A complete resource allocation framework for flexible high throughput satellite constellations

by

Nils Pachler de la Osa

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics and Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Author . . . .
Department of Aeronautics and Astronautics
May 6, 2022

Certified by . . . .
Edward F. Crawley
Professor of Aeronautics and Astronautics
Thesis Supervisor

Certified by . . . .
Bruce G. Cameron
Director of the System Architecture Group
Thesis Supervisor

Accepted by . . . .
Jonathan P. How
R. C. Maclaurin Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

A complete resource allocation framework for flexible high throughput satellite constellations

by

Nils Pachler de la Osa

Submitted to the Department of Aeronautics and Astronautics
on May 6, 2022, in partial fulfillment of the
requirements for the degree of
Master of Science in Aeronautics and Astronautics

Abstract

The past few years have been a witness for the rise of the new era of satellite communications. The interest for providing broadband access from space has reached levels that remind of those in the late 90s with Iridium and Globalstar among others. The novel mega-constellation designs will rely on thousands of highly capable satellites to provide service to the ever-growing communications market. Nevertheless, the new payload flexibilities and the larger space segment involve a level of complexity and dimensionality that this sector has not seen before. While manual allocation of resources was feasible and efficient in early stages of this industry, it becomes unfeasible under the new conditions. To exploit the capabilities of the new spacecrafts to its full potential, new automatic and optimized tools for the Resource Allocation (RA) problem in the context of satellite communications need to be developed.

While individual tools and methods for the specific sub-problems have been proposed during these last years, most approaches fail to address the interactions between different sub-problems, and those who do rely on simplified assumptions that do not capture the reality of modern operations. To close this gap, this Thesis proposes an adaptive framework to solve the long-horizon RA problem under high dimensionality conditions for highly flexible satellite constellations. The proposed framework uses a divide-and-conquer approach where the RA problem is decomposed into different sub-problems, then solved via state-of-the-art optimization techniques, and integrated back to obtain a valid, feasible, and efficient solution for the long-horizon RA problem. The performance of this framework is then analyzed using different user distributions, model parameters, and solution algorithms to understand the capabilities and robustness of the obtained solutions, as well as the sensitivity to the different variables.

The executed analyses prove the validity and effectiveness of the framework to deal with the incumbent problem. Specifically, the proposed method and algorithms prove to be robust against a variety of user distributions and model parameters, being always able to obtain a feasible plan. In addition, the tests performed in this work demonstrate that the state-of-the-art algorithms significantly outperform simple techniques, being able to multiply the capacity of the constellation by 4 with the same payload characteristics, while reducing to a third the power consumption. Furthermore, the sensitivity tests prove that optimized solutions are able to achieve improved coverage even with limited hardware compared to heuristic techniques.

Thesis Supervisor: Edward F. Crawley
Title: Professor of Aeronautics and Astronautics

Thesis Supervisor: Bruce G. Cameron
Title: Director of the System Architecture Group

Acknowledgments

This Thesis has been a product of 2+ years of effort under extreme conditions due to the COVID-19 global pandemic. First of all, I want to use this opportunity to thank all healthcare and essential workers that have kept our society moving, and all members of society who have collaborated under this conditions so that this pandemic can end. Thanks to all of you.

I would like to sincerely thank my advisor, Prof. Edward F. Crawley, and Dr. Bruce G. Cameron for their support, advice, and guidance since I started at MIT in 2019 as a visiting student. Ed, Bruce, we have been working on this project for the last three years and I really appreciate your input and implication to make this research project succeed. It is truly an honor to work side by side with you.

I would really like to thank everyone who has actively or passively participated in this research project. First, I would like to express my gratitude to my labmate and mentor Juanjo Garau, who has guided and instructed me during these two years of hard work, but also accompanied me along with many moments of joy. Juanjo, thank you for your wisdom and teachings about how to be a better researcher, I sincerely wish you all the best in your future endeavors. Next, I really appreciate the support received by SES, especially the feedback and motivation from Joel Grotz, Valvanera Moreno, Íñigo del Portillo, and Markus Guerster. I would like to acknowledge the rest of my labmates who have also been a part of this project at some point during these two years. Skylar, James, Guillem, Sergi, Damon, and Rubén, you have been a source of inspiration and stimulating discussions, best of luck on your next steps.

Although my lab operations have been heavily disrupted by the ongoing pandemic, I would also like to thanks the rest of my labmates who have been a key to succeed in this first part of my graduate studies. Johannes, Chloé, Elwyn, Adam, Justin, thank you for your constant encouragement and for all the fun times together. I would like to express my sincere gratitude and appreciation for Anne-Marlene Rüede, with whom I have shared many experiences. Anne-Marlene, thank you for showing me that graduate life is more than just problem solving in front of a screen, and thank you for all the time we have spent together. I do not want to miss the chance to say thank you for the invaluable administrative support I always get from Amy Jarvis, Beth Marois, Beata Shuster, and Ping Lee.

Finally, but not less important, I would like to thank my parents, Jørn and Nina, for their unconditional support and love during my whole life. Thank you both for all you gave me and the knowledge and skills that have helped me achieve my dreams. I am also deeply grateful to my closest friends – you know who you are – who always show me that distance is nothing when it comes to our friendship.

Contents

1 Introduction **1**

1.1 Motivation . . . 1

1.2 General objectives . . . 3

1.3 Background . . . 4

1.3.1 A Communications Satellite . . . 4

1.3.2 The link budget equation . . . 10

1.3.3 On the Resource Allocation framework . . . 15

1.4 Specific objective . . . 18

1.5 Overview . . . 19

2 Literature review **20**

2.1 Power Allocation . . . 20

2.2 Short-horizon Frequency Assignment . . . 23

2.2.1 TDMA . . . 23

2.2.2 Beam Hopping . . . 24

2.3 Long-horizon Frequency Assignment . . . 26

2.3.1 FDMA or Bandwidth Allocation . . . 26

2.3.2 Frequency Assignment . . . 28

2.4 User Grouping and Beam Placement 30
2.4.1 Beam Placement 30
2.4.2 User Grouping 30
2.5 Beam Shaping 31
2.6 Satellite Routing 32
2.7 Gateway and Inter-Satellite Routing 35
2.7.1 Gateway Routing problem 35
2.7.2 Inter-Satellite Routing 35
2.8 Joint problems 37
2.8.1 Research on two sub-problems 38
2.8.2 Research on more than two sub-problems 45
2.9 Literature summary 48

3 A Divide and Conquer approach to the Resource Allocation Problem 52

3.1 On the Resource Allocation framework 53
3.2 User Grouping and Beam Shaping 58
3.2.1 General problem description 58
3.2.2 Specific problem description 58
3.2.3 Subsequent problem interface 60
3.3 Satellite Routing 62
3.3.1 General problem description 62
3.3.2 Specific problem description 62
3.3.3 A brief discussion on the joint User Grouping and Satellite Routing problem 65
3.3.4 Subsequent-problem interface 66

3.4 Gateway Routing 66
3.4.1 General problem description 66
3.4.2 Specific problem description 67
3.4.3 A brief discussion on the joint Satellite Routing and Gateway Routing problem 72
3.4.4 Subsequent-problem interface 72
3.5 Frequency Assignment 73
3.5.1 General problem description 73
3.5.2 Specific problem description 74
3.5.3 A brief discussion on the joint Gateway Routing and Frequency Assignment problem 78
3.6 System metrics 78
3.6.1 Power consumption as a system metric 79
3.6.2 User Satisfaction as a system metric 80
3.7 Assumptions and Challenges 81
3.7.1 General assumptions and relaxations 81
3.7.2 Specific assumptions 83

4 Heuristics, practical implementations, and efficient optimizations for the Resource Allocation sub-problems 89
4.1 User Grouping 90
4.1.1 One beam per user 90
4.1.2 Minimum number of beams 91
4.1.3 Genetic Algorithm 91
4.1.4 Coverage grid 93
4.2 Satellite Routing 94

4.2.1 Closest satellite 95
4.2.2 Mixed Integer Linear Programming 96
4.2.3 Particle Swarm Optimization 96
4.3 Gateway Routing 98
4.3.1 Closest gateway 99
4.3.2 Mixed Integer Linear Programming 99
4.4 Frequency Assignment 100
4.4.1 Heuristic approach 101
4.4.2 Integer Linear Programming 101
4.5 Power Allocation 104

5 A complete Resource Allocation Process for a long-term operations plan 106

5.1 Integration validation 107
5.2 Test procedures 108
5.3 User distribution 109
5.3.1 Satellite operator model (SES) 109
5.3.2 Proportional to population (Population) 109
5.3.3 Proportional to uncovered population (Uncovered) 112
5.4 Constellation and Gateway Model 114
5.4.1 Constellation Model 114
5.4.2 Gateway Model 115
5.5 Experiments 115
5.6 Baseline Comparison and Performance Analysis (Experiment Set A) . 118
5.6.1 Low system capacity 118
5.6.2 High system capacity 123

5.6.3 Run-time analysis 127

6 Sensitivity and Robustness analysis on the Resource Allocation Process 129

6.1 Robustness Analysis (Experiment Set B) 129

6.1.1 Robustness to dimensionality on the SES distribution 130

6.1.2 Robustness to dimensionality on the Population distribution . 132

6.1.3 Robustness to dimensionality on the Uncovered distribution . 135

6.2 Sensitivity Analysis (Experiment Set C) 137

6.2.1 Sensitivity test: Number of satellites 138

6.2.2 Sensitivity test: Number of gateways 140

6.2.3 Sensitivity test: Number of beam channels 142

6.2.4 Sensitivity test: Bandwidth per beam channel 144

6.2.5 Sensitivity test: Frequency reuse factor 146

6.2.6 Sensitivity test: Half cone angle 148

6.2.7 Sensitivity test: Number of and bandwidth per beam channel 149

6.2.8 Sensitivity summary 151

6.3 Sensitivity of the framework results to minor user changes 152

6.3.1 Sensitivity of the framework results to multiple executions . . 153

6.3.2 Sensitivity of the framework results to minor user changes . . 155

7 Conclusions 156

7.1 Thesis summary 156

7.2 Main findings 158

7.3 Future work 161

Appendices 163

A Mathematical Transformations **164**

A.1 Logic operations 164

A.1.1 OR 164

A.1.2 AND 165

A.2 Activation variables 165

A.2.1 Transforming an inequality into a binary variable 165

A.2.2 Transforming an equality into a binary variable 166

B Complete results **167**

List of Figures

1-1 Division of frequency spectrum 6
1-2 Gain of a parabolic antenna 9
1-3 Minimum elevation angle definition 10

3-1 Long-horizon Resource Allocation framework 57
3-2 The joint User Grouping and Beam Shaping problem 61

4-1 A possible grid based solution for the User Grouping problem 94

5-1 User Population distribution 111
5-2 User Uncovered distribution 113
5-3 Performance comparison under low system capacity 119
5-4 Impact of each resolution procedure decision on the low capacity system 122
5-5 Performance comparison under high system capacity 124
5-6 Impact of each resolution procedure decision on the high capacity system 126

6-1 Performance comparison under three dimensionality scenarios 131
6-2 Performance comparison under the population proportional user distribution 132

6-3 Performance comparison under the uncovered population user distribution . . . 135
6-4 Performance comparison under different number of satellites 139
6-5 Performance comparison under different number of gateways 141
6-6 Performance comparison under different number of beam channels . . 143
6-7 Performance comparison under different bandwidths 145
6-8 Performance comparison under different frequency reuse factors . . . 147
6-9 Performance comparison under different half cone angles 149
6-10 Performance comparison under different number of beam channels and bandwidths . . . 150

List of Tables

1.1 Summary of the six different sub-problems for NGSO constellations . 16
1.2 Short-horizon sub-problems for NGSO constellations 18
1.3 Long-horizon sub-problems for NGSO constellations 18
1.4 Sub-problems considered in this work 19

2.1 Summary of the literature's scope 50
2.2 Summary of the literature's characteristics 51

3.1 N^2 diagram on the flow of information on the long-horizon RA problem 54

5.1 Constellation parameters. Those parameters are either taken directly from the filings or based on the author's assumptions. 114
5.2 Summary of the resolution procedures used in the experiments 117
5.3 Summary of the experiments of this work. 117
5.4 Detailed numbers on the performance comparison under low system capacity. 120
5.5 Detailed numbers on the performance comparison under high system capacity. 125
5.6 Execution time for each optimization algorithm considered in this work128

6.1 Detailed numbers on the performance comparison under different dimensionality scenarios on the Population dataset 133
6.2 Detailed numbers on the performance comparison under different dimensionality scenarios on the Uncovered dataset 136
6.3 Impact of choosing the fully optimized pipeline over the heuristic solution when analyzing different model parameters 152
6.4 Sensitivity of the optimization algorithms on multiple independent executions under identical inputs 154
6.5 Sensitivity of the optimization algorithms on two independent executions under slightly different initial conditions 155

B.1 Detailed numbers on the performance comparison under different dimensionality scenarios on the SES dataset 167
B.2 Detailed numbers on the performance comparison under different dimensionality scenarios on the Population dataset 168
B.3 Detailed numbers on the performance comparison under different dimensionality scenarios on the Uncovered dataset 169
B.4 Detailed numbers on the performance comparison under different number of satellites 170
B.5 Detailed numbers on the performance comparison under different number of gateways 170
B.6 Detailed numbers on the performance comparison under different number of beam channels 171
B.7 Detailed numbers on the performance comparison under different bandwidth per beam channel 172

B.8 Detailed numbers on the performance comparison under different frequency reuse factors . . . 172

B.9 Detailed numbers on the performance comparison under different half cone angles . . . 173

B.10 Detailed numbers on the performance comparison under different number of and bandwidth per beam channel . . . 173

Nomenclature

Acronyms

ABC	Artificial Bee Colony
ACM	Adaptive Modulation and Coding
ACO	Ant Colony Optimization
BSh	Beam Shaping
DRL	Deep Reinforcement Learning
$EIRP$	Effective Isotropic Radiated Power
FA	Frequency Assignment
$FDMA$	Frequency-division multiple access
FSL	Free Space Loss
GA	Genetic Algorithm
GEO	Geosynchronous Earth Orbit
GR	Gateway Routing
ILP	Integer Linear Programming
$ISL(s)$	Inter satellite link(s)
ITU	International Telecommunication Union
LEO	Low Earth Orbit
LSE	Least Squared Error
MD	Met Demand
MEA	Minimum Elevation Angle
MEO	Medium Earth Orbit
$MILP$	Mixed Integer Linear Programming
$MODCOD$	Modulation and Coding
$NGSO$	Non GeoSynchronous Orbits
NN	Neural Network
PA	Power Allocation
PSO	Particle Swarm Optimization

RA	Resource Allocation
SA	Simulated Annealing
$SINR$	Signal to Interference plus Noise Ratio
SR	Satellite Routing
$TDMA$	Time-division multiple access
UD	Unmet Demand
UG	User Grouping
USC	Unmet System Capacity

Symbols

δ	Beam cone angle
η_i	Number of channels needed for beam i
Γ	Spectral efficiency
λ	Wavelength
$\mathcal{V}_i$	Set of users covered by beam i
θ	Angle with reference to the beam pointing
A	Area
BW_i	Bandwidth needed for beam i
c	Speed of light / Bandwidth of a channel (context dependent)
D	Antenna diameter
d_u	Demand of user u
f	Frequency
G	Gain
k	Boltzmann constant
N	Noise power
P	Power / Period (context dependent)
p_i	Position of element i
R	Data rate
r	Antenna radius
T	Temperature
$t_{start,i}$	Initial point in time where beam i is visible
$t_{stop,i}$	Final point in time where beam i is visible
W	Watt
x, y	Auxiliary variables (problem dependent)

Chapter 1

Introduction

1.1 Motivation

In an era driven by information, staying connected becomes a top priority. Over the past few years, the world's network infrastructure has been rapidly expanding to cover the novel necessity for such information [1]. While terrestrial nets handle most of this load, satellite networks have recently re-emerged to absorb the traffic where the land infrastructure is poor or non-existent [2]. After a severe drawback in the early 2000's with the failure or cancellation of most space constellations for broadband communications [3], newborn projects have materialized to deal with the new market's necessities. Both old competitors (e.g., SES, Telesat), and new entrants (e.g., SpaceX, Amazon) aspire to take a piece of the communications cake.

The new era of satellite communications will be driven by two main factors: an expanding user base with larger traffic requirements, and a technological improvement that boosts the capacity of the network. For the former, several recent studies [2, 4] analyze the economic aspects of the new mega-constellations and what parameters and factors are the key drivers of the financial viability of such systems. New users with a variety of demand requirements will enter the market, and to maximize profit the satellite operators will need to manage highly variable traffic for tens of thousands of users in real-time. For the latter, the digitalization of on-board resources, usage of phased-array antennas, and the adoption of novel adaptive control and modulation allow for an enhanced spacecraft performance. Furthermore, reduced launch costs and mass production of satellites evoke a larger and highly flexible space segment.

Moreover, managing the diverse pool of resources of a contemporary satellite to guarantee service and maximize throughput is not an easy task. Modern spacecrafts are able to reconfigure the position, shape, frequency, and power of each beam in almost real-time. Each of these resources is in nature very distinct and involves a particular and unique set of constraints. On top of that, satellites are able to project from tens to thousands of beams and each constellation is constituted from a range between tens to thousands of satellites. For example, SpaceX Starlink constellation [5–8] has 4408 satellites with around 32 spot beams per spacecraft, while O3b mPower constellation [9, 10] is conformed by 10 satellites with thousands of beams per spacecraft. Both of those systems will need to continuously monitor tens of thousands of beams.

While the new system capabilities boost the total throughput and increases the range and availability of the space network, the new flexibilities and dimensionality involve a greater degree of complexity that hasn't been seen before in satellite operations. While manual and static resource allocation mechanisms worked well for early stages of this industry, these techniques become obsolete when dealing with thousands of software-driven satellites and tens of thousands of demand-varying users in real-time. New automated tools and methods for the high dimensional resource allocation problem in satellite communications need to be developed. Furthermore, optimizing the resource allocation process allows for a more effective usage of the satellite flexibilities, increasing the overall performance of the system. Thus, operators not only aim for a feasible operations plan, but rather the best possible operations plan that maximizes throughput. However, the combination of resource-specific constraints plus the high dimensionality element make the resource allocation problem a complex conundrum difficult to solve with traditional optimization techniques.

For this reason, recent literature has started looking at artificial intelligence (AI) approaches to deal with the naturally dissimilar resources and/or the high-dimensionality of the problem due to its performance in similar fields [11]. While state-of-the-art studies have investigated individual resources in high-dimensional environments, or pairs or triples of resources in dimensionality-restricted scenarios, there is no clear path of how to implement those methods in real operations. This Thesis aims to close this gap by developing a framework to solve the resource allocation problem for satellite communications in high-dimensional scenarios using a decomposition plus integration framework in combination with state-of-the-art algorithms to achieve optimized system-level performance.

1.2 General objectives

The previous Section identified the key drivers of the next generation of satellite communications and the necessary methods and tools that need to be developed to achieve maximum system's performance. Specifically, before implementing the academia methods into industry, further research needs to:

- Develop a method to automate the complete resource allocation for satellite communications. This tool needs to take into account all the flexibilities and dimensionality aspects of the new constellations and provide a feasible plan for the satellite operators to execute. Furthermore, this method should be based on an optimization framework that tries to maximize system's performance by efficiently assigning the available resources.

- Assess the capabilities and limitations of such tool. It is important to understand to what limits can the tool be extended to and in which contexts can it be applied. It is necessary to benchmark the resource allocation method under a representative set of use cases to provide the satellite operator with enough confidence and knowledge to cast informed decisions.

1.3 Background

The objective of this section is to introduce the reader to general concepts of satellite communications that are going to be used throughout this Thesis. This does not intend be a profound analysis of space communications or a replacement of other more detailed work, but rather a brief summary of the most important concepts and relevant definitions. For more knowledge about satellite communications, refer to Maral et al. [12].

1.3.1 A Communications Satellite

A communications satellite can be defined as a device orbiting around the Earth that is able to transmit information from one point of the Earth to another. This transmission follows a two-step process connecting three entities: the transmission from a ground station (i.e., a station on the surface of the Earth) to the satellite (a.k.a. satellite uplink) and from the satellite to a second ground station (a.k.a.

satellite downlink). Optionally, the information can travel through multiple satellites before returning to the ground. The satellite to satellite connections are commonly referred to as inter-satellite links (ISLs). In addition, the ground stations can be defined into two sub-groups: users (or user terminals) and gateways. Users are small ground stations that pay for a specific service and ask for information to be provided. Gateways are generally larger and are responsible of providing that information. The two flows of communication are: the forward link, which follows the path gateway - satellite - user, and the return link, which follows user - satellite - gateway. For services such as Internet provision, both flows tend to be asymmetric, as Internet requests are generally smaller than the data being requested. Other types of services such as phone communications may be more symmetric.

The Frequency Spectrum

Contrary to terrestrial cable communications based on electrical impulses, satellite communications rely on electromagnetic waves. The information is encoded into the wave at the transmitter, and decoded at the receiver. These waves are defined by three main factors:

- *Power*: Determines the wave's travel distance. As power decays with distance, there is a point where the signal is no longer distinguishable from ambient noise, which restricts the wave's reachable destinations.

- *Frequency*: Rate of oscillation of the wave. Given the large amount of frequency users and the scarcity of the frequency resource, the International Telecommunication Union (ITU) is responsible of dividing the frequency space into the different sectors. The bands assigned for space communications are shown in Figure 1-1.

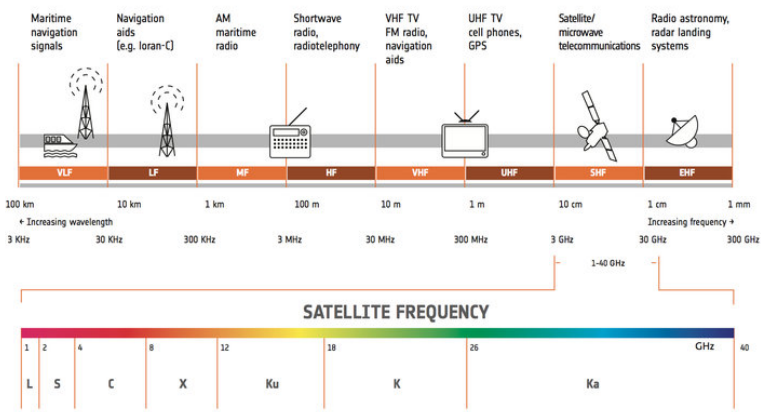


Figure 1-1: Division of frequency spectrum [13]

- *Polarization*: Any electromagnetic wave is formed by an electric field and a magnetic field. By nature, both fields are orthogonal and perpendicular to the propagation of the wave and the direction of these fields is decided at the transmission station. By physics, two perpendicular electric or magnetic fields do no interfere. This allows for two waves with perpendicular fields with the same frequency to be transmitted at the same time without information loss, which permits doubling the effective capacity of the communication link. The direction of the electric and magnetic fields is known as polarization.

Rather than encoding the information in the amplitude of the wave, as commonly done in terrestrial links, the information is encoded in the frequency of the wave. This allows for a much reliable link, at the expense of being limited by the amount of frequency spectrum available. To transform the information bits into frequencies, what is known as modulation and coding schemes (MODCODs) are commonly used. These MODCODs are standard processes that, depending on the quality of the link, decide how to encode the transmission information and how many error correction bits to include. A higher quality link needs less error correction, and thus the information transmission rate is higher. On the contrary, if the signal is barely distinguishable from surrounding noise, the link needs many error correction bits, which reduces considerable the efficiency of the connection. As a simple example, with a perfect link we only need to transmit the information once with no error correction, which allows for 100% efficiency. On the contrary, if the link is very unreliable, we may need to transmit the same information twice or thrice, which implies efficiencies of 50% or 33%, respectively. The MODCODs used in this Thesis are the ones in the standard DBV-S2X [14].

Satellite architecture

In simple words, a communications satellite is no more than a very sophisticated mirror: a ground station sends a wave with information encoded and expects the satellite to redirect that information to another ground station somewhere else. Conventional satellites have often used what is known as the bent-pipe model: the received signal is amplified and redirected directly into the transmitting antenna without modification. This allows for a simplified payload model, but imposes additional constraints regarding the uplink and downlink, since the signal is inherently the same. On the other hand, modern satellites are able to decode the signal in the satellite payload and encode it again with a different MODCOD. This makes the uplink and downlink independent, and allows for a higher flexibility and spectrum usage. This latter architecture is the one used in this work. However, the methods, algorithms, and implementations developed can be applied to bent-pipe architectures with minimal modifications.

Antennas

The entities responsible for transforming a electrical signal into an electromagnetic wave and vice versa are the antennas. By definition, an isotropic antenna is an antenna that transmits the electromagnetic wave to all directions with the same intensity. On the contrary, the most used type of antennas in space communications are parabolic, which allow to concentrate most part of the power into a specific direction or set of directions. Usually, these antennas create lobes as shown in Figure 1-2. The receiver antennas located inside the lobe receive a much higher signal intensity that the ones outside. Since these antennas use power much more efficiently by directing it into the correct direction, they help save power compared to an isotropic antenna. The amount of power saved is known as the gain of the antenna. Formally defined, the gain is the amount of power needed by an isotropic antenna to reach the same signal intensity as a parabolic antenna at the center of the main lobe. To determine the size of the lobe, the 3dB angle ($\theta_{3dB}/2$) is defined as the angle at which the power is half of the power at the center of the lobe. A beam is defined as the cone produced by the direction of the center of the lobe and a cone angle of θ_{3dB}. The center of the beam is the point where the axis of the cone intersects the surface of the Earth. This definition can be seen as a hard constraint: users that fall within $\theta_{3dB}/2$ of the center of the lobe can be served by the beam, while users outside cannot.

Going a step beyond parabolic antennas, modern satellites use what is known as phased array antennas. These antennas are able to create thousands of very narrow beams that can reach a wide angle of directions with high gain without the need

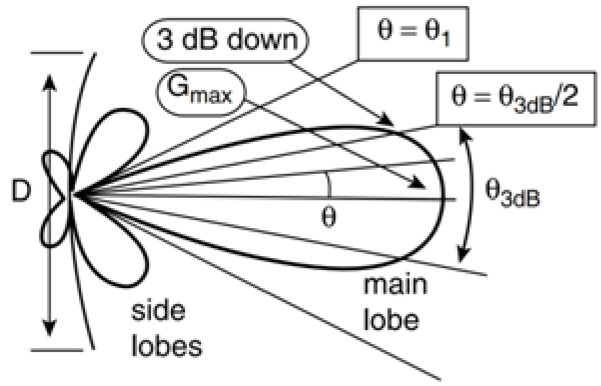


Figure 1-2: Gain of a parabolic antenna [12]

of moving the antenna. Although this allows for improved information rates, the operational burden of managing all those beams is not negligible.

Satellite coverage

While modern satellites can create thousands of beams, not all positions in the surface of the Earth can be reached by a single satellite at the same time. Given that the Earth has a somewhat spherical shape, it acts as an occluding body to the satellite. Ideally, the portion of the Earth that can be reached from a satellite is the set of points that form an angle with the Earth's surface higher than 0 degrees. In practice, due to additional occluding elements such as buildings or natural formations, this angle (known as minimum elevation angle, MEA) is higher than 0. Figure 1-3 shows the coverage of a satellite with 0 degrees and ϵ degrees MEA (in blue and green, respectively). The MEA is a given parameter of the constellation, as it must be specified when filing for a new constellation. Formally, the coverage of a satellite at any given time is the portion of the Earth that forms an elevation angle higher or equal than the predefined MEA.

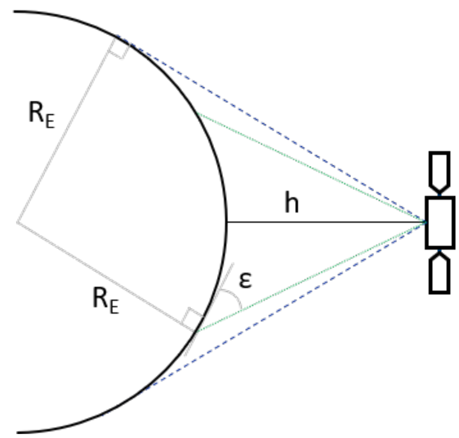


Figure 1-3: Satellite coverage with 0 and ϵ minimum elevation angle, with blue dashed and green dotted lines, respectively. R_E denotes the radius of the Earth, and h is the altitude of the orbit.

Given that the new applications for satellite communications follow NGSO, satellites drift over time with respect to the Earth. Therefore, regions that were not visible at some point in time may become visible after a while and vice versa. Starting from a reference time, the visibility window of a specific surface position can be computed as the period of time where the elevation angle between that position and the satellite is higher than the MEA. This range can be computed with classical orbital equations by determining the (at most) two moments for every period where the elevation angle is equal to the MEA.

1.3.2 The link budget equation

To allow distinguishing the signal from surrounding noise, it needs to have a significantly higher intensity than such noise. As mentioned in the previous section, the intensity of an electromagnetic signal is equivalent to the received power. The following lines detail how to compute the power to noise ratio, which will determine the quality of the link.

First, the transmission antenna emits the signal with a specific power P_T. Due to the use of non-isotropic antennas, this power gets amplified by a gain G_T. The resulting quantity is known as effective isotropic radiated power (EIRP):

$$EIRP = P_T G_T \text{ [W]} \tag{1.1}$$

To match common communications notation, values are usually transformed to the logarithmic domain to keep track of large numbers. The common transformation is: [dB] $= 10\log_{10}[-]$. Next, for parabolic antennas, the gain at the center of the beam G_{max} can be computed as:

$$G_{max} = \mu \left(\frac{\pi D}{\lambda}\right)^2 \; [-] \tag{1.2}$$

Where μ is the efficiency of the antenna, λ is the wavelength of the signal (which can be computed with the frequency f and the speed of light c as $\lambda = \frac{c}{f}$), and D is the diameter of the antenna. The gain at an angle θ off the center axis, the total gain is:

$$G(\theta) = G_{max} - 12\left(\frac{\theta}{\theta_{3dB}}\right)^2 \text{ [dB]} \tag{1.3}$$

Note that this definition is already in the logarithmic domain (i.e., G_{max} needs to be converted before the operation). Then, the signal travels through the space and atmosphere to reach the receiver antenna. This travel implies a loss commonly known as the free space loss (FSL):

$$L_{FSL} = \left(\frac{4\pi r}{\lambda}\right)^2 \; [-] \tag{1.4}$$

Where r is the distance between both antennas. Note that this value appears as a consequence of the loss of the signal due to the natural propagation of the wave. No atmospheric losses have been included until this point. After such distance, the signal reaches the receiver antenna and it gets amplified due to an additional gain on the receiving end:

$$P_R = P_T + G_T + G_R - L_{FSL} \text{ [dB]} \tag{1.5}$$

P_R represents the power received at the receiver antenna. However, this is the ideal power (i.e., if every element is ideal and there are no further losses). Reality tends to be less perfect. The effects of additional losses can be reflected with the inclusion of an additional generic term L. Those losses can have many sources: atmospheric attenuation, non-ideal receivers and transmitters, pointing losses, signal disturbances, etc. For a complete list of sources and a method for quantifying each one refer to [12]. Although the received power is relevant to establish a successful connection, its relation to noise is what determines if the signal can be understood or not. In this context, noise is usually defined as:

$$N = kTB \text{ [W]} \tag{1.6}$$

Where k is Bolzmann's constant, B is the bandwidth of the signal, and T is the equivalent temperature at which a standard resistor would produce noise N. Then the signal to noise ratio (SNR) can be computed as:

$$SNR = P_T + G_T + G_R - L_{FSL} - L - N \text{ [dB]} \tag{1.7}$$

Interference and how to avoid it

In addition to surrounding noise, signals can receive and produce interference from and to other signals if they operate on the same frequency range and polarization. The four classical interference sources are:

- *Carrier to adjacent beam interference*: two beams may interfere if they point to nearby locations.
- *Carrier to adjacent satellites interference*: two nearby satellites pointing at nearby locations may cause mutual interference.
- *Carrier cross polarization interference*: since polarization is not a perfect process, signals with different polarizations may still have an interfering effect between them.
- *Carrier to third order inter-modulation products of interference*: electronic manipulation of signals create residual signals on nearby frequency bands, which can interfere with other beams.

From this four elements, the most relevant one on continuous operation is the carrier to adjacent beam interference. Nevertheless, it can be efficiently dealt with in the resource allocation process. For the purposes of this work, it will be assumed that nearby beams only produce harmful interference if they occupy the same frequency band and they are *geographically close*. This latter factor will be determined using the minimum angle difference between both beams when they are in the same satellite and a threshold angle, after which interference is considered to be negligible. Assuming this element is effectively managed in the RA process, for the each of the interference components, an average interference of value I will be assumed. Then, the signal to noise plus interference can be computed as:

$$SINR = \frac{1}{\frac{1}{SNR} + \frac{4}{I}} \ [-] \tag{1.8}$$

Satellite throughput

For a signal to be correctly decoded by a receiver, two additional factors must be taken into account: first, each MODCOD has an associated required margin (called OBO) needed to perform the operation, and second, operators usually include a margin factor m as a way to avoid signal loss under unforeseen disturbances. Then, a signal can be properly understood if:

$$SINR - OBO - m \geq 0 \ [\mathrm{dB}] \tag{1.9}$$

Where the acronym $SINR$ stands for signal to interference plus noise ratio. Additionally, each MODCOD is directly related with a spectral efficiency Γ, that quantifies how many bits can be transmitted per second for each Hertz of bandwidth used. For a complete list of MODCODs, OBO, and spectral efficiency, see [14]. The total data-rate for a specific beam i can then be computed as:

$$R_i = BW_i \ \Gamma_i \tag{1.10}$$

Where BW_i corresponds to the bandwidth allocated to beam i, and Γ_i is the associated spectral efficiency of the MODCOD used in the link of beam i. The value R_i determines the data-rate provided to beam i and it is the way a communications satellite delivers performance.

The procedure presented to compute the data-rate is usually a more academic approach that what reality demands. In practice, the data-rate is given as a constraint, and the MODCOD can be adapted to achieve the desired value. For this purpose, this work will use an adaptive modulation and coding (ACM) scheme in which the MODCOD with lowest OBO that matches the desired data-rate (i.e., that has a spectral efficiency equal or higher than the requested value) is chosen. In case the required spectral efficiency is too high and no MODCOD can match that value, the highest spectral efficiency MODCOD will be used (so that the maximum amount of data-rate can be provided). Note that this method assumes no constraints in power. Power constrained approaches will be discussed later in this work.

1.3.3 On the Resource Allocation framework

The Resource Allocation problem (RA) for satellite communications can be defined as finding the most efficient resource distribution that meets the users' requirements. This definition is driven purely by the general objective of maximizing profit: we want to serve current users the best we can (since this maximizes current revenue), while leaving as many resources as possible for future users (since this maximizes expected revenue). To understand how to solve this problem, first we have to define two important concepts: how are the users' requirements defined and what are the satellite resources. In this work, we consider the users' requirements as some demand that needs to be met. Each user is considered a point on the surface of the Earth and the objective for the satellite constellation is to serve the demand of all users.

On the other hand, the satellites resources are all the elements that need to be allocated so that the connection between user and gateway can take place. Guerster et al. [15] consider four different resources to allocate: the radio-frequency transmission power, the central frequency and bandwidth, the position or direction of the beam, and the beam shape. In general, these are the classical elements on a GEO satellite. However, given the capacity increase of modern satellites, operators usually need multiple gateways per satellite to offload all their data traffic. Deciding which gateway to use for each beam is an additional factor that was not considered in earlier stages of this industry. Moreover, given the time-dependency in NGSO, there are additional resources or decisions that need to be considered, such as when to handover between different satellites. In general, we can decompose the RA problem for satellite communications for modern NGSO systems into six different decisions, summarized in Table 1.1. Given the complexity of this joint problem, the RA problem in satellite communications is often subdivided into different sub-problems, each associated with a different satellite resource.

Resource or decision to make	**Common name(s)**	**Question to answer**
Transmission power (or radio-frequency transmission power)	Power Allocation	What is the necessary power transmitted to the antenna to achieve the desired link margin and data rate for a specific beam?
Frequency	Frequency Assignment	How much bandwidth and which frequency range to use for each beam and user?
Position and user grouping	Beam Placement / User Grouping	Where to place the beam and which users to serve?
Beam shape	Beam shaping	Which specific footprint shape to use?
Satellite	Satellite Routing / Beam-to-Satellite Scheduling	Which satellite is going to serve each beam and when to handover between satellites?
Gateway	Gateway Routing	Which gateway to route beam's data to?

Table 1.1: Summary of the six different sub-problems for NGSO constellations

In addition to the six sub-problems, it is important to define the time-horizon of each problem. While transmission power is something that can be changed almost in real time as it only involves one entity (the transmission antenna), frequency changes need to be planned ahead of time as we have to synchronize the satellite antenna and its frequency subsystems with the ground antenna and its frequency subsystems. Due to the high amount of synchronization, this Thesis will denote those sub-problems as long-horizon sub-problems. On the other hand, the power sub-problem can be seen as a short-horizon sub-problem.

Even more, depending on the definition of a sub-problem, it can be seen as a long- or short-horizon sub-problem: if we define the frequency problem as determining the frequency and bandwidth per beam or user (by using frequency-division multiple access, FDMA), the problem is long-horizon, if, on the other hand, we only want to split the frequency of a beam into multiple users in time (by using time-division multiple access, TDMA), or to decide when to turn on and off a specific beam in a beam hopping manner, the problem is short-horizon, since we are not changing frequency and, thus, do not need specific synchronization.

In the case of the beam positioning and user grouping, deciding the exact position of the beam is a short-horizon sub-problem, but changing the users from one beam to another is a long-horizon sub-problem (due to the need to adapt the frequency of those users). The beam shape sub-problem is a special case, since in theory it could be a short-horizon sub-problem, but in practice modern satellites cannot usually select whichever footprint they want, but rather they need to decide between a sub-set of all possible footprints. Deciding which sub-set of footprints to use is a long-horizon problem. Deciding the satellite and gateway to serve each beam is always a long-horizon problem due to the need to synchronize the users with the different satellites and gateways. However, choosing the path from the initial or user satellite to the final or gateway satellite can be done dynamically in a short-horizon formulation. Tables 1.2 and 1.3 summarize the different sub-problems depending on the short-/long-horizon definition.

Short-horizon		
Resource	**Sub-problem**	**Brief description**
Transmission power	Power Allocation	Determine the power transmitted to the antenna
Frequency	TDMA	Determine the time-window for each user within the beam
	Beam Hopping	Determine when to turn on/off each beam
Position and user grouping	Beam Placement	Determine where to place the beam
Beam shape	Beam Shaping	Determine which specific footprint shape to use
Satellite	-	
Gateway	Inter-Satellite Routing	Determine the best path from the initial to final satellites

Table 1.2: Short-horizon sub-problems for NGSO constellations

Long-horizon		
Resource	**Sub-problem**	**Brief description**
Transmission power	-	
Frequency	FDMA	Determine the bandwidth for each user within the beam
	Frequency Assignment	Determine the bandwidth and frequency for each beam
Position and user grouping	User Grouping	Determine which users to serve for each beam
Beam shape	Shape Selection	Determine which sub-set of shapes to use
Satellite	Satellite Routing	Determine the time window for each satellite
Gateway	Gateway Routing	Determine which gateway to route each beam to

Table 1.3: Long-horizon sub-problems for NGSO constellations

1.4 Specific objective

This Thesis has two main objectives:

- **To** develop a framework to solve the long-horizon resource allocation problem in satellite communications **by** decomposing this complex conundrum into smaller sub-problems (summarized in Table 1.4) and sequentially solving them **using** state-of-the-art optimization algorithms for each individual resource.

- **To** test the performance of such framework under different operational conditions **by** identifying the main inputs and parameters of the model and varying them **using** sensibility and robustness schemes.

Resource	Sub-problem	Brief description
Transmission power	Power Allocation	Determine the power transmitted to the antenna
Frequency	Frequency Assignment	Determine the bandwidth and frequency for each beam
Position and user grouping	User Grouping	Determine which users to serve for each beam
Beam shape	Shape Selection	Determine which sub-set of shapes to use
Satellite	Satellite Routing	Determine the time window for each satellite
Gateway	Gateway Routing	Determine which gateway to route each beam to

Table 1.4: Sub-problems considered in this work

1.5 Overview

The remainder of this Thesis is organized as follows: Chapter 2 briefly describes the relevant literature on the resource allocation problem for satellite communications and highlights the research gaps that need to be addressed moving forward, Chapter 3 details the decomposition framework developed in this work, identifies the building blocks of such framework and provides a mathematical formulation for each one, and describes the metrics and assumptions that drive the performance and limitations of this method, Chapter 4 outlines different resolution methods for each block, ranging from simple heuristics to state-of-the-art optimization algorithms, Chapter 5 defines the integration and performance tests used in this work for the assessment of the framework and provides a first comparison of the different resolution procedures under different system configurations, Chapter 6 contains the main results of the analysis, where the framework has been evaluated against different inputs and model parameters to understand the effects of each factor on the final allocation, and, finally, Chapter 7 summarizes the main findings of this work and possible directions of future research.

Chapter 2

Literature review

The Resource Allocation (RA) problem in satellite communications has received the attention of many researchers in the recent years. Although most of the work focuses on how to allocate one specific resource or how to solve one specific sub-problem, there is also active research in the joint distribution of two or more resources. Given the increasing complexity, the amount of research decreases as the number of active resources increase. This chapter summarizes the work done on each of the individual sub-problems, as well as the literature combining multiple sub-problems.

2.1 Power Allocation

Power Allocation is, together with the Frequency Assignment, the most studied sub-problem from the six considered in this work. When constrained by amplifiers (i.e., when the number of amplifiers is less than the number of beams), it is shown to be NP-hard and hard to approximate by Aravanis et al. [16]. To solve this problem, the authors in that work propose a two-step multi-objective optimization process to minimize power utilization and unmet demand (UD, called Unmet System Capacity, USC, in that work) using a combination of simulated annealing (SA) and genetic algorithm (GA). They show more than a 50% improvement in system capacity and a 20% reduction in power over uniform allocation techniques in a 37-beam scenario. In a recent study, Efrem et al. [17] show that they are able to obtain similar Pareto-optimal solutions as Aravanis using successive convex approximation in a more resource-restricted environment. Regarding other metaheuristics, Durand et al. [18] use particle swarm optimization (PSO) to solve the Power Allocation problem and show their approach drastically reduces the operational complexity of the implementation, making it more suitable for higher dimensional scenarios, while achieving close-to-optimal solutions.

Using the same 37-beam scenario, Liu et al. [19] propose a game-based approach that outperforms uniform and proportional approaches over incremental and random traffic demands. Their objective function is defined by the least squared error (LSE) between the capacity requested and offered, which may lead to under serving users when the system is overbooked, but ensures a better fairness between users compared to the UD metric. Using the same formulation, Zhang et al. [20] propose a deep reinforcement learning (DRL) architecture and shows a 5.3% improvement over the game-based approach.

Another DRL implementation was proposed by Garau et al. [21]. In this case, the authors used power, UD, and computation time metrics to show that their DRL architecture was able to find solutions with comparable performance at least 1,300 times faster than previous GA approaches, making this formulation more suitable for real-time operations. In a further study, Garau et el. [22] comprehensively compares previous GA, SA, PSO, and DRL approaches together with some hybrid combinations. The authors conclude that the DRL approach performs best in highly time-restricted scenarios, as it gives a close-to-optimal solution in seconds. However, a hybrid PSO-GA is preferred when time is not critical as it outperforms DRL in the long term executions. Nevertheless, this latter algorithm suffers some drawbacks in uncertain scenarios, for which a more robust implementation, i.e., a non-hybrid GA approach, is suggested.

While all these approaches cover the amplifier-constrained problem, the non-amplifier-constrained problem can be solved optimally by manipulating the formulation. Hong et al. [23] transform an LSE formulation of the non-amplifier-constrained problem into a monotonically increasing function using the Lagrangian to include the demand constraints. By doing so, they can obtain the optimal power using a bisection search. Qi et al. [24], arguing that the objective function, although monotonic, may not be differentiable, improve this formulation by exchanging the bisection search with a sub-gradient optimization. In a similar fashion, Wang et al. [25] start from an LSE formulation and the Lagrangian to obtain the dual problem, and solve the remaining formulation using duality theory.

In a slightly different formulation, Destounis et al. [26] propose a user satisfaction metric that only tracks how many users have been satisfied. Given that this approach is again hard to solve, they propose a greedy approach that tries to satisfy as many users as possible and show improvement over static allocations. In a later study, Srivastava et al. [27] improve over this formulation by grouping the users with similar power into clusters and applying a heuristic allocation to the clusters. This method shows significant improvements over the greedy technique.

Finally, some works include external factors into the Power Allocation formulation for a better representation of real operations and an improved optimal point: Lagunas et al. [28] developed a formulation under interference-restricted scenarios as a multi-objective approach, which they later transform into a single-objective formulation. They show that their proposed max-min implementation improves fairness and overall data-rate over previous approaches. On the other hand, Kapsis et al. [29] include rain and other weather-type perturbations to their model and propose a predictive-based water-filling-like method to maximize system capacity. Their algorithm outperforms uniform power distributions over all scenarios considered.

2.2 Short-horizon Frequency Assignment

As explained in section 1.3, the Frequency Assignment problem can be defined in multiple ways, depending on the specific variables considered. This work distinguished between long- and short-horizon problems, depending if the variable being allocated does or does not need further synchronization between different entities, respectively. This section reviews the literature concerning the Frequency Assignment that fall under the short-horizon definition. For the long-horizon sub-problems, refer to the following section.

2.2.1 TDMA

The time-division multiple access (TDMA) sub-problem is defined as how to allocate the different time-slots to the users within a beam so that the demand of each user is met. By comparing it to the bin-packing problem, Park et al. [30, 31] shows that the problem is NP-complete. In order to solve it, they divide the problem into two steps: a first step where they compute the amount of resources available given the atmospheric conditions and predictions, and a second step where they allocate those resources following an heuristic approach. The authors show that their proposed algorithm outperforms classical fitting heuristics, such as first-fit and best-fit, both in terms of time-slot utilization and user satisfaction. In a similar approach, Dong et al. [32] exploit the concept of recursion trees to come up with a different heuristic algorithm. Although they also show improvements over first-fit, it is difficult to assess the performance of both heuristics due to the lack of comparison studies. In a further work, Feng et al. [33] propose a refined heuristic that presents improved performance over Park's heuristic in both voice and stream services when the number of user terminals is greater than the number of frequency channels.

A more recent study by Bejarano et al. [34] includes interference mitigation in their time-slot allocation algorithm. They propose two heuristics, a fast one and a fair one, that present a 20% system throughput increase against pseudo-random allocations. The authors also compare their algorithm with methods for bandwidth manipulation and show that they achieve slightly better solutions in the same computation window. Finally, Lee et al. [35] approach use a different formulation for the TDMA problem and derive a binary integer programming description that can be solved optimally with traditional optimization techniques. By decomposing the problem into smaller sub-problems, they are able to deal with the high-dimensionality of the problem and find an optimal solutions in less than a second.

2.2.2 Beam Hopping

The beam hopping sub-problem is defined as when to turn on/off each beam to avoid interference with nearby beams while achieving the desired data-rate to satisfy the users. It can be seen as the beam-wise version of the TDMA approach. As one of the first studies in this field, Angeletti et al. [36] compare a GEO satellite with classical power and bandwidth flexibilities with a system using beam hopping. In order to solve the inherent time-slot allocation problem, they propose a genetic algorithm approach, and show that the new system can offer up to a 30% capacity increase with respect to the classical approach, but recognize a trade-off between system performance and power consumption. In a posterior study, Anzalchi et al. [37] further refine this results, concluding that a beam hopped system can achieve both improved capacity and reduced power consumption compared to classical systems when both use a single carrier scheme. When using multiple carriers, the beam hopped satellite still performs slightly better than the classical one, but the improvements are reduced.

Focused on a different metric, Han et al. [38] propose an optimization based on user satisfaction, rather than system throughput. In their formulation, the objective is to minimize the delay of packets to the users and are able to derive close-form solutions based on stochastic gradient theory. The authors show how their implementation outperforms other techniques in terms of fairness between the users. Hu et al. [39] refine even further this idea, and develop a DRL algorithm to solve the beam hopping problem. The authors conclude that their method outperforms classical optimization methods, as well as a genetic algorithm approach, giving less overall packet delay. Using a multiobjective formulation combining user delay and system throughput, Wang et al. [40] propose a genetic algorithm approach and show how their algorithm outperforms other classical optimization techniques in both metrics.

Zhang et al. [41] slightly change the problem formulation and group the beams into clusters. Then, they propose a mathematical approach to optimize over the clusters, instead of the beams, which reduced the overall dimensionality of the problem. By comparing different approaches, they conclude that one of their implementations is able to follow closely the user demand and maximizes throughput. Lei et al. [42] also use a different formulation. Their objective function relies on maximizing fairness between the users, so that all of them get similar levels of throughput versus demand ratios. To solve this approach, they propose a fully connected neural network

(NN) that provides a close-to-optimal solution.

2.3 Long-horizon Frequency Assignment

This section encloses all the frequency related sub-problems that require time-consuming synchronization between different entities. Specifically, frequency changes either in bandwidth or central frequency require both, satellite and users, to reach an agreement on the link characteristics.

2.3.1 FDMA or Bandwidth Allocation

The frequency-division multiple access (FDMA) problem, also known as the Bandwidth Allocation problem, is defined as how to subdivide a given frequency range into multiple users or carriers to satisfy the users' demand. Similarly to the Power Allocation works, Park et al. [43, 44] propose an LSE metric that tries to maximize the allocation overall fairness between the users. In their works, the authors derive a mathematical formulation of the problem and define an heuristic that proves to be superior compared to water-filling methods in terms of user fairness. They also show that the water filling method is the one that maximizes capacity. In a later work, Wang et al. [45] further refine the previous formulation with the help of the Lagrangian and show that the problem is convex. Therefore, they are able to obtain an optimal solution that is able to achieve an optimal trade-off between the system capacity and the distribution's fairness compared to uniform and proportional allocations.

Given the nature of the problem where different users or groups compete for the unique pool of resources of the satellite, different game-based approaches have been proposed to solve the Bandwidth Allocation problem. Li et al. [46] propose an asymmetric monopoly model to address the spectrum allocation. In their work, they introduce the concept of *spectrum price* to model the value of the bandwidth and derive their heuristic allocation based on the notion of Bayesian equilibrium. Su et al. [47] propose a Stackelberg differential game. The authors divide the users into two sets, higher demanding and lower demanding, and formulate and simulate their game-based approach with the bandwidth as the available resource. They show optimality by achieving Nash equilibrium. Wang et al. [48] also use a game-based approach that relies on the Nash equilibrium to achieve optimality. In this work, however, the authors propose an iterative learning algorithm that proves to outperform other methods and achieve close-to-optimal solutions in all scenarios considered.

Other works on this problem change the original formulation to better adapt the operations reality. Bisio et al. [49] describe the problem as a combination of virtual and physical entities that compete for resources. Their formulation involves one objective function per virtual-physical relation, which derives into a multi-objective problem. The authors propose a minimum distance-based solution that proves to save power with respect to other approaches. Liu et al. [50] define a user satisfaction metric that includes traffic priority and propose a bee colony-based algorithm as a solution. The authors define all the necessary elements for the bee colony optimization and show that their approach is superior in all metrics compared to a fair water-based and a utility maximization methods. Kawamoto et al. [51] include inter-beam interference and show how to mitigate it while solving the Bandwidth Allocation problem. They compare their heuristic approach to two commonly used simple algorithms and show that their method improves both system capacity and flexibility under the four scenarios considered. Abe et al. [52] go a step beyond and include control inputs in their implementation. Their formulation is based on model predictive control and a sparse optimization and proves to reduce the amount of traffic loss within the simulated day.

2.3.2 Frequency Assignment

The Frequency Assignment problem is defined as what frequency range to assign to each beam to meet the users' demand. When the bandwidth of the beam is fixed, it is also known as the beam coloring problem or graph coloring for satellite communications. When considering co-channel interference, it is proven to be NP-complete by Mizuike et al. [53]. To deal with this complexity, the authors in that work propose an heuristic lexicographic minimization to achieve a solution. Throughout the years, many researchers have proposed solutions to deal with this particular problem as formulated in Mizuike's work, most of which rely on modern optimization techniques, which do not guarantee optimality, but achieve *good* solutions efficiently. Funabiki et al. [54] describe a gradual NN in combination with heuristic approaches for a faster convergence. The authors describe in detail the implementation and the simulations and argue how their approach can also be applied to similar problems. Salcedo et al. [55, 56] propose a Hopfield NN combined with simulated annealing (SA) and GA, respectively. Both solutions present better scalability properties and improved performance compared to previous proposals.

Wang et al. [57] apply the concepts of stochasticity and noise to perturb static states for a better solution. By combining these concepts with the Hopfield NN, Wang et al. [58] prove how their multi-start stochastic Hopfield NN proposal is able to escape local minima. By a detailed comparison with previous techniques, the authors show how their approach outperform all other NN presented so far for the Frequency Assignment Problem. Although most of the approaches presented so far rely on NN as the core of the computation, other artificial intelligence techniques have also shown the potential to perform well. Salman et al. [59] propose several differential evolution algorithms combined with heuristics that make a better use of the domain knowledge of the problem. The authors prove how their implementation provide solutions with similar or better quality in less time. Wang et al. [60] go a step beyond and propose a multi-objective approach that expands the previous differential evolution implementation. By comparing their method with previous techniques using the same benchmark, they prove the superiority of their algorithm.

By transforming the non-linear interference constraints into cumulative interference, Houssin et al. [61] develop an integer linear programming (ILP) formulation of the problem. This allows the authors to solve the problem using off-the-shelf mathematical solvers to find an optimal solution. They also propose a greedy approach that provides comparable solutions to the ILP for low number of users, but with significantly lower computation time. Focusing on mobile networks, Hu et al. [62] propose a channel-based approach to the Frequency Assignment problem and develop a DRL method to solve it. They show improved results over a 37-beam grid compared to other techniques. As a final note on the Frequency Assignment problem, it is important to remark that this issue is not limited to space communications. Many approaches have been proposed for the general Frequency Assignment problem (i.e., not specialized in satellite communications), including a variety of applications and implementations. Aardal et al. [63] summarize more than 100 methods and techniques proposed for the problem.

2.4 User Grouping and Beam Placement

As we move outside of the Power Allocation and Frequency Assignment sub-problems, the amount of literature dwindles. This section focuses on the sub-problems affecting the relation between users and beams (a.k.a. User Grouping) and the position of the beam (a.k.a. Beam Placement).

2.4.1 Beam Placement

The Beam Placement problem is defined as where to place the center of each beam, or where to direct each beam, such that the demand of the users is met. Xu et al. [64] propose a tracking-based system for LEO mobile constellations in which the beam follows the user until the connection finishes. This allows for significant savings in power while improving the quality of service for the user. Ivanov et al. [65] develop a method that creates a virtual and optimized grid over the Earth. The existence of this grid allows for a reduced power allocation scheme and a reduced number of handovers that only add overhead to the network. In addition, the low complexity of their approach eases the implementation in real operations.

2.4.2 User Grouping

The User Grouping problem (a.k.a. Beam Arrangement problem) is defined as how to group the users into beams such that the demand requirements of the users are met. When considering the beam shapes as fixed sized circles, the problem can be transformed into the Minimum Geometric Disk Cover, which Fowler et al. [66] proved to be NP-hard. To deal with this complexity, Yao et al. [67] propose a modified version of the heuristic K-means algorithm in which they include load balancing considerations (i.e., how to balance the users between the beams). Their algorithm presents improvements against random allocations in terms of throughput and system capacity. In a previous publication [68], this Thesis' author developed a formulation for the User Grouping problem that considers the real footprint of the fixed-shape beam, which is an ellipsoid due to the projection of the conic beam over the Earth. Due to the complexity of the problem, the authors in that work develop a GA implementation to solve the problem, and prove how their proposal outperforms previous techniques showing improved capacity while achieving a significant power reduction.

2.5 Beam Shaping

Since the distinction between the Beam Shaping problem and the Shape Selection problem is not as clear as in the previous sub-problems and most approaches deal with both at the same time, this section will cover both simultaneously. In general terms, the Shape Selection decides which sub-set of beam shapes to use in the entire constellation, while the Beam Shape decides the beam-shape relation. The objective of both problems is still to meet the requirements of the users.

As one of the first works on this field, Sherman [69] develops a formulation based on beam directivity and ring-shaped circular layouts. They propose a GA implementation that decides how many beams to place on each ring and their shape and prove that it achieves the desired margin. Zhao et al. [70] further refine this approach by adding a dual coding to the GA to achieve better results in a more complex model. In a different approach, Okello et al. [71] start from a uniform beam distribution and propose an heuristic that iteratively modifies the position and shape to be able to increase the balance between beams while maintaining the throughput.

Qian et al. [72] also start from a uniform grid Beam Placement with a fixed shape, and propose a shape modification mechanism that aims to balance the load between beams. The authors develop and alternative configuration to the standard grid where the central beam shrinks in size, while the adjacent beams grow, achieving a more spread demand between the beams. This allows for an increased system throughput. Wenqian et al. [73] propose a very similar approach where they iteratively modify the shape of the beam until they achieve maximum performance. They show how this approach outperforms conventional Beam Shaping techniques independently on the systems characteristics. To avoid wasting the resources allocated to a beam, Zhang et al. [74] develop a very similar approach where they keep increasing the size of the central beam and serving the additional users with the exceeding resources, until there are no more resources left. They show that their proposal increases user satisfaction when the amount of traffic is low given the reduced complexity of the layout.

Finally, Camino et al. [75] formulate the problem as a mixed-integer linear programming optimization problem in which the exact placement and shape of the beams are decision variables. The objective is to maximize the balance between the different beams while serving the maximum number of users. Although this formulation can be optimally solved with off-the-shelf mathematical solvers, the authors show that the problem grows exponentially with the number of beams and users, which has scalability issues in high-dimensional scenarios.

2.6 Satellite Routing

In order to accomplish the communication goal, we need to ensure that the data being transported can reach its destination, which can be seen as a routing problem involving users, satellites, and gateways. This work divides this sub-problem into two: the Satellite Routing sub-problem, i.e., starting from a user, deciding which satellite is going to serve that user, and the Gateway Routing sub-problem, i.e., once the data has reached the satellite, decide which gateway to use and how to transport it there. This section covers the literature on the Satellite Routing sub-problem, refer to the next section for the Gateway Routing sub-problem.

The Satellite Routing problem, most known in literature as the Satellite Handover problem or Satellite Scheduling problem, consists of deciding which satellite is going to serve each user at each point in time. Note that the difficulty of this problem scales with the number of satellites in line of sight (LoS), which makes it trivial for GEO satellites or sparse MEO constellations where only one satellite is visible, and quite complex in LEO mega-constellations where each user sees more than 10 satellites at all times.

A simple and fast technique for Satellite Routing is to always reach for the nearest satellite. Krewel et al. [76] presents a detailed comparison of several heuristics, including going to the nearest satellite, as well as going to the satellite with maximal signal power, longest visible satellite, or least congested satellite. They show that the different techniques offer different trade-offs in terms of waiting time and network load, and propose the last method for multimedia services given that it achieves the lowest network load. Papapetrou et al. [77, 78] propose a different heuristic based on channel reservation that includes the dynamic of the Earth and the satellite movements to improve the user satisfaction. To reduce the complexity that involves dealing with a large number of users, Zhu et al. [79] develop a grouping-based approach, where they cluster the different users based on similar characteristics and assign handover times to the groups.

In a slight different approach, Wu et al. [80] pre-compute the time-window for each satellite and develop a graph based formulation to solve the problem. The authors explain how all the other techniques can be easily included in their graph-based approach for a better solution, and show how well-known algorithms for path planning already yield good results on the considered problem. He et al. [81] go a step beyond and develop a DRL framework that relies on the network load to achieve improved user satisfaction. The authors compare their approach with other state-of-the-art methods and show large improvements over all metrics considered, specially user satisfaction. In a previous work [82], this Thesis' author developed a formulation that aims to balance the load for thousands of users distributed across the world in a global constellation. The implementation relies on PSO and shows improvements in terms of the total number of constraints a load balancing between the regions.

The problem of Satellite Routing is also common in areas close to the satellite communications field. As an example, this problem is very prominent in imaging satellites, where the visibility windows are varying and the data load is large. Under this conditions, authors have proposed many different implementations: Pemberton et al. [83] develop a constraint satisfaction-based approach, Xhafa et al. [84] and Kolici et al. [85] adapt a generic GA to solve this specific problem, Zhuang et al. [86] use an artificial bee colony (ABC) optimization, Tharmarasa et al. [87] tailor a markov decision process to suit their necessities, and Chen et al. [88] rely on a mixed ILP (MILP). A survey of different methods for Satellite Scheduling is presented by Xhafa et al. [89]. Finally, this problem is also relevant in mobile satellite networks, where many different approaches and techniques have been proposed. Chowdhury et al. [90] summarize some of the presented approaches.

2.7 Gateway and Inter-Satellite Routing

The previous section summarized the works on how to transport the data from the user to an initial satellite. In this section, we deal with the subsequent problems: how to decide the final satellite (the Gateway Routing problem, i.e., how to choose the satellite that will serve as the ground connection to the user data), and how to reach that satellite from the initial one (the Inter-Satellite Routing problem).

2.7.1 Gateway Routing problem

The Gateway Routing problem is defined and deciding which satellite and gateway to use that will serve as the final destination of the user data in the satellite network. This can be requested from the user, in which case the problem is trivial, or can be dynamically assigned by the network, for example when the ground station serves as the connection to the internet. Since in most cases the users have a pre-allocated gateway, the amount of literature on this problem is scarce. The only example on the literature that could be found is Crosnier et al. [91]. In that work, the authors develop a load balancing heuristic that divides the data rate between the different gateways to increase the capacity of the network while improving user satisfaction. Their results show improvements over random and simple allocations.

2.7.2 Inter-Satellite Routing

The Inter-Satellite Routing problem is defined as, knowing the starting and final satellite of the user-gateway connection, decide which route to follow in the satellite network. Note that this problem is trivial when the initial and final satellites are the same, which is generally the case in MEO and GEO constellations. In general, it can be transformed into a common network-flow problem with variable topology, for which multiple solutions have been proposed. Following such transformation, Werner [92] develops a virtual topology formulation, and a dynamic heuristic solution to deal with the network variability. The author proves the feasibility of the new design and how it can reduce delay. Going a step beyond, Sigel et al. [93] develop an ant colony optimization (ACO) to reduce packet delays. Over a series of detailed simulations, the authors show how their implementation outperforms other common path planning algorithms and are able to reduce delay significantly. Using the same metric, Li et al. [94] propose a Hopfield NN in combination with SA to achieve minimum delay. The authors prove to achieve a lower delay than previous approaches and a better optimum. Similarly, Rao et al. [95] improve over previous results with their GA adaptation. They prove to achieve better solutions not only in packet delay, but also in other user satisfaction metrics such as packet dropping probability. In a recent study, Rajagopal et al. [96] propose a beetle swarm optimization combined with extreme learning that relies on traffic prediction to improve performance. This complex implementation proves to reduce delay even further and achieve state-of-the-art optimality.

Sun et al. [97] propose adopting a throughput-centric view, instead of delay-centric, to determine the goodness of the implementations. The authors then develop three heuristic algorithms that prove to increase the system's capacity, especially when the satellites have buffering capabilities. In a different approach, Rao et al. [98] present a formulation based on balancing the throughput throughout the network. They propose an implementation that relies on agents transporting the information through the network and direct the flow based on the obtain knowledge. This concept is also used by Liu et al. [99], where nearby satellite exchange information of the network and the transmission algorithm is a probabilistic-based implementation that evolves with this information. The authors prove that this approach reduces complexity and increases performance in all scenarios considered. Usign a combination of all previous metrics, including transmission delay, system capacity, and network flow balance, Zhao et al. [100] develop an ACO implementation that aims to merge and outperform all previous formulations. With detailed simulations, the authors show how this approach presents a balanced result and improved trade-off in all metrics.

Finally, some works extend the original routing problem to better adapt real-case scenarios. Wang et al. [101] propose an adapting routing scheme for a three-tier GEO-MEO-LEO system. They improve over existing path-planning algorithms and show better results than other commonly used methods. Fraire et al. [102] assess not only the performance of their algorithm, but also the robustness and sensibility of routing networks. The authors provide insightful analysis on different failure rates and how the algorithms perform in each case. As a summary, Alagoz et al. [103] comprise many more different routing approaches that have been proposed throughout the years, and the idiosyncrasies of each one.

2.8 Joint problems

As mentioned before, although most of the literature on the RA problem focuses on single sub-problems or specific instances, there is also research in joint problems. This section covers the works combining multiple sub-problems. As emphasized at the beginning of this chapter, the amount of literature rapidly decreases as the amount of problems increase.

2.8.1 Research on two sub-problems

Power Allocation + TDMA

Wang et al. [104] propose a formulation for the joint Power Allocation and timeslot allocation (TDMA) problem that maximizes system capacity under fairness constraints. The authors then develop a hybrid GA - PSO implementation that solves the joint problem and achieves increased throughput and fairness compared to simple heuristics in scenarios with up to 45 users.

Power Allocation + Beam Hopping

For the joint Power Allocation and Beam Hopping problem, Alberti et al. [105] develop a formulation that includes weather impairments and maximizes met demand. The authors present an iterative algorithm to provide an refined solution and compare their approach against traditional systems, which they prove to outperform. Lei et al. [106] propose two different objectives to approach the problem: one based on the difference with the desired demand and the other based on the relative fairness between the users. The authors develop a mathematical formulation for both and make use of the Lagrangian to solve the problem. Their approach allows them to obtain a better optimum point with reduced power usage. Using the same objective functions and mathematical formulation, Shi et al. [107] increase the complexity by introducing non-uniform channels to the problem. To solve this new expression, the authors divide the two sub-problems and propose a mathematical solution for each one, which proves to outperform classical methods.

Wang et al. [108] propose a low-complexity solution by grouping the beams into clusters and then applying a simple mathematical optimization procedure to resolve over the clusters. They show to improve over non-optimized schemes, but it is hard to assess their performance against previous studies due to the lack of comparisons. Finally, Wang et al. [109] develop an extensive mathematical formulation for this joint problem. Due to the complexity of the search space, they propose a heuristic implementation that proves to behave well and provides close-to-optimal solutions with reduced complexity.

Power Allocation + Bandwidth Allocation

The joint problem of Power Allocation and Bandwidth Allocation is one of the most studied, given their intrinsic dependence in the link budget equations. Cocco et al. [110] develop a single-objective formulation that aims to maximize the user satisfaction while meeting user requirements. The authors show that this problem is non-convex and present a SA strategy to achieve a solution. They show how this approach is able to follow closely the requested demand compared to more simple techniques. Following a different framework, Zhong et al. [111] formulates the problem as a bargaining game, in which the different beams and users compete for resources. This method proves to achieve a better trade-off between total capacity and user fairness compared to other approaches. The metric of UD that was widely under the Power Allocation framework is reused in the joint problem in the work of Paris et al. [112]. The authors motivate this metric due to the non-symmetry of user penalties when over- or under-serving. Based on this and the complexity of the problem itself, they propose a GA to solve the model. Within their extensive results, they show that including the bandwidth flexibility in the problem further helps to increase system capacity and yield better results.

Jia et al. [113] formulate the joint problem including inter-beam interference and use the LSE metric to achieve maximum capacity fit. The authors develop a detailed mathematical formulation and use the Lagrangian and dual to develop an iterative algorithm that reaches a satisfactory solution. Liao et al. [114] propose a more complex framework with a dual objective that aims to maximize fairness and system capacity. To deal with this complexity, they introduce a DRL model based on a large training set that proves to achieve good solutions and match traffic demand.

Power Allocation + Frequency Assignment

Jahn [115] develops an extensive work where he investigates how to apply graph theory to the joint Power Allocation and Frequency Assignment problem. He proposes different sub-graph families for each one and reports their impact in both MEO and LEO configurations. He concludes that the problem is well framed under graph theory and that known graph optimization algorithms already obtain good solutions on this problem. Under a different framework, Lei et al. [116] develop a formulation based on carrier allocation. The authors introduce the demand requirements as constraints and present a non-convex mathematical formulation, which they solve using an iterative algorithm. In a similar framework, Abdu et al. [117] divide the problem into two, and present a successive convex approximation to solve each one of them. The authors show extensive comparative results and prove that their approach is able to match closely the user requirements, which results in an increase of total capacity and a reduction in unmet demand. Vidal et al. [118] propose a SA implementation for the joint problem, which allows them to include non-linear elements in their formulation. They show that their approach scales well and outperforms previous techniques in terms of total throughput.

Power Allocation + Beam Placement

In a series of works, Choi et al. [119–121] propose a formulation to optimize Power Allocation and Beam Placement at the same time using an LSE fairness metric. In their multiple papers, the authors develop a realistic satellite model with beam steering capabilities and a mathematically found solution that can turn off or redistribute certain beams to maximize the objective. They show how this combined approach can increase the total capacity and serve more users. In a different framework, Takahashi et al. [122–124] allow beam movement to achieve maximum throughput. They develop a mathematical formulation that directly yields the optimum arrangement and show how this method outperforms the approaches that deal with the individual Power Allocation and Beam Placement problems.

Power Allocation + Beam Shaping

In a detailed study, Schubert et al. [125] refine a formulation that combines Power Allocation with beam-forming capabilities. The authors also show how to obtain the global optimum with an algebraic analysis on the formulation and an iterative approach that converges to the desired value. They show how this approach outperforms conventional beam-forming techniques and is able to serve more users with the same power limitation.

Power Allocation + Satellite Routing

The combined problem of Power Allocation and Satellite Handover in LEO is studied by Liu et al. [126]. Similarly to approaches on the individual problem, the authors develop a graph-based formulation for the handover scheme in which they include multi-satellite connections and power management. The authors show that this method drastically increases the systems capacity over time and a achieves better balance in the satellite network. Abdelsadek et al. [127] investigate the possibility of including handover management in mobile satellite services. The authors develop an extensive mathematical framework for this problem, which they convert into MILP. To deal with the increased complexity, they propose a GA implementation that increases the overall throughput compared to simple heuristics.

Bandwidth Allocation + Beam Shaping

For the joint Bandwidth Allocation and Beam Shaping problem, Kyrgiazos et al. [128] develop a formulation where the implementation needs to decide between two possible beam sizes, as well as the allocated bandwidth, for each beam. With an heuristic iterative algorithm, the authors show that their method increases the system capacity by 11% compared to constant shape allocations in a scenario with 200 beams.

Frequency Assignment + Beam Placement

In a series of works, Kiatmanaroj et al. [129–131] develop a refined mathematical formulation for the Frequency Assignment problem based on ILP where they allow the beams to move if that results in increased benefits. The beams are only allowed to move slightly and never beyond some margin that ensures user coverage, but this is enough to increase the number of covered users given the frequency constraints. As a comparison, the authors develop an heuristic algorithm and prove that the ILP implementation is able to achieve improved results in less computation time.

Frequency Assignment + User Grouping

In a previous work [132], this Thesis' author developed a formulation for allocating beams to users and frequency to beams in a large LEO constellation. The proposed solution in that work was based on a series of heuristics that proved to reduce the amount of beams in the system and increase the frequency reuse, which allows to serve more users. The authors showed results for the SpaceX constellation, for which almost 10.000 beams were needed, and around 97.7% of those could be assigned a frequency.

Frequency Assignment + Beam Shaping

The joint problem of Frequency Assignment and Beam Shaping is studied by Camino et al. [133]. The authors propose a two step approach where first the beam shape is selected, and then the frequencies are chosen. The first step is performed with a greedy method, while the second one is based on a modified first-fit search. The authors show the feasibility of the implementation and their results over Africa. In a different approach, Zhong et al. [134] investigates how to obtain a uniform beam layout with non-uniform shapes. The authors transform the problem into a circle fitting problem (i.e., how to fit circles inside circles), which are allocated a frequency *a posteriori*. They show how their proposal meets the imposed system requirements.

Frequency Assignment + Satellite Routing

Based on waiting queues and channel reservations, Wan et al [135] analyzes the joint problem of Frequency Assignment and Satellite Handover for mobile satellite networks. The authors propose a new scheme that aims to ensure a higher quality of service compared to commonly used techniques in handover-only schemes. Based on a single comparison, the authors conclude that their approach can help reduce the user overload and increase the systems performance.

User Grouping + Beam Shaping

Given the close relation between both problems, the joint User Grouping and Beam Shaping algorithm has been often studied in the last years. Alinque [136] proposes a greedy starting configuration, followed by a gradient descent optimization to reorganize the beam layout. The author proves to achieve a higher gain and lower loss per beam, which leads to an increased throughput. In a different approach, Liu et al. [137] develop a mathematical formulation to maximize the total capacity. Given the complexity of this approach, the authors propose a two-step heuristic approximation to achieve a solution, which proves to outperform standard techniques in terms of system capacity and number of beams. Tang et al. [138] elaborate their formulation on the joint problem based on a combination of the individual problems. Similarly to the User Grouping methods, the authors assume circular beam shapes and transform the problem into the Minimum Geometric Disk Cover, which they propose a p-center algorithm to solve it. On top of this, the authors allow the radius of the circle to vary, so that the performance is maximized. Based on the comparison of their system with classical methods, the authors prove to achieve increased throughput and reduced user delay in all cases.

In a more mathematical approach, Honnaiah et al. [139] propose Voronoi maps and ellipsoidal shapes to adapt the beam layout to the user demand. This approach provides a higher antenna gain which resolves in higher throughput. By comparing it to fixed beams, the authors show that it also increases the fairness between beams and users. In a similar framework, Camino et al. [140] develop three MILP approximations of the problem, which can be solved with commercial mathematical solvers. The authors show how the implementations improve over classical layouts and are able to increase the total coverage.

2.8.2 Research on more than two sub-problems

Power Allocation + Beam Hopping + Bandwidth Allocation

The joint problem combining Power Allocation, Beam Hopping, and Bandwidth Allocation has been studied by Tian et al. [141]. The authors detail a mathematical formulation where the objective is to maximize throughput. Given that this approach is hard to resolve, they propose a greedy algorithm there only the maximum demanding beams are served. With a comparison with other simpler methods, the authors conclude that their approach is superior given that it provides higher throughput. Although it proves that higher flexibility allows for increased performance, the authors in this work oversee some crucial aspects of satellite operation: 1) user terminals cannot adapt to bandwidth changes instantaneously due to the synchronization required, 2) satellites have frequency reuse mechanisms that allow operators to use the same frequency several times, which complicates the bandwidth allocation and requires from interference mitigation mechanisms, 3) modern satellites can manage hundreds or even thousands of satellites, while this work presents results for up to 12.

Power Allocation + Beam Hopping + Frequency Assignment

To deal with the joint Power Allocation, Beam Hopping, and Frequency Assignment problem, Zuo et al. [142] present a 3-level decomposition that deals with each problem individually. The authors propose a combination of heuristics and mathematical optimization to solve the different sub-problems and prove how this approach can match the users' requirements in terms of system throughput. They show results for up to 20 beams. In a similar decomposition framework, Tang et al. [143] develop a mathematical formulation for each of the individual sub-problems. They present comparison results between a beam-hopping satellite and a multi-beam satellite under different interference-constrained scenarios. The authors model the multi-beam satellite with up to 91 beams. It is important to note that, in both works, the authors assume that the satellite needs negligible time to send the new configuration to the users, and that the user terminals need negligible time to adapt, which may not be representative of real operations. In addition, the results presented deal with a low number of beams, the satellites are assumed to be of low complexity with few spectrum divisions, and the scalability of the different sub-problems is not discussed, which poses questions on the validity of this approach for high-dimensional scenarios.

Power Allocation + Frequency Assignment + User Grouping

Deng et al. [144] study the joint Power Allocation, Frequency Assignment, and User Grouping problem. Their implementation relies on two steps: first, the users are grouped based on a heuristic clustering algorithm, and second, the joint Power Allocation and Frequency Assignment problem is resolved optimally by applying the Lagrangian to the mathematical formulation. The authors show how this approach increases the capacity of the satellite compared to more inflexible allocations. Within this work, there are some assumptions that could make the implementation in real operations hard: 1) only fixed-size frequency channels can be assigned at a time, which eases the formulation, but does not reflect the reality that satellites can use multiple channels for a single beam at a time, 2) only GEO satellites are analyzed, which makes difficult to assess the extension of the approach to MEO or LEO orbits, where beam footprints and satellite visibility is constantly changing, and 3) the simulation provided analyzes a low number of beams, which poses questions to the scalability of the presented algorithms.

A similar divide and conquer approach is presented by Angeletti et al. [145]. In their extensive report, the authors explain how to obtain a combined solution by dividing the joint problem into the individual sub-problems, solving them using a dense algebraic formulation, and obtain the final allocation by union of the singular solutions. The User Grouping is solved by adjusting the position of all the beams at the same time, the Power Allocation is resolved optimally with mathematical transformations, and the Frequency Assignment is transformed into a color reuse allocation for easier interpretation. By showing simulations with thousands of users, the authors prove the feasibility of their design in high-dimensional scenarios. However, similar to the previous work, the color reuse transformation only allows for fixed bandwidth for all beams, and the results are only shown for static GEO satellites.

Power Allocation + Bandwidth Allocation + Frequency Assignment + Beam Placement

The work of Lagunas et al. [146] solves two different joint problems in a single paper: the joint beam forming and carrier allocation problem for the satellite uplink, and the joint carrier, power, and bandwidth allocation for the satellite downlink. For the former, the authors divide the joint problem into the individual sub-problems: the Beam Placement is solved using a linear formulation, and the Frequency Assignment is solved by a beam-per-beam resolution. For the latter, they propose an optimal solution to the joint carrier and power allocation at the beam level, followed by an optimal division of the allocated frequency into multiple users. To prove the feasiblity of this approach, the authors show results for a GEO satellite over Europe with up to 250 beams for both uplink and downlink. Due to the clear separation in the paper, it is difficult to assess how the resolution on the uplink affects the resolution on the downlink and vice versa. For instance, questions like how does beam forming techniques affect the downlink transmission remain unanswered. In addition, the extension of this work to MEO or LEO is also not immediate. Time-varying beam footprints and visibility windows poses additional constraints with unclear resolution. Finally, modern satellites are expected to have thousands of beams, which is an order of magnitude higher of what is explored in this work. The high-dimensionality is an additional factor that needs to be resolved moving forward.

2.9 Literature summary

The previous sections presented a dissected literature review on each on of the specific sub-problems within the more general resource allocation problem. Table 2.1 presents a summary of the different works and their focus. As shown, although there is vast research on the individual sub-problems, there is also some studies combining multiple problems. However, the amount of literature decreases proportionally to the number of sub-problems considered. The research is scarce when considering three sub-problems, and almost non-existent when considering more. In addition, these higher-order works propose solutions for highly restricted scenarios. As highlighted in the previous section, the solutions proposed tend to consider only GEO satellites, reduced flexibility, such as fixed bandwidth, and/or low number of beams. From this, we can observe a clear gap in existent work:

- How to include more than three sub-problems when solving the RA to achieve an improved solution.

- How to extend previous works to more dynamic LEO/MEO constellations
- How to include further flexibility within the sub-problems, such as uneven or variable bandwidth distributions
- How to adapt the implementations for realistic high-dimensional scenarios.

The purpose of this work is to close part of this gap by developing a divide and conquer-based approach for the joint problem combining all of the long-horizon sub-problems within the RA problem plus Power Allocation and excluding Beam Shaping (i.e., the joint problem combining Power Allocation, Frequency Assignment, User Grouping, Satellite Routing, and Gateway Routing). Each problem is considered as a building block within the wall that is the RA problem. This allows or a modification the formulation of individual sub-problems without altering the overall structure, as long as the interfaces within sub-problems remain the same. The implementation proposed to solve the joint problem rises from combination of mathematical optimization plus metaheuristic algorithms, which are the two most common techniques in literature as shown in Table 2.2. To provide further insight, this solution is then tested in very high-dimensional scenarios with tens of thousands of users on a simulated equatorial MEO constellation. In addition, an extended study on the sensibility and robustness of the proposed method is presented.

Works	Power Allocation	TDMA	Beam Hopping	Bandwidth Allocation	Frequency Assignment	Beam Placement	User Grouping	Beam Shaping	Satellite Routing	Gateway Routing	Inter-Satellite Routing
[16–29]	$\mathcal{X}$										
[30–35]		$\mathcal{X}$									
[36–42]			$\mathcal{X}$								
[43–52]				$\mathcal{X}$							
[53–62]					$\mathcal{X}$						
[64, 65]						$\mathcal{X}$					
[67, 68]							$\mathcal{X}$				
[69–75]								$\mathcal{X}$			
[76–88]									$\mathcal{X}$		
[91]										$\mathcal{X}$	
[92–101]											$\mathcal{X}$
[104]	$\mathcal{X}$	$\mathcal{X}$									
[105–109]	$\mathcal{X}$		$\mathcal{X}$								
[110–114]	$\mathcal{X}$			$\mathcal{X}$							
[115–118]	$\mathcal{X}$				$\mathcal{X}$						
[119–124]	$\mathcal{X}$					$\mathcal{X}$					
[125]	$\mathcal{X}$							$\mathcal{X}$			
[126, 127]	$\mathcal{X}$								$\mathcal{X}$		
[146]				$\mathcal{X}$		$\mathcal{X}$					
[128]				$\mathcal{X}$				$\mathcal{X}$			
[129–131]					$\mathcal{X}$	$\mathcal{X}$					
[132]					$\mathcal{X}$		$\mathcal{X}$				
[133, 134]					$\mathcal{X}$			$\mathcal{X}$			
[135]					$\mathcal{X}$				$\mathcal{X}$		
[136–140]							$\mathcal{X}$	$\mathcal{X}$			
[141]	$\mathcal{X}$		$\mathcal{X}$	$\mathcal{X}$							
[142, 143]	$\mathcal{X}$		$\mathcal{X}$		$\mathcal{X}$						
[144, 145]	$\mathcal{X}$				$\mathcal{X}$		$\mathcal{X}$				
[146]	$\mathcal{X}$			$\mathcal{X}$	$\mathcal{X}$						
This work	$\mathcal{X}$				$\mathcal{X}$		$\mathcal{X}$		$\mathcal{X}$	$\mathcal{X}$	

Table 2.1: Summary of the literature's scope

Works	**Dimensionality***			**Proposed solution**				
	≤ 10	$10 < x \leq 500$	> 500	Heuristic	Mathematical Optimization	Meta-heuristic	Machine Learning	Game-based
[33, 51, 64, 71–74, 76, 97, 135]	$\mathcal{X}$			$\mathcal{X}$				
[17, 23–25, 28, 29, 38, 45, 49, 52, 80, 83, 108, 113, 117, 122–124]	$\mathcal{X}$				$\mathcal{X}$			
[39, 81]	$\mathcal{X}$						$\mathcal{X}$	
[46, 47]	$\mathcal{X}$							$\mathcal{X}$
[48]	$\mathcal{X}$						$\mathcal{X}$	$\mathcal{X}$
[26, 27, 30–32, 34, 43, 44, 53, 65, 67, 77–79, 91, 92, 98, 99, 105, 109, 116, 126, 128, 133, 137, 138, 141]		$\mathcal{X}$		$\mathcal{X}$				
[35, 41, 61, 75, 87, 101, 106, 107, 115, 119–121, 125, 129–131, 134, 139, 140, 143, 146]		$\mathcal{X}$			$\mathcal{X}$			
[142, 144]		$\mathcal{X}$		$\mathcal{X}$	$\mathcal{X}$			
[16, 18, 36, 37, 40, 50, 59, 60, 69, 70, 84–86, 93, 95, 100, 104, 110, 112, 118, 127]		$\mathcal{X}$				$\mathcal{X}$		
[20, 21, 42, 54, 57, 58, 62, 114]		$\mathcal{X}$					$\mathcal{X}$	
[22, 55, 56, 94, 96]		$\mathcal{X}$				$\mathcal{X}$	$\mathcal{X}$	
[19, 111]		$\mathcal{X}$						$\mathcal{X}$
[132]			$\mathcal{X}$	$\mathcal{X}$				
[88, 136, 145]			$\mathcal{X}$		$\mathcal{X}$			
[68, 82]			$\mathcal{X}$			$\mathcal{X}$		
This work			$\mathcal{X}$		$\mathcal{X}$	$\mathcal{X}$		

Table 2.2: Summary of the literature's characteristics. *Dimensionality refers to the cardinality of the characteristic set (i.e., number of beams for frequency-related sub-problems, number of users for User Grouping, etc)

Chapter 3

A Divide and Conquer approach to the Resource Allocation Problem

The first step towards a successful resolution of any issue is a genuine and robust definition of the problem. This Chapter describes the RA problem for satellite communications as a composition of multiple sub-problems and outlines a framework to approach such complex conundrum. Subsequently, each sub-problem is dissected and analyzed, and a mathematical formulation for each one is detailed. Finally, this Chapter concludes summarizing the overarching assumptions and challenges that govern the described model. Note that this Chapter is agnostic to the resolution procedure and is centered around defining the different sub-problems. For specific solutions, refer to the next Chapter.

3.1 On the Resource Allocation framework

As highlighted in Section 1.3.3, the Resource Allocation (RA) problem for satellite communications can be defined as finding the most efficient resource distribution that meets the users' requirements. It can be subdivided into six different sub-problems: Power Allocation, Frequency Assignment, Beam Placement and User Grouping, Beam Shaping, Satellite Routing, and Gateway Routing. Furthermore, each sub-problem can be analyzed from a long-horizon perspective (i.e., when the changes in allocation must be synchronized between different entities), and a short-horizon perspective (i.e., when changes do not need to be synchronized). Long-horizon problems need more time for planning and deploying, and can be assimilated with long-term plans. Short-horizon problems have a faster deploy time and are the ones that can be dealt with in real-time and solved via continuous operation. This Thesis focuses on the long-horizon problem: how to obtain a static feasible plan that can satisfy the users' requirements. Specifically, the objective is to develop a framework for the joint Frequency Assignment, User Grouping, Beam Shaping, Satellite Routing, and Gateway Routing and the necessary formulation to solve it. In addition, this work discusses the role of and presents a formulation for the Power Allocation problem, as it will serve as a measure of performance for the defined system.

The long-horizon RA problem is defined in this work as assigning the users into beams and finding the shape, frequency, bandwidth, handover time, and offload gateway for each beam so that the users' demand is met and the resources are allocated efficiently. In order to achieve this goal, the problem is decomposed into smaller interconnected sub-problems: the joint Beam Placement and Beam Shaping problem, the Satellite Routing problem, the Gateway Routing problem, and the Frequency Assignment problem.

User Distribution	User position and demand				
	User Grouping & Beam Shaping	Possible visibility windows	Demand per beam	Demand per beam	User grouping and beam shape
	Beam footprint limits	**Satellite Routing**	Real visibility windows	Beam restrictions	Handover times
	Gateway-user restrictions	Gateway visibility window	**Gateway Routing**	Gateway beam restrictions	Offload gateways
	Beam frequency consumption	Beam frequency overlapping	Gateway frequency load	**Frequency Assignment**	Central frequency and bandwidth
	Beam power consumption	Exact beam position	Gateway power load	Beam power consumption	**Operations**

Table 3.1: N^2 diagram on the flow of information on the long-horizon RA problem. The elements on the diagonal correspond to the different sub-problems. The elements on the upper diagonal denote the flow of information from the element directly to the left to the element directly below. The elements on the lower diagonal denote the flow of information from the element directly to the right to the element directly above.

The flow of information between the different elements is shown in Table 3.1. Note that, without further assumptions, all elements are interconnected and the resolution procedure would need to resolve all of them at once. However, since the nature of each decision is extremely different, and modern operators need to deal with high-dimensional scenarios, a unified solution that deals with all these factors is unlikely to succeed in reasonable time. In order to reduce the complexity of the problem, several assumptions help eliminate the lower diagonal triangle in the information flow, which allows for a chain resolution of the problem. Specifically, the following simplifications are introduced:

- Beam footprint limits: assume worse case scenario, i.e., the footprint limit corresponds to the minimum elevation angle.
- Gateway-user restrictions: include the known restrictions in the user grouping

process (e.g., if two users have to be served by different gateways, do not allow them to be in the same beam)

- Gateway visibility window: restrict visibility windows where no gateway is available and reward windows where more gateways are visible.

- Beam frequency and power consumption, beam frequency overlapping, and gateway frequency and power load: assume all beams have the same spectral efficiency, which allows the calculation of necessary bandwidth per beam based on demand.

- Exact beam position: use an approximate beam position that is in practice close to the real one.

These assumptions eliminate the lower triangle on the matrix shown in Table 3.1. Consequently, there are no loops in the information flow, which allows for a chain-based framework. Starting from the users' distribution, the framework developed in this work is the following:

1. Solve the User Grouping + Beam Shaping problem: group the users into shaped beams.

2. Transform the data from user-centric to beam-centric and compute the possible visibility window for each beam on each satellite.

3. Solve the Satellite Routing problem: decide the handover time for each beam.

4. Compute the real visibility window of each beam and the possible matching between beams and gateways.

5. Solve the Gateway Routing problem: decide the offloading gateway for each beam.

6. Include the gateway beams to the model and compute the interference restrictions between beams.

7. Solve the Frequency Assignment problem: decide the central frequency and bandwidth of each beam.

8. Operate the plan: solve the short-horizon problems and serve the users.

Figure 3-1 visually represents the previous transitions. Note that the diagram described is an abstract representation of the problem and there is no condition on the resolution procedure. This implies that the techniques used to solve each block can vary as long as the interface remains the same, which makes the framework adaptive to innovative sub-problem improvements. However, given the sequential dependence between blocks, each action cannot be started before the previous one is finished. The following sections discuss each of the individual sub-problems. Each sub-problem is divided into four sub-sections: a general problem description, that defines the problem and is independent on the formulation and resolution procedures, a specific problem description, in which the specific formulation used in this work is explained, a brief discussion on how to formulate the joint problem combining the current one with the previous in the chain, and the necessary interface description, which connects each sub-problem to the next one in the chain. As mentioned, the Power Allocation formulation will be discussed at the end of this Chapter, since it will be used as a metric of performance for the full resource allocation.

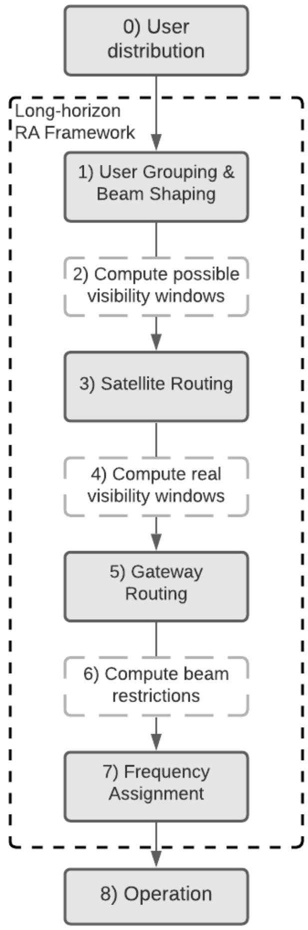


Figure 3-1: Long-horizon Resource Allocation framework. Shaded boxed with non-dashed borders imply decisions taken or to be made, while white boxes with dashed lines imply computations (i.e., no decision). The white dashed box surrounding the points 1-7 represents the scope of the framework described in this work.

3.2 User Grouping and Beam Shaping

3.2.1 General problem description

The objective of the joint User Grouping and Beam Shaping problem is to cluster the users into beams and determine the shape of each beam such that all users are covered and the resources are minimized. However, as noted in the previous section, the resource consumption is only known after the long- and short-horizon RA problem is resolved. Also, knowing if a user can be served by a beam depends on the footprint of the beam, which is strictly related to the shape, position, and handover time of the beam. Following previous assumptions and in order to break the information loops, it will be assumed that a user can only be served by a beam if it *always* falls within the footprint for the full window of satellite visibility. In addition, it will be assumed that the spectral efficiency for all beams is the same and it is known. This two assumptions make the joint User Grouping and Beam Shaping problem independent and allows for a standalone resolution.

3.2.2 Specific problem description

To reduce the complexity on this issue, in this work it will be assumed that all beams have a constant a fixed circular beam shape with a cone angle of δ. Then, the User Grouping is formulated in a similar manner as the author's previous work [68]. To provide the reader with enough knowledge to understand how the problem was defined, the following lines skim through the mathematical formulation of the User Grouping problem. Refer to the mentioned work for additional insights.

The objective of the User Grouping problem is to provide a set of beams and a collection of user-beam relations such that the users always fall within the footprint of the beam and all users are covered by a beam. In addition, the optimal solution is the one that will incur in the lowest resource consumption. Given that the actual consumption cannot be computed, a proxy multi-objective metric is used. First, assuming a fixed spectral efficiency for all beams Γ, the bandwidth needed to serve a specific beam b is:

$$BW_b = \frac{\sum_{u \in \mathcal{V}_b} d_u}{\Gamma} \tag{3.1}$$

Where $\mathcal{V}_b$ is the set of users served by beam b and d_u represents the data rate demanded by user u. Given that most satellites have predefined frequency channels of size c, the amount of channels η_b needed for beam b is:

$$\eta_b = \left\lceil \frac{BW_b}{c} \right\rceil \tag{3.2}$$

The total amount of channels consumed by all beams ($\sum_b \eta_b$) is the first proxy metric and aims to represent the amount of spectrum needed by the users. Note that this value is not fixed due to the fixed-bandwidth channels imposed by the satellites' payload. As an example, two separate users served by different beams that need half a channel each would need a total of two channels ($\lceil 0.5 \rceil + \lceil 0.5 \rceil = 2$). However, when both users are served by the same beam, we only need one channel ($\lceil 0.5 + 0.5 \rceil = 1$). As opposed to this, beams that serve many users require larger bandwidth, which makes them more difficult to allocate within the frequency spectrum as they have tighter constraints. The second metric is to minimize the amount of channels per beam ($\frac{\sum_b \eta_b}{\sum_b 1}$). However, given that the amount of channels is already an optimization objective, this is equivalent to maximizing the amount of beams ($\sum_b 1$). The final formulation is as follows:

$$
\begin{aligned}
\min \quad & \sum_b \eta_b \\
\max \quad & \sum_b 1 \\
\text{s.t.} \quad & C1 : \mathcal{V}_i \cap \mathcal{V}_j = \emptyset \text{ if } i \neq j \quad \forall i, j \\
& C2 : \bigcup_{\forall i} \mathcal{V}_i = \mathcal{U} \\
& C3 : \alpha_i \leq \frac{\delta}{2} \quad \forall i
\end{aligned}
\tag{3.3}
$$

Where $C1$ ensures that a user can only be covered by one beam, $C2$ that each user in the set of all users $\mathcal{U}$ is assigned a beam, and $C3$ that the angle α between each user and the center of their respective beam is lower than the half cone angle of the beam, which ensures user coverage. The solution to this problem is a collection of $\mathcal{V}_b$ that represent the users served by each beam b. In fact, given the multi-objective formulation, there are several optimal solutions that represent different trade-offs between the metrics. To avoid the complexity of splitting the framework's flow for each solution, it will be assumed that there exists a selection function that chooses one of the options provided by the formulation based on some criteria. A scheme for the selection process will be discussed later in this work. Figure 3-2 dissects the resolution procedure followed in this Thesis for the joint User Grouping and Beam Shaping problem.

3.2.3 Subsequent problem interface

Once the collection of beams is obtained, the objective is to proceed with the Satellite Routing problem. Before that, however, we need to compute the visibility window for each beam and satellite, i.e., the range in time when a beam and a satellite

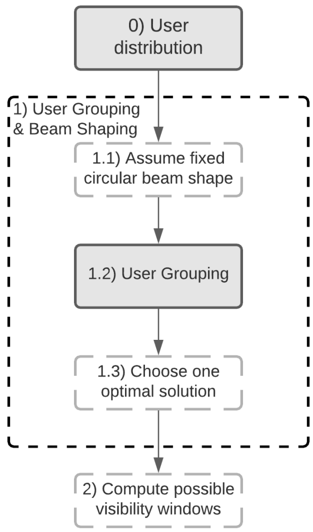


Figure 3-2: A procedure to solve the joint User Grouping and Beam Shaping problem

form an elevation angle higher than the minimum elevation angle. To perform this calculation, the position of the beam is necessary, which implies solving the short-horizon Beam Placement problem. Nevertheless, given the set of users a beam is serving, the possible location of the center of the beam such that all users are covered is limited. The variation in the visibility window depending on the exact position is relatively low if we can obtain an approximate position of the beam center. Therefore, it is not necessary to solve the Beam Placement problem, and it is enough to obtain an approximate solution. For this work, the beam center p_b will be approximated as the weighted sum of the position of the users within the beam p_u, where the weight w_u is proportional to the sum of the distance of that user with respect to the others:

$$p_b = \sum_{u \in \mathcal{V}_b} w_u p_u \tag{3.4}$$

$$\tilde{w}_u = \sum_{v \in \mathcal{V}_b} ||p_u - p_v|| \tag{3.5}$$

$$w_u = \frac{\tilde{w}_u}{\sum_{u \in \mathcal{V}_b} \tilde{w}_u} \tag{3.6}$$

The visibility windows can now be computed with the procedure explained in Section 1.3.

3.3 Satellite Routing

3.3.1 General problem description

The objective of the Satellite Routing problem is to determine the handover times for each beam so that the resource consumption across the constellation is minimized, a handover being the time when the satellite in charge of serving a specific beam changes to a different satellite in the constellation. In single-plane constellations, this implies computing the points in time where the beams are handed over a subsequent satellite in the plane. Multi-plane constellations, especially multi-shell constellations (i.e., constellations with multiple shells of satellites at different altitudes), are inherently more complex since they also have to decide which plane is serving each beam at each point in time.

3.3.2 Specific problem description

Although the nature of this problem is highly dynamic due to NGSO orbital characteristics, in this work the scope of the handover scheme will be limited to one (or a few) orbital period(s) (this period of time will be referred to as a cycle). This implies the need to recompute the solution to the Satellite Routing problem periodically if the configuration of the constellation is significantly different after a cycle (e.g., if the constellation suffers from severe orbital shifting, or if the characteristics of the set of beams change considerably). This Thesis follows a Satellite Routing formulation similar to the one presented in the author's previous work [82]. For the purpose of completeness, the following lines will give a brief description of the formulation used. Refer to the aforementioned work for additional details.

To reduce complexity, this work will assume the utilization of (almost-)stationary single-planar constellations with evenly spaced satellites. In this context, the term almost-stationary means that the shifting of the constellation (i.e., the variation of the right ascension of the ascending node) is low. In addition, the users' demand (or beams' demand) is assumed to be constant and equal to the maximum demand. This ensures that if we can serve the maximum demand, we can always meet the users requirements. As mentioned in the previous section, it will be assumed that the center of the beams is known and, additionally, that it is fixed for the duration of the cycle. From these conditions, it can be shown that each satellite observes the same conditions in terms of beam visibility and demand with a time-shift of $\frac{P}{N_{sat}}$ with respect to the previous satellite in the plane, where P refers to the period of the orbit and N_{sat} refers to the number of satellites in the constellation. Given that the conditions are the same, the solution that optimizes the resource consumption of a satellite is the same as the previous satellite with the mentioned time-shift. Thus, it is only necessary to solve the handover scheme for one period of one satellite (called the reference satellite) and propagate the results to the rest of the constellation.

Similarly to what happened in the previous joint User Grouping and Beam Shaping formulation, *minimizing the resource consumption* is a hard task since the real resource consumption is only known when the subsequent RA sub-problems are resolved. To break the information loop, a proxy function for the resource consumption will be used. Specifically, this function relies on the fact that two beams that are served by the same satellite compete for the resources of the satellite, which produces a consumption overhead compared to if they were served by different satellites. This is specially true if the beams are geographically close, since they may interfere between each other. Formally, two beams i and j overlap if their time-windows on the reference satellite overlap (y_{ij} is defined to be 1 if i and j overlap, 0 otherwise). In addition, the *cost* z_{ij} of two beams overlapping can be quantified based on the demand of each beam and their geographical positions. The objective of the Satellite Routing problem is to minimize the cost of overlapping by deciding the time-windows of each beam. Mathematically, the formulation is as follows (extracted from [82]):

$$\begin{aligned} \min \quad & \sum_{i,j,i\neq j} y_{ij} z_{ij} \\ \text{s.t.} \quad & y_{ij} = \begin{cases} 1 \text{ if } & \begin{Bmatrix} t_i < t_j + T_s \\ t_j < t_i + T_s \end{Bmatrix} \\ & 0 \text{ otherwise} \end{cases} \\ & t_{start,i} \leq t_i \leq t_{stop,i} - T_s \end{aligned} \tag{3.7}$$

Where t_i denotes the initial serving time of the beam i, $T_s = \frac{P}{N_{sat}}$, and $t_{start,i}$ and $t_{stop,i}$ represent the limits in time when the beam i can be served by the reference satellite. The solution to this problem is the set of t_i that minimize the total overlapping. Once a solution has been found, the handover scheme for the entire constellation for the desired cycle can be derived.

While this is the original formulation as presented in [82], it does not consider gateway visibility windows due to the lack of an overarching framework. It could happen, for example, that a beam is assigned a time-window that no gateway can serve, which would mean that the users on that beam cannot be served. To avoid this issue, a slight modification is proposed. Instead of considering the entirety of the valid range of satellite visibility, only the range where a gateway is available will be considered. This range will be denoted as $\tau_{gat,i} \subseteq [t_{start,i}, t_{stop,i} - T_s]$. Note that if there is no gateway available in the entire range, $\tau_{gat,i} = \emptyset$ and the beam cannot be served. However, this is consistent with real operations. The new problem definition is:

$$\begin{aligned} \min \quad & \sum_{i,j,i\neq j} y_{ij} c_{ij} \\ \text{s.t.} \quad & y_{ij} = \begin{cases} 1 \text{ if } \begin{Bmatrix} t_i < t_j + T_s \\ t_j < t_i + T_s \end{Bmatrix} \\ 0 \text{ otherwise} \end{cases} \\ & t_i \in \tau_{gat,i} \end{aligned} \tag{3.8}$$

3.3.3 A brief discussion on the joint User Grouping and Satellite Routing problem

While this work presents the User Grouping and Satellite Routing sub-problems as different entities, a contrasting approach could be to address the joint User Grouping and Satellite Routing problem. While the decision variables can be included into a joint formulation easily, the complex part of this technique is to deal with the variable footprints of the beams: knowing if a user can belong to a beam or not depends on the footprint over all times, which depends on the serving window of the satellite. In situations where the serving window can be placed within a wide range of possibilities, the footprints may change significantly depending on the exact placement. Obtaining the exact footprint requires expensive simulation. This would create a costly evaluation function, which in turn slows down the optimization procedure. A *good* joint problem formulation needs to ensure that the computation of the footprint does not make the computing time unreasonable. In addition, combining the metrics of both problems is not straightforward: while the User Grouping tries to reduce virtual demand while keeping a low demand per beam, the Satellite Routing objective is to decrease overlapping between beams. How to deal with all three elements at the same time falls out of the scope of this Thesis and needs to be resolved before approaching the joint problem.

3.3.4 Subsequent-problem interface

Given that the Satellite Routing problem directly reports the serving window for each beam, the only element left to compute in order to proceed with the next sub-problem is the set of gateways that can serve each beam according to the matching between the gateways' visibility windows and the beams' serving windows. In particular, for constellations that do not use inter-satellite links (ISLs), the beam's serving window has to completely fall within a gateway's visibility window so that the beam can be served by that gateway. When the payload on the satellites include ISLs, virtually any connection beam-gateway is possible as the data can travel through the satellite network. The set of possibilities represents the flow of information between the Satellite Routing problem and the subsequent Gateway Routing problem.

3.4 Gateway Routing

3.4.1 General problem description

The objective of the Gateway Routing problem is to determine the offloading gateway for each beam such that the resources on the network can be distributed efficiently.

In this work, it will be assumed that only one gateway is assigned to each beam. Although assigning multiple gateways is technically possible, this carries an additional level of complexity, as it implies deciding how to distribute the load across the gateways at each point in time. Consequently, and as highlighted in the previous section, for constellations that don't use ISLs, the only possible gateways are the ones that are visible from the satellite for the full duration of the serving window of a beam. If the constellation uses ISLs, any beam can connect to any gateway as long as the gateway is visible to at least one satellite at all points in time. As happened in previous sub-problems, an efficient resource distribution relies on knowing the resource consumption of each element, which is unknown at this point. To break the information loop it will be assumed that all beams use the same spectral efficiency, which allows the computation of the bandwidth needed to satisfy the user's requirements.

3.4.2 Specific problem description

Since the literature on the Gateway Routing problem is scarce, and there are no works under a user-centric spot-beam configuration, this work will detail a novel formulation for this problem. For the purpose of simplicity, it will be assumed that the constellation under consideration does not use ISLs, which limits the possibilities of each beam and excludes the complexity of the satellite network flow assessment.

As mentioned, the objective of the Gateway Routing problem is to determine the beam-to-gateway connections such that the resource consumption is minimized, while the users' requirements are met. For a proper resolution, this needs to be framed as a mathematical optimization problem where the variables, constraints, and objectives are well defined. First, y_{ij} will denote the possible connection between beam i and gateway j. Specifically:

$$y_{ij} = \begin{cases} 1 \text{ if } t_{start,j} \leq t_i \leq t_{stop,j} - T_s \\ \quad\quad 0 \text{ otherwise} \end{cases} \tag{3.9}$$

Where, following the same notation as for the Satellite Routing problem, t_i represents the initial time of the serving window for beam i, $T_s = \frac{P}{N_{sat}}$, and $t_{start,j}$ and $t_{stop,j}$ represent the limits in time when gateway j can be served by the reference satellite. Note that, although this is a general definition, further business constraints may alter the value of y_{ij}. For example, if the contract with a user specifies a concrete gateway k to route the data to, y_{ik} follows previous definition and $y_{ij} = 0 \; \forall j, j \neq k$. Following this, x_{ij} is defined as a binary variable indicating if the offloading gateway for beam i is gateway j ($x_{ij} = 1$), or if it is not ($x_{ij} = 0$). Beam i can only be served by gateway j if it is visible, so:

$$0 \leq x_{ij} \leq y_{ij}; \; x_{ij} \text{ binary} \tag{3.10}$$

As a restriction, we impose that each beam has to be served by at most one gateway, which can be encoded as:

$$\sum_j x_{ij} \leq 1 \; \forall \; i \tag{3.11}$$

The reason for the inequality will be made clear in the following lines. At this point, we need to define the objective of the optimization problem. Similarly to previous problems, a proxy that represents the resource consumption will be used. Since this problem allocates beams (and, thus, demand) to gateways, it seems natural to account for the demand allocated to each gateway. Specifically, it is assumed that each gateway j has a total capacity of μ_j, where the capacity is measured as the number of frequency channels it can handle at the same time. Under ideal conditions, the number of channels for each gateway is bounded by the total number of channels used on the satellite, without accounting for frequency reuse since the same frequency cannot be reused for the same gateway. Assuming a satellite can handle N_{ch} number of channels and it uses dual polarization, the capacity is bounded by $\mu_j \leq 2N_{ch}$ ($\mu_j \leq N_{ch}$ for single polarization). In this context, the inequality allows for an *ad hoc* definition of the capacity of each gateway according to further constraints, such as frequency limitations in certain regions, as μ_j is considered a given value of the model. To accommodate the capacity limitation and similarly to the formulation in Section 3.2, the bandwidth of each beam i is defined as:

$$BW_i = \frac{\sum_{u \in \mathcal{V}_i} d_u}{\Gamma} \tag{3.12}$$

Where $\mathcal{V}_i$ is the set of users served by beam i, d_u represents the data rate demanded by user u, and Γ denotes the spectral efficiency (assumed constant and given). Furthermore, the number of channels needed for beam i is defined as η_i and computed as:

$$\eta_i = \left\lceil \frac{BW_i}{c} \right\rceil \tag{3.13}$$

Where c denotes the bandwidth of each frequency channel. With this definition, the restriction on the capacity on each gateway can be defined as:

$$\sum_i x_{ij}\eta_i \leq \mu_j \ \forall \ j \tag{3.14}$$

The previous equation is the reason for the necessity of an inequality in Equation 3.11: since the total capacity of the network of gateways may be lower than the demand of the beams, it is possible that some beams cannot be served by a gateway.

Once all the variables and restrictions have been defined, the only aspect left to detail is the optimization function. Similarly to the Satellite Routing problems and other problems in industry, an uneven load distribution can cause resource consumption overhead. Therefore, it is generally more efficient to aim at a well balanced distribution so that the resources on all gateways are used evenly. For this purpose, g_{max} is defined as the load of the most loaded gateway in the system:

$$g_{max} = \max_j \sum_i x_{ij} \eta_i \tag{3.15}$$

Then, the objective is to spread the load between gateways, which can be translated as minimizing the load on the most loaded gateway:

$$\min \quad g_{max} \tag{3.16}$$

Without further intervention, the optimal solution is to not serve any beam (i.e., $x_{ij} = 0 \ \forall i, j$). To avoid this trivial solution, and to maximize the met demand, another factor is added to the minimization:

$$\min \quad g_{max} - M \sum_i \sum_j x_{ij} \tag{3.17}$$

Where M follows the big-M notation and represents a big number (i.e., a number sufficiently big such that the right part of the equation is always prioritized against the left part). Now, the focus of the optimization is to maximize the balance between the gateways, while keeping as many beams active as possible. Note that this a min-max problem, which may require deep computation to solve it. However, a simple mathematical transformation converts the problem into an integer linear programming (ILP) problem:

$$\begin{aligned} \min \quad & \gamma - M \sum_i \sum_j x_{ij} \\ \text{s.t.} \quad & \gamma \geq \sum_i x_{ij} \eta_i \ \forall j \end{aligned} \tag{3.18}$$

The complete formulation of the problem is as follows:

$$\begin{aligned} \min \quad & \gamma - M \sum_i \sum_j x_{ij} \\ \text{s.t.} \quad & \gamma \geq \sum_i x_{ij} \eta_i \ \forall j \\ & \sum_i x_{ij} \eta_i \leq \mu_j \ \forall \ j \\ & \sum_j x_{ij} \leq 1 \ \forall \ i \\ & 0 \leq x_{ij} \leq y_{ij}; \ x_{ij} \text{ binary} \end{aligned} \tag{3.19}$$

The solution to the Gateway Routing problem is a set of x_{ij} that determine which gateway is serving each beam. On the complexity analysis, the amount of decision variables x_{ij} grows proportionally to the number of beams to allocate and the number of gateways considered. The input's asymptotic complexity is $\mathcal{O}(nm)$ where n is the number of beams and m the number of gateways. On the algorithmic side, the problem can be solved optimally using traditional mathematical optimization methods (e.g., problem relaxation plus bounding strategies to avoid non-integer solutions). For large instances where computing time is an issue, modern optimization techniques such as artificial intelligence or machine learning implementations can achieve a close-to-optimal solution in reasonable time.

3.4.3 A brief discussion on the joint Satellite Routing and Gateway Routing problem

While the joint User Grouping and Satellite Routing problem needed further study for a successful resolution, the joint Satellite Routing and Gateway Routing problem is much more approachable: both individual problems follow similar formulations with resembling decision variables and non-antagonistic metrics. A combined formulation could involve a joint objective where both single-objectives are weighted with respect to their importance, and an aggregation of the constraints of each individual formulation. While the study of the joint problem falls outside the scope of this work, it may be investigated in future research.

3.4.4 Subsequent-problem interface

Once the Gateway Routing problem is resolved, the last resource to allocate is the frequency spectrum. In order to do so, the next objective is to solve the Frequency Assignment problem. Since the key aspect of this problem is interference mitigation and the position and serving windows for all user and gateway beams is known, it is theoretically possible to compute the possible interference between beams beforehand. However, the formulation of interference is inconsistent across works: some consider interference as a binary factor that either exists or not, while others consider a larger interference spectrum and advocate for minimizing the aggregated or maximum value. Nevertheless, the common consensus is that the interference depends on the angular separation between beams with reference to the satellite and the frequency of such beams. While the latter is paradoxically the result of the Frequency Assignment itself, the former can be pre-computed by simulating the orbit and computing the minimum separation angle between each pair of beams. For beams that are never served by the same satellite, the angular separation between beams is irrelevant since they are not going to suffer from inter-beam interference (although they might suffer from inter-satellite interference, this work assumes that this type of interference is negligible).

3.5 Frequency Assignment

3.5.1 General problem description

The objective of the Frequency Assignment problem is to assign a central frequency and bandwidth to each beam so that the resource consumption is minimized and the users' requirements are met. The assigned frequency and bandwidth must fall within the frequency spectrum allocated to the constellation, and must respect ground regulations appropriately. For the purpose of the current problem, it is important to highlight the differences between uplink and downlink, since the frequency spectrum allocated to each connection may be different. Specifically, the uplink connections tend to be generally less power and frequency constrained, since they are powered by the Earth's resources, while the downlink connections only rely on the power from the satellite's batteries and solar arrays and the frequency usage pointing to Earth tends to be more restricted. For previous problems this distinction was irrelevant since the problems for uplink and downlink were decoupled and could be solved independently. However, the introduction of gateway beams introduces new relation: for a given user beam, the associated gateway beam has the same demand but with opposite direction. The strength of the coupling depends on the satellite's architecture: a bent-pipe architecture implies a stronger coupling, since the uplink and downlink beam's pair need to have the same bandwidth, while satellites with on-board processing capabilities allow different bandwidths and evoke a weaker coupling. Both the definition of interference, and the satellite's architecture play a significant role in how to formulate and resolve the Frequency Assignment problem.

3.5.2 Specific problem description

Although the formulation only varies slightly for bent-pipe architectures, this work will assume the usage of satellites with on-board processing capabilities. In addition, interference will be understood as a binary constraint with a threshold: a threshold angle α defines the limit for inter-beam interference. Specifically, if the minimum angle between two beams is lower than α, the beams cannot have the same frequency, while a minimum angle higher than α allows for frequency reuse. Although this approach is somewhat conservative in terms of interference mitigation, it admits a strong mathematical formulation that allows to solve the Frequency Assignment problem independently from operations. Finally, it will be assumed that for the duration of the cycle, the frequency and bandwidth assigned to each beam is fixed (i.e., only one frequency can be assigned and it cannot change). Variable frequencies are technically possible, but require higher synchronization. This Thesis follows a Frequency Assignment formulation similar to the one presented by the author [132] and extended by Garau et al. [147]. The following lines give a brief description of the formulation. Refer to the aforementioned works for additional insight.

Let f_i and b_i denote the initial channel and number of channels assigned to beam i, respectively. Note that assigning central frequency is analogous to assigning initial channel and assigning bandwidth is analogous to assigning number of channels for a bandwidth-fixed channel system. In addition to these decision variables, let g_i be the index of the frequency reuse for beam i. The purpose of this index is to avoid reusing the same frequency more times that what the system can support. For completeness, let p_i denote the polarization of beam i. Note that frequency reuse and polarization accomplish the same thing, determining the number of times a specific frequency can be used, but it will be a pragmatic separation that allows for simpler constraint definition.

Let β_{ij} denote a binary variable that specifies if two beams i and j are served by the same satellite at some point in time (1 if they are, 0 if not). In addition, let γ_{ij} be the minimum angle between beams i and j over all points in time where they are on the same satellite. The value of γ_{ij} when $\beta_{ij} = 0$ is irrelevant. Then we can force the following logical restrictions:

$$\beta_{ij} = 0 \text{ or } f_i + b_i \leq f_j \text{ or } f_j + b_j \leq f_i \text{ or } g_i \neq g_j \text{ or } p_i \neq p_j \tag{3.20}$$

$$\beta_{ij} = 0 \text{ or } \gamma_{ij} \geq \alpha \text{ or } f_i + b_i \leq f_j \text{ or } f_j + b_j \leq f_i \text{ or } p_i \neq p_j \tag{3.21}$$

Equation 3.20 ensures that if two beams are at some point in the same satellite, either the frequency range, reuse index, or polarization is different. This secures that the available resources are restricted to the satellite payload capabilities. On the other hand, Equation 3.21 ensures that if two beams could potentially interfere, they need to have non-overlapping frequencies or different polarization. This protects the beams from harmful interference. Although these equations can be transformed into more formal mathematical expressions, those require many additional auxiliary variables and working through the algebra is not the main purpose of this work. See [147] for how this transformation can be accomplished.

The next step is to define the objective function. As mentioned, the purpose of the optimization should be to minimize the usage of resources while serving all the users. However *minimize the usage of resources* is an ambiguous definition that leads to many different formulations. In this work, power consumption is defined as the metric that determines the resource usage. It comes from two main sources: the necessary power to serve the beams, i.e., the required power to close the link budget equation as detailed in Section 1.3, and the necessary power for frequency reuse, i.e., the extra power required on the satellite to perform additional frequency reuses. The total power consumption can be defined as:

$$P = h_i(cb_i, d_i) + \epsilon|\mathcal{G}| \tag{3.22}$$

$$\mathcal{G} = \bigcup_i \{g_i\} \tag{3.23}$$

Where $h_i(\cdot,\cdot)$ is a function that computes power based on the bandwidth, demand, and user characteristics of beam i (note that this can be seen as an average or reference power, since the link conditions are unknown at this point), c is the bandwidth assigned to each channel, and $\mathcal{G}$ denotes the set of all frequency reuse indexes that are used in at least one beam. ϵ acts as a flexible optimization parameter to adjust the weight of each factor according to the characteristics of the satellite's payload. A higher ϵ puts more emphasis into reducing the amount of frequency reuses, while a lower ϵ tries to minimize power for each beam. The formulation for the Frequency

Assignment problem is, then:

$$
\begin{aligned}
\min \quad & P \\
\text{s.t.} \quad & \text{Eq. } 3.20 \\
& \text{Eq. } 3.21 \\
& 0 \leq f_i \ \forall i \\
& f_i + b_i \leq N_{ch} \ \forall i \\
& b_{min,i} \leq b_i \leq b_{max,i} \ \forall i \\
& 0 \leq g_i < N_r \ \forall i \\
& 0 \leq p_i < p_{max} \ \forall i \\
& f_i,\ b_i,\ g_i,\ p_i \text{ integer } \forall i
\end{aligned}
\tag{3.24}
$$

Where N_{ch} is the total number of available channels, $b_{min,i}$ and $b_{max,i}$ denote the minimum and maximum number of channels for beam i so that there exists a power that closes the loop, N_r is the number of frequency reuses, and p_{max} represents the number of different polarizations (either 1 or 2). In addition, it is important to note that $\beta_{ij} = 0$ by definition if beam i and beam j have different directions since they will never interfere. The solution to this optimization problem gives the desired frequency, bandwidth, and polarization for each beam. Although this is strictly for architectures with on-board processing capabilities, the formulation for bent-pipe payloads only need to add an additional constraint imposing that the bandwidth of the user beam has to be the same as the bandwidth on the associated gateway beam. Also, this formulation assumes that all beams can be assigned a frequency, which may not be the case for highly constrained scenarios. To adapt this case, it will be assumed that the bandwidth that can be assigned to a beam can be 0, and that the power associated with 0 bandwidth is infinity (which can be transformed into a big-M notation for practical purposes). Note that this is just an mathematical approximation: in practice, the beams assigned 0 bandwidth will not be served and need to be dropped, so the power consumption for those beams will also be 0.

3.5.3 A brief discussion on the joint Gateway Routing and Frequency Assignment problem

Similar to the joint User Grouping and Satellite Routing discussion, how to merge the individual Gateway Routing and Frequency Assignment sub-problems into one is a complex task with unclear outcome. Although neither of the formulations present opposed variables or metrics, the Gateway Routing problem introduces new gateway links while the Frequency Assignment assumes fixed beams. How to deal with these clashing views needs to be investigated before the joint problem can be successfully resolved. This falls out of the scope of this Thesis and is left as an open challenge for future work.

3.6 System metrics

As highlighted at the beginning of this Chapter, the objective of the Resource Allocation problem is to assign resources to users efficiently while meeting the demand requirements. From this general clause, two specific objectives arise: 1) optimize the resource allocation, and 2) meet the users requirements. While the latter objective could be addressed as a constraint (i.e., imposing that the users are always served), the limitation on satellite resources may lead to unfeasible solutions (e.g., if the demand of the users surpass the capacity of the network, not all users can be served and no feasible allocation exists). Therefore, it will be considered that all allocations that satisfy the technical constraints are feasible, while the ones that maximize user satisfaction are desired. When comparing two allocations that achieve the same user satisfaction, the one with lower resource consumption is preferred. To quantitatively assess two different allocations, each objective needs to be formally defined.

3.6.1 Power consumption as a system metric

Although assessing the efficiency of the RA process is ambiguous due to the different resources involved and the flexibility within each resource, this Thesis will quantify the adequacy of a RA from a power consumption perspective. Similar to the Frequency Assignment sub-problem, the power consumption comes from two different sources: the necessary power to meet the demand of each beam, and the additional required power for frequency re-use. Note that the joint User Grouping and Beam Shaping as well as the Satellite Routing and Gateway Routing sub-problems play a significant role in the power consumption: the shape of the beam affects the gain, the set of users that it is serving defines the demand, and the serving window determines the distance between antennas, which in turn affects the link quality. Thus, power consumption reflects the effect of the complete resource allocation. For the first factor, the power consumption for each beam can be computed by closing the link budget equation of Section 1.3.

To this beam power, it is necessary to add the power consumption due to the usage of reuse groups. For the purpose of this work, each additional resource group has a power consumption of ϵ, which drives the total power consumption metric to:

$$P = \sum_i P_i + \epsilon|\mathcal{G}| \tag{3.25}$$

It is important to highlight that, while physical systems are constrained by a maximum power, the explained optimization metric does not reflect this hard constraint. In real operations, the system may need to drop certain demand in order to meet the power requirements. However, it is apparent that procedures that obtain a lower power consumption will be inherently more attractive, since the amount of demand that needs to be dropped due to power limitations is lower (and, thus, more demand can be met).

3.6.2 User Satisfaction as a system metric

In this work, user satisfaction is defined as the amount of user demand that can be met. As mentioned, while ideally all the demand is satisfied, the limitation of on-board resources may make this impossible. Specifically, this Thesis uses an unmet demand (UD) metric, as defined in [16]. This metric is defined as:

$$UD = \sum_i \max\left(\sum_{u \in \mathcal{V}_i} d_u - R_i, 0\right) \tag{3.26}$$

Where R_i is the actual data-rate offered by beam i, which can be computed based on the power of the beam P_i computed in the previous section. $\sum_{u \in \mathcal{V}_i} d_u - R_i$ represents the amount of demand in beam i that could not be served. Note that beams that could not be served by default (i.e., beams that could not find an offload gateway, or beams that do not have assigned frequency due to constraints) have a data-rate of 0 by definition ($R_i = 0$). As a final remark, some results will discuss the *coverage* of the solution, which refers to the met demand (MD) and can be computed as $MD = \sum_i \sum_{u \in \mathcal{V}_i} d_u - UD$.

Although other satisfaction metrics might be interesting for certain applications (e.g., minimum delay for better quality of service or maximum fairness to avoid prioritization of users), this work puts emphasis on guaranteeing service to the maximum number of users, which is equivalent to maximize capacity. The reasoning behind this view is that it is less relevant to talk about user delay if not all users can be served, and guaranteeing fairness among users is not useful if the minimum contracted demand is not satisfied.

3.7 Assumptions and Challenges

This Section summarizes the assumptions on the general framework and the concrete sub-problems that were highlighted through this Chapter. Moreover, it discusses how to relax each assumption and emphasizes the remaining open challenges and how future work can approach them.

3.7.1 General assumptions and relaxations

The following lines detail the general assumptions that the framework relies on, as well as a brief discussion of the implications of each assumption.

A User distribution and demand is known: as this framework is user-centric, the position and (maximum) demand need to be a parameter of the model.

- For uncertain user position, the framework is still valid if the uncertainty is low (i.e., if the user position is known approximately with a specific confidence level). For high uncertainty, a user-centric model is a less attractive option. Note, however, that given that satellite communications mostly involve a somewhat large antenna, the positions of such antennas tend to be known. Only for new market models or new services the uncertainty may be higher. For uncertain user demand, usually contracts specify a minimum and maximum ceiling. Without any more information about real demand, contract values can be used without modifying the framework. Introducing uncertainty in the user distribution remains an open challenge and its effects may be studied in a future work.

B Gateway position and maximum capacity are known: this is a required parameter of the model.

- The purpose of this framework is to resolve the RA problem for a given space and ground configuration. Choosing the gateway positions falls out of the scope of this work. However, there is research specifically focused on the Gateway Placement problem [148, 149]. Note that once the gateway position and maximum demand is known, the framework can be applied effectively again.

C The framework has no prior constraints or allocation that it needs to match. However, specific business constraints are allowed as long as they fit within the defined formulation for each sub-problem. For example, defining regional spectrum limitation is relatively uncomplicated knowing the position of the beam and setting limits to the frequency variables.

- Introducing a *warm start* does not disrupt the current framework as long as the imposed constraints can be defined within the individual sub-problem definition. However, the effects of a warm start are outside the scope of this work and may be investigated in future research.

D Users are allowed to not be served, but maximizing served users is a priority.

- Allowing users not to be served is an engineering convenience. While in reality users need to be served at all costs (due to contract penalties), the user distribution needs to ensure that there exists a feasible plan. Since this work does not put constraints on the user distribution, not all distributions can be fully served. To avoid falling in unfeasible scenarios, allocations with unmet demand are allowed.

E Satellites have virtually no power constraint.

- In reality, all satellites rely on some power generation with limited capacity. By relaxing the power constraint we may end up in two possible situations: either the power is lower than the power generation capabilities, in which case the plan is technically feasible and there are no further constraints, or the power is larger than the power capacity, in which case means that the framework could not come up with a better distribution to serve the users, and thus some demand needs to be dropped. For this latter case, either the individual optimization resolutions need to be improved enough to achieve a lower power consumption, or the user distribution needs to be reevaluated according to the real system capacity.

3.7.2 Specific assumptions

To deal with the complexity of each individual sub-problem, specific assumptions have been made. The relaxation of these assumptions can be addressed through individual or multiple sub-problem formulation redefinitions. However, they usually involve a higher level of complexity. The implications of these relaxations and possible approaches will be discussed in the following lines. Since each assumption derives from a specific problem, the indentation mark encodes which concrete sub-problem the assumption applies to: UG refers to User Grouping and Beam Shaping, SR refers to Satellite Routing, GR refers to Gateway Routing, and FA refers to Frequency Assignment.

UG1 Users have a fixed position for the duration of the allocation cycle. However, this position can be virtually anywhere in the world (if the position cannot be served because of coverage constraints, the demand of the user will simply be added to the UD metric).

- Although the mobility sector (i.e., planes and ships) is very attractive for satellite communications, it involves a much complex resolution of the allocation problem. If the velocity of the movement is low, or the users tend to spend a lot of time in the same place (e.g., maritime users), the problem can be solved within the framework assuming a shorter cycle. For other cases, mobile users need an *ad hoc* problem resolution from the sub-problem perspective for the joint User Grouping and Beam Shaping, Satellite Routing, and Gateway Routing sub-problems (the Frequency Assignment problem may not require additional modifications). How to address mobility users for a spot-beam configuration has not yet been studied in research. This remains as an open challenge and future works may address this issue.

UG2 Each user is served by *one* beam, and the user has to fall within the beam footprint at all times.

- Potentially, each user could be served by multiple beams if they have the necessary antennas to support multiple operations. In practice, this implies solving an additional decision: how to split the demand of each user into several beams. In addition, it involves additional constraints for certain type of services (e.g., some services require a specific offloading gateway, which would need to be addressed for each beam containing that user). Clearly, this is a much complex problem and it falls out of the scope of this work. However, once the splitting decision has been made, the physical user could be divided into as many different virtual users as beams to be split to, and the same framework could be applied at a virtual level.

UG3 The shape of the beams is predefined and known.

- Beam Shaping is a complex problem in itself and assuming fixed shape is a convenient simplification. To include a more complex Beam Shaping resolution, it is enough to modify the joint User Grouping and Beam Shaping block, without need of changing the framework. This Thesis does not study the effect of Beam Shaping in the final allocation, and this open challenge is left for future studies.

UG4 Satellites have fixed-bandwidth channels of known size.

- Fixed-bandwidth channels are generally a restriction of the payload, which in turn depends on the advances of hardware engineering. Dealing with channels of variable bandwidth is currently not considered on satellite communications. In a scenario where this is no longer the case, it is sufficient to exploit a mathematical peculiarity: variable bandwidth channels are equivalent to fixed-bandwidth channels with size 0. For a practical purpose, it is enough to assign a very small channel size while maintaining the same framework.

SR1 The visibility windows for each beam can be approximated with an estimate of the beam center, which is known to be fixed, without incurring into technical constraints.

- Given that the position of the center of the beam is limited to the region where it covers all the users, the variation of the visibility window of the approximated center versus the real center is relatively low. However, if the technological constraints do not allow for *any* variation of this window, a useful approximation is to be conservative in serving window assignation (i.e., instead of considering the full window, consider a reduced window where all possible real beam positions fall within technical feasibility).

SR2 The constellation lies on a single plane and the satellites are evenly spaced.

- Multi-plane constellations, especially multi-shell constellations, in addition to the serving window, involve deciding the satellite that serve each beam at each point in time. Although this falls out of the scope of this work, it is enough adapting the formulation of the Satellite Routing subproblem without any additional change. The analysis on the individual sub-problem of Satellite Routing of multi-plane constellations remains an open challenge and is left as possible future research.

SR3 The constellation has low drifting across the duration of the cycle.

- Most terrestrial constellations tend to have a relatively low variation of the right ascension for the ascending node. In case this implies additional problems, it is enough to change the duration of a resource allocation cycle. If the cycle turns out to be extremely short, the Satellite Routing sub-problem formulation needs to be reevaluated to include a more robust definition.

SR4 Handovers between satellites happen instantly.

- Instant handovers are an engineering simplification. In practice, handovers usually take a time T_h to happen. However, the formulation can include this factor by extending the serving window to $T_s = \frac{P}{N_{sat}} + 2T_h$, where the initial and final handovers are included in the serving window.

GR1 The constellation does not use ISLs.

- The usage of ISLs involve a much higher complexity Gateway Routing problem that requires network flow assessment. Given the scarcity of research in the Gateway Routing problem in general, this work focuses on constellations without ISLs, and systems including this technology are left as an open challenge for future studies.

GR2 Each beam is served by *one* gateway for the full duration of the cycle.

- Similar to splitting users into several beams, splitting users into several gateways is a precarious proposition that involves a much more complex resolution. Moreover, the implication and benefits of such split remain to be studied. Similarly to the user case, this falls out of the scope of this Thesis.

FA1 Interference can be modelled with a threshold after which interference is negligible.

- Interference calculation is a complex topic. Simplifying it to a binary variable allows for a stronger mathematical formulation, at the expense of a weaker representation of reality. If this approach is considered to be insufficient or too conservative for practical purposes, it is enough to modify the formulation for the Frequency Assignment problem for a more accurate representation of the interference.

FA2 Satellites have on-board demodulation capabilities and allow frequency and bandwidth changes from uplink to downlink.

- The implications of on-board computation against bent-pipe architectures have been discussed in Section 3.5. As a summary, this Thesis will focus on full flexible architectures, but minor changes in the formulation allow for an extension to more constrained systems.

FA3 Each beam is assigned *one* frequency, bandwidth, and polarization for the full duration of the cycle.

- Assigning one resource per cycle is again an engineering simplification. While assigning multiple resources is possible, the benefits of such allocation under stationary conditions are unclear. This analysis, as well as how to obtain a robust formulation to address this context, fall out of the scope of this Thesis and remains as an open challenge.

Chapter 4

Heuristics, practical implementations, and efficient optimizations for the Resource Allocation sub-problems

While the previous Chapter focused on explaining what the problem is and setting up a comprehensive framework for the RA, this Chapter is centered around how to solve each piece of the puzzle. The following sections detail resolution procedures for each individual sub-problem that range from simple heuristics to more complex techniques such as linear mathematical optimization and metaheuristic implementations.

4.1 User Grouping

As highlighted in Section 3.2, the User Grouping problem can be formulated as:

$$
\begin{aligned}
\min \quad & \sum_b \eta_b \\
\max \quad & \sum_b 1 \\
\text{s.t.} \quad & \mathcal{V}_i \cap \mathcal{V}_j = \emptyset \text{ if } i \neq j \quad \forall i, j \\
& \bigcup_{\forall i} \mathcal{V}_i = \mathcal{U} \\
& \alpha_i \leq \frac{\delta}{2} \quad \forall i
\end{aligned}
\tag{4.1}
$$

Where the objective of the problem is to group users to achieve less virtual demand while keeping the demand of each beam as low as possible to avoid frequency restrictions. In [68], the author highlights how this problem is at least **NP-hard**, which means that there does not exist yet an optimal algorithm in polynomial time. Thus, the reasonable resolutions for this problem come in the form of heuristics or more complex approximations that inherently give a sub-optimal (but, hopefully, close-to-optimal) solution.

4.1.1 One beam per user

The first trivial solution to the User Grouping problem is simply to center the objective of the problem as maximizing the number of beams, while ignoring the virtual demand. This translates directly into assigning one beam per user. Note that, although this resolution basically ignores the purpose of the optimization, it is widely used in similar fields such as phone communications (where each phone is given a channel on demand). Also, this heuristic gives the optimal solution when the users are widely spread (e.g., where there are very few users distributed around the world) or when the beams are very narrow (e.g., when instead of electromagnetic waves, optical links are used). In such cases, no beam can cover more than one user due to geographical limitations and assigning one beam per user is the only feasible solution.

4.1.2 Minimum number of beams

As opposed to the previous solution, the other extreme is to group the users as much as possible to reduce the amount of virtual demand. However, as shown in [68], the problem of minimizing the number of beams can be transformed into an Edge Clique Cover problem, which is known to be **NP-hard**. This work will follow an heuristic developed by the author in [132] in which the beams are selected based on the number of users per beam in descending order.

4.1.3 Genetic Algorithm

While the multi-objective problem as presented above cannot be directly solved using traditional mathematical optimization methods due to its complexity, it can be approximated using modern artificial intelligence (AI) techniques. Specifically, this work uses a genetic algorithm (GA) approach developed by the author in [68]. The following lines give a brief description of how this method works. Refer to the aforementioned work for additional insight.

Genetic algorithms are a subclass of evolutionary algorithms inspired by the evolution of a population over time [150]. The population consists of a set of individuals which evolve based on two operators: crossing and mutation. Each individual represents a solution to the problem and, over multiple iterations, only the *best* individuals are kept, while the others are discarded. In order to apply a genetic algorithm to a specific problem, four elements need to be defined: how to encode a solution to the problem into an individual, and the crossover, mutation, and selection operators.

First, each individual is defined as a mapping from users to beams. Note that this element has to satisfy the constraints imposed by the problem: each user has to be covered by a beam and each beam has to cover all of its assigned users at all times. The initial definition, as well as the subsequent operators applied to the individual, need to make sure to not break these restrictions. Then, the operators are defined as follows:

- *Crossover* (i.e., merging of two individuals to create a new solution): select a non-colliding sub-set of beams from each individual and create a new solution based on the aggregation of both sets. To avoid invalid assignments, allocate a new beam for each user that remains uncovered.

- *Mutation* (i.e., change of a single individual to create a new solution): create, alter, or remove a beam from the set of beams and tweak the remaining set to ensure constraint satisfaction.

- *Selection* (i.e., filter the entire population by selecting only the individuals that perform best under the metrics of the problem): given the multi-objective definition, the selection process in this work is based on the NSGA-II [151], which searches for Pareto-Front solutions while keeping a wide population diversity.

Upon execution, the GA iterates over the population to obtain better and better solutions. Note that, due to the multi-objective formulation, the final set of individuals might contain multiple optimal solutions with different trade-offs between the metrics. As highlighted in the definition of NSGA-II, the solutions that deliver the best performance in one metric are always part of the final set of individuals (i.e., the solutions that maximize the number of beams and minimize the total demand are always part of the final set of solutions). Therefore, the heuristics that assign one beam per user or minimize the number of beams are included within the solutions of the GA. Since the previous two heuristics are a sub-set of the GA solutions, there is no need to run such heuristics.

On a related note, the selector function described in Section 3.2 will consist of a simple decision based on the number of beams. Specifically, different solutions with different number of beams will be compared and their performance analyzed. For example, the solutions with minimum and maximum number of beams will be executed and tested within the framework. For completeness, other intermediate solutions will also be studied.

4.1.4 Coverage grid

Instead of assigning beams to users, the User Grouping problem can also be seen as assigning users to beams. Specifically, it is possible to detail a set of beams without any constraints about the users, and then assign each user to each beam based on proximity. To avoid empty regions while having a sparse net of beams, the simplest technique is to divide the world into cells and assign a beam to each cell. Each user is assigned to the beam in the corresponding cell. Note that the spherical shape of the Earth makes this resolution slightly more complex since the division of the Earth's surface into equally shaped cells is not a simple problem. However, since nearby beams are allowed to have colliding footprints as long as they do not interfere in frequency, the conditions on this problem can be relaxed and an approximate solution can be found. For consistency with the other approaches, if a beam has no users assigned, it can be safely removed since its demand is zero.

Figure 4-1: A possible grid based solution for the User Grouping problem. Image extracted from Amazon's LEO constellation (Kuiper) filings [152]

Figure 4-1 shows the beam grid followed by Amazon's constellation described in [152]. While this allows for easier beam management, it may create severe unbalance between the beams. For example, beams on populated areas will inherently be more loaded than other beams, which may pose additional problems in the frequency and power allocation.

4.2 Satellite Routing

As detailed in Section 3.3, the objective of the Satellite Routing problem is to find the handover times for each beam and satellite to satisfy the users requirements while minimizing the resource consumption. The problem can be defined with a single-objective formulation as follows:

$$
\begin{aligned}
\min \quad & \sum_{i,j,i\neq j} y_{ij} c_{ij} \\
\text{s.t.} \quad & y_{ij} = \begin{cases} 1 \text{ if } & \begin{Bmatrix} t_i < t_j + T_s \\ t_j < t_i + T_s \end{Bmatrix} \\ & 0 \text{ otherwise} \end{cases} \\
& t_i \in \tau_{gat,i}
\end{aligned}
\tag{4.2}
$$

Where $y_{ij} = 1$ when two beams overlap and c_{ij} represents the cost of two beams overlapping. While the natural linearity of the presented equations may suggest a simple mathematical approach, the author in [82] showed that the traditional approaches have scalability problems in high dimensional scenarios. Specifically, and as noted in that work, there is one variable per beam pair, which increases quadratically with the number of beams and may pose computational problems when solving the problem for modern multi-beam constellations.

4.2.1 Closest satellite

One simple solution to the Satellite Routing problem for single-plane constellations is to always reach for the nearest satellite, which theoretically benefits the link budget equation since the ground antennas are closer to the satellite on average compared to other more complex approaches. Specifically, in terms of the previous formulation, the key would be to always reach for the middle point of the visibility window (i.e., $t_i = \frac{t_{start,i}+t_{stop,i}-T_s}{2}$). Note that, although this ignores the gateway visibility constraint encoded in $\tau_{gat,i}$, most users are going to be in similar regions of gateways by design, which allows to relax this requirement while still obtaining a feasible solution.

This approach would not allow to serve users whose serving window does not fall within a visibility window of a gateway. Note that this is an optimal solution when both the user demand and gateways are perfectly distributed across the coverage of the constellation, since the cost of overlapping is the same everywhere, or when the number of satellites is low, since the visibility windows and the serving windows are approximately the same size.

4.2.2 Mixed Integer Linear Programming

As described in [82], the Satellite Routing formulation can be transformed into a mixed integer linear programming (MILP) formulation which can then be solved using commercial software. Note that, however, the mentioned transformation does not include the gateway restrictions and considers that the possible serving window initial time is a continuous interval. Nevertheless, the restriction of having multiple non-continuous intervals can be adapted by including additional variables that decide which continuous interval to choose, and resolve the same formulation by treating each interval independently.

As explained in the aforementioned work, although providing the optimal solution, the MILP formulation has scalability issues for cases with more than 200 beams. For higher dimensional approaches, other less computationally intensive techniques that provide a sub-optimal solution are better suited to carry out the optimization.

4.2.3 Particle Swarm Optimization

While traditional mathematical optimization suffers from computational restrictions in high dimensional cases, more modern optimization approaches, such as meta-heuristics, fill the gap of providing a close-to-optimal solution with low computational complexity. Specifically, this work builds upon the author's previous work in [82] where a particle swarm optimization (PSO) implementation was proposed. The following lines intend to give a brief description of the PSO as detailed in the aforementioned paper. Refer to that work for additional insight.

Particle swarm optimization is a population- and iteration-based metaheuristic algorithm inspired by the movement of bird flocks or insect swarms [153]. Similarly to the GA, each individual in the population (a.k.a. particle) represents a solution and the population changes over time to achieve improved results. However, instead of encoding the solutions as a chain of genes as done in the GA, particles encode the solution as a set of coordinates that represent their position in the search space. The particles change position based on forces received from other particles and from previous history of the set of particles (a.k.a. swarm). The swarm evolves according to the dynamics of each particle. The objective of the algorithm is to direct the swarm to interesting regions in the search space to obtain an optimized solution.

Specifically, the dynamics of the particle is described as follows: the position of the particle encodes a solution to the problem, while the velocity represents the rate of change of each coordinate. The velocity is usually composed by three factors: the global pull, that represents the particle's attraction towards the best particle in the swarm, the local pull, that represents the particle's attraction towards the best position the particle has been so far, and the inertia, which represents the decay of the velocity with respect to the previous iteration. In addition to these elements, most implementations limit the maximum velocity to ensure that the search space is explored enough before convergence. At each iteration, the velocity is computed according to these definitions and the previous value, and the position is updated accordingly based on the new velocity.

For this particular problem, we have one coordinate for each beam and the value of the coordinates represents the initial time of the serving window (t_i). Given that y_{ij} just encode auxiliary variables, we do not need to encode them within the particles. The position of each particle is initialized randomly respecting the limits imposed by $\tau_{gat,i}$. Then, the PSO is executed until convergence to achieve an optimized set of handover times that minimize overlapping. Note that, since $\tau_{gat,i}$ may represent multiple non-continuous intervals, the position of each particle may need to be corrected to adapt to a feasible interval.

4.3 Gateway Routing

The objective of the Gateway Routing problem is to distribute the load of the beams across gateways to maximize resource efficiency while meeting the users requirements and satisfying the gateway constraints. Specifically, it can be mathematically defined as (extracted from Section 3.4):

$$\begin{aligned} \min \quad & \gamma - M \sum_i \sum_j x_{ij} \\ \text{s.t.} \quad & \gamma \geq \sum_i x_{ij} \eta_i \ \forall j \\ & \sum_i x_{ij} \eta_i \leq \mu_j \ \forall \ j \\ & \sum_j x_{ij} \leq 1 \ \forall \ i \\ & 0 \leq x_{ij} \leq y_{ij}; \ x_{ij} \text{ binary} \end{aligned} \tag{4.3}$$

Where x_{ij} represents the decision of assigning beam i to gateway j, γ is an auxiliary variable, and all the other parameters are given by the model. Note that this formulation is already a MILP formulation. However, the amount of variables scale proportionally to the number of beams times the number of gateways, which may pose computational problems for scenarios with thousands of gateways (as is the case of LEO mega-constellations [154, 155]).

4.3.1 Closest gateway

Similarly to the Satellite Routing closest satellite solution, a simple heuristic for deciding the offloading gateway for each beam is to reach for the geographically closest gateway. Although this ignores the optimization framework, this heuristic provides a simple and scalable solution for the problem. Moreover, it achieves the optimal solution when the beam demand is perfectly distributed near the gateways. As the demand across regions becomes more unbalanced, the solution provided by the heuristic moves away from the optimal. To keep the solution within the constraints, if the theoretical load of the gateway surpasses its technical limits, the beams that are geographically furthest out from the gateway are dropped.

4.3.2 Mixed Integer Linear Programming

Since the formulation is already in linear form, it can be directly fed into a commercial solver to obtain the optimal solution. As mentioned, the optimal solution may need to drop some beams to ensure feasiblity. If a beam does not have any gateway assigned, it cannot acquire the desired data and must be deactivated. As a technical remark, note that if $y_{ij} = 0$ there is no need to represent variable x_{ij} in the implementation, since its value will always be 0. This can help alleviate computational pressure in high dimensional scenarios.

4.4 Frequency Assignment

As highlighted in Section 3.5, the Frequency Assignment problem aims to minimize power consumption by efficiently allocating the frequency spectrum to the beams. Specifically, the objective is to solve the following formulation:

$$\begin{aligned} \min \quad & P \\ \text{s.t.} \quad & \text{Eq. } 3.20 \\ & \text{Eq. } 3.21 \\ & 0 \leq f_i \ \forall i \\ & f_i + b_i \leq N_{ch} \ \forall i \\ & b_{min,i} \leq b_i \leq b_{max,i} \ \forall i \\ & 0 \leq g_i < N_r \ \forall i \\ & 0 \leq p_i < p_{max} \ \forall i \\ & f_i,\ b_i,\ g_i,\ p_i \text{ integer } \forall i \end{aligned} \tag{4.4}$$

Where $f_i,\ b_i,\ g_i,\ p_i$ represent the initial frequency, bandwidth, frequency reuse, and polarization of beam i, respectively. Note that the formulation is non-linear due to the non-linearities in the power consumption and the *or* operation in Equations 3.20 and 3.21. Similarly, it is important to remark that there are multiple solutions that could incur the same power consumption, which makes the optimal solution non-unique. This factor may pose certain problems depending on the resolution procedure.

4.4.1 Heuristic approach

Given that the Frequency Assignment problem involves many decision variables and the search space is highly non-linear, a reasonable first approach to solve the problem is fall back to a heuristic procedure. This work will use the heuristic proposed by the author in [132], in which the objective is to maximize the spectrum usage while keeping a proportional distribution across beams. Specifically, the method is based on consecutive allocation of frequency following a demand decreasing sequence of the beams. If some beam cannot be assigned a frequency due to restrictions, it is dropped from the sequence. Note that, while the objective is strictly different that the formulation presented above, the aim is to increase the average bandwidth per beam in an attempt to decrease the necessary power to reach the desired data-rate. This approach ignores the second source of power coming from frequency reuse, which is equivalent of assuming $\epsilon = 0$.

4.4.2 Integer Linear Programming

While the original formulation falls in the non-linear division, the problem can be linearized following standard mathematical transformations. Specifically, this work presents an integer linear programming (ILP) formulation that follows the one presented by Garau et al. [147]. The following lines give a brief description of the solution proposed by Garau et al. Refer to the aforementioned work for additional details. Moreover, Appendix A describes how to perform linearizations of simple operations such as *or*, *and*, and inequalities and equalities transformations.

First, *or* operations, such as the ones from Equations 3.20 and 3.21, can be linearized in an uncomplicated way using auxiliary binary variables. Since this is only part of the sources of non-linearity, it is also necessary to linearize the power function. As mentioned, power can be computed as:

$$P = h_i(cb_i, d_i) + \epsilon|\mathcal{G}| \tag{4.5}$$

$$\mathcal{G} = \bigcup_i \{g_i\} \tag{4.6}$$

Where h_i is refers to the function that returns the solution to the link budget equation, d_i is the demand of beam i and $|\mathcal{G}|$ refers to the amount of frequency reuses used. To deal with these expressions, let us define x_{ij} as a binary variable that is 1 if $b_i = j$, 0 otherwise. Then, we can precompute the power used by beam i with bandwidth j as $P_{ij} = h_i(cj, d_i)$. With this, we can rewrite the power consumption as:

$$P = \sum_i \sum_j x_{ij} P_{ij} + \epsilon|\mathcal{G}| \tag{4.7}$$

$$\mathcal{G} = \bigcup_i \{g_i\} \tag{4.8}$$

Finally, let us define y_k as a binary variable that is 1 when any $g_i = k \ \forall i$, 0 otherwise. Note that this can be transformed into a chain of *or* operations, which can be easily linearized. Then, $|\mathcal{G}|$ can be computed as $|\mathcal{G}| = \sum_k y_k$, which leads to:

$$P = \sum_i \sum_j x_{ij} P_{ij} + \epsilon \sum_k y_k \tag{4.9}$$

Now all sources of power consumption, as well as constraints, are linear, which allows the formulation to be directly fed into a commercial linear integer solver. Note that, in this formulation, the non-activated beams are represented as beams with $b_i = 0$. To be consistent with the general formulation, $P_{i0} = M \ \forall i$, in which M follows the big-M notation.

It is important to remark that while the number of decision variable scales linearly with the number of beams, the number of auxiliary variables does so quadratically, which makes a brute-force approach computationally infeasible for high dimensional scenarios. Therefore, Garau et al. propose an iteration based method where only some beams are changed at each time, while the rest are left fixed. With this, the complexity scales proportionally to the number of beams times the number of beams that are allowed to change. By playing with the latter value, it is possible to achieve reasonably low iteration times even in cases with thousands of beams. As a remark, the initial value for all beams is obtained from the author's heuristic explained in [132].

Given that there are many solutions that achieve the same power consumption, the problem is degenerated, which causes computational problems to linear solvers. To deal with this, Garau et al. include an additional factor in the objective function that penalizes lower frequencies and higher frequency reuse indexes. In addition, for better scalability properties, the authors propose a formulation transformation where the decision variables are encoded as binary decisions, and each variable encodes one option for one beam. By exploiting the fact that some beams are static at each iteration, the authors rank the best feasible options and only allow the algorithm to take values within those preselected options. This greatly reduces the search space and allows for faster run times. Including the current option in the search space ensures that there is always at least one feasible option, which avoids running into unfeasibilities. Now the number of variables increases linearly with the number of options and the number of changes allowed, and the number of constraints scales quadratically with the number of options and the number of changes allowed.

4.5 Power Allocation

As mentioned in Section 3.6.1, this work uses power consumption as a metric to assess the optimality of the resource allocation. Thus, it considers no further constraints regarding power generation, and the objective is to determine, from two different allocations, which one is more attractive. Following this, the resolution procedure does not need to consider power restrictions. Specifically, based on Equation 1.9, a simple resolution follows:

1. Compute necessary spectral efficiency based on beam demand $\sum_{u\in\mathcal{V}_i} d_u$ and bandwidth assigned cb_i (the notation follows the one described for previous sub-problems).
$$\Gamma_i = \frac{\sum_{u\in\mathcal{V}_i} d_u}{cb_i} \tag{4.10}$$

2. Compute the MODCOD, and its associated link margin, that achieves at least the required spectral efficiency with the lowest link quality. If no available MODCOD exists that can match the desired spectral efficiency, assume highest MODCOD.

3. Compute the necessary power P_i that achieves the required link margin by solving equation 1.9.

Note that this resolution does not take into account interference between the different beams and assumes that all demand is concentrated at the center of the beam. For the former, the formulation for the Frequency Assignment sub-problem ensures that the interference between beams with the same frequency is below a predefined threshold and, thus, assumed negligible, which allows for an independent resolution for each beam. For the latter, the actual pointing loss due to the users not being at the center of the beam can be computed based on the approximate position of the beam center and the real position and demand of the users. This loss is added to the link budget equation before obtaining the necessary power.

At this point, it is no longer necessary to assume maximum demand on the user side. On the contrary, for a better representation of real operations, the full constellation can be simulated and the actual link budget solutions computed. The total power consumption is then the aggregated power over the simulated time. As a technical note, although the optimization process has no preference and tries to optimize power as a whole, only the *downlink* power will be included in the metric, since it is usually the most constrained in satellite systems.

Chapter 5

A complete Resource Allocation Process for a long-term operations plan

By combining the framework described in Chapter 3 with the methods detailed in Chapter 4, one obtains a powerful tool to solve the long-horizon Resource Allocation problem for Satellite Communications. However, to understand the effects, results, and implications of this tool, it must be subjected to rigorous testing. This Chapter dissects the testing procedures, user scenarios, and model parameters that will be used to analyze the performance of this tool. In addition, this Chapter presents a first analysis on the behaviour of the framework under different system conditions.

5.1 Integration validation

Prior to analyzing the results and performance of the proposed framework, correctness needs to be assessed. In this case, a plan is considered correct when it does not break any restriction of the model. Specifically, we need to ensure that the different constraints of the individual problems are not violated. For this purpose, the simulation includes additional validation functions that ensures the validity of the solution:

- User Grouping: we need to ensure that the users assigned to a beam fall within the footprint of the beam at all points. This can be checked by simulating an orbital period of the constellation and computing the angle between the user terminal and the center of the beam. If all users fall within the half cone angle of the beam for all simulation steps, the User Grouping is considered valid.

- Satellite Routing: we need to ensure that the time window allocated to a beam falls within the visibility window of the satellite. This can be checked by computing the specific visibility window of each satellite and assessing if all assigned beams fall within that window. If they do, the Satellite Routing is considered valid.

- Gateway Routing: we need to ensure that the gateway assigned to a beam is visible to the assigned satellite during the time window of the beam. This can be checked by computing the specific visibility window of each satellite with respect to each gateway and assessing if all assigned beams fall within that window. If they do, the Gateway Routing is considered valid.

- Frequency Assignment: we need to ensure that beams do not occupy the same spectrum resource on the same satellite at any point in time. For this purpose, we assess each interference restriction and ensure that if there is a restriction, the beams cannot occupy the same frequencies. If no constraints are violated, the Frequency Assignment is considered valid.

5.2 Test procedures

As highlighted in Chapter 3, the framework described in this work consists of a sequential resolution of the different long-horizon RA sub-problems. Given a specific constellation model and once the user distribution and characteristics are known, the first aspect to resolve is the User Grouping sub-problem. After deciding how to group the users into beams, the next step is to decide the mapping between beams, satellites, and gateways (i.e., resolve the Satellite Routing and Gateway Routing sub-problems). Finally, the frequency characteristics of each beam need to be assessed and assigned in a way that allows for feasible operation, which translates to resolving the Frequency Assignment problem. Once all the decisions have been made, the plan is ready for real-time operations.

Following this scheme, one can construct different tests by considering different user distributions, constellation models, or resolution procedures. Specifically, while user distributions and constellation models are considered fixed and given by the operating conditions, the resolution procedures can be tuned by the satellite operator. To better understand how each method performs under different conditions, each test will consist of a specific combination of user distribution and constellation model in which different algorithms are evaluated. Each configuration will be assessed based on the metrics previously presented: Unmet Demand (UD) and Power consumption (P).

5.3 User distribution

This Section describes the three user distributions considered in this work. Each user distribution consists of a list of users with fixed and known position and variable demand. In addition, the antenna characteristics, as well as ground operating conditions, are also given by the model.

5.3.1 Satellite operator model (SES)

This user distribution is given by SES S.A. and consists of around 20000 users distributed across the world. This model represents a realistic operational scenario for the satellite operator and, therefore, will be the most used user distribution in this work. It contains information of the location of the users as well as the requested demand across a day (24h) of operations. This is the same user model as used in the author's previous works [22, 68, 82, 132, 147].

5.3.2 Proportional to population (Population)

This user distribution is automatically generated based on a combination of the world population distribution and the SES dataset. Specifically, data from the Gridded Population of the World v4 dataset [156] (published by the NASA Socioeconomic Data and Applications Center, SEDAC) has been used to create a grid with 0.1° resolution where each cell represents the amount of population living in that cell. Then, a total of 20000 points have been sampled from that grid, where the probability of each point being sampled is proportional to the population in that cell. Each one of those points represents a user in the new distribution. In order to obtain the user demand, each point additionally samples a demand from a Gaussian distribution where the mean is set to be average demand per user in the distribution given by SES S.A., which allow for both distributions to have similar demand needs, although different geographical distributions. For correctness, demands that fall below a minimum threshold are set to that threshold value. The additional characteristics of the user, such as the antenna properties, are decided by copying the user parameters from the user in the SES distribution for which the demand is closest to the sampled value. This is done to avoid large discrepancies between user capabilities and demand (e.g., assigning a large demand to a low-capability antenna may result in link unfeasibilities that are unrelated with the framework). The variable demand is also extracted from the SES dataset and adjusted accordingly to the sampled demand. Figure 5-1 shows the obtained distribution of users across the world.

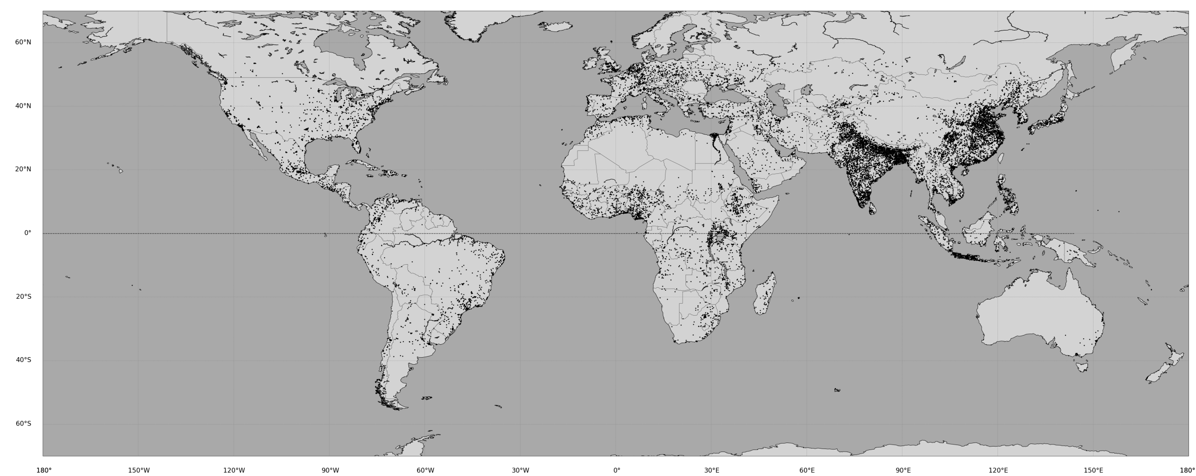


Figure 5-1: User Population distribution

5.3.3 Proportional to uncovered population (Uncovered)

This user distribution follows a similar construction mechanism as the previous distribution, with the only difference being that the data is taken from the uncovered plus badly covered population, instead of the entire population. Specifically, the user terminals are sampled from the data used in [2], which estimates the population of the world that is currently uncovered or poorly covered by the terrestrial network infrastructure. Figure 5-2 shows the obtained distribution of users across the world.

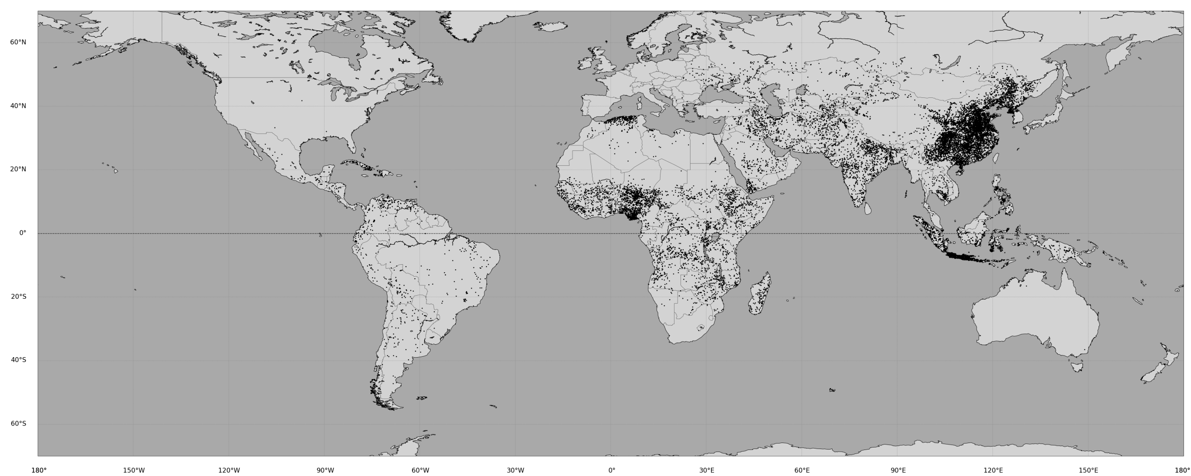


Figure 5-2: User Uncovered distribution

Parameter	Short name	Unit	Typical value
Number of satellites	N_s	-	10
Number of beam channels	N_{ch}	-	150
Number of polarizations	N_p	-	2
Frequency reuse factor	N_r	-	12
Bandwidth per beam channel	BW_{ch}	MHz	15
Beam half cone angle	δ	°	1

Table 5.1: Constellation parameters. Those parameters are either taken directly from the filings or based on the author's assumptions.

5.4 Constellation and Gateway Model

5.4.1 Constellation Model

The constellation chosen for the experiments is the constellation O3b mPower [9,10], which consists of 10 satellites in equatorial orbits at 8062 kilometers above the Earth's surface (MEO orbit). As specified in the filings, the minimum elevation angle is 10°. In addition to the parameters specified in the filings, the technology included in the satellite determines the capacity and capabilities of the constellation. Specifically, the payload on-board the satellite combined with the authorizations obtained by the operator determines the number of beam channels, number of polarizations, frequency reuse factor, bandwidth per beam channel, and beam half cone angle. Although the specific characteristics of the constellation depend on the experiment to be run, the common numbers used in this work are given in Table 5.1. In addition, and as highlighted in Section 3.7, it is assumed that the satellites have sufficient on-board processing capabilities to decouple uplink and downlink.

5.4.2 Gateway Model

In addition to the constellation model, the exact position and capabilities of the gateways is necessary to determine the quality and feasibility of the link. In this work, the position of the gateways is determined following a similar procedure as the one presented by del Portillo et al. [157], in which a large pool of possible gateway locations (a total of 181) is down-selected to obtain the desired number of gateways with the maximum coverage. Since this procedure uses a genetic algorithm implementation for the down-selection process, which can induce deviations if computed multiple times, the best set of 10/20/40/80 gateways is precomputed given the characteristics of the described constellation. This prefixed sets of gateways will be used repeatedly during the experiments.

5.5 Experiments

The experiments in this work are divided into three main categories:

- Baseline Comparison (A): initial tests on the performance of the framework. This category includes two analysis: a general test under low capacity conditions (i.e., when the capacity of the system is likely not enough to satisfy all the demand), and a general test under high capacity conditions (i.e., when the capacity of the system is likely enough to satisfy all the demand).

- User distribution sensitivity (B): tests to determine the capacity of the framework to deal with different user distributions, by selecting users from different sets (SES, Population, and Uncovered), and scenario dimensionalities, by randomly selecting only a subset of users to serve.

- Model sensitivity (C): tests to determine the capacity of the framework to adapt to different constellation characteristics and assess how those affect performance. These experiments include one-at-a-time variations on the model parameters and sensitivity testing on specific modifications.

The different resolution procedures considered in the experiments are shown in Table 5.2. A summary of the different experiments performed in this work is given by Table 5.3. Throughout the experiments, the UD will be normalized against the total aggregated demand according to the user distribution, while the Power metric will be normalized against the estimated power capabilities of the system. It is important to remark that if $P > 1$, the real system would not be able to perform such a plan, and additional beams would need to be dropped to fall within the capabilities of the system (i.e., the satellite operator would need to increase UD to reduce P so that $P \leq 1$). As a final remark, some results discuss the *coverage* of the solution, which refers to the met demand (MD) and can be computed as $MD = 1 - UD$ (in normalized units).

Identification number	User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment
1	GA	PSO	IP	IP
2	GA	PSO	IP	HEU
3	GA	PSO	HEU	IP
4	GA	HEU	IP	IP
5	GA	HEU	HEU	HEU
6	Grid	PSO	IP	IP
7	Grid	PSO	IP	HEU
8	Grid	PSO	HEU	IP
9	Grid	HEU	IP	IP
10	Grid	HEU	HEU	HEU

Table 5.2: Summary of the resolution procedures used in the experiments. The acronyms on the table refer to: GA-Genetic Algorithm, PSO-Particle Swarm Optimization, IP-(Mixed) Integer (Linear) Programming, and HEU-Heuristic.

Inputs											Outputs	
Experiment Set	User Dataset	Approximat number of users	Approximate number of beams	Number of satellites	Number of gateways	Number of beam channels	Bandwidth per beam channel [Mhz]	Frequency reuse factor	Half cone angle [°]	Resolution procedures	Reference Figure(s)	Reference Table(s)
	SES	20k	2.5k/6k/10k/20k	10	40	150	15	12	1	1-5	5-3, 5-4	5.4
A	SES	20k	2.5k	10	40	150	15	12	1	6-10	5-3, 5-4	5.4
	SES	20k	2.5k/6k/10k/20k	10	40	300	15	12	1	1-5	5-5, 5-6	5.5
	SES	20k	2.5k	10	40	300	15	12	1	6-10	5-5, 5-6	5.5
	SES	200/2k/20k	200/2k/2.5k	10	40	150	15	12	1	1-5	6-1	B.1
B	Populatio	200/2k/20k	200/2k/4k	10	40	150	15	12	1	1-5	6-2	6.1
	Uncovere	200/2k/20k	200/2k/3k	10	40	150	15	12	1	1-5	6-3	6.2
	SES	20k	2.5k	6/10/14	40	150	15	12	1	1,10	6-4	B.4
	SES	20k	2.5k	10	10/20/40/80	150	15	12	1	1,10	6-5	B.5
	SES	20k	2.5k	10	40	100/150/200	15	12	1	1-4,10	6-6	B.6
C	SES	20k	2.5k	10	40	150	10/15/20	12	1	1,10	6-7	B.7
	SES	20k	2.5k	10	40	150	15	8/12/16	1	1,10	6-8	B.8
	SES	20k	2.5k	10	40	150	15	12	0.5/1/1.5	1,10	6-9	B.9
	SES	20k	2.5k	10	40	100/150/200	11.25/15/22	12	1	1,10	6-10	B.10

Table 5.3: Summary of the experiments of this work.

5.6 Baseline Comparison and Performance Analysis (Experiment Set A)

This Section presents a first initial performance comparison and evaluation on the framework. For this purpose, two different scenarios have been considered: one with low system capacity, where the main objective of the framework should be to accommodate as many users as possible, and one with sufficient system capacity to include all users, and where the objective of the framework should be to minimize power consumption. The exact parameters of each run is summarized in Part A of Table 5.3.

5.6.1 Low system capacity

This first initial analysis evaluates the performance of the system under low system capacity conditions. The maximal throughput of the system is set to be lower than the expected capacity, and the optimization framework should try to allocate the resources as efficiently as possible to maximize the amount of users that can be served. This exercise revolves around two ideas: 1) test how a fully optimized framework (i.e., a framework in which all problems are optimized with state-of-the-art algorithms) compares to a simple heuristic solution (i.e., a solution consisting of aggregation of heuristics), and 2) assess the effect of each algorithm and each decision on the final solution. In this context, a fully optimized algorithm corresponds to selecting the GA for the User Grouping and choosing the solution with the least amount of beams, and selecting the PSO, IP, and IP resolutions for the Satellite Routing, Gateway Routing, and Frequency Assignment, respectively. A simple heuristic resolution procedure corresponds to selecting one beam per user (which is equivalent to running the GA for the User grouping and selecting the option with the largest amount of beams), choosing the closest satellite and gateway for the Satellite Routing and Gateway Routing problems, respectively, and using the heuristic method for the Frequency Assignment. Then, to test how decisions on the resolution process affects the final solution, the fully optimized pipeline is tuned one factor at a time: by changing only the User Grouping solution, or the Satellite Routing, Gateway Routing, or Frequency Assignment algorithms, we can understand the role of those elements in the performance of the framework.

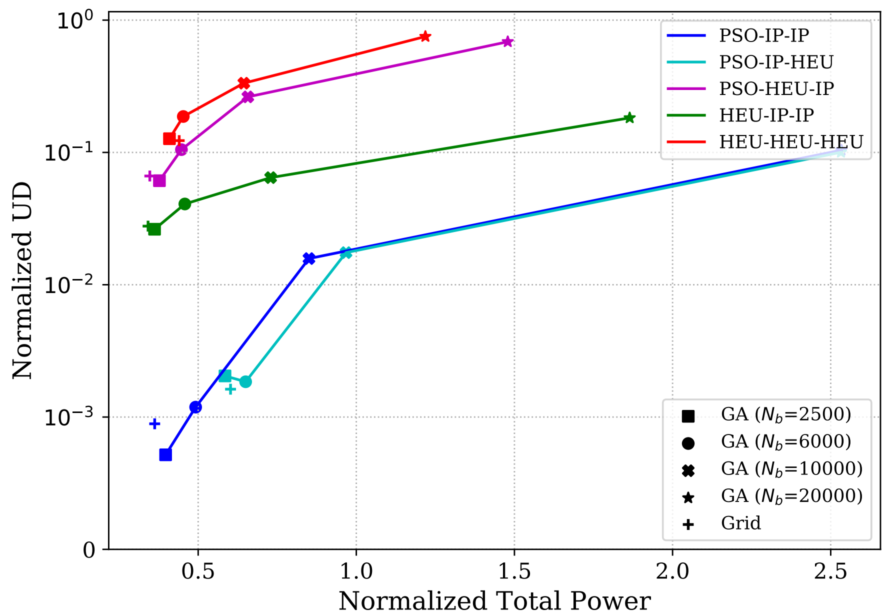


Figure 5-3: Performance comparison under low system capacity. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with the number of beams. The upper legend encodes each resolution procedure with a color, where each element is defined by a Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes a resolution procedure for the User Grouping and the number of beams selected (if applicable) with a shape.

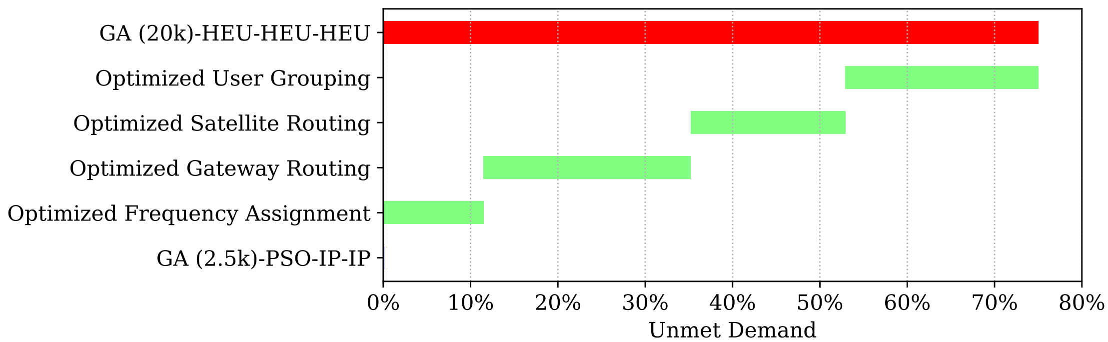
Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Number of beams	Power	UD
GA	PSO	IP	IP	2500	0.397	**0.001**
GA	PSO	IP	HEU	2500	0.585	0.002
GA	PSO	HEU	IP	2500	0.378	0.061
GA	HEU	IP	IP	2500	0.362	0.026
GA	HEU	HEU	HEU	2500	0.409	0.127
GA	PSO	IP	IP	6000	0.492	0.001
GA	PSO	IP	HEU	6000	0.650	0.002
GA	PSO	HEU	IP	6000	0.447	0.105
GA	HEU	IP	IP	6000	0.458	0.041
GA	HEU	HEU	HEU	6000	0.453	0.187
GA	PSO	IP	IP	10000	0.850	0.016
GA	PSO	IP	HEU	10000	0.968	0.017
GA	PSO	HEU	IP	10000	0.657	0.263
GA	HEU	IP	IP	10000	0.729	0.065
GA	HEU	HEU	HEU	10000	0.644	0.334
GA	PSO	IP	IP	20000	2.534	0.104
GA	PSO	IP	HEU	20000	2.532	0.100
GA	PSO	HEU	IP	20000	1.479	0.686
GA	HEU	IP	IP	20000	1.864	0.182
GA	HEU	HEU	HEU	20000	1.218	0.752
Grid	PSO	IP	IP	2500	0.363	0.001
Grid	PSO	IP	HEU	2500	0.602	0.002
Grid	PSO	HEU	IP	2500	0.347	0.066
Grid	HEU	IP	IP	2500	**0.342**	0.028
Grid	HEU	HEU	HEU	2500	0.440	0.123

Table 5.4: Detailed numbers on the performance comparison under low system capacity.

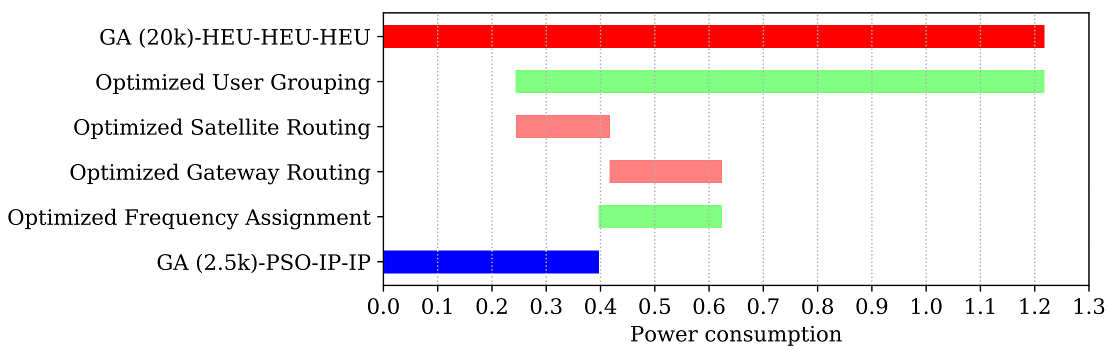
This leads to a total of 25 evaluations, which are illustrated in Figure 5-3 and Table 5.4. The fully optimized solution corresponds to the blue square, while the simple heuristic resolution corresponds to the red star. All the other points correspond to intermediate solutions. As shown, the simple heuristics have approximately a 75% of UD, which corresponds to covering only 25% of the demand. On the other hand, the fully optimized pipeline achieves the lowest UD, at only 0.1%. In other words, the optimized framework is able to cover around 4 times the demand covered by the heuristic resolution. In terms of power, the solution given by purely heuristics achieves a value of 1.218, which is 21.8% higher than the actual power achievable by the system. To obtain an operational plan, even more demand would need to be dropped so that the power does not exceed the capabilities of the satellites. On the other hand, the optimized framework not only manages to practically cover all the demand, but does so using only 40% of the total power available.

While the two-point comparison clearly implies that an efficient usage of resources is desirable in all fronts, it is also important to notice which factors generate the larger gains. If we turn our look into the User Grouping solution and the number of beams selected, we observe that selecting a lower number of beams is desirable in most cases. As mentioned in Chapter 3, reducing the number of beams helps reduce the amount of *virtual* demand (demand that only appears due to integer restrictions on the allocation of resources), which increases the overall capacity of the system. Additionally, dealing with beams with higher demand does not pose additional problems to the resolution procedures (with one exception being the cyan square, which corresponds to the 2.5k beam solution from the GA for the User Grouping and the PSO, IP, and HEU procedures for the Satellite Routing, Gateway Routing, and Frequency Assignment, respectively). If we compare the general low beam solution to the grid approach, we observe that both of them perform fairly similar, with barely any noticeable change in power or UD. The average improvement in both power and UD when selecting the 2.5k beam solution over the 20k is shown in the first section of Figure 5-4. As mentioned, choosing the 2.5k beams solution over a grid solution only barely modifies performance, but selecting 2.5k beams over

20k beams leads to almost 90% average reduction in UD and almost 80% average reduction in power.



(a)



(b)

Figure 5-4: Impact of each resolution procedure decision on the low capacity system. The upper and lower bars represent the heuristic and optimized solutions, respectively. The intermediate bars denote the independent improvement (or decline) of the optimized algorithms on each sub-problem.

When shifting focus to the other algorithms, the information contained in Figures 5-3 and 5-4 and Table 5.4 suggest that the problem with the highest impact on the UD metric is the Gateway Routing problem. Specifically, an optimized algorithm for this problem can achieve up to 90% average reductions in UD. This can be explained by the fact that gateways tend to be a bottleneck of the system, so a robust and efficient allocation mechanism to map beams to gateways is needed to maximize system capacity. On a secondary level, the Satellite Routing algorithm can help reduce around 60% the UD. Similar to the previous case, satellites can also be a bottleneck in dense regions, and spreading the demand throughout the constellation helps in alleviating this pressure. Although less impactful than the previous two resolution procedures, selecting a suitable technique for the Frequency Assignment problem can reduce the UD around 40%. In terms of power, however, the same Frequency Assignment algorithm can help reduce power consumption around 20% on average. Selecting an optimized resolution for the Satellite Routing or Gateway Routing problems actually increases the power consumption between 15% and 20%, as a trade-off for improved system throughput. This is explained by the fact that the heuristic of going to the nearest satellite or gateway is always the one that translates to the lowest power consumption, since the satellite is nearest to the terminal and less power is required for the communication to happen. Any other algorithm that modifies this allocation will therefore incur in a larger power consumption.

5.6.2 High system capacity

The objective of this second analysis is to assess the properties of the system in similar aspects as the previous one, but in a high capacity scenario. In this case, the framework should try to minimize the overall power consumption of the system. Similar to the previous case, the experiment revolves around 25 evaluations of the framework under different resolution procedures, showcased in Figure 5-5 and Table 5.5. As highlighted in the previous scenario, the red star represents an simple solution constituted by classic heuristics. In this case, even this straightforward resolution is able to cover more than 85% of the demand, due to the extraordinarily high capabilities of the system. Furthermore, most algorithms are able to achieve total coverage (0 UD) with minimal optimization. Given this conditions, the best algorithm is the one that provides the solution with lower power, which turns out to be the one that uses a Grid approach for the User Grouping, and a HEU, IP, and IP for the Satellite Routing, Gateway Routing, and Frequency Assignment problems, respectively (denoted as a green plus symbol).

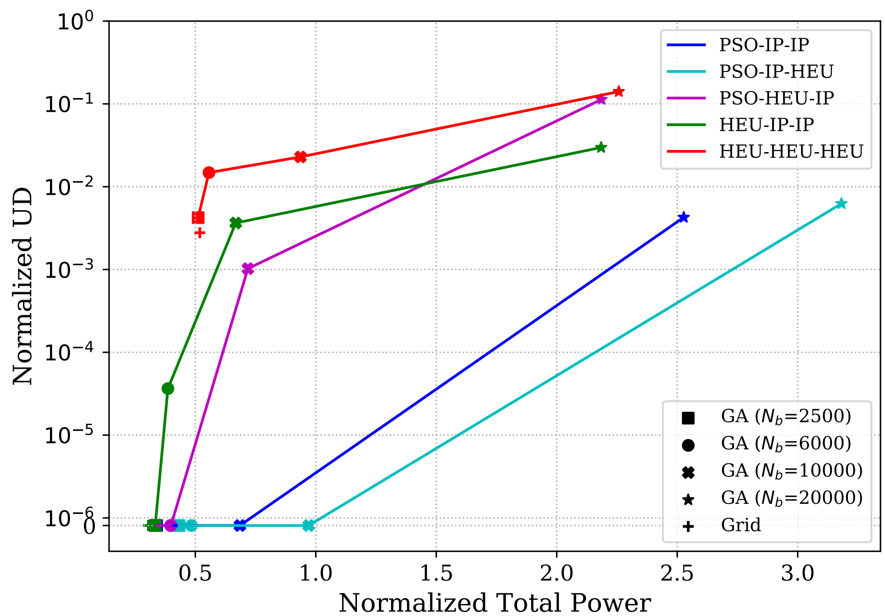


Figure 5-5: Performance comparison under high system capacity. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with the number of beams. The upper legend encodes each resolution procedure with a color, where each element is defined by a Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes a resolution procedure for the User Grouping and the number of beams selected (if applicable) with a shape.

Similar to the previous case, reducing the number of beams always helps in re-

Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Number of beams	Power	UD
GA	PSO	IP	IP	2500	0.339	**0.000**
GA	PSO	IP	HEU	2500	0.433	**0.000**
GA	PSO	HEU	IP	2500	0.332	**0.000**
GA	HEU	IP	IP	2500	0.334	**0.000**
GA	HEU	HEU	HEU	2500	0.512	0.004
GA	PSO	IP	IP	6000	0.398	**0.000**
GA	PSO	IP	HEU	6000	0.484	**0.000**
GA	PSO	HEU	IP	6000	0.399	**0.000**
GA	HEU	IP	IP	6000	0.386	0.000
GA	HEU	HEU	HEU	6000	0.557	0.015
GA	PSO	IP	IP	10000	0.685	**0.000**
GA	PSO	IP	HEU	10000	0.970	**0.000**
GA	PSO	HEU	IP	10000	0.718	0.001
GA	HEU	IP	IP	10000	0.668	0.004
GA	HEU	HEU	HEU	10000	0.937	0.023
GA	PSO	IP	IP	20000	2.527	0.004
GA	PSO	IP	HEU	20000	3.182	0.006
GA	PSO	HEU	IP	20000	2.185	0.113
GA	HEU	IP	IP	20000	2.184	0.030
GA	HEU	HEU	HEU	20000	2.258	0.141
Grid	PSO	IP	IP	2500	0.341	**0.000**
Grid	PSO	IP	HEU	2500	0.482	**0.000**
Grid	PSO	HEU	IP	2500	0.328	**0.000**
Grid	HEU	IP	IP	2500	**0.305**	**0.000**
Grid	HEU	HEU	HEU	2500	0.519	0.003

Table 5.5: Detailed numbers on the performance comparison under high system capacity.

ducing both the UD and power consumption, since it helps reducing the amount of virtual demand on the system. Likewise, the grid method performs comparable to the lowest beam solution from the GA approach, both in terms of UD and power. As shown in the first part of Figure 5-6, we can achieve significant gains in power by reducing the number of beams. Specifically, we can reduce power around 85% when using 2.5k beams over 20k beams. As shown, the 2.5k beam solution for the User Grouping performs on average slightly better than the Grid approach, but the reduction in power is not significant enough (less than 5%) to determine which approach is better.

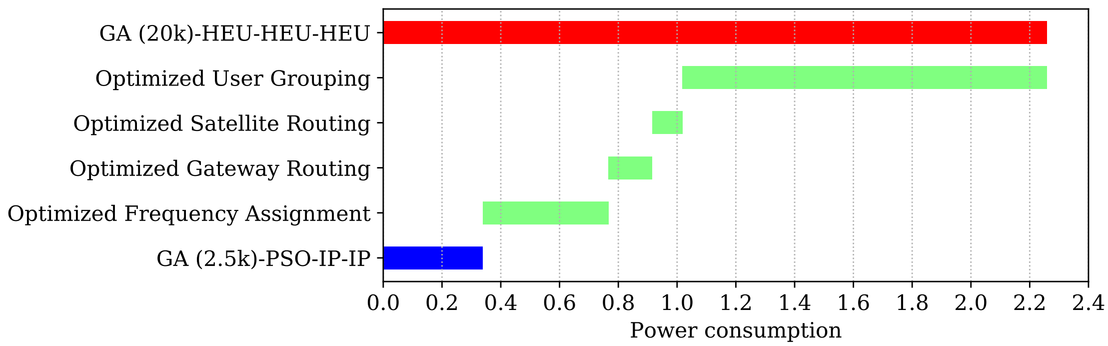


Figure 5-6: Impact of each resolution procedure decision on the power for the high capacity system. The upper and lower bars represent the heuristic and optimized solutions, respectively. The intermediate bars denote the independent improvement of the optimized algorithms on each sub-problem.

The largest difference of this high capacity scenario versus the previous low capacity one is shown in the resolution procedures for the Satellite Routing, Gateway Routing, and Frequency Assignment. While an optimized algorithm for the Frequency Assignment had the lowest impact on UD in the previous experiment, it is highly relevant to choose a good Frequency Assignment method if the objective is solely to reduce power, and the results show that an optimized version can reduce power in the order of 30% on average with respect to simple heuristics. It is also important to note that, while the optimized methods for the Satellite Routing and Gateway Routing problems implied an increase in power consumption on the low capacity case, here they have the opposite effect and they actually help in decreasing the power consumption. This can be explained by the fact that, while the consumption per beam is larger due to being further from the satellite with respect to the heuristic, a more balanced allocation can favor a better distributed spectrum management, which compensates for the power loss of those methods. However, their impact is generally small and they can only achieve reductions up to 10%.

5.6.3 Run-time analysis

While the previous Section showed that the optimized pipeline is desirable to achieve improved UD and power consumption, it comes at a computation cost: while simple heuristics tend to be fast to compute and evaluate, complex algorithms incur in significantly higher computing cost, which may make them impractical for real operations. For this reason, it is important to analyze the cost of the optimization algorithms in terms of computation and demonstrate that the full framework can be executed in reasonable time. The results in Table 5.6 correspond to the average computing time of 4 independent runs of the algorithms in the same conditions as the previous Section. The code has been executed using common python libraries for Artificial Intelligence, such as numpy or gurobi®, and a 16 CPU modern computer operating at 2.3 GHz. The results show that the full optimization framework can be executed in around 3.5h and produces a plan that is valid for at least an orbital period. Note that using a Grid algorithm for the User Grouping problem instead of a GA can reduce the computing time to around 1h at the expense of a slight decrease in system performance. All other algorithms not presented in the table have an execution time < 1 minute due to their heuristical nature.

Algorithm	Average Computing Time (h)
Genetic Algorithm for User Grouping	2.560
Grid for User Grouping	0.041
Particle Swarm Optimization for Satellite Routing	0.020
Mixed Integer Programming for Gateway Routing	0.001
Mixed Integer Programming for Frequency Assignment	0.988

Table 5.6: Execution time for each optimization algorithm considered in this work. Results are reported as the average of 4 independent runs on a standard 16 CPU computer at 2.3 GHz.

Chapter 6

Sensitivity and Robustness analysis on the Resource Allocation Process

Following the pattern established by Table 5.3, this Chapter will cover the experiments noted as B, which refer to the robustness and performance of the model against different user distributions, and C, which refer to the sensitivity of the model against different operational parameters. While the Figures will be covered in the following lines, only some Tables will be presented in this Section to ease comprehension. A complete list of Tables with all the detailed results can be found in Appendix B.

6.1 Robustness Analysis (Experiment Set B)

This Section details the results for the robustness analysis performed in this work, where robustness refers to the capacity of the framework to adapt to different inputs of the model. In this case, the inputs of the model corresponds to the user distribution. This Section will be decomposed into three different tests, one for each user distribution (SES, Population, and Uncovered). In each test, each resolution procedure[1] will be evaluated under three different scenarios with 200, 2k, and 20k users. Since the results that involved the Grid and the GA with 2.5k beams for the User Grouping have been shown to be practically identical, only the solution with 2.5k beams will be considered in the experiments of this Section. Similarly, since solutions with higher number of beams have been shown to perform worse in an overwhelming amount of the experiments in set A, they have been left out of the experiments in set B.

6.1.1 Robustness to dimensionality on the SES distribution

This first robustness tests assesses the capability of the framework to adapt to different dimensionality scenarios under the SES user distribution. Since the results for Section 5.6 already report the values of the framework on the 20k user input, only the heuristic and the fully optimized resolution procedures have been evaluated in this scenario.

Figure 6-1 shows the results for the heuristic and fully optimized pipelines on the 200, 2k, and 20k user inputs on both power and UD. If we focus on the points with 20k users (marked as a cross), we observe that the main focus of the framework is to prioritize the reduction in unmet demand, while power plays a secondary role. By observing the tendency of the heuristic solutions (red line), we can assess that even the heuristics where no objective function is defined tend to try to maximize the system's capacity, often disregarding power aspects. As mentioned in Section

[1]A resolution procedure represents a specific combination of algorithms that resolve each of the sub-problems in the framework

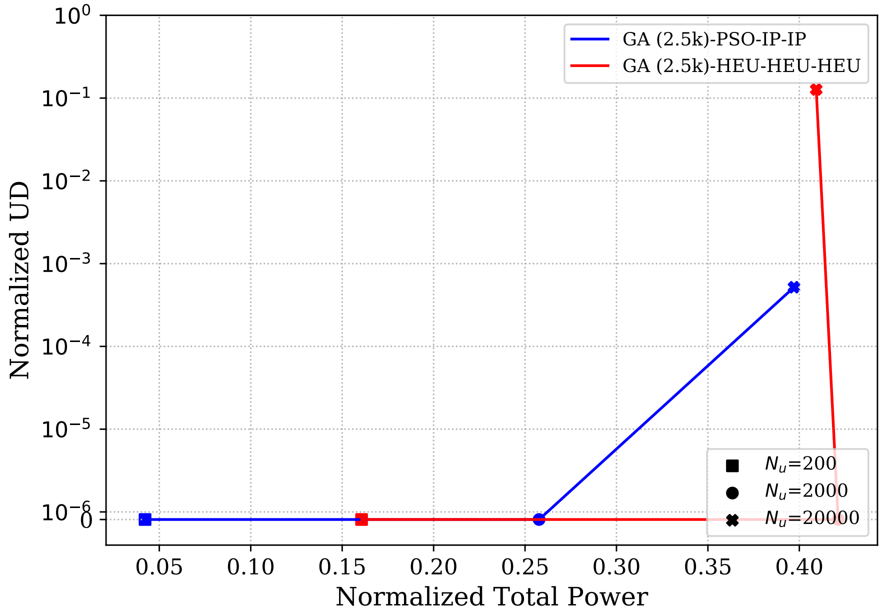


Figure 6-1: Performance comparison under three dimensionality scenarios. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with the number of users. The upper legend encodes each resolution procedure with a color, where each element is defined by a User Grouping, Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes the number of users considered in each evaluation with a shape.

5.6, the optimized pipeline is able to quadruple the amount of demand met with similar power levels. As we reach 0 UD (square and circle shapes), the focus shifts towards minimizing power, where the optimized pipeline is able to reduce power by 75% and 40% for the 200 and 2k user input over the heuristic resolution, respectively. This first analysis entails that having an optimized resolution procedure for the full long-horizon resource allocation pipeline yields highly improved results over heuristic procedures independently on the dimensionality of the problem.

6.1.2 Robustness to dimensionality on the Population distribution

Together with the next experiment, this second robustness test addresses the capability of the framework to adapt to different user distributions. Specifically, this experiment uses the Population proportional user dataset, which represents a user scenario where the user terminals are sampled from the global population distribution, as explained in Section 5.3.

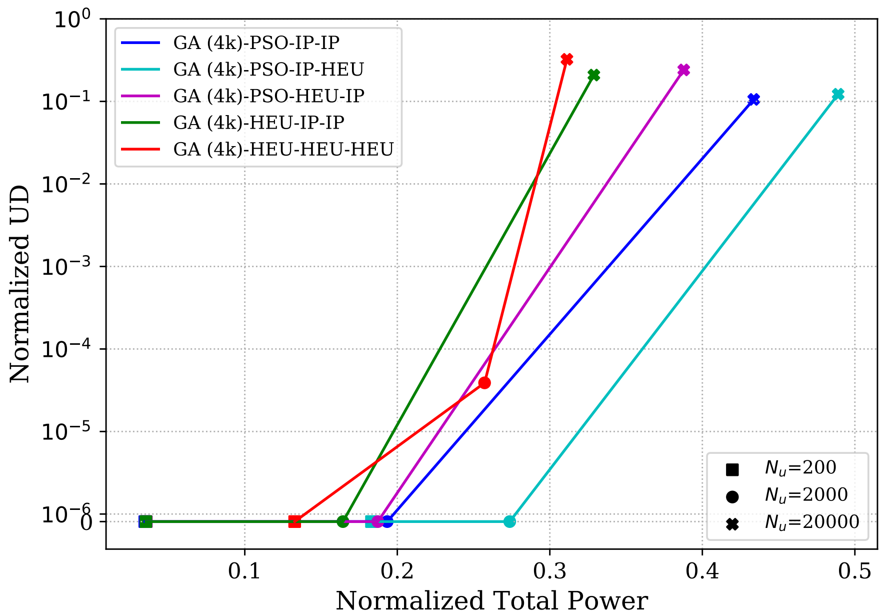


Figure 6-2: Performance comparison under the population proportional user distribution. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with the number of users. The upper legend encodes each resolution procedure with a color, where each element is defined by a User Grouping, Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes the number of users considered in each evaluation with a shape.

Figure 6-2 and Table 6.1 show results for 5 different resolution procedures under

Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Number of users	Power	UD
GA (200)	PSO	IP	IP	200	**0.034**	**0.000**
GA (200)	PSO	IP	HEU	200	0.183	**0.000**
GA (200)	PSO	HEU	IP	200	0.035	**0.000**
GA (200)	HEU	IP	IP	200	0.035	**0.000**
GA (200)	HEU	HEU	HEU	200	0.133	**0.000**
GA (2k)	PSO	IP	IP	2000	0.193	**0.000**
GA (2k)	PSO	IP	HEU	2000	0.274	**0.000**
GA (2k)	PSO	HEU	IP	2000	0.187	**0.000**
GA (2k)	HEU	IP	IP	2000	**0.164**	**0.000**
GA (2k)	HEU	HEU	HEU	2000	0.257	0.000
GA (4k)	**PSO**	**IP**	**IP**	20000	0.434	**0.106**
GA (4k)	PSO	IP	HEU	20000	0.489	0.122
GA (4k)	PSO	HEU	IP	20000	0.388	0.241
GA (4k)	HEU	IP	IP	20000	0.329	0.209
GA (4k)	HEU	HEU	HEU	20000	**0.311**	0.323

Table 6.1: Detailed numbers on the performance comparison under different dimensionality scenarios on the Population dataset. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

3 different user dimensionality scenarios: 200, 2k, and 20k users. If we focus on the 20k user input data (cross-shaped points), we observe a similar behaviour as observed in the set of experiments A in Section 5.6: the fully optimized pipeline achieves the lowest UD, and each algorithm in the chain helps contribute in reducing this factor. However, this reduction comes at a power cost: the fully optimized solution achieves a 67% reduction in UD over the heuristics resolution, at an expense of a 40% increase in power consumption. Still, the final solution is below power limits of the system capacity (represented by a normalized value of 1). Even more, results show that

both the Satellite Routing and Gateway Routing formulations trade a lower UD for a higher power, while the Frequency Assignment algorithm achieves a reduction in both factors simultaneously.

Similarly to the results of experiment set A, once we move to the 0 UD region a fully optimized pipeline may not yield the lowest power. This can be observed in the 200 and 2k user inputs (square and circle shapes, respectively): since the optimized algorithms for Satellite Routing and Gateway Algorithm tend to put beams further away from the satellite, the system capacity rises at the cost of an increased power consumption. Thus, if we have enough capacity to cover all users, the results suggest that we can use heuristic resolutions for the Satellite Routing and Gateway Routing algorithms in order to save power.

It is important to highlight that, while the heuristics perform better under this user scenario than under the SES dataset for the 20k user input (from 75% UD in the previous case to 32% in this one), the fully optimized pipeline performs worse (from 0.1% UD to 11%). This can be explained by the nature of the user distributions: the SES dataset tends to have highly dense areas and large sparse zones, while the Population proportional distribution has a more spread demand. Concentrating the user base implies very high demand peaks, which are not well dealt by the heuristic algorithms, but are well exploited by the optimized techniques. Thus, more dense environments advocate for an improved and more complex resolution procedure, while simpler algorithms can be developed for more demand-spread scenarios. To assess this conclusion further, the next experiment should offer a middle ground between density and spread in demand. If this statement is correct, the results should show an improvement in UD that is between the one seen for the SES dataset and the one for the Population proportional dataset.

6.1.3 Robustness to dimensionality on the Uncovered distribution

This third and final test evaluates the performance of the different resolution procedures under the Uncovered population dataset. Similarly to the Population proportional user distribution used in the previous experiment, the user terminals are sampled from the uncovered areas of the Earth according to population distribution, as explained in 5.3.

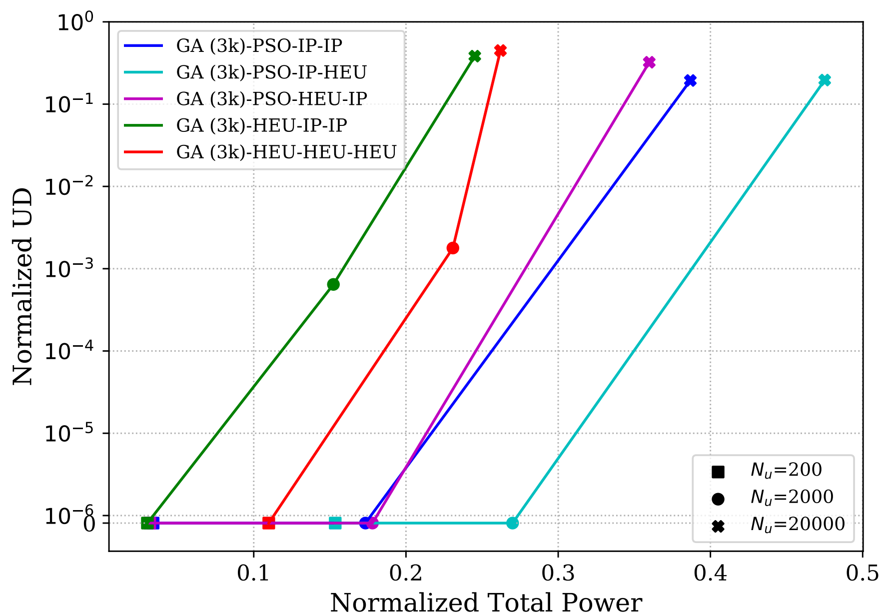


Figure 6-3: Performance comparison under the uncovered population user distribution. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with the number of users. The upper legend encodes each resolution procedure with a color, where each element is defined by a User Grouping, Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes the number of users considered in each evaluation with a shape.

Figure 6-3 and Table 6.2 show the results for the 5 resolution procedures and

Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Number of users	Power	UD
GA (200)	PSO	IP	IP	200	0.034	**0.000**
GA (200)	PSO	IP	HEU	200	0.154	**0.000**
GA (200)	PSO	HEU	IP	200	0.032	**0.000**
GA (200)	HEU	IP	IP	200	**0.030**	**0.000**
GA (200)	HEU	HEU	HEU	200	0.110	**0.000**
GA (2k)	PSO	IP	IP	2000	0.173	**0.000**
GA (2k)	PSO	IP	HEU	2000	0.270	**0.000**
GA (2k)	PSO	HEU	IP	2000	0.178	**0.000**
GA (2k)	HEU	IP	IP	2000	**0.152**	0.001
GA (2k)	HEU	HEU	HEU	2000	0.231	0.002
GA (3k)	**PSO**	**IP**	**IP**	20000	0.387	**0.194**
GA (3k)	PSO	IP	HEU	20000	0.475	0.196
GA (3k)	PSO	HEU	IP	20000	0.360	0.323
GA (3k)	HEU	IP	IP	20000	**0.245**	0.383
GA (3k)	HEU	HEU	HEU	20000	0.262	0.448

Table 6.2: Detailed numbers on the performance comparison under different dimensionality scenarios on the Uncovered dataset. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

3 dimensionality scenarios on the Uncovered population dataset. Similarly to the previous cases, the 20k results (cross-shape points) show that the optimized resolution procedures are able to highly reduce the UD compared to simple heuristics, at the expense of an increase in power consumption. Specifically, the fully optimized pipeline reduces the unmet demand to 19%, which translates to a 57% reduction over the heuristic result. This follows the logic presented in the previous section about user distributions and proves the validity of the claim: the heuristic solutions presented in this work are highly affected by user density and are unable to balance

the load between regions, which leads to a poorer performance. On the other hand, optimized algorithms are able to exploit the regional characteristics of dense regions and greatly boost performance compared to more simple techniques.

For lower dimensionality scenarios (square and circle shapes), the results state once again the conclusions of the previous sections: once we move into 0 UD regions, the optimized algorithms tend to achieve solutions with slightly higher power compared to more simple approaches. Specifically, in scenarios with excess capacity, simple solutions allow for power savings due to the power-efficiency approaches considered, while in scenarios with low capacity, optimized resolution procedures are able to efficiently manage the system's resources to significantly raise the satellite's capabilities, thus highly increasing the number of users that can be served.

As a closing remark for experiment set B, the results show that, although the performance of the algorithms varies depending on the user distribution, the framework is able to provide a feasible and operable solution for the resource allocation problem in all cases. Even more, optimized implementations prove to improve system throughput compared to simple heuristics, independently on the user input and dimensionality scenario.

6.2 Sensitivity Analysis (Experiment Set C)

This Section details the results for the sensitivity analysis performed in this work, where sensitivity refers to the capacity of the framework to adapt to different parameters of the model. In this case, the parameters of the model denote all those operational variables defined by the hardware characteristics. All evaluations in this Section will be performed on the SES user distribution. This Section will be decomposed into seven different tests: one for each model parameter (i.e., number of satellites, number of gateways, number of beam channels, frequency reuse factor, bandwidth per beam channel, and half cone angle) plus one performing a combined analysis (number of beam channels plus bandwidth per beam channel). In each test, each resolution procedure will be evaluated under three or four parameter values, as specified in Table 5.3. Since solutions with higher number of beams have been shown to perform worse in an overwhelming amount of the experiments in set A, they have been left out of the experiments in set C. Finally, to ease comprehension of the results, most analysis will only include two versions of the resolution pipeline: one heuristic version that includes the Grid solution for the User Grouping, and a heuristic algorithm for the rest of the sub-problems, and one optimized version that corresponds to the best pipeline determined in Section 5.6.

6.2.1 Sensitivity test: Number of satellites

While the number of satellites in a constellation is usually an early decision in the development of a space project, analyzing how coverage varies depending on this factor can help assess the performance and validity of the framework under different conditions while informing future design decisions. This experiment evaluates the framework under a constellation with 6, 10, and 14 satellites. The rest of the satellite and orbital parameters are kept constant.

Figure 6-4 and Table B.4 present the results of the framework for the heuristic and fully optimized pipelines under different number of satellites. Since increasing the number of satellites actively boosts the capacity of the constellation, the coverage always improves when adding more spacecrafts. While the relation with power is not so apparent at first, it is important to note that the capacity of the system in terms of power also increases with the number of satellites. Therefore, while the total power

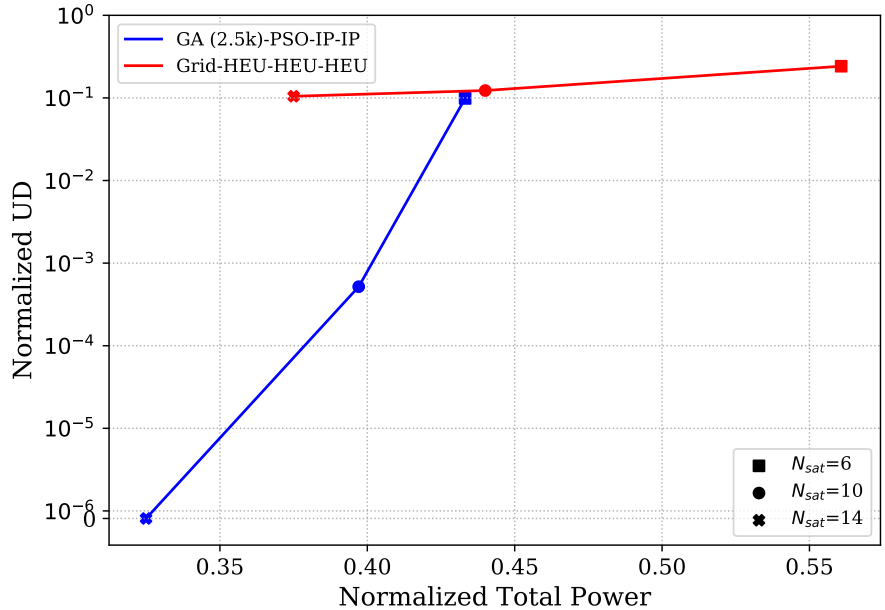


Figure 6-4: Performance comparison under different number of satellites. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with the number of satellites. The upper legend encodes each resolution procedure with a color, where each element is defined by a User Grouping, Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes the number of satellites considered in each evaluation with a shape.

may increase, it does so at a lower rate than the total power capacity, and thus the power per satellite will always be lower.

Turning our attention to the heuristic pipeline (red line), we observe that while the UD for 6 satellites is at 24%, increasing the number of satellites to 14 only reduces UD by 50% and the final configuration still has almost 12% of uncovered users. In other words, the heuristic allocations are unable to exploit the increased capacity of the system and are unable to deliver the expected performance when increasing the number of satellites. On the other hand, an optimized allocation is already able to achieve a 10% UD with only 6 satellites. It is important to emphasize that the optimized framework is able to achieve better performance with 6 satellites than the heuristics with 14 in terms of coverage. Even more, when increasing the number of satellites to 10, the system is able to achieve virtually total coverage, with only 0.1% of uncovered demand. In terms of power consumption, the optimized solution is also able to reduce this value between 10% and 23% over simple heuristics.

6.2.2 Sensitivity test: Number of gateways

Deciding how many gateways to use in a satellite constellation is one of the key drivers of system throughput [155]. Thus, analyzing how the framework performs under different ground station configurations can give significant insight on the behaviour of each resolution procedure and the overall capabilities of the framework. For this purpose, this experiment evaluates the different algorithm pipelines under four gateway configurations with 10, 20, 40, and 80 gateways.

Figure 6-5 and Table B.5 illustrate the results for the heuristic and fully optimized pipelines under different number of gateways. Similar to the previous case, increasing the number of gateways has always a positive impact in the total coverage. However, the effect on the power consumption is less apparent: while increasing the number of gateways in the low range (between 10 and 20) reduces the power consumption by allowing for a better use of the frequency spectrum, adding more ground stations after this point has the opposite effect, as the increase in capacity allows for more users to be added, which incur in a larger power value.

If we shift our focus on the heuristic pipeline (red line), we observe that the number of gateways plays a crucial role in the capacity of the system: while only 14% of the demand is covered with 10 gateways, this number rises up to 57% with

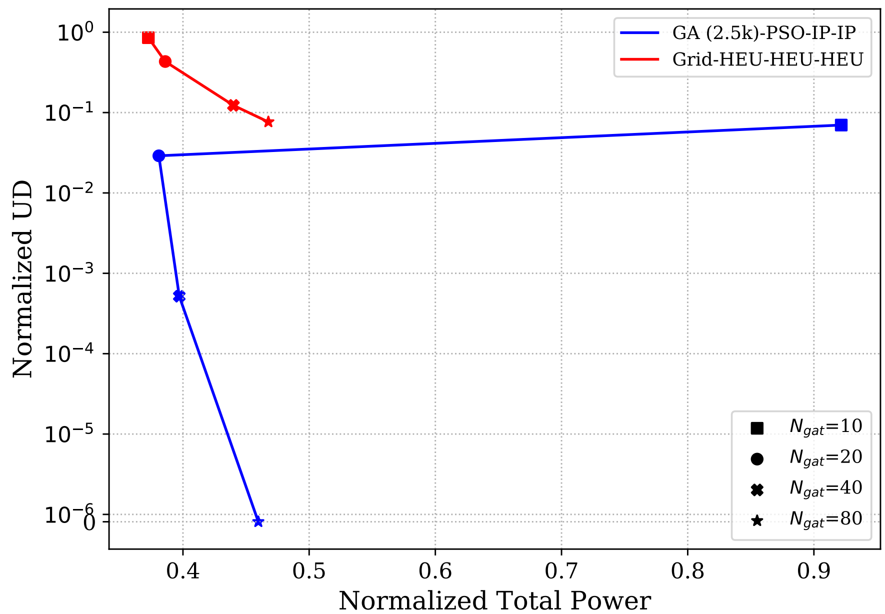


Figure 6-5: Performance comparison under different number of gateways. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with the number of gateways. The upper legend encodes each resolution procedure with a color, where each element is defined by a User Grouping, Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes the number of gateways considered in each evaluation with a shape.

20, 88% with 40, and 92% with 80. In other words, the system is able to multiply by more than 4 the capacity of the constellation only by increasing by a factor of 2 the number of gateways in the range from 10 to 20. Gains in the 20 to 40 range continue to be significant, being a factor of 1.5. After that, further increasing the capabilities of the ground segment report minimal gains on the throughput of the system. On the other hand, the optimized pipeline is able to obtain better coverage with 10 gateways than the heuristic results with 80, which demonstrates the capabilities of the optimization framework. Going from 10 to 20 gateways shifts the UD from 7%

to 3%, while the system achieves virtually total coverage with 40 gateways.

6.2.3 Sensitivity test: Number of beam channels

While the previous two tests cover important parameters outside the spacecraft, the on-board technology is what determines the capacity and flexibility of each satellite. Together with power, the frequency pool is the most important factor in determining the available resources. Specifically, this pool is limited by the total amount of frequency available (which connects to the number of beam channels and the bandwidth per beam channel), and the number of frequency reuses allowed by the hardware. The following tests analyze each of these three elements and their impact in the framework's performance.

As a pure measure of the total capacity, the number of beam channels determine how many frequency slots are available to use. This test assesses the performance of different algorithms with different number of beam channels (100, 150, and 200). Exceptionally, this test will evaluate 5 different pipelines, since this will ease the comprehension of how this parameter affects each optimization method.

Figure 6-6 and Table B.6 present the results for the 5 different resolution procedures under 3 different number of beam channels. As expected, increasing this factor (which corresponds to increasing capacity) always has a positive impact in the coverage of the constellation and reduces UD. However, its effect in power is unclear and depends on the exact characteristics of the scenario and resolution procedure. As a general trend, increasing the number of beam channels in solutions with higher UD will tend to trade power for UD, while in solutions with low UD will tend to reduce both metrics.

If we shift our focus onto the full heuristic resolution (red line), we observe a

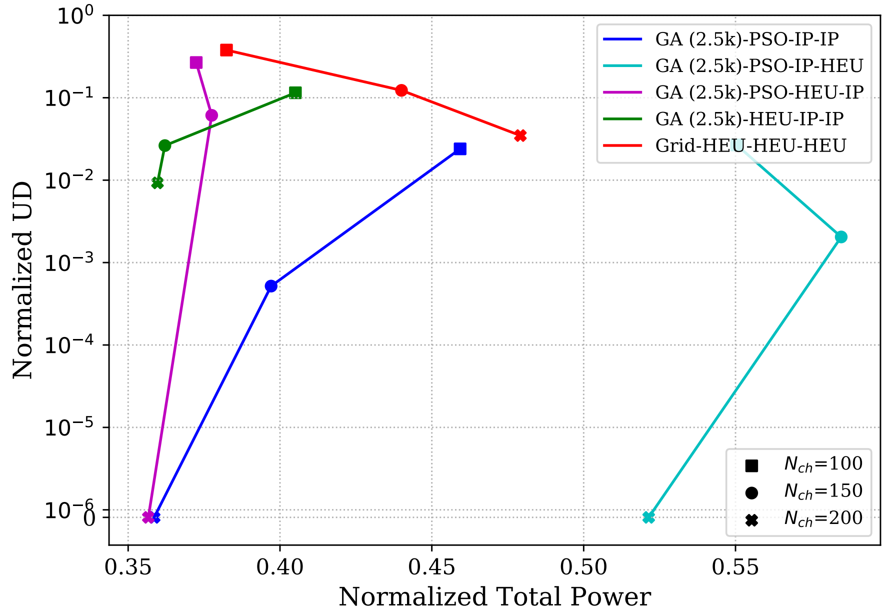


Figure 6-6: Performance comparison under different number of beam channels. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with the number of beam channels. The upper legend encodes each resolution procedure with a color, where each element is defined by a User Grouping, Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes the number of beam channels considered in each evaluation with a shape.

net increase of 41% (from 62% to 88%) in coverage when increasing the frequency pool to 1.5 its initial value, which practically corresponds to a linear gain. The gain observed when increasing up to 200 beam channels is significantly reduced, but still noticeable, reducing the UD to only 4%. All these improvements come at a power expense. Turning our attention to the fully optimized pipeline (blue line), we can make a similar statement as previous experiments regarding the capabilities of the optimization framework: the results show that a fully optimized pipeline with 100 beam channels is able to achieve better performance than the heuristic implementations with twice the number of channels. Even more, increasing the number of channels also implies a reduction in power of 22% when doubling the system's capacity.

The results involving different resolution procedures (cyan, purple, and green lines) reinforce the conclusions from Section 5.6. The most crucial sub-problem to reduce UD is the Gateway Routing algorithm, followed by the Satellite Routing and Frequency Assignment. On the other hand, the Frequency Assignment is critical to reduce the power consumption of the system, being able to achieve power reductions between 17% and 32% solely by optimizing frequency. Once more, if the system is able to cover all the users (i.e., the solution is in the 0 UD region), the best resolution procedure may involve heuristics that improve the link budget link quality in a way that reduces power consumption (in this case the best solution for the 200 beam channel scenario corresponds to the purple pipeline).

6.2.4 Sensitivity test: Bandwidth per beam channel

Similar to the number of beam channels, the bandwidth per beam channel has a direct impact in the total capacity of the system. In this case, instead of the number of slots available, the bandwidth per beam channel affects the size of each slot. This test assesses the performance of different algorithms with different bandwidth per beam channel (10 MHz, 15 MHz, and 20 MHz).

Figure 6-7 and Table B.7 show the results for the the heuristic and the fully optimized pipelines under 3 different bandwidths per beam channel. As expected, an increase in the total bandwidth per channel has a positive impact in the total coverage of the system, since the system throughput is boosted. On the other hand, the effect on power is less clear and depends on the exact configuration and resolution

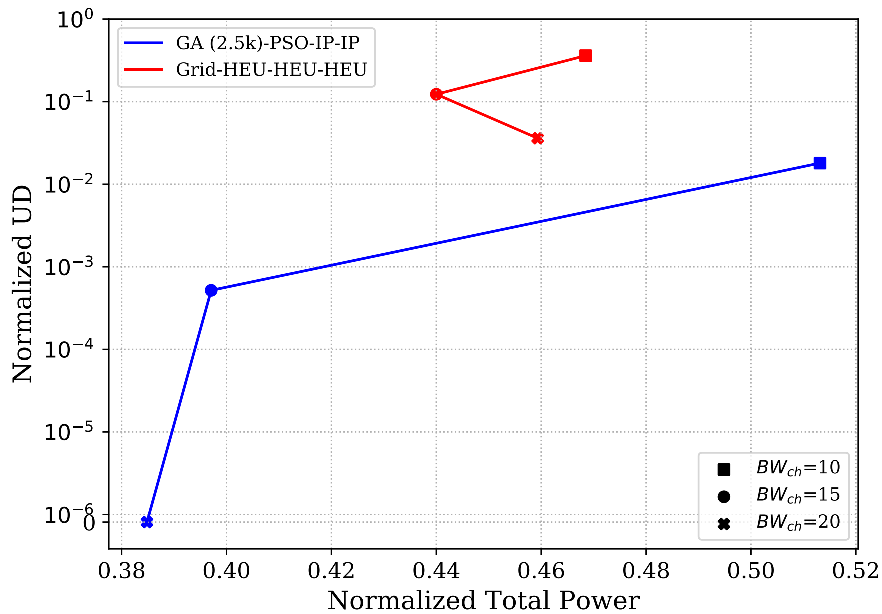


Figure 6-7: Performance comparison under different bandwidths. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with the bandwidth. The upper legend encodes each resolution procedure with a color, where each element is defined by a User Grouping, Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes the bandwidth considered in each evaluation with a shape.

procedure. While increasing the bandwidth always helps reducing power in the optimized resolution procedure, its effect on the heuristic solution does not follow a clear pattern.

If we turn our attention to the heuristic solution (red line), we observe that the change from 10 MHz per beam channel to 15 MHz reduces the UD from 36% to 12%, which translates to a 33% increase in total coverage. Further increases in bandwidth also achieve up to 9% improvement in coverage. However, similarly as the previous case, the optimized algorithm is able to achieve better coverage and power consumption with 10 MHz per beam channel compared to the heuristic solution with 20 MHz per beam channel. In other words, even with half the system capacity, the optimized pipeline obtains improved performance over the heuristic solutions. Furthermore, increasing the bandwidth has a positive effect in both UD and power consumption, as the optimization algorithms are able to fully exploit the flexibilities of the system.

6.2.5 Sensitivity test: Frequency reuse factor

This test evaluates the sensitivity of the results against different reuse factor values (6, 12, and 18). While the number of beam channels and the bandwidth per beam channel have clear implications in the total capacity of the system, the impact of the frequency reuse factor depends on the user distribution and its effectiveness depends on the user density. Specifically, if all users were to be clumped into the same region, a frequency reuse factor larger than 1 would have virtually no impact, since all the users would have interference restrictions between them. On the other end, when users are highly spread amongst zones, the frequency reuse factor has a significant contribution to the total capacity.

Figure 6-8 and Table B.8 illustrate the results for the fully optimized and heuristic pipelines on different frequency reuse factors. As expected, the frequency reuse factor has a positive impact in UD. However, this effect is severely limited when considering the fully heuristic resolution procedure, since those techniques tend to obtain a larger number of frequency-related constraints, which limits the impact of reusing frequency. Specifically, when doubling the frequency reuse factor from 6 to 12, the constellation is able to improve coverage by 7% when using the heuristic resolution procedure. Furthermore, there is no significant improvement when changing the frequency reuse factor from 12 to 18. On the other end, the fully optimized algorithms

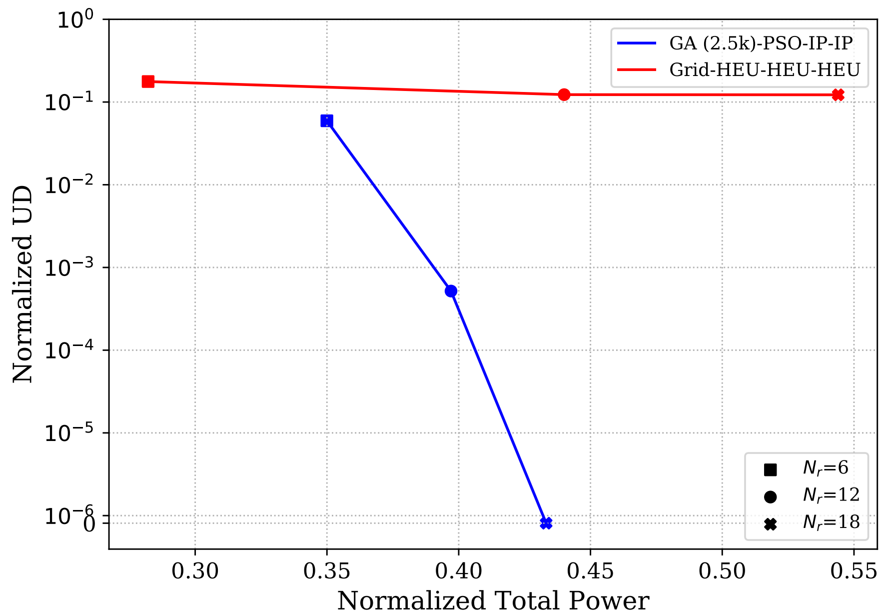


Figure 6-8: Performance comparison under different frequency reuse factors. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with the frequency reuse factors. The upper legend encodes each resolution procedure with a color, where each element is defined by a User Grouping, Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes the frequency reuse factors considered in each evaluation with a shape.

can make better use of this flexibility and exploit the regional characteristics of the user distribution to maximize system capacity. An optimized framework is able to achieve higher coverage with a frequency reuse factor of 6 compared to an heuristic resolution with three times the reuse factor. Even more, when the frequency reuse factor is increased to 12, virtually all the demand is met.

6.2.6 Sensitivity test: Half cone angle

While the relation between the frequency pool and total capacity is fairly clear, the effects of the beam's half cone angle are not apparent. On one hand, a larger footprint allows for a better grouping of beams and, thus, less beams, which has a significant contribution to performance as discussed in Section 5.6. On the other hand, smaller footprints allow for higher gains and better link quality while reducing the overall interference with nearby beams. The objective of this experiment is to assess the impact of the half cone angle on the developed framework and its implications in performance. The experiment will test the fully optimized pipeline and the heuristic resolution on three different half cone angle configurations: 0.5°, 1°, and 1.5°.

Figure 6-9 and Table B.9 present the results for the two main resolution procedures on the three half cone angle configurations. According to the results, reducing the half cone angle has a positive impact on UD and offers a significant increase in coverage without changing most of the configuration of the payload. The advantages of having a reduced footprint in terms of higher gain and lower interference outweigh the benefits of grouping the users further in terms of total coverage. On the other hand, reducing the half cone angle implies a higher power requirement, since we need more individual beams. This implies that once we reach 0 UD, we should not decrease the half cone angle further without trying to increase demand, since that would only result in a higher power consumption.

If we focus on the heuristic results (red line), we observe a 19% increase in coverage when changing from 1.5° to 1°. There is still benefit in further reducing the half cone angle, but the impact is less significant. Similarly to the previous results, the optimized pipeline is able to achieve improved performance in both metrics with a half cone angle of 1.5° (blue cross) compared to the heuristic resolution with a half

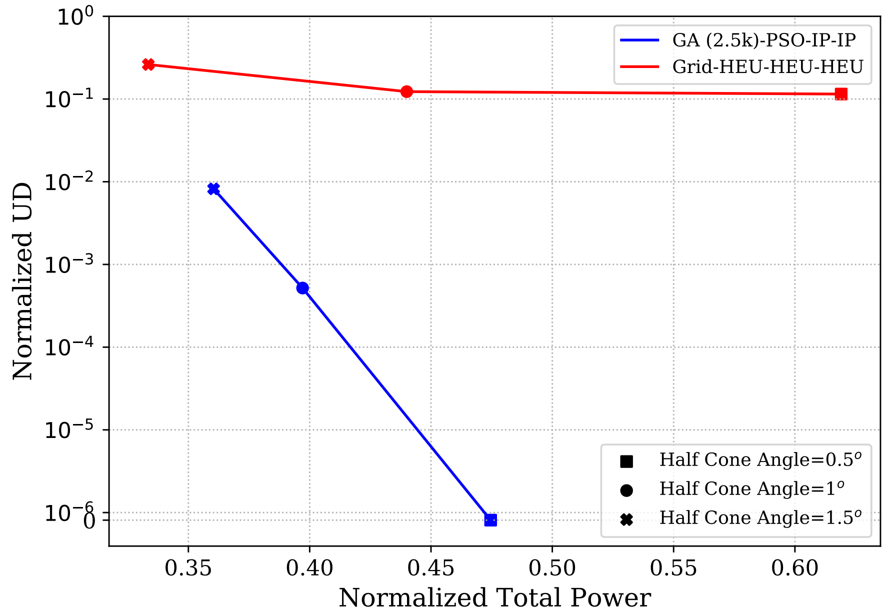


Figure 6-9: Performance comparison under different half cone angles. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with the half cone angle. The upper legend encodes each resolution procedure with a color, where each element is defined by a User Grouping, Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes the half cone angle considered in each evaluation with a shape.

cone angle of 0.5° (red square).

6.2.7 Sensitivity test: Number of and bandwidth per beam channel

While one-at-a-time experiments evaluate the tendency of change of single factors, tests combining multiple elements are necessary to understand how the different components interact. This last experiment assesses the interaction between the number of beam channels and the bandwidth assigned to each channel. Specifically, the ob-

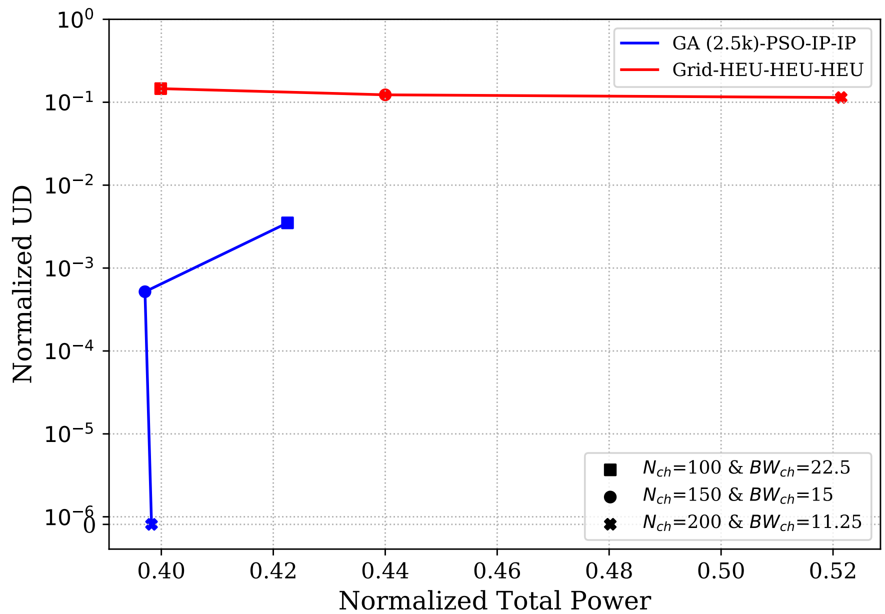


Figure 6-10: Performance comparison under different number of beam channels and bandwidths. Each point represents a different experiment, and each line represents the tendency of each resolution procedure with both the number of beam channels and bandwidth. The upper legend encodes each resolution procedure with a color, where each element is defined by a User Grouping, Satellite Routing, Gateway Routing, and Frequency Assignment algorithm, in order. The bottom legend encodes the number of and bandwidth per beam channel considered in each evaluation with a shape.

jective is to understand the tendency of the system when the total capacity is fixed (i.e., when the product of the number of beam channels and the bandwidth per beam channel is predefined), but we have the ability to decide how to split the frequency pool between those two factors. For this purpose, the fully optimized an heuristic resolution procedures have been evaluated in three different configurations: 1) 100 beam channels and 22.5 MHz per channel, 2) 150 beam channels and 15 MHz per channel, and 3) 200 beam channels and 11.25 MHz per channel.

Figure 6-10 and Table B.10 illustrate the results for the optimized and heuristic pipelines under the three aforementioned configurations. The first aspect to notice is that reducing the size of the channel and increasing the number of channels has a clear positive impact on the UD. This can be explained by the fact that, since frequency can only be split in a finite number of channels but the demand is continuous, configurations with higher flexibility will be able to better match the demand and make more appropriate use of the available resources. Thanks to this, the results exhibit minor improvements when selecting configurations with more and smaller beam channels. On the other hand, the effects on power are unclear and depends on the exact configuration. Increasing the number of slots implies a significant power increase when using the heuristic resolution while obtaining only a minor improvement in coverage. On the other hand, increasing the number of slots evokes a reduction in power when using the optimized pipeline.

6.2.8 Sensitivity summary

The experiments detailed in the previous sections meticulously explain the influence of each factor on each resolution procedure and on the final operational plan. The results show that an optimized pipeline is able to outperform heuristic solutions even with limited hardware. Specifically, Table 6.3 summarizes these findings by presenting the comparison between a heuristic pipeline and an optimized resolution with reduced hardware (i.e., with less number of satellites, number of gateways, number of beam channels, bandwidth per beam channel, or frequency reuse factor, or with a larger half cone angle). In particular, using state-of-the-art algorithms for the different sub-problems can yield solutions with improved coverage with 57% less satellites, 88% less gateways, half the frequency spectrum, a third of the frequency reuse factor, or three times the half cone angle. These results showcase the importance of

Model Parameter	Number of satellites	Number of gateways	Number of beam channels	Bandwidth per beam channel	Frequency reuse factor	Half cone angle
Hardware factor	1	1	1	1	1	1
UD	-100%	-100%	-100%	-100%	-100%	-100%
Power	-13%	-2%	-25%	-16%	-20%	-23%
Hardware factor	2.3	8	2	2	3	3
UD	-6%	-8%	-31%	-50%	-52%	-93%
Power	15%	97%	-4%	12%	-36%	-42%

Table 6.3: Impact of choosing the fully optimized pipeline over the heuristic solution when analyzing different model parameters. The hardware factor denotes the improvement in hardware (i.e., reduction in number of satellites, number of gateways, number of beam channels, bandwidth per beam channel, or frequency reuse factor, or increase in half cone angle) on the heuristic resolution compared to the optimized algorithm chain. For example, a hardware factor of 2 on the number of beam channels implies that the cell is comparing the performance of the optimized pipeline against a heuristic solution which has twice the number of beam channels.

choosing the right algorithm to carry out the resource allocation, especially when the capacity of the system is constrained.

6.3 Sensitivity of the framework results to minor user changes

Up until this point, this Thesis has discussed how different resolution procedures achieve different performances and trade-offs on a system level. However, once the framework produces a feasible and operable plan, it is important to understand how that plan may change under minor changes on the user distribution. For example, if the framework computed a plan that assumed a certain user position, but it turns out to be slightly different, it is important to understand how different would the new plan be compared to the pre-computed one. For this purpose, this Section analyzes the effects of minor changes in the user distribution on each of the optimization algorithms.

6.3.1 Sensitivity of the framework results to multiple executions

Since most of the optimization algorithms considered in this work rely on random effects to achieve improved solutions, it is important to first understand how solutions change under independent executions. In other words, if we execute the framework repeatedly, how different are the solutions that we obtain. To analyze this effect, each optimization algorithm has been executed multiple times under the same initial conditions, and the similarity of solutions has been studied. Specifically, each method has been executed a total of 5 times and the solutions have been compared using a *distance* metric that varies from problem to problem:

- User Grouping: Since the User Grouping problem revolves around clustering the user terminals, this work uses the Rand Index (RI, [158]) to determine how similar are two User Grouping solutions. Specifically, this metric performs a pairwise comparison between all the user terminals to determine how similar the clusterings are. The value presented in the results correspond to $1-\text{RI}$ as it reflects better the distance between two solutions.

- Satellite Routing: Since the result of the Satellite Routing problem is an array of serving times, the intuitive distance metric is to determine the difference between the two arrays using Euclidean distance.

- Gateway Routing: The results of the gateway Routing problem does not include any random effect and can be solved optimally under the formulation presented.

Thus, it can be left out of this analysis.

- Frequency Assignment: Determining the difference between two frequency plans is no simple task, since one could argue that two plans in which two beams are swapped are effectively the same. Thus, this work will consider only the difference in bandwidth: once each frequency plan is computed, an array containing the bandwidth for each beam can be created and compared to other plans using Euclidean distance.

Table 6.4 illustrates the difference between solutions on multiple independent executions of the algorithms under identical input parameters. As shown, the GA for User Grouping produces solutions that are basically identical, with only a few beams changing between solutions. The PSO for Satellite Routing has slightly more variability between different runs due to the fast converge of the PSO to a local optima. Still, the results show a $< 10\%$ in difference between solutions. On the other hand, the MILP for the Frequency Assignment proves to be highly variable in terms of specific solution, while keeping similar solution quality. In other words, since many Frequency Assignment solutions achieve the same objective value, the

Algorithm	**Mean distance between solutions**	**Standard deviation**
Genetic Algorithm for User Grouping	0.002	4×10^{-5}
Particle Swarm Optimization for Satellite Routing	0.089	0.001
Mixed Integer Programming for Frequency Assignment	0.212	0.062

Table 6.4: Sensitivity of the optimization algorithms on multiple independent executions under identical inputs. All other algorithms not shown in this Table but considered in this work are deterministic, and therefore have no difference between executions.

Algorithm	**Distance between solutions**
Genetic Algorithm for User Grouping	0.236
Particle Swarm Optimization for Satellite Routing	0.158
Mixed Integer Programming for Frequency Assignment	0.465

Table 6.5: Sensitivity of the optimization algorithms on two independent executions under slightly different initial conditions. All other algorithms not shown in this Table but considered in this work are deterministic, and therefore have no difference between executions.

algorithm may obtain qualitatively very different solutions while achieving a similar result. This produces a high variance in the final operational plan.

6.3.2 Sensitivity of the framework results to minor user changes

In order to analyze how the resource allocation plan changes with minor changes on the users, this Section will compare two solutions to the framework, one corresponding to the full optimized pipeline under the same conditions as Experiment A in Section 5.6, and a second scenario based on the exact same use case, but where 1% of the users suffer some minor changes in location and demand. Following the same distance metrics as the previous Section, Table 6.5 presents the distance between the two executions on the different steps of the framework. As shown, changing the position and demand of 1% of the users produces significant changes in the final solution. These variations imply a 24% difference in the User Grouping plan, a 16% different in the Satellite Routing plan, and a 47% difference in the Frequency Assignment plan. Note that while the Satellite Routing algorithm is the least affected, the Frequency Assignment can be deemed as significantly different between the two executions.

Chapter 7

Conclusions

7.1 Thesis summary

This Thesis describes a comprehensive framework to solve the long-horizon Resource Allocation (RA) problem in the context of High Throughput Satellite Constellations and analyses its performance over different user distributions and constellation configurations. The first Chapter starts with an overview of the evolution of the space communication market over the past years and a brief outline on the characteristics of the next generation of space constellations. After discussing the role of autonomous decision making algorithms, the Thesis proceeds with a synopsis of the most relevant satellite communications concepts and ideas to give an engineering background based on which this work builds upon. Then, the RA problem is described and decomposed into smaller sub-problems, which are grouped into two categories (short or long) based on the time-horizon that they aim to make decisions for. This Chapter ends by highlighting the main objectives of this Thesis.

Chapter 2 provides the relevant literature for the RA problem in satellite communications and briefly describes the contributions and limitations of each work. After analyzing the literature related to each individual sub-problem as decomposed in Chapter 1, works combining multiple sub-problems are dissected and discussed. Based on this study, this Chapter highlights the research gap that needs to be examined before the academia knowledge can be implemented onto industry. The Chapter concludes with an explanation of how this Thesis addresses this disparity and the specific topics that drive the construction of the long-horizon RA framework.

Next, Chapter 3 constitutes the theoretical back-bone of this work. First, this Chapter starts with an initial discussion of how the long-horizon RA problem can be addressed with a sequential resolution of the individual sub-problems. Based on this concept, the variables, constraints, and objectives of each individual sub-problem are detailed. In addition, to ensure proper interaction between the different elements, the interfaces between the components are examined. Then, the evaluation metrics used for the framework are introduced. This Chapter ends with an extensive discussion on the different assumptions that the framework or the different sub-problems build upon and how those can be addressed to improve the generality of the results.

The fourth Chapter builds upon the previous one and introduces all the resolution algorithms that aim to solve each individual problem. For each single component in the framework, this Chapter details techniques that range from simple heuristics to state-of-the-art optimization methods. This Chapter concludes with a simple method to compute the metrics which will allow to assess the potential and performance of the discussed framework.

Then, Chapter 5 details the validity checks and experiment configuration that determine the behaviour of the framework under different conditions. First, this Chapter starts with a brief discussion on how to ensure correctness on the final solution. Second, the test procedures are explained. Third, the main inputs and configurations of the model for the different experiments are specified. Fourth, a summary of the full set of experiments is introduced. And, fifth, this Chapter concludes with a first analysis of the framework under two representative conditions and a detailed discussion on the results and implications.

Finally, Chapter 6 constitutes the main performance analysis of this Thesis. The framework and the different sub-problem algorithms have been evaluated under different user distributions and constellation model configurations to determine their robustness and sensitivity against the different parameters. The first set of experiments validates the framework against different user arrangements and dimensionality, while the second set checks the importance and effectiveness of each system design variable.

7.2 Main findings

The objective of this Thesis has been to develop a framework to solve the long-horizon Resource Allocation (RA) problem for satellite communications and assess the validity and performance of the framework and the underlying algorithms under different user and model conditions. The results prove the soundness and effectiveness of such framework to obtain a feasible plan for representative satellite operations in an automated way.

As a first conclusion, decomposing the long-horizon RA problem into smaller sub-problems and sequentially solving them has been proven to be a valid, effective, and scalable, tool to obtain an operable resource allocation plan. The assumptions and simplifications explained along the framework allow for the transformation of the complex long-horizon RA conundrum into a tractable and solvable problem, which can be optimized using independent state-of-the-art optimization techniques.

According to the reviewed literature, this Thesis is the first public work that attempts to solve the complete long-horizon RA problem, and it does so using modern optimization techniques to maximize the system's performance in a variety of dimensionality scenarios.

Second, optimized algorithms have proven to be key to maximize system's capacity. While heuristic approaches can give a feasible solution and a valid operational plan in all scenarios, state-of-the-art techniques prove to significantly increase the performance of the allocation, being able to quadruple the system's capacity with a third of the power consumption in low-capacity scenarios, and obtaining an 86% power reduction while maximizing coverage in high-capacity situations. Using a refined User Grouping algorithm compared to trivial approaches has proven to be the key driver in those reductions. Specifically, using a solution with low number of beams shows to reduce both UD and power consumption between 75% and 90% in all cases with respect to the simple solution of assigning one user per beam. Improving the Gateway Routing allocation using an optimized algorithm offers significant reductions in UD, at the expense of a slight increase in power. At a lower scale, optimizing the Satellite Routing solution also trades higher power for lower UD. Finally, using optimized tools for the Frequency Assignment problem offers meaningful gains in both metrics and it is a key power driver in high capacity scenarios.

Third, the framework and algorithms tested prove to be robust against different user distributions and dimensionality scenarios. The sequential resolution of independent sub-problems proves to give satisfactory results independently of the user input. Even more, optimized solutions prove to systematically outperform heuristic resolutions in terms of UD in high demand scenarios and in terms of power when $UD = 0$, independently of the number of users considered or where those users are located. Nonetheless, the user distribution does affect the performance of the resolution procedure: while heuristics tend to deal better with scenarios with spread demand, the optimized techniques are able to exploit the regional benefits of dense regions and maximize system's throughput in scenarios where the users are concentrated in small areas. While simple heuristic might provide *good enough* results in widely spread user distributions, optimization techniques are mandatory to accommodate demand in dense regions.

Fourth, the framework and algorithms tested prove to be robust against different model parameters and design decisions. Furthermore, the state-of-the-art resolution algorithms consistently prove to improve coverage with reduced system capabilities over simple heuristics with improved hardware. Specifically, results show that the fully optimized resolution procedure is able to find a solution with better:

- Coverage with 6 satellites than a heuristic approach with 14.
- Coverage with 10 gateways than a heuristic approach with 80.
- Performance in both power and UD with 100 beam channels than a heuristic approach with 200 beam channels.
- Coverage with 10 MHz per channel than a heuristic approach with 20 MHz per channel.
- Performance in both power and UD with a reuse factor of 6 than a heuristic approach with a reuse factor of 18.
- Performance in both power and UD with a half cone angle of 1.5° than a heuristic approach with a half cone angle of 0.5°.

Which highlights the potency of using optimized algorithms for the individual subproblems. Furthermore, while increasing the frequency spectrum of the spacecraft is an obvious way of improving the performance of the constellation, the results show that enlarging the ground or space segment can also give interesting opportunities to maximize capacity. Improving the beam-forming hardware to achieve higher gains and smaller footprints can also yield significant gains. On the other hand, the frequency reuse factor has a lower impact on performance and should be considered only in certain scenarios.

7.3 Future work

Based on the results and main conclusions of this work, different directions of future research have been identified:

- Explore relaxations of the different assumptions and simplifications to better understand the limits of the framework and extend its functionality to more general configurations. While the framework has been constructed using both general and problem-specific assumptions, a future study could enlighten ways to extend the capabilities of the sequential resolution to allow for improved or more general solutions. Specifically, constellations with non-static users (such as planes or ships), multiple orbital planes (with potentially multiple altitudes), or that use inter-satellite links may be of high interest in the upcoming years.

- Extension of the results with different algorithms. While this work has considered state-of-the-art optimization techniques for each individual sub-problem, future research may investigate the impact of different techniques on the overall long-horizon resource allocation problem. Furthermore, an interesting research direction is to assess how the improvements on the individual sub-problems affect the general solution.

- Extension and generalization of the framework with different formulations. While the problem-specific mathematical formulations have been taken from recent literature, the independence between problems allows for an adaptive framework in which different formulations can be considered. Future research may include different formulations that tackle each sub-problem with a more realistic formulation.

- Extension of the results with different user cases and operational scenarios. While this Thesis has evaluated the framework under different conditions and configurations, further intelligence on the real conditions will allow for a better assessment of the capabilities of the framework. Furthermore, while this Thesis has considered the O3b mPower constellation as the reference for experiments, future work may include other modern constellations to assess the capabilities of the framework to adapt to different environments.

- Inclusion of uncertainty. While this work has considered that the demand of the users is fixed and known, operational reality is usually more complex. Exploring the capabilities of the framework to adapt to different uncertainty levels is a candidate direction of future work.

- Extension of the framework to the short-horizon RA problem. While this work has focused on the long-horizon RA problem, the real-time short-horizon problem still needs to be addressed to maximize the system's performance. Studying how both problems interact and how one can inform the other will be necessary to exploit the system's flexibilities to the limit.

Appendices

Appendix A

Mathematical Transformations

The objective of this Appendix is to introduce the reader to some useful mathematical adjustments to transform a non-linear formulation to a linear one.

A.1 Logic operations

A.1.1 OR

Let us define a binary variable x that is 1 only when either y or z are 1, 0 otherwise. This can be linearly formulated as:

$$x \leq y + z \leq 2x \tag{A.1}$$

While this formulation is very simple, it can be extended to N variables. That is, suppose that x denotes the or operation between $y_i,\ i \in \{1, ..., N\}$. Then:

$$x \leq \sum_i y_i \leq Nx \tag{A.2}$$

Note that, in this case, x denotes an auxiliary variable. If x is not a decision or variable, but rather fixed and given by the problem, it is sufficient to fix the numerical value. For example, if we need to enforce that either y or z to be 1, then we can write this restriction as:

$$1 \leq y + z \tag{A.3}$$

A.1.2 AND

Let us define a binary variable x that is 1 only when both y and z are 1, 0 otherwise. This can be linearly formulated as:

$$2x \leq y + z \leq x + 1 \tag{A.4}$$

If instead on two factors, we have an array of N elements denoted as $y_i,\ i \in \{1, ..., N\}$:

$$Nx \leq \sum_i y_i \leq x + N - 1 \tag{A.5}$$

Similarly to the previous case, if x is given by the model or by the definition of the problem, the it is enough to fix the numerical value of the variable.

A.2 Activation variables

A.2.1 Transforming an inequality into a binary variable

Let us define x as a binary variable that is equal to 1 when $a \leq b$ and 0 otherwise. Mathematically, this can be linearly expressed with a big-M notation:

$$-Mx \leq a - b \leq M(1 - x) \tag{A.6}$$

Where M is a sufficiently large number so that $M \geq |a-b| \; \forall a, b$.

A.2.2 Transforming an equality into a binary variable

Let us define x as a binary variable that is equal to 1 when $a = b$ and 0 otherwise. Although the formulation is slightly more complex than the previous one, there is a simple trick that allows us to use the same notation: x can be defined as the union of $a \leq b$ and $a \geq b$. Then, we can use the previous results to obtain a binary variable such that $x^- = 1$ when $a \leq b$, 0 otherwise, and $x^+ = 1$ when $a \geq b$, 0 otherwise. Finally, the value of x can be derived as the AND operation between $x^- = 1$ and $x^+ = 1$.

Appendix B

Complete results

This chapter presents the complete results for the optimization framework presented under the different conditions explained in Chapter 5.

Inputs					**Outputs**	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Number of users	Power	UD
GA (2.5k)	PSO	IP	IP	200	**0.042**	**0.000**
Grid	HEU	HEU	HEU	200	0.161	**0.000**
GA (2.5k)	PSO	IP	IP	2000	**0.258**	**0.000**
Grid	HEU	HEU	HEU	2000	0.421	**0.000**
GA (2.5k)	PSO	IP	IP	20000	**0.397**	**0.001**
Grid	HEU	HEU	HEU	20000	0.409	0.127

Table B.1: Detailed numbers on the performance comparison under different dimensionality scenarios on the SES dataset. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Number of users	Power	UD
GA (200)	PSO	IP	IP	200	**0.034**	**0.000**
GA (200)	PSO	IP	HEU	200	0.183	**0.000**
GA (200)	PSO	HEU	IP	200	0.035	**0.000**
GA (200)	HEU	IP	IP	200	0.035	**0.000**
GA (200)	HEU	HEU	HEU	200	0.133	**0.000**
GA (2k)	PSO	IP	IP	2000	0.193	**0.000**
GA (2k)	PSO	IP	HEU	2000	0.274	**0.000**
GA (2k)	PSO	HEU	IP	2000	0.187	**0.000**
GA (2k)	HEU	IP	IP	2000	**0.164**	**0.000**
GA (2k)	HEU	HEU	HEU	2000	0.257	0.000
GA (4k)	PSO	IP	IP	20000	0.434	**0.106**
GA (4k)	PSO	IP	HEU	20000	0.489	0.122
GA (4k)	PSO	HEU	IP	20000	0.388	0.241
GA (4k)	HEU	IP	IP	20000	0.329	0.209
GA (4k)	HEU	HEU	HEU	20000	**0.311**	0.323

Table B.2: Detailed numbers on the performance comparison under different dimensionality scenarios on the Population dataset. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

		Inputs			Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Number of users	Power	UD
GA (200)	PSO	IP	IP	200	0.034	**0.000**
GA (200)	PSO	IP	HEU	200	0.154	**0.000**
GA (200)	PSO	HEU	IP	200	0.032	**0.000**
GA (200)	HEU	IP	IP	200	**0.030**	**0.000**
GA (200)	HEU	HEU	HEU	200	0.110	**0.000**
GA (2k)	PSO	IP	IP	2000	0.173	**0.000**
GA (2k)	PSO	IP	HEU	2000	0.270	**0.000**
GA (2k)	PSO	HEU	IP	2000	0.178	**0.000**
GA (2k)	HEU	IP	IP	2000	**0.152**	0.001
GA (2k)	HEU	HEU	HEU	2000	0.231	0.002
GA (3k)	PSO	IP	IP	20000	0.387	**0.194**
GA (3k)	PSO	IP	HEU	20000	0.475	0.196
GA (3k)	PSO	HEU	IP	20000	0.360	0.323
GA (3k)	HEU	IP	IP	20000	**0.245**	0.383
GA (3k)	HEU	HEU	HEU	20000	0.262	0.448

Table B.3: Detailed numbers on the performance comparison under different dimensionality scenarios on the Uncovered dataset. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Number of satellites	Power	UD
GA (2.5k)	PSO	IP	IP	6	0.433	0.099
Grid	HEU	HEU	HEU	6	0.561	0.242
GA (2.5k)	PSO	IP	IP	10	0.397	0.001
Grid	HEU	HEU	HEU	10	0.440	0.123
GA (2.5k)	PSO	IP	IP	14	**0.325**	**0.000**
Grid	HEU	HEU	HEU	14	0.375	0.105

Table B.4: Detailed numbers on the performance comparison under different number of satellites. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Number of gateways	Power	UD
GA (2.5k)	PSO	IP	IP	10	0.922	0.070
Grid	HEU	HEU	HEU	10	**0.373**	0.855
GA (2.5k)	PSO	IP	IP	20	0.381	0.029
Grid	HEU	HEU	HEU	20	0.386	0.432
GA (2.5k)	PSO	IP	IP	40	0.397	0.001
Grid	HEU	HEU	HEU	40	0.440	0.123
GA (2.5k)	PSO	IP	IP	80	0.460	**0.000**
Grid	HEU	HEU	HEU	80	0.468	0.076

Table B.5: Detailed numbers on the performance comparison under different number of gateways. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Number of beam channels	Power	UD
GA (2.5k)	PSO	IP	IP	100	0.459	0.024
GA (2.5k)	PSO	IP	HEU	100	0.550	0.027
GA (2.5k)	PSO	HEU	IP	100	0.372	0.268
GA (2.5k)	HEU	IP	IP	100	0.405	0.115
Grid	HEU	HEU	HEU	100	0.382	0.379
GA (2.5k)	PSO	IP	IP	150	0.397	0.001
GA (2.5k)	PSO	IP	HEU	150	0.585	0.002
GA (2.5k)	PSO	HEU	IP	150	0.378	0.061
GA (2.5k)	HEU	IP	IP	150	0.362	0.026
Grid	HEU	HEU	HEU	150	0.440	0.123
GA (2.5k)	PSO	IP	IP	200	0.359	**0.000**
GA (2.5k)	PSO	IP	HEU	200	0.521	**0.000**
GA (2.5k)	PSO	HEU	IP	200	**0.357**	**0.000**
GA (2.5k)	HEU	IP	IP	200	0.360	0.009
Grid	HEU	HEU	HEU	200	0.479	0.035

Table B.6: Detailed numbers on the performance comparison under different number of beam channels. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Bandwidth per beam channel [Mhz]	Power	UD
GA (2.5k)	PSO	IP	IP	10	0.513	0.018
Grid	HEU	HEU	HEU	10	0.468	0.361
GA (2.5k)	PSO	IP	IP	15	0.397	0.001
Grid	HEU	HEU	HEU	15	0.440	0.123
GA (2.5k)	PSO	IP	IP	20	**0.385**	**0.000**
Grid	HEU	HEU	HEU	20	0.459	0.036

Table B.7: Detailed numbers on the performance comparison under different bandwidth per beam channel. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Frequency reuse factor	Power	UD
GA (2.5k)	PSO	IP	IP	6	0.350	0.059
Grid	HEU	HEU	HEU	6	**0.282**	0.177
GA (2.5k)	PSO	IP	IP	12	0.397	0.001
Grid	HEU	HEU	HEU	12	0.440	0.123
GA (2.5k)	PSO	IP	IP	18	0.433	**0.000**
Grid	HEU	HEU	HEU	18	0.544	0.122

Table B.8: Detailed numbers on the performance comparison under different frequency reuse factors. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Half cone angle [^{o}]	Power	UD
GA (2.5k)	PSO	IP	IP	0.5	0.475	**0.000**
Grid	HEU	HEU	HEU	0.5	0.619	0.115
GA (2.5k)	PSO	IP	IP	1.0	0.397	0.001
Grid	HEU	HEU	HEU	1.0	0.440	0.123
GA (2.5k)	PSO	IP	IP	1.5	0.360	0.008
Grid	HEU	HEU	HEU	1.5	**0.334**	0.261

Table B.9: Detailed numbers on the performance comparison under different half cone angles. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

Inputs					Outputs	
User Grouping	Satellite Routing	Gateway Routing	Frequency Assignment	Number of and Bandwidth per beam channel	Power	UD
GA (2.5k)	PSO	IP	IP	100 / 22.5	0.423	0.004
Grid	HEU	HEU	HEU	100 / 22.5	0.400	0.146
GA (2.5k)	PSO	IP	IP	150 / 15	**0.397**	0.001
Grid	HEU	HEU	HEU	150 / 15	0.440	0.123
GA (2.5k)	PSO	IP	IP	200 / 11.25	0.398	**0.000**
Grid	HEU	HEU	HEU	200 / 11.25	0.521	0.114

Table B.10: Detailed numbers on the performance comparison under different number of and bandwidth per beam channel. The leftmost columns indicate the resolution procedure chosen for each sub-problem (plus the number of beams selected for the User Grouping problem). The fifth column indicates the variable that is being tuned in the experiments. The two rightmost columns indicate the metrics of the framework under the established conditions.

Bibliography

[1] Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022," tech. rep., Cisco, 2019.

[2] I. del Portillo, S. Eiskowitz, E. F. Crawley, and B. G. Cameron, "Connecting the other half: Exploring options for the 50% of the population unconnected to the internet," *Telecommunications Policy*, vol. 45, no. 3, p. 102092, 2021.

[3] E. W. Ashford, "Non-Geo systems—where have all the satellites gone?," *Acta Astronautica*, vol. 55, no. 3-9, pp. 649–657, 2004.

[4] K. T. Li, C. A. Hofmann, F. Völk, and A. Knopp, "Techno-economic design aspects of satellite mega-constellations for 6g services," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, pp. 01–06, IEEE, 2021.

[5] Space Exploration Holdings, LLC, "SAT-LOA-20161115-00118." `http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATLOA2016111500118`, 2016. Accessed: 2020-12-09.

[6] Space Exploration Holdings, LLC, "SAT-MOD-20181108-00083." `http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2018110800083`, 2018. Accessed: 2020-12-09.

[7] Space Exploration Holdings, LLC, "SAT-MOD-20190830-00087." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2019083000087, 2019. Accessed: 2020-12-09.

[8] Space Exploration Holdings, LLC, "SAT-MOD-20200417-00037." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2020041700037, 2020. Accessed: 2020-12-09.

[9] SES S.A., "O3b mPower." https://o3bmpower.ses.com/, 2021. Accessed: 2021-08-27.

[10] O3b Limited, "SAT-MOD-20200526-00058." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2020052600058, 2019. Accessed: 2021-02-01.

[11] N. Abbas, Y. Nasser, and K. E. Ahmad, "Recent advances on artificial intelligence and learning techniques in cognitive radio networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, pp. 1–20, 2015.

[12] G. Maral and M. Bousquet, *Satellite communications systems: systems, techniques and technology*. John Wiley & Sons, 2011.

[13] European Space Agency, "Satellite frequency bands." https://www.esa.int/spaceinimages/Images/2013/11/Satellite_frequency_bands, 2019. Accessed: 2019-07-12.

[14] Digital Video Broadcasting (DVB), "Second generation framing structure, channel coding and modulation systems for broadcasting; part 2: DVB-SE extensions (DBVS2X)," Tech. Rep. 6, 2014.

[15] M. Guerster, J. J. G. Luis, E. F. Crawley, and B. G. Cameron, "Problem representation of dynamic resource allocation for flexible high throughput satellites," in *2019 IEEE Aerospace Conference*, 2019.

[16] A. I. Aravanis, B. S. MR, P.-D. Arapoglou, G. Danoy, P. G. Cottis, and B. Ottersten, "Power allocation in multibeam satellite systems: A two-stage multi-objective optimization," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3171–3182, 2015.

[17] C. N. Efrem and A. D. Panagopoulos, "Dynamic energy-efficient power allocation in multibeam satellite systems," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 228–231, 2019.

[18] F. R. Durand and T. Abrao, "Power allocation in multibeam satellites based on particle swarm optimization," *AEU-International Journal of Electronics and Communications*, vol. 78, pp. 124–133, 2017.

[19] S. Liu, Y. Fan, Y. Hu, D. Wang, L. Liu, and L. Gao, "Ag-dpa: Assignment game–based dynamic power allocation in multibeam satellite systems," *International Journal of Satellite Communications and Networking*, vol. 38, no. 1, pp. 74–83, 2020.

[20] P. Zhang, X. Wang, Z. Ma, S. Liu, and J. Song, "An online power allocation algorithm based on deep reinforcement learning in multibeam satellite systems," *International Journal of Satellite Communications and Networking*, vol. 38, no. 5, pp. 450–461, 2020.

[21] J. J. G. Luis, M. Guerster, I. del Portillo, E. F. Crawley, and B. G. Cameron, "Deep reinforcement learning architecture for continuous power allocation in

high throughput satellites," in *Reinforcement Learning for Real Life Workshop at 2019 International Conference on Machine Learning*, 2019.

[22] J. J. G. Luis, N. Pachler, M. Guerster, I. del Portillo, E. F. Crawley, and B. G. Cameron, "Artificial intelligence algorithms for power allocation in high throughput satellites: A comparison," in *2020 IEEE Aerospace Conference*, 2020.

[23] Y. Hong, A. Srinivasan, B. Cheng, L. Hartman, and P. Andreadis, "Optimal power allocation for multiple beam satellite systems," in *2008 IEEE Radio and Wireless Symposium*, pp. 823–826, IEEE, 2008.

[24] F. Qi, L. Guangxia, F. Shaodong, and G. Qian, "Optimum power allocation based on traffic demand for multi-beam satellite communication systems," in *2011 IEEE 13th International Conference on Communication Technology*, pp. 873–876, IEEE, 2011.

[25] H. Wang, A. Liu, X. Pan, and J. Yang, "Optimization of power allocation for multiusers in multi-spot-beam satellite communication systems," *Mathematical Problems in engineering*, vol. 2014, 2014.

[26] A. Destounis and A. D. Panagopoulos, "Dynamic power allocation for broadband multi-beam satellite communication networks," *IEEE Communications letters*, vol. 15, no. 4, pp. 380–382, 2011.

[27] N. K. Srivastava and A. Chaturvedi, "Flexible and dynamic power allocation in broadband multi-beam satellites," *IEEE communications letters*, vol. 17, no. 9, pp. 1722–1725, 2013.

[28] E. Lagunas, S. Maleki, S. Chatzinotas, M. Soltanalian, A. I. Pérez-Neira, and B. Oftersten, “Power and rate allocation in cognitive satellite uplink networks,” in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2016.

[29] T. T. Kapsis and A. D. Panagopoulos, “Optimum power allocation based on channel conditions in optical satellite downlinks,” *Wireless Personal Communications*, vol. 116, no. 4, pp. 2997–3013, 2021.

[30] J. M. Park, U. R. Savagaonkar, E. K. Chong, H. Siegel, and S. D. Jones, “Efficient resource allocation for qos channels in mf-tdma satellite systems,” in *MILCOM 2000 Proceedings. 21st Century Military Communications. Architectures and Technologies for Information Superiority (Cat. No. 00CH37155)*, vol. 2, pp. 645–649, IEEE, 2000.

[31] J.-M. Park, U. Savagaonkar, E. K. Chong, H. J. Siegel, and S. D. Jones, “Allocation of qos connections in mf-tdma satellite systems: a two-phase approach,” *IEEE Transactions on vehicular technology*, vol. 54, no. 1, pp. 177–190, 2005.

[32] Q. Dong, J. Zhang, and T. Zhang, “Optimal timeslot allocation algorithm in mf-tdma,” in *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–4, IEEE, 2008.

[33] S.-D. Feng, G.-X. Li, F. Wang, G.-X. Zhang, Y. Lin, and X. Jie, “A multiple frequency channels reserved timeslots assignment algorithm for real-time traffic in mf-tdma satellite system,” *Wireless personal communications*, vol. 73, no. 3, pp. 803–817, 2013.

[34] J.-M. R. Bejarano, C. M. Nieto, and F. J. R. Piñar, "Mf-tdma scheduling algorithm for multi-spot beam satellite systems based on co-channel interference evaluation," *IEEE Access*, vol. 7, pp. 4391–4399, 2018.

[35] K.-D. Lee, Y.-H. Cho, H.-J. Lee, and H. Jeong, "Optimal scheduling for timeslot assignment in mf-tdma broadband satellite communications," in *Proceedings IEEE 56th Vehicular Technology Conference*, vol. 3, pp. 1560–1564, IEEE, 2002.

[36] P. Angeletti, D. Fernandez Prim, and R. Rinaldo, "Beam hopping in multibeam broadband satellite systems: System performance and payload architecture analysis," in *24th AIAA International Communications Satellite Systems Conference*, p. 5376, 2006.

[37] J. Anzalchi, A. Couchman, P. Gabellini, G. Gallinaro, L. D'agristina, N. Alagha, and P. Angeletti, "Beam hopping in multi-beam broadband satellite systems: System simulation and performance comparison with non-hopped systems," in *2010 5th Advanced Satellite Multimedia Systems Conference and the 11th Signal Processing for Space Communications Workshop*, pp. 248–255, IEEE, 2010.

[38] H. Han, X. Zheng, Q. Huang, and Y. Lin, "Qos-equilibrium slot allocation for beam hopping in broadband satellite communication systems," *Wireless Networks*, vol. 21, no. 8, pp. 2617–2630, 2015.

[39] X. Hu, S. Liu, Y. Wang, L. Xu, Y. Zhang, C. Wang, and W. Wang, "Deep reinforcement learning-based beam hopping algorithm in multibeam satellite systems," *IET Communications*, vol. 13, no. 16, pp. 2485–2491, 2019.

[40] L. Wang, X. Hu, S. Ma, S. Xu, and W. Wang, "Dynamic beam hopping of multi-beam satellite based on genetic algorithm," in *2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*, pp. 1364–1370, IEEE, 2020.

[41] C. Zhang, X. Zhao, and G. Zhang, "Joint precoding schemes for flexible resource allocation in high throughput satellite systems based on beam hopping," *China Communications*, vol. 18, no. 9, pp. 48–61, 2021.

[42] L. Lei, E. Lagunas, Y. Yuan, M. G. Kibria, S. Chatzinotas, and B. Ottersten, "Beam illumination pattern design in satellite networks: Learning and optimization for efficient beam hopping," *IEEE Access*, vol. 8, pp. 136655–136667, 2020.

[43] U. Park, H. W. Kim, D. S. Oh, and B.-J. Ku, "A dynamic bandwidth allocation scheme for a multi-spot-beam satellite system," *Etri Journal*, vol. 34, no. 4, pp. 613–616, 2012.

[44] U. Park, H. W. Kim, D. S. Oh, and B. J. Ku, "Flexible bandwidth allocation scheme based on traffic demands and channel conditions for multi-beam satellite systems," in *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pp. 1–5, IEEE, 2012.

[45] H. Wang, A. Liu, X. Pan, and L. Jia, "Optimal bandwidth allocation for multi-spot-beam satellite communication systems," in *Proceedings 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)*, pp. 2794–2798, IEEE, 2013.

[46] F. Li, X. Liu, K.-Y. Lam, Z. Na, J. Hua, J. Wang, and L. Wang, "Spectrum allocation with asymmetric monopoly model for multibeam-based cognitive satellite networks," *IEEE Access*, vol. 6, pp. 9713–9722, 2018.

[47] J. Su, S. Yang, H. Xu, and X. Zhou, "A stackelberg differential game based bandwidth allocation in satellite communication network," *China Communications*, vol. 15, no. 8, pp. 205–214, 2018.

[48] J. Wang, B. Zhang, B. Zhao, G. Ding, and D. Guo, "A game-theoretical learning approach for spectrum trading in cognitive satellite-terrestrial networks," *IEEE Communications Letters*, 2021.

[49] I. Bisio and M. Marchese, "Power saving bandwidth allocation over geo satellite networks," *IEEE Communications Letters*, vol. 16, no. 5, pp. 596–599, 2012.

[50] Y. Liu, Q. Zhang, X. Xin, G. Cao, Y. Tao, and Y. Shen, "Dynamic bandwidth allocation for multi-qos guarantee based on bee colony optimization," in *2020 IEEE Computing, Communications and IoT Applications (ComComAp)*, pp. 01–05, IEEE, 2020.

[51] Y. Kawamoto, T. Kamei, M. Takahashi, N. Kato, A. Miura, and M. Toyoshima, "Flexible resource allocation with inter-beam interference in satellite communication systems with a digital channelizer," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 2934–2945, 2020.

[52] Y. Abe, H. Tsuji, A. Miura, and S. Adachi, "Frequency resource allocation for satellite communications system based on model predictive control and its application to frequency bandwidth allocation for aircraft," in *2018 IEEE Con-*

ference on Control Technology and Applications (CCTA), pp. 165–170, IEEE, 2018.

[53] T. Mizuike and Y. Ito, "Optimization of frequency assignment," *IEEE Transactions on Communications*, vol. 37, no. 10, pp. 1031–1041, 1989.

[54] N. Funabiki and S. Nishikawa, "A gradual neural-network approach for frequency assignment in satellite communication systems," *IEEE transactions on neural networks*, vol. 8, no. 6, pp. 1359–1370, 1997.

[55] S. Salcedo-Sanz, R. Santiago-Mozos, and C. Bousoño-Calzón, "A hybrid hopfield network-simulated annealing approach for frequency assignment in satellite communications systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 2, pp. 1108–1116, 2004.

[56] S. Salcedo-Sanz and C. Bousoño-Calzón, "A hybrid neural-genetic algorithm for the frequency assignment problem in satellite communications," *Applied Intelligence*, vol. 22, no. 3, pp. 207–217, 2005.

[57] L. Wang, W. Liu, and H. Shi, "Noisy chaotic neural networks with variable thresholds for the frequency assignment problem in satellite communications," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 2, pp. 209–217, 2008.

[58] J. Wang, Y. Cai, and J. Yin, "Multi-start stochastic competitive hopfield neural network for frequency assignment problem in satellite communications," *Expert systems with applications*, vol. 38, no. 1, pp. 131–145, 2011.

[59] A. A. Salman, I. Ahmad, M. G. Omran, and M. G. Mohammad, "Frequency assignment problem in satellite communications using differential evolution," *Computers & Operations Research*, vol. 37, no. 12, pp. 2152–2163, 2010.

[60] J. Wang and Y. Cai, "Multiobjective evolutionary algorithm for frequency assignment problem in satellite communications," *Soft Computing*, vol. 19, no. 5, pp. 1229–1253, 2015.

[61] L. Houssin, C. Artigues, and E. Corbel, "Frequency allocation problem in a sdma satellite communication system," *Computers & Industrial Engineering*, vol. 61, no. 2, pp. 346–351, 2011.

[62] X. Hu, S. Liu, R. Chen, W. Wang, and C. Wang, "A deep reinforcement learning-based framework for dynamic resource allocation in multibeam satellite systems," *IEEE Communications Letters*, vol. 22, no. 8, pp. 1612–1615, 2018.

[63] K. I. Aardal, S. P. Van Hoesel, A. M. Koster, C. Mannino, and A. Sassano, "Models and solution techniques for frequency assignment problems," *Annals of Operations Research*, vol. 153, no. 1, pp. 79–129, 2007.

[64] Y. Xu, Y. Zhang, H. Zhou, and M. Yang, "Staring beam forming method for leo satellite communication system," in *International Conference in Communications, Signal Processing, and Systems*, pp. 415–422, Springer, 2017.

[65] A. Ivanov, M. Stoliarenko, S. Kruglik, S. Novichkov, and A. Savinov, "Dynamic resource allocation in leo satellite," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pp. 930–935, IEEE, 2019.

[66] R. J. Fowler, M. S. Paterson, and S. L. Tanimoto, "Optimal packing and covering in the plane are np-complete," *Information processing letters*, vol. 12, no. 3, pp. 133–137, 1981.

[67] R. Yao, Y. Zhang, P. Jiang, L. Yao, and X. Zuo, "Load balanced user grouping scheme for multibeam multicast satellite communications," in *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 1042–1046, IEEE, 2018.

[68] N. Pachler, E. F. Crawley, and B. G. Cameron, "A genetic algorithm for beam placement in high-throughput satellite constellations," in *2021 IEEE Cognitive Communications for Aerospace Applications Workshop (CCAAW)*, pp. 1–6, IEEE, 2021.

[69] K. N. Sherman, "Phased array shaped multi-beam optimization for leo satellite communications using a genetic algorithm," in *Proceedings 2000 IEEE International Conference on Phased Array Systems and Technology (Cat. No. 00TH8510)*, pp. 501–504, IEEE, 2000.

[70] H. Zhao, Z. Xie, H. Wang, and J. Jin, "Beam shaping for satellite phased array antenna using dual coding genetic algorithm," in *2009 5th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–4, IEEE, 2009.

[71] D. K. Okello and M. Kaplan, "Adaptive beam allocation for multimedia ka-band satellite networks," in *2004 IEEE 59th Vehicular Technology Conference. VTC 2004-Spring (IEEE Cat. No. 04CH37514)*, vol. 5, pp. 2843–2847, IEEE, 2004.

[72] C. Qian, S. Zhang, and W. Zhou, "Traffic-based dynamic beam coverage adjustment in satellite mobile communication," in *2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, IEEE, 2014.

[73] B. Wenqian, W. Weidong, L. Shuaijun, and C. Gaofeng, "Beam coverage dynamic adjustment scheme based on maximizing system capacity for multi-beam satellite communication system," in *International Conference on Space Information Network*, pp. 288–298, Springer, 2017.

[74] T. Zhang, L. Zhang, and D. Shi, "Resource allocation in beam hopping communication system," in *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, pp. 1–5, IEEE, 2018.

[75] J.-T. Camino, C. Artigues, L. Houssin, and S. Mourgues, "Mixed-integer linear programming for multibeam satellite systems design: Application to the beam layout optimization," in *2016 Annual IEEE Systems Conference (SysCon)*, pp. 1–6, IEEE, 2016.

[76] W. Krewel and G. Maral, "Analysis of the impact of handover strategies on the qos of satellite diversity based communications systems," in *18th International Communications Satellite Systems Conference and Exhibit*, p. 1220, 2000.

[77] E. Papapetrou and F.-N. Pavlidou, "Qos handover management in leo/meo satellite systems," *Wireless Personal Communications*, vol. 24, no. 2, pp. 189–204, 2003.

[78] E. Papapetrou, S. Karapantazis, G. Dimitriadis, and F.-N. Pavlidou, "Satellite handover techniques for leo networks," *International Journal of Satellite Communications and Networking*, vol. 22, no. 2, pp. 231–245, 2004.

[79] K. Zhu, C. Hua, P. Gu, and W. Xu, "User clustering and proactive group handover scheduling in leo satellite networks," in *2020 IEEE Computing, Communications and IoT Applications (ComComAp)*, pp. 1–6, IEEE, 2020.

[80] Z. Wu, F. Jin, J. Luo, Y. Fu, J. Shan, and G. Hu, "A graph-based satellite handover framework for leo satellite communication networks," *IEEE Communications Letters*, vol. 20, no. 8, pp. 1547–1550, 2016.

[81] S. He, T. Wang, and S. Wang, "Load-aware satellite handover strategy based on multi-agent reinforcement learning," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2020.

[82] N. Pachler, E. F. Crawley, and B. G. Cameron, "Beam-to-satellite scheduling for high throughput satellite constellations using particle swarm optimization," in *IEEE Aerospace Conference*, 2022.

[83] J. C. Pemberton and F. Galiber, "A constraint-based approach to satellite scheduling," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 57, pp. 101–114, 2001.

[84] F. Xhafa, J. Sun, A. Barolli, A. Biberaj, and L. Barolli, "Genetic algorithms for satellite scheduling problems," *Mobile Information Systems*, vol. 8, no. 4, pp. 351–377, 2012.

[85] V. Kolici, X. Herrero, F. Xhafa, and L. Barolli, "Local search and genetic algorithms for satellite scheduling problems," in *2013 Eighth International Confer-*

ence on Broadband and Wireless Computing, Communication and Applications, pp. 328–335, IEEE, 2013.

[86] S. Zhuang, Z. Yin, Z. Wu, and Z. Shi, "The relay satellite scheduling based on artificial bee colony algorithm," in *2014 International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 635–640, IEEE, 2014.

[87] R. Tharmarasa, A. Chatterjee, Y. Wang, T. Kirubarajan, J. Berger, and M. C. Florea, "Closed-loop multi-satellite scheduling based on hierarchical mdp," in *2019 22th International Conference on Information Fusion (FUSION)*, pp. 1–7, IEEE, 2019.

[88] X. Chen, G. Reinelt, G. Dai, and A. Spitz, "A mixed integer linear programming model for multi-satellite scheduling," *European Journal of Operational Research*, vol. 275, no. 2, pp. 694–707, 2019.

[89] F. Xhafa and A. W. Ip, "Optimisation problems and resolution methods in satellite scheduling and space-craft operation: a survey," *Enterprise Information Systems*, vol. 15, no. 8, pp. 1022–1045, 2021.

[90] P. K. Chowdhury, M. Atiquzzaman, and W. Ivancic, "Handover schemes in satellite networks: State-of-the-art and future research directions," *IEEE Communications Surveys & Tutorials*, vol. 8, no. 4, pp. 2–14, 2006.

[91] M. Crosnier, R. Dhaou, F. Planchou, and A.-L. Beylot, "A cluster-based load balancing between satellite gateways in a manet," in *2012 6th Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC)*, pp. 303–307, IEEE, 2012.

[92] M. Werner, "A dynamic routing concept for atm-based satellite personal communication networks," *IEEE journal on selected areas in communications*, vol. 15, no. 8, pp. 1636–1648, 1997.

[93] E. Sigel, B. Denby, and S. Le Hégarat-Mascle, "Application of ant colony optimization to adaptive routing in a leo telecomunications satellite network," in *Annales des télécommunications*, vol. 57, pp. 520–539, Springer, 2002.

[94] H. Li and X. Gu, "Application of hopfield neural network routing algorithm in nongeostationary satellite communication networks," in *Proceedings. 2005 International Conference on Communications, Circuits and Systems, 2005.*, vol. 2, IEEE, 2005.

[95] Y. Rao and R. Wang, "Performance of qos routing using genetic algorithm for polar-orbit leo satellite networks," *AEU-International Journal of Electronics and Communications*, vol. 65, no. 6, pp. 530–538, 2011.

[96] A. Rajagopal, A. Ramachandran, K. Shankar, M. Khari, S. Jha, and G. P. Joshi, "Optimal routing strategy based on extreme learning machine with beetle antennae search algorithm for low earth orbit satellite communication networks," *International Journal of Satellite Communications and Networking*, vol. 39, no. 3, pp. 305–317, 2021.

[97] J. Sun and E. Modiano, "Routing strategies for maximizing throughput in leo satellite networks," *IEEE journal on selected areas in communications*, vol. 22, no. 2, pp. 273–286, 2004.

[98] Y. Rao and R.-c. Wang, "Agent-based load balancing routing for leo satellite networks," *Computer networks*, vol. 54, no. 17, pp. 3187–3195, 2010.

[99] X. Liu, Z. Jiang, C. Liu, S. He, C. Li, Y. Yang, and A. Men, "A low-complexity probabilistic routing algorithm for polar orbits satellite constellation networks," in *2015 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 1–5, IEEE, 2015.

[100] N. Zhao, X. Long, and J. Wang, "A multi-constraint optimal routing algorithm in leo satellite networks," *Wireless Networks*, pp. 1–12, 2021.

[101] F. Wang, D. Jiang, and S. Qi, "An adaptive routing algorithm for integrated information networks," *China Communications*, vol. 16, no. 7, pp. 195–206, 2019.

[102] J. A. Fraire, P. Madoery, S. Burleigh, M. Feldmann, J. Finochietto, A. Charif, N. Zergainoh, and R. Velazco, "Assessing contact graph routing performance and reliability in distributed satellite constellations," *Journal of Computer Networks and Communications*, vol. 2017, 2017.

[103] F. Alagoz, O. Korcak, and A. Jamalipour, "Exploring the routing strategies in next-generation satellite networks," *IEEE Wireless Communications*, vol. 14, no. 3, pp. 79–88, 2007.

[104] X. Wang, Y. Li, S. Zhao, Y. Zheng, Z. Zhu, and G. Cao, "A tradeoff resource allocation based on mf-tdma scheme in the multibeam data relay satellite systems," *International Journal of Satellite Communications and Networking*, vol. 37, no. 3, pp. 200–212, 2019.

[105] X. Alberti, J. Cebrian, A. Del Bianco, Z. Katona, J. Lei, M. Vazquez-Castro, A. Zanus, L. Gilbert, and N. Alagha, "System capacity optimization in time

and frequency for multibeam multi-media satellite systems," in *2010 5th Advanced Satellite Multimedia Systems Conference and the 11th Signal Processing for Space Communications Workshop*, pp. 226–233, IEEE, 2010.

[106] J. Lei and M. A. Vazquez-Castro, "Multibeam satellite frequency/time duality study and capacity optimization," *Journal of Communications and Networks*, vol. 13, no. 5, pp. 472–480, 2011.

[107] S. Shi, G. Li, Z. Li, H. Zhu, and B. Gao, "Joint power and bandwidth allocation for beam-hopping user downlinks in smart gateway multibeam satellite systems," *International Journal of Distributed Sensor Networks*, vol. 13, no. 5, p. 1550147717709461, 2017.

[108] L. Wang, C. Zhang, D. Qu, and G. Zhang, "Resource allocation for beam-hopping user downlinks in multi-beam satellite system," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pp. 925–929, IEEE, 2019.

[109] A. Wang, L. Lei, E. Lagunas, S. Chatzinotas, A. I. P. Neira, and B. Ottersten, "Joint beam-hopping scheduling and power allocation in noma-assisted satellite systems," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, 2021.

[110] G. Cocco, T. De Cola, M. Angelone, Z. Katona, and S. Erl, "Radio resource management optimization of flexible satellite payloads for dvb-s2 systems," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 266–280, 2017.

[111] X. Zhong, H. Yin, Y. He, and H. Zhu, "Joint transmit power and bandwidth allocation for cognitive satellite network based on bargaining game theory," *IEEE Access*, vol. 7, pp. 6435–6449, 2018.

[112] A. Paris, I. Del Portillo, B. Cameron, and E. Crawley, "A genetic algorithm for joint power and bandwidth allocation in multibeam satellite systems," in *2019 IEEE Aerospace Conference*, pp. 1–15, IEEE, 2019.

[113] M. Jia, X. Zhang, X. Gu, Q. Guo, Y. Li, and P. Lin, "Interbeam interference constrained resource allocation for shared spectrum multibeam satellite communication systems," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6052–6059, 2018.

[114] X. Liao, X. Hu, Z. Liu, S. Ma, L. Xu, X. Li, W. Wang, and F. M. Ghannouchi, "Distributed intelligence: A verification for multi-agent drl-based multibeam satellite resource allocation," *IEEE Communications Letters*, vol. 24, no. 12, pp. 2785–2789, 2020.

[115] A. Jahn, "Resource management model and performance evaluation for satellite communications," *International journal of satellite communications*, vol. 19, no. 2, pp. 169–203, 2001.

[116] J. Lei and M. A. Vazquez-Castro, "Joint power and carrier allocation for the multibeam satellite downlink with individual sinr constraints," in *2010 IEEE International Conference on Communications*, pp. 1–5, IEEE, 2010.

[117] T. S. Abdu, S. Kisseleff, E. Lagunas, and S. Chatzinotas, "Flexible resource optimization for geo multibeam satellite communication system," *IEEE Transactions on Wireless Communications*, pp. 1–15, 2021.

[118] F. Vidal, H. Legay, and G. Goussetis, "Joint power, frequency and precoding optimisation in a satellite sdma communication system," in *2020 10th Advanced Satellite Multimedia Systems Conference and the 16th Signal Processing for Space Communications Workshop (ASMS/SPSC)*, pp. 1–8, IEEE, 2020.

[119] J. P. Choi and V. W. Chan, "Optimum multibeam satellite downlink power allocation based on traffic demands," in *Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE*, vol. 3, pp. 2875–2881, IEEE, 2002.

[120] J. P. Choi and V. W. Chan, "Optimum power and beam allocation based on traffic demands and channel conditions over satellite downlinks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2983–2993, 2005.

[121] J. P. Choi and V. W. Chan, "Resource management for advanced transmission antenna satellites," *IEEE transactions on wireless communications*, vol. 8, no. 3, pp. 1308–1321, 2009.

[122] M. Takahashi, Y. Kawamoto, N. Kato, A. Miura, and M. Toyoshima, "Adaptive power resource allocation with multi-beam directivity control in high-throughput satellite communication systems," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1248–1251, 2019.

[123] M. Takahashi, Y. Kawamoto, N. Kato, A. Miura, and M. Toyoshima, "Adaptive multi-beam arrangement for improving throughput in an hts communication system," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2020.

[124] M. Takahashi, Y. Kawamoto, N. Kato, A. Miura, and M. Toyoshima, "Dbf-based fusion control of transmit power and beam directivity for flexible resource

allocation in hts communication system toward b5g," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 95–105, 2021.

[125] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual sinr constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18–28, 2004.

[126] Y. Liu, L. Feng, L. Wu, Z. Zhang, J. Dang, B. Zhu, and L. Wang, "Joint optimization based satellite handover strategy for low earth orbit satellite networks," *IET Communications*, 2021.

[127] M. Y. Abdelsadek, H. Yanikomeroglu, and G. K. Kurt, "Future ultra-dense leo satellite networks: A cell-free massive mimo approach," *arXiv preprint arXiv:2106.09837*, 2021.

[128] A. Kyrgiazos, B. Evans, and P. Thompson, "Irregular beam sizes and non-uniform bandwidth allocation in hts multibeam satellite systems," in *31st AIAA International Communications Satellite Systems Conference (ICSSC)*, 2013.

[129] K. Kiatmanaroj, C. Artigues, L. Houssin, and F. Messine, "Hybrid discrete-continuous optimization for the frequency assignment problem in satellite communication system," *IFAC Proceedings Volumes*, vol. 45, no. 6, pp. 1419–1424, 2012.

[130] K. Kiatmanaroj, C. Artigues, L. Houssin, and F. Messine, "Frequency allocation in a sdma satellite communication system with beam moving," in *2012 IEEE International Conference on Communications (ICC)*, pp. 3265–3269, IEEE, 2012.

[131] K. Kiatmanaroj, C. Artigues, L. Houssin, and F. Messine, "Frequency assignment in a sdma satellite communication system with beam decentring feature," *Computational Optimization and Applications*, vol. 56, no. 2, pp. 439–455, 2013.

[132] N. Pachler de la Osa, M. Guerster, I. del Portillo Barrios, E. Crawley, and B. Cameron, "Static beam placement and frequency plan algorithms for leo constellations," *International Journal of Satellite Communications and Networking*, vol. 39, no. 1, pp. 65–77, 2021.

[133] J.-T. Camino, S. Mourgues, C. Artigues, and L. Houssin, "A greedy approach combined with graph coloring for non-uniform beam layouts under antenna constraints in multibeam satellite systems," in *2014 7th Advanced Satellite Multimedia Systems Conference and the 13th Signal Processing for Space Communications Workshop (ASMS/SPSC)*, pp. 374–381, IEEE, 2014.

[134] K. Y. Zhong, Y. J. Cheng, H. N. Yang, and B. Zheng, "Leo satellite multi-beam coverage area division and beamforming method," *IEEE Antennas and Wireless Propagation Letters*, vol. 20, no. 11, pp. 2115–2119, 2021.

[135] P.-J. Wan, V. Nguyen, and H. Bai, "Advance handovers arrangement and channel allocation in leo satellite systems," in *Seamless Interconnection for Universal Services. Global Telecommunications Conference. GLOBECOM'99.(Cat. No. 99CH37042)*, vol. 1, pp. 286–290, IEEE, 1999.

[136] R. Alinque Dianez, "Joint optimization of beam placement and shaping for multi-beam high throughput satellite systems using gradient descent," B.S. thesis, Universitat Politècnica de Catalunya, 2020.

[137] B. Liu, C. Jiang, L. Kuang, and J. Lu, "Joint user grouping and beamwidth optimization for satellite multicast with phased array antennas," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2020.

[138] J. Tang, D. Bian, G. Li, J. Hu, and J. Cheng, "Optimization method of dynamic beam position for leo beam-hopping satellite communication systems," *IEEE Access*, vol. 9, pp. 57578–57588, 2021.

[139] P. J. Honnaiah, N. Maturo, S. Chatzinotas, S. Kisseleff, and J. Krause, "Demand-based adaptive multi-beam pattern and footprint planning for high throughput geo satellite systems," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1526–1540, 2021.

[140] J.-T. Camino, C. Artigues, L. Houssin, and S. Mourgues, "Milp formulation improvement with k-means clustering for the beam layout optimization in multibeam satellite systems," *Computers & Industrial Engineering*, vol. 158, p. 107228, 2021.

[141] F. Tian, L. Huang, G. Liang, X. Jiang, S. Sun, and J. Ma, "An efficient resource allocation mechanism for beam-hopping based leo satellite communication system," in *2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–5, IEEE, 2019.

[142] P. Zuo, T. Peng, W. Linghu, and W. Wang, "Resource allocation for cognitive satellite communications downlink," *IEEE Access*, vol. 6, pp. 75192–75205, 2018.

[143] J. Tang, D. Bian, G. Li, J. Hu, and J. Cheng, "Resource allocation for leo beam-hopping satellites in a spectrum sharing scenario," *IEEE Access*, vol. 9, pp. 56468–56478, 2021.

[144] B. Deng, C. Jiang, L. Kuang, N. Ge, S. Guo, and S. Zhao, "Resource allocation of multibeam communication satellite systems in sparse networks," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2019.

[145] P. Angeletti and R. De Gaudenzi, "A pragmatic approach to massive mimo for broadband communication satellites," *IEEE Access*, vol. 8, pp. 132212–132236, 2020.

[146] E. Lagunas, S. K. Sharma, S. Maleki, S. Chatzinotas, and B. Ottersten, "Resource allocation for cognitive satellite communications with incumbent terrestrial networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, no. 3, pp. 305–317, 2015.

[147] J. J. Garau-Luis, S. Aliaga, G. Casadesus, N. Pachler, E. Crawley, and B. Cameron, "Frequency plan design for multibeam satellite constellations using linear programming," 2022.

[148] Y. Cao, Y. Shi, J. Liu, and N. Kato, "Optimal satellite gateway placement in space-ground integrated network for latency minimization with reliability guarantee," *IEEE Wireless Communications Letters*, vol. 7, no. 2, pp. 174–177, 2017.

[149] Q. Chen, L. Yang, X. Liu, J. Guo, S. Wu, and X. Chen, "Multiple gateway placement in large-scale constellation networks with inter-satellite links," *In-*

ternational Journal of Satellite Communications and Networking, vol. 39, no. 1, pp. 47–64, 2021.

[150] M. Mitchell, *An introduction to genetic algorithms*. The MIT Press, 1996.

[151] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[152] Kuiper Systems LLC, "SAT-LOA-20190704-00057." `http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATLOA2019070400057`, 2019. Accessed: 2020-12-09.

[153] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *MHS'95. Proceedings of the sixth international symposium on micro machine and human science*, pp. 39–43, Ieee, 1995.

[154] I. del Portillo, B. G. Cameron, and E. F. Crawley, "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband," *Acta Astronautica*, 2019.

[155] N. Pachler, I. del Portillo, E. F. Crawley, and B. G. Cameron, "An updated comparison of four low earth orbit satellite constellation systems to provide global broadband," in *IEEE International Workshop*, 2021.

[156] Center for International Earth Science Information Network, "Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11." `https://doi.org/10.7927/H4JW8BX5`, 2018.

[157] I. del Portillo, B. Cameron, and E. Crawley, "Ground segment architectures for large leo constellations with feeder links in ehf-bands," in *2018 IEEE Aerospace Conference*, pp. 1–14, IEEE, 2018.

[158] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical association*, vol. 66, no. 336, pp. 846–850, 1971.