N} Log f_k(x_n|\theta) \tag{5.30}$$

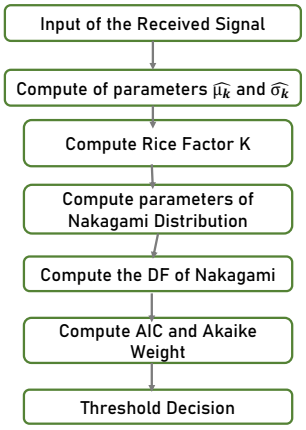

Figure 5.1 – Flowchart of our approach on Nakagami distribution signal

Consequently, the expression of the maximum likelihood in our case is,

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}} \frac{1}{N} \sum_{n=1}^{N} Log(g_{\theta}^{k}(x_n)) \tag{5.31}$$

The maximum of this expression may then be found numerically using varied optimization algorithms. This contrasts with seeking an unbiased estimator of θ, which may not necessarily yield the MLE but which will yield a value that (on average) will neither tend to over-estimate nor under-estimate the true value of θ.

To compute the MLE, we will use a work already done about the estimation of the parameter of Nakagami distribution [92][93].

5.5.2 DAD on Nakagami Distribution Signal

In this section, we present the approach to detect the best handover based on exploiting model selection approach, especially AIC in a Nakagami distribution signal.

The sequential diagram of the proposed algorithm is shown in 5.1, which can be implemented in seven steps:

We consider that the initial signal can be modeled using Gaussian distribution and its norm can be modeled using Nakagami distribution. Following the input of the values of the Received Signal for each BS (observations), in the first step we compute the Rice distribution parameters [4], then using the relations between Rice and Nakagami distribution to compute Rice Factor K and shape parameter m , then the pdf for the Received Signal for each BS k.

Once we get g_{θ}^{k}, we calculate AIC_k and W_k for each BS.

If the Akaike weight of Nakagami distribution of the BS_k is higher than the Akaike weights of other BSs, then there is no Handover, and if the Akaike weight of BS_k is lower than the Akaike weight of BS_i where $i \in \{1, 2, 3, 4, 5\}$ then there is Handover from BS_k to BS_i,

$$\lambda_{threshold}(x_n) = \begin{cases} W_k - W_i < \lambda_{threshold} & \text{Handover } (H_0) \\ W_k - W_i \geq \lambda_{threshold} & \text{No Handover } (H_1) \end{cases} \tag{5.32}$$

To compute the decision threshold we use the equation of the probability of false alarm P_{FA} [75], so, the threshold $\lambda_{threshold}$ for a given false alarm probability [75] is determined by solving the equation,

$$P_{FA} = P(\lambda_{threshold}(x) < \lambda_{threshold}|H_1) \tag{5.33}$$

5.6 Channel Capacity and Outage Probability

The Handover between base stations and User effect the capacity of the network and there is a correlation between the cell-quality and network performance. So, in these section, we will estimate the network performance in term of Channel capacity and Outage probability for two signal distribution models, Rice and Nakagami.

5.6.1 Rice Distribution

First, let us give an entrance on Channel Capacity. In Communication theory, it is generally expected that the transmitted signals are ruined by some noise. To see how much information is possible to transmit over the channel, we maximize the Mutual Information between the transmitted variable X and the received variable Y, with the condition that the power is limited by P. Without the power constraint in the definition we would be disposed to select as many signal alternatives as far apart as we like. Then we would be capable to transmit as much information as we like in a single channel use. With the power constraint we obtain a more realistic system where we need to find other mechanisms that growing the power to get a greater information throughput over the channel.

So, the first definition of the "information" channel capacity of a discrete channel is,

$$C = \max_{E(X^2) \leq P} I(X, Y) \tag{5.34}$$

The Mutual Information can be expressed as,

$$I(X, Y) = H(Y) - H(Y|X) \tag{5.35}$$

where $H(Y)$ is the marginal entropy and $H(Y|X)$ is the conditional entropy.

The most common Channel Model, which is the so-called Gaussian Channel, which can be presented as,

$$Y_i = X_i + Z_i, Z_i \sim \mathcal{N}(0, N) \tag{5.36}$$

This is a time-discrete channel with output Y_i at time i, where Y_i is the sum of the input X_i and the noise Z_i. The noise Z_i is drawn i.i.d. from a Gaussian distribution with variance N. The noise Z_i is assumed to be independent of the signal X_i.

So now, we can calculate the Mutual Information as follows,

$$\begin{aligned} I(X, Y) &= H(Y) - H(Y|X) && (5.37)\\ &= H(Y) - H(X + Z|X) && (5.38)\\ &= H(Y) - H(Z|X) && (5.39)\\ &= H(Y) - H(Z) && (5.40) \end{aligned}$$

since Z is independent of X, where h is the marginal entropy.

In statistics, for a given $X \sim \mathcal{N}(\sigma, \mu^2)$, the entropy is defined as,

$$\begin{aligned} H &= E(-Log(f(x))) && (5.41)\\ &= \int_{-\infty}^{+\infty} f(x)(-Log(f(x)))dx && (5.42)\\ &= \int_{-\infty}^{+\infty} -Log((2\pi\sigma^2)^{-\frac{1}{2}})f(x)dx + \int_{-\infty}^{+\infty} \frac{1}{2\sigma^2}(x-\mu)^2 f(x)dx && (5.43)\\ &= \frac{1}{2}Log(2\pi\sigma^2)\int_{-\infty}^{+\infty} f(x)dx + \frac{1}{2\sigma^2}\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx && (5.44)\\ &= \frac{1}{2}(Log(2\pi\sigma^2) + 1) && (5.45)\\ H &= \frac{1}{2}Log(2\pi\sigma^2 e) && (5.46) \end{aligned}$$

Applying this result to bound the mutual information, we obtain,

$$\max_{E(X^2)\leq P} I(X,Y) = maxH(Y) - H(Z) \tag{5.47}$$

$$= \frac{1}{2}Log(2\pi e(P+N)) - \frac{1}{2}Log(2\pi e(N)) \tag{5.48}$$

$$= \frac{1}{2}Log(\frac{2\pi e(P+N)}{2\pi eN}) \tag{5.49}$$

$$= \frac{1}{2}Log(1+\frac{P}{N}) \tag{5.50}$$

Definition

The Channel Capacity of a Gaussian Channel with Power Constraint P, and Noise variance N,

$$C = \frac{1}{2}Log(1+\frac{P}{N}) \tag{5.51}$$

5.6.2 Nakagami Distribution

In this part we want to estimate the performance of the Nakagami distribution signal in terms of Channel Capacity and Outage Probability, that are often based on concepts from Information Theory. The Nakagami fading model is described in Section II and the channel capacity under Nakagami fading is then investigated in terms of its density function in this section.

In a Wireless communication, when a group of channels is active at the same time, the interference from the other channels is considered as noise, which mean the presence of Interference boundary.

To examine how much information is possible to transmit over the channel, we maximize the Mutual Information between the transmitted variable X and the received variable Y, with the condition that the power is limited by P . The Channel Capacity formula is given by,

$$C = \sum_{0}^{\infty} Log(1+\gamma)P(\gamma) \tag{5.52}$$

where γ is the SNR.

The outage probability is another desired QoS parameter of communication systems. The outage happens when the SNR becomes lower than the allowed threshold level. In the proposed system model, the outage occurs when all multiple-relay links between transmitters and receivers of SU are damaged, ie, the SNR corresponding to each link must be less than the threshold.

The protection for the channel between user and desired BS must be guaranteed in a Cellular Network. This protection is guaranteed if the sum of all other BSs Transmitters' powers is not greater than the Interference constraint P_T . Then, the desired BS verifies the Outage probability constraint.

For a channel k, no data is sent when $C_k < T_k$ where T_k is the transmitted data rate, then, the channel maybe breakdown with an outage probability P_{out}, which is well known as a performance metric in fading channels, equal to probability of not being able to successfully send a signal on a channel, defined by the probability that the capacity of the user is under the transmitted code rate, given by,

$$P_{out} = P(C_k < T_k) \leq P_{outmax}, \forall k = 1, ..., L \tag{5.53}$$

where P_{outmax} is the Maximum Outage Probability.

5.7 Study Cases

For the numerical application of the approach presented above, we use the software package Matlab R2019a. Each simulation setup is running several times in order to smooth up the results.

We summarize the Estimated parameters for each BS_i and their Akaike Weights in the figure 5.2.

The complexity of Handover detection is a major concern in Handover management. Using the implementation steps of the DAD and the DAD with Compressive Sampling, we will study the complexity required for each detector to derive their Handover algorithm.

The complexity of the algorithm is measured through the number of complex multiplications that the algorithms has to perform for the calculation of the test statistics. We summarize the

Index of BS_i	$\hat{\mu}_i$	$\hat{\sigma}_i$	Akaike Weight
1	5.7302	1.6420	0
2	2.7013	3.2151	0
3	5.8207	1.6942	0
4	0	2.7821	0
5	6.3030	2.5196	0.05
6	0	1.3727	0
7	5.5804	1.9369	0
8	6.7584	2.5385	0.7
9	6.2326	2.2945	0.05
10	7.1991	2.7916	0.2

Figure 5.2 – Estimated parameters for each BS_i and their Akaike Weights

Handover Detection Technique	Complexity
Distribution Analysis Detector	2N
DAD + Compressive Sampling	2M

Table 5.2 – Complexity Comparison of the two Handovers Detection Techniques

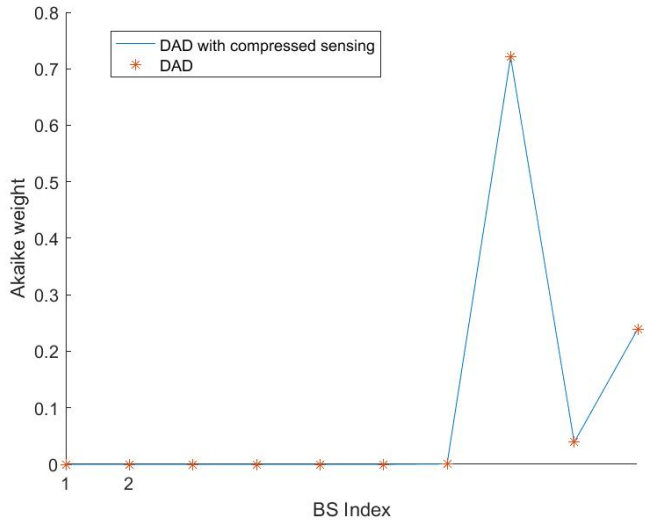

Figure 5.3 – Akaike Weight vs BS index

number of multiplications required for each technique in Table 5.2. Note that N refers to the number of samples of the received signal and M is the number of samples after compressive sampling. From these results, we find that compressive sampling decreases the complexity as compared to the simple Distribution Analysis Detector.

In term to select the best BS for the user, we apply the approach in Figure 5.1, we compute the AW for the BSs with Nakagami distribution of the signal between the user and the BSs. Figure 5.3 illustrates the AW with Nakagami distribution of 10 BSs. From the figure we can see that the BS which has the maximum Akaike weight is the BS 8, so the best BS for the user is the BS 8.

Figure 5.3 depicts also the comparison between two detectors. This figure shows the Akaike Weight for 10 Base Stations calculated with two techniques DAD and DAD with Compressive Sampling. From the numerical application results, we show that the two detectors give the same Akaike weight which means that the use of Compressive Sampling do not change the final result.

After that, we consider a cellular network with a user and $L = 10$ BSs, trying to communicate at the same time as a transmission, subject to mutual interference. For the numerical application, we take, $P_T = 50dBm$ (Power constraint).

We summarize the Estimated parameters for each BS_i and the Channel Capacity & Outage

BS_i	$\hat{\mu}_i$	$\hat{\sigma}_i$	C_i	P_{out}
1	6.3030	1.9866	**0.792481255**	$0.9\ 10^{-4}$
2	4.7226	1.1153	**0.79248126**	$2\ 10^{-4}$
3	5.5804	1.9369	**0.792481255**	$1.4\ 10^{-4}$
4	6.5761	2.5385	**0.79248126**	10^{-4}
5	6.2326	2.2945	**0.792481200**	$0.8\ 10^{-4}$
6	7.1991	2.7916	**0.792481175**	$0.77\ 10^{-4}$
7	5.5804	1.9369	**0.79248126**	$0.7\ 10^{-4}$
8	6.7584	2.5385	**0.792481265**	$0.5\ 10^{-4}$
9	6.2326	2.2945	**0.79248126**	$0.2\ 10^{-4}$
10	7.1991	2.7916	**0.792481277**	$0.4\ 10^{-4}$

Figure 5.4 – Estimated parameters for each BS_i and the Channel Capacity & Outage probability for Rice distribution model

BS_i	$\hat{\mu}_i$	$\hat{\sigma}_i$	C_i	P_{out}
1	6.3030	1.9866	**0.792481200**	$0.62\ 10^{-4}$
2	4.7226	1.1153	**0.792481199**	$0.42\ 10^{-4}$
3	5.5804	1.9369	**0.792481175**	$0.8\ 10^{-4}$
4	6.5761	2.5385	**0.792481275**	$0.4\ 10^{-4}$
5	6.2326	2.2945	**0.792481127**	$0.8\ 10^{-4}$
6	7.1991	2.7916	**0.792481127**	$0.3\ 10^{-4}$
7	5.5804	1.9369	**0.792481129**	$0.4\ 10^{-4}$
8	6.7584	2.5385	**0.79248128**	$0.3\ 10^{-4}$
9	6.2326	2.2945	**0.79248128**	$0.4\ 10^{-4}$
10	7.1991	2.7916	**0.7924811295**	$0.3\ 10^{-4}$

Figure 5.5 – Estimated parameters for each BS_i and the Channel Capacity & Outage probability for Nakagami distribution model

probability for Rice and Nakagami distribution models in the tables 5.4 5.5.

Firstly, to estimate the performance of networks, we compute the estimation of channel capacity and outage probability for Rice and Nakagami distribution signals between the user and BSs. Figure 5.6 and Figure 5.7 show the behavior of estimated values for 10 Channels, we can see that when the channel capacity increases, the Outage probability decreases and vice versa, which confirm the correlation between the channel capacity and outage probability that we have mentioned.

After that, we compare the performance of the two models, in term of channel capacity Figure 5.6, we can see that the channel capacity for Nakagami distribution model have greater values than Rice distribution model, which is normal, because Nakagami channels can be seen as multiple Rice channels, and in term of outage probability Figure 5.7, we can see that the outage probability of Nakagami distribution model is lower than the Rice , so, we can conclude that the Nakagami channels are more efficient than Rice channels.

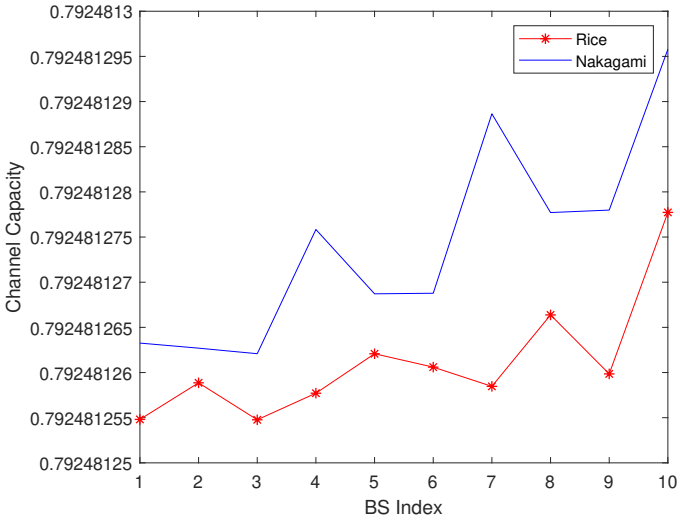

Figure 5.6 – Channel Capacity of Rice and Nakagami Distribution Channels (bits/s/Hz)

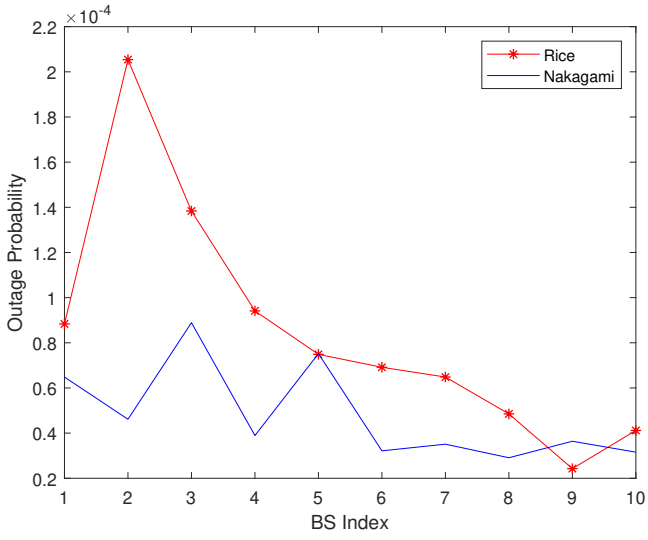

Figure 5.7 – Outage probability of Rice and Nakagami Distribution Channels

5.8 Conclusion

In this chapter, we applied a novel approach to manage handovers in a Nakagami distribution signal. This approach is based on analyzing the DF of the received signal between the BS and the user using KLD, AIC and Akaike Weight, to pick the suitable handover for each user. After that we combine this approach with compressive sampling in order to make the detection possible with less number of samples. The analysis of the complexity of the proposed technique shows that it can be reduced. The numerical comparison at different sampling rates shows that the new designed scheme achieves the same performance as the DAD while preserving a low computational complexity.

We also presented a new approach to estimate the performance of network channel, just by analyzing received signal DF and a comparison between Nakagami and Rice distribution models. In the first step, we computed an estimation of channel capacity and the outage probability for Rice distribution model and Nakagami distribution model. After that, we realized numerical applications of the proposed estimation of channel capacity and outage probability in order to show that Nakagami distribution model is better than Rice distribution model.

Conclusion

Summary of research contributions

The importance of Information and Communication Technology has increased significantly over the last few years. The continuous growth of mobile data traffic and the rise of 5G Technology concept have raised new challenges. In this thesis we have addressed the Handover Management challenge. The goal is to investigate a new technique to manage the handovers based on some analytical approaches and estimate the performance of the network.

In our first contribution, we have proposed a new approach to manage the handovers between a number of users and the base stations of small cells. Our idea has been based on analyzing the density function of the Received Signal for each BS (DAD), to provide an indication of the intensity of the Received Signal, and exploit KL Distance, Akaike Information Criterion and Akaike Weight in order to decide the best handover and the best BS for each user. The proposed Blind Detection Approach is evaluated using the software package Matlab R2018a.

In our second contribution, we have combined the compressive sampling and the DAD to detect Handovers. In the first step, we designed a sensing matrix that keeps the linear properties of the sampled primary signal. Then, we applied the compressed measurements on the DAD, in order to make the detection possible with a smaller number of samples or smaller sampling rate. The numerical analysis of the complexity of the proposed approach shows that it achieves the same performance as the DAD while preserving a low computational complexity.

In the last contribution, we have applied the approach of DAD with the Compressive Sampling in a Nakagami distribution signal. We also presented a new approach to estimate the network channel performances, just by analyzing the received signal DF and we compared between Nakagami and Rice distribution models. In the first step, we computed a channel capacity estimation and the outage probability for Rice distribution model and Nakagami distribution model. After that, we did numerical evaluations of the proposed channel capacity estimation and the outage probability in order to show that Nakagami distribution model is better than Rice distribution model.

Future works

As perspectives of our research work, we believe that it still has many challenges to be addressed, among them we can give the following ones:

- Test the proposed approaches by simulation and implementing them in real networks.

- Estimate other measures of performances of the effectiveness of the Handover management approach such as average throughput, delay, Bit error rate and Packet loss ratio.
- Design of the sensing matrix: Doing a good design of the sensing matrix for the CS in order to avoid the loss of the important information by reducing the dimensionality after the signal compression. The CS reconstruction performance can be substantially improved by optimizing the choice of the sensing matrix.
- Close expressions for detection performance: It is important to find the best threshold for the detector. For a good detection algorithm, the probability of detection should be high and the probability of false alarm should be low, which usually depends on the choice of the threshold. It would be interesting to find close expressions of the probability distributions of each detection hypothesis.
- Explore if there is a possibility to use some Artificial Intelligence Algorithms in our prosposed approaches.
- Compare our approach with the classical approaches.
- Apply our approach on Vertical Handovers.

Thesis publications

Publications in journals

- **Adnane EL HANJRI**, Aawatif HAYAR and Abdelkrim HAQIQ, ***"Blind Handover detection based on KLD and Channel Capacity, Outage Probability Estimation for Rice and Nakagami Models"***.IAENG International Journal of Computer Science, vol. 48, no. 4, pp1087-1094, 2021.
- **Adnane EL HANJRI**, Aawatif HAYAR and Abdelkrim HAQIQ, ***"Features detection based blind handover using kullback leibler distance for 5G HetNets systems"***. IAES International Journal of Artificial Intelligence (IJ-AI), Vol. 9, No. 2, pp. 193-202, June 2020.

Participations in international conferences

- **Adnane EL HANJRI**, Aawatif HAYAR and Abdelkrim HAQIQ, ***"A Novel Handover Management Evaluation based on Channel Capacity and Outage Probability"***. In: The 1st Online Arab Post Graduate Students Conference in Emerging Technologies and Applications (June 21-23, 2020)
- Amine ERROUTBI, **Adnane EL HANJRI** and Abderrahim SEKKAKI, ***"Secure and Lightweight HMAC Mutual Authentication Protocol for Communication between IoT Devices and Fog Nodes"***. In: The 5th IEEE International Smart Cities Conference (ISC2), IEEE, 2019.
- **Adnane EL HANJRI**, Aawatif HAYAR and Abdelkrim HAQIQ, ***"Combined Compressive Sampling Techniques and Features Detection using Kullback Leibler Distance to Manage Handovers"***. In: The 5th IEEE International Smart Cities Conference (ISC2), IEEE, 2019.
- **Adnane EL HANJRI**, Abdellah ZAALOUL and Abdelkrim HAQIQ, ***"Analytic approach using Continuous Markov chain to improve the QoS of a wireless network"***. In: Abraham A., Haqiq A., Muda A., Gandhi N. (eds) Innovations in Bio-Inspired Computing and Applications. IBICA 2017. Advances in Intelligent Systems and Computing, vol 735. Springer, Cham.

Bibliography

[1] United Nations Department of Economic and Social Affairs. World population prospects. In *United Nations Publications*, 2015. 2, 10

[2] Buhaug and Urdal. An urbanization bomb? population growth and social disorder in cities. *Global Environmental Change*, pages 1–10, 2013. 2, 10

[3] A. El Hanjri, A. Hayar, and A. Haqiq. Features detection based blind handover using kullback leibler distance for 5g hetnets systems. In *IAES International Journal of Artificial Intelligence*, pages 193–202, 2020. 3, 58

[4] A. El Hanjri, A. Hayar, and A. Haqiq. Combined compressive sampling techniques and features detection using kullback leibler distance to manage handovers. *IEEE International Smart Cities Conference (ISC2)*, pages 504–507, 2019. 3, 88, 93, 97

[5] A. El Hanjri, A. Hayar, and A. Haqiq. Blind handover detection based on kld and channel capacity, outage probability estimation for rice and nakagami models. *IAENG International Journal of Computer Science*, 48:1087–1094, 2021. 4, 88

[6] D. Pinedal and C. Hernandez. Cognitive radio for tvws usage. In *TELKOMNIKA*, pages 2735–2746. IEEE, 2019. 6, 87

[7] Cisco. Cisco visual networking index: Global mobile data traffic forecast update, 2017–2022. *White paper*, 2019. 6

[8] A. Gupta and R. K. Jha. A survey of 5g networks: Architecture and emerging technologies. In *IEEE Access*, pages 1206–1232. IEEE, 2015. 6, 9, 57, 87

[9] C. K. Agubor, I. Akwukwuegbu, M. Olubiwe, Nosiri, Ehinomen, Olukunle, Okozi, Ezema, and Okeke. A comprehensive review on the feasibility and challenges of millimeter wave in emerging 5g mobile communication. In *Advances in Science, Technology and Engineering Systems Journal*, pages 138–144, 2019. 6, 9, 57, 87

[10] J. Hoydis, M. Kobayashi, and M. Debbah. Green small-cell networks. In *IEEE Vehicular Technology Magazine*, pages 37–43, 2011. 6, 57, 71, 87

[11] A. Damnjanovic, J. Montojo, Y. Wei, T. Luo, Vajapeyam, Yoo, Song, and Malladi. A survey on 3gpp heterogeneous networks. In *IEEE Wireless Communications*, 2011. 7, 8, 57

[12] R. Ahamad, E. Sundavarajan, and N. Ismail. Handover in lte- advanced wireless networks: state of art and survey of decision algorithm. In *Springer Science + Business Media*, 2017. 7, 57, 71, 87

[13] I. Akyildiz, J. McNair, Ho, J. Uzunalioglu, and H. Wang. Mobility management in next-generation wireless systems. In *Proceedings of the IEEE*, pages 1347–1384, 1999. 7, 57, 71, 87

[14] M. Ekpenyong, D. Asuquo, S. Robinson, I. Umoren, and E. Isong. Soft handoff evaluation and efficient access network selection in next generation cellular systems. In *Advances in Science, Technology and Engineering Systems Journal*, pages 1616–1625, 2017. 7, 57, 71, 87

[15] Yuzhe Zhou and Bo Ai. Handover schemes and algorithms of high-speed mobile environment: A survey. In *Comput. Commun.*, 2014. 7

[16] Richard Heeks. Information and communication technologies, poverty and development. In *Development Informatics Working Paper*, 1999. 8

[17] Mudit Ratana Bhalla and Anand Vardhan Bhalla. Generations of mobile wireless technology: A survey. In *International Journal of Computer Applications*, 2010. 8

[18] W. R. Young. Advanced mobile phone service: Introduction, background, and objectives. In *Bell System Technical Journal*, 1979. 8

[19] J. Lehenkari and R. Miettinen. Standardisation in the construction of a large technological system—the case of the nordic mobile telephone system. In *Telecommunications Policy*, pages 109–127, 2002. 8

[20] Altamura, Toscano, Gentilucci, Ammirati, Castro, Pandozi, and Satini. Influence of digital and analogue cellular telephones on implanted pacemakers. In *European Heart Journal*, page 1632–1641, 1997. 8

[21] M. Rahnema. Overview of the gsm system and protocol architecture. In *IEEE Communications Magazine*, pages 92–100, 1993. 8

[22] P. Datta and S. Kaushal. Exploration and comparison of different 4g technologies implementations: A survey. *Recent Advances in Engineering and Computational Sciences (RAECS)*, 2014. 9

[23] A. Aswal A. Kumar and L. Singh. 4g wireless technology: A brief review. *International Journal of Engineering and Management Research*, 2013. 9

[24] R. Arshad, H. ElSawy, S. Sorour, T. Al-Naffouri, and M. Alouini. Handover management in 5g and beyond: A topology aware skipping approach. In *IEEE Access*, page 9073–9081, 2016. 9, 28

[25] T. K. Sawanobori. The next generation of wireless: 5g leadership in the u.s. In *Technical Report, CTIA Everything Wireless: Washington, DC, USA*, 2016. 9, 12

[26] M. Eremia, L. Toma, and M. Sanduleac. The smart city concept in the 21st century. *Procedia Engineering*, 181:12–19, 2017. 10

[27] V. Pinto C. Gargiulo and F. Zucaro. Eu smart city governance. In *Journal of Land Use, Mobility and Environment*, 2013. 11

[28] S. Eldridge K. Rose and L. Chapin. The internet of things: An overview understanding the issues and challenges of a more connected world. In *The Internet Society (ISOC)*, 2015. 12

[29] F. Appio, M. Lima, and S. Paroutis. Understanding smart cities: Innovation ecosystems, technological advancements, and societal challenges. *Technol. Forecast. Soc. Chang*, pages 1–14, 2019. 12

[30] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani. Internet-of-things-based smart cities: Recent advances and challenges. *IEEE Commun. Mag*, pages 16–24, 2017. 12

[31] S. Zhang, J. Rong, and B. Wang. A privacy protection scheme of smart meter for decentralized smart home environment based on consortium blockchain. *Int. J. Electr. Power Energy Syst*, 2020. 14

[32] S. Nikam and R. Ingle. Criminal detection system in iot. In *Int. J. Control Theory Appl*, pages 11–16, 2017. 15

[33] P. Ray, M. Mukherjee, and L. Shu. Internet of things for disaster management: State-of-the-art and prospects. In *IEEE Access*, 2017. 16

[34] H. Ishii, Y. Kishiyama, and H. Takahashi. A novel architecture for lte-b: C-plane/u-plane split and phantom cell concept. In *IEEE GLOBECOM Wksps*, 2012. 28

[35] H. Zhang, X. Wen, B. Wang, W. Zheng, and Y. Sun. A novel handover mechanism between femtocell and macrocell for lte based networks. In *IEEE 2nd Int'l. Conf. Commun. Software and Networks*, pages 228–231, 2010. 28, 30

[36] J. Moon, J. Jung, S. Lee, A. Nigam, and S. Ryoo. On the trade-off between handover failure and small cell utilization in heterogeneous networks. *IEEE ICC Wksp*, pages 2282–2287, 2015. 28, 30

[37] H. Leem, J. Kim, D. Sung, Y. Yi, and B. Kim. A novel handover scheme to support small-cell users in a hetnet environment. *IEEE Wireless Commun. and Networking Conf*, page 1978–1983, 2015. 29, 30

[38] D. López-Pérez, I. Guvenc, and X. Chu. Theoretical analysis of handover failure and ping-pong rates for heterogeneous networks. pages 6774–6779, 2012. Proc. 2012 IEEE ICC. 29, 30

[39] 3GPP. Overview of 3gpp release 12. 2015, lien: https://www.3gpp.org/specifications/releases/68-release-12: :text=The 20Mobile,Standardtext=Other 20important 20features 20completed 20in,in 20to 20mobile 20operator's 20offerings. (last visit: 11-11-2021). 29

[40] Uwe Hassler. *Stochastic Processes and Calculus*. Springer Texts in Business and Economics, 2016. 32, 38

[41] Achim Klenke. *Probability Theory*. Springer Universitext, 2014. 32

[42] Jim Pitman. *Probability*. Springer Texts in Statistics, 1993. 32

[43] O. Barndorff-Nielson. On a formula for the distribution of the maximum likelihood estimator. 1983. Biometrika. 40

[44] Thomas Cover and Joy. *Elements of Information Theory*. John Wiley & Sons, Inc, 2006. 41, 87

[45] Raymond W. Yeung. *Information Theory and Network Coding*. Springer, 2008. 41

[46] C. E. Shannon. A mathematical theory of communication. pages 379–423, 1948. The Bell System Technical Journal. 41

[47] Ramirez, Via, Santamaria, and Crespo. Entropy and kullback-leibler divergence estimation based on szego's theorem. 2009. 17th European Signal Processing Conference (EUSIPCO). 44

[48] S. Verdu and Te Sun Han. A general formula for channel capacity. 1994. IEEE Transactions on Information Theory. 46

[49] Kullback and Leibler. On information and sufficiency. 1951. The Annals of Mathematical Statistics. 47

[50] Dmitry Belov and Ronald Armstrong. Distributions of the kullback–leibler divergence with applications. 2011. British Journal of Mathematical and Statistical Psychology. 47

[51] H. Akaike. Information theory and an extension of the maximum likelihood principle. 1973. 2nd International Symposium on Information Theory. 51, 52, 58, 62, 63, 64, 66, 93

[52] Hamparsum Bozdogan. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. pages 345–370, 1987. Psychometrika. 51, 52, 58

[53] K. Takeuchi. Distribution of informational statistics and a criterion of model fitting. *Suri-Kagaku (Mathematical Sciences)*, 1976. 52

[54] W. Pan. Akaike's information criterion in generalized estimating equations. *Biometrics*, 2001. 52

[55] Hurvich and Tsai. Regression and time series model selection in small samples. *Biometrika*, 1989. 52

[56] Broersen. Finite sample criteria for auto regressive order selection. *IEEE Trans. Signal Process*, 2000. 52

[57] G. Schwarz. Estimating the dimension of a model. *Ann. Statist*, 1978. 52

[58] Andrew A. Neath and Joseph E. Cavanaugh. The bayesian information criterion: background, derivation, and applications. pages 199–203, 2011. wires computational statistics. 53

[59] Eric-Jan Wagenmakers and Simon Farrell. Aic model selection using akaike weights. pages 192–196, 2004. Psychonomic Bulletin Review. 54, 58

[60] Li, Zhang, Duan, Jia, , and Liu. Performance evaluation and analysis on group mobility of mobile relay for lte advanced system. 2012. IEEE Vehicular Technology Conference (VTC Fall). 58

[61] Zhang and Wang. Group-based collaborative handover in lte macro-femtocell network. 2014. 10th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2014). 58

[62] Zhao, Huang, Zhang, and Fang. Handoff for wireless networks with mobile relay stations. 2011. IEEE Wireless Communications and Networking Conference. 58

[63] Luo, Zhang, and Fang. A comp soft handover scheme for lte systems in high speed railway. 2012. EURASIP Journal on Wireless Communications and Networking. 58

[64] Pan, Chen, and Lin. An enhanced handover scheme for mobile relays in lte-a high-speed rail networks. 2014. IEEE Transactions on Vehicular Technology. 58

[65] Van Quang, Prasad, and Niemegeers. A survey on handoffs—lessons for 60 ghz based wireless systems. 2010. IEEE Communications surveys and tutorials. 59

[66] Bing, He, and Jiang. Performance analysis of vertical handover in a umtswlan integrated network. 2003. 14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications. 59

[67] J. Lv, Y. Ma, and S. Yoshizawa. Intelligent seamless vertical handoff algorithm for the next generation wireless networks. 2008. Proceedings of the 1st international conference on MOBILe Wireless MiddleWARE, Operating Systems, and Applications. 59

[68] Yang, Gondal, Qiu, and Dooley. Combined sinr based vertical handoff algorithm for next generation heterogeneous wireless networks. 2007. IEEE GLOBECOM 2007-IEEE Global Telecommunications Conference. 59

[69] Zhu and McNair. Multiservice vertical handoff decision algorithms. 2006. Journal on Wireless Communications and Networking. 59

[70] Chou and Huang. Dynamic vertical handover control algorithm for wlan and umts. 2006. IEEE Wireless Communications and Networking Conference. 59

[71] N. Mouhassine? M. Moughit and F. Laassiri. Improving the quality of service of voice over ip in wireless sensor networks by centralizing handover management and authentication using the sdn controller. Third International Conference on Intelligent Computing in Data Sciences (ICDS). 59

[72] M. Pischella. Blind handover technique. In *Patent NO: US 7, 224, 972 B2*, 2007. 59

[73] B. Zayen, A. Hayar, and D. Nussbaum. Blind spectrum sensing for cognitive radio based on model selection. In *CrownCom08, 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, Singapore*, pages 15–17, 2008. 60, 64, 81

[74] H. Holma and A. Toskala. Wcdma for umts, radio access for third generation mobile communication. In *John Wiley and Sons, Inc*, 2000. 64

[75] B. Zayen, A. Hayar, and K. Kansanen. Blind spectrum sensing for cognitive radio based on signal space dimension estimation. In *ICC09, IEEE International Conference on Communications*, pages 14–18. IEEE, 2009. 66, 98

[76] E. Candes and M. Wakin. An introduction to compressive sampling. page 2130, 2008. IEEE Signal Processing Magazine. 71, 73, 78, 79, 90

[77] E. J. Cands. Compressive sampling. 2006. Proceedings of the International Congress of Mathematicians, (Madrid, Spain). 71, 73

[78] E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *arXiv*, 2005. 71, 79, 90

[79] E. J. Candes and T. Tao. Decoding by linear programming. *math/0502327*, 2005. 71

[80] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 2006. 71

[81] D. Donoho. Compressed sensing. *IEEE Transactions*, pages 1289–1306, 2006. 79, 90

[82] W. Guibene, H. Moussavinik, and A. Hayar. Combined compressive sampling and distribution discontinuities detection approach to wideband spectrum sensing for cognitive radios. *Journal of Communications and Networks*, 2010. 79, 90

[83] Do, Le, Nguyen, Kieu, and Nguyen. Wireless power transfer enabled noma relay systems: two sic modes and performance evaluation. *TELKOMNIKA*, 17(6):2697–2703, 2019. 87

[84] M. Nakagami. The m-distribution a general formula of intensity distribution of rapid fading. *W. G. Hoffman, editor, Statistical Methods in Radio Wave Propagation. Oxford, U.K.: Pergamon*, 1960. 88, 91, 92, 120

[85] Y. Wang and H. Haas. Dynamic load balancing with handover in hybrid li-fi and wi-fi networks. *Journal of Lightwave Technology*, pages 4671–4682, 2015. 88

[86] Chang and Chen. Cross-layer-based adaptive vertical handoff with predictive rss in heterogeneous wireless networks. *IEEE Transactions on Vehicular Technology*, pages 3679–3692, 2008. 88

[87] Yan, Mani, and Sekercioglu. A traveling distance prediction based method to minimize unnecessary handovers from cellular networks to wlans. *IEEE Communications Letters*, pages 14–16, 2008. 89

[88] Ben-Mubarak, Ali, Noordin, and al. Fuzzy logic based self-adaptive handover algorithm for mobile wimax. *Wireless Personal Communications*, pages 1421–1442, 2013. 89

[89] Pischella. Blind handover technique. *Patent NO: US 7*, 2007. 89

[90] Jaynes. Information theory and statistical mechanics. In *Physical Review*, 1957. 93

[91] Jiao, Venkat, Han, and Weissman. Maximum likelihood estimation of functionals of discrete distributions. In *IEEE Transactions on Information Theory*, 2017. 95, 96

[92] Huang and Lin. The estimation of the m parameter of the nakagami distribution. *WSEAS Transactions on Biology and Biomedecine*, 13, 2016. 97

[93] Gaeddert and Annamalai. Further results on nakagami-m parameter estimation. In *IEEE Communications Letters*. IEEE, 2005. 97

[94] M. M. Siddiqui. Some problems connected with rayleigh distributions. 1962. JOURNAL OF RESEARCH of the National Bureau of Standards. 116

[95] S. O. Rice. Mathematical analysis of random noise. 1944. Bell System Technical Journal. 118

Appendix

Probability Distributions

The characteristics of the signal propagation vary greatly with the operating frequency, and the mode of propagation, e.g., line-of-sight (LoS) radio links, diffraction/scatter, and satellite links. Typically, a non-line-of-sight (NLoS) radio propagation path will exist between a BS and mobile station (MS), because of natural and man-made objects that are situated between the BS and MS. At the MS, plane waves arrive from many different directions and with different delays. This property is called multipath propagation. The multiple plane waves combine vectorially at the receiver antenna to produce a composite received signal.

Rayleigh

A MS in a typical macrocellular environment is usually surrounded by local scatterers so that the plane waves will arrive from many directions without a direct LoS component. Two-dimensional isotropic scattering where the arriving plane waves arrive in from all directions with equal probability is a very commonly used scattering model for the forward channel in a macrocellular system. For this type of scattering environment the received envelope is Rayleigh distributed at any time, and is said to exhibit **Rayleigh fading**[94].

Rayleigh fading is a statistical model for the effect of a propagation environment on a radio signal, such as that used by wireless devices. Rayleigh fading models assume that the magnitude of a signal that has passed through such a transmission medium (also called a communication channel) will vary randomly, or fade, according to a Rayleigh distribution — the radial component of the sum of two uncorrelated Gaussian random variables.

Rayleigh fading is viewed as a reasonable model for tropospheric and ionospheric signal propagation as well as the effect of heavily built-up urban environments on radio signals. Rayleigh fading is most applicable when there is no dominant propagation along a line of sight between the transmitter and receiver. If there is a dominant line of sight, Rician fading may be more applicable. Rayleigh fading is a special case of two-wave with diffuse power (TWDP) fading. Rayleigh fading is caused by multipath reception. The mobile antenna receives a large number, say N, reflected and scattered waves. Because of wave cancellation effects, the

instantaneous received power seen by a moving antenna becomes a random variable, dependent on the location of the antenna.

Rayleigh fading is a reasonable model when there are many objects in the environment that scatter the radio signal before it arrives at the receiver. The central limit theorem holds that, if there is sufficiently much scatter, the channel impulse response will be well-modelled as a Gaussian process irrespective of the distribution of the individual components. If there is no dominant component to the scatter, then such a process will have zero mean and phase evenly distributed between 0 and 2π radians. The envelope of the channel response will therefore be Rayleigh distributed.

The requirement that there be many scatterers present means that Rayleigh fading can be a useful model in heavily built-up city centres where there is no line of sight between the transmitter and receiver and many buildings and other objects attenuate, reflect, refract, and diffract the signal. Experimental work in Manhattan has found near-Rayleigh fading there. In tropospheric and ionospheric signal propagation the many particles in the atmospheric layers act as scatterers and this kind of environment may also approximate Rayleigh fading. If the environment is such that, in addition to the scattering, there is a strongly dominant signal seen at the receiver, usually caused by a line of sight, then the mean of the random process will no longer be zero, varying instead around the power-level of the dominant path. Such a situation may be better modelled as Rician fading. Note that Rayleigh fading is a small-scale effect. There will be bulk properties of the environment such as path loss and shadowing upon which the fading is superimposed.

The Rayleigh distribution 5.8 is defined by

$$p(R) = \frac{R}{\sigma^2} exp(-\frac{R^2}{2\sigma^2}), R \geq 0$$

where σ controls the spread of the distribution. It has a mean value defined by

$$E(R) = \sigma\sqrt{\frac{\pi}{2}}$$

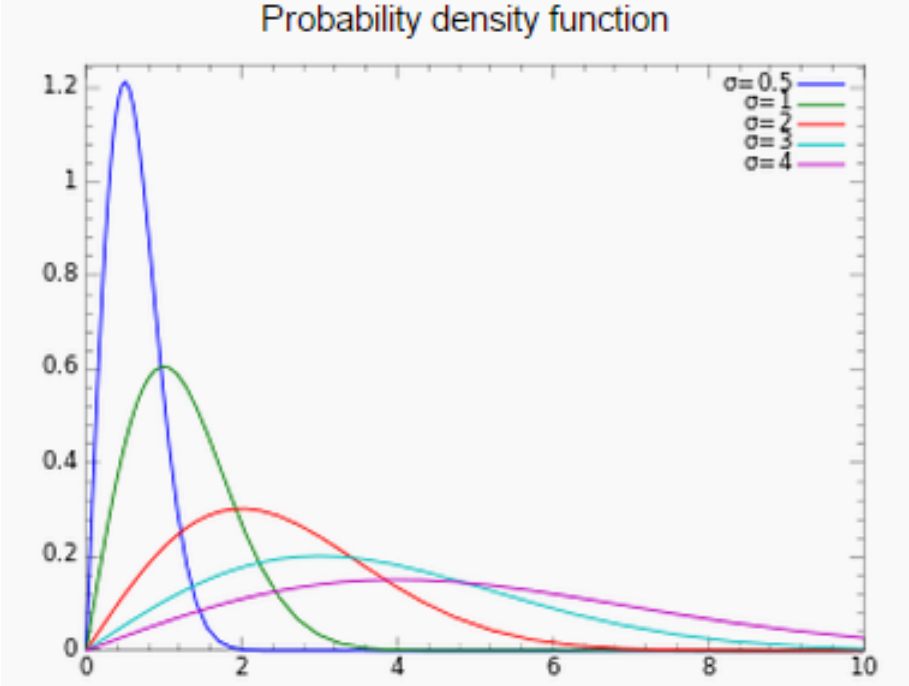

Figure 5.8 – DF Rayleigh

Rician

The LoS or dominant reflected or diffracted path produces the specular component and the multitude of weaker secondary paths contribute to the scatter component of the received envelope. In this type of propagation environment, the received signal envelope still experiences fading. However, the presence of the specular component changes the received envelope distribution, and very often a Ricean distributed envelope is assumed. In this case the received envelope is said to exhibit **Ricean fading**[95].

The model behind Rician fading is similar to that for Rayleigh fading, except that in Rician fading a strong dominant component is present. This dominant component can for instance be the line-of-sight wave. Refined Rician models.

- That the dominant wave can be a phasor sum of two or more dominant signals, e.g. the line-of-sight, plus a ground reflection. This combined signal is then mostly treated as a deterministic (fully predictable) process.
- The dominant wave can also be subject to shadow attenuation. This is a popular assumption in the modelling of satellite channels.

Rician fading or Ricean fading is a stochastic model for radio propagation anomaly caused by partial cancellation of a radio signal by itself — the signal arrives at the receiver by several different paths (hence exhibiting multipath interference), and at least one of the paths is changing (lengthening or shortening). Rician fading occurs when one of the paths, typically a line of sight signal or some strong reflection signals, is much stronger than the others. In Rician fading, the amplitude gain is characterized by a Rician distribution. Rayleigh fading is sometimes considered a special case of Rician fading for when there is no line of sight signal.

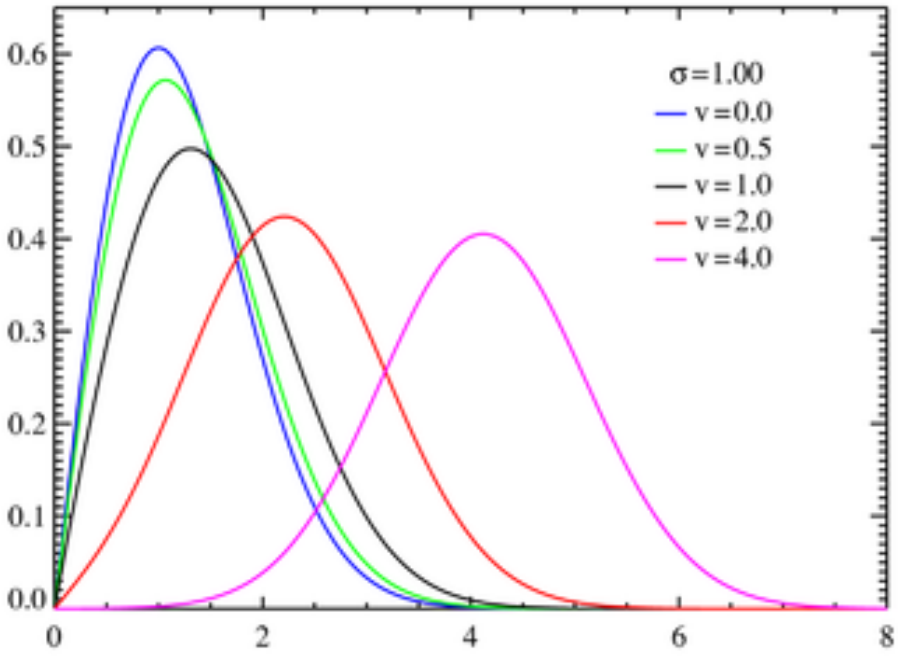

Figure 5.9 – DF Rice

In such a case, the Rician distribution, which describes the amplitude gain in Rician fading, reduces to a Rayleigh distribution. Rician fading itself is a special case of two-wave with diffuse power (TWDP) fading. A Rician fading channel can be described by two parameters. The first one, is the ratio between the power in the direct path and the power in the other, scattered, paths. The second one, is the total power from both paths, and acts as a scaling factor to the distribution.

The Ricean distribution 5.9 is defined by

$$p(R) = \frac{R}{\sigma^2} exp(-\frac{R^2 + s^2}{2\sigma^2}) I_0(\frac{R_s}{\sigma^2}), R \geq 0$$

where s controls the mean of the distribution , a its variance, and $I_0()$ the modified Bessel function of the first kind, order zero. It has a mean value defined by

$$E(R) = \frac{\sqrt{\frac{\pi}{2}}\sigma^2 L_{\frac{1}{2}}(-\frac{s^2}{2\sigma^2})}{|\sigma|}$$

where $L_n(x)$ is a Laguerre function.

The mean square value of the distribution is given by

$$E(R^2) = s^2 + 2\sigma^2$$

Nakagami

Besides Rayleigh and Rician fading, refined models for the pdf of a signal amplitude exposed to mobile fading have been suggested. The distribution of the amplitude and signal power can be used to find probabilities on signal outages.

- If the envelope is Nakagami distributed, the corresponding instantaneous power is gamma distributed.
- The parameter m is called the 'shape factor' of the Nakagami or the gamma distribution.
- In the special case $m = 1$, Rayleigh fading is recovered, with an exponentially distributed instantaneous power.
- For $m > 1$, the fluctuations of the signal strength reduce compared to Rayleigh fading.

The **Nakagami distribution** [84] was introduced by Nakagami in the early 1940's to characterize rapid fading in long distance HF channels. The Nakagami distribution was selected to fit empirical data, and is known to provide a closer match to some experimental data than either the Rayleigh, Ricean, or lognormal distributions.

The Nakagami fading model was initially proposed because it matched empirical results for short wave ionospheric propagation. In current wireless communication, the main role of the Nakagami model can be summarized as follows:

- It describes the amplitude of received signal after maximum ratio diversity combining. After k-branch maximum ratio combining (MRC) with Rayleigh-fading signals, the resulting signal is Nakagami with $m = k$. MRC combining of m-Nakagami fading signals in k branches gives a Nakagami signal with shape factor mk.
- The sum of multiple independent and identically distributed (i.i.d.) Rayleigh-fading signals have a Nakagami distributed signal amplitude. This is particularly relevant to model interference from multiple sources in a cellular system.
- The Nakagami distribution matches some empirical data better than other models.
- Nakagami fading occurs for multipath scattering with relatively large delay-time spreads, with different clusters of reflected waves.

Within any one cluster, the phases of individual reflected waves are random, but the delay times are approximately equal for all waves. As a result the envelope of each cumulated cluster signal is Rayleigh distributed. The average time delay is assumed to differ significantly between clusters. If the delay times also significantly exceed the bit time of a digital link, the different clusters produce serious intersymbol interference, so the multipath self-interference then approximates the case of co-channel interference by multiple incoherent Rayleigh-fading signals.

- The Rician and the Nakagami model behave approximately equivalently near their mean value.

This observation has been used in many recent papers to advocate the Nakagami model as an approximation for situations where a Rician model would be more appropriate. While this may be accurate for the main body of the probability density, it becomes highly inaccurate for the tails. As bit errors or outages mainly occur during deep fades, these performance measures are mainly determined by the tail of the probability density function (for probability to receive a low power).

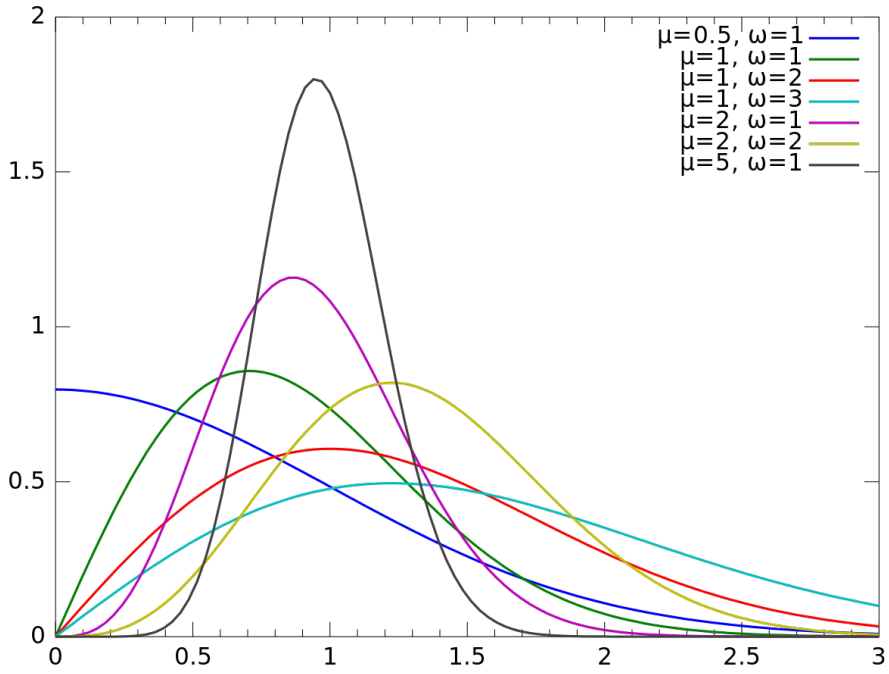

Figure 5.10 – DF Nakagami

The m-distribution covers a wide range of fading scenarios by varying its fading parameter m, includes the Rayleigh and one-sided Gaussian distributions as special cases for the respective fading of $m = 1$ and $m = 0.5$. The m-distribution can also closely approximate the Rician distributions.

The Nakagami distribution 5.10 is defined by

$$p(R) = \frac{2m^m}{\Gamma(m)\Omega^m} R^{2m-1} exp(-\frac{m}{\Omega}R^2), R \geq 0$$

where m is a shape parameter, and Ω controls the spread of the distribution.

The distribution has a mean value defined by

$$E(R) = \frac{\Gamma(m+\frac{1}{2})}{\Gamma(m)}(\frac{\Omega}{m})^{\frac{1}{2}}$$

and a mean square value defined by

$$E(R^2) = \frac{\Gamma(m+1)}{\Gamma(m)}(\frac{\Omega}{m})$$

resulting in a variance

$$Var(R) = \Omega(1 - \frac{1}{m}(\frac{\Gamma(m+\frac{1}{2})}{\Gamma(m)})^2)$$