Optimizing resource allocation in large communications 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

Doctor of Philosophy in Aeronautics and Astronautics
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

April 2024

Author .....
Department of Aeronautics and Astronautics
April 16, 2024

Thesis Committee Chair .....
Edward F. Crawley
Ford Professor of Engineering

Thesis Committee Member .....
Vincent W.S. Chan
Joan and Irwin M. Jacobs Professor

Thesis Committee Member .....
Joel Grotz
Senior Manager, Technology Development, Systems Engineering, SES

External Evaluator .....
Bruce G. Cameron
Director, System Architecture Group

External Evaluator .....
Iñigo del Portillo Barrios
Senior Software Engineer, Google

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

Optimizing resource allocation in large communications satellite constellations

by

Nils Pachler de la Osa

Submitted to the Department of Aeronautics and Astronautics
on April 16, 2024
PhD Thesis

Abstract

Satellite communications are becoming a key technology for maintaining connectivity in a world driven by information. In the recent years, established players (such as SES and Telesat), as well as new competitors (such as SpaceX and Amazon) have proposed constellations able to serve hundreds of thousands of users, using thousands of satellites. While the orbital configuration of each design is different, the next generation of satellite communications relies on highly flexible digital payloads, such as phased array antennas, on-board processing, and adaptive modulation and coding schemes. Several approaches have been proposed to deal with the complexity of the added flexibilities at the spacecraft level. Nevertheless, how to address the flexibilities at the constellation level, critical to operate the next generation of systems, remains an open question.

This dissertation develops optimization-based decision-making frameworks for designing and operating the next generation of communication constellations. In particular, novel methods for the Beam Shaping, User Grouping, Satellite Routing, Frequency Assignment, and Gateway Routing problems are proposed, tailored for large non-geostationary orbit constellations with satellites at multiple altitudes, referred to as hybrid systems. The methods leverage optimization to find an optimized set of decisions that maximize capacity and quality of service and minimize necessary ground infrastructure, all while avoiding interference. The proposed methods are then combined, tested, and evaluated using existing constellation designs under representative operational conditions with hundreds of thousands of users.

The reported results prove that the proposed techniques are able to multiply by two the capacity of these systems, with favorable trade-offs in quality of service and necessary ground infrastructure. By testing existing designs, it is concluded that the number of satellites, as well as the link quality are the main drivers of performance. Furthermore, the analysis shows that hybrid constellations offer advantages over other designs, thanks to the combination of high quality links on low altitude satellites, and high coverage on high altitude satellites. Additionally, this dissertation studies the optimal proportion of satellites across various altitudes in hybrid LEO-MEO constellations. Results show that hybrid constellations are desirable when the cost of MEO and LEO satellites are comparable and interference is minimal.

Thesis Committee Chair: Edward F. Crawley
Title: Ford Professor of Engineering

Acknowledgments

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 five 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 part of my time at MIT, 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, and Valvanera Moreno. I also want to mention Íñigo del Portillo and Markus Guerster, who have shared the wisdom and advice that has helped me succeed. I would like to acknowledge the rest of my labmates who have also been a part of this project at some point during these five years. Skylar, James, Guillem, Sergi, Joel, Damon, and Rubén, you have been a source of inspiration and stimulating discussions, best of luck on your next steps.

I would also like to thanks the rest of my labmates who have been a key to succeed in this last part of my graduate studies. 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, Tamires Meireles, Fran Marrone, Shokofeh Khadivi, Erinn Taylor Barroso, 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

List of Figures

List of Tables

Acronyms

ACM Adaptive Coding and Modulation

C3IM Carrier to Third Order Inter-Modulation Products of Interference

CABI Carrier to Adjacent Beam Interference

CASC China Aerospace Science and Technology Corporation

CASI Carrier to Adjacent Satellites Interference

CXPI Carrier Cross Polarization Interference

DRA Dynamic Resource Allocation

DRL Deep Reinforcement Learning

DRM Dynamic Resource Management

EIRP Effective Isotropic Radiation Power

FCC Federal Communications Commission

FDMA Frequency-Division Multiple-Access

FoV Field of View

FSLP Free Space Loss Propagation

GA Genetic Algorithm

GSO Geostationary Orbit

HEO High Earth Orbit

ILP Integer Linear Programming

ISL Inter-Satellite Links

ITU International Telecommunication Union

LEO Low Earth Orbit

LoS Line of Sight

MEO Medium Earth Orbit

MF-TDMA Multi-Frequency Time-Division Multiple-Access

MILP Mixed-integer linear programming

MODCOD Modulation and coding scheme

NGSO Non-Geostationary Orbit

NN Neural Network

NP Nondeterministic Polynomial time

OBO Output back-off

PSO Particle Swarm Optimization

QoS Quality of Service

RAP Resource Allocation Problem

RF Radio Frequency

SA Simulated Annealing

TDMA Time-Division Multiple-Access

UD Unmet Demand

Chapter 1

Introduction

1.1 Motivation

Satellite communications have emerged as a vital technology for sustaining connectivity in our information-driven world. With the increasing demand for ubiquitous data access anytime [1], traditional static satellite communication systems are evolving into dynamic software-defined constellations, able to respond to the needs of the rapidly fluctuating user demand. Furthermore, their reduced dependence on physical infrastructure allows satellite operators to offer effective solutions for markets where terrestrial networks are impractical or inefficient, such as isolated regions or *connectivity on-the-go* [2]. These novel conditions have spurred interest from new entrants (such as SpaceX and Amazon), as well as established players (such as SES and Viasat). Together, they are driving the development of the next generation of satellite communications, which is expected to materialize in the upcoming years [3–6].

The introduction of highly adaptable satellite payloads, combined with a substantial increase in the number of satellites at lower altitudes, is expected to enhance the capacity of satellite operators, potentially reaching tens of terabits per second (Tbps), [7, 8]. This elevated capacity will be accompanied by latency levels comparable to ground-based networks [9]. Achieving this level of performance, however, hinges on the deployment of automated and scalable frameworks capable of managing the heightened operational complexity arising from software-defined components and a larger user base. In line with this, an incipient area of research focuses on automated and optimized

decision-making frameworks [10–12]. State-of-the-art approaches demonstrate that leveraging modern optimization techniques can increase resource efficiency by a notable factor of 12 compared to heuristic methods, as noted in reference [12]. However, current autonomous frameworks do not encompass all flexibilities inherent in modern constellations [13].

In a parallel line of research, extensive analyses of megaconstellations provide valuable insights on the technical and economic aspects of new proposals [7, 8, 14]. This information is instrumental for stakeholders and regulatory bodies in understanding the impact of a specific constellation design within the space environment. Nonetheless, due to the one-sided nature of these studies, there are limited insights that can effectively guide the development of future proposals. Consequently, the focus of this study centers on the development of autonomous decision-making frameworks that can be used to inform the future design of megaconstellations.

1.2 Background

The objective of this section is to provide brief technical background on satellite communications. For a more in-depth exploration of the models and mathematical expressions employed in this study, refer to [15].

1.2.1 The satellite communications landscape

Throughout much of its history, the satellite communication industry has focused on TV broadcasting and voice-over-phone communications [16]. Seeking expansion into new markets, initiatives such as Iridium and Globalstar in the 1990s experimented with the concept of providing broadband services through space connectivity [17]. However, despite significant investments, these original projects faced early cancellations or bankruptcy filings, mainly attributed to technological limitations and overestimations of market potential [18]. Following a 15-year lull with no new proposals, the past decade has witnessed a resurgence in applications aiming to provide Internet services from space [19].

Fueled by advancements in payload technology and a reduction in manufacturing and launch costs, nine companies have submitted applications to the International Telecommunication Union (ITU) or the Federal Communications Commission (FCC) for communication constellations exceeding 100 satellites. These companies are OneWeb, SpaceX, Boeing, Amazon, Telesat, Viasat,

SES, China Aerospace Science and Technology Corporation (CASC)[1], and Intelsat. Refer to Table 1.1 for an overview of the orbit characteristics of the first generation of each project[2]. Should all these projects come to fruition, the new space environment is expected to host more than 100,000 satellites [20], with more than 90% dedicated to communication purposes. It is noteworthy that the first generation of projects is already projected to surpass the 30,000-satellite milestone. Furthermore, certain systems have achieved notable milestones, such as SpaceX Starlink with more than a million subscribers [21]. This figure is expected to grow in the coming months. The success of these new systems hinges on two primary factors: a technology push facilitated by innovative, flexible components, and a societal pull driven by the increasing necessity for ubiquitous connectivity.

Technology push

The technological push propelling the next generation of satellite communications can be categorized into three key areas: advancements in satellite payload technology, the development of cost-efficient launch options and satellite miniaturization, and the enhancement of ground infrastructure.

(a) *Advancements in satellite payload*: Satellite payload advaancements are driven by three primary components. First, phased array antennas, capable of directing the satellite signal to specific user regions without the need for mechanical movement. This allows for a faster, more precise, and less mechanically taxing control of the signal, compared to previous mechanically-steered parabolic antennas. Second, on-board processing and routing capabilities, enabling more flexible data management. This allows for a higher degree of control on the flow of information through the network, compared to existing bent-pipe architectures. And third, Adaptive Coding and Modulation (ACM) schemes, optimizing spacecraft power utilization by only utilizing the necessary power to close the link, as opposed to a fixed setting. These technological improvements significantly enhance spacecraft communication capabilities by maximizing the efficient use of limited satellite resources.

(b) *Cost-efficient launch options and satellite miniaturization*: The reduction in launch costs, facilitated by reusable launchers, results in a lower cost per mass launched into space (as low as 1,520 $ / kg using a reusable Falcon Heavy rocket for Low Earth Orbit (LEO) launches [39]). Additionally, satellites are now considerably smaller and lighter than their predecessors, exemplified by the con-

[1]Note that, while initially three independent Chinese proposals were filed, they ultimately merged into a single mega-constellation.

[2]Only the characteristics for the first generation of these systems are presented

System	Altitude (km)	Inclination (°)	Planes	Satellites per plane	State	Number of satellites	ISL	References
OneWeb	1,200	87.9	36	49	P	6,372	No	[22–24]
	1,200	55	32	72	P			
	1,200	40	32	72	P			
SpaceX	540	53.2	72	22	A	4,408	Yes	[3, 25–28]
	550	53	72	22	A			
	560	97.6	6	58	A			
	560	97.6	4	43	A			
	570	70	36	20	A			
Boeing	670	82.9	20	30	P	5,936	Yes	[29, 30]
	680	54.9	40	35	P			
	690	37.9	46	34	P			
	1,040	37.2	28	30	P			
	1,056	54	11	12	A			
	1,070	48.8	35	28	P			
	1,085	79.6	11	26	P			
	9,000	0	1	39	P			
	10,000	41.2	10	8	P			
	35,786	63.4	5	3	A			
Amazon	590	33	28	28	A	3,236	Yes	[4]
	610	42	36	36	A			
	630	51.9	34	34	A			
Telesat	1,015	98.98	27	13	P	1,671	Yes	[31, 32]
	1,325	50.88	40	33	P			
Viasat	1,300	45	8	36	P	288	Yes	[6, 33]
SES-O3b	507	97.4	4	9	P	112	Yes	[5, 34, 35]
	8,062	70	2	5	P			
	8,062	90	4	6	P			
	8,062	0	1	42	P			
CASC	508	55	60	60	*	12,992	Not specified	[36, 37]
	590	85	16	30	*			
	600	50	40	50	*			
	1,145	30	48	36	*			
	1,145	40	48	36	*			
	1,145	50	48	36	*			
	1,145	60	48	36	*			
Intelsat	8,600	47	6	12	P	216	No	[38]
	8,600	62.9	6	12	P			
	8,600	89	6	12	P			

Table 1.1: Summary of the orbit characteristics of the proposed constellations. *A* represents shells approved by the FCC, *P* represents pending changes and * represents proposals submitted only to the ITU. Table adapted from [8] with permission of the authors.

trast between Starlink (260 kg) [40] and SES-17 (6,411 kg) [41]. This dual reduction in launch costs and satellite size implies that operators can deploy between 100 and 1,000 times more satellites for the same cost compared to a decade ago.

(c) *Enhancement of Ground Infrastructure*: The realization of a more extensive network of highly capable satellites is contingent on the presence of robust ground infrastructure. Each new appli-

cation depends on a substantial ground network comprising over 1,000 antennas [7, 8] to provide Internet access to users within the constellation.

Societal pull

Meeting projections [1], Internet traffic has surged at a rate of 29% over the past four years [42], nearly reaching the 1,000 Tbps milestone in 2022. Beyond the demands imposed by data-intensive services, such as video streaming, this growth is propelled by factors like the expanding requirements in historically underserved regions [43] and the imperative for seamless connectivity while *on-the-go* [44] (e.g., in aviation or maritime environments). Expected growth persists in these markets in the coming years. Notably, in the Asia-Pacific region, video streaming services are forecasted to expand at a rate of 22.4% annually [43], while in-flight connectivity services are projected to reach 1,000 Gbps by 2031 [45]. Given the inefficiency of terrestrial networks in catering to these specialized markets, satellite communications stand out as a viable alternative to meet the increasing demands for Internet services.

Beyond sheer traffic growth, latency emerges as a critical consideration with the increasing prominence of live video and voice streaming [43]. Although conventional satellite networks relied on Geostationary Orbit (GSO) satellites for data transmission, with round-trip-latencies up to 0.5 seconds, contemporary demands necessitate much swifter connections achievable only with proximate satellites. Addressing this challenge, new proposals hinge on Non-Geostationary Orbit (NGSO) in low or medium Earth orbits (LEO/Medium Earth Orbit (MEO)), delivering latencies ranging from 40 ms to 180 ms [9].

Current state

In summary, the novel constellations rely on thousands of highly adaptable satellites in NGSO, designed to serve hundreds of thousands, and potentially millions, of users globally, including mobile users such as those on planes or ships. Supporting this intricate network are an array of thousands of ground antennas. Given the transformative shift entailed by this paradigm, the traditional manual operation, once the industry gold standard, is no longer feasible. To navigate this new reality, operators rely on automated and scalable tools capable of harnessing the payload flexibilities under high-dimensional conditions to maximize the constellations performance.

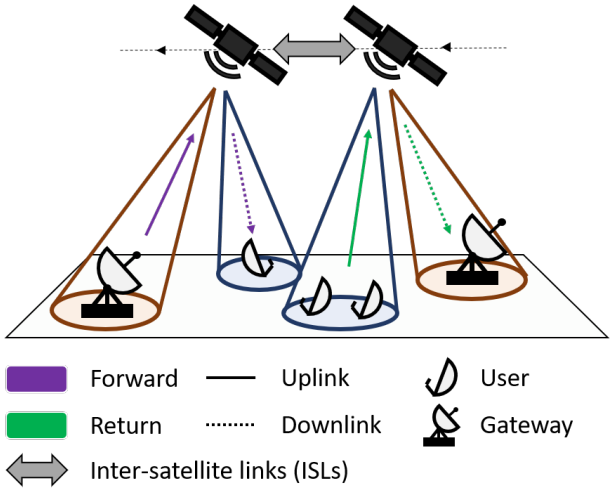


Figure 1-1: Definition of satellite communications links

1.2.2 The satellite communications environment

While current satellite constellations are significantly more complex than their predecessors, they still rely on the same underlying principles. Communication, in any form, arises from the fundamental need to transmit data between two points. Specifically, in the realm of broadband satellite communications, both the source and destination are situated within the Earth atmosphere, utilizing a network of satellites as relays for data transmission. In the context of providing Internet access, either the source or destination must be a *gateway*, a ground antenna facilitating access to the terrestrial Internet network. When multiple gateways are positioned in the same location, they collectively form a *ground station*. The counterpart in the communication link is referred to as the *user*, defined as the entity requesting communication services and remunerating for the service provided. The connection from gateways to users is commonly termed the *forward link*, while the link from users to gateways is designated as the *return link*. The transmission from Earth to space is denoted as the *uplink*, while the reverse, from space to Earth, is identified as the *downlink*. Figure 1-1 illustrates these definitions. Furthermore, satellite connections are denoted as ISLs.

Electromagnetic waves

Satellite communications are fundamentally built on the transmission of electromagnetic waves —disturbances in electric and magnetic fields— via antennas. Antennas, entities capable of con-

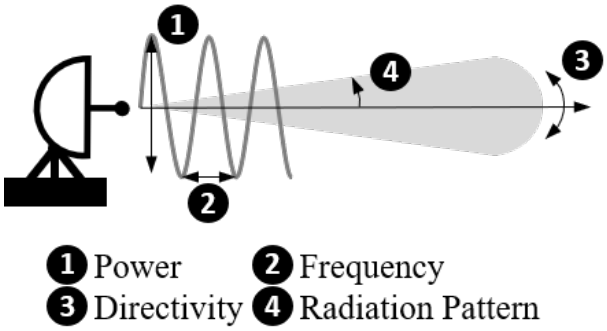


Figure 1-2: Characteristics of an electromagnetic wave

verting electric impulses into electromagnetic disturbances, play a pivotal role in this process. Electromagnetic waves are characterized by four primary elements:

1. *Power (or transmission power)*: This factor governs the signal strength, akin to how far the signal can propagate before merging with ambient noise. It is directly proportional to the electric power supplied to the source antenna.

2. *Frequency, bandwidth, and polarization*: These aspects determine how waves interact with surrounding signals and are dictated by the sequence of electric filters at the source antenna. In contemporary communications, information is encoded in the amplitude and phase of the wave using Modulation and coding scheme (MODCOD)s. Waves with overlapping frequencies and the same polarization may cause detrimental interference, resulting in information loss.

3. *Directivity*: This parameter dictates the spatial direction in which the signal is stronger, making it easier to decode. It is determined by the directivity of the source antenna. The ratio of signal strength in the main direction compared to signal strength when transmitting in all directions is termed the *gain* of the antenna.

4. *Radiation Pattern*: This characteristic defines the profile of the signal, representing its strength in directions other than the main direction. It is determined by the attributes of the source antenna.

The depiction of these four characteristics is provided in Figure 1-2. In the context of satellite communications, each electromagnetic wave is commonly referred to as a *beam*. In most contempo-

rary systems, all four factors are software-defined for each beam, affording a high degree of control over the communication parameters.

Link budget equation

To ensure successful satellite communications, the signal transmitted by the source must be accurately identified and decoded at the destination. This necessitates the signal being significantly larger compared to surrounding noise, a criterion encapsulated by the link budget equation:

$$\begin{aligned} \frac{C}{N} &= P_T + G_T + G_R - L - N && \text{[dB]} \\ L &= L_{FSPL} + L_{atm} + L_{pT} + L_{pR} + L_T + L_R + OBO && \text{[dB]} \\ N &= 10\log_{10}(kT_{sys}) + 10\log_{10}(BW) && \text{[dB]} \end{aligned} \tag{1.1}$$

Where $\frac{C}{N}$ corresponds to the carrier-to-noise ratio, P_T stands for the transmission power, G_T, G_R represent the antenna gains in the main direction in the transmitter and receiver antenna, respectively, L represents the losses of the system, N represents the noise of the system, and BW refers to the communication bandwidth. System loss (L) is a composite of: 1) Free Space Loss Propagation (FSLP) L_{FSPL}, contingent on the distance between the source and destination, 2) atmospheric losses L_{atm}, dependent on the angle at which the atmosphere is traversed, as well as the wave frequency, 3) pointing losses L_{pT}, L_{pR}, occurring during transmission and reception, respectively, and influenced by the direction of the wave and the directionality of the antennas, 4) antenna losses L_T, L_R, occurring during transmission and reception, respectively, and dependent on the internal characteristics of the antennas, and 5) power-amplifier Output back-off (OBO), dependent on the chosen MODCOD and other factors. The system noise is determined by the noise bandwidth BW, the Boltzmann constant k, and the system temperature T_{sys}, which can be computed according to the Friis equation when accounting for antenna (T_{ant}), atmosphere (T_{atm}), and waveguide (T_w) temperatures:

$$T_{sys} = T_{ant}10^{-\frac{L_R}{10}} + T_{atm}(1 - 10^{-\frac{A_t}{10}}) + T_w(1 - 10^{-\frac{L_R}{10}}) + 290(10^{\frac{NF}{10}} - 1) \text{ [K]} \tag{1.2}$$

Where A_t is the total atmospheric loss, and NF is the noise figure at the input receiver.

Interference

In addition to noise, the correct decoding of the signal may be influenced by interference. This study takes into account four types of interference:

- Carrier to Adjacent Beam Interference (CABI): Interference arising from two nearby beams.
- Carrier to Adjacent Satellites Interference (CASI): Interference resulting from two nearby satellites.
- Carrier Cross Polarization Interference (CXPI): Interference caused by imperfections in the polarization process.
- Carrier to Third Order Inter-Modulation Products of Interference (C3IM): Interference caused by imperfections in the signal filtering.

The final energy to noise plus interference spectral density $\frac{E_b}{N+I}$ ratio can be computed as:

$$\begin{aligned} \frac{C}{N+I} &= \left(\frac{1}{CABI} + \frac{1}{CASI} + \frac{1}{CXPI} + \frac{1}{C3IM} + \frac{1}{C/N}\right)^{-1} \\ \frac{E_b}{N+I} &= \frac{C}{N+I}\frac{BW}{R_b} \end{aligned} \tag{1.3}$$

Where R_b is the data-rate of the beam.

Data-rate

This study adopts ACM strategies, enabling the dynamic adjustment of the MODCOD and power of the beam based on the attenuation and interference values. To compute the MODCOD, the aim is to resolve the following optimization problem:

$$\begin{aligned} &\min_{MODCOD} OBO \\ &s.t.\ \left.\frac{E_b}{N_0}\right|_{th} \geq \frac{E_b}{N+I} + \gamma \end{aligned} \tag{1.4}$$

Where $\left.\frac{E_b}{N_0}\right|_{th}$ represents the energy per bit to noise spectral density of the MODCOD, and γ denotes the desired margin of the link, determined by the satellite operator. Note that the MODCOD chosen allows operators to consume only the necessary resources, with a slight margin to respond to traffic

needs and external attenuations. The final data rate can then be computed using:

$$R_b = \frac{BW}{1+\alpha_r}\Gamma\left(\frac{E_b}{N+I}\right) \text{[bps]} \tag{1.5}$$

Where $\Gamma\left(\frac{E_b}{N+I}\right)$ is the spectral efficiency (in bps/Hz) of the associated MODCOD, and α_r is the roll-off factor, which this work assumes fixed, equal to 0.1. Despite the circular dependency between these equations, it can be resolved iteratively by assuming an initial MODCOD, solving the system of equations, and computing the new MODCOD until convergence is achieved.

1.2.3 The resource allocation problem

In order to provide broadband access and, more generally, communication services to users, satellite operators need to decide how to distribute the resources available on the network to the individual users. Collectively, these decisions are known as the Resource Allocation Problem (RAP) and involve a series of *sub-problems*. To explain the specific decisions, an example providing broadband connectivity, specifically regarding the return link (users to gateways) is subsequently explained. First, the information starts on the user antenna, and needs to be transmitted to the satellite network. When users have more than one visible satellite, the operator needs to decide which satellite or set of satellites will provide services to the user at each point in time. Throughout this dissertation, this is referred as the **Satellite Routing** problem, although it is also known as the Satellite Scheduling or User Association problem. Once the information reaches the satellite and under presence of inter-satellite links, the operator needs to decide the routing protocols to transmit the information through the satellite network and to the final satellite, i.e., the satellite that will transmit the information to a gateway. Note that, under no inter-satellite links, the initial and final satellite must be one and the same. Deciding how to transmit the information through the satellite network will be referred to as the **Inter-Satellite Routing** problem. Note that the architecture of the satellite influences how the data and signal must be handled between nodes. For instance, for bent-pipe architectures, the MODCOD is maintained throughout the entire communication, while architectures with on-board processing might allow signals to change MODCOD for better transmission, and possible fragment the data packets into multiple, smaller packets which can follow different paths. After reaching the final satellite, the information must be transmitted to the gateway, so that it can reach the Internet. Deciding how individual gateways are mapped to

satellites, as well as deciding how to map users and gateways is referred to in this dissertation as the **Gateway Routing** problem. Note that for the forward link, the decisions are the same, although the meaning of initial and final satellites is reversed.

On top of this, each time a signal is transmitted from one antenna to another, operators must decide the 4 characteristics of the beam: shape, direction, frequency, and power. Deciding the shape of each beam is referred to as the **Beam Shaping** problem. Deciding the direction for each beam is split into two decisions: 1) deciding the users served by each beam, referred to as the **User Grouping** problem, and 2) deciding the central direction of each beam, referred to as the **Beam Placement** problem. Deciding the frequency of each beam is split among four sub-problems. Deciding the carrier frequency and bandwidth for each beam is referred to as the **Frequency Assignment** problem. Deciding the activation time for each beam is referred to as the **Beam Hopping** problem. Note that the Frequency Assignment and Beam Hopping attempt to split the spectrum in space and time, respectively. Since a beam can cover multiple users, deciding how to split the spectrum in time and space within the beam is referred to as the **Time-Division Multiple-Access (TDMA)** and **Frequency-Division Multiple-Access (FDMA)** problems, respectively. Finally, deciding the power for each beam is referred to as the **Power Allocation** problem. Note that all these decisions assume that the operator has a high degree of control on the spacecraft payload and they can be reassessed during operations.

These eleven sub-problems, visually depicted in Figure 1-3 compose the RAP and are grouped into six critical decisions, depending on the resource the address, represented in Table 1.2. To effectively serve users, operators must determine the power, frequency, direction, shape, relaying satellite, and offloading gateway for each beam. It is important to note that, beyond the four intrinsic characteristics of the beam detailed in Section 1.2.2, the substantial increase in both space and ground segments necessitates considering multiple satellites and gateways in Line of Sight (LoS), introducing additional decision variables. Furthermore, the dynamic nature of NGSO satellites and mobile users results in the physical positions of antennas changing over time, making the equations presented in Section 1.2.2 time-dependent.

Despite the prior split among sub-problems, certain decisions may impact the same segment of the overall system, resulting in interdependencies among sub-problems. Two types of dependencies are identified: circular dependencies, where determining the feasibility of a decision for two sub-problems necessitates addressing both simultaneously, and linear dependencies, where a sub-

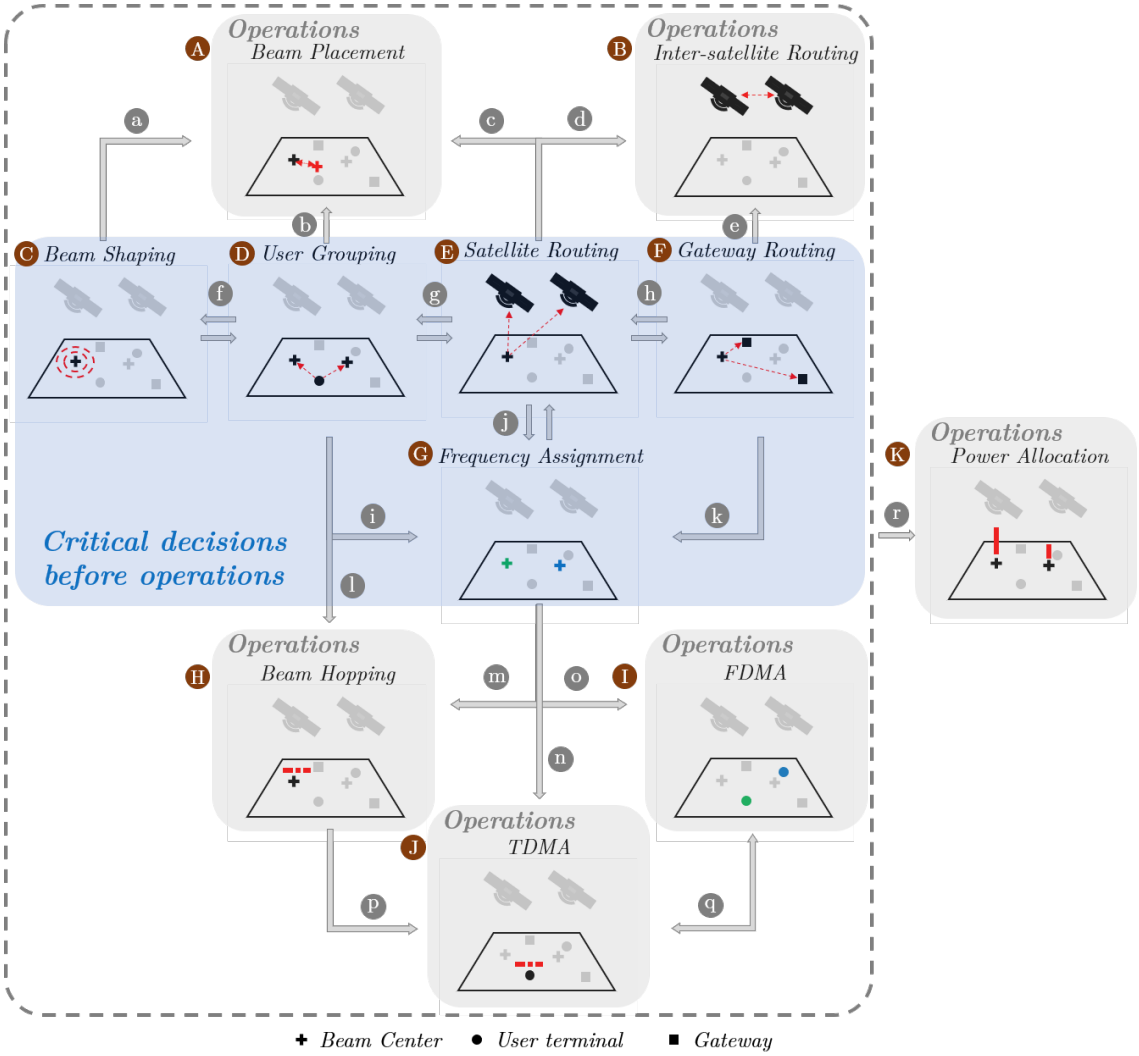


Figure 1-3: Resource Allocation sub-problems and dependencies between sub-problems. Each sub-problem is associated with a letter, which do not represent a specific resolution sequence. Each image represents an RAP sub-problem. Dependencies between sub-problems are indicated using grey arrows. The sub-problems in the blue regions need to be addressed before operations due to the complexity to find a feasible solution, while the sub-problems shaded in grey can be addressed during operations.

problem requires input from another sub-problem. The dependencies, as illustrated in Figure 1-3, include:

- Circular dependency between User Grouping and Beam Shaping (letter f): Determining which

Resource decision	Sub-problem	What is decided?	Addressed before operations	Letter
Shape	Beam Shaping	Shape of each beam	Yes	C
Direction	User Grouping	Users served by each beam	Yes	D
	Beam Placement	Position of each beam	No	A
Frequency	Frequency Assignment	Central frequency and bandwidth for each beam	Yes	G
	Beam Hopping	Activation time of each beam	No	H
	Bandwidth Allocation - FDMA	Bandwidth for each user within the beam	No	I
	TDMA	Time-window for each user within the beam	No	J
Transmission Power	Power Allocation	Power transmitted to the antenna	No[3]	K
Relaying Satellite	Satellite Routing	Relaying satellite for each beam	Yes	E
Offloading Gateway	Gateway Routing	Offloading gateway for each beam	Yes	F
	Inter-Satellite Routing	Best path from the initial to final satellites	No	B

Table 1.2: Resources and their associated sub-problems on the RAP for NGSO constellations. The letter corresponds to the associated image in Figure 1-3.

users can be served by each beam depends on the shape of the beam, while determining the shape of the beam relies on knowledge of the users served by the beam.

- Circular dependency between User Grouping and Satellite Routing (g): Knowing the footprint of the beam depends on the satellite serving each beam, while the demand of each beam, influencing the satellite load, can only be determined once the mapping between users and beams is decided.

- Circular dependency between Satellite Routing and Gateway Routing (h): Beams can only be mapped to satellites that have access to the associated gateway, but the mapping between gateways and beams depends on the satellites visible to the beam.

- Circular dependency between Satellite Routing and Frequency Assignment (j): Both sub-problems influence the potential signal interference between beams.

- Circular dependency between FDMA and TDMA (q): Both sub-problems influence the po-

[3]Power Allocation can be solved in real-time as long as there exists limitations on the user demand which guarantee that a certain threshold will not be surpassed

tential data-rate of the beam. Note that these two sub-problems tend to be addressed simultaneously as the Multi-Frequency Time-Division Multiple-Access (MF-TDMA) sub-problem.

- Linear dependency of Frequency Assignment on User Grouping (i): The mapping of users to beams, determined by User Grouping, influences the demand of each beam.
- Linear dependency of Frequency Assignment on Gateway Routing (k): The available bandwidth is contingent on the mapping of beams to gateways.
- Linear dependency of Beam Placement on Beam Shaping (a), User Grouping (b), and Satellite Routing (c): These three elements collectively limit the feasible locations of the beam center.
- Linear dependency of Inter-satellite Routing on Satellite Routing (d) and Gateway Routing (e): Routing the information requires knowledge of the initial and final satellites in the network.
- Linear dependency of Beam Hopping on User Grouping (l): The mapping of users to beams influences the demand of each beam.
- Linear dependency of Beam Hopping on Frequency Assignment (m): The active duration of a beam depends on the available bandwidth.
- Linear dependency of FDMA on Frequency Assignment (o): The spectrum available for users within the beam depends on the spectrum available for the beam.
- Linear dependency of TDMA on Frequency Assignment (n) and Beam Hopping (p): The active duration of a user depends on the available spectrum and time for the beam.
- Linear dependency of Power Allocation (r) on all other sub-problems: All decisions collectively affect the link budget equation.

To address the complete RAP for satellite communications, all outlined decisions and sub-problems need to be systematically resolved.

The temporal dynamics inherent in modern satellite systems pose a challenge for automated resource allocation tools. These tools must either make instantaneous decisions based on the current state of the system or generate decisions valid for longer time horizons, reducing the need for real-time control. However, certain decisions are particularly challenging to address in real-time due

to their inherent complexity. For example, to prevent harmful interference, a prior agreement on the frequency spectrum is necessary for all beams. This presents a high-dimensional, intricate, and computationally expensive problem [46]. Generally, real-time methodologies exist to address the Beam Placement, Beam Hopping, FDMA, TDMA, Power Allocation, and Inter-Satellite Routing. It is envisioned that real-time methodologies allow operators to reassess these decisions in the scale of miliseconds or seconds. On the other hand, while the rest of decisions can be reassessed over time, solutions for the other five sub-problems, shaded in blue in Figure 1-3, which are the Beam Shaping, User Grouping, Frequency Assignment, Satellite Routing, and Gateway Routing, must be valid for long time-horizons. It is envisioned that operators will reassess these decisions in a weekly or monthly basis, depending on the computation and telemetry time required to obtain a new allocation and transmit it to the satellites, gateways, and users involved. Their resolution is essential before operations commence due to the complexity of the underlying sub-problems, making real-time decision-making challenging.

1.3 General Objectives

The aim of this dissertation is to expand upon existing frameworks that solve the complete RAP to efficiently leverage the novel flexibilities and give insight that can be used by satellite operators when developing a new satellite communications constellation. To achieve this goal, the following general objectives have been defined:

- To create effective autonomous decision-making methods tailored to modern systems, aimed at assisting satellite operators in managing communication constellations.
- To provide insights into design trade-offs for the next generation of constellations, encompassing unique opportunities presented by novel configurations within operational contexts.

1.4 Thesis structure

The remainder of this dissertation is structured as follows: Chapter 2 reviews existing literature addressing all or some of the sub-problems contained in the RAP, as well as current works addressing the design of satellite constellations; Chapter 3 describes the simulation models necessary to recreate realistic operational conditions, as well as appropriate data sources for each model; Chapter 4

proposes a new methodology for the joint Beam Shaping and User Grouping sub-problem, tailored for hybrid constellations; Chapter 5 proposes a new framework combining Satellite Routing and Frequency Assignment methods for interference mitigation in large constellations; Chapter 6 proposes a new technique to address the Gateway Routing sub-problem, and studies the design of supporting infrastructure; Chapter 7 proposes a novel RAP framework and uses it to study the performance of current megaconstellation designs; Chapter 8 studies the design of hybrid megaconstellations through the usage of the RAP framework; and Chapter 9 summarizes the work conducted in this dissertation, states the main contributions, and describes potential areas of future research.

Chapter 2

Literature review

Before diving into the contributions of this dissertation, it is important to understand what are existing practices on the field and what are the research gaps when operating large mega-constellations. To that end, this chapter discusses the relevant literature for this dissertation, delving deeper into methods to address the RAP. In particular, approaches that solve one or more of the RAP sub-problems, as well as studies on the novel megaconstellations are reviewed. Based on this, the research gaps and contributions are identified.

The research questions that this section aims to address are:

Research question 2.1

What flexibilities are missing on current automatic RAP frameworks to be able to address the complexities of the novel satellite constellation operations?

Hypothesis: The purpose of this question is exploratory.

Research question 2.2

What critical data is lacking for satellite operators to strategically plan and design the upcoming satellite communications infrastructure?

Hypothesis: The purpose of this question is exploratory.

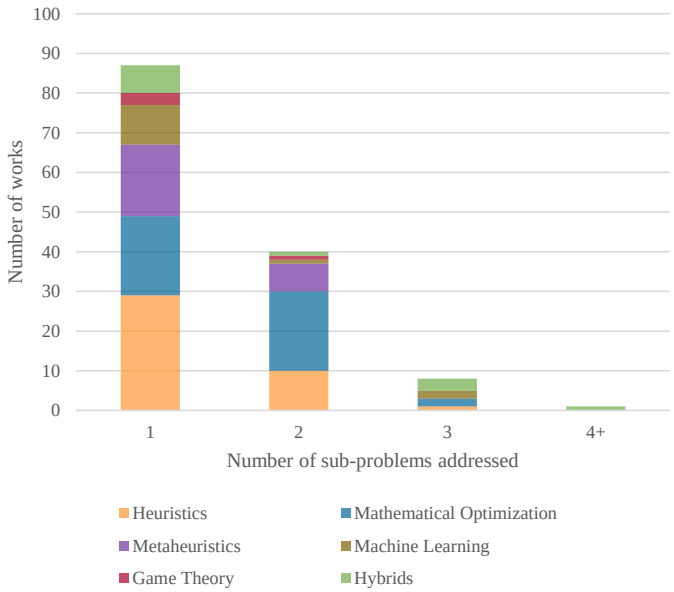


Figure 2-1: Literature on the RAP sub-problems sorted by solution method

2.1 Research on RAP frameworks

The individual sub-problems that compose the RAP[1] are an active area of research (see Figure 2-1). Nevertheless, due to the intricate dependencies between resources, the numbers of works decrease when increasing the number of sub-problems addressed simultaneously. Currently, only eight papers attempt to solve three sub-problems at the same time, and only one attempts to solve more than three.

In particular, Tian [47] addresses the joint Beam Hopping, Bandwidth Allocation, and Power Allocation problem by proposing a greedy heuristic. Their method allocates power and bandwidth to one beam at a time until there are no more resources left to assign, where beams are sorted by decreasing traffic demand. Since they assume static MODCODs, the power is proportional to the bandwidth, which simplifies the link budget equation. The authors show how this straightforward method can multiply the throughput of the satellite by a factor of two compared to fixed and randomized allocations, while maintaining the Quality of Service (QoS) when managing satellites with up to 12 beams. It is important to highlight that the spot beam configuration, which includes

[1]The RAP is also referred in the literature as Dynamic Resource Allocation (DRA) or Dynamic Resource Management (DRM)

the shape and direction of each beam, as well as the users being covered, is assumed to be given by the parameters of the model. Furthermore, the proposed technique assumes that there is no possible interference between the allocated satellite and nearby satellites, which requires a conservative preemptive Frequency Assignment, as well as no frequency reuse mechanisms that allow to use each bandwidth more than once. In addition, modern technology allows for thousands of beams to be managed at the same time, which poses questions to the scalability of this method as it is not investigated in this work.

The joint Beam Hopping, Frequency Assignment, and Power Allocation problem is studied by Zao [48], Tang [49], and Leng [50]. The first two works adopt decomposition strategies where each sub-problem is addressed individually. Particularly, in both works, the Beam Hopping and Frequency Assignment sub-problems are solved using heuristic algorithms that address more interesting beams first, while Power Allocation is solved using a Lagrangian dual formulation that yields an optimized solution. However, Zao addresses first Beam Hopping, then Frequency Assignment, and finally Power Allocation. The authors show how these techniques can improve total throughput compared to uniform allocations, even under uncertain information, by benchmarking the performance of a LEO satellite with 12 beams serving 240 users. On the other hand, Tang addresses first Frequency Assignment, then Beam Hopping, and finally Power Allocation. Here, the authors prove how considering Beam Hopping methods can outperform techniques without this consideration by running simulations on a LEO satellite under realistic conditions for up to 91 beams. Finally, Leng addresses all three problems simultaneously by training a proximal policy optimization method that learns the best policy for each beam. By comparing it against randomized and proportional approaches, the authors show the policy learning model can reduce the total latency by 30%. Again, it is important to note that the spot beam configuration, as well as the mapping of beams to satellites are assumed known before applying the proposed methods. Furthermore, while interference considerations with external satellites are included, there is no mechanism for interference avoidance between satellites that include the proposed techniques, since this would require additional communication between the satellites to avoid frequency overlapping. In addition, while frequency reuse mechanisms are included in their respective formulations, the flexibility of the satellite payload is assumed to be low enough so that the systems can be managed in real time. Finally, similar to previous works, modern systems allow for thousands of beams to be managed at the same time, which questions the scalability of these methods.

The joint Beam Shaping, Bandwidth Allocation, and Power Allocation problem is studied by Ortiz-Gomez in [51]. There, the authors propose a multi-agent deep reinforcement learning to address all three sub-problems at the same time, where the actions of each agent are tied to the actions to the specific resource decisions. Specifically, the authors allow each agent to increase/reduce power, bandwidth, and beamwidth, or do nothing. By training all agents at the same time against a share pool of resources, they are able to learn the best allocation of resources for each beam. When simulating a GSO satellite with 82 beams, the results show how the agents are able to follow the required demand closely, with 30% of the power usage compared to uniform allocations. Similar to previous frameworks, the proposed method relies on the assumption that the beam configuration, satellite mapping, and frequency per beam are given and ensure no interference between nearby beams and satellites independently on the final allocation of resources. Similarly, since bandwidth is only assigned from a single pool of resources and no interference mechanisms are considered, it is unclear how to apply the proposed method under scenarios with frequency reuse. Finally, the expansion of this technique to cases with variable footprints, dynamic satellite assignments, and a larger user base remains to be studied.

Next, the joint User Grouping, Frequency Assignment, and Power Allocation problem is addressed by Deng [52] and Angeletti [53]. Similarly to the frameworks mentioned previously, both works divide the joint problem into simplified sub-problems. For the former, Deng first solves the User Grouping method using an extended version of k-means, and then approaches the joint Frequency Assignment and Power Allocation using the Lagrange dual method. By simulating 2 GSO satellites serving 10,000 users grouped in 40 beams, the authors show an increase of 33% in capacity over an existing baseline, which consists of a randomized allocation for User Grouping and proportional allocation for the joint Frequency Assignment and Power Allocation problem. For the latter, Angeletti first solves User Grouping by proposing a Voronoi tesselation of the satellite Field of View (FoV) and assigning each user to the nearest beam, then addresses the Power Allocation by proposing a formulation for which a greedy heuristic gives optimal solutions, and finally solves the Frequency Assignment by converting the problem into a coloring problem, which can be solved easily for grid-like architectures. By simulating thousands of users grouped in hundreds of beams, the authors show how novel beam forming technologies can be successfully implemented in satellite communications. Although these two approaches address the problem of grouping users into beams, they assume beams of fixed shape. Therefore, it is unclear how to apply the proposed methods

in cases with variable footprints, which occurs when considering NGSO satellites, variable beam shapes, or satellites at various altitudes. Even more, methods for the mapping between beams and satellites are not explored. Furthermore, the proposed formulations do not allow for beams with non-fixed bandwidth. Finally, although they consider cases with thousands of users, the number of beams simulated still poses questions on how these systems would apply in cases with tens of thousands of beams.

Lagunas [10] solves two problems: the joint Beam Placement and Frequency Assignment problem for satellite downlink, and the joint Bandwidth Allocation, Frequency Assignment, and Power Allocation problem for satellite uplink. For the former, the authors propose a two-level decomposition based on solving the Beam Placement using prior literature [54], and the Hungarian algorithm for beam per beam resolution of the Frequency Assignment. For the latter, they also propose a two-level decomposition based on solving the joint Frequency Assignment and Power Allocation problem using an adapted Hungarian algorithm, followed by a greedy heuristic that can be proven to give optimal bandwidth to the beams. The authors test their methods for a GSO satellite over Europe with 250 beams and they achieve increase capacity against more static methods where not all flexibilities are considered. Again, the models used in this work assume fixed footprints, which might not be realistic for modern constellations. In addition to this, given the nature of GSO satellites, the mapping between beams and satellites is not addressed, and requires from operators input. Finally, the order of magnitude of the beams in the simulations poses questions on how these systems scale to the dimensionality of modern constellations.

Finally, the author [12] presented a five-level decomposition framework to address the joint User Grouping, Satellite Routing, Gateway Routing, Frequency Assignment, and Power Allocation problem by addressing each problem one at a time. In this framework, each decision is addressed independently by including known information from the rest of sub-problems, such as the gateway visibility windows into the Satellite Routing problem. Specifically, User Grouping is solved using a multi-objective Genetic Algorithm (GA), where the objectives are to minimize the amount of virtual demand while maximizing the number of beams, Satellite Routing is solved using Particle Swarm Optimization (PSO) that minimizes the amount of interference between nearby beams, Gateway Routing is solved by balancing the load between gateways using Integer Linear Programming (ILP), Frequency Assignment is solved using a previous ILP implementation proposed by Garau-Luis and the author in [55], and Power Allocation is solved optimally using a greedy heuris-

tic under relaxed system constraints. By comparing the optimization framework against heuristic approaches, the authors show an improvement by a factor of 12 in resource efficiency in scenarios with 20,000 users using the O3b mPower constellation[2]. Additionally, the impact of each step in the optimization implementation is assessed by running a set of experiments with different characteristics and comparing the average improvement. As a summary, this work reveals that using an optimized User Grouping has no substantial advantage against grid-like methods, that reducing the number of beams is always beneficial, that the Satellite Routing and Gateway Routing formulations offer trade-offs in terms of power and system capacity, and that using a Frequency Assignment optimization is always beneficial. Nevertheless, this work has several drawbacks that prevent it from applying it to modern systems: 1) While the framework can be adapted to include variable footprints, only fixed-shape beams are considered, and the impact of variable footprints is not assessed, 2) NGSO constellations are allowed as long as they are composed of a single equatorial plane, which does not match most proposals, 3) Although ISLs are included in most proposals, the Gateway Routing formulation presented does not allow for such technology.

Summary Most current RAP frameworks rely on decomposition strategies were each sub-problem is addressed individually or as a pair. Given their relevance in prior stages of the satellite communications industry, most of the focus has been on GSO satellites with limited number of users (<10,000) or limited number of beams (<1,000), although there have been steps in extending the frameworks towards high dimensional NGSO systems. Note that, due to the increase in number of variables and possible solutions, high dimensionality might render existing solutions for low dimensional scenarios unfeasible. The solution techniques are diverse, going from greedy heuristics to machine learning and mathematical optimization. Although valid for older constellations, current RAP frameworks need to address major gaps before being applicable to the next generation of satellite communications. According to the new proposals (see Table 1.1), automated decision making frameworks for the RAP need to account for: 1) variable footprints, 2) inclined orbits, 3) ISLs, 4) NGSO conditions, 5) high dimensionality (i.e., >10,000 users/beams). Table 2.1 summarizes the capabilities of current RAP frameworks compared to the design characteristics of current constellation designs. Note that the variable footprints introduced by satellites at multiple altitudes, multiple planes in inclined orbits, and constellations with ISLs have not yet been addressed

[2]Which corresponds to 10 satellites in equatorial orbits at 8062km

System characteristics	Current literature				Recent constellation proposals			
	[10, 51]	[52, 53]	[47–49]	[12]	SES-17, Viasat GSO	O3b mPower	OneWeb, Intelsat, CASC	SpaceX, Amazon, Telesat, O3b†, Viasat†, Boeing
1-3 GSO	✓	✓	✓	✓	✓			
1-30 MEO			✓	✓		✓		
100+ LEO ∪ MEO ∪ GSO							✓	✓
ISLs								✓
$<$ 10,000 users/beams	✓	✓	✓	✓	✓			
$\geq$ 10,000 users/beams		✓		✓		✓	✓	✓

Table 2.1: Capabilities of current literature against the recent constellation proposals. The designs marked in green correspond to the constellations solvable with current RAP frameworks. The designs marked in red correspond to the constellations **not** solvable with current RAP frameworks. † Represents the new MEO/LEO designs of the denoted companies

by current literature. Furthermore, novel RAP frameworks need to account for the NGSO and high dimensional conditions of modern systems.

2.2 Research on individual RAP sub-problems

Given that current RAP frameworks cannot address the flexibilities of novel systems, a novel refined RAP framework needs to be developed. However, given that the new framework can be a combination of previously proposed methods for the individual sub-problems, or pairs of sub-problems, an exhaustive search is necessary to understand if the main drawbacks of general RAP frameworks have been addressed at a sub-problem level. To that end, works that address one or two sub-problems of the RAP are summarized in the following sections.

2.2.1 Beam Shaping

The Beam Shaping sub-problem (letter C in 1-3) is defined as finding the shape for each beam that maximizes performance. The works of Sherman [56] and Zhao [57] attempt to cover a circular layout using ring-type subdivisions and a GA implementation. Similarly, Okello [58] aims to cover any type of shape using linear divisions of the user layout. Although all three achieve the desired link margin, this type of definition is intended for fixed antennas that always see the same portion of the Earth, or antennas with fixed projections that rotate with the satellite. Given the dynamic nature of modern NGSO systems and the usage of spot beam technology (i.e., a beam always covers the same potion of the ground), these methods are not suited for the next generation of constellations.

Qian [59], Wenqian [60], Zhang [61], and Camino [62] attempt to solve the same layout covering

problem, but using circles of varying shape, which is more suitable for modern technology, since the covering of a circular shape can be easily transferred between satellites. Specifically, Qian proposes to shrink beams in hot spot areas to achieve a more balanced user allocation between the beams, Wenqian proposes an iterative approach were the beams are changed in size at each iteration until all resources are utilized, Zhang proposes to allocate excess resources to certain beams by making them bigger, so that they can serve more users, and Camino proposes a Mixed-integer linear programming (MILP) formulation that decides the shape of each beam to achieve the best balance between the beams. Although they effectively address beams of varying size to cover a set of users by trying to achieve balance between the beams, all these works assume that the same satellite is serving all the beams. The reality of satellites at different altitudes orbits entails that the Earth footprints are different because they come from different satellites, which means that assigning shape and assigning altitude cannot be solved as independent problems.

Summary There is limited research addressing the problem of variable footprints in the context of constellations with satellites at different altitudes. To address the reality of novel applications, a new method that assigns shape and altitude at the same time needs to be developed.

2.2.2 User Grouping

Generally, the User Grouping sub-problem (D) is defined as finding the mapping between users and beams that maximizes performance. Note that grouping multiple users into single beams allow for a better statistical multiplexing of user traffic, enabling a larger utilization of the beams individual resources. When considering fixed circular shapes, the problem can be transformed into a minimum geometric disk cover, which Fowler [63] proves to be Nondeterministic Polynomial time (NP) hard. Three approaches attempt to solve this problem for satellite communications: Yao [64], the author [65], and Dinh [66]. Specifically, Yao develops an adapted k-means implementation to include load balancing considerations in the clustering problem. By comparing against random allocations, or no grouping (i.e., where each user receives their own beam), the authors show how their method can yield between 5% and 10% improvement in spectral efficiency for a GSO satellite with up to 100 beams. On the other hand, the author proposes a multi-objective GA implementation to reduce the amount of virtual demand (i.e., demand as a consequence of integer constraints, rather than real demand), while maximizing the number of beams. By including considerations about the

projection of the beam over the Earth, the author proves how this approach can be applied to high dimensional NGSO constellations (tested on the O3b mPower constellation with 20,000 users) and leads to solutions with at least 40% power reduction compared to previous heuristic methods. Finally, Dinh proposes using Quantum Annealing, which takes advantage of the exponential time reduction of quantum computers, to solve this problem. By showing results with up to 200 users and comparing against more simple heuristic, the authors show the effectiveness of this solution. However, they acknowledge that, due to the capacity limitation of current quantum computers, the method cannot scale for more than 200 users. Despite the techniques proposed by the authors proving efficient in addressing the User Grouping sub-problem, they fail to account for the fact that assigning users to beams in constellations composed of different altitudes, also entails assigning users to altitudes. Even when assuming that all altitudes have similar footprints, grouping users into individual beams means that all the grouped users will be served by the same satellite, which implies that they will consume resources from the same pool.

In [12], the author shows how the User Grouping can reduce the Unmet Demand (UD) by a factor of 10, while reducing radiated power by 80%. However, using a grid-like algorithm performs similarly to optimized approaches. An effective grid-like approach for NGSO satellite communications has been proposed by Ivanov [67].

Summary To maximize capacity, grouping users into clusters is always beneficial. However, grid-like techniques offer similar performance compared to optimized approaches where the objective is to minimize the number of beams. When addressing the User Grouping sub-problem individually, there is limited research addressing the fact that grouping users into beams also entails grouping users into satellites.

2.2.3 Beam Placement

The Beam Placement sub-problem (A) is defined as finding the exact location of each beam center such that the power consumption is minimized. Finding a solution for the Beam Placement sub-problem for fixed terminals given a user-beam mapping (i.e., once the User Grouping sub-problem has been addressed), can be done by finding the center of smallest-circle that encloses all users [68], which can be done using polynomial algorithms. For moving terminals, Xu [69] proves how user-staring technology can be used effectively to track the user.

Summary Given a User Grouping solution, finding the center of the beam can be done efficiently for both fixed and mobile users.

2.2.4 Frequency Assignment

The Frequency Assignment sub-problem (G) consists of assigning the central frequency and bandwidth for each beam such that the capacity is maximized while minimizing the interference between beams. Under fixed bandwidth, Mizuike [70] prove that the sub-problem is NP-complete, and develops a heuristic that assigns the beams in descending order based on the available frequencies. Due to the complexity of the sub-problem, many approaches rely on machine learning or metaheuristic approaches, which prove to obtain allocations efficiently: Funabiki [71] proposes a gradual Neural Network (NN), Salcedo [72, 73] describes a Hopfield NN combined with Simulated Annealing (SA) and GA implementations, Wang [74] addresses the concepts of stochasticity and noise detailed by Wang [75] with an adapted Hopfield NN, Hu [76] and Garau-Luis [77] discuss the implementation of Deep Reinforcement Learning (DRL), and Salman [78] details differential evolution algorithms, which Wang [79] expands upon by incorporating several objectives. All these implementations address the Frequency Assignment sub-problem efficiently and provide *good* solutions. However, due to the lack of insight about how *good* a solution is, research has shifted towards mathematical programming methods, which can give mathematical bounds as to how far the optimal solution is. In this line, Houssin [80] transforms the interference into cummulative interference, and proposes an ILP approach, which can be used with off-the-shelf mathematical solvers for cases with up to 200 users. Furthermore, Garau-Luis and the author [55] develop another ILP formulation where interference is treated as a binary constraint. Given the poor scalability of the formulation, they propose a iterative-based approach that is able to provide optimized solutions in under 20 minutes in cases with 20,000 users, while including the flexiblity of modern constellations, such as variable bandwidth, multiple frequency reuses, and multiple polarizations.

Summary Current methods to solve the Frequency Assignment sub-problem have proven to be effective to address high-dimensional conditions in highly flexible configurations.

2.2.5 Beam Hopping

The Beam Hopping sub-problem (H) consists of deciding when to turn on/off each beam so that the demand is met, while avoiding harmful interference. Note that the Beam Hopping is analogous to the Frequency Assignment, but in the time-domain, with no clear advantages of one over the other. Several studies by Angeletti [81] and Anzalchi [82] use GA to prove that Beam Hopping yield increased capacity compared to systems that use classical Frequency Assignment methods. In view of these results and to assess which algorithm gives the best results for the Beam Hopping sub-problem, Li [83] compares several metaheuristic and machine learning techniques using a LEO constellation with tens of thousands of beams and ranks them in terms of system throughput. Other research has focused on proposing Beam Hopping techniques that provide optimized solutions for a variety of objectives: Han [84] and Hu [85] center their stochastic gradient descent and DRL approaches, respectively, in maximizing user satisfaction, Wang [86] proposes a GA implementation to deal with both user delay and system capacity, and Lei [87] focuses on maximizing user fairness using NN. Finally, Zhang [88] discusses a cluster based formulation that can give better understanding about the found solution compared to less intuitive methods.

Summary A variety of methods considering different objectives and techniques have been proposed for the Beam Hopping sub-problem. Current methods prove to be effective to maximize system capacity under high dimensional NGSO conditions.

2.2.6 Bandwidth Allocation - FDMA

The Bandwidth Allocation or Frequency Division Multiple Access sub-problem (I) consists of dividing the frequency assigned to a beam into multiple sub-sets of frequencies and bandwidths (commonly referred to as *carriers*) for each user. Given that this technology was already utilized in earlier generations of the satellite communications industry, many formulations to address the different aspects of each configuration have been proposed.

First, the standard method to solve this sub-problem is the water-filling algorithm, which revolves around the idea of allocating more resources to those beams that require a higher throughput. This type of allocation aims at maximizing capacity by linearizing the equations and assumes that bandwidth is continuous. Although this approach proves to maximize capacity, it ignores other considerations:

- Approaches with other objectives: Park [89, 90] presents an heuristic that maximizes user fairness, instead of capacity. In a further work, Wang [91] refines the previous formulation to obtain a multi-objective formulation that can derive the best trade-off between fairness and capacity. Liu [92] defines a user-satisfaction metric that includes data-rate, quality of service, and user priority, among other factors, while solving the problem using ant colony optimization.

- Approaches with discrete bandwidth: Since all bandwidth resources are drawn from the same pool, research has focused on approaches that follow game theory, where bandwidth is the resource that agents must compete for. To this end, Li [93] proposes an asymmetric monopoly model, Su [94] develops a Stackelberg differential game, and Wang [95] combines learning algorithms with game theory to achieve a bandwidth allocation. All approaches are based on reaching the Nash equilibrium to determine the best distribution.

- Approaches with different decisions: Other than the bandwidth for each user, Bisio [96] describes the entities of the problem as a combination of virtual and physical entities that compete for resources, which proves to save power, while Abe [97] includes control input that tries to regulate the output of the system based on closed-loop control.

- Approaches with different constraints: Kawamoto [98] describes how to include interference constraints in the Bandwidth Allocation sub-problem and proposes a heuristic to mitigate the potential loss of information.

- Approaches with different time-horizons: Xiao [99] discusses how using a long term approach combined with short term variations increase the total capacity.

Summary The Bandwidth Allocation sub-problem (or FDMA) has been studied under a variety of objectives, decisions, and constraints. When maximizing capacity, current approaches prove to be effective in splitting bandwidth between users of a beam.

2.2.7 TDMA

The Time Division Multiple Access (TDMA) sub-problem (J) consists of assigning a time-slot to each user in a beam such that the demand of all users is met while avoiding harmful interference.

Similar to FDMA, this technology has already been used in prior stages of this industry and has been proven to increase capacity against more static approaches. Given that the problem is NP-complete [100], most of the proposed techniques revolve around heuristic algorithms: Park [100,101] proposes an iterative method in which each user is allocated the minimum amount of slots necessary to serve their demand until all resources are exhausted, Dong [102] proposes splitting the time slots using a recursive tree, and assign nodes, not necessarily leaves, to users, and Feng [103] explores reserving certain slots for users, and allocating the rest of a first-come first served basis. Build on top of this, other works explore the addition of beam-to-beam interference to the constraints (Bejarano [104]), or how expand optimal solutions onto high dimensional environments (Lee [105]).

Summary The solutions proposed for the TDMA sub-problem have effectively addressed the complexity of including high dimensionality, beam-to-beam interference, and real-time in the formulation.

2.2.8 Power Allocation

The Power Allocation sub-problem (K) consists of assigning the power to each beam such that the demand requirements are met. Since power variation was one of the first flexibilities to be introduced to the satellite, Power Allocation is one of the most studied sub-problems. Depending on the technological constraints considered, most approaches can be divided into four categories:

- Amplifier-constrained problems for power consumption minimization: where the number of power amplifier blocks is less than the number of beams. Since this formulation has been proven to be NP-complete [106], many approaches rely on modern optimization techniques such as metaheuristics and machine learning to obtain *good* solutions in reasonable time: Aravanis [106] proposes a combination of SA and GA, Efrem [107] details a successive convex optimization, Durand [108] uses particle swarm optimization (PSO), Zhang [109] and Garau-Luis [110] explore a DRL application, and Liu [111] describes a game-based approach where different beams are competing for power. The objective of all these works is to match the demand as closely as possible. As a comparison, Garau-Luis and the author [112] benchmark different approaches and conclude that, while machine learning techniques perform better under time-restricted scenarios, GA implementations tend to be more robust and slightly reduce the power consumed.

- Non amplifier-constrained problems for power consumption minimization: where each beam has its own amplifier, and the only restriction is the total power of the system. This formulation can be solved optimally by using the Lagrangian dual [113–115].
- Non amplifier-constrained problems for user satisfaction: since this formulation is again hard to solve, some heuristics have been proposed that maximize the Quality of Service, instead of the total capacity [116, 117].
- Including additional constraints: other works explore the addition of interference-based constraints [118] and weather-type perturbations [119] on the formulation.

Summary The Power Allocation sub-problem has been studied under a variety of objectives, decisions, and constraints. When maximizing capacity, current approaches prove to be effective in under both constrained and unconstrained considerations.

2.2.9 Satellite Routing

The Satellite Routing sub-problem (E) consists on deciding, at every point in time, the mapping between beams and satellites so that the demand of the users is met. Research is divided into three main lines of work depending on the nature of the communications:

- Mobile networks: where the users of the network want to establish a link to another user, which will span over some minutes. The demand for communication can happen anywhere and anytime, and thus the system must be prepared to handle users on the spot and consistently across satellites. Given the historic relevance of satellite transmission for phone communications, this is the most studied system. Given the necessity for a quick consistent action, several works propose the usage of heuristic techniques to address the user-satellite mapping (Krewel [120], Papapetrou [121, 122], Chowdhury [123]). Recently, research has focused in including the evolution of the network over time (Wu [124]), dealing with high dimensionality (Zhu [125]), and including load balancing considerations (He [126]).
- Task-based networks: where users on the network want to transmit large amounts of data in packages, but how much data needs to be transmitted is known in advance. Typical applications for this type of networks are imaging satellites and internet of things (IoT)

applications. Compared to the previous network, these systems are not so restricted in terms of computation time, which allows for other more time-demanding approaches to be used. Specifically, Xhafa [127] develops a GA implementation, Zhuang [128] proposes an artificial bee colony, Tharmasa [129] details a markov decision process formulation, Chen [130] discusses a MILP formulation that can be solved with off-the-shelf solvers, and Wang [131] explains a SA implementation combined with Monte-Carlo analysis. Xhafa [132] presents a survey including other approaches.

- Stationary networks: where users of the network require a link to be active during a long period of time, either for data transmission (back-hauling) or data reception (Internet-like communications). Note that these are the type of networks defined by broadband communication systems, which are the focus of this work. To address this sub-problem, Dai [133, 134] proposes a greedy approach where each user selects its own best set of satellites. Using a satellite-centric perspective, Yin [135] develops a DRL approach to learn the policy of which user stations to connect to. To deal with high dimensionality, Jiang [136] details a multi-objective iterative sub-gradient approach that aims to maximize coverage, capacity, and balance between the satellites simultaneously. Note that these approaches assume that there will be no interference between the users connected to a satellite, which corresponds to developing a conservative Frequency Assignment plan that ensures that no interference can happen. Such a conservative plan undermines the flexibilities of new satellites, leading to an under-utilization of resources. In terms of interference-based methods, the author [137] developed a MILP formulation to minimizes possible interference between beams on single-plane, equatorial constellations.

Summary The Satellite Routing sub-problem has been addressed effectively for historical types of networks, such as phone communications or task-based architectures. However, for broadband satellite networks, there is limited research including beam-to-beam and satellite-to-satellite interference under multi-plane multi-altitude NGSO constellations. To address the reality of novel operations, where satellites and users are constantly moving, a new method to include interference-based constraints under NGSO configurations needs to be developed.

2.2.10 Gateway Routing

The Gateway Routing sub-problem (F) consists of deciding, at each point in time, the mapping between beams and gateways such that the demand of the users is met. Depending on the characteristics of the network, different types of problems can be devised:

- Fixed beam-gateway association: each beam is linked to one gateway, and all the traffic is routed through that gateway. For configurations without ISLs, the author [12] developed an MILP formulation that balances the load between gateways for a more efficient usage of resources. This approach relies on the assumption that the satellite assigned to the beam has the gateway in FoV at all times, which requires a conservative formulation for the Satellite Routing sub-problem.

- Flexible beam-gateway association: each beam is linked to one gateway at each point in time, but the gateway may change over time. For configurations without ISLs, Crosnier [138] developed a heuristic that attempts to assign the best gateway based on the current configuration of the network.

- Satellites in on-board networking: the satellite network behaves like a terrestrial network, and packets are transmitted from node to node. Once the data from the user reaches the satellite, it is routed following routing protocols for dynamic networks [139]. The satellite to gateway connection behaves just like any other edge in the network. To decide the satellite-gateway mapping at each point in time, del Portillo [7] proposes a graph-based approach, where the number of gateways is upper bounded, but not optimized for.

Summary Configurations without ISLs have been studied under both fixed and flexible beam-gateway associations. However, there is limited research regarding solutions for fixed or flexible beam-gateway associations with ISLs. Since many applications have stated the usage of this technology, novel approaches that address the beam-gateway association in constellations with ISLs need to be developed. Furthermore, when on-board networking is present, no approach includes the satellite-gateway mapping as a variable to optimize. Similarly to the Satellite Routing sub-problem, the novel approaches should include beam-to-beam and satellite-to-satellite interference, and well as handling high dimensional scenarios under NGSO conditions.

2.2.11 Inter-Satellite Routing

The Inter-Satellite Routing sub-problem (B) is defined as finding the best path between the initial and final satellites assuming the usage of ISLs. Note that this is analogous to a classical flow routing problem with dynamic topology. Under ideal conditions, it can be solved using a classical network-flow assessment (del Portillo [7]). However given the packet delay, different approaches have been proposed to deal with these types of networks under realistic conditions: Werner [140] proposes a shortest-path approach that includes the network changes over time, Sigel [141] details an ant colony optimization to determine the best path, Li [142] details a Hopfield NN combined with SA, Rao [143] explains a GA implementation, and Rajagopal [144] proposes a beetle swarm optimization. All these approaches manage to effectively include the changes of the network over time to minimize the package delay. Techniques to maximize throughput have also been proposed (Sun [145]). Alagoz [146] surveys more different routing approaches that have been proposed during the years. A sub-set of studies also explore the possibility of multi-altitude inter-satellite routing by including LEO, MEO, and GSO satellites (Wang [147]). How to operate the ISL network to maintain a robust topology was studied by Qiao [148]. Lin [149] describes an Internet-like topology to efficiently route packages using predefined paths. Finally, Lee [150] analyzes the impact of ISLs in terms of average distance and average number of hops to target under different constellation sizes.

Summary The Inter-Satellite Routing sub-problem has been studied under a variety of objectives, algorithms, and topologies. Current approaches are able to effectively manage the reality of dynamic NGSO constellations in multiple altitudes.

2.2.12 Joint problems including Beam Shaping

Given the dependency between the Beam Shaping and the User Grouping sub-problems, several studies propose solutions that address both issues at the same time. To resolve this issue, some works propose first deciding the position of the beams, and then altering the shape to maximize performance. Specifically, Alinque [151] starts with a greedy solution to minimize the number of beams, followed by gradient descend optimization, Liu [152] details an adapted k-means implementation, where each user is assigned to the closest beam center, and the centers are updated using randomized search, Tang [153] proposes to first solve the Minimum Geometric disk cover, and then

adapt the shape of each beam using a p-center algorithm, and Liu [154] uses hierarchical clustering to decide the User Grouping allocation, and semidefinite programming to select the shape for each beam. Other approaches propose Voronoi maps mapped to ellipsoidal shapes (Honnaiah [155]), and MILP formulations that adapt both the shape and position of the beam at the same time (Camino [156]). Although each approach proves to effectively address the joint sub-problem, no study addresses the fact that variable footprints might be imposed by different altitudes, so that the mapping between users, beams, and shapes is also related to the mapping between users and altitudes.

In terms of other sub-problems, some works explore the possibility of solving other joint problems involving the Beam Shaping sub-problem: the joint Beam Shaping and Frequency Assignment problem is addressed by Camino [157] and Zhong [158], where both works propose to first decide the shape of the beam, and then the central frequency using heuristics, the joint Beam Shaping and Bandwidth Allocation problem is addressed by Kyrgiazos [159], where the authors propose an iterative algorithm that tries between two different beam shapes to achieve a better allocation, and the joint Beam Shaping and Power Allocation problem is addressed by Schubert [160], where the authors propose an optimal iterative approach for both shape and power at the same time.

Summary Several studies investigate how to effectively allocate the beam shape along with either frequency bandwidth or power. Despite including the user-to-beam mapping in the decision process, studies that research the joint Beam Shaping and User Grouping problem do not address the shortcomings of individual Beam Shaping methods: there is limited research solutions that effectively handle variable footprints coming from satellites at different altitudes. A new method that assigns shape, beams, and altitudes to users needs to be developed.

2.2.13 Joint problems including Frequency Assignment

The Frequency Assignment sub-problem is most commonly studied along the Power Allocation sub-problem. Current approaches include: graph theory (Jahn [161]) which transforms he problem into a coloring problem, for which efficient implementations are known, and iterative heuristics (Lei [162]), successive convex approximations (Abdu [163]), and SA (Vidal [164]) to address non-convex formulations. All these techniques show improvements in total throughput compared to addressing the problems independently.

In addition to the studying the effects of addressing the Beam Shaping and Frequency Assignment sub-problems at the same time shown in the previous section, research has also investigated the impact of:

- Changing the beam position: Kiatmanaroj [165–167] develops an ILP formulation for the joint Beam Placement and Frequency Assignment problem that proves to increase higher coverage compared to simple heuristics where each sub-problem is solved independently.
- Changing the user grouping: the author [168] shows how an improved User Grouping formulation can help increase the capabilities of Frequency Assignment algorithms to allocate frequency, thus increasing the total capacity of the system.
- Including routing considerations: Wan [169] analyzes how reserving channels can benefit both the Frequency Assignment and Satellite Routing problems to achieve increased capacity.

Summary Several studies investigate the effect of addressing the Frequency Assignment sub-problem along other decisions, which proves to increase total capacity. However, despite providing interesting results independently and being deeply correlated, no study attempts to solve the joint Frequency Assignment and Beam Hopping problem. Although this is not a necessity to address the RAP in modern constellations, future studies could research the potential impact of solving the Frequency Assignment and Beam Hopping sub-problems simultaneously.

2.2.14 Joint problems including Power Allocation

Given the historical relevance of the Power Allocation sub-problem, joint problems that include the power resource are the most studied:

- Power Allocation + TDMA: Wang [170] proposes a hybrid GA-PSO implementation to maximize system capacity.
- Power Allocation + Beam Hopping: Most approaches for this joint problem rely on non-convex mathematical formulations followed by heuristics that address power and time slots for one beam or group of beams at a time (Lei [171], Shi [172], Wang [173], Wang [174]). By considering the joint formulation, this works prove to achieve improved capacity compared

to methods that address each sub-problem independently. Finally, Alberti [175] studies the inclusion of weather impairments in the formulation for a more robust application.

- Power Allocation + Bandwidth Allocation: This joint problem is proven to be non-convex and hard to approximate [176]. Thus, many approaches rely on metaheuristics or machine learning moderns to yield *good* solutions: Cocco [176] propose a SA implementation, Paris [177] details a GA algorithm, the author [178] explains a PSO approach, Liao [179] describes a DRL model, and Zhong [180] treats the problem as a bargaining game. These works show how introducing the bandwidth flexibility to the Power Allocations sub-problem increases the system capacity. A comparison of these and more metaheuristics was done by Gao [181]. The inclusion of inter-beam interference was discussed by Jia [182], where a Lagrangian dual formulation followed by an iterative algorithm is proposed, and by Gao [183], where the authors propose a multiobjective formulation and find Pareto-Front solutions using PSO.

- Power Allocation + Beam Placement: Choi [184–186] and Takahashi [187–189] show how allowing certain movement on the beam centers increases the total capacity of the system.

- Power Allocation + Satellite Routing: Including power considerations on the Satellite Routing problem proves to increase system capacity by avoiding connections that require significant power (Liu [190]). Abdelsadek [191] draw similar conclusions where applying an adapted formulation to mobile networks using a GA implementation. Note that these works do not address beam-to-beam or satellite-to-satellite interference, which are the main shortcomings of current Satellite Routing approaches.

Summary The Power Allocation sub-problem in combination other sub-problems has been effectively addressed. Novel RAP frameworks can make use of the proposed methods to handle power in modern NGSO constellations under high dimensionality conditions.

2.3 Summary of literature on the RAP sub-problems

Table 2.2 summarizes the scope of the works on the RAP sub-problems. As shown, most works focus on an individual sub-problem, although there is some research addressing multiple sub-problems. Although all sub-problems have been addressed individually, research addressing multiple resources

simultaneously is limited. Furthermore, some individual sub-problems have not yet been explored under the novel conditions imposed by modern constellations.

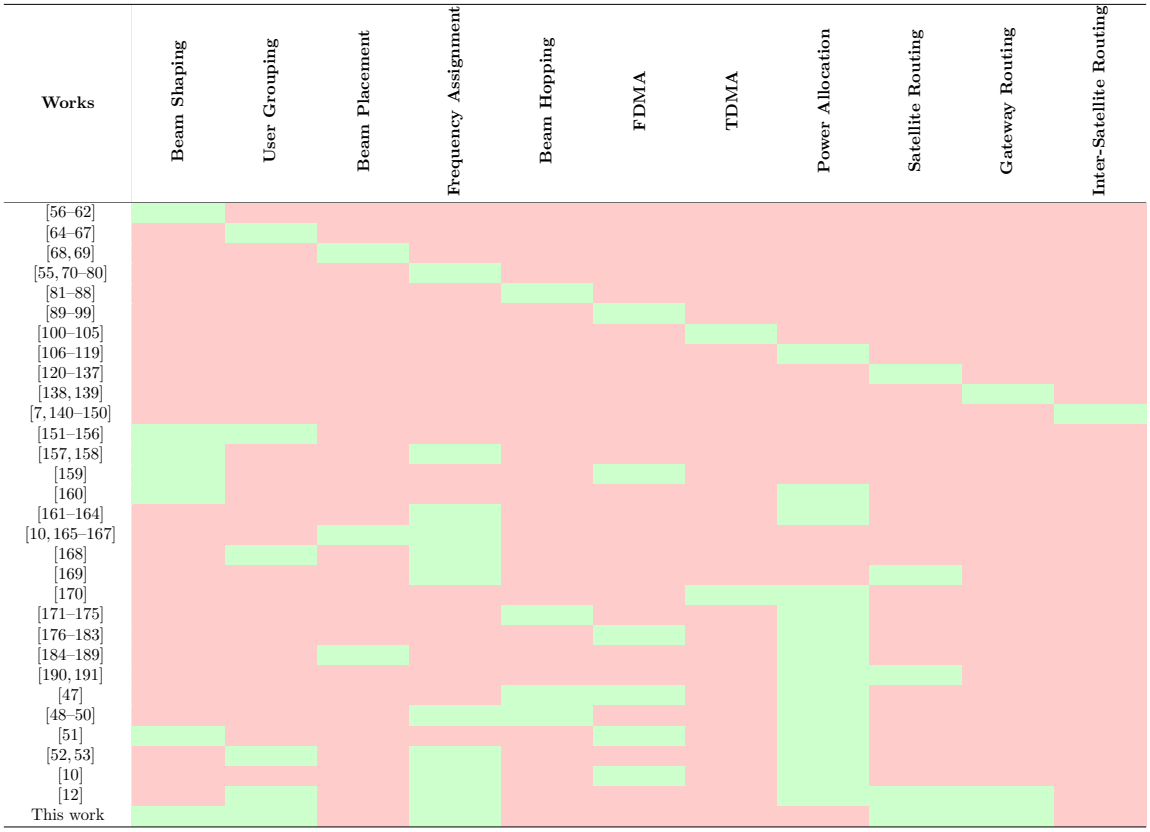


Table 2.2: Summary of the literature's scope. Table adapted from [12].

2.4 Megaconstellation analysis

Following the recent wave of proposals for satellite megaconstellations (i.e., NGSO constellations with thousands of satellites), several studies have arised investigating the impact and performance of the new systems. Many stakeholders rely on this knowledge to cast informed decisions: governmental regulators need to understand how new operating the new architectures will affect the established constellations, investors need to understand what is the expected performance of new systems to leverage possible benefits and risks, and satellite operators need to understand how the

different constellations will interact between them, and how to make best use of the spacecraft.

Following these necessities, del Portillo [7] and the author [8] compare the performance of OneWeb, Telesat, SpaceX, and Amazon in terms of constellations throughput and satellite utilization. Specifically, they assume user demand based on population density, optimized the gateway positions, and Monte-Carlo simulations to address weather effects and compute the total expected throughput on each bottleneck link. These studies show that the next generation of constellations is expected to reach the tens of Tbps in capacity, with Amazon reaching around 50 Tbps. Note that these models are ideal and assume perfect utilization of on-board resources. From a more abstract perspective, Lin [192] proposes a model to analyze satellite utilization in megaconstellations in the presence of ISLs. In this work, the authors discuss the importance of a good resource allocation method, and show how traditional heuristic techniques imply <10% satellite utilization.

Future megaconstellation design has also received the attention of many researchers. Deng [193] proposes an optimization approach to minimize the number of satellites while guaranteeing a minimum coverage. Specifically, this work describes a heuristic procedure to select a sub-set of satellites, starting from a polar constellation, which yields the desired coverage. Jia [194] analyzes different constellation sizes from an interference perspective. In this work, the authors conclude that there exists a threshold after which increasing the constellation size results in less capacity due to inter-satellite interference. This threshold is believed to be located between 10,000 and 30,000 satellites. Vidal [195] studies the effect of different architectural decisions on the constellation performance. In particular, the authors discuss the performance of bent-pipe, beam steering, and beam hopping architectures in terms of power consumption and total mass, and conclude that beam hopping payloads achieve the best payload with minimal reduction in performance compared to beam steering techniques, and similar mass compared to bent-pipe systems. It is important to highlight that the authors only consider optimization approaches for the Beam Hopping sub-problem, which could influence the results. Okati [196] proposes a model based on stochastic geometry to analyze the performance of satellite constellations without the need for simulations. By comparing their model to traditional simulations, the authors show that the outcome of the model matches almost perfectly the outcome of the simulations, with significant decrease in computation time. Ouyang [197] studies the robustness of megaconstellations against geographical failures, concluding that the large amount of satellites combined with ISLs allow for a very low drop-off rate in case of failure in these systems.

Finally, besides technical characteristics of the constellations, other works study other implications of the new systems. Del Portillo [198] analyzes the evolution of the satellite communication market over the next years and estimates which types of constellations are more likely to succeed. Specifically, MEO and GSO constellations have the highest likelihood of profitability, although smart LEO designs could be competitive. From a design perspective, Li [199] examines the design search space and surveys over 33,000 configurations to determine the different trade-offs between the designs. The authors obtain a Pareto-Front of solutions that balance cost and total capacity. Other papers investigate collision avoidance designs (Jia [200]), and the debris impact of modern systems (Sanchez [201]). Finally, other studies also assess the political impact of the proposed constellations (Bhatia [202]).

Summary Several studies analyze the performance and impact of the next generation of satellite constellations. Furthermore, some works discuss orbital and payload design decisions, which can help operators when developing future architectures. However, there is limited research using realistic RAP frameworks for the satellite simulations, which will be a crucial factor in determining the performance of the satellite constellations. Furthermore, several design parameters remain unaddressed: no research studies the impact of using hybrid constellations at different altitudes, and no research studies analyze the impact of different ground infrastructure and ISLs configurations.

2.5 Research gap

Based on the literature analysis, the following gaps are identified:

- There is limited research on the problem of variable footprints in the context of hybrid constellations with satellites at different altitudes.
- There is limited research on beam-to-beam and satellite-to-satellite interference under multi-plane multi-altitude NGSO constellations.
- There is limited research regarding solutions for fixed or flexible beam-gateway associations with ISLs. Furthermore, when on-board networking is present, the satellite-gateway mapping as a variable to optimize is yet to be considered.

- In the context of megaconstellations, there is limited research using realistic RAP frameworks constellations performance analysis.
- There is limited research on the impact of using hybrid constellations with satellites at different altitudes.
- There is limited research on the impact of different ISLs configurations.
- There is limited research on the impact of different kind of users on the network.

The gaps in the current literature arise primarily from the rapid development and deployment of new technologies and architectures. The introduction of highly flexible payloads, along with a significant increase in the number of satellites in the last decade, has led to more complex operations. While ongoing research is making progress in addressing aspects some of these new complexities, the highlighted gaps remain unexplored.

2.6 Thesis Statement

The objective of this dissertation is **to** aid satellite operators in designing and managing the next generation of communications constellations **by**:

- Developing novel RAP methodologies for the Beam Shaping, User Grouping, Satellite Routing, Frequency Assignment, and Gateway Routing tailored to hybrid megaconstellations, aimed at maximizing capacity.
- Studying the performance of existing megaconstellation proposals under realistic operational scenarios.
- Studying the design of LEO-MEO megaconstellations, particularly on the optimal proportion across altitudes.

Using mathematical optimization and realistic simulation environments.

2.7 Chapter summary and conclusions

This chapter has detailed the relevant literature of this dissertation, focusing on the individual and joint techniques to address the RAP, as well as studies regarding the design of megaconstellations.

Additionally, this chapter has highlighted the research gaps and contributions of this dissertation.

2.7.1 Chapter Summary

The first part of this chapter concentrated on methodologies to address the RAP. Section 2.1 provided an overview of existing methods aimed at tackling at least three sub-problems within the RAP, emphasizing their limitations in addressing novel constellation designs. Building on this, Section 2.2 examined existing approaches focused on addressing one or two sub-problems within the RAP, while also identifying areas where current methodologies fall short in accommodating new designs. Subsequently, Section 2.3 summarized existing literature, categorizing them according to the sub-problems they solve.

The second part of this chapter delved into studies analyzing the designs of megaconstellations, as outlined in Section 2.4. Drawing from the literature review, Section 2.5 underscored the existing research gap. Finally, Section 2.6 delineated the contributions of this dissertation, primarily revolving around megaconstellation design and operation.

2.7.2 Response to Research Questions

Research question 2.1

What flexibilities are missing on current automatic RAP frameworks to be able to address the complexities of the novel satellite constellation operations?

Existing RAP frameworks are lacking in two main areas: 1) They are not adapted for hybrid constellations, where satellites might exist at different altitudes, 2) They are not adapted for large constellations, where users and gateways might see multiple satellite simultaneously, and satellites might interfere with each other. These gaps come mainly from the rapid development of satellite constellation design over the past year, which has focused on a significant space segment increase.

Research question 2.2

What critical data is lacking for satellite operators to strategically plan and design the upcoming satellite communications infrastructure?

The information missing for satellite operators mainly evolve from the rapid changes in satellite payload and constellation design over the past years. Studies on the following aspects are lacking:

1) Studies on the design and performance of hybrid constellations, and 2) Studies on the design and impact of different ground infrastructure and ISLs architectures on service quality and performance.

2.7.3 Specific chapter contributions

The specific contributions of this chapter are as follows:

- Defined the research gap in existing RAP methods to address operations in large megaconstellations.
- Defined the research gap in existing studies regarding several aspects of megaconstellation design.
- Described the contributions of this dissertation.

Chapter 3

Simulation environment

Recreating realistic operational conditions in simulations does not only imply developing an efficient Resource Allocation framework that represents the logic of satellite operators, but also appropriately modelling all the adjacent physical phenomena that characterize the system. The objective of this chapter is to detail the surrounding models necessary to simulate realistic operations, as well as reasonable data sources to reproduce these models.

The research questions that this chapter aims to address are:

Research question 3.1

What are the necessary models, in addition to the RAP framework, to recreate realistic operational conditions regarding satellite communications constellations?

Hypothesis: The necessary models will be the ones that determine the physical characteristics around operations. It will likely involve a characterization of the antennas, both in ground and space, and a characterization of the link quality, related to signal loss from several sources.

Research question 3.2

What are appropriate sources of data to simulate these models?

Hypothesis: For antenna characterization, ideal data sources would be obtained directly from satellite operators. If no public information can be found, estimations based on historical data are likely close to the real values. For link quality characterization, ideal data sources would come

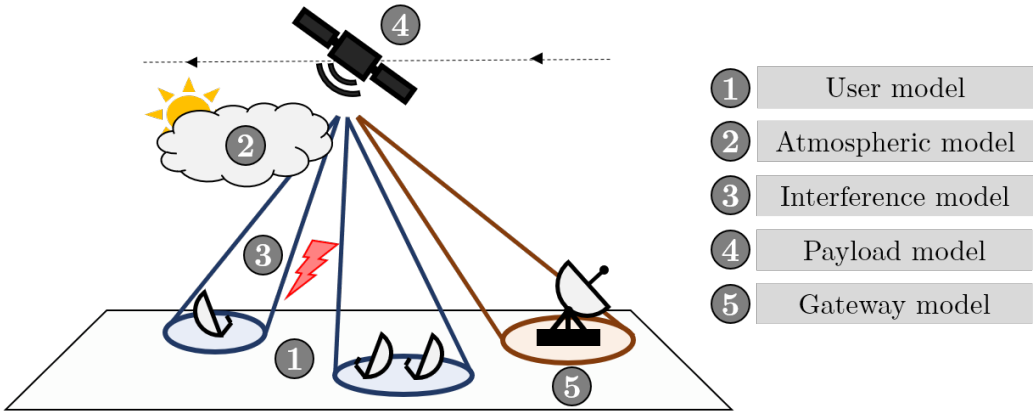


Figure 3-1: Illustration of the five physical models necessary to recreate realistic operational conditions

from real operations. Since most systems are still under development, standard communication link models are likely close to the real values.

3.1 Model identification

The first step to use Internet services via satellite communications involves a **user** transmitting a signal to a satellite through a ground terminal. The signals strength and quality are subject to **atmospheric** conditions and potential **interference** from other sources. Upon reaching the satellite, the signal undergoes reception, transformation, and re-transmission using the satellites **payload**. The redirected signal then reaches a **gateway** responsible for processing the users request and retrieving the desired information from the internet. Subsequently, the information needs to be transmitted back to the user through a similar mechanism. To comprehensively model and understand the communication process, five essential models influencing the outcome are identified (see Figure 3-1):

1. User model: This model defines a strategy to determine the position, demand, and antenna characteristics of users that is representative of modern operations.

2. Atmospheric model: This model formulates a strategy to determine the impact of atmospheric conditions at each point on the Earth surface.

3. Interference model: This model outlines a strategy to determine the existence of interference between signals.

4. Satellite payload model: This model encapsulates the hardware capabilities of the satellites involved in the communication process.

5. Gateway model: This model establishes a strategy for determining the placement and antenna characteristics of gateways, acting as a portal to the Internet.

Note that these five models correspond to physical entities or interactions within the realm of satellite communications, distinct from the decision-making process emphasized in the RAP.

3.2 User model

To simulate realistic operational conditions, a model that encompasses information about the position, demand, and antenna characteristics of each user is needed. Given the necessity for distributions with varying numbers of users, a flexible strategy is required to adjust the model size. The subsequent explanation outlines the approach employed to obtain the necessary information for generating user models of diverse sizes.

3.2.1 User distribution

Concerning **user positioning**, the positions are determined based on population distribution data [203]. Specifically, a public database featuring a world grid with a resolution of 0.1° is utilized. Each grid cell provides information about the population within that cell. When a new user position is required, the strategy selects a random cell based on roulette wheel selection process, where the probability of selecting each cell equal to the percentage of population in the cell. This method enables the generation of any desired number of user positions, facilitating the creation of user models with varying sizes. As an illustrative example, Figure 3-2 displays a random sample of 20,000 points chosen using the described method.

However, employing population distribution as a reference source presents three primary drawbacks:

1. This reference exclusively focuses on residential customers, neglecting other user segments such as aviation or maritime sectors. Additionally, it overlooks the fact that densely populated

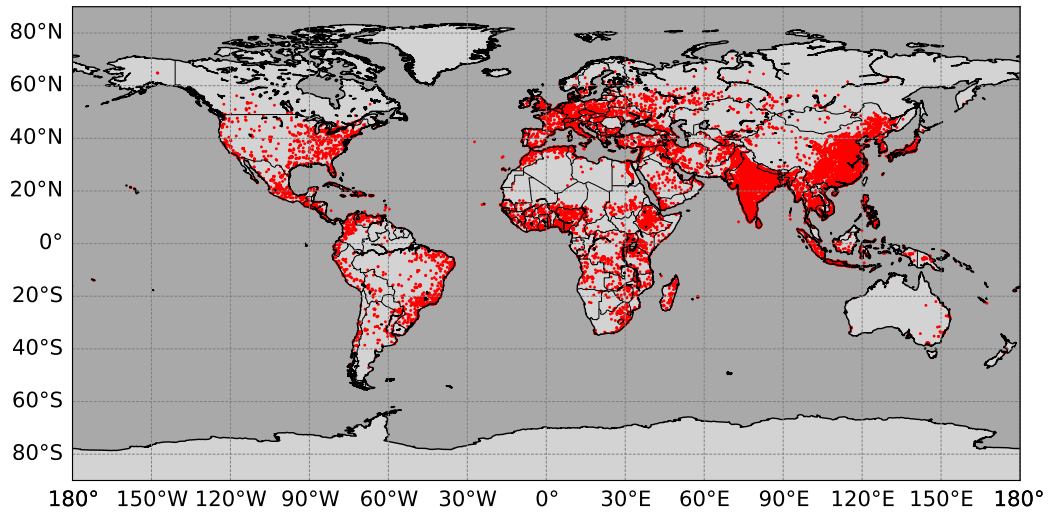


Figure 3-2: Random sample of 20,000 points using the wheel selection on the world population distribution. Each user location is indicated using a red dot.

areas are likely to have alternative terrestrial communication networks, potentially diminishing the need for satellite solutions. Finally, it does not consider economic power or political factors within the distribution, which might drive operators into specific regions more than others. Nevertheless, interviews with key players in the satellite communications market reveal that representing users based on population distribution has historically proven effective in estimating the general behavior of satellite constellations. Furthermore, recent works [7, 8, 192], leverage user models based on population distribution to create realistic simulation environments for modern constellations.

2. These models exhibit lower accuracy when compared to local demand models [10]. Nevertheless, given the strong global component of modern constellations, the use of local simulation tools becomes impractical. Utilizing population distribution as a reference enables the generation of user models that remain consistent globally.

Concerning user demand, each user is assigned a demand of 100 Mbps, aligning with the offerings of satellite operators [204, 205], as well as the objectives outlined by governmental agencies [206]. Note that communications systems tend to be over-subscribed by a factor between 10-200 [207] to account for the fact that users are not consuming demand continuously. As a simplification, this

Satellite Operator	Equivalent antenna diameter [m]	Source
OneWeb	1	Public filing [22]
SpaceX	0.6	Product sheets [204]
Boeing	1	Estimation*
Amazon	0.7	Public filing [4]
Telesat	1	Public filing [31]
Viasat	0.6	Public filing [33]
SES-O3b	1	Estimation*
CASC	1	Estimation*
Intelsat	1	Estimation*

Table 3.1: Antenna sizes for each satellite operator. The values given correspond to the diameter of an equivalent parabolic antenna with the same area as specified in the information source. Values marked with * correspond to estimations by the author based on similarities with other systems.

dissertation assumes that each user is consuming 100 Mbps continuously with no over-cubscription factor. While recognizing the diversity of products offered by different companies, employing a standardized user demand ensures a fair performance comparison across constellations.

3.2.2 User antenna

Concerning **user antenna characteristics**, these characteristics are contingent upon the specific satellite constellation. Notably, the primary distinction among operators lies in the size of the antenna and performance aspects directly linked to size, including gain and gain over temperature. A larger antenna facilitates higher link quality but comes with increased costs and inconvenience for the client. Conversely, a smaller antenna offers practicality and reduced cost but at the expense of poorer link quality. As a general guideline, constellations operating in LEO typically employ smaller antennas, as the proximity to Earth negates the need for high antenna gains. In contrast, constellations operating in MEO or High Earth Orbit (HEO) opt for larger antennas to compensate for the increased distance to the satellite. The specific dimensions of user antennas for each constellation are summarized in Table 3.1. Other antenna characteristics, such as efficiency and accuracy losses, are assumed to be uniform across all operators and are presented in Table 3.2.

Notably, in certain scenarios, it is interesting to simulate users with higher demand levels. This approach proves beneficial when modeling clusters of customers situated in close proximity or users outside the residential sector, such as those on oil platforms or local Internet providers. To replicate these scenarios, instead of generating individual users, a collective entity of $N_{us/loc}$ users

Parameter	Value	Parameter	Value
Transmission Power	20 dBW	G/T	25 dB
Antenna efficiency	0.65	LNB noise figure	2 dB
Pointing loss	1 dB	Feeder loss	1.1 dB
Waveguide loss	0.2 dB	Additional loss	1 dB

Table 3.2: Operator-independent user antenna parameters. LNB: Low Noise Block converter

for a specific location is established. All $N_{us/loc}$ users operate as a unified entity, yet each possesses the demand and antenna characteristics akin to $N_{us/loc}$ distinct antennas. The specific value of $N_{us/loc}$ is contingent upon the experiment and will be explicitly delineated for each simulation. In instances where it is not explicitly stated, it is assumed that only individual residential customers are considered (i.e., $N_{us/loc} = 1$).

3.3 Atmospheric model

To assess signal quality under realistic conditions, a comprehensive atmospheric model that accounts for potential signal attenuation due to diverse atmospheric factors is essential. In this dissertation, the adopted model is from [208], which adheres to the ITU-R P.618-13 standard and incorporates considerations for gaseous, clouds, tropospheric scintillation, and rain impairments. This model takes the following parameters as **input**:

- Geographical characteristics of the ground antenna: longitude, latitude, and altitude. In cases where altitude is not provided, an estimate can be derived using ITU-R P.1511 guidelines [209].
- Environmental characteristics of the location: temperature, pressure, humidity, vapor density, and rain conditions. In instances where specific values are absent, temperature estimates can be derived from ITU-R P.1510 [210], pressure and humidity estimates can be obtained from ITU-R P.453 [211], vapor density estimation can be acquired from ITU-R P.836 [212], and rain conditions can be approximated using ITU-R P.837 [213].
- Physical characteristics of the communications link: frequency, antenna diameter, and elevation angle.
- Percentage of availability required. This parameter signifies the percentage of time during which the actual attenuation must be lower than the specified thresholds.

With this, this module provides the following **output**:

- Rain and scintillation attenuation, following ITU-R P.618 [214].
- Gaseous attenuation, following ITU-R P.676 [215].
- Cloud attenuation, following ITU-R P.840 [216].

3.4 Interference model

In addition to the losses due to attenuation, a model to determine the losses due to interference between signals is needed. While interference was not a limiting factors in earlier constellation designs due to the large separation between satellites, and between beams within a satellite, it plays a crucial role in determining the feasibility of operations in modern systems.

3.4.1 Determining the existence of interference

Interference between two signals depends primarily on two factors: 1) the relative strength of each signal, and 2) the portion of the spectrum they utilize along with their polarization. Furthermore, the relative strength is a consequence of the power at the transmission of each signal, the atmospheric losses of each signal, and the relative position of the transmitters and receivers. Note that many of the decisions within the RAP have a direct influence on the interference. In particular, the Frequency Assignment sub-problem affects the frequency spectrum and polarization of each signal, the Satellite Routing and Gateway Routing sub-problems affect the relative position of the antennas, and all sub-problems contribute to the power required at the transmission antenna to balance the link budget. Given that certain RAP decisions, like TDMA, FDMA, and Power Allocation, depend on real-time considerations, such as the atmospheric attenuation, determining the existence of interference **cannot** be concluded before operations commence.

Nevertheless, as highlighted in Chapter 2, a key objective of this dissertation is to formulate a methodology for addressing interference during the resource allocation process, i.e., pre-operations. Consequently, the methodology cannot involve an exact interference computation. To reconcile this challenge, the following observation is underscored: while the existence of interference can only be ascertained during operations, there are specific conditions that, if met, indicate that two signals do **not** interfere. Two signals will not interfere under the following circumstances:

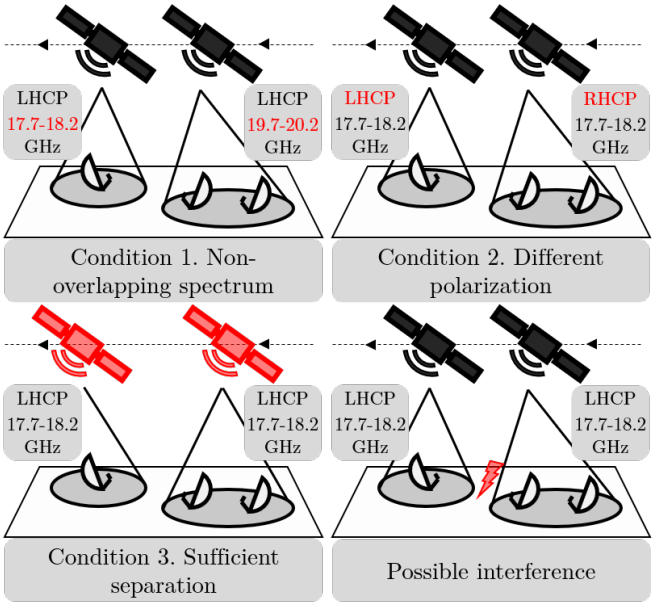


Figure 3-3: Conditions to mitigate possible interference

1. They occupy non-overlapping portions of the frequency spectrum.
2. They have opposite polarization.
3. They are *sufficiently* geographically separated.

Figure 3-3 provides a visual representation of the conditions influencing interference. Note that the first two conditions can be exclusively determined as part of the Frequency Assignment sub-problem, while the last one relies on the Beam Shaping, Satellite Routing, and Gateway Routing sub-problems. As highlighted in Section 1.2.3, all these decisions must be resolved prior to operations, eliminating the reliance on real-time operations. Consequently, two signals will **not** interfere if at least one of the specified conditions is active, and *potential interference* arises when neither condition is met. To assess interference impact in the results, a conservative approach is adopted: if two signals have the potential to interfere, the results are computed under the assumption that they **do** interfere.

Identifying instances where two signals occupy non-overlapping frequency spectrum portions involves checking if the frequency bands occupied by each signal overlap. Determining when two signals exhibit opposite polarization is a binary comparison between two values. Both of these con-

ditions can be easily incorporated into the Frequency Assignment formulation and resolution [55]. However, determining when two signals are sufficiently separated poses a greater challenge. To address this, potential interference is categorized into two complementary subtypes: intra-satellite interference, occurring when two signals are transmitted or received within the same satellite (CABI), and inter-satellite interference, occurring when two signals are transmitted or received in different satellites (CASI).

3.4.2 Addressing inter-satellite interference

Recent literature has presented methodologies for estimating when two signals are adequately separated in the context of intra-satellite interference. Notably, prior research has employed the concept of *angular separation* between beams [55, 168]. In essence, two beams are considered sufficiently separated if the angular distance between their main directions exceeds a predefined threshold. Note that this definition exclusively addresses intra-satellite interference and does not tackle the challenge of inter-satellite interference, where both signals might have different sources and destinations.

Inter-satellite interference can be reasonably neglected when the constellation comprises a low number of satellites, and the separation between them is substantial. However, this premise no longer holds in light of the current industry trend favoring constellation designs with thousands of satellites. Given these novel conditions, there arises a necessity for new models capable of incorporating inter-satellite interference. To address this, this dissertation uses the concept of *isolation* (refer to Figure 3-4), which generalizes the concept of angular distance to encompass inter-satellite interference.

In particular, instead of being contingent on the angle, the isolation of signal S_2 at the reception point of signal S_1 is computed. The isolation serves as a measure of a signal strength relative to another, excluding power or atmospheric considerations. Similar to the angular distance concept, two signals are considered sufficiently separated if their isolation surpasses a predefined threshold I_{thres}; they might interfere if the isolation falls below this threshold. It is worth noting that the concept of isolation is analogous to angular distance when both signals share the same source. Note than the concept of *splash zone*, indicating the area around a footprint in which signals might interfere, is now dependent on the origin of the signals and, thereby, variable. Therefore, using this concept to tackle inter-satellite interference becomes impractical.

Formally defining the isolation involves introducing variables r_1 and $G_1^R(p)$, representing the

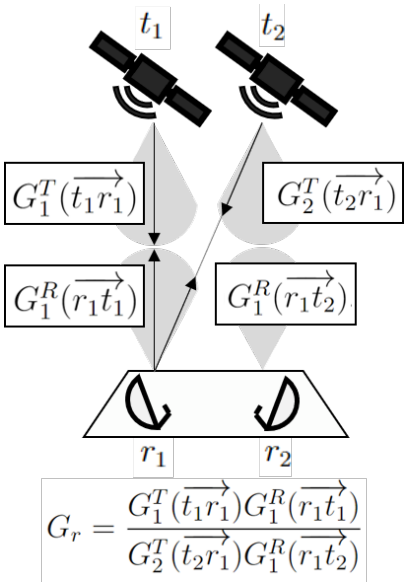


Figure 3-4: Definition of isolation. The isolation is used to evaluate when two beams have sufficient separation (condition 3).

position and gain at direction p of the receptor of signal S_1, respectively. Similarly, t_1 and $G_1^T(p)$ denote the position and main direction gain of the transmitter of signal S_1, respectively. The added gain from signal S_1 at r_1 is the sum of gains (in dB): $G_1 = G_1^T(\overrightarrow{t_1r_1}) + G_1^R(\overrightarrow{r_1t_1})$. Corresponding variables r_2, $G_2^R(p)$, t_2, and $G_2^T(p)$ are defined for S_2. The gain of S_2 at the reception point of S_1 is computed as the gain of the receiver of S_1 in the direction marked by p_{t_2} plus the gain of the transmitter of S_2 in the direction marked by r_1 (in dB): $G_2 = G_1^R(\overrightarrow{r_1t_2}) + G_2^T(\overrightarrow{t_2r_1})$. This gain is equal to or lower than the main direction gain, contingent on the relative direction and the radiation pattern of the beam. The isolation is then computed as the subtraction of both values (in dB): $G_{rel} = G_1 - G_2$. If $G_{rel} \leq I_{thres}$, the signals are adequately separated and do not interfere; if $G_{rel} > I_{thres}$, interference is plausible. The calculation of G_r is illustrated in Figure 3-4. For the purpose of this work, circular beams with fixed aperture angles are employed. The radiation patterns utilized align with the ITU recommendations: ITU-R S.465 for Earth station antennas [217] and ITU-R S.1528 for non-geostationary orbit satellite antennas [218].

3.5 Satellite payload model

To determine how the satellite processes the data transmitted by users or gateways, an accurate satellite payload model encompassing currently deployed technologies is indispensable. The satellite payload model is divided into antenna characteristics and requirements, along with on-board processing capabilities.

3.5.1 Satellite antenna model

Concerning the satellite antenna model, the majority of deployed constellations opt for phased array antennas. These antennas enable operators to electronically steer beams without the need for mechanical movement and may facilitate different software-defined beam shapes. Given that phased array antennas are now the prevailing choice on satellites, all simulations will exclusively consider this type of antenna. While these antennas permit arbitrary beam shapes, this work will focus on circular beams with a fixed aperture angle. The specific aperture angle will be contingent upon the simulation parameters and will be elucidated in each simulation details. Despite a consensus among operators on the type of antenna, variations in antenna characteristics persist across different constellations. Table 3.3 compiles the satellite antenna characteristics for each considered constellation, based on publicly available filings. When servicing a specific beam, it is assumed that the maximum beam power is linear with the bandwidth allocated to the beam, where the linearity factor is the Effective Isotropic Radiation Power (EIRP) density outlined in Table 3.3. It is assumed that satellites can produce sufficient power to serve all beams based on this beam-wise limitation. Not modelling a satellite-wise power limit allows for a simplified Power Allocation resolution, following the link budget equations in Section 1.2.2, with the maximum power limit established above.

3.5.2 On-board processing model

Regarding the on-board processing model, two primary architectures prevail: bent-pipe and on-board processing. In a bent-pipe architecture, the satellite essentially mirrors the received signal, transmitting it without modifications in MODCOD or bandwidth (though the central frequency may change). Conversely, in an architecture with on-board processing capabilities, the signal undergoes decoding, potential aggregation, and subsequent encoding for transmission. This architecture

Operator	Altitude	Link	Transmission				Reception				Ref.
			Gain [dB]	EIRP per Hz [dBW/Hz]	Pointing Loss [dB]	Rotation Loss [dB]	Gain [dB]	G/T [dB/K]	Pointing Loss [dB]	Rotation Loss [dB]	
OneWeb	-	User GW	37 33	-49.4 -52	0.4 0.6	1	38.6 38	11.6 11.4	0.4 0.6	1	[24]
SpaceX	-	User GW	34 34.5	-51.1 -54.3	0.1	0.1	35.7 38.5	8.4 11.5	0.1	0.1	[28]
Boeing	LEO MEO HEO	- - -	39.9 49.59 55.9	-31.3 -17.8 -4.3	0.03	0	39.96 49.69 55.9	9.26 19 25.3	0.03	0	[30]
Amazon	-	User GW	39 36.9	-43.9 -51	0.1	0.1	39 40.7	12.4 14.1	0.1	0.1	[4]
Telesat	-	User GW	32.5 32	-50 -50	0.1	0	35 31.8	4.5 1	0.1	0	[32]
Viasat	-	-	52.4	-43.7	0.1	0.1	51	24	0.1	0.1	[6]
SES-O3b	-	-	44.25	-34.28	0.6	0.5	44.25	7	0.6	0.5	[35]
CASC	-	-	38.6	-49.15	0.1	0.1	38.6	10	0.1	0.1	[37]
Intelsat	-	User GW	42.6 42.1	-39.5	0.1	0.1	44.00	14.2	0.1	0.1	[38]

Table 3.3: Satellite antenna parameters for each satellite operator. Values extracted from public filings. GW: Gateway

decouples uplink and downlink, allowing flexibility in bandwidth and MODCOD changes. In such a setup, the satellite functions akin to a terrestrial Internet node, and the constellation operates as a network of interconnected nodes. As a general trend, LEO constellations commonly employ architectures with on-board processing capabilities, whereas MEO and HEO designs tend to favor bent-pipe architectures.

Given the prevailing shift towards more advanced satellites, this dissertation exclusively considers satellites equipped with on-board processing capabilities. Specifically, it is assumed that upon reaching the satellite, the signal is decoded into its original data. If the constellation incorporates inter-satellite links, the data may traverse these links to the destination, potentially being fragmented based on package size (akin to terrestrial Internet connections). Upon reaching the destination satellite, the data is re-aggregated into a single stream for each beam, encoded using adaptive modulation and coding schemes, and then transmitted. This aligns with contemporary trends that view the communication network as a "network of networks" [219, 220]. It is assumed that each satellite can modify the frequency, bandwidth, and MODCOD of each signal prior to

transmission, as well as deciding the following node in the network. The set of MODCODs used in this dissertation follows the standard DBV-S2X [221]. It is assumed that the routing policy is controlled via telemetry and telecommand, and can be changed over time, albeit not at a higher rate than the telecommand allows for. It is assumed that bandwidth is channelized into channels of up to 250 MHz. The last channel in a specific frequency range might be smaller than 250 MHz to fit the frequencies specified in Table 3.4.

For simplicity, this dissertation assumes that the satellite capabilities are fully dedicated to communication applications. Exploring other simultaneous applications, such as orbital edge computing or power beaming, lies beyond the scope of this study. Beyond on-board computation, Table 3.4 outlines the hardware capabilities of the satellites for each operator.

3.6 Gateway model

To simulate realistic operational scenarios, it is essential to incorporate a representative gateway model that includes information about the location and antenna characteristics of ground stations. In the context of this dissertation, a distinction is made between ground stations and gateways. A ground station refers to a location on the Earth surface that furnishes reliable and high-speed Internet connectivity to the constellation. On the other hand, a gateway is an antenna situated within a ground station that establishes connections with satellites in the constellation. A ground station can host multiple gateways, but there is a physical limitation on the number of antennas at a single location due to interference considerations. Based on available public information, Table 3.5 provides a summary of gateway antenna characteristics organized by operator.

To furnish Internet connectivity to users, satellite operators must make critical decisions not only about the space segment but also regarding the locations of ground stations and the number of gateways to install at each station. This complex decision-making process is commonly known as the *Gateway Placement* problem. When tackling this problem, operators must consider technical factors, such as the relative position of the station with respect to the satellites and the technical capabilities of the antennas, as well as logistical and regulatory considerations like available Internet speed, terrain availability, and atmospheric conditions. It is important to note that this problem is intricately linked to the constellation design rather than its day-to-day operation, and as such, it is not incorporated within the RAP.

Operator	General		User			GW		
	Num. of GW antennas	Frequency reuse factor	MEA [°]	Uplink freq. [GHz]	Downlink freq. [GHz]	MEA [°]	Uplink freq. [GHz]	Downlink freq. [GHz]
OneWeb	1	4	50*	12.75 - 13.25 14 - 14.5	10.7 - 12.7	20*	27.5 - 30	17.8 - 18.6 18.8 - 20.2
SpaceX	1	4*	25	12.75 - 13.25 14 - 14.5	10.7 - 12.7	25	14 - 14.5 27.5 - 29.1 29.5 - 30	10.7 - 12.7 17.8 - 18.55 18.8 - 19.3 19.7 - 20.2
Boeing	†	10*	35	48.2 - 50.2	40 - 42	5	47.2 - 50.2 50.4 - 52.4	37.5 - 42
Amazon	2	3	35	28.35 - 29.1 29.5 - 30	17.7 - 18.6 18.8 - 19.4 19.7 - 20.2	20	27.5 - 30	17.7 - 18.6 18.8 - 20.2
Telesat	2	4	10	27.5 - 29.1 29.5 - 30	17.8 - 18.6 18.8 - 19.3 19.7 - 20.2	10	27.5 - 29.1 29.5 - 30	17.8 - 18.6 18.8 - 19.3 19.7 - 20.2
Viasat	†	3*	25	27.5 - 29.1 29.5 - 30 47.2 - 50.2 50.4 - 52.4	17.8 - 18.6 18.8 - 20.2 37.5 - 42	25	27.5 - 29.1 29.5 - 30 47.2 - 50.2 50.4 - 52.4	17.8 - 18.6 18.8 - 20.2 37.5 - 42
SES-O3b	†	10*	25	27.5 - 30	17.8 - 18.6 18.8 - 20.2	10	27.5 - 30	17.8 - 18.6 18.8 - 20.2
CASC	2*	4*	50*	12.75 - 13.25 14 - 14.5 27.5 - 30 47.2 - 50.2 50.4 - 51.4	10.7 - 11.7 12.5 - 12.75 17.7 - 18.6 18.8 - 20.2 37.5 - 42.5	25*	12.75 - 13.25 14 - 14.5 27.5 - 30 47.2 - 50.2 50.4 - 51.4	10.7 - 11.7 12.5 - 12.75 17.7 - 18.6 18.8 - 20.2 37.5 - 42.5
Intelsat	3	10*	25	12.5 - 13.25 14.0 - 14.5 48.2 - 50.2	10.7 - 12.7 37.5 - 42.0	35	47.2 - 50.2 50.4 - 52.4	37.5 - 42

Table 3.4: Satellite hardware capabilities for each satellite operator. (*) Values estimated by the author, the rest are extracted from public information. † Antennas are shared between users and gateways. GW: Gateway, MEA: Minimum Elevation Angle

Operator	Diameter [m]	Gain [dB]	EIRP density [dB/Hz]	Source
OneWeb	3.5	58	-13	[222]
SpaceX	1.85	52.6	-24.2	[223, 224]
Boeing	4.5	60	-24.7	Estimated
Amazon	2.4	53.8	-20.2	[225]
Telesat	4	57.1	-3.1	Size from [31], rest estimated
Viasat	2.4	53.1	-25.8	[226]
SES-O3b	5.5	62.6	-10.9	Advised from SES
CASC	1.5	49.5	-26.3	Estimated
Intelsat	4.5	60	-24.7	Estimated

Table 3.5: Gateway antenna characteristics for each operator

While this dissertation does **not** aim to solve the Gateway Placement problem, it necessitates realistic ground station placements to accurately recreate operational conditions within the constellation. To address this need, existing solutions in the literature are explored. The Gateway Placement problem can be subdivided into two smaller sub-problems: 1) Selecting locations for ground stations, and 2) Determining the number of antennas to install at each location. Most existing literature has primarily focused on the first sub-problem, as it carries the most substantial economic impact: establishing a new location on the globe is significantly more costly than installing a new antenna in an existing location.

3.6.1 Literature on the selection of ground station locations

To tackle the challenge of selecting ground station locations, existing methodologies suggest initiating from a predetermined list of potential locations provided by the operator [7, 227–229]. This list serves as a pre-selection of favorable locations meeting various requirements, including regulatory and logistical constraints, alongside performance considerations. Utilizing graph-based formulations and down-selecting algorithms, different authors propose diverse mechanisms to obtain a subset of ground stations with desired properties. For instance, Cao [228] formulates the problem as a flow-based one, aiming to maximize reliability, and employs particle swarm optimization to solve it. Torkzaban [227], using a similar formulation, presents a MILP implementation to minimize the cost of deploying ground infrastructure, subject to performance constraints. While both studies demonstrate promising results, two main drawbacks impede their practical implementation: 1) They provide no guidance on obtaining the initial set of ground station locations, as they select benchmarks from the graph domain, and 2) Their benchmarks consider only up to 53 possible locations, while prior work suggests that modern systems may rely on hundreds of ground stations to optimize performance [8].

Chen [229] suggests gridding the world and treating each cell as a potential ground station location, assuming that each cell contains a realistic, valid location. They employ a simulation-based approach, simulating the behavior of the constellation, including routing mechanisms and ISLs. Using GA, they down-select to grid cells that offer optimal performance. While the authors illustrate how a constellation like SpaceX Starlink can be optimized, the main drawback is the necessity of knowing routing and ISLs in advance. Since routing is part of the RAP, and therefore partially addressed during operations, the solution may not be known in the design phase. Finally,

del Portillo [7] proposes starting from existing ground station locations based on historical data and down-selecting to a set that provides the highest expected throughput based on constellation load. For the down-selecting algorithm, the authors use GA to obtain a Pareto-Front of solutions with different trade-offs between the number of locations and total capacity.

While the requirements and outcomes of the approach proposed by del Portillo [7] align with the objectives of this dissertation, there are two main considerations when applying this methodology in a general scenario:

1. Satellite Altitudes: The methodology assumes that all satellites in the constellation are positioned at similar altitudes. While this assumption suits certain designs like SpaceX and OneWeb, it may not be suitable for hybrid constellations with satellites in LEO, MEO, and HEO, such as Boeing.

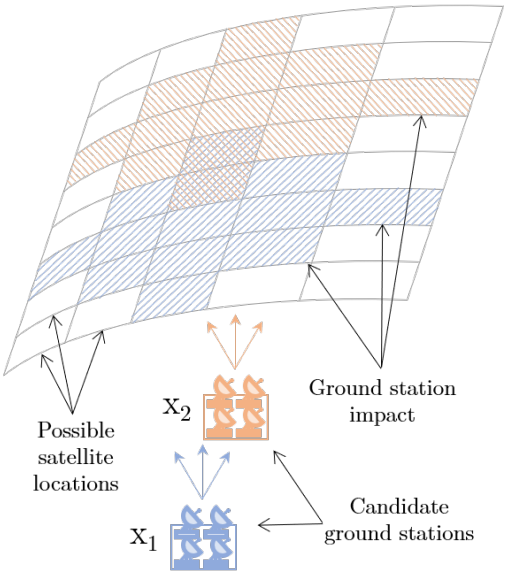
2. Computational Complexity: The authors assert that obtaining the optimal solution is computationally intractable for the required dimensionality, necessitating the use of a suboptimal method. However, given that all constraints and objectives are linear, and recent advancements in mathematical solvers enable the resolution of problems with hundreds of thousands of integer variables and linear constraints, there is a belief that an optimal solution to the problem can be achieved in a reasonable time using a similar formulation.

3.6.2 Addressing the selection of the ground station locations

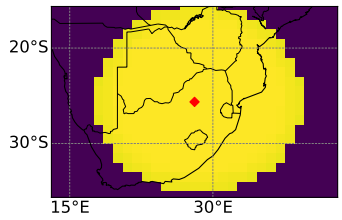
To address the gaps identified in the previous section, the following lines outline how the formulation from [7] has been extended to incorporate hybrid considerations and achieve an optimal solution using mathematical solvers.

Firstly, let $\mathcal{G}$ be defined as the initial set of potential ground station locations, containing geographical information for each site. For this study, historical ground station sites are included in this set, expanded from [230]. The full list of 204 potential locations is provided in Appendix B. Each ground station is assumed to have unlimited capacity. Subsequently, x_g is introduced as a binary variable indicating whether site g has been selected ($x_g = 1$) or not ($x_g = 0$). It is assumed that at most k locations can be chosen, encoded as $\sum_g x_g \leq k$.

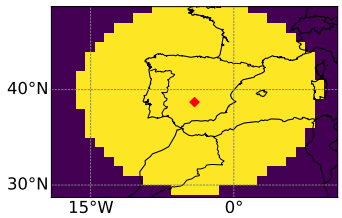
Following this, $\mathcal{U}$ is defined as the set of users that need to be served, encompassing information about the geographical position p_u and demand d_u of each user u. $\mathcal{S}_a$ is defined as the set of



(a) Ground station impact



(b) Ground station in Pretoria, South Africa



(c) Ground station in Puertollano, Spain

Figure 3-5: Impact of selecting a particular ground station, and examples over Pretoria, South Africa, and Puertollano, Spain.

satellites at altitude a. Next, v_p is defined as the minimum number of satellites visible from point p on the Earth surface across all altitudes. Now, the impact of user u over a satellite is defined as:

$$\nu_u = \frac{d_u}{v_p} \tag{3.1}$$

ν_u can be interpreted as the amount of demand expected in a satellite by user u. Next, for each altitude, a grid of $1° \times 1°$ is created. For each cell, the expected demand that will be absorbed if a satellite was present in the cell is computed as:

$$d_{a,p} = \sum_u \mathbb{1}_{LoS(p_u,p,a)} \nu_u \tag{3.2}$$

Where $\mathbb{1}_{LoS(p_u,p,a)}$ represents if user u is visible by a satellite at altitude a at position p, where p represents the position at the center of the cell. The exact definition of $\mathcal{S}_a$, v_p, and $LoS(p_u, p, a)$ can be found in Appendix A.

Subsequently, $c^{ava}_{g,p,a}$ is introduced as the capacity provided by a gateway at ground station g to

a satellite at altitude a and position p with availability ava. This value can be computed based on link budget simulations, utilizing the equations described in Chapter 1. Note that if site g is not visible to a satellite at altitude a and position p, the provided capacity is 0 by definition. Therefore, the impact of selecting a particular ground station is contingent on the link quality at the site, combined with the visible satellite positions, as illustrated in Figure 3-5. Now, $d_{a,p}$ signifies the demand of the users, and $c^{ava}_{g,p,a}$ represents the supply of the ground stations. To combine both values, $r^{ava}_{p,a}$ is defined as the amount of **uncovered** demand in a satellite at altitude a and position p with availability ava:

$$\begin{aligned} r^{ava}_{p,a} &\geq d_{a,p} - \sum_{g} c^{ava}_{g,p,a} x_g \\ r^{ava}_{p,a} &\geq 0 \end{aligned} \tag{3.3}$$

Based on the previous definitions, the objective function is defined similarly to [7]:

$$\min_{x_g} \sum_{p,a} (r^{0.95}_{p,a} + r^{0.99}_{p,a}) \tag{3.4}$$

With this, the complete formulation becomes:

$$\begin{aligned} \min_{x_g} \quad & \sum_{p,a} (r^{0.95}_{p,a} + r^{0.99}_{p,a}) \\ s.t. \quad & r^{0.95}_{p,a} \geq d_{a,p} - \sum_{g} c^{0.95}_{g,p,a} x_g \\ & r^{0.99}_{p,a} \geq d_{a,p} - \sum_{g} c^{0.99}_{g,p,a} x_g \\ & \sum_{g} x_g \leq k \\ & x_g \in \{0,1\}, r^{0.95}_{p,a} \geq 0, r^{0.99}_{p,a} \geq 0 \end{aligned} \tag{3.5}$$

This formulation aims to minimize the uncovered demand across all possible points and altitudes. Note that, in the presence of very high demand, the solution will try to allocate as many ground stations as possible to serve as much demand as it can. Given that it incorporates both binary (x_g) and continuous ($r^{0.95}_{p,a}, r^{0.99}_{p,a}$) variables, it corresponds to a MILP. Such problems fall into the NP-hard category and consequently have an exponential complexity concerning the number of

integer variables. However, due to the relatively low number of integer variables ($|\mathcal{G}| = 164$), this formulation can be optimally solved using commercial mathematical solvers. Note that, unlike [7], the proposed formulation is single-objective. Nonetheless, a similar Pareto-Front can be obtained by solving the formulation multiple times with different values for k.

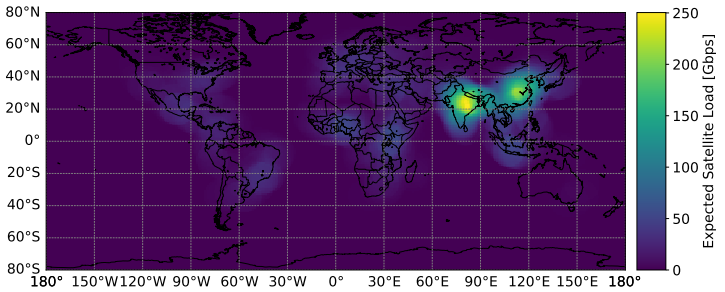
3.6.3 Verification of the ground station selection process

As the methodology for selecting ground station locations proposed in this dissertation differs from [7], it is imperative to verify and validate that the proposed approach achieves the desired objectives. To accomplish this, SpaceX constellation with a user distribution of 20,000 users is simulated using the proposed ground station selection method for $k = \{40, 80, 120, 160\}$. The user and atmospheric models used are the ones explained in Sections 3.2 and 3.3, respectively.

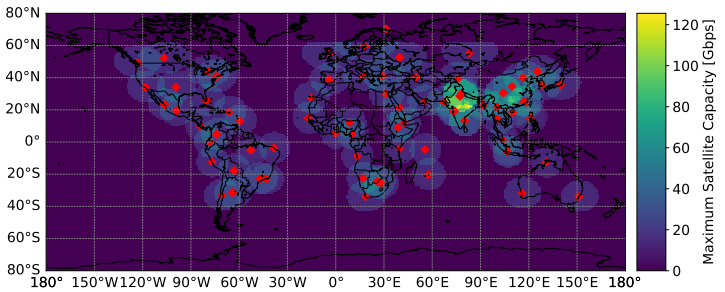
Figure 3-6 illustrates the expected satellite load, satellite maximum capacity, and expected unmet demand based on user demand for $k = 80$. As evident, the supply provided by the selected ground station locations aligns with the expected demand of users. Notably, in densely populated regions like India and China, the provided capacity may fall short of the required demand due to physical limitations induced by interference from nearby stations. Furthermore, Table 3.6 show the unmet demand, optimality gap in the solver, and computation time when solving the formulation based on commercial solvers for different k. As shown, the solver is able to obtain an optimal solution within a few seconds. The results demonstrate that the proposed method effectively selects a subset of ground stations that minimizes uncovered demand, thereby validating the approach.

# of ground sites (k)	Unmet demand [Tbps]	Gap [%]	Total time [s]	# of variables (binary)	# of variables (total)	# of constraints
40	23.954	0	3.506	251	87739	87489
80	13.139	0	9.107			
120	11.894	0	2.821			
160	11.825	0	1.377			

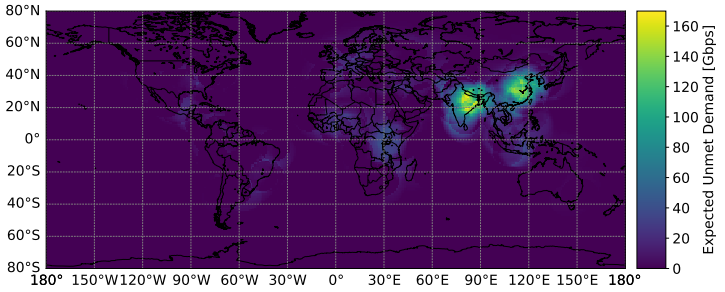
Table 3.6: Summary of results when executing the Gateway Placement MILP model



(a) Expected satellite load



(b) Maximum satellite capacity



(c) Expected unmet demand

Figure 3-6: Expected satellite load, satellite capacity, and unmet demand at each position over the Earth for 20,000 users on the SpaceX constellation with 80 ground stations

3.6.4 Addressing the selection of the number of gateways per ground station

While the previous methodology successfully achieves the desired objective of identifying suitable ground station locations, it does not provide information about the number of gateway antennas per ground station. Up to this point, this aspect has not been explored as an optimization parameter within the expected dimensionality of modern constellations. On one hand, [227, 228] attempt to optimize the location for each individual gateway but using a low-dimensional graph. It remains unclear how their proposed methodology can scale to the dimensionality of modern systems. On the other hand, [7, 229] use modern designs as a reference but only assume a maximum number of gateways per station without optimizing the exact value. To address this gap, Chapter 6 will detail a novel methodology for the Gateway Routing problem that includes the number of gateways per station as a parameter to optimize.

3.7 Validation of the link budget model

Note that the five models presented so far hinge on the link budget equation presented in Section 1.2.2. Since this equation is crucial to study the performance of satellite constellations, this section validates the link budget model through simulations. In particular, the link budget equation is evaluated for three specific constellations: SpaceX, OneWeb, and SES-O3b. One user is simulated, located at 0° latitude and 0° longitude. For the SpaceX and OneWeb constellations, one satellite altitude has been chosen as representative for the entire constellation, at 550 km and 1,200 km, respectively. For the SES-O3b constellations, their two altitudes have been chosen, at 507 km and 8,062 km. For each altitude, two elevation angles have been simulated, at 40° and 90°. The physical models regarding the user antenna, satellite antenna, and atmospheric conditions are the ones presented in this chapter. Interference is assumed to be 30 dB for each type (CABI, CASI, CXPI, C3IM). Only the user downlink has been simulated.

The link budget results, as well as intermediate values, are shown in Table 3.7. As shown, with a bandwidth of 250 MHz, all three systems achieve around 1 Gbps on all links. SpaceX requires approximately 0.7 W to close the link, OneWeb around 0.55 W, and SES-O3b between 0.5 W and 1 W for the lower altitude satellites, and between 2.9 W and 3.4 W for the higher altitude satellites.

Parameter	**SpaceX**		**OneWeb**		**SES-O3b**			
Altitude [km]	550		1,200		507		8,062	
Elevation angle [°]	90	40	90	40	90	40	90	40
Central frequency* [GHz]	11.7	11.7	11.7	11.7	19.0	19.0	19.0	19.0
Bandwidth [MHz]	250	250	250	250	250	250	250	250
EIRP [dB]	30.9	31.4	31.6	31.9	39.4	42.8	46.9	47.5
MODCOD [-]	64APSK 32/45-L	32APSK 11/15	32APSK 7/9	32APSK 2/3-L	256APSK 2/3-L	256APSK 2/3-L	32APSK 11/15	16APSK 5/6
Spectral efficiency [bps/Hz]	3.82	3.29	3.49	2.99	4.77	4.77	3.29	3.00
Distance [km]	550.0	813.3	1,200.0	1,692.8	507.0	751.2	8,062.0	9,487.6
FSLP [dB]	168.6	172.0	175.4	178.4	172.3	175.8	196.4	197.8
User Diameter* [m]	0.6	0.6	1.0	1.0	1.0	1.0	1.0	1.0
User Gain [dB]	35.5	35.5	39.9	39.9	44.3	44.3	44.3	44.3
$\frac{C}{N_0}$ [dB]	17.1	14.2	15.5	12.7	30.8	30.8	14.2	13.4
$\frac{E_b}{N_0}$ [dB]	11.2	9.0	10.1	7.98	24.0	24.0	9.0	8.6
$\frac{E_b}{N_I}$ [dB]	8.7	7.5	8.1	6.84	11.0	11.0	7.5	7.3
Margin [dB]	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Power [W]	0.67	0.75	0.53	0.56	0.53	1.16	2.94	3.41
Throughput [Gbps]	0.956	0.823	0.873	0.748	1.191	1.191	0.823	0.750

Table 3.7: Link budget model validation. Values marked with * are extracted from the models presented in this chapter.

Parameter	**SpaceX**	**OneWeb**
Altitude [km]	1,200	1,200
Central frequency* [GHz]	13.5	13.5
Bandwidth [MHz]	250	250
Spectral efficiency [bps/Hz]	2.7	2.4
Distance [km]	1,684	1,504
FSLP [dB]	179.6	178.6
User Diameter* [m]	0.7	0.75
$\frac{C}{N_0}$ [dB]	12	10.5
CASI [dB]	25	25
CXPI [dB]	22	20
$\frac{E_b}{N_0}$ [dB]	6.7	5.9
Margin [dB]	0.82	0.76
Throughput [Gbps]	0.674	0.599

Table 3.8: Values for the link budget model extracted from [7]

Assuming SpaceX uses the full 2 GHz spectrum available, with 2 polarizations and 4 frequency reuses, they could reach around 61 Gbps on a single satellite, at around 43 W of Radio Frequency (RF) power on the user downlink. These numbers reach 56 Gbps at 34 W for OneWeb, 214 Gbps at 95 W for the lower altitude satellites on SES-O3b, and 148 Gbps at 529 W for the higher altitude satellites on SES-O3b. Note that SES-O3b achieves such a high capacity due to the fact that the frequency reuse has been assumed on the order of 10.

Compared to existing literature [7], shown in Table 3.8, the numbers obtained through the model used in this dissertation are higher by a factor of 1.4 and 1.5 for the SpaceX and OneWeb constellations, respectively. Regarding the SpaceX constellation, the numbers are higher as the altitude of the constellation is significantly lower in current proposals compared to prior designs (550 km compared to 1,200 km). This reduces the FSLP by around 10 dB, which allows for higher spectral efficiencies, enabling higher throughput. Regarding the OneWeb constellation, the diameter of the user antenna has been considered slightly higher, following current filings (1 m compared to 0.75 m). This factor, combined with lower interference and margin, allow for a higher throughput. Based on this, it is reasonable to assume that the link budget model as presented in this work is valid.

3.8 Model assumptions

The models presented above outline the physical attributes of various entities essential for recreating realistic operational conditions. However, certain assumptions have been made during the description of these models, limiting their applicability to a subset of realistic scenarios. The primary assumptions include:

- Fixed Users: Users are assumed to be fixed and distributed according to the population distribution. While this overlooks emerging markets like aviation or maritime segments, interviews with satellite operators indicate that using population distribution has proven effective in estimating the general performance of satellite systems.

- Fixed User Demand: The assumption of fixed user demand over time is made. Despite the variable nature of Internet connectivity demands, assuming a fixed demand provides an effective means of estimating the performance tendency of the satellite network, rather than

instantaneous performance.

- Ideal User Terminals: User terminals are presumed to be capable of tracking any visible satellite at any point, and handovers between satellites occur instantly. Moreover, it is assumed that there are no outages caused by foliage or other obstructions. The time required for tracking and handover purposes, as well as the analysis of these factors, is beyond the scope of this dissertation.

- No Interference from Other Systems: Constellations are assumed to operate in a vacuum without interference from other satellite systems. This simplification allows for an optimistic estimation of the real capabilities of each constellation.

- Uniform Satellite Capabilities: It is assumed that all satellites are equipped with phased array antennas, on-board processing capabilities, use adaptive modulation and coding, and only employ circular beams with a fixed aperture angle. These assumptions simplify the resolution of the RAP while representing the characteristics of most satellite systems.

- Constant Satellite Power: Satellites are assumed to produce enough power to communicate at maximum effective isotropic radiation power (EIRP) whenever required. While power generation is influenced by orbital dynamics, this simplification allows for an optimistic estimation of the real capabilities of each constellation. The battery power and power over time management are out of the scope of this dissertation.

- Historical Ground Station Placement: Representative ground station placement can be estimated from historical data. Although each constellation requires its own ground segment network, addressing the Gateway Placement problem falls outside the scope of this dissertation. Interviews with satellite operators support the use of historical data as an acceptable alternative.

3.9 Chapter summary and conclusions

This chapter has described the five necessary models, in addition to the RAP framework to recreate realistic operational conditions. In addition, it has detailed appropriate data sources for each model, which allows for the simulation of system.

3.9.1 Chapter summary

The first section of this chapter, Section 3.1, outlined the identification of the five essential models integral to the RAP framework. Subsequent sections then proceeded to elaborate on each of these models in detail. Section 3.2 delineated the user model, which encompasses three primary elements: location, demand, and antenna characteristics. Following this, Section 3.3 provided a comprehensive overview of the atmospheric model, drawing upon recommendations from the ITU and leveraging openly available software. In Section 3.4, methods for determining the absence of *interference* were explained, considering parameters of the constellation and resource allocation. Section 3.5 delved into the characterization of satellite payload, contingent upon the specific constellation under consideration. Moreover, Section 3.6 introduced a novel approach for determining *gateway* locations, improving over existing literature. Since these five models hinge on the link budget equation of Chapter 1, Section 3.7 validates this equation by simulating different constellations and comparing the results against prior literature. Lastly, Section 3.8 elucidated the key assumptions underlying the described models.

3.9.2 Response to Research Questions

Research question 3.1

What are the necessary models, in addition to the RAP framework, to recreate realistic operational conditions regarding satellite communications constellations?

There are 5 additional models necessary to recreate realistic operational conditions: 1) User model, 2) Atmospheric model, 3) Interference model, 4) Satellite payload model, and 5) Gateway model. Each of these models represents a unique physical aspect that surrounds the RAP framework.

Research question 3.2

What are appropriate sources of data to simulate these models?

Due to the different nature of the models, each model must be simulated using their own data source. For the user model, appropriate data sources can be extracted from the world population distribution, and antenna product sheets from the operator. For the atmospheric model, the ITU provides standard models for each atmospheric effect. For the interference model, existing literature

relies on the separation between signals and their frequency spectrum to determine interference. For the satellite payload model, the public filings from operators might be used as a source of data. For the gateway model, existing literature relies on historical locations to determine possible ground station positions.

3.9.3 Specific chapter contributions

The specific contributions of this chapter are as follows:

- Detailed appropriate data sources for each model to represent realistic operational scenarios in modern megaconstellations.
- Applied the concept of isolation for large constellations, accounting for both intra- and inter-satellite interference. This model generalizes previous literature, which only focused on intra-satellite interference.
- Improved upon existing methods to determine ground station locations based on historical data and integer programming.

Chapter 4

Addressing hybrid LEO-MEO-HEO constellations through joint Beam Shaping and User Grouping

Hybrid constellations, which consist of satellite orbits at varying altitudes, are emerging as a prominent trend in constellation design (refer to Table 1.1). However, existing literature, as outlined in Section 2.2.1, has largely overlooked the inclusion of hybrid considerations in the resource allocation process. Instead, it typically assumes uniform altitude across all satellites within the constellation. Given that hybrid constellations afford operators greater flexibility, it stands to reason that integrating these considerations into resource allocation processes would enhance system performance. Therefore, the primary objective of this chapter is to address the identified gaps by: 1) proposing a formulation for joint Beam Shaping and User Grouping tailored to hybrid constellations, 2) presenting a methodology to handle the computational complexity associated with this formulation, addressing both low and high dimensionality scenarios, and 3) validating and assessing the performance of these proposed methods under realistic operational conditions. Throughout this chapter, geometric concepts described in Appendix A will be employed.

The research questions that this chapter aims to address are:

Research question 4.1

What are the complexities related to hybrid constellations w.r.t single-altitude designs? What are the mechanisms with which satellite operators can address these complexities?

Hypothesis: The purpose of this question is exploratory.

Research question 4.2

Can we increase the performance of satellite communications by including hybrid considerations in the resource allocation process?

Hypothesis: Yes. Including hybrid considerations during the resource allocation process will entail a higher flexibility, which will likely imply a higher utilization of resources. Since these consideration imply an additional level of complexity, it is possible that computational tractability becomes an issue to obtain high quality solutions in reasonable time.

4.1 Complexities of hybrid constellations

The complexities inherent in hybrid constellations, as opposed to single-altitude designs, stem primarily from the geometric intricacies involved. These complexities manifest in two key aspects:

Varying path loss

Satellites positioned at different altitudes experience varying levels of path loss, impacting the quality of communication links. In instances where the altitude variation is minimal (e.g., the SpaceX constellation with satellites between 540 km and 570 km), any additional path loss can typically be mitigated by adjusting the link margin without necessitating hardware modifications. However, in scenarios with significant altitude disparity (e.g., the LEO-MEO-HEO Boeing constellation), operators may need to compensate for heightened loss by augmenting antenna gain and EIRP for satellites at higher altitudes. It is noteworthy that in cases where user antennas have limited gain, certain users may be unable to establish communication with satellites at the highest altitudes. In such instances, the resource allocation process must ensure these users are assigned to nearby satellites where path loss is minimized.

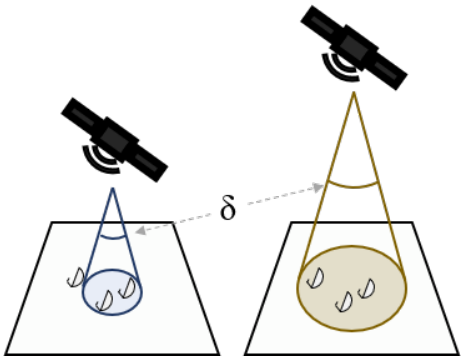


Figure 4-1: Footprint of satellites at different altitudes for the same beam shape

Varying footprints

Despite employing identical beam shapes, satellites at differing altitudes yield distinct footprints due to their varying distances from the Earth's surface. Consequently, for a given beam shape, beam center, and user, the user coverage within the beam may fluctuate depending on the altitude of the serving satellite, as depicted in Figure 4-1. Operators typically address this issue through two approaches: 1) By maintaining a fixed footprint and adjusting the beam shape accordingly to ensure consistent coverage, irrespective of altitude, and 2) By categorizing users based on altitude, thereby guaranteeing continuous coverage by assigning them to suitable altitudes. Currently, the former approach of adjusting beam shapes is more prevalent in literature. However, it fails to fully capitalize on the advantages afforded by satellites at lower altitudes, which can utilize smaller footprints with higher gain. The latter approach, focusing on altitude-based user assignments, remains underexplored in literature. To bridge this gap, the following section outlines a methodology for incorporating altitude-based user assignments into the mapping between users, beams, and shapes, a problem known as the joint Beam Shaping and User Grouping problem.

4.2 Problem set-up

The objective of the joint Beam Shaping and User Grouping problem is to establish a mapping between users and beams, and between beams and shapes, ensuring users remain within the footprint of their assigned beam with the specific shape at all times. Additionally, mappings that maximize capacity are preferred. Formally, the set of users to be covered, denoted as $\mathcal{U}$, includes information

regarding each user position p_u and demand d_u. User positions are assumed to remain fixed over time (see discussion in Section 4.13.1). Furthermore, users require continuous coverage to meet their demand. For considerations regarding users with non-continuous coverage, please refer to Section 4.13.4. For the purpose of this formulation, demand represents a reference data-rate expected to be received by each user, which could be the average demand, or the constant or peak information rate specified in the user contract.[1]. Following the user model in Section 3.2, in this study, this data-rate is the peak information rate, assumed to be consumed continuously at 100 Mbps with no over-subscription.

To serve the beams, operators manage a set of satellites denoted as $\mathcal{S}$, each with orbital information summarized by a reference position $p_s = lon_s, lat_s, alt_s$ and velocity v_s, alongside a reference capacity c_s. The mean altitude of each satellite remains fixed, implying that satellites with zero eccentricity orbits maintain a consistent altitude. Given that satellites at varying altitudes often possess different payloads, the parameter c_s represents the spacecraft ideal throughput relative to other satellites in the constellation. The satellite constellation is assumed to be hybrid, with satellites divided into distinct orbital planes potentially at different altitudes. However, it is assumed that each satellite orbital plane remains static. Additionally, satellites are grouped into *altitudes*, where each altitude $\mathcal{A}$ consists of satellites covering continuous portions of the Earth's surface at a specific mean altitude $alt_{\mathcal{A}}$. It is further assumed that the capacity of all satellites within the same altitude is equivalent ($c_s = c_{\mathcal{A}}\ \forall s \in \mathcal{A}$).

In modern systems, each satellite has the capability to project multiple beams toward the Earth's surface, each characterized by a distinct shape. Geometrically, each shape is defined by three components: an origin, corresponding to the satellite position p_s; a principal direction, representing the vector from the satellite to the center of the beam p_b; and an array of rays defining the boundary within which the signal strength is halved compared to the strength along the principal direction (called the 3 dB angle for circular shapes). In the context of this formulation, users positioned within this defined shape are eligible to be served by the corresponding beam, while those situated outside are not. Although this study predominantly focuses on circular shapes, where the rays form a cone around the principal direction, a discussion on incorporating various shapes is provided in

[1]Contracts between satellite operators and individual users are historically known as service level agreements

Section 4.13.2. Now, a conical shape with an aperture angle δ can be formally defined as:

$$\mathcal{C}_{b,s,\delta} = \left\{ p \in \mathbb{R}^3 \mid (p - p_s)(p_b - p_s) - ||p - p_s||\ ||p_b - p_s|| \cos \delta/2 \geq 0 \right\} \tag{4.1}$$

Where the positions p, p_s, and p_b are represented in the Cartesian space.

The projection on the Earth surface of a shape is known as the footprint ϕ. Assuming the Earth is a perfect sphere, the footprint of a conic shape is:

$$\phi_{b,s,\delta} = \{ p \in \mathcal{C}_{b,s,\delta} \mid ||p|| = R_E, LoS(p_s, p, \mathcal{A}_s) \} \tag{4.2}$$

Note that, due to the dynamic nature of NGSO constellations, the footprint of a beam, even when directed towards the same point, varies over time depending on the relative position of the satellite. However, since the user-beam mapping must remain valid over extended time periods, any viable mapping must ensure continuous user coverage within the beam footprint **at all times**. To address this requirement, the concept of *footprint contour* is introduced. A footprint contour Φ for a specific beam center p_b, shape $\mathcal{C}_{b,s,\delta}$, and altitude $\mathcal{A}$, is defined as the intersection of all footprints projected towards p_b with the specified shape $\phi_{b,s,\delta}$ by any visible satellite $s \in \mathcal{A} | LoS(s, b, \mathcal{A})$ at any given time $\mathcal{T}$ (refer to Figure 4-2):

$$\Phi_{b,\mathcal{A},\delta} = \left\{ p \in \mathbb{R}^3 \mid p \in \phi_{b,s,\delta}\ \forall\ s \in \mathcal{A} | LoS(p_s, p_b, \mathcal{A}), t \in \mathcal{T} \right\} \tag{4.3}$$

Note that satellites at the same altitude utilizing the same shape will generate identical footprint contours. However, differing altitudes or shapes will produce distinct footprint contours.

In addition to the beam center p_b, a beam b is characterized by an altitude $\mathcal{A}_b$, a footprint contour Φ_b, and a set of associated users $\mathcal{V}_b$. A beam is considered valid only if:

$$\begin{aligned} \mathcal{V}_b \neq \emptyset \quad & C1: \text{ The beam is not empty} \\ p_u \in \Phi_b \forall u \in \mathcal{V}_b \quad & C2: \text{ All users fall within the footprint contour} \\ v_{p_b,\mathcal{A}} \geq 1 \quad & C3: \text{ The beam has always at least 1 satellite } s \in \mathcal{A} \text{ in LoS} \end{aligned} \tag{4.4}$$

In this study, it is assumed that a beam associated with a specific altitude $\mathcal{A}$ can only be serviced by satellites within that altitude. Consequently, users assigned to a particular beam can

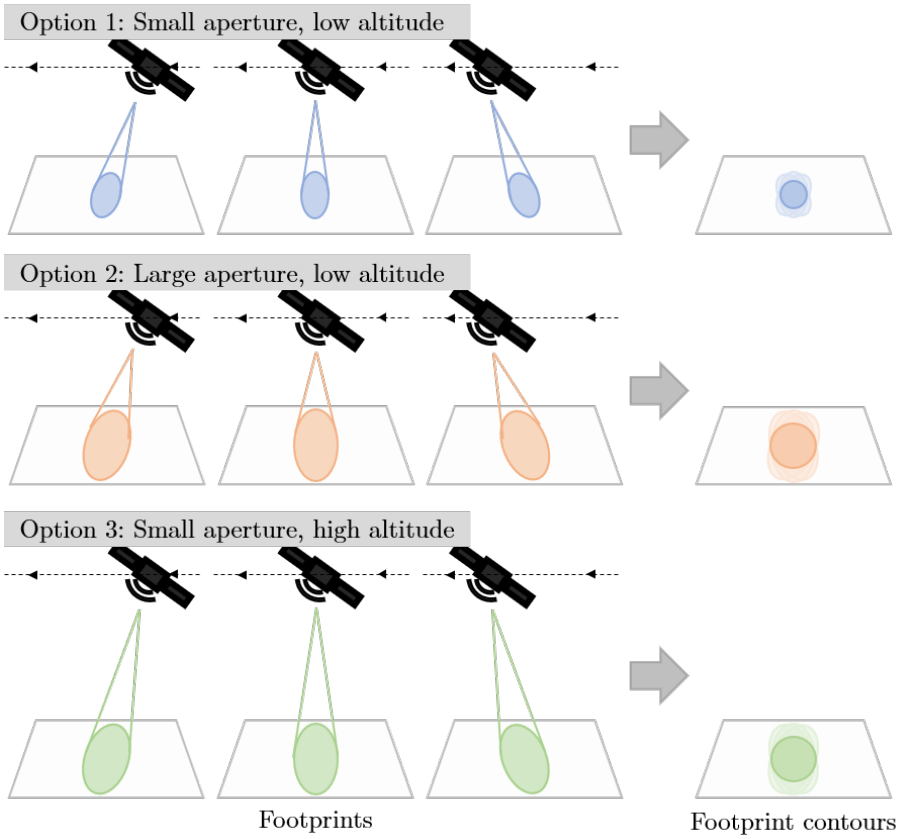


Figure 4-2: Difference between footprint and footprint contour. A footprint corresponds to the intersection of a shape with the Earth's surface at a specific point in time, while a footprint contour corresponds to the intersection of all footprints at any point in time.

only be served by the subset of satellites affiliated with that beam. This constraint arises from the definition of footprint contours: each altitude generates a distinct set of footprint contours, and each beam corresponds to a specific contour. Consequently, transferring beams between altitudes is not feasible unless the set of associated users $\mathcal{V}_b$ also falls within the footprint contour Φ of the new altitude, which is generally not the case.

4.3 Problem formulation

Based on this premise, the joint Beam Shaping and User Grouping problem for hybrid constellations is defined as identifying a set of valid beams that maximizes total capacity while ensuring coverage

for all users. To effectively address this problem, it is crucial to understand how the joint Beam Shaping and User Grouping problem impacts system capacity. Previous research [12] illustrates that reducing the number of beams correlates with higher throughput. This occurs due to two phenomena: 1) fewer beams imply more users per beam, which enables a higher utilization of the resources allocated to the beam, and 2) fewer beams imply a greater distance between them, reducing possible interference and enabling a more efficient utilization of the frequency spectrum. Note that, while a larger number of beams allows for a higher spectrum reuse, multiple beams can use the same frequency band only if they are sufficiently separated, as detailed in Section 3.4.

Therefore, grouping nearby users, which cannot reuse frequency, into single beams increases capacity. This aligns with existing literature on the User Grouping sub-problem [151, 153, 168]. However, when determining the mapping of beams to satellites, [12] also demonstrates that distributing beams across satellites positively influences total capacity. Given that the mapping of users to altitudes essentially represents a form of beam-to-satellite mapping, it is imperative to consider the balancing across satellites as an additional objective for optimization. Finally, *ensuring coverage for all users* is included in the formulation in the form of a constraint, similar to existing literature [151, 153, 168]. To that end, it is assumed that the satellites have an unlimited capacity to generate beams. Restricting the number of beams to the technical capabilities of the system will be addressed during the Satellite Routing and Frequency Assignment sub-problems in Chapter 5.

Subsequently, the joint Beam Shaping and User Grouping problem encompasses two primary objectives: minimizing the number of beams and balancing the load across altitudes. Note that these objectives can sometimes conflict: while reducing the number of beams might entail using larger footprints, corresponding to satellites at higher altitudes, such a strategy could result in an imbalance in demand towards those satellites, potentially underutilizing spacecraft at lower altitudes. Hence, striking a balance between these two objectives is paramount. It is important to emphasize that a user u cannot be covered only if an altitude that covers p_u continuously does not exist (e.g., when an equatorial constellation attempts to cover a user located at the North Pole). $\mathcal{U}^*$ denotes the subset of users that can be continuously served by at least one altitude within the constellation.

Assuming knowledge of the set of all possible valid beams $\mathcal{B}$ (i.e., beams satisfying the constraints in Equation 4.4), we define $x_{u,b}$ as a binary variable mapping user u to beam $b \; \forall \; u \in \mathcal{U}, b \in \mathcal{B}$. A user can only be assigned to a beam if it falls within the footprint contour ($p_u \in \Phi_b$) and possesses

the hardware capability to connect to the altitude associated with the beam $\mathcal{A}_b$. To distribute demand across altitudes, d_b represents the demand of beam b, computed as $d_b = \sum_{u \in \mathcal{V}_b} d_u x_{u,b}$. To fulfill this demand, the resource allocation process must allocate specific resources (e.g., frequency channels and power) to each beam. To decouple the joint Beam Shaping and User Grouping from other resource allocation sub-problems, η_b is introduced as a measure of the resources consumed by beam b. In particular, η_b is defined as the required number of channels per beam, and estimated using a static spectral efficiency Γ, a fixed bandwidth per channel BW_{ch}, and the data rate equation (Equation 1.5):

$$\eta_b = \left\lceil \frac{\sum_{u \in \mathcal{V}_b} d_u x_{u,b}}{BW_{ch}\Gamma} \right\rceil \tag{4.5}$$

Note that, while Γ could depend on the altitude to account for the fact that higher altitudes tend to have lower spectral efficiency, studying the implications of varying Γ falls out of the scope of this study. Furthermore, η_b is upper bounded by the maximum number of channels assigned to a single beam, denoted as η_{max}. Next, the *impact* of beam b at point p of altitude $\mathcal{A}$ is defined as:

$$I_{b,\mathcal{A},p} = \mathbb{1}_{\mathcal{A}_b = \mathcal{A}} \mathbb{1}_{LoS(p,p_b,\mathcal{A})} \frac{1}{v_{\mathcal{A},p_b}} \tag{4.6}$$

The impact denotes the fraction of demand that beam b is expected to transmit to a satellite positioned at point p and altitude $\mathcal{A}$. By definition, it is 0 if the beam is not assigned to altitude $\mathcal{A}$ or if the point is not within the line of sight of the beam. If non-zero, the demand is assumed to be evenly distributed among all visible satellites. A visualization illustrating the impact of candidate beams is depicted in Figure 4-3. Subsequently, given that the capacity of the constellation is localized (i.e., the capacity at each point on the Earth's surface predominantly depends on the satellites visible at that point), the demand at point p and altitude $\mathcal{A}$ is calculated as:

$$\eta_{\mathcal{A},p} = \sum_b I_{b,\mathcal{A},p} \eta_b \tag{4.7}$$

Note that, by definition, only $p_b \in \mathcal{P}^E_{\mathcal{A}}$ are valid. Now the maximum and minimum weighted

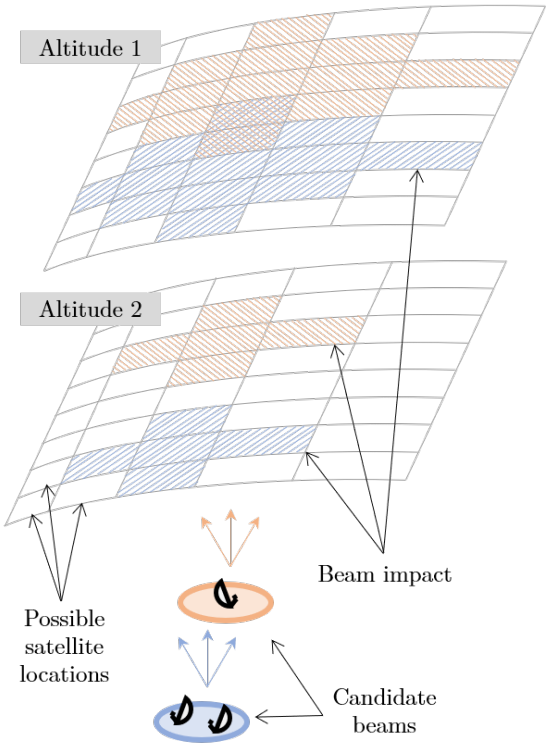


Figure 4-3: Illustration of the impact of candidate beams on each altitude

demand over all altitudes at each point can be computed as:

$$\begin{aligned} \mu_p^+ &= \max_{\mathcal{A} \ni p \in \mathcal{P}_\mathcal{A}^E} \frac{\eta_{\mathcal{A},p}}{c_\mathcal{A}} \\ \mu_p^- &= \min_{\mathcal{A} \ni p \in \mathcal{P}_\mathcal{A}^E} \frac{\eta_{\mathcal{A},p}}{c_\mathcal{A}} \end{aligned} \tag{4.8}$$

Now, the maximum imbalance between the different altitudes can be computed as:

$$\gamma = \max_p \mu_p^+ - \mu_p^- \tag{4.9}$$

Finally, to include the minimization of the number of beams in the formulation, y_b is defined as a binary auxiliary variable that indicates whether a beam is active or not, computed as:

$$y_b = \max_u x_{u,b} \tag{4.10}$$

Thus, the complete formulation is:

$$
\begin{aligned}
\min \quad & \omega_l \gamma + \omega_d \sum_b y_b & & \\
s.t. \quad & y_b = \max_u x_{u,b} & & \forall b \in \mathcal{B} \\
& \gamma = \max_p \mu_p^+ - \mu_p^- & & \\
& \mu_p^+ = \max_{\mathcal{A} \ni p \in \mathcal{P}_{\mathcal{A}}^E} \frac{\eta_{\mathcal{A},p}}{c_{\mathcal{A}}} & & \forall p \in \mathcal{P} \\
& \mu_p^- = \min_{\mathcal{A} \ni p \in \mathcal{P}_{\mathcal{A}}^E} \frac{\eta_{\mathcal{A},p}}{c_{\mathcal{A}}} & & \forall p \in \mathcal{P} \\
& \eta_{\mathcal{A},p} = \sum_b \mathbb{1}_{\mathcal{A}_b = \mathcal{A}} \mathbb{1}_{LoS(p,p_b,\mathcal{A})} \frac{\eta_b}{v_{\mathcal{A},p_b}} & & \forall \mathcal{A}, \forall p \in \mathcal{P} \\
& \eta_b = \left\lceil \frac{\sum_{u \in \mathcal{V}_b} d_u x_{u,b}}{BW_{ch}\Gamma} \right\rceil & & \forall b \in \mathcal{B} \\
& \eta_b \leq \eta_{max} & & \forall b \in \mathcal{B} \\
& \sum_b x_{u,b} = 1 & & \forall\ u \in \mathcal{U}^* \\
& x_{u,b} \in \{0,1\} & & \forall\ u \in \mathcal{U}^*, b \in \mathcal{B} \ni p_u \in \Phi_b \\
& x_{u,b} = 0 & & \forall\ u \in \mathcal{U}^*, b \in \mathcal{B} \ni p_u \notin \Phi_b \\
& y_b \in \{0,1\} & & \forall b \in \mathcal{B}
\end{aligned}
\tag{4.11}
$$

The aim of this formulation is to distribute the load across different altitudes in a way that is proportional to the capacity of each altitude at each point, while minimizing the number of beams. As highlighted before, it is assumed that the payload can generate an unlimited number of beams, and that restricting the number of beams to the capabilities of the payload is part of the Satellite Routing and Frequency Assignment, addressed in Chapter 5. The weight of each objective, represented as ω_l and ω_d, respectively, will be discussed in Section 4.12.

Problem linearization

While the previous formulation is a composition of several optimization problems, it can be linearized by manipulating the expressions. First, the ceiling operation in Equation 4.5 can be trans-

formed into three linear constraints:

$$\begin{aligned}
\eta_b &\geq \frac{\sum_{u \in \mathcal{V}_b} d_u x_{u,b}}{BW_{ch}\Gamma} && \forall b \in \mathcal{B} \\
\eta_b &\leq \frac{\sum_{u \in \mathcal{V}_b} d_u x_{u,b}}{BW_{ch}\Gamma} + 1 - \epsilon && \forall b \in \mathcal{B} \\
\eta_b &\in \mathbb{Z} && \forall b \in \mathcal{B}
\end{aligned} \tag{4.12}$$

Where ϵ is a small tolerance value. Finally, since they do not represent disjoint spaces, the maximum and minimum operations can be replaced by inequalities:

$$\begin{aligned}
\gamma &\geq \mu_p^+ - \mu_p^- && \forall p \\
\mu_p^+ &\geq \frac{\sum_b I_{b,\mathcal{A},p}\eta_b}{c_{\mathcal{A}}} && \forall \mathcal{A} \ni p \in \mathcal{P}_{\mathcal{A}}^E \\
\mu_p^- &\leq \frac{\sum_b I_{b,\mathcal{A},p}\eta_b}{c_{\mathcal{A}}} && \forall \mathcal{A} \ni p \in \mathcal{P}_{\mathcal{A}}^E \\
y_b &\geq x_{u,b} && \forall u
\end{aligned} \tag{4.13}$$

Now the complete linear formulation is:

$$
\begin{aligned}
\min \quad & \omega_l \gamma + \omega_d \sum_b y_b & & \\
s.t. \quad & y_b \geq x_{u,b} & & \forall u \in \mathcal{U}^*, b \in \mathcal{B} \\
& \gamma \geq \mu_p^+ - \mu_p^- & & \forall p \in \mathcal{P} \\
& \mu_p^+ \geq \frac{\sum_b I_{b,\mathcal{A},p} \eta_b}{c_{\mathcal{A}}} & & \forall \mathcal{A} \ni p \in \mathcal{P}_{\mathcal{A}}^E \\
& \mu_p^- \leq \frac{\sum_b I_{b,\mathcal{A},p} \eta_b}{c_{\mathcal{A}}} & & \forall \mathcal{A} \ni p \in \mathcal{P}_{\mathcal{A}}^E \\
& \eta_b \geq \frac{\sum_{u \in \mathcal{V}_b} d_u x_{u,b}}{BW_{ch}\Gamma} & & \forall b \in \mathcal{B} \\
& \eta_b \leq \frac{\sum_{u \in \mathcal{V}_b} d_u x_{u,b}}{BW_{ch}\Gamma} + 1 - \epsilon & & \forall b \in \mathcal{B} \\
& \sum_b x_{u,b} = 1 & & \forall\ u \in \mathcal{U}^* \\
& x_{u,b} \in \{0,1\} & & \forall\ u \in \mathcal{U}^*, b \in \mathcal{B} \ni p_u \in \Phi_b \\
& x_{u,b} = 0 & & \forall\ u \in \mathcal{U}^*, b \in \mathcal{B} \ni p_u \notin \Phi_b \\
& \eta_b \in \mathbb{Z} & & \forall b \in \mathcal{B} \\
& y_b \in \{0,1\} & & \forall b \in \mathcal{B}
\end{aligned}
\tag{4.14}
$$

Note that, although linear, this formulation contains both integer and continuous variables, and thus corresponds to a MILP problem, which is NP-hard. Since beams are associated to altitudes and shapes, this formulation addresses the user-beam, beam-shape, and beam-altitude mappings at the same time, breaching the gap found in Section 2.2.1.

4.4 Proof of NP-hardness

While a MILP formulation always conforms an NP-hard problem, there is no guarantee that the original problem is NP-hard. The following lines prove that the joint Beam Shaping and User Grouping problem as described in this work is NP-hard by proving that the associated decision problem is NP-Complete. The NP-completeness of the joint Beam Shaping and User Grouping decision problem can be demonstrated using two steps: 1) Proving that the problem is a combination of the set cover problem and a generalized generalized multiway number partitioning problem, indicating NP-hardness, and 2) Describing a polynomial-time verifier, which confirms that the

Case	$\omega_l = 0$	$\omega_d = 0$
Simplified formulation	$\min \quad \sum_b y_b$ $s.t. \quad y_b \geq x_{u,b}$ $\sum_b x_{u,b} = 1$ $x_{u,b} \in \{0,1\}$ $x_{u,b} = 0$ $y_b \in \{0,1\}$ (4.15)	$\min \quad \gamma$ $s.t. \quad \gamma \geq \mu^+_{p_0} - \mu^-_{p_0}$ $\mu^+_{p_0} \geq \sum_b I_{b,\mathcal{A},p_0} \sum_{u \in \mathcal{V}_b} d_u x_{u,b}$ $\mu^-_{p_0} \leq \sum_b I_{b,\mathcal{A},p_0} \sum_{u \in \mathcal{V}_b} d_u x_{u,b}$ (4.16) $\sum_b x_{u,b} = 1$ $x_{u,b} \in \{0,1\}$
Specific scenario		$\mathcal{P} = \{p_0\}$ $p_u \in \Phi_b \ \forall u \in \mathcal{U}^*, b \in \mathcal{B}$ $d_u \in \mathbb{N}, \nu_{b,p,\mathcal{A}} \in \mathbb{N}$ (4.17) $c_{\mathcal{A}} = c \forall \mathcal{A}$ $BW_{ch} = \Gamma = 1$
Equivalent problem	Set cover	Number partitioning

Table 4.1: Simplified joint Beam Shaping and User Grouping formulations and equivalent problems

problem belongs to NP.

To establish NP-hardness, three cases are considered based on the objective weights: 1) $\omega_l = 0$, 2) $\omega_d = 0$, and 3) $\omega_l \neq 0, \omega_d \neq 0$. A summary of the equivalent formulations is shown in Table 4.1.

When $\omega_l = 0$, the objective is to find the minimum set of beams y_b that cover all users. Beam b can cover user u iff $p_u \in \Phi_b$. $\mathcal{U}_b$ is defined as the set of users covered by beam b. The objective of the problem is to find the smallest collection of $\mathcal{U}_b$ such that their union equals $\mathcal{U}$. This corresponds exactly to the set cover problem, a known NP-complete problem [231].

When $\omega_d = 0$, the objective is to find a partition of users that minimizes the difference between altitudes. The following assumptions are introduced: $\mathcal{P}$ contains only one position ($\mathcal{P} = \{p_0\}$), any user can be mapped to any beam ($p_u \in \Phi_b$ is always true), the demand of each user and impact of a beam are integers ($d_u \in \mathbb{N}, \nu_{b,p,\mathcal{A}} \in \mathbb{N}$), the capacity of all altitudes are equal ($c_{\mathcal{A}} = c \forall \mathcal{A}$), and the bandwidth per channel and spectral efficiency are 1 ($BW_{ch} = \Gamma = 1$). Given these conditions, the objective of the problem is to find a user partition that minimizes the difference between the most

and least loaded altitudes. This problem is a strict generalization of number partitioning, which is known to be NP-hard [232].

When $\omega_l \neq 0$ and $\omega_d \neq 0$, the problem combines the previous two cases, which completes the proof for NP-hardness.

To establish NP-completeness, the polynomial-time verifier that checks, given $x_{u,b}\ \forall u \in \mathcal{U}, b \in \mathcal{B}$, whether $\omega_l \gamma + \omega_d \sum_b y_b \leq k$ for a given $k \in \mathbb{R}_+$ consists of resolving the auxiliary variables based on the formulation described in Equation 4.14, resulting in linear-time verification with respect to the input size. This demonstrates that the joint Beam Shaping and User Grouping decision problem is NP-Complete, proving that the optimization problem is NP-Hard.

4.5 Direct approach

The method proposed in this work to solve this formulation relies on using mathematical solvers that can address the complexity of MILP formulations. However, current MILP solvers cannot deal with continuous constraint spaces such as $\mathcal{P}$ or $\mathcal{B}$. Thus, the following lines indicate how to effectively discretize the search space to be able to find solutions using off-the-shelf mathematical solvers.

The first step involves obtaining a representative list of potential beams. However, the complexity of the previous formulation scales with the number of beams, and since the space of possible beams is continuous, enumerating all beams is computationally prohibitive, as observed in prior literature [168]. To mitigate this challenge, previous studies [151,153,168] have focused on identifying the minimum number of beams required to cover the set of users. Additionally, for constellations with all the satellites at the same altitude, [12] demonstrates that minimizing the number of beams consistently yields resource consumption reduction, and grid-like approaches perform comparably to optimization frameworks. Consequently, the proposed approach entails initially identifying possible beam centers using a grid-like methodology. However, to account for hybrid systems, multiple grids are generated, each one representing a different footprint contour. Specifically, spherical tessellation is employed [233]. To achieve this, the Earth's surface is represented using an icosahedral grid, which involves dividing a perfect sphere into a series of triangles. Figure 4-4 illustrates the construction of such a representation. Initially, the Earth's surface is covered using a known, high-granularity grid. Subsequently, each initial triangle is recursively subdivided into smaller triangles

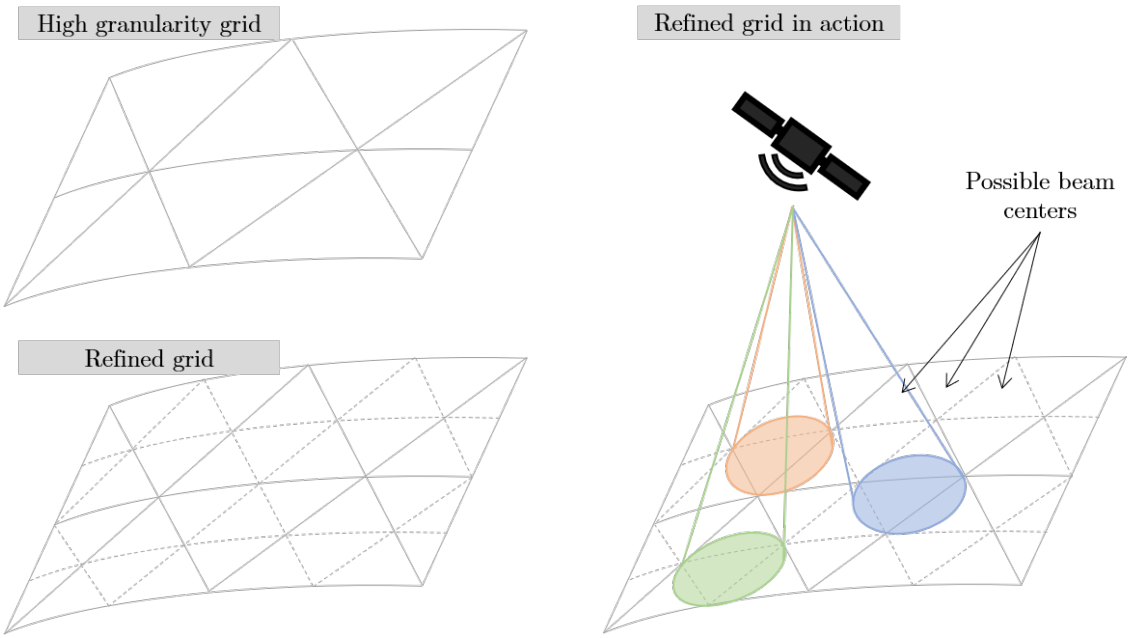


Figure 4-4: Example of a grid construction using an icosahedral tessellation

until the size of the sub-triangles is smaller than a specific footprint contour. Each resultant triangle center defines a potential beam center. This procedure is repeated for all existing footprint contours. Since the specific beam centers hinge on the choice of the initial grid, N_{tess} randomized lists of initial positions are utilized to generate multiple grids. The set of possible beams $\mathcal{B}$ corresponds to the collection of beam centers derived from each initial position. From this set, the users covered by each beam ($\mathcal{V}_b$) can be determined. Given that pinpointing the exact center of the beam necessitates solving the Beam Placement sub-problem, which falls beyond the scope of this approach, once the set of users has been computed for each beam, the beam center is approximated as a weighted sum of all users within the beam:

$$\begin{aligned} p_b &= \frac{\sum_{u \in \mathcal{V}_b} w_u p_u}{\sum_{u \in \mathcal{V}_b} w_u} \\ w_u &= \sum_{u' \in \mathcal{V}_b} |p_{u'} - p_u| \end{aligned} \tag{4.18}$$

Hence, two beams covering the same set of users share the same beam center. Note that the list of potential beams scales with the number of altitudes ($N_{\mathcal{A}}$), the number of footprint contours

Algorithm 1 Find possible set of beams $\mathcal{B}$

Input: N_{tess} ▷ Number of random initializations
Input: $\{\mathcal{A}\}$ ▷ Set of altitudes
Input: $\{\Phi_{\mathcal{A}}\}$ ▷ Set of footprint contours per altitude
Output: $\mathcal{B}$ ▷ List of possible beams

$\mathcal{B} = \emptyset$ ▷ Initialize empty solution ($\mathcal{B}$)
for $\mathcal{A}$ **do** ▷ For each altitude
  for $\Phi_{\mathcal{A}}$ **do** ▷ For each footprint contour in each altitude
    $\mathcal{B}_c = \emptyset$ ▷ Initialize temporary set
    for $n \in \{1, ..., N_{tess}\}$ **do** ▷ For each random initialization
      $\mathcal{B}_{grid} = \text{ComputeGrid}(\Phi_{\mathcal{A}})$ ▷ Create a grid with footprint contour $\Phi_{\mathcal{A}}$ and random origin
      for $b_1 \in \mathcal{B}_{grid}$ **do** ▷ For each beam in the grid
        $\mathcal{A}_{b_1} = \mathcal{A}$ ▷ Initialize altitude
        $\Phi_{b_1} = \Phi_{\mathcal{A}}$ ▷ Initialize footprint contour
        $\mathcal{V}_{b_1} = \{u \ni p_u \in \Phi_{b_1}\}$ ▷ Compute the set of users that can be covered
        if $|\mathcal{V}_{b_1}| > 0$ **then** ▷ If the user set is not empty
          p_{b_1} = Equation 4.18 ▷ Update beam center according to Equation 4.18
          include = True
          for $b_2 \in \mathcal{B}_c$ **do** ▷ For each beam in the temporal set
            if $\mathcal{V}_{b_1} \subseteq \mathcal{V}_{b_2}$ **then** ▷ If the beam is covered, skip it
              include = False
            if $\mathcal{V}_{b_2} \subset \mathcal{V}_{b_1}$ **then** ▷ If any of the previous beams becomes covered, extract it
              $\mathcal{B}_c = \mathcal{B}_c \setminus b_2$
          if include **then** ▷ Add the new beam to the set if not covered
            $\mathcal{B}_c = \mathcal{B}_c \cup b_1$
    $\mathcal{B} = \mathcal{B}_c \cup \mathcal{B}$ ▷ Update beam set

considered per altitude (N_Φ), the granularity of the grid, and the number of different initializations (N_{tess}). As a reference, a constellation with 4 altitudes, 1 contour per altitude, grid cell area of 400 km^2, and 100 random initializations yields approximately 2×10^8 possible beams[2]. To mitigate the potential number of beams, this dissertation leverages the concept of *cover*, as utilized in previous literature [168]. Specifically, it assumes that a beam b_1 is covered by beam b_2 if b_1 covers a subset of users of b_2 ($\mathcal{V}_{b_1} \subset \mathcal{V}_{b_2}$), and both b_1 and b_2 are assigned to the same altitude and share the same footprint contour ($\mathcal{A}_{b_1} = \mathcal{A}_{b_2}$ and $\Phi_{b_1} = \Phi_{b_2}$). Moreover, two beams are considered *equal* if they cover the same set of users, are assigned to the same altitude, and employ the same footprint contour ($\mathcal{V}_{b_1} = \mathcal{V}_{b_2}$, $\mathcal{A}_{b_1} = \mathcal{A}_{b_2}$, and $\Phi_{b_1} = \Phi_{b_2}$). Removing duplicated and covered beams from the set effectively reduces the list of beams. Algorithm 1 outlines the pseudo-code for identifying the list of potential beams.

To solve the problem using modern mathematical solvers, $\mathcal{P}$ is discretized using a similar tessellation approach. Here, the size of each cell serves as a hyperparameter of the optimization, rather than being predetermined by the problem. A smaller cell size enhances accuracy but at the

[2]This order of magnitude estimation aligns with expectations for a constellation similar to SpaceX

Algorithm 2 Iterative approach to find optimal γ for each position

Input: $\mathcal{B}$ ▷ Set of possible beams
Input: $\mathcal{U}^*$ ▷ Set of valid users
Input: $\{\mathcal{A}\}$ ▷ Set of altitudes
Input: BW_{ch}, Γ ▷ Model parameters
Output: $x_{u,b} \forall u \in \mathcal{U}^*, b \in \mathcal{B}$ ▷ Solution to the problem
$\mathcal{F} = \emptyset$ ▷ Initialize set of fixed users
$\mathcal{P}^- = \emptyset$ ▷ Initialize set of fixed positions
while $\mathcal{F} \subset \mathcal{U}^*$ **do** ▷ While some users are not fixed
    Solve Equation 4.14 using commercial solvers and assuming $\mathcal{F}$ and $\mathcal{P}^-$ are fixed ▷ Solve the problem
    Find p such that $\gamma = \mu_p^+ - \mu_p^-$ and $p \notin \mathcal{P}^-$ ▷ Find a position that cannot be balanced further
    for $b \in \mathcal{B}$ **do** ▷ For each beam
        if $w_{b,\mathcal{A},p} > 0$ **then** ▷ If the beam is active in position p
            $\mathcal{F} = \mathcal{F} \cup \mathcal{V}_b$ ▷ Add the users of beam b to the fixed user set
    $\mathcal{P}^- = \mathcal{P}^- \cup p$ ▷ Fix p

expense of computational speed. Once discretized, this formulation can be directly inputted into a commercial solver. However, note that, after determining γ, not all positions p may exhibit a load difference of γ. Certain regions might be able to achieve a more balanced distribution locally. Consequently, the problem is solved iteratively by identifying γ, fixing the users, and subsequently re-solving, as outlined in Algorithm 2.

4.6 Scalable approach

Although the discretized formulation can be fed into commercial solvers and solved optimally for a low number of users, it may have scalability issues for large number of users. To address this, an iterative approach is proposed, where only some variables at optimized at the same time, while the rest remain fixed. Note that this assumption sacrifices optimality for tractability. In particular, at each iteration, the procedure follows:

1. Initialize $\mathcal{B}'$, $\mathcal{B}^i$, $\mathcal{U}'$, $\mathcal{U}^i$ as empty sets.

2. Randomly select a user from $\mathcal{U}$, and add it to $\mathcal{U}'$.

3. Choose a user u from $\mathcal{U}'$, and add all the beams b such that u fall within coverage of beam b ($p_u \in \Phi_b$) and $b \notin \mathcal{B}^i$ to $\mathcal{B}'$. Extract u from $\mathcal{U}'$ and add it to $\mathcal{U}^i$.

4. For each beam b in $\mathcal{B}'$, add all the users u that fall within coverage of b ($p_u \in \Phi_b$) and $u \notin \mathcal{U}^i$ to $\mathcal{U}'$. Set $\mathcal{B}^i = \mathcal{B}^i \cup \mathcal{B}'$ and $\mathcal{B}' = \emptyset$.

5. If the size of $\mathcal{B}^i$ or $\mathcal{U}^i$ exceed predefined hyperparameters N_b or N_u, respectively, or if $\mathcal{U}'$ is empty, proceed with the next step. Otherwise go to step 3.

6. Solve the previous formulation by allowing the variables related to $\mathcal{B}^i$ and $\mathcal{U}^i$ to change, while keeping the remaining variables fixed.

The algorithm converges when there has been an improvement lower than I_{thres} for at least N_{conv} iterations. Note that this iterative procedure solves on top of the iterative algorithm for the optimal problem that addresses the differences between regions. Algorithm 3 shows how to address both fixed regions and fixed users simultaneously. To accelerate computation time, a warm-start can be obtained by running the same algorithm with a lower quality grid.

4.7 Complexity analysis

The complexity analysis of the proposed approaches can be performed at two levels: 1) the necessary memory needed to store the model, and 2) the potential size of the search space and complexity of the search. To analyze these aspects, M is defined as the maximum number of beams covering a user, computed as $M = \max_u \sum_b \mathbb{1}p_u \in \Phi_b$.

From a memory perspective, the formulation has two types of variables: decision variables ($x_{u,b}$), which scale with M and the number of users $|\mathcal{U}|$, and auxiliary variables ($y_b, \mu_p^+, \mu_p^-, \gamma$), which scale with the number of beams $|\mathcal{B}|$ and the total number of points $|\mathcal{P}|$. When using the direct approach, the total memory requirement is $\mathcal{O}(M|\mathcal{U}| + |\mathcal{B}| + 2|\mathcal{P}| + 1)$. When using the scalable approach, the number of users and number of beams is limited to N_u and $N_b + M$, respectively. Thus, the total memory requirement is $\mathcal{O}(MN_u + N_b + M + 2|\mathcal{P}| + 1)$.

From a computation perspective, both problems are NP-Hard, which can be proven using a similar logic as Section 4.4. The worst-case solution scales with the number of possible solutions. Since the auxiliary variables are fixed once the decision variables are computed, the size of the search space for the direct approach is $\mathcal{O}(M^{|\mathcal{U}|})$, which is the complexity of the algorithm. In the case of the scalable approach, the number of users is bounded by N_u, which leads to a total complexity of $\mathcal{O}(M^{N_u})$. By adjusting hyperparameters N_u and N_b, the complexity of the problem solved by the mathematical solver can be regulated.

Algorithm 3 Iterative, scalable approach to find optimized γ for each position

Input: $\mathcal{B}$ ▷ Set of possible beams
Input: $\mathcal{U}^*$ ▷ Set of valid users
Input: $\{\mathcal{A}\}$ ▷ Set of altitudes
Input: BW_{ch}, Γ ▷ Model parameters
Input: N_{users} ▷ Algorithm parameters
Output: $x_{u,b} \forall u \in \mathcal{U}^*, b \in \mathcal{B}$ ▷ Solution to the problem

$\mathcal{F} = \emptyset$ ▷ Initialize set of fixed users
$\mathcal{P}^- = \emptyset$ ▷ Initialize set of fixed positions
$x_{u,b}$ = Random Initialization or Warm-Start ▷ Initialize user-beam mapping
while $\mathcal{F} \subset \mathcal{U}^*$ **do** ▷ While some users are not fixed
    i=0, Prev=Inf ▷ Initialize counter and previous objective
    $\mathcal{U} = \mathcal{U}^* \setminus \mathcal{F}$ ▷ Initialize set of free users
    converged=False
    while not converged **do**
        $\mathcal{B}' = \mathcal{B}^i = \mathcal{U}' = \mathcal{U}^i = \emptyset$ ▷ Initialize auxiliary sets
        $\mathcal{U}' = \mathcal{U}' \cup \{Rand(\mathcal{U})\}$ ▷ Randomly select a user from $\mathcal{U}$, and add it to $\mathcal{U}'$
        while $|\mathcal{B}^i| < N_b$ and $|\mathcal{U}^i| < N_u$ and $\mathcal{U}' \neq \emptyset$ **do** ▷ While not converged
            $u = Rand(\mathcal{U}')$ ▷ Choose a user u from $\mathcal{U}'$
            for $b \in \mathcal{B}$ **do**
                if $p_u \in \Phi_b$ and $b \notin \mathcal{B}^i$ **then** ▷ If within coverage and not in $\mathcal{B}^i$
                    $\mathcal{B}' = \mathcal{B}' \cup \{b\}$ ▷ Add it to $\mathcal{B}'$
            $\mathcal{U}' = \mathcal{U}' \setminus \{u\}$ ▷ Extract u from $\mathcal{U}'$
            $\mathcal{U}^i = \mathcal{U}^i \cup \{u\}$ ▷ Add u from $\mathcal{U}^i$
            for $b \in \mathcal{B}'$ **do**
                for $u \in \mathcal{U}$ **do**
                    if $p_u \in \Phi_b$ and $u \notin \mathcal{U}^i$ **then** ▷ If u is within coverage and not in $u \notin \mathcal{U}^i$
                        $\mathcal{U}' = \mathcal{U}' \cup \{u\}$ ▷ Add it to $\mathcal{U}'$
            $\mathcal{B}^i = \mathcal{B}^i \cup \mathcal{B}'$ ▷ Add the beams to the already checked beams
            $\mathcal{B}' = \emptyset$ ▷ Empty the pending beams
        Solve Equation 4.14 using commercial solvers and assuming $\mathcal{F}$, $\mathcal{P}^-$, $\mathcal{U} \setminus \mathcal{U}^i$, and $\mathcal{B} \setminus \mathcal{B}^i$ are fixed
        if Prev - Current Objective $> I_{thres}$ **then** ▷ Check for convergence
            i=0
        else
            i = i + 1
            if i = N_{conv} **then**
                converged=True
        Prev = Current Objective
    Find p such that $\gamma = \mu_p^+ - \mu_p^-$ and $p \notin \mathcal{P}^-$ ▷ Find a position that cannot be balanced further
    for $b \in \mathcal{B}$ **do** ▷ For each beam
        if $w_{b,\mathcal{A},p} > 0$ **then** ▷ If the beam is active in position p
            $\mathcal{F} = \mathcal{F} \cup \mathcal{V}_b$ ▷ Add the users of beam b to the fixed user set
    $\mathcal{P}^- = \mathcal{P}^- \cup p$ ▷ Fix p

4.8 Experimental set-up

The following lines detail the validation and performance analysis on the proposed methodology. In particular, 4 different experiments that test the capabilities of the novel formulation have been executed, summarized in Table 4.2.

Analysis	Objective	Constellation	Number of users	Objective weight
Validation & Verification	Demonstrate that the proposed methodology produces valid solutions	SpaceX	20,000	$\omega_l = 1$ $\omega_d = 0$
Convergence	Demonstrate that the proposed methodology produces high quality feasible solutions in reasonable time	SpaceX	20,000	$\omega_l = 1$ $\omega_d = 0$
Tradespace	Study the trade-off between the difference objectives	SpaceX & Boeing	5,000	Variable
Performance	Analyze the performance with respect to established techniques	Boeing	Up to 100,000	$\omega_l = \hat{\omega}_l$ $\omega_d = \hat{\omega}_d$

Table 4.2: Experimental set-up to evaluate the proposed methodology on the joint Beam Shaping and User Grouping problem

To conduct the experiments, it is necessary to define a realistic set of users $\mathcal{U}$ and a realistic constellation configuration. The user distribution generation process is outlined in Section 3.2. For the constellation configuration, existing designs are selected to represent realistic scenarios. Each altitude $\mathcal{A}$ is defined as comprising satellites with the same mean altitude, and all satellites are assumed to be equal in payload. Two existing constellation designs are chosen:

- SpaceX constellation: This constellation comprises two altitudes at 540 km and 550 km with similar configurations (72 planes with 22 satellites per plane at around 53° inclination), a polar altitude at 560 km (with two sets of 6 and 4 planes with 58 and 43 satellites per plane, respectively), and a final altitude at 570 km (with 36 planes and 20 satellites per plane), featuring less than half the number of satellites compared to the altitudes at 540 km and 550 km. The complete configuration is outlined in Table 1.1.

- Boeing constellation: This constellation incorporates 10 different altitudes: 7 in LEO, 2 in MEO, and 1 in HEO (refer to Table 1.1 for specific configurations). It serves as an ideal example for testing the performance of the novel methodology under hybrid constellations, as it integrates LEO, MEO, and HEO satellites.

For standardized comparison, only circular beam shapes with an aperture angle of 2° are permitted. The footprint contour of each shape and altitude is determined by simulating the constellation every second over a 24-hour period and identifying the contour across all times. Additionally, the hyperparameters for all experiments are specified in Table 4.3. All simulations utilize the commercial solver Gurobi [234] (version 9.1.2) on an Intel(R) Xeon(R) Platinum 8160 CPU @ 2.10GHz,

Parameter	Value	Gurobi Parameter	Value
N_u	1,000	Threads	16
N_b	3,000	MIPGap	10^{-4}
I_{thres}	1%	Iteration	300s
N_{conv}	20	Time Limit*	

Table 4.3: Summary of the parameters of the simulation. * The Time Limit is only applicable for the scalable approach.

allowing for up to 16 simultaneous threads.

To assess the performance of the proposed methodology, it is compared against existing methods, particularly the formulation presented in [168]. In this method, the User Grouping problem is transformed into a clique-based problem, where the objective is to find the smallest set of cliques that cover a graph. In this context, a clique represents a set of nearby users that can be covered by a footprint contour. Given its complexity, the authors then describe a heuristic to find a solution to the problem. Given that this method was initially designed for single-altitude constellations, it is assumed that the footprint of each beam remains fixed and equal to the largest possible footprint. While this approach does not fully capitalize on the potential benefits of having smaller beams with higher gain, it allows the operator the flexibility to redistribute traffic between altitudes during operations.

4.9 Validation and verification analysis

The validation and verification analysis focuses on the SpaceX constellation, comprising four altitudes at 540 km, 550 km, 560 km (polar), and 570 km. Based on the formulation and orbital configuration, a satisfactory allocation should distribute approximately equal demand across the 540 km, 550 km, and 570 km altitudes, with only polar users (i.e., those around the polar circle) assigned to the polar altitude. To ensure the results align with expectations, the scalable methodology is executed for 20,000 users.

The results of the scalable methodology for this experiment are depicted in Figures 4-5, 4-6, 4-7, and 4-8. Specifically, these figures illustrate the expected satellite load of each altitude of the constellation at each coordinate point. The observations confirm the expectations:

- The 540 km, 550 km, and 570 km altitudes exhibit roughly equivalent demand across all

regions of the Earth, peaking at approximately 60 Gbps above India and China. Moreover, the demand is evenly distributed across the globe, rather than being concentrated solely in densely populated areas. Notably, the algorithm successfully balances the load even when the altitudes have significantly different numbers of satellites (approximately 1,500 satellites for the 540 km and 550 km altitudes, and 700 satellites for the 570 km altitude).

- The 560 km altitude exclusively serves polar users. Due to the altitude configuration and the problem formulation, only users above 58° or below -58° latitude receive continuous coverage, as evidenced in Figure 4-7. Note that, in practical scenarios, it is likely that SpaceX will use larger footprints on this altitude to cover continuously more meaningful portions of the Earth, thereby utilizing the satellites at this altitude.

4.10 Convergence analysis

The convergence analysis aims to ascertain how the objective value changes with each iteration to determine if the solution has converged. This analysis provides insight into the algorithm performance if computation time is constrained. Similar to the previous scenario, the SpaceX constellation is utilized. The scalable methodology is executed for 20,000 users, and the objective value and computation time for each iteration are recorded. As mentioned, an initial solution is obtained using a low-quality grid, which is subsequently refined with a higher resolution.

Figure 4-9 depicts the evolution of demand imbalance using the scalable algorithm with a warm-

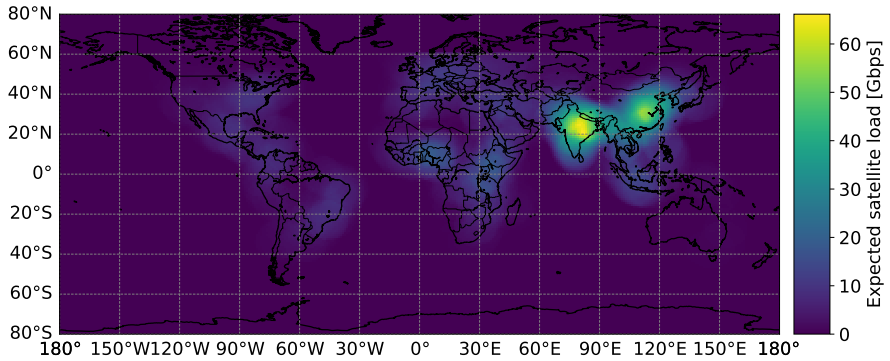


Figure 4-5: Expected satellite load on the 540 km altitude of the SpaceX constellation with 20,000 users. The expected load might be higher than the achievable load.

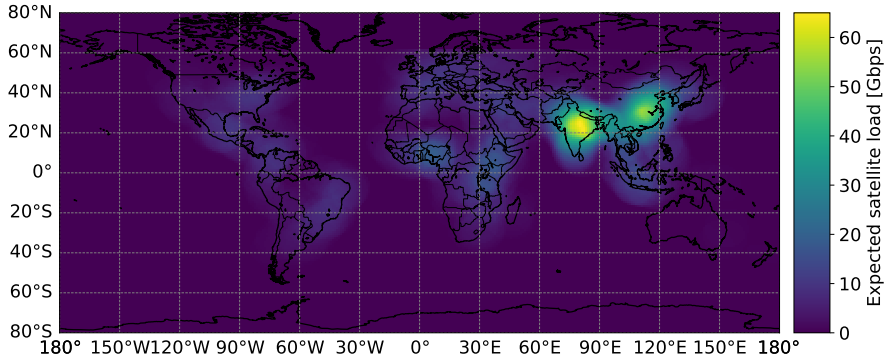


Figure 4-6: Expected satellite load on the 550 km altitude of the SpaceX constellation with 20,000 users. The expected load might be higher than the achievable load.

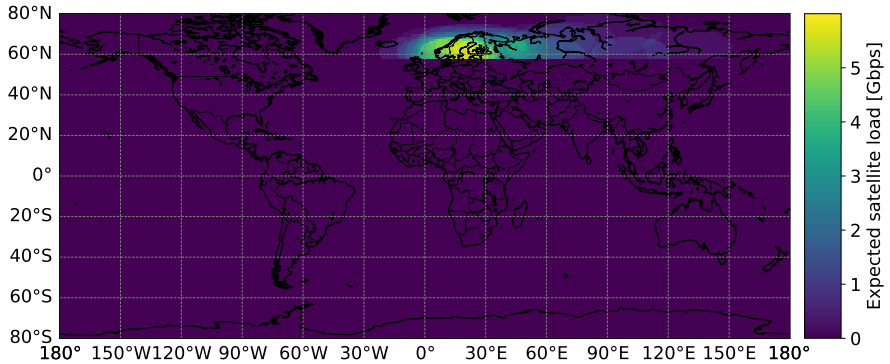


Figure 4-7: Expected satellite load on the 560 km altitude of the SpaceX constellation with 20,000 users. The expected load might be higher than the achievable load.

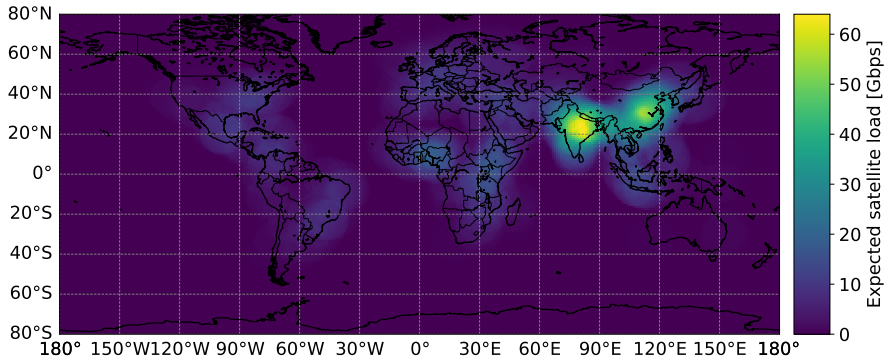
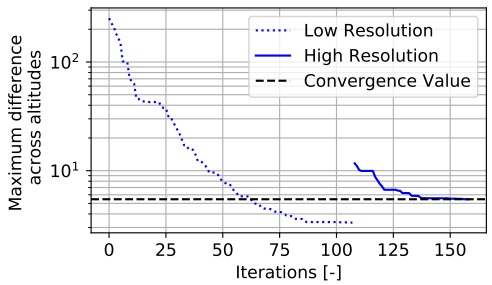
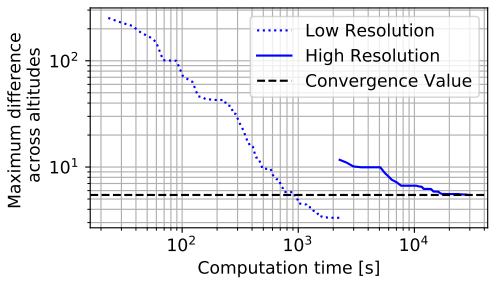


Figure 4-8: Expected satellite load on the 570 km altitude of the SpaceX constellation with 20,000 users. The expected load might be higher than the achievable load.



(a) Maximum difference vs. Iterations



(b) Maximum difference vs. Computation time

Figure 4-9: Maximum difference across altitudes per iteration and point in time for a single execution. The dotted line corresponds to the initial low resolution run, which is then used as a warm-start for the high resolution (continuous line).

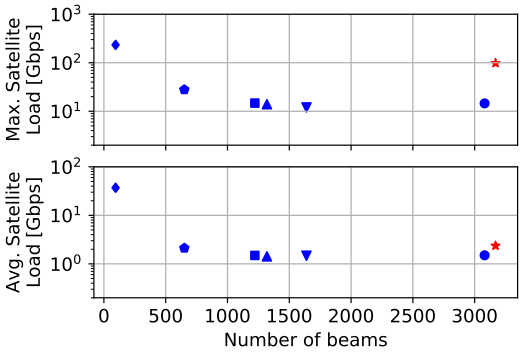
start approach for the SpaceX constellation comprising 20,000 users. The displayed values only represent the convergence process related to the solution where neither users nor regions are fixed. However, since the algorithm iteratively fixes users and regions upon convergence, the total computation time may exceed what is depicted in the figure. As shown, the initial low-resolution algorithm swiftly directs the solution towards the focal point of the search space. Within approximately 2,000 seconds, the low-resolution algorithm converges to a local optimum. Analyzing the transitional phase reveals that the convergence point attained by the low-resolution grid indeed represents a local optimum in the higher resolution, which is further refined in subsequent iterations, albeit at the expense of an additional computation time of around 28,000 seconds. Interestingly, although the higher-resolution execution involves fewer iterations, each iteration requires more time on average compared to the low-resolution start. The convergence of the algorithm is considered achieved when the solution stabilizes. As a final remark regarding convergence, when executing the same experiment 5 times, the objective value varies, but the standard deviation regarding multiple solutions is around 1% of the objective value. This means that, while multiple solutions attain different values, all solutions perform similarly with respect to the metrics.

4.11 Tradespace analysis

The objective of the tradespace analysis is to study the trade-off between the two objectives considered in this work: minimizing the number of beams and distributing the load across different

Table 4.4: Summary of the weights tested on the tradespace analysis

w_l (Boeing)	w_l (SpaceX)	w_d
$w_l = 1$	$w_l = 1$	$w_d = 0$
$w_l = 10\hat{w}_l$	$w_l = 100\hat{w}_l$	$w_d = \hat{w}_d$
$w_l = \hat{w}_l$	$w_l = 10\hat{w}_l$	$w_d = \hat{w}_d$
$w_l = 0.1\hat{w}_l$	$w_l = \hat{w}_l$	$w_d = \hat{w}_d$
$w_l = 0.01\hat{w}_l$	$w_l = 0.5\hat{w}_l$	$w_d = \hat{w}_d$
$w_l = 0$	$w_l = 0$	$w_d = 1$

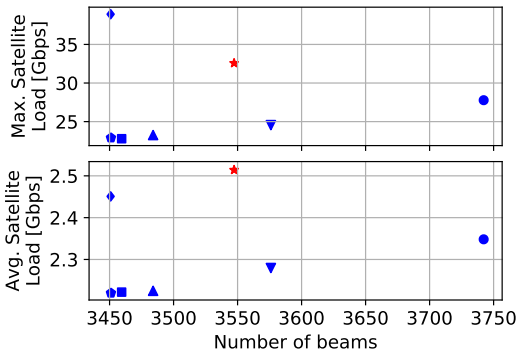


	w_d	w_l	Number of beams	Max. load [Gbps]	Avg. load [Gbps]
●	0	1	2993	13.41	1.395
▼	$\hat{w}_d$	$10\hat{w}_l$	2157	12.84	1.299
▲	$\hat{w}_d$	$\hat{w}_l$	1816	12.44	1.238
■	$\hat{w}_d$	$0.1\hat{w}_l$	1431	12.70	1.289
⬟	$\hat{w}_d$	$0.01\hat{w}_l$	782	27.30	1.905
◆	1	0	82	233.7	32.38
★	Single-altitude		3166	103.0	2.421

Figure 4-10: Maximum and average satellite load for the joint Beam Shaping and User Grouping problem using different weight factors on the Boeing constellation. The red star represents the solution obtained with single-altitude methods. The average satellite load ignores unused satellites. This visualization does not take into account the capacity limits imposed by the satellite payload. Therefore, the expected load might be higher than the achievable load. The y-axis is in logarithmic form

altitudes. To that end, the scalable algorithm is executed using different weights for the different objectives, and the results are compared against the method for the single-altitude constellations.

To evaluate both objectives as a single metric, it is necessary to normalize each objective in the formulation presented in Equation 4.14. When normalizing objectives, a common approach involves using a worst-case scenario to determine a normalization value. However, for the load balance objective, utilizing a worst-case scenario may result in biased outcomes, as it can yield a significantly higher value compared to the average solution. Therefore, an alternative normalization strategy is proposed. First, the solution corresponding to $\omega_l = 1$ and $\omega_d = 0$ is obtained, and the load balance γ_l and the number of beams N_l for this solution are computed. Subsequently, the normalization weights are defined as follows: $\hat{\omega}_l = \frac{1}{\gamma_l}$ and $\hat{\omega}_d = \frac{1}{N_l}$. Based on this approach, the



	w_d	w_l	Number of beams	Max. load [Gbps]	Avg. load [Gbps]
●	0	1	3742	27.78	2.348
▼	$\hat{w}_d$	$100\hat{w}_l$	3576	24.54	2.280
▲	$\hat{w}_d$	$10\hat{w}_l$	3484	23.26	2.225
■	$\hat{w}_d$	$\hat{w}_l$	3459	22.78	2.222
⬟	$\hat{w}_d$	$0.5\hat{w}_l$	3451	22.96	2.221
◆	1	0	3450	38.92	2.451
★	Single-altitude		3547	32.64	2.515

Figure 4-11: Maximum and average satellite load for the joint Beam Shaping and User Grouping problem using different weight factors on the Boeing constellation. The red star represents the solution obtained with single-altitude methods. The average satellite load ignores unused satellites. This visualization does not take into account the capacity limits imposed by the satellite payload. Therefore, the expected load might be higher than the achievable load.

simulations are executed with the weights highlighted in Table 4.4.

Before delving into the results, it is crucial to establish unbiased metrics for determining the actual performance between the proposed method and existing practices. Utilizing metrics such as minimizing the number of beams or achieving load balancing can introduce bias towards methods directly optimizing these metrics. To mitigate this, the real system capacity is computed using two metrics: the average satellite load across used satellites (i.e., ignoring unused satellites), and the maximum satellite load. Note that this dissertation does not use spectrum or power usage as comparison metrics (see discussion in Section 4.13.5). A lower average and maximum satellite load indicate better resource utilization and a higher constellation capacity. These metrics are computed by assigning each beam to the satellite with the highest elevation angle at the corresponding altitude, and aggregating the loads of all beams for each satellite.

Figures 4-10 and 4-11 depict the outcomes concerning the maximum and average satellite loads for the Boeing and SpaceX constellations, respectively, encompassing 5,000 users. These results stem from employing the scalable algorithm with multiple weight factors, alongside the single-altitude mechanism delineated in [168], which prioritizes minimizing the number of beams.

First, it is important to note that focusing solely on a single objective does not necessarily yield the highest capacity. When concentrating solely on balancing the beams across satellites ($\omega_d = 0, \omega_l = 1$, depicted by the blue circle), the algorithm generates solutions characterized by a large

number of beams, thereby resulting in a less efficient utilization of satellite resources. Conversely, solely emphasizing the minimization of the number of beams leads to an oversaturation of the highest altitudes, consequently diminishing overall performance. This effect is more pronounced in constellations spanning a wide range of altitudes, such as the Boeing constellation, although it remains notable even in constellations situated in close proximity, such as the SpaceX constellation.

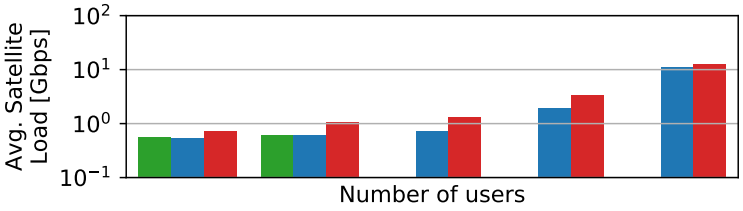
The optimal outcomes for achieving the lowest average satellite load are attained when $\omega_l = \hat{\omega}_l$ and $\omega_d = \hat{\omega}_d$ for the Boeing constellation, and when $\omega_l = 0.5\hat{\omega}_l$ and $\omega_d = \hat{\omega}_d$ for the SpaceX constellation. Furthermore, the most effective approach for minimizing the maximum satellite load is achieved when $\omega_l = \hat{\omega}_l$ and $\omega_d = \hat{\omega}_d$ in both cases. Lastly, the results underscore the importance of leveraging the flexibilities inherent in hybrid constellations, as opposed to solely relying on approaches tailored for single-altitude designs. Although the latter may offer more operational flexibilities, the findings suggest that it is advantageous to incorporate these flexibilities into the decision-making process while limiting real-time capabilities.

4.12 Performance analysis

The performance analysis focuses on the Boeing constellation, which comprises 10 altitudes: 7 in LEO, 2 in MEO, and 1 in HEO. The primary objective of this experiment is to evaluate the superiority of the proposed methodology compared to standard practices. Additionally, as a secondary objective, the experiment aims to ascertain the threshold at which the direct approach becomes impractical in terms of dimensionality. Similar to the previous section, the results are juxtaposed against those of an existing single-altitude method [168], utilizing the average and maximum satellite load as evaluation metrics.

The anticipated outcomes are as follows: 1) The direct approach is expected to yield the highest capacity, albeit with a higher computational cost, especially as the number of users increases, and 2) The scalable approach is expected to perform better than the single-altitude method but may underperform compared to the direct method. However, it should remain computationally feasible.

To evaluate these expectations, each approach is executed under the Boeing configuration for varying numbers of users: 100, 500, 1,000, 10,000, and 100,000. The user distribution follows the methodology outlined in Section 3.2. Subsequently, the satellite average and maximum load are computed based on the simulations. The algorithm executions are conducted on 16 cores of

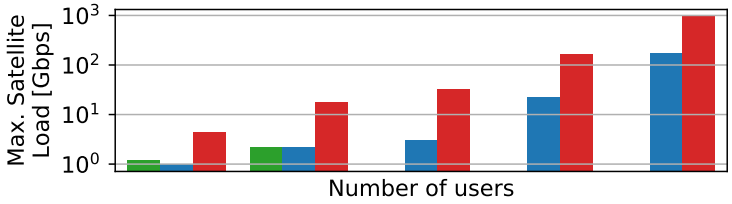


	100	500	1000	10000	100000
Hybrid - Optimal	0.5528	0.5995	-	-	-
Hybrid - Scalable	0.5346	0.6121	0.7291	1.89	11.13
Single-altitude	0.7302	1.044	1.32	3.381	12.72

Figure 4-12: Average satellite load on the proposed methods for the joint Beam Shaping and User Grouping problem. Note that, for high number of users, the direct method does not provide a solution in less than 48h, and, therefore, has not been considered.

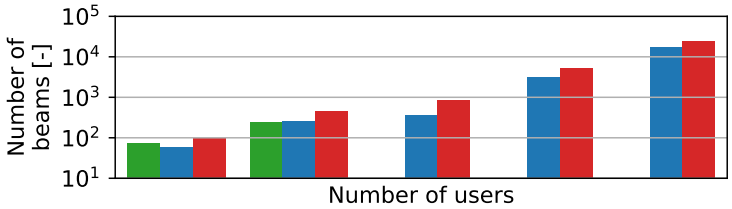
an Intel(R) Xeon(R) Platinum 8160 CPU, with a maximum execution time of 48 hours (172,800 seconds). It is important to note that the initial distribution of users across altitudes needs to be performed only once at the beginning of operations, and subsequently whenever there is a change in the user distribution. Therefore, an extended execution time is permissible.

The results for the average and maximum satellite load are displayed in Figures 4-12 and 4-13, respectively. Additionally, the results illustrating the number of beams and the scaling of computation time against the number of users are presented in Figures 4-14 and 4-15, respectively. One notable observation is that both the direct and scalable approaches exhibit similar performance for low numbers of users (≤ 500). However, the scalable approach demonstrates its superiority by providing commendable results within a feasible computation time even for up to 100,000 users. Secondly, it is evident that the scalable algorithm consistently outperforms the single-altitude method across all user counts in terms of both maximum and average satellite load. This superiority stems from its capability to distribute the demand across various altitudes proportionally to the number of satellites at each altitude, thereby enhancing satellite utilization. Furthermore, the scalable algorithm effectively reduces the number of beams by leveraging diverse footprints in response to the geographical distribution of users. It is important to highlight that the direct algorithm is not able to provide allocations for high number of users due to the computation time.



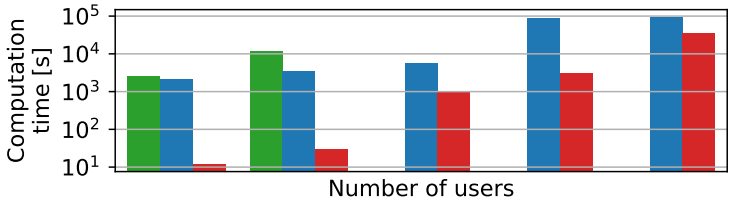
	100	500	1000	10000	100000
Hybrid - Optimal	1.185	2.2	-	-	-
Hybrid - Scalable	1.0	2.18	3.065	22.16	168.6
Single-altitude	4.42	17.76	32.67	164.4	936.3

Figure 4-13: Maximum satellite load on the proposed methods for the joint Beam Shaping and User Grouping problem. Note that, for high number of users, the direct method does not provide a solution in less than 48h, and, therefore, has not been considered.



	100	500	1000	10000	100000
Hybrid - Optimal	76.0	243.0	-	-	-
Hybrid - Scalable	60.0	255.0	372.0	3091.0	17150.0
Single-altitude	99.0	471.0	868.0	5265.0	24280.0

Figure 4-14: Number of beams on the proposed methods for the joint Beam Shaping and User Grouping problem. Note that, for high number of users, the direct method does not provide a solution in less than 48h, and, therefore, has not been considered.



	100	500	1000	10000	100000
Hybrid - Optimal	2545.0	11820.0	-	-	-
Hybrid - Scalable	2155.0	3378.0	5531.0	89650.0	94700.0
Single-altitude	11.92	30.34	994.9	3143.0	36060.0

Figure 4-15: Scaling of the computation time for each algorithm with the number of users considered. Note that, for high number of users, the direct method does not provide a solution in less than 48h, and, therefore, has not been considered.

4.13 Discussion

4.13.1 Inclusion of mobile users

Due to the mobility of users, a fixed ground beam serving them continuously may not suffice if the user movement exceeds the footprint contour coverage area. In such scenarios, two options exist for covering users effectively.

User tracking

This approach involves employing tracking techniques [69] that follow the user movement. Each tracking beam is associated with only the user it serves. To adapt the prior formulation, one beam per mobile user, altitude, and footprint contour is included. The beam center aligns with the user position ($p_b = p_u$). For formulation purposes, a position p is considered in line of sight of beam b if it maintains line of sight at any point along the beam trajectory:

$$\mathbb{1}LoS(p, p_b, \mathcal{A}) = \max t \in \mathcal{T}\, \mathbb{1}_{LoS(p, p_b(t), \mathcal{A})} \tag{4.19}$$

With these adjustments, the previous formulation and resolution methods can be applied.

Grid-like approach

This method involves using fixed ground beams covering the Earth's surface and associating each mobile user with the geographically closest beam. To modify the prior formulation, instead of associating each mobile user with a specific beam ($x_{u,b}$), they are associated with an altitude ($x_{u,\mathcal{A}}$). The specific association at any given time depends on the closest beam to the user and the altitude it corresponds to. The demand of each beam d_b can be computed as the average over all associated users across all time instances. With these adjustments, the prior formulation and resolution methods can be applied once again.

4.13.2 Inclusion of different shapes

It is worth noting that although the formulation assumes conical shapes for simplicity, any shape can be utilized. To integrate the prior formulation with different shapes, the shape $\mathcal{C}$ must be defined first. With this shape, Equations 4.2 and 4.3 facilitate the definition of the footprint and footprint contour for the new shape at each altitude $\mathcal{A}$. Once the new footprint contours are determined, the formulation can be applied once again.

4.13.3 Mapping between users and altitudes

During simulations, users have been assumed to be able to connect to any altitude. However, hardware constraints may prevent certain users from connecting to specific altitudes due to high propagation losses. Adapting the previous formulation involves restricting the mapping between users and beams to only those beams at valid altitudes. Although analyzing the impact of this effect on results is beyond the scope of this dissertation, it poses an interesting avenue for future research.

4.13.4 Inclusion of users without continuous coverage

Depending on the service requested, some users may not require continuous coverage to fulfill their requests. For example, users who only upload server contents to an external database once a day may not need constant connection. However, the current formulation only accommodates users continuously visible to at least one altitude. To incorporate users not needing continuous service, only Equation A.4 requires adaptation:

$$v_{p,\mathcal{A}} = \max_{\mathcal{T}_n \in \mathcal{T}} \min_{t \in \mathcal{T}_n} \sum_{s \in \mathcal{A}} \mathbb{1}_{LoS(p_s,p,\mathcal{A})} \tag{4.20}$$

Where $\mathcal{T}_n$ corresponds to a smaller subset of all times when the position is covered. This subset must be sufficiently large to provide the service requested by the client (e.g., if a customer needs 15 minutes a day to transmit their data, $\mathcal{T}_n$ has to be at least 15 minutes per day of service). To accommodate different types of beams (i.e., beams that can serve continuously versus beams that can only serve for a sub-portion of the time), one set of beams per type of customer is defined. The minimum number of visible satellites per beam ($v_{\mathcal{A},p_b}$) depends on the type of service requested by the users. To ensure accurate mapping of beams to users, users of a specific type can only be associated with beams of the same type. Once the different sets of beams and possible beam-to-user mappings are defined, the formulation and resolution presented previously can be applied again.

4.13.5 Alternative objective functions

In contrast to most literature in satellite communications, the methodology provided does not evaluate results based on spectrum or power usage. This choice is motivated by two factors:

1. Spectrum and power usage are a suitable reference for evaluating resource allocation methods on a single satellite, but they do not fully capture the effect of multiple pools of resources. It is more appropriate to assess the impact of allocation based on the utilization of multiple resource pools rather than just one.

2. No Frequency Assignment or Power Allocation techniques have been assumed. Without allocation for these resources, spectrum and power usage cannot be computed. While existing literature offers approaches to solve these problems [55,112], addressing them within the joint Beam Shaping and User Grouping falls beyond the scope of this dissertation and serves as potential future research.

Other metrics such as latency or quality of service are also not included in the formulation. This decision is justified by arguing that, when assuming homogeneous users, the primary objective is to maximize the number of users served. Exploring the impact of other metrics falls beyond the scope of this work.

4.14 Chapter summary and conclusions

This chapter has proposed a novel methodology to address the joint Beam Shaping and User Grouping problem in hybrid constellations. The new methodology includes the flexibilities of hybrid constellations, i.e., satellites at varying altitudes, in the decision of mapping users, beams, shapes, and altitudes, thereby addressing the research gap found in Section 2.2.1.

4.14.1 Chapter summary

The theoretical part of the chapter commenced with Section 4.1, which delved into the characteristics of hybrid constellations, highlighting the complexities inherent in such designs compared to single-altitude configurations. Subsequently, Section 4.2 explored the geometric properties of the problem, establishing the foundational concepts utilized throughout the chapter. Building upon this groundwork, Section 4.3 introduced a novel mathematical formulation for the joint Beam Shaping and User Grouping problem, elucidating how the two objectives—minimizing the number of beams and balancing load across altitudes—are integrated into the formulation. The formulation was further transformed into a MILP formulation to facilitate solution using mathematical solvers. Concluding the theoretical segment, Section 4.4 rigorously proved the NP-Hardness of the joint Beam Shaping and User Grouping problem, highlighting the impracticality of brute-force and commercial solvers for high-dimensional scenarios.

The applied segment of the chapter began with Section 4.5, outlining the necessary discretizations to tackle the MILP formulation using mathematical solvers, referred to as the direct approach. To address high-dimensional scenarios, Section 4.6 introduced a scalable, sub-optimal approach that decomposed the problem into smaller, more manageable instances. Subsequently, Section 4.7 discussed the computational complexity of both the direct and scalable approaches, indicating the memory and worst-case computation requirements.

Leading into the simulations, Section 4.8 described four distinct experiments and the corresponding constellation configurations. Section 4.9 validated the proposed methodology, demonstrating its capability to yield feasible and valid solutions with desired properties. In a similar vein, Section 4.10 established the feasibility of obtaining high-quality solutions within reasonable timeframes for realistic operational scenarios. Investigating the impact of different objectives, Section 4.11 explored various weighting factors to underscore the relevance of both objectives in maximizing performance.

Concluding the simulation section, Section 4.12 presented performance results of the methodology vis-à-vis existing approaches. Finally, Section 4.13 discussed the significance and applicability of various assumptions, as well as possible relaxations.

4.14.2 Response to Research Questions

Research question 4.1

What are the complexities related to hybrid constellations w.r.t single-altitude designs? What are the mechanisms with which satellite operators can address these complexities?

Satellite operators face two primary complexities: variation in path loss and footprint variation. Addressing path loss variation involves ensuring sufficient link margin or enhancing hardware capabilities, depending on the magnitude of the variation. Footprint variation can be managed by adjusting beam aperture angles according to altitude or by assigning users to specific altitudes. While adjusting beam aperture angles provides operational flexibility, it may not fully leverage the capabilities of satellites at different altitudes during decision-making processes.

Research question 4.2

Can we increase the performance of satellite communications by including hybrid considerations in the resource allocation process?

Traditional single-altitude formulations restrict ground footprints to a uniform shape regardless of altitude. In contrast, hybrid considerations allow for the allocation of multiple shapes and the potential use of smaller beams with higher gains. This flexibility enables more efficient load distribution across satellites, thereby enhancing overall utilization. Although operational flexibility is reduced, empirical evidence indicates that incorporating hybrid constellation flexibilities into resource allocation processes yields higher capacity compared to single-altitude methodologies.

4.14.3 Specific chapter contributions

The specific contributions of this chapter are as follows:

- Formulated the joint Beam Shaping and User Grouping problem as a mixed integer linear problem, and proved NP-hardness.

- Developed a novel methodology to address the joint problem based on splitting the global problem into smaller instances and solving them using commercial solvers.
- Studied the validation, convergence, and performance of the proposed methodology, proving that it provides high quality solutions in feasible time, improving upon standard practice.
- Studied the trade-off between the different objectives, proving that both number of beams and load distribution are relevant when maximizing capacity on the joint problem.
- Discussed assumption relaxation and implications on the methodology.

Chapter 5

Avoiding interference through coordinated Satellite Routing and Frequency Assignment

Avoiding self-interference in large megaconstellations is becoming a priority for satellite operators to maintain seamless operations. With a significantly larger satellite segment compared to previous designs, modern constellations will rely on automated and scalable mechanisms to map users to satellites, while avoiding interference within and between spacecraft. However, as discussed in Section 2.2.9, current research has not proposed techniques to bridge this gap. The objective of this chapter is to address the identified gaps by proposing a coordination framework that integrates Satellite Routing and Frequency Assignment techniques to minimize interference and enhance performance. To that end, this chapter introduces novel approaches for solving the Satellite Routing and Frequency Assignment sub-problems, with the aim of mitigating interference in realistic operational scenarios. Additionally, the chapter evaluates the validity and performance of the proposed coordination framework and provides a detailed analysis on how to integrate the methodology into operations.

The research questions that this chapter aims to address are:

Research question 5.1

How do the Satellite Routing and Frequency Assignment decisions influence the existence of interference between two signals?

Hypothesis: The purpose of this question is exploratory.

Research question 5.2

Can we increase performance in satellite communications by applying optimization to the Satellite Routing and Frequency Assignment problems?

Hypothesis: Yes. Taking smart decisions in the allocation of resources is likely to increase the overall performance as less resources are wasted. Nevertheless, the degree of improvement of optimization is unclear.

5.1 Coordination framework

As discussed in Section 3.4, this dissertation uses the concept of *isolation* to address interference between and within satellites. In particular, isolation hinges on three conditions that ensure two signals will **not** interfere during operations (see Figure 5-1):

1. Condition 1: The signals occupy non-overlapping portions of the frequency spectrum.
2. Condition 2: The signals have opposite polarization.
3. Condition 3: The signals are sufficiently spatially separated (see Section 3.4.2)

It is important to note that the first two conditions are determined exclusively as part of the Frequency Assignment sub-problem, as they pertain to spectrum usage. When focusing solely on the user link and assuming a predetermined beam-user mapping, the third condition depends solely on the geographical positions of the antennas, determined as part of the Satellite Routing problem. Additionally, if one condition is active, the value of the other two becomes irrelevant, as the signals will not interfere. Based on these observations, it is logical to propose a coordinated approach that aims to maintain at least one condition active for all pairs of signals, guaranteeing non-interference. To that end, it is important to define the Satellite Routing and Frequency Assignment sub-problems, and how they might be used to address interference.

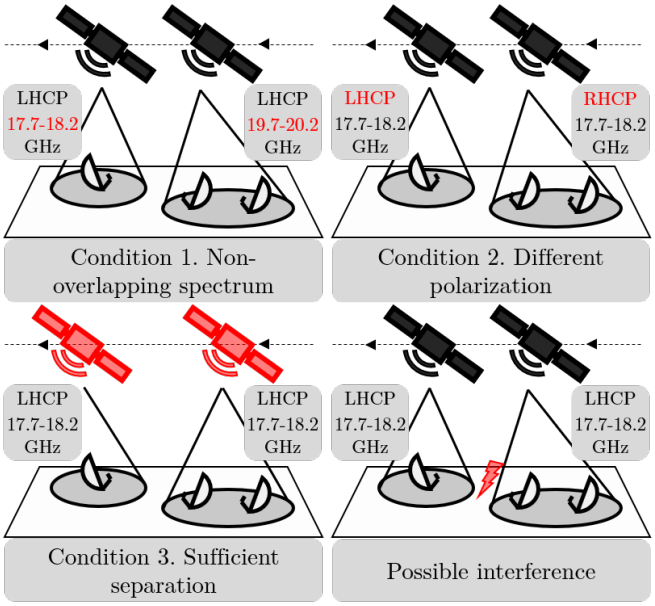


Figure 5-1: Conditions to mitigate possible interference. Replicated from Figure 3-3 for convenience.

On one hand, the Satellite Routing sub-problem involves determining the mapping between beams and satellites to maximize capacity. A solution to this problem is only valid if the assigned satellite is in line of sight of all users within the beam, ensuring coverage. On the other hand, the Frequency Assignment sub-problem entails assigning the central frequency and bandwidth to each beam to maximize capacity. If the system permits frequency reuse or multiple polarizations, these aspects must also be considered in the Frequency Assignment resolution.

Additionally, Frequency Assignment must ensure that the assigned resources do not exceed the technical capabilities of the system. For instance, a satellite cannot reuse a frequency more times than permitted by hardware limitations. However, determining which beam is assigned to which satellite is an integral part of the Satellite Routing sub-problem. This creates an inter-dependency between both problems: while the Satellite Routing decides how many and which beams will be mapped to each satellite, the Frequency Assignment needs to ensure that the assigned resources do not surpass the physical limits of the system.

In summary, the Satellite Routing and Frequency Assignment sub-problems intersect at two focal points: 1) the efficacy of Frequency Assignment depends on decisions made during Satellite Routing resolution, and 2) both problems impact interference conditions. However, since each individual

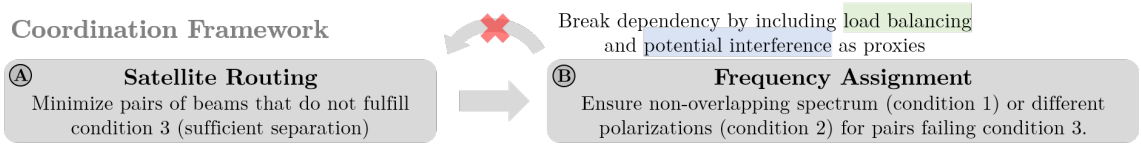


Figure 5-2: Proposed coordination framework for interference mitigation

problem is proven to be NP-Hard [46, 137], addressing both simultaneously at the scale required in modern constellations is computationally intractable. Therefore, the proposed approach is to sequentially resolve both sub-problems: first, Satellite Routing and then Frequency Assignment, leveraging the following observations:

- The expected load of each satellite, which determines the effectiveness of the Frequency Assignment method, can serve as a proxy during Satellite Routing resolution. By finding solutions with a more balanced distribution, the frequency spectrum utilization can be optimized, leading to higher overall capacity.
- Given that both sub-problems impact interference through independent conditions, it is possible to maximize throughput through coordinated efforts. Specifically, the proposed approach involves: 1) Using geographical separation (third condition) as a proxy during Satellite Routing resolution, and 2) When addressing Frequency Assignment, assign non-overlapping spectrum (first condition) or different polarizations (second condition) to all pairs of beams not meeting condition 3.

Based on this, the proposed framework consists of (see Figure 5-2):

A Satellite Routing: the quality of the solution is assessed based on two proxies: 1) the load balance across satellites, and 2) the potential interference between beams. A good solution aims to distribute the beams across satellites while ensuring sufficient geographical separation between beams.

B Frequency Assignment: the resolution procedure must focus on minimizing potential interference (i.e., assigning non-overlapping spectrum or different polarizations) between beams that do not have sufficient geographical separation. Furthermore, a high quality solution must be able to utilize the resources effectively, maximizing capacity.

5.2 Satellite Routing formulation

The Satellite Routing sub-problem involves finding the mapping between beams and satellites at each point in time. Following the framework outlined previously, mappings that distribute the load across satellites and minimize potential interference are preferred. Formally, consider the set of beams $\mathcal{B}$, containing information about each beam position p_b and demand d_b. Additionally, $\mathcal{B}$ may include information about the valid satellites that can serve each beam, denoted as $\mathcal{S}_b$. If provided, it is assumed that beam b can only connect to satellites in $\mathcal{S}_b$, even if other satellites in the constellation are in line of sight. Next, the set of satellites $\mathcal{S}$ contains information about each satellite orbital trajectory in the form of position $p_s(t)$ and velocity $v_s(t)$. If $\mathcal{S}_b$ is not provided, it is assumed that $\mathcal{S}_b = \mathcal{S}$.

To formally formulate this problem, let $x_{b,s}(t)$ be a binary variable indicating if satellite s serves beam b at time $t \in \mathcal{T}$. At any given time, each beam can only be served by one satellite, which can be encoded as: $\sum_s x_{b,s}(t) = 1$. To distribute beams across satellites, $y^A_{b_1,b_2}(t)$ is defined as an auxiliary binary variable determining, at time t, whether b_1 and b_2 are on the same satellite:

$$y^A_{b_1,b_2}(t) = \max_s x_{b_1,s}(t) + x_{b_2,s}(t) - 1 \tag{5.1}$$

By eliminating time dependencies, $y^A_{b_1,b_2}$ denotes if b_1 and b_2 are on the same satellite at some point in time:

$$y^A_{b_1,b_2} = \max_{t \in \mathcal{T}} y^A_{b_1,b_2}(t) \tag{5.2}$$

Similarly, $y^E_{b_1,b_2}(t)$ is an auxiliary binary variable that determines if b_1 and b_2 are not sufficiently separated at time t:

$$\begin{aligned} y^E_{b_1,b_2}(t) &= \max_{s_1,s_2} I_{b_1,s_1,b_2,s_2}(x_{b_1,s_1}(t) + x_{b_2,s_2}(t) - 1) \\ I_{b_1,s_1,b_2,s_2}(t) &= \mathbb{1}_{G_{rel} \leq I_{thres}} \end{aligned} \tag{5.3}$$

Where G_{rel} can be computed based on the definition of isolation in Section 3.4.2. Then, $y^E_{b_1,b_2}$ determines if b_1 and b_2 do not have sufficient separation at some point in time:

$$y^E_{b_1,b_2} = \max_{t \in \mathcal{T}} y^E_{b_1,b_2}(t) \tag{5.4}$$

Then, the Satellite Routing sub-problem consists of finding the values $x_{b,s}(t)$ that minimize the total constraints:

$$\min_{x_{b,s}(t)} \sum_{b_1,b_2} (\omega_A y^A_{b_1,b_2} + \omega_E y^E_{b_1,b_2}) \tag{5.5}$$

Where ω_A and ω_E are the weights of the distribution of load and potential interference, respectively. Then, the complete formulation for the Satellite Routing sub-problem with interference considerations is as follows:

$$\begin{aligned}
\min_{x_{b,s}(t)} \quad & \sum_{b_1 \in \mathcal{B}, b_2 \in \mathcal{B}} (\omega_A y^A_{b_1,b_2} + \omega_E y^E_{b_1,b_2}) \\
s.t. \quad & y^A_{b_1,b_2} = \max_{t \in \mathcal{T}} y^A_{b_1,b_2}(t) && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, t \in \mathcal{T} \\
& y^A_{b_1,b_2}(t) = \max_{s} x_{b_1,s}(t) + x_{b_2,s}(t) - 1 && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, t \in \mathcal{T} \\
& y^E_{b_1,b_2} = \max_{t \in \mathcal{T}} y^E_{b_1,b_2}(t) && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, t \in \mathcal{T} \\
& y^E_{b_1,b_2}(t) = \max_{s_1,s_2} I_{b_1,s_1,b_2,s_2}(x_{b_1,s_1}(t) + x_{b_2,s_2}(t) - 1) && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, t \in \mathcal{T} \\
& \sum_{s} x_{b,s}(t) = 1 && \forall b \in \mathcal{B}, t \in \mathcal{T} \\
& x_{b,s}(t) \in \{0,1\} && \forall b \in \mathcal{B}, s \in \mathcal{S}_b, t \in \mathcal{T} \\
& x_{b,s}(t) \leq \mathbb{1}_{LoS(p_b(t),p_s(t))} && \forall b \in \mathcal{B}, s \in \mathcal{S}_b, t \in \mathcal{T} \\
& y^A_{b_1,b_2}(t), y^E_{b_1,b_2}(t) \in \{0,1\} && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, t \in \mathcal{T} \\
& y^A_{b_1,b_2}, y^E_{b_1,b_2} \in \{0,1\} && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}
\end{aligned} \tag{5.6}$$

Where $LoS(p_b(t), p_s(t))$ determines if satellite s is in LoS of beam b at time t. By including interference considerations, this formulation addresses the gaps found in Section 2.2.9.

Problem linearization

While the previous formulation includes several internal optimizations and non-linear constraints, it can be linearized as follows. In particular, since it is a minmax problem, the maximization operations can be transformed into simple inequalities:

$$
\begin{aligned}
y^A_{b_1,b_2}(t) &\geq x_{b_1,s}(t) + x_{b_2,s}(t) - 1 && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, s \in \mathcal{S}_{b_1} \cap \mathcal{S}_{b_2}, t \in \mathcal{T} \\
y^A_{b_1,b_2} &\geq y^A_{b_1,b_2}(t) && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, t \in \mathcal{T} \\
y^E_{b_1,b_2}(t) &\geq I_{b_1,s_1,b_2,s_2}(x_{b_1,s_1}(t) + x_{b_2,s_2}(t) - 1) && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, s_1 \in \mathcal{S}, s_2 \in \mathcal{S}, t \in \mathcal{T} \\
y^E_{b_1,b_2} &\geq y^E_{b_1,b_2}(t) && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, t \in \mathcal{T}
\end{aligned}
\tag{5.7}
$$

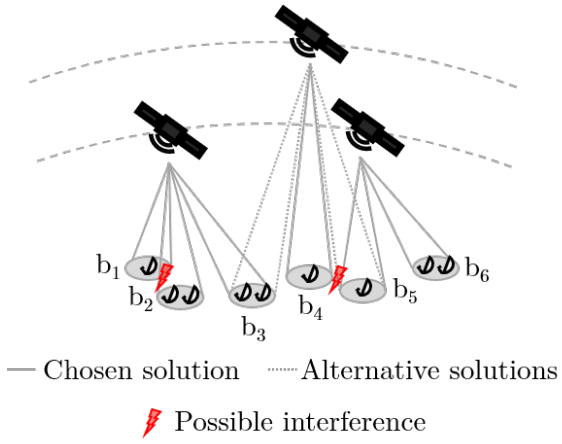
Then, the constraints can be grouped:

$$
\begin{aligned}
y^A_{b_1,b_2} &\geq x_{b_1,s}(t) + x_{b_2,s}(t) - 1 && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, s \in \mathcal{S}_{b_1} \cap \mathcal{S}_{b_2}, t \in \mathcal{T} \\
y^E_{b_1,b_2} &\geq I_{b_1,s_1,b_2,s_2}(x_{b_1,s_1}(t) + x_{b_2,s_2}(t) - 1) && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, s_1 \in \mathcal{S}, s_2 \in \mathcal{S}, t \in \mathcal{T}
\end{aligned}
\tag{5.8}
$$

Thus, the complete linear formulation for the Satellite Routing sub-problem is:

$$
\begin{aligned}
\min_{x_{b,s}} \quad & \sum_{b_1 \in \mathcal{B}, b_2 \in \mathcal{B}} (\omega_A y^A_{b_1,b_2} + \omega_E y^E_{b_1,b_2}) \\
s.t. \quad & y^A_{b_1,b_2} \geq x_{b_1,s}(t) + x_{b_2,s}(t) - 1 && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, s \in \mathcal{S}_{b_1} \cap \mathcal{S}_{b_2}, t \in \mathcal{T} \\
& y^E_{b_1,b_2} \geq I_{b_1,s_1,b_2,s_2}(t)(x_{b_1,s_1}(t) + x_{b_2,s_2}(t) - 1) && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, s_1 \in \mathcal{S}_{b_1}, s_2 \in \mathcal{S}_{b_2}, t \in \mathcal{T} \\
& \sum_s x_{b,s}(t) = 1 && \forall b \in \mathcal{B}, t \in \mathcal{T} \\
& x_{b,s}(t) \in \{0,1\} && \forall b \in \mathcal{B}, s \in \mathcal{S}_b, t \in \mathcal{T} \\
& x_{b,s}(t) \leq LoS(p_b(t), p_s(t)) && \forall b \in \mathcal{B}, s \in \mathcal{S}_b, t \in \mathcal{T} \\
& y^A_{b_1,b_2}, y^E_{b_1,b_2} \in \{0,1\} && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}
\end{aligned}
\tag{5.9}
$$

Note that, although linear, this formulation contains both integer and continuous variables, and thus corresponds to a MILP problem, which is NP-hard. To be used during the Frequency Assignment resolution, $\mathcal{R}_A$ and $\mathcal{R}_E$ are defined as the sets of beam pairs that are at some point in the same satellite, or present potential interference, respectively. These sets can be computed from $y^A_{b_1,b_2}$ and $y^E_{b_1,b_2}$, respectively. An example of a Satellite Routing solution accompanied by the $\mathcal{R}_A$ and $\mathcal{R}_E$ sets is shown in Figure 5-3.



$\mathcal{R}_A$	b_1	b_2	b_3	b_4	b_5	b_6
b_1	-	1	1			
b_2	1	-	1			
b_3	1	1	-			
b_4				-		
b_5					-	1
b_6					1	-

$\mathcal{R}_E$	b_1	b_2	b_3	b_4	b_5	b_6
b_1	-	1				
b_2	1	-				
b_3			-			
b_4				-	1	
b_5				1	-	
b_6						-

Figure 5-3: Satellite Routing example accompanied by the $\mathcal{R}_A$ and $\mathcal{R}_E$ sets

5.3 Satellite Routing proof of NP-Hardness

While an integer linear formulation is always NP-Hard, at this point is unclear whether the problem itself is NP-Hard, or whether there is another formulation that allows to solve the problem optimally in polynomial time. To that end, the following lines prove that the Satellite Routing sub-problem, as described in this work, is NP-Hard by proving that the associated decision problem is NP-Complete. Specifically, it is demonstrated that the decision problem is a generalization of the maximum independent set problem, which is NP-Complete. Using a polynomial time-veryfier, it is concluded that the problem is in NP, proving NP-Completeness.

To prove NP-hardness, an oracle that provides the complete sets $\mathcal{R}_A$ and $\mathcal{R}_E$ for an optimal solution is assumed. Then, each potential beam-to-satellite mapping ($x_{b,s}(t)$) is represented as a vertex in a graph. An edge between two nodes is introduced if any of the following conditions are true:

- Both nodes address the same beam at the same time: $b_1 = b_2$ and $t_1 = t_2$.
- Both nodes address the same satellite at the same time, and the pair is not in $\mathcal{R}_A$: $s_1 = s_2$, $t_1 = t_2$, and $\{b_1, b_2\} \notin \mathcal{R}_A$
- Both nodes present potential interference at the same time, and the pair is not in $\mathcal{R}_E$: $I_{b_1,s_1,b_2,s_2}(t_1) = 1$, $t_1 = t_2$, and $\{b_1, b_2\} \notin \mathcal{R}_E$.

Once the graph is constructed, solving the Satellite Routing problem is equivalent to finding an independent set of size $|\mathcal{B}||\mathcal{T}|$. Since all nodes related to beam b at time t form a clique, this is equivalent to finding the maximum independent set of the graph, which is NP-complete [235]. Since $\mathcal{R}_A$ and $\mathcal{R}_E$ offer an additional level of freedom in the problem, this proves that the problem is NP-Hard.

To prove NP-completeness, it is necessary to prove that, given $x_{b,s}(t)\ \forall b, s, t$, a verifier that checks if $\omega_A|\mathcal{R}_A| + \omega_E|\mathcal{R}_E| \leq k$ for $k \in \mathbb{R}_+$ in polynomial time can be found. This can be accomplished by traversing the graph edges and constructing $\mathcal{R}_A$ and $\mathcal{R}_E$ based on $x_{b,s}(t)$, which can be done in linear time with respect to the edges, and at most quadratic time with respect to the nodes. This proof establishes the Satellite Routing decision problem as being in NP, thus demonstrating its NP-Completeness.

5.4 Satellite Routing approach

Addressing the previous formulation in realistic operational scenarios involves two main complexities: 1) The range of potential solutions scales proportionally with the time horizon, which can extend indefinitely, and 2) The NP-Hardness means that grappling with high-dimensionalities can quickly become computationally intractable. To mitigate these challenges, two simplifications are proposed: focus on only a single time-step at a time, and group the beams into clusters while mapping clusters to satellites. Note that these changes trade computational efficiency for optimality. The following lines detail each one of these simplifications.

5.4.1 Addressing indefinite time horizons

Regarding the previous formulation, the definition of the time variable $\mathcal{T}$ entails two complexities: 1) It is continuous, and 2) It might be arbitrarily large. To address these complexities, the proposed strategy is to discretize time into individual time-steps and addressing one time-step at a time, thereby facilitating manageable computation. Note that this method may yield different values for $y^A_{b_1,b_2}$ and $y^E_{b_1,b_2}$ across iterations. However, for the Frequency Assignment resolution, consistent $\mathcal{R}_A$ and $\mathcal{R}_E$ over time are necessary. To reconcile this, an iterative process is defined, aimed at identifying the smallest, time-independent sets $\mathcal{R}_A$ and $\mathcal{R}_E$, without requiring solutions for every time-step. To achieve this, each time-step selectively considers pairs of beams that either 1) share

Algorithm 4 Addressing the continuous and indefinite time-horizon in the Satellite Routing sub-problem

Input: $\mathcal{B}$ ▷ Set of beams
Input: $\mathcal{S}$ ▷ Set of satellites
Input: $\mathcal{T}$ ▷ Valid times
Input: N_{conv}^{SR} ▷ Iterations without change until convergence
Output: $\mathcal{R}_A, \mathcal{R}_E$ ▷ Set of existing constraints between the different beam pairs
$\mathcal{R}_A = \emptyset$ ▷ Initialize set related to the distribution of beams across satellites
$\mathcal{R}_E = \emptyset$ ▷ Initialize set related to the potential interference between beams
continue = True ▷ Initialize convergence
iterationsWithoutImprovement = 0 ▷ Initialize convergence factor
while continue **do** ▷ While not converged
    Find $t \in \mathcal{T}$ where $\mathcal{T}$ is discrete ▷ Obtain current time-step
    Solve 5.9 for t, fixing $y^A_{b_1,b_2} = 0 \forall \{b_1, b_2\} \in \mathcal{R}_A$ and $y^E_{b_1,b_2} = 0 \forall \{b_1, b_2\} \in \mathcal{R}_E$ ▷ Solve the formulation, ignoring already found pairs
    iterationsWithoutImprovement = iterationsWithoutImprovement + 1 ▷ Increase the counter
    for $b_1 \in \mathcal{B}$ **do** ▷ For each beam
        for $b_2 \in \mathcal{B}$ **do** ▷ For each pair of beams
            if $y^A_{b_1,b_2} \geq 0$ **then** ▷ If there is a new constraint of beams going to the same satellite
                $\mathcal{R}_A = \mathcal{R}_A \cup \{(b_1, b_2)\}$ ▷ Add it to the set
                iterationsWithoutImprovement = 0 ▷ Reset the counter
            if $y^E_{b_1,b_2} \geq 0$ **then** ▷ If there is a new constraint regarding potential interference between beams
                $\mathcal{R}_E = \mathcal{R}_E \cup \{(b_1, b_2)\}$ ▷ Add it to the set
                iterationsWithoutImprovement = 0 ▷ Reset the counter
    if iterationsWithoutImprovement $\geq N_{conv}^{SR}$ **then** ▷ Check convergence
        continue = False

a satellite yet remain outside $\mathcal{R}_A$, or 2) potentially interfere with each other while not belonging to $\mathcal{R}_E$. This can be implemented by adapting Equation 5.8 accordingly:

$$
\begin{aligned}
y^A_{b_1,b_2} &\geq x_{b_1,s}(t) + x_{b_2,s}(t) - 1 && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, s \in \mathcal{S}_{b_1} \cap \mathcal{S}_{b_2}, \underline{\{b_1, b_2\} \notin \mathcal{R}_A} \\
y^E_{b_1,b_2} &\geq I_{b_1,s_1,b_2,s_2}(x_{b_1,s_1}(t) + x_{b_2,s_2}(t) - 1) && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, s_1 \in \mathcal{S}_{b_1}, s_2 \in \mathcal{S}_{b_2}, \underline{\{b_1, b_2\} \notin \mathcal{R}_E}
\end{aligned}
\tag{5.10}
$$

Note that t now represents the current time-step, instead of all time-steps. The idea is to solve the original problem iteratively, until the algorithm converges. The algorithm is deemed to converge after N_{conv}^{SR} iterations without a new pair added to $\mathcal{R}_A$ or $\mathcal{R}_E$. The logic of the iterative approach is highlighted in Algorithm 4. Note that, at this point, there is no further restrictions across time-steps, which means that a beam could switch satellites at each time-step. This aspect will be addressed in Section 5.14.

5.4.2 Addressing an individual time-step

While attempting to directly solve the formulation may yield satisfactory results for scenarios with a low number of beams, scalability concerns arise in high-dimensional contexts. The vast search space inherent in such scenarios can impede the efficient attainment of high-quality solutions. However, the following is observed: once a pair of beams enters $\mathcal{R}_A$, it is irrelevant whether they coincide in the same satellite once, or multiple times. Moreover, optimal solutions are likely to involve aggregating beams already within $\mathcal{R}_A$, as this minimizes the objective function. Building upon this rationale, the proposed approach entails clustering beams into groups and then assigning these groups to satellites. This strategy reduces the complexity of the search space, facilitating faster computation.

The general idea involves assigning each satellite and beam with a cluster, whereby only beams sharing a cluster with a given satellite can be linked to that satellite. The aim of this concept is to help distributing nearby beams into different clusters, which means that they will never reach for the same satellite, thereby reducing the amount of potential pairs in $\mathcal{R}_A$. If appropriately executed, this distribution among clusters can also alleviate potential interference between beams (e.g., by assigning two very close beams to different clusters). The proposed approach comprises two steps: 1) Assign each beam a cluster, and 2) Assign each cluster to a satellite. Note that the first step is a one-time operation, while the second must be repeated for each time-step.

5.4.3 Beam clustering

The proposed approach first assigns each beam to a cluster. The potential number of clusters for each beam corresponds to the minimum count of satellites visible at any given moment (represented by $v_{p,\mathcal{A}}$ in Equation A.4). Put simply, the availability of clusters depends upon the observable satellites. It is important to note that nearby beams may have access to different clusters. However, satellites can only be assigned to one cluster, potentially restricting certain beams from accessing specific satellites if they fall outside their cluster spectrum. To circumvent this limitation, the concept of hierarchical clustering is introduced, as illustrated in Figure 5-4. This clustering framework is structured as a tree, with a single cluster at the root. At each layer, one branch divides into two, generating new clusters.

Beams may have constraints only if they reside on the same branch of the tree. In Figure 5-4,

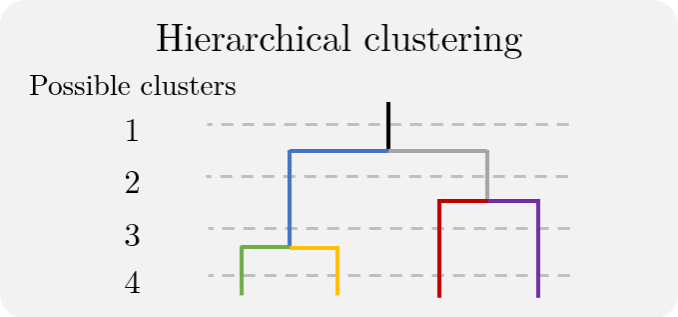


Figure 5-4: Definition of hierarchical clustering for up to 4 clusters

for instance, a beam within the blue cluster might have constraints with beams in the black, green, or yellow clusters, but not with those in the grey, red, or purple clusters. The set $\mathcal{R}_B$ denotes pairs c_1, c_2 where c_1 and c_2 belong to the same branch of the tree and thus may have constraints.

Furthermore, a beam can only be assigned to satellites within the same cluster or a cluster downstream within the same branch. In the illustration, a beam in the blue cluster can solely be associated with a satellite in either the blue, green, or yellow clusters, but not in the black, grey, red, or purple clusters. This restriction ensures that only beams subject to the same constraints can be assigned to the same satellite. The set $\mathcal{R}_D$ comprises pairs c_1, c_2 where the clusters are on the same branch, with c_2 being equal to or further downstream from c_1.

To facilitate the assignment of beams to clusters, a clustering problem is formulated. Initially, $x_{b,c}$ is defined as a binary variable denoting whether beam b is assigned to cluster c. Each beam must be assigned to a single cluster, expressed as $\sum_c x_{b,c} = 1$. It is crucial to note that each beam may only be associated with a subset of clusters $\mathcal{C}_b$, contingent upon the minimum number of satellites visible to beam b ($|\mathcal{C}_b| = v_{p_b,\mathcal{L}}$). As previously mentioned, each cluster has a subset of clusters with which it exhibits constraints, designated as $\mathcal{R}_B$ (where $\{c, c\} \in \mathcal{R}_B \forall c$)).

Now, two beams can share the same satellite, thus forming a pair in $\mathcal{R}_A$, if they are linked with interfering clusters ($c_1, c_2 \in \mathcal{R}_B$) and, at some point in time, they both observe the same satellite:

$$\begin{aligned} z_{b_1,b_2} =& \max_{s,t} \mathbb{1}_{s\in LoS(b_1,t)} \mathbb{1}_{s\in\mathcal{S}_{b_1}} \mathbb{1}_{s\in LoS(b_2,t)} \mathbb{1}_{s\in\mathcal{S}_{b_2}} \\ y^A_{b_1,b_2} \geq& z_{b_1,b_2} \mathbb{1}_{\{c_1,c_2\}\in\mathcal{R}_B} (x_{b_1,c_1} + x_{b_2,c_2} - 1) \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, c_1 \in \mathcal{C}_{b_1}, c_2 \in \mathcal{C}_{b_2} \end{aligned} \tag{5.11}$$

Note that, since z_{b_1,b_2} is solely dependent on the problem characteristics, it can be precomputed.

This implies that the preceding equation enables the identification of beam pairs that could potentially share a satellite at some point in time, without resolving the beam-to-satellite mapping. While this facilitates the computation of beam pairs that may share a satellite, accurately determining potential interference necessitates precise knowledge of the beam-to-satellite mapping. Nonetheless, the conditions for when two beams might be assigned to the same or neighboring satellites based on the distance between their centers can be approximated. To this end, the distance factor d_{b_1,b_2} is defined as follows:

$$\begin{aligned} a_{b_1} &= \max_{b_2} ||p_{b_1} - p_{b_2}|| z_{b_1,b_2} \\ d_{b_1,b_2} &= \left(1 - \frac{||p_{b_1} - p_{b_2}||}{a_{b_1}}\right) z_{b_1,b_2} \end{aligned} \tag{5.12}$$

By definition, if $p_{b_1} = p_{b_2}$, $d_{b_1,b_2} = 1$, and if $z_{b_1,b_2} = 0$, $d_{b_1,b_2} = 0$. Similar to z_{b_1,b_2}, d_{b_1,b_2} can be precomputed. This factor will be used to weight which beams should be mapped to different clusters, thereby reducing the likelihood of potential interference. The formulation for the beam clustering problem is as follows:

$$\begin{aligned} \min_{x_{b,c}} \quad & \sum_{b_1 \in \mathcal{B}, b_2 \in \mathcal{B}} (d_{b_1,b_2} y^A_{b_1,b_2}) \\ s.t. \quad & y^A_{b_1,b_2} \geq z_{b_1,b_2} \mathbb{1}_{\{c_1,c_2\} \in \mathcal{R}_B} (x_{b_1,c_1} + x_{b_2,c_2} - 1) \quad \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, c_1 \in \mathcal{C}_{b_1}, c_2 \in \mathcal{C}_{b_2} \\ & \sum_{c \in \mathcal{C}_b} x_{b,c} = 1 \qquad \forall b \in \mathcal{B} \\ & x_{b,c} \in \{0,1\} \qquad \forall b \in \mathcal{B}, c \in \mathcal{C}_b \\ & y^A_{b_1,b_2} \in \{0,1\} \qquad \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B} \end{aligned} \tag{5.13}$$

The mathematical solvers currently available can effectively handle this as a MILP formulation. The output of this allocation process includes both the beam-to-cluster associations and the set $\mathcal{R}_A$, which contains pairs of beams that might share a satellite at some point in time. Note that, while potential interference is incorporated into the formulation through the distance factor, computing the set $\mathcal{R}_E$ requires precise information on the mapping of beams to satellites.

5.4.4 Cluster-to-satellite mapping

Once the beam-to-cluster mapping is established, the remaining task is to determine the cluster-to-satellite mapping. For this purpose, $x_{s,c}(t)$ is defined as a binary variable describing the mapping between satellite s and cluster c at time t. This mapping is time-dependent, allowing satellites to change clusters over time.

Additionally, it is assumed that each beam b possesses a prioritized list of satellites, denoted as $\mathcal{S}_b^* = s_1, s_2, ..., s_k$. It is presumed that, whenever feasible, the beam will attempt to connect to the first satellite on this list. If not feasible, it will proceed to the second, then the third, and so forth. This list may evolve over time based on the geographical positions of the satellites. For this study, it is assumed that the list corresponds to the set of satellites ordered by descending elevation angle. However, satellite operators may choose to implement alternative policies, such as prioritizing the current satellite servicing the beam to avoid beams switching satellites continuously, are beyond the scope of this work. This assumption aims to streamline the mapping of beams to satellites with matching cluster, thereby simplifying the problem complexity. Specifically, once the mapping of satellites to clusters is known, the beam to satellite mapping can be directly retrieved using the known beam-cluster mapping.

Formally, building upon the original formulation, $x_{b,s}(t)$ represents a binary variable mapping beam b to satellite s at time t. A beam will be mapped to satellite s_i if it shares the same cluster and none of the satellites preceding s_i in the list have a matching cluster:

$$x_{b,s_i} = \sum_{\{c_b,c\}\in\mathcal{R}_D} (x_{s_i,c} - \sum_{j=1}^{i-1} x_{s_j,c}) \forall b \in \mathcal{B}, s_i \in \mathcal{S}_b^* \tag{5.14}$$

Note that the satellites are ordered based on the priority list $\mathcal{S}_b^*$. In this Equation, x_{b,s_i} might attain values lower than 1. It is assumed that beam b is mapped to satellite s_i only when $x_{b,s_i} = 1$. To ensure coverage, each beam must have at least one valid and visible satellite with a matching cluster: $\sum_{\{c_b,c\}\in\mathcal{R}_D} \sum_{s\in\mathcal{S}_b^*} x_{s,c} \geq 1$. Modifying the original formulation, the formulation for the cluster-to-satellite mapping is:

$$
\begin{aligned}
\min_{x_{s,c}} \quad & \sum_{b_1 \in \mathcal{B}, b_2 \in \mathcal{B}} y^E_{b_1,b_2} \\
s.t. \quad & I_{b_1,s_1,b_2,s_2}(t)(x_{b_1,s_1}(t) + x_{b_2,s_2}(t) - 1) - y^E_{b_1,b_2} \leq 0 && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, \\
& && s_1 \in \mathcal{S}_{b_1}, s_2 \in \mathcal{S}_{b_2}, \{b_1, b_2\} \notin \mathcal{R}_E \\
& x_{b,s_i} = \sum_{\{c_b,c\} \in \mathcal{R}_D} (x_{s_i,c} - \sum_{j=1}^{i-1} x_{s_j,c}) && \forall b \in \mathcal{B}, s_i \in \mathcal{S}^*_b \\
& \sum_{\{c_b,c\} \in \mathcal{R}_D} \sum_{s \in \mathcal{S}^*_b} x_{s,c} \geq 1 && \forall b \in \mathcal{B} \\
& \sum_{c \in \mathcal{C}} x_{s,c} = 1 && \forall s \in \mathcal{S} \\
& x_{s,c} \in \{0,1\} && \forall s \in \mathcal{S}, c \in \mathcal{C} \\
& y^E_{b_1,b_2} \in \{0,1\} && \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}
\end{aligned}
\tag{5.15}
$$

Similar to the original formulation, this formulation corresponds to a mixed integer linear problem. Note that the variables that determine the pairs of beams in the same satellite ($y^A_{b_1,b_2}$) do not appear, as they have already been accounted for in the beam clustering. By combining both approaches, a solution for the complete Satellite Routing sub-problem is obtained.

5.4.5 Addressing infeasible scenarios

In certain cases, depending on the geometric properties of the problem, the previous formulation might not yield feasible solutions. This is because the possible clusters for each beam has been decided upon optimistically, leading to no valid solution. For instance, see Figure 5-5 as a reference of an infeasible cluster-to-satellite problem.

To address these cases, a new binary variable is introduced, x_b, which takes the value of 1 when a beam cannot find a valid and visible satellite with a matching cluster, 0 otherwise. This value can be computed using:

$$
\sum_{\{c_b,c\} \in \mathcal{R}_D} \sum_{s \in \mathcal{S}^*_b} x_{s,c} + x_b \geq 1 \tag{5.16}
$$

Based on this, a new formulation is defined as:

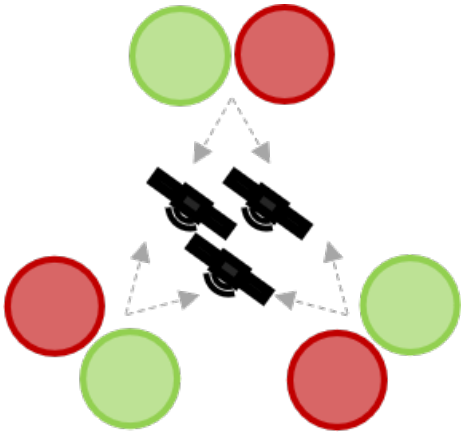


Figure 5-5: Example of impossible cluster-to-satellite

$$
\begin{aligned}
\min_{x_{s,c}} \quad & \sum_{b \in \mathcal{B}} x_b \\
s.t. \quad & x_{b,s_i} = \sum_{\{c_b,c\} \in \mathcal{R}_D} (x_{s_i,c} - \sum_{j=1}^{i-1} x_{s_j,c}) && \forall b \in \mathcal{B}, s_i \in \mathcal{S}_b^* \\
& \sum_{\{c_b,c\} \in \mathcal{R}_D} \sum_{s \in \mathcal{S}_b^*} x_{s,c} + x_b \geq 1 && \forall b \in \mathcal{B} \\
& \sum_{c \in \mathcal{C}} x_{s,c} = 1 && \forall s \in \mathcal{S} \\
& x_{s,c} \in \{0,1\} && \forall s \in \mathcal{S}, c \in \mathcal{C} \\
& x_b \in \{0,1\} && \forall b \in \mathcal{B}
\end{aligned}
\tag{5.17}
$$

This formulation tries to maximize the number of beams with a valid satellites. For any beam that could not be matched to a valid satellite, a *downgrading* process is defined, which reassigns the cluster of the beam to the one immediately above in the hierarchy. Then the previous formulation is re-solved iteratively, until all beams are matched with a cluster. Note that beams assigned to the root cluster will always have valid satellites (assuming that the minimum number of visible beams is always at least 1), so this technique will always converge. The resulting solution $x_{s,c}$ can be used as a warm start for the formulation in Equation 5.15. It is important to note that upon downgrading, the set $\mathcal{R}_A$ must be recomputed, since pairs of beams might now be reassigned to the same satellite. As a note, in the following experiments, around 20% of the beams needed downgrading at some

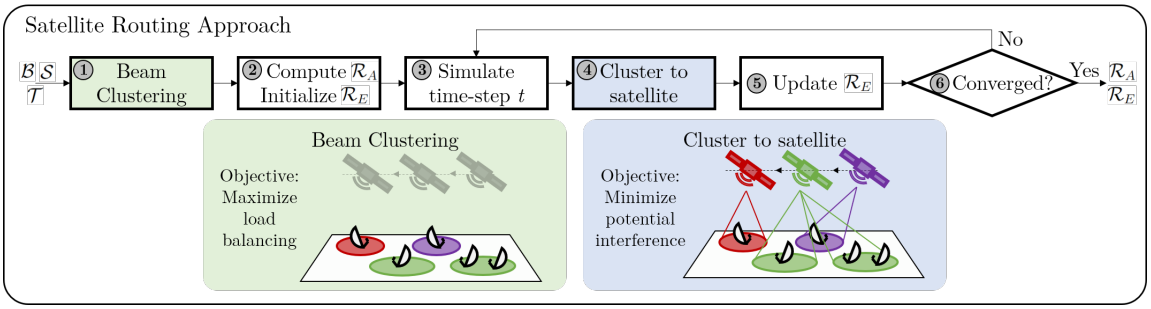


Figure 5-6: Illustration of the Satellite Routing approach

point during the simulation.

5.4.6 Complete approach for the Satellite Routing sub-problem

The proposed approach involves the following sequential steps (see Figure 5-6):

1. Resolve the beam clustering problem using the formulation in Equation 5.13.
2. Compute $\mathcal{R}_A$, initialize $\mathcal{R}_E = \emptyset$, and set t to the initial time-step.
3. Simulate time t, computing relevant geometrical parameters.
4. Resolve the cluster-to-satellite mapping using the formulation in Equation 5.15.
5. Add any new pairs of beams with $y^E_{b_1,b_2} = 1$ to $\mathcal{R}_E$.
6. Upon convergence, finish, otherwise set t to the next time-step and return to step 3.

For the resolution of the formulations (step 1 and 4), commercial mathematical solvers can be used.

5.5 Satellite Routing complexity analysis

To compute the complexity of the proposed formulation and approach, M is defined as the maximum number of visible and valid satellites for a single beam at any given time, computed as $M = \max_{b,t} \sum_s \mathbb{1}_{s \in \mathcal{S}_b} \mathbb{1}_{s \in LoS(b,t)}$. N_{CT} and N_{CB} are also introduced, representing the total number of clusters and the maximum number of clusters a single beam can be mapped to, respectively. The complexity analysis can be performed both on memory and execution time.

Regarding the memory requirement, the original formulation has a complexity of $\mathcal{O}(|\mathcal{B}|M|\mathcal{T}| + |\mathcal{B}|^2)$, where the first term accounts for the decision variables ($x_{b,s}(t)$) and the second term represents the auxiliary variables ($y^A_{b_1,b_2}$ and $y^E_{b_1,b_2}$). For the beam clustering problem, the total necessary memory scales with $O(|\mathcal{B}|N_{CB} + |\mathcal{B}|^2)$. For the cluster-to-satellite problem, the total necessary memory scales with $O(|\mathcal{S}|N_{CT} + |\mathcal{B}|^2)$. It can be seen that memory requirements of the proposed approach is much smaller than the original formulation, as it is not dependent on the time-horizon.

Regarding the computation, the complexity of the search space scales with $O(M^{|\mathcal{B}||\mathcal{T}|})$, as each beam and time requires selecting one (1) variable from a maximum of M options. Since the problem is NP-Hard, this is the expected complexity of the resolution. For the beam clustering problem, the search space scales with $O(N_{CB}^{|\mathcal{B}|})$. For the cluster-to-satellite problem, the search space scales with $O(N_{CT}^{|\mathcal{S}|})$. Assuming N_T iterations are needed for convergence, the total complexity is $O(N_{CB}^{|\mathcal{B}|} + N_T N_{CT}^{|\mathcal{S}|})$. Note that, while the complexity is still exponential with the number of beams, it is much smaller than the original formulation due to the reduction in options per beam (with N_{CB} at most M, often smaller) and by addressing one time-step at a time. It is important to note that the formulations for the proposed approach are also NP-Hard, which can be proven similarly to the original formulation. However, the reduction in complexity is sufficient to be able to be addressed using mathematical solvers, as will be shown in subsequent sections.

5.6 Frequency Assignment formulation

Once a solution for the Satellite Routing sub-problem is obtained, the subsequent task is to address the Frequency Assignment sub-problem. To this end, an existing approach from the literature (reference [55]) that utilizes equivalent definitions of $\mathcal{R}_A$ and $\mathcal{R}_E$ is used. However, given the original method inability to scale to the required dimensionality in realistic operational scenarios, the following lines outline a scalable approach that builds upon the original formulation proposed in [55] to accommodate high-dimensional scenarios. Notably, although initially developed as part of this dissertation, the proposed method is also detailed in Appendix A of [55].

The objective of the Frequency Assignment sub-problem is to assign the initial frequency and bandwidth to each beam. If the system permits frequency reuse or multiple polarizations, these aspects must also be incorporated into the formulation. The solution must adhere to the hardware limitations of the satellite, ensuring they are not violated. Furthermore, the solution should strive

to maximize the total throughput, thereby optimizing performance. It is assumed that each beam retains its spectrum allocation indefinitely.

To formalize this problem, $x_{f,bw,r,p,b}$ is defined as a binary variable indicating whether beam b is assigned the initial frequency f, bandwidth bw, reuse factor r, and polarization p. Notably, f and bw are discretized based on the available channels onboard the spacecraft. The combination of frequency f and bandwidth bw is confined to the spectrum allocation for communication purposes. In other words, if F_i and F_f represent the initial and final frequencies of the frequency band designated for communication, respectively, it must adhere to $F_i \leq f$ and $f + bw \leq F_f$. Moreover, r must be less than the total allowable frequency reuses permitted by the hardware. Finally, the potential values for polarization p are constrained by the available polarization determined by the hardware.

To evaluate the quality of a solution, $l_{f,bw,r,p,b}$ is defined as a metric indicating the suitability of the allocation for each beam b. For the purposes of this study, $l_{f,bw,r,p,b}$ corresponds to the estimated radiated power of beam b when assigned the initial frequency f, bandwidth bw, reuse factor r, and polarization p. While beyond the scope of this work, this parameter can be defined according to various objectives, as demonstrated in [55].

If a pair of beams is included in $\mathcal{R}_A$, it signifies the possibility of them being simultaneously active on the same satellite during operations. Consequently, they cannot occupy the same frequency range, share the same reuse factor, and polarization simultaneously. This constraint can be expressed as follows:

$$\mathbb{1}_{f_1 \leq f_2 + bw_2} \mathbb{1}_{f_2 \leq f_1 + bw_1} (x_{f_1,bw_1,r,p,b_1} + x_{f_2,bw_2,r,p,b_2}) \leq 1 \tag{5.18}$$

Similarly, if a pair of beams is present in $\mathcal{R}_E$, it indicates potential interference between them. In such cases, they must utilize non-overlapping portions of the spectrum or different polarizations to mitigate interference. This requirement can be encoded as:

$$\mathbb{1}_{f_1 \leq f_2 + bw_2} \mathbb{1}_{f_2 \leq f_1 + bw_1} (x_{f_1,bw_1,r_1,p,b_1} + x_{f_2,bw_2,r_2,p,b_2}) \leq 1 \tag{5.19}$$

Note that the equation might look similar, but in this case different frequency reuse factors are

not allowed. Based on this, the complete formulation of the problem is:

$$
\begin{aligned}
\min_{x_{f,bw,r,p,b}} \quad & \sum_{f,bw,r,p,b} (l_{f,bw,r,p,b} - M) x_{f,bw,r,p,b} \\
s.t. \quad & \mathbb{1}_{f_1 \leq f_2 + bw_2} \mathbb{1}_{f_2 \leq f_1 + bw_1} (x_{f_1,bw_1,r,p,b_1} + x_{f_2,bw_2,r,p,b_2}) \leq 1 && \forall f_1, bw_1, f_2, bw_2, \\
& && \{b_1, b_2\} \in \mathcal{R}_A \\
& \mathbb{1}_{f_1 \leq f_2 + bw_2} \mathbb{1}_{f_2 \leq f_1 + bw_1} (x_{f_1,bw_1,r_1,p,b_1} + x_{f_2,bw_2,r_2,p,b_2}) \leq 1 && \forall f_1, bw_1, r_1, f_2, bw_2, r_2, \\
& && \{b_1, b_2\} \in \mathcal{R}_E \\
& \sum_{f,bw,r,p} x_{f,bw,r,p,b} \leq 1 && \forall b \in \mathcal{B} \\
& x_{f,bw,r,p,b} \in \{0, 1\} && \forall f, bw, r, p, b \in \mathcal{B}
\end{aligned}
\tag{5.20}
$$

Where M is a large number with the objective of maximizing the number of active beams, defined as $M = \max_{f,bw,r,p,b} l_{f,bw,r,p,b} |\mathcal{B}|$. This formulation corresponds to a integer linear formulation of the problem.

5.7 Frequency Assignment proof of NP-Hardness

Proving NP-Hardness on the proposed formulation is straightforward. In particular, it is observed that the problem is the optimization version of the independent set problem, which has been proven to be NP-Hard [235]. Specifically, the nodes on the graph correspond to the decision variables, and each constraint represents and edge between a set of nodes.

5.8 Frequency Assignment approach

Understanding the previous formulation is relatively straightforward; however, the computational complexity stemming from the high dimensionality of the number of beams and the myriad combinations of spectrum, reuse factor, and polarization renders the problem computationally challenging. In response, two mechanisms to streamline the search space are proposed: 1) Addressing only a subset of beams simultaneously, and 2) Focusing solely on pertinent sections of the search space. Note that these adjustments no longer guarantee optimality but enable faster computation.

To address a subset of beams simultaneously, an iterative approach is defined wherein only N_{ch}

Algorithm 5 Frequency Assignment approach

Input: $\mathcal{B}$ ▷ Set of beams
Input: $\mathcal{R}_A$ ▷ Pairs of beams that, at some point, coincide in the same satellite
Input: $\mathcal{R}_E$ ▷ Pairs of beams that, at some point, present potential interference
Input: N_{ch} ▷ Number of beams allowed to change per iteration
Input: N_{cutoff} ▷ Number of options allowed per beam and bandwidth
Input: N_{conv}^{FA} ▷ Iterations without change until convergence
Output: f_b, bw_b, r_b, p_b ▷ Initial frequency, bandwidth, reuse factor, and polarization for each beam

$f_b, bw_b, r_b, p_b \forall \mathcal{B}$ = Warm start [168] ▷ Initialize the solution using a warm start
iterationsWithoutImprovement = 0 ▷ Initialize convergence factor
previousObjective = 0 ▷ Initialize previous objective
while iterationsWithoutImprovement $< N_{conv}$ **do** ▷ While not converged
    $b_0 = Rand(\mathcal{B})$ ▷ Select one beam randomly
    $\mathcal{B}' = Neighbors(b_0, N_{ch})$ ▷ Select N_{ch} neighbors to b
    $\mathcal{X} = \emptyset$ ▷ Initialize pool of variables
    for $b \in \mathcal{B}'$ **do** ▷ For each non-fixed beam
        Find $l_{f,bw,r,p,b}$ for each valid option ▷ Find objective value for valid options
        $vars = RankAndCutOff(l_{f,bw,r,p,b}, N_{cutoff})$ ▷ Restrict options to the most interesting ones
        if $Valid(f_b, bw_b, r_b, p_b)$ **then** ▷ If current solution is valid
            $vars = vars \cup \{(f_b, bw_b, r_b, p_b)\}$ ▷ Add it to the pool
        $\mathcal{X} = \mathcal{X} \cup vars$
    Solve 5.20 only with variables in $\mathcal{X}$, fixing $\mathcal{B} \setminus \mathcal{B}'$ ▷ Solve the reduced problem
    for $b \in \mathcal{B}'$ **do**
        $f_b, bw_b, r_b, p_b = Solution$ ▷ Update the values with the new solution
    if currentObjective == previousObjective **then**
        iterationsWithoutImprovement = iterationsWithoutImprovement + 1 ▷ Increase the counter
    else
        iterationsWithoutImprovement = 0
    previousObjective = currentObjective

beams are permitted to change in each iteration. Determining which beams undergo modification in a single iteration involves randomly selecting one beam from the pool and including its $N_{ch} - 1$ closest geographic neighbors. The algorithm reaches convergence when there are no changes in the objective function for N_{conv}^{FA} consecutive iterations. Initially, the algorithm initializes with the solution proposed by [168].

To focus on the interesting part of the search space, at each iteration, $l_{f,bw,r,p,b}$is computed for all possible **valid** options of beam b. An option is only valid if its assignment is not constricted by the fixed beams. Next, for each valid bandwidth bw, each option is ranked based on the value of $l_{f,bw,r,p,b}$. Following this, a cutoff is applied, eliminating all options below the N_{cutoff} rank for each bandwidth. Furthermore, only three bandwidth options are considered: reducing the bandwidth by one channel, maintain current bandwidth, or increase bandwidth to the next interesting bandwidth, i.e., the one that reduces the objective function. With these two changes, the total available options are reduced for each beam to $3N_{cutoff}$. In addition, the current assignment, if existent, is also

considered as a possibility. The pseudocode for the complete algorithm is shown in Algorithm 5.

5.9 Frequency Assignment complexity analysis

The complexity analysis of the proposed formulation and approach can be performed at two levels: on the memory side, and on the computation side. For this analysis, N_{BW} is defined as the total number of available channels, N_R as the frequency reuse factor, and N_P as the number of polarizations.

Regarding the memory requirement, the complexity is $\mathcal{O}(|\mathcal{B}|N_{BW}^2 N_R N_P + |\mathcal{B}|^2)$ for the original formulation, as it has one variable for each option per beam, and one constraint for each pair of beams in $\mathcal{R}_A$ and $\mathcal{R}_E$, which can be at most size $|\mathcal{B}|^2$. Compared to this, the proposed approach has a complexity of $\mathcal{O}(N_{ch}N_{cutoff} + N_{ch}|\mathcal{B}|)$, which evolves from two factors: 1) for the variables, the number of options per beam is limited to $3N_{cutoff}$, and the number of beams is limited to N_{ch}, and 2) for the constraints, there are at most $N_{ch}|\mathcal{B}|$ constraints regarding pairs of beams where at least one is allowed to change.

Regarding the computation, the size of the search space is $\mathcal{O}((N_{BW}^2 N_R N_P)^{|\mathcal{B}|})$ for the original formulation, since each beam can choose from, at most, $N_{BW}^2 N_R N_P$ options. This scaling factor is equivalent to the complexity of the problem since it has proven to be NP-Hard. For the proposed approach, assuming that the algorithm needs N_T as the iterations needed for convergence, the total complexity is $\mathcal{O}(N_T(3N_{cutoff})^{N_{ch}})$. Note that, while this value is still NP-Hard by a similar reasoning as the original one, the complexity of the problem fed into the commercial solver can be regulared by adjusting the values of N_{cutoff} and N_{ch}.

5.10 Experimental set-up

The following lines detail the validation and performance analysis on the proposed methodology. In particular, 4 different experiments that test the capabilities of the novel formulation are executed, summarized in Table 5.1.

To conduct the experiments, it is necessary to define a realistic set of users $\mathcal{U}$ and a representative constellation configuration. For the set of users, Section 3.2 provides details on how the user distribution is generated in this dissertation. For the constellation configuration, two existing

Analysis	Objective	Constellation	N_{loc}	$N_{us/loc}$
Validation & Verification	Demonstrate that the proposed methodology produces valid solutions	SpaceX	5,000	1
Convergence	Demonstrate that the proposed methodology produces high quality feasible solutions in reasonable time	SpaceX	10,000	1
Performance	Analyze the performance with respect to established techniques	O3b	2,000	1
		O3b	20,000	1
		SpaceX	20,000	1
		SpaceX	20,000	10
Operations	Study the impact of the proposed methodology during operations	SpaceX	5,000	10

Table 5.1: Experimental set-up to evaluate the proposed methodology on the coordinated Satellite Routing and Frequency Assignment.

constellation designs have beem chosen:

- O3b Constellation: This chapter utilizes a compact version of the O3b mPower constellation, featuring 10 satellites in equatorial orbits at 8,062 km. Despite its smaller size compared to other designs, it allows for a comparison with previous methodologies developed for equatorial, single-plane constellations. Each user has a maximum of two satellites in line of sight at any given time, which should lead to reduced benefit when using an optimized Satellite Routing approach. For this constellation, only users within $+/-50^{\circ}$ latitude are considered.

- SpaceX constellation: The configuration for this constellation corresponds to the one outlined in Table 1.1. As depicted in Figure 5-7 (extracted from [8]), each user maintains line of sight with between 20 and 50 satellites, irrespective of time. An optimized Satellite Routing approach can leverage this flexibility to significantly enhance system capacity.

For a standard comparison, it is assumed that only circular beam shapes with aperture angle of 2° are allowed. The users are organized into beams using the algorithm outlined in [168]. Furthermore, the hyperparameters for all experiments are detailed in Table 5.2. All simulations have used the commercial solver Gurobi [234] (version 9.1.2) in an Intel(R) Xeon(R) Platinum 8160 CPU @ 2.10GHz, allowing up to 16 simultaneous threads.

To evaluate the efficacy of the proposed methodology, a comparative analysis against established methods is conducted. Specifically, two distinct Satellite Routing techniques are examined: a heuristic approach, where each beam is directed towards the satellite with the highest elevation

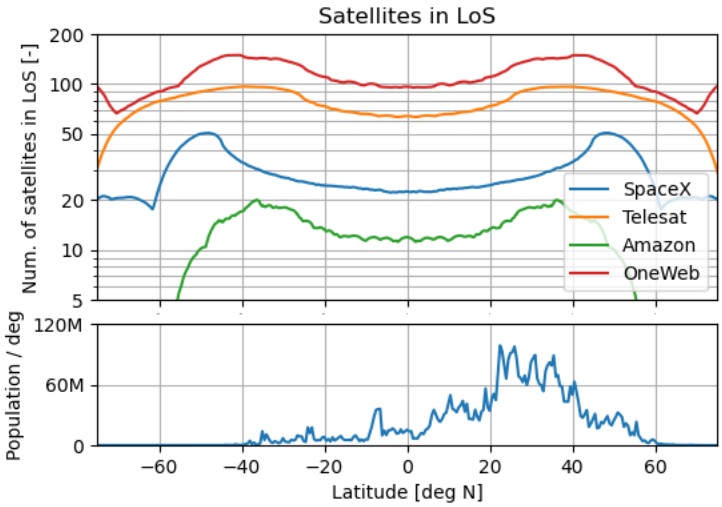


Figure 5-7: Number of satellites in lines of sight as a function of the latitude for SpaceX, Telesat, Amazon, and OneWeb. Image extracted from [8] with permission of the authors.

Parameter	Value	Parameter	Value	Gurobi Parameter	Value
N_{conv}^{SR}	10	N_{ch}	100	Threads	16
N_{conv}^{FA}	15	N_{cutoff}	25	MIPGap	10^{-4}

Table 5.2: Summary of the parameters of the simulation.

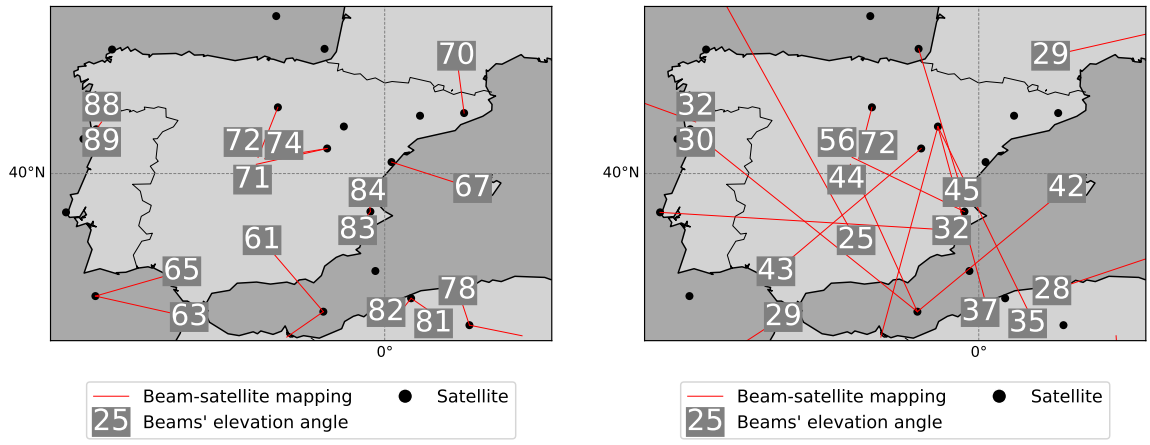
angle, as outlined in references [236, 237], and an optimized strategy employing PSO to attain an optimal solution, as described in reference [137]. In the case of the heuristic method, the sets $\mathcal{R}_A$ and $\mathcal{R}_E$ are derived through iterative simulation of the constellation, identifying beam pairs within the same satellite and those potentially susceptible to interference, until no new pairs are discovered for N_{conv}^{SR} iterations. Note that the PSO approach is specifically tailored for equatorial, single-plane constellations, rendering it inapplicable to alternative configurations. Additionally, regarding the Frequency Assignment, a water-filling approach as delineated in reference [168] is included. A summary of the diverse implementations is presented in Table 5.3.

5.11 Validation and verification analysis

The validation and verification analysis aims to determine the validity of the results derived from the proposed methodology. Given that the proposed Frequency Assignment method is essentially an

Satellite Routing	Frequency Assignment
Maximum elevation angle	Water filling [168]
Particle Swarm Optimization [137]	Integer Optimization
Integer Optimization (this dissertation)	(this dissertation, also in [55])

Table 5.3: Summary of the implementations used.



(a) Highest elevation angle

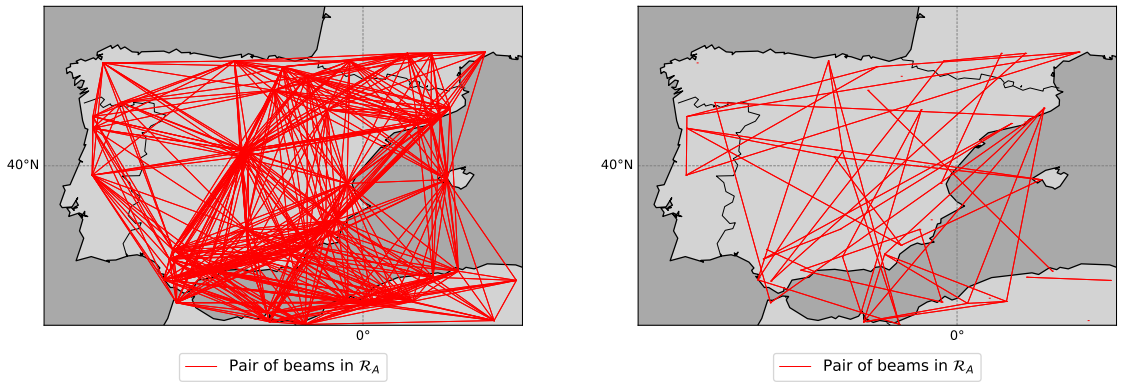
(b) Proposed method

Figure 5-8: Mapping between beams and satellites under the SpaceX constellation with 5,000 users. Each line corresponds to a connection between a beam and a satellite. For clarity, only half of the beams are shown.

adaptation of an existing approach [55], the validation and verification outcomes remain consistent. Consequently, this section exclusively concentrates on the outcomes generated by the Satellite Routing approach. Specifically, the objectives are as follows:

1. Feasibility of the solution: Ensuring that each beam establishes connections solely with satellites within line of sight. Practically, this entails that the elevation angle of the beam center w.r.t. the satellite exceeds the minimum elevation angle stipulated by the operator (i.e., 25° for SpaceX).

2. Attainment of desired properties: Confirming that the solution effectively reduces the number of beams serviced by the same satellite and mitigates potential interference, in comparison to established techniques. To achieve this, a comparative analysis is conducted against the maximum elevation angle heuristic, focusing on the sizes of sets $\mathcal{R}_A$ and $\mathcal{R}_E$.



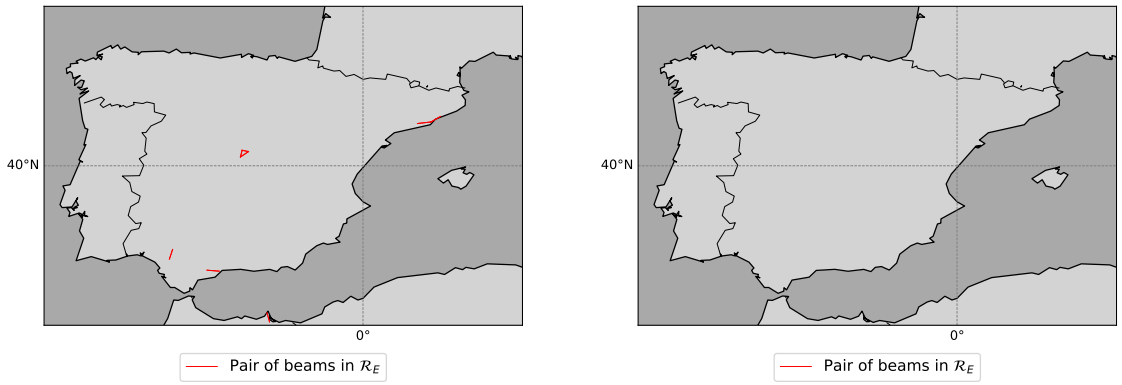
(a) Highest elevation angle

(b) Proposed method

Figure 5-9: Beam pairs assigned to the same satellite under the SpaceX constellation with 5,000 users. Each line corresponds to a pair of beams in $\mathcal{R}_A$, where the origin and destination correspond to the beam center of each beam. Only pairs where both beams fall within the limits of the image are shown.

To fulfill this objective, a simulation of the SpaceX constellation with 5,000 users is conducted. Figure 5-8 illustrates the correspondence between beams and satellites for both the highest elevation heuristic and the proposed approach, along with the elevation angle of each beam. As anticipated, the highest elevation heuristic yields higher elevation angles compared to the proposed method. Notably, all elevation angles surpass the minimum threshold set by the operator, affirming the solution validity.

However, opting for the satellite with the highest elevation angle results in larger sets $\mathcal{R}_A$ and $\mathcal{R}_E$. This disparity is evident in Figures 5-9 and 5-10, depicting the beam pairs within $\mathcal{R}_A$ and $\mathcal{R}_E$ respectively, overlaid on a map of the Iberian peninsula. By dispersing the beams across satellites, the proposed approach markedly diminishes the occurrence of beam pairs mapped to the same satellite or those with potential interference. Specifically, Table 5.4 highlights that the novel methodology achieves a fourfold reduction in the size of $\mathcal{R}_A$ and reduces the size of $\mathcal{R}_E$ by two orders of magnitude. This substantiates that the proposed approach indeed yields the desired solution, thereby validating the technique.



(a) Highest elevation angle

(b) Proposed method

Figure 5-10: Beam pairs with potential interference under the SpaceX constellation with 5,000 users. Each line corresponds to a pair of beams in $\mathcal{R}_E$, where the origin and destination correspond to the beam center of each beam. Only pairs where both beams fall within the limits of the image are shown. On the right image, no pair of beams was found.

Method	Number of pairs in $\mathcal{R}_A$	Number of pairs in $\mathcal{R}_E$
Highest elevation angle	318,942	1,756
Proposed method	83,105	26

Table 5.4: Number of pair of beams served by the same satellite or with potential interference. The results shown correspond to a simulation of the SpaceX constellation with 5,000 users.

5.12 Convergence analysis

The convergence analysis aims to track the evolution of sets $\mathcal{R}_A$ and $\mathcal{R}_E$ over time, providing insights into whether the algorithm has reached convergence and how its performance might be affected by time constraints. Similar to the preceding section, this analysis is confined to the Satellite Routing technique.

Figure 5-11 illustrates the progression of sets $\mathcal{R}_A$ and $\mathcal{R}_E$ over computation time for the SpaceX constellation with 10,000 users. Notably, pairs of beams directed towards the same satellite ($\mathcal{R}_A$) predominantly emerge during the beam clustering phase, while pairs potentially subject to interference ($\mathcal{R}_E$) primarily surface in the satellite-to-cluster phase of the methodology. This deliberate partitioning is intentional: by segmenting the beam-to-satellite problem into these distinct phases, the computation of objectives is separated, enabling faster resolution. Note that resolving the beam

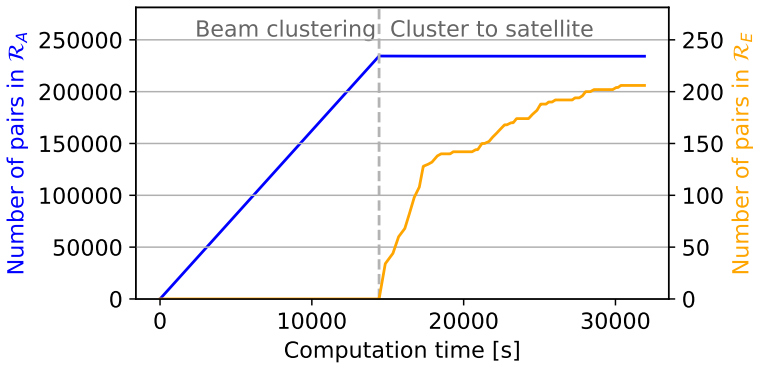


Figure 5-11: Evolution of sets $\mathcal{R}_A$ and $\mathcal{R}_E$ over computation time using the proposed approach. The left portion of the graph corresponds to the time spent on the beam clustering, while the right portion corresponds to the satellite to cluster mapping.

clustering phase demands a substantial time investment but provides all constraints simultaneously, whereas the satellite-to-cluster phase relies on simulations, providing only a fraction of beam pairs in each iteration. It can be seen that the algorithm is deemed to converge when the solution stabilizes.

5.13 Performance analysis

The objective of this analysis is to elucidate the performance advantages and efficiency of the cooperative approach compared to previously established techniques. To achieve this, simulations for two constellations are conducted: a scaled-down version of the O3b mPower constellation, serving 2,000 and 20,000 users respectively, and the SpaceX Starlink constellation, accommodating 20,000 and 200,000 users.

Figure 5-12 illustrates the simulation outcomes concerning throughput and transmission power for the O3b mPower constellation with 2,000 and 20,000 users, as well as the SpaceX Starlink constellation with 20,000 and 200,000 users. Additionally, Table 5.5 provides detailed insights into key performance metrics for scenarios involving 20,000 and 200,000 users in the mPower and Starlink constellations, respectively.

The first aspect to highlight is that the proposed coordination framework (depicted by the blue dot in Figure 5-12) achieves the highest throughput for the SpaceX constellation, but not for the O3b mPower constellation. The reason for that lies in two factors. On one hand, a low number of visible

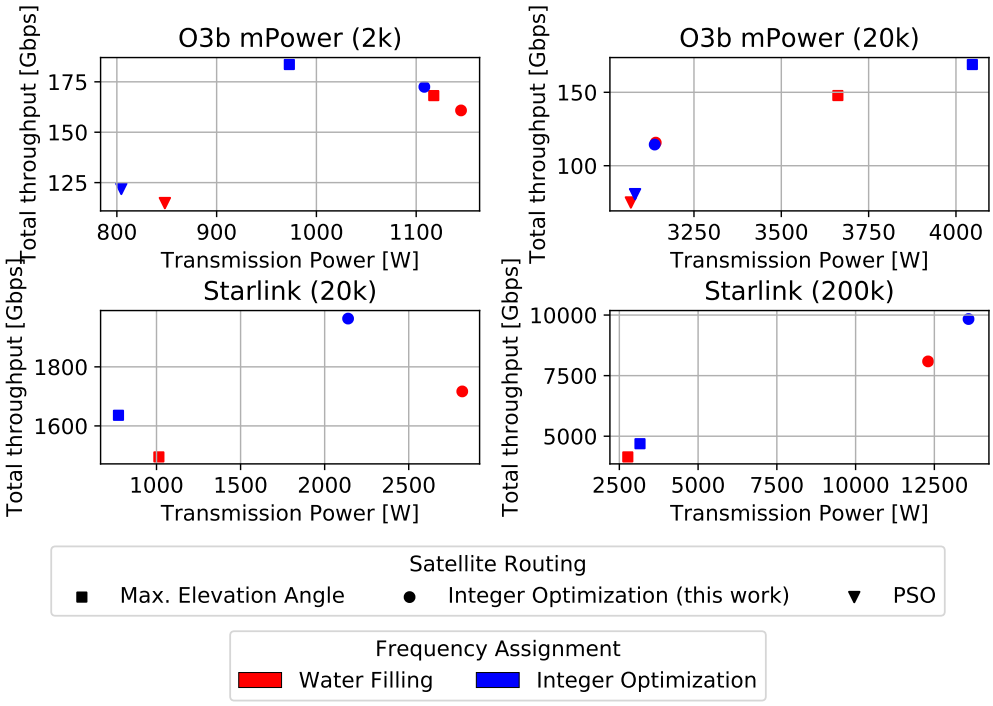


Figure 5-12: Throughput and transmission power for different combinations of interference-aware implementations. The two upper images correspond to simulations on the O3b mPower constellation with 2,000 and 20,000 users, respectively. The two lower images corresponds to simulations on the SpaceX constellation with 20,000 and 200,000 users, respectively

satellites (e.g., 1 to 2 in O3b mPower) entails a lower flexibility during the Satellite Routing, leading to reduced benefit when using optimization. On the other hand, a complex Satellite Routing implies higher transmission power compared to heuristics like maximum elevation angle, as the distance between antennas is higher and, therefore, the free space loss. The reduced benefit combined with the increased transmission power leads to the fact that, for small constellations, heuristics perform slightly better than the proposed optimization techniques.

As anticipated, the impact of the proposed framework is more pronounced within the SpaceX constellation. This heightened effect is attributable to the added flexibility afforded by thousands of satellites, enabling a more effective distribution of demand. Consequently, this leads to an increase in total throughput of up to 137%, all without necessitating additional payload. Notably, this enhancement stems from the utilization of more satellites concurrently, reaching up to 1,480, rather than from a substantial increase in utilization on individual satellites. It is important to

acknowledge that these improvements come at the expense of increased transmission power, which increases by almost a factor of 5.

It is important to note that employing individual optimization techniques in isolation may yield undesirable outcomes. For instance, relying solely on the Satellite Routing technique proposed in this work (represented by the red dot) can lead to increased transmission power and potentially diminished throughput, as users might be assigned to satellites farther away. Similarly, utilizing only the Frequency Assignment approach proposed herein may result in reduced transmission power but could also lead to lower throughput. This discovery challenges existing literature, which predominantly focuses on individual techniques rather than integrated approaches. The results underscore the necessity of a coordinated methodology to attain superior performance.

Constellation	Satellite Routing	Frequency Assignment	Avg. Throughput [Gbps]	Avg. Active Sats.	Total Power [kW]	Avg. Power on active sat. [W]	Satellite utilization [%] (all)	Satellite utilization [%] (active)	Spectrum usage [GHz]	Computation Time [s]
O3b mPower	Max. El.	WF	148	10	3.66	366	21.6	21.6	262	384
	Max. El.	IO (new)	**169**	10	4.05	405	24.6	24.6	290	479
	IO (new)	WF	116	10	3.14	314	16.9	16.9	224	2360
	IO (new)	IO (new)	114	10	3.14	314	16.7	16.7	224	2450
	PSO	WF	74.9	10	3.07	307	10.9	10.9	219	558
	PSO	IO (new)	80.6	9.8	3.08	314	11.8	12.0	221	644
SpaceX Starlink	Max. El.	WF	4150	937	2.75	2.93	1.87	8.79	2170	99700
	Max. El.	IO (new)	4690	896	3.13	3.50	2.12	10.4	2810	100000
	IO (new)	WF	8090	1460	12.3	8.42	3.66	11.0	4750	117000
	IO (new)	IO (new)	**9830**	1480	13.6	9.19	4.45	13.2	7290	118000

Table 5.5: Results for various figures of merit for the simulations of the O3b mPower and SpaceX constellations with 20,000 and 200,000 users, respectively. Each combination of Satellite Routing and Frequency Assignment implementations is represented in a row. Acronyms: Max. El. - Maximum Elevation Angle; IO - Integer Optimization; PSO - Particle Swarm Optimization; WF - Water Filling.

In the case of the SpaceX constellation, employing the proposed methodology for Satellite Routing results in a higher increase in spectrum usage compared to individual Frequency Assignment techniques. This suggests that distributing beams across satellites yields greater benefits than attempting to optimize the performance of each individual satellite. When combining both approaches, the total spectrum usage can be multiplied by over a factor of three. However, due to the limited number of spacecraft, these observations do not hold true for the mPower constellation. Furthermore, it is noteworthy that the proposed framework outperforms the particle swarm optimization technique introduced in reference [137], underscoring its effectiveness and superiority, and demonstrating its potential for real-world implementation.

5.14 Operations implications

The objective of this section is to analyze the implications of implementing the proposed methodology into real operations. To that end, three aspects are studied: 1) How and where should the computation of the solution be performed, 2) When should the information be transmitted to the satellites through telemetry, and 3) What is the size of the updates necessary.

5.14.1 Centralized computation

The proposed methodology relies on a centralized computation system, where the algorithm is executed on a single machine, potentially located on Earth, and subsequently distributed to the network through ground-to-satellite or satellite-to-satellite links. However, the two components of the system require varying levels of computation:

- Satellite Routing: requires continuous updates for each new time-step to be considered. Given that the solution is computed centrally and propagated to all satellites, it is necessary to ensure that the computation time remains sufficiently low to maintain smooth operation. Note, however, that each time-step can be computed well in advance, so it is only necessary that the sequential time between time-steps, if any, is maintained below the time per time-step, on average.

- Frequency Assignment: requires only one computation, as the solution remains valid indefinitely. Therefore, the computation is not constrained by specific operational timelines.

5.14.2 Telemetry updates

After a solution has been computed, it needs to be transmitted to the spacecraft. The two components imply different levels of telemetry:

- Satellite Routing: involves three blocks of information. First, the cluster for each beam must be computed and transmitted once. This can be executed prior to operations or during low-telemetry intervals to ensure operations remain unaffected. Second, the priority list for each beam and time ($\mathcal{S}_b$) needs to be computed and transmitted at every time-step. If the computation solely relies on geometrical parameters (such as using the highest elevation angle), it can be performed on-board without telemetry. Other alternatives might impact operations, but these are beyond the scope of this work. Finally, the cluster for each satellite and time must also be computed and transmitted at every time-step. Since this information is derived from centralized computation, it cannot be computed on-board and must be accounted for during telemetry operations. However, note that it is not necessary to transmit the full mapping of clusters to satellites at each instance; only changes in the cluster need to be transmitted. Given the likelihood of obtaining similar solutions in sequential time-steps, it is important to analyze the frequency of changes to ensure smooth operations.
- Frequency Assignment: This comprises a single block of information containing the initial frequency, bandwidth, reuse factor, and polarization of each beam. This information can be transmitted prior to operations to ensure operations are not altered.

5.14.3 Operations formulation

During operations, the satellite operator must compute and transmit the Satellite Routing solution for each time-step to the satellites. While two sequential time-steps may exhibit similar geometrical conditions, there is no guarantee that the solution will be similar when using the proposed methodology. In fact, due to the symmetries of the Satellite Routing formulation, multiple solutions may be valid at each point, and they may differ significantly from one another. To address this issue, a slight modification of the previous formulation is proposed, aiming to minimize changes between time-steps while ensuring solution feasibility.

Specifically, during operations, given a time t, the first step is to compute $x_{s,c}^{t-1}$ and $x_{s,c}^{t}$ by solving Equation 5.15 for time $t-1$ and time t, respectively. Assuming a complete set $\mathcal{R}_E$ has been

obtained prior to operations, both solutions will guarantee that $\sum_{b_1 \in \mathcal{B}, b_2 \in \mathcal{B}} y^E_{b_1,b_2} = 0$. Next, an auxiliary binary variable x_s is introduced which takes the value of 1 when the cluster of satellite s changes from time $t-1$ to time t. x_s can be computed using the following equation:

$$x_s \geq x^{t-1}_{s,c} - x^t_{s,c} \forall c \in \mathcal{C} \tag{5.21}$$

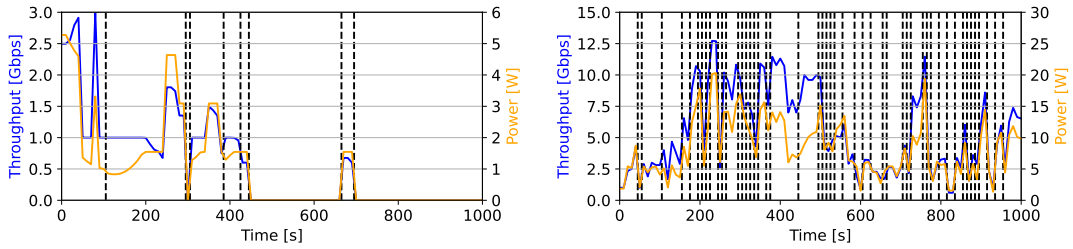
Now, to minimize the number of changes, the formulation of the problem becomes:

$$\begin{aligned}
\min_{x^t_{s,c}} \quad & \sum_{s \in \mathcal{S}} x_s & & \\
s.t. \quad & y^E_{b_1,b_2} \geq I_{b_1,s_1,b_2,s_2}(t)(x_{b_1,s_1} + x_{b_2,s_2} - 1) & & \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B}, \\
& & & s_1 \in \mathcal{S}_{b_1}, s_2 \in \mathcal{S}_{b_2}, \{b_1, b_2\} \notin \mathcal{R}_E \\
& \sum_{c \in \mathcal{C}} x^t_{s,c} = 1 & & \forall s \in \mathcal{S} \\
& x_{b,s_i} = \sum_{\{c_b,c\} \in \mathcal{R}_D} (x^t_{s_i,c} - \sum_{j=1}^{i-1} x^t_{s_j,c}) & & \forall b \in \mathcal{B}, s_i \in \mathcal{S}^*_b \\
& \sum_{\{c_b,c\} \in \mathcal{R}_D} \sum_{s \in \mathcal{S}^*_b} x^t_{s,c} \geq 1 & & \forall b \in \mathcal{B} \\
& x_s \geq x^{t-1}_{s,c} - x^t_{s,c} & & \forall c \in \mathcal{C} \\
& y^E_{b_1,b_2} = 0 & & \forall b_1 \in \mathcal{B}, b_2 \in \mathcal{B} \\
& x^t_{s,c} \in \{0,1\} & & \forall s \in \mathcal{S}, c \in \mathcal{C} \\
& x_s \in \{0,1\} & & \forall s \in \mathcal{S}
\end{aligned} \tag{5.22}$$

Note that $x^{t-1}_{s,c}$ is considered fixed in this formulation. Furthermore, the original formulation can be used to compute a warm-start for this problem. This approach reduces the amount of cluster changes between consecutive operations, optimizing the stability and continuity of the satellite constellation performance over time.

5.14.4 Operations simulation

To provide a clearer understanding of the implications of implementing the proposed methodology during operations, the SpaceX constellation with $N_{loc} = 10,000$ and $N_{us/loc} = 10$ is simulated.



(a) Satellite 0 (7 changes)

(b) Satellite 748 (49 changes)

Figure 5-13: Throughput and transmission power for two satellites during operations. The dashed lines represent changes in the cluster to satellite mapping. The left image corresponds to the first satellite, while the right image corresponds to the satellite with the maximum number of changes.

Simulation Parameters		**User Parameters**	
Total simulation time	1,000s	N_{loc}	5,000
Time Step	10s	$N_{us/loc}$	10
Results			
Total number of changes	21,365	Total number of handovers	121,669
Total computation time	990s	Total number of handovers using maximum elevation	87,594

Table 5.6: Summary of parameters and results in the operations simulation

The objective of this simulation is to analyze the telemetry requirements and computation time required for the proposed approach.

To address this, Figure 5-13 focuses on two particular satellites in the constellation (the first one and the one with the most changes), depicting the evolution of throughput and transmission power alongside the time placement of cluster-to-satellite changes. As anticipated, when a satellite changes cluster, it results in a significant variation in throughput and transmission power, as all connections to the beams need to be switched. Table 5.6 provides specific results from the simulation. For notation, a handover is the moment a user switches between satellites.

During 1,000 seconds of operations, the satellites switched clusters a total of 21,365 times. Assuming clusters can be encoded using 2 Bytes of information and the time at which the cluster changes can be encoded using 8 Bytes, this implies a total of 1,709 bps to transmit the changes to the whole constellation. Furthermore, this translates to an average of around 21 changes per second across the constellation or 206 seconds per change on a single satellite. Note that this is in the same order of magnitude compared to the average visibility time for each satellite at any given

point on Earth, which is around 4 minutes (240 seconds). For the satellite with the most changes, it changed clusters approximately every 20.4 seconds on average. Note that operators could further reduce the number of handovers by choosing a different policy regarding the priority list of each beam, such as one that prioritizes the current satellite. Studying such policies, however, falls out of the scope of this study.

Comparing with the heuristic approach of reaching the highest elevation angle satellite, the results indicate an increase of approximately 40% in the number of handovers, but they remain within the same order of magnitude. Note that these numbers indicate a handover every 27 seconds for each beam, on average. Additionally, note that the computation time is lower than the simulation time, enabling the method to sustain continuous operations. It is important to note that these computations can be performed well in advance, so single time-steps consuming more time than required or temporary outages on the telemetry link should not constrain operations. These findings underscore the feasibility and practicality of the proposed methodology for real-world applications, ensuring efficient resource utilization and maintaining optimized performance over time.

5.15 Chapter summary and conclusions

This chapter has proposed a novel methodology to address interference in the context of large megaconstellations. The new methodology is based on a coordination framework between Satellite Routing and Frequency Assignment techniques, which addresses interference both within and between satellite, thereby addressing the research gap found in Section 2.2.9.

5.15.1 Chapter summary

The theoretical part of the chapter began with Section 5.1, which examined the relationship between the Satellite Routing and Frequency Assignment sub-problems and proposed a coordination framework to mitigate interference. Subsequently, Section 5.2 delved into the Satellite Routing sub-problem, presenting a novel formulation aimed at distributing load across satellites while minimizing interference. The NP-Hardness of this formulation was established in Section 5.3 by demonstrating its generalization of the maximum independent set problem. Given the complexity of solving this problem in high-dimensional scenarios, Section 5.4 outlined a decomposition approach that divided the problem into beam clustering and cluster-to-satellite mapping sub-problems. Addition-

ally, this section elucidated how each sub-problem can be addressed using commercial mathematical solvers. Section 5.5 concluded the Satellite Routing methodology by evaluating the formulation and approach complexity in terms of memory and computation requirements.

Transitioning to the Frequency Assignment sub-problem, Section 5.6 presented a novel formulation based on selecting the optimal option for each beam using binary variables. The NP-Hardness of this formulation was established by comparing it to the independent set problem in Section 5.7. To tackle high-dimensional scenarios, Section 5.8 proposed an iterative algorithm that reduced the search space by ranking options and focusing solely on the best ones. Finally, Section 5.9 concluded the Frequency Assignment discussion by examining the complexity of the formulation and proposed approach.

The practical section of the chapter commenced with Section 5.10, which meticulously outlined the simulation conditions and experimental setups for subsequent sections. Following this, Section 5.11 validated the proposed approach by demonstrating its feasibility and its ability to achieve the desired objectives. Similarly, Section 5.12 studied the convergence of the approach, affirming its capability to yield high-quality solutions within reasonable timeframes. Through comparative analysis with existing techniques, Section 5.13 showcased the superiority and necessity of the coordination framework in maximizing system performance. Finally, Section 5.14 delved into the implications of applying the proposed methodology in real operational scenarios, underscoring its feasibility to maintaining continuous operations.

5.15.2 Response to Research Questions

Research question 5.1

How do the Satellite Routing and Frequency Assignment decisions influence the existence of interference between two signals?

Satellite Routing influences the distance between signals, as well as the relative position of the satellites, thereby impacting their relative strength at reception. Conversely, Frequency Assignment dictates spectrum usage and signal polarization, influencing signal overlap.

Research question 5.2

Can we increase performance in satellite communications by applying optimization to the Satellite Routing and Frequency Assignment problems?

Employing optimized techniques to address both problems has resulted in reduced system capacity in small constellations, but up to 137% capacity increase in large megaconstellations. The variance in these outcomes is attributed to the operational flexibility in large systems, increasing efficacy of optimized methods. While these methods necessitate increased telemetry and handovers, this study demonstrates that they remain within operational constraints, allowing seamless operations.

5.15.3 Specific chapter contributions

The specific contributions of this chapter are as follows:

- Developed a novel coordination framework based on Satellite Routing and Frequency Assignment to address interference within and between the satellites.
- Formulated the Satellite Routing sub-problem as a mixed integer linear problem, and proved NP-hardness.
- Developed a novel methodology to address the Satellite Routing sub-problem by decomposing it into two smaller sub-problems and solving them using commercial solvers.
- Formulated the Frequency Assignment sub-problem as a mixed integer linear problem, and proved NP-hardness.
- Developed a novel methodology to address the Frequency Assignment sub-problem by reducing the search space to interesting solutions and solving the problem iteratively using commercial solvers.
- Studied the validation, convergence, and performance of the proposed methodology, proving that it provides high quality solutions in feasible time, improving upon standard practice.
- Studied the applicability of the novel methodology when applied to realistic operational scenarios, discussing aspects related to computation and telemetry.

Chapter 6

Designing ground infrastructure and ISL architecture through Gateway Routing

With the proliferation of satellites in modern constellations, designing the appropriate ground and satellite network infrastructure has become a complex and important issue. Satellite operators now manage numerous ground stations and thousands of gateway antennas [8], which are essential for providing connectivity to the satellites. Moreover, modern satellites can exchange information among themselves using inter-satellite links (ISLs), further complicating operations. In such a scenario, operators rely on automated and scalable mechanisms to route information from the ground to the satellites and among the satellites. Current studies, as highlighted in Section 2.2.10, offer routing strategies for transmitting data through the network, employing techniques akin to those used in terrestrial networks. However, given the large number of satellites and gateways, determining the appropriate mapping between satellites and gateways remains a crucial yet unresolved issue. Additionally, as discussed in Section 2.4, the impact of this mapping on design aspects such as the number of gateways per ground station or the ISL architecture has not been thoroughly explored. This chapter aims to address these gaps by: 1) proposing a novel methodology to determine the optimal mapping between gateways and satellites, considering fixed and flexible beam-to-gateway

constraints, and 2) investigating the impact of this mapping on the required number of gateways per ground station and the ISL architecture to maximize performance. This chapter focuses solely on network design aspects, such as ground infrastructure and ISLs design, and does not delve into orbital design considerations.

The research questions this chapter aims to address are:

Research question 6.1

How do the ground infrastructure and ISL architecture affect the constellation performance?

Hypothesis: The purpose of this question is exploratory.

Research question 6.2

Does the capacity scale linearly with the amount of ground station locations?

Hypothesis: No. Having more ground station locations implies a larger supply of information, which should translate into a higher capacity. However, diminishing returns are expected as the number of ground stations increase.

Research question 6.3

When changing the ISL configuration (i.e., the number of links and the satellite each link connect to), what is the best ISL configuration that maximizes throughput and quality of service, while minimizing the required ground infrastructure?

Hypothesis: Is it likely that using a deeply interconnected ISL architecture will lead to a higher overall performance. In situations with limited hardware, operators have traditionally relied on simple network designs. It is unclear if a more complex design will offer benefits with respect to simpler architectures.

6.1 Designing supply infrastructure for satellite communications

The objective of this chapter is to examine the impact and design trade-offs of various architectures for ground and space network infrastructure. Specifically, the aim is to analyze the effects of the

number of gateways per ground station and the chosen ISL architecture on system performance, aspects that have not yet been thoroughly explored in the literature. To achieve this, it is crucial to understand how these elements contribute to performance.

Both the ground infrastructure and ISL architecture influence the *supply* side of the constellation, affecting the capacity of each satellite to deliver the necessary connectivity. This supply side is contingent on three main factors:

The first aspect to consider is the spatial distribution of demand on the sky, determined by 1) the spatial location of the satellites, which rests on the orbital design of the constellation; 2) the distribution of users; 3) the policy mapping users to satellites, established during the Satellite Routing resolution; and 4) the ISL architecture, dependent on the satellite payload design. When combining these factors, designers can estimate the supply required on the sky to meet the desired performance criteria. For the purpose of this chapter, the orbital constellation design and user distribution are assumed given. Studying alternative methods for the Satellite Routing sub-problem has been addressed in Chapter 5. As highlighted in Section 2.4, studying different ISL architectures has not yet been addressed, and will be a core contribution of this chapter.

The second aspect to consider is the availability of potential locations that meet logistical and performance requirements. These requirements typically encompass factors such as the availability of fast broadband connectivity, adherence to governmental regulations, terrain and physical constraints, and favorable weather conditions. In addition to evaluating possible locations, it is also important to consider the number of gateway antennas per location, as this directly influences supply availability. By combining these two factors, designers can estimate the supply provided by the ground infrastructure. Since studying possible ground station locations, often referred to as the Gateway Placement problem, lies beyond the scope of this dissertation, this chapter will rely on a subset of historically chosen sites, as outlined in Section 3.6. As highlighted in that section, while literature offers insights into how to select the placement of ground stations, determining the number of gateways per station has not been extensively studied and will represent a key contribution of this chapter.

The third and final aspect to consider is the policy governing the mapping of gateways to satellites. This policy must accommodate the physical constraints of each site and anticipate the expected link quality when connecting to the satellites. This aspect enables designs to bridge the ground supply with the satellite demand. Given the lack of attention to this aspect in existing

literature, as demonstrated in Section 2.2.10, this chapter will develop a novel methodology to address this problem. This methodology will be adapted to accommodate fixed and flexible beam-to-gateway mappings, tailored for ground and space network design.

6.2 Problem formulation

Before delving into the various design aspects related to ground and space network design, it important to develop an efficient algorithm for mapping satellites and gateways, enabling the system to maximize performance. This decision is commonly addressed as part of the Gateway Routing sub-problem, which is itself a component of the RAP. Formally, the Gateway Routing sub-problem aims to determine the optimal flow of information through the satellite-terrestrial network to maximize total throughput while maintaining an appropriate quality of service. An illustration of the Gateway Routing sub-problem is depicted in Figure 6-1. While some solutions for Gateway Routing have recently been proposed [7], these techniques do not treat the mapping as a variable to optimize and do not accommodate fixed beam-gateway relations. Moreover, they overlook potential self-interference, leading to overly optimistic solutions. To bridge this gap, the following lines outline a novel formulation for the Gateway Routing sub-problem, aiming to determine the most efficient satellite-gateway mapping at each point, while considering self-interference and allowing fixed beam-gateway constraints.

For the purposes of this dissertation, and as detailed in Section 3.5.2, it is assumed that the satellites possess onboard processing capabilities and that the payload is regenerative. This implies that once the data reaches the satellite, it is decoded, potentially fragmented, and transmitted through the satellite network in a manner similar to a traditional terrestrial network. The routing mechanisms for information within the network have been extensively studied in previous literature [139] and are therefore not the focus of this section. Furthermore, this section assumes the utilization of optical ISLs, which do not interfere with ground-to-space or space-to-ground communications.

6.2.1 Problem set-up

Formally, consider the set of beams $\mathcal{B}$, containing information regarding the demand d_b and the satellite $s_b(t)$ serving beam b at time t. Additionally, $\mathcal{B}$ might contain information about the ground station $g_b(t)$ serving beam b at time t. If not explicitly specified, it is presumed that beam b can

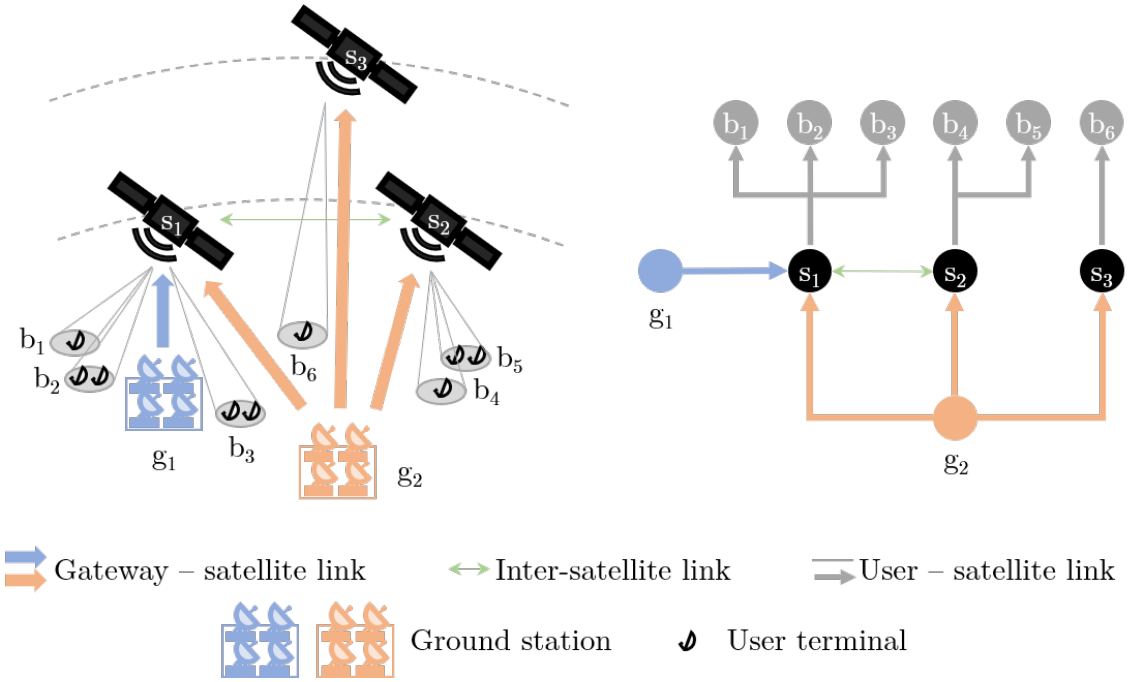


Figure 6-1: Gateway Routing illustration and equivalent graph

establish a connection with any ground station accessible through the satellite network. In addition, set of satellites $\mathcal{S}$ contains information about the orbital position of the satellite at each point $p_s(t)$. Finally, the set of ground stations $\mathcal{G}$ contains information about the ground station location p_g, the specifications of the gateway antennas, and the maximum allowed simultaneous gateway antennas, denoted as N_g. This latter restriction comes from the minimum necessary angular separation between gateways to ensure seamless operations.

To formulate the problem, the following formulation focuses on a specific time instance t. Then, binary variable $x_{g,s}$ is defined as the mapping between ground station g and satellite s at time t. Note that a satellite can only connect to a ground station if it is in line of sight of the station, which can be encoded according to the LoS definition in Appendix A:

$$C1 : x_{g,s} \leq \mathbb{1}_{LoS(p_g, p_s, a_s)} \tag{6.1}$$

Each ground station is connected to, at most, N_g satellites simultaneously, which is ensured by

defining the following constraint:

$$C2 : \sum_{s \in \mathcal{S}} x_{g,s} \leq N_g \tag{6.2}$$

The total number of ground stations a single satellite can connect to is limited by the number of gateway antennas onboard the spacecraft, denoted as N_{ant}. This can be encoded as:

$$C3 : \sum_{g \in \mathcal{G}} x_{g,s} \leq N_{ant} \tag{6.3}$$

To optimize the satellite-gateway mapping for efficiency, it is important to estimate the flow within the satellite network. To that end, it is assumed that each beam can connect to a specific ground station through designated *paths*. Here, a path represents a sequence of satellites directly linked via ISLs. For the forward link, the starting satellite denotes the one connected to the ground station, while the final satellite indicates the one linked to the beam. Conversely, for the return link, this order is reversed. For simplicity, the following formulation exclusively addresses the forward link, but extending it to the return link follows a similar approach. In cases where the constellation lacks ISLs, all paths are necessarily of length 1 (i.e., initial and final satellites must be the same).

Given the dynamic nature of the network, the set of available paths evolves over time. Formally, a path is represented as $p^k_{b,g} = s_1, s_2, ..., s_b$, where k indexes the path. The collection of paths connecting beam b and ground station g is denoted as $\mathcal{P}_{b,g}$. Notably, in scenarios involving ISLs, the potential number of paths between b and g can be infinite due to possible cycles within the satellite network. To mitigate this, only the complete set of shortest paths, in terms of hop count, connecting b and g are considered. In instances where a beam must specifically link to a particular ground station (e.g., customers connecting to their designated gateway), only paths originating from that ground station are considered. This allows for fixed mappings between beams and ground stations.

Now, $y^k_{b,g}$ is defined as a continuous variable ranging from 0 to 1, indicating the proportion of flow directed from ground station g to beam b via path $p^k_{b,g}$. It is necessary to ensure that the flow proportion directed to beam b does not exceed 100%:

$$C4 : \sum_{g,k} y^k_{b,g} \leq 1 \tag{6.4}$$

Furthermore, the flow can only be positive if the first satellite in the path is connected to the ground station. To satisfy this limitation, $s^k_{b,g}$ is defined as the first satellite in $p^k_{b,g}$ (i.e., the satellite receiving the data directly from the gateway). Now, the flow might be positive only when the path is valid:

$$C5 : y^k_{b,g} \leq x_{g,s^k_{b,g}} \tag{6.5}$$

Next, the flow of ground station g towards satellite s is bounded by the maximum data-rate given by the link budget equation $c_{g,s}$. This information can be computed using the real-time conditions, or estimated through historical data. With $s^k_{b,g}$, the total flow of ground station g to satellite s is defined as:

$$f_{g,s} = \sum_{b,k} d_b y^k_{b,g} \mathbb{1}_{s^k_{b,g}=s} \tag{6.6}$$

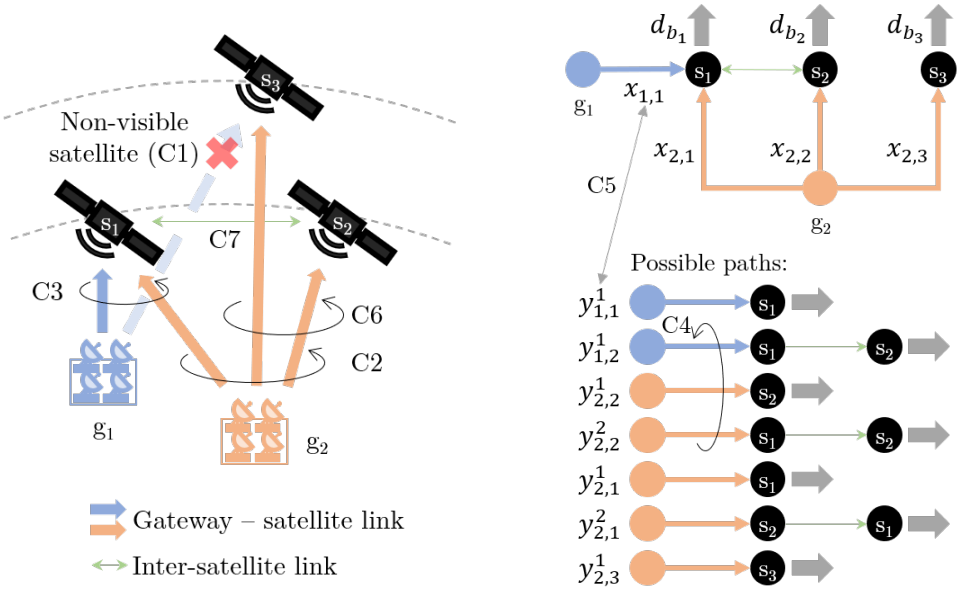
In the context of megaconstellations, it is possible that the signal from two nearby satellites connecting to nearby ground stations might interfere. To compute when two signals might interfere, I_{g_1,s_1,g_2,s_2} is defined as a binary parameter indicating if the isolation of the signals from g_1 to s_1 and g_2 to s_2 surpasses a specific threshold, as defined in Section 3.4. To ensure non-interfering signals, it is assumed that, if two signals present potential interference, each signal will receive bandwidth proportional to their data-rate. Now, it is ensured that the flow does not surpass the physical limitations of the system by using the following constraint:

$$C6 : \sum_{g',s'} I_{g,s,g',s'} \frac{f_{g',s'}}{c_{g',s'}} \leq 1 \tag{6.7}$$

Note that the possible interference of a signal with itself is always 1. This equation limits the possible total flow of interfering signals. Note that, when the signal does not present potential interference with any other signal, the flow is limited to the total capacity of the link.

Another aspect to be considered is that the total flow from satellite s_1 to satellite s_2 through ISLs cannot surpass the capacity of the available ISLs, denoted as c_{ISL}. The pairs of satellites connected by ISLs are collected in set $\mathcal{R}_S$. This set is determined by the ISL architecture. The limited capacity of ISLs can be ensured by:

$$C7 : \sum_{s_1,s_2,b,g,k} d_b y^k_{b,g} \mathbb{1}_{\{s_1,s_2\} \in p^k_{b,g}} \leq c_{ISL} \tag{6.8}$$



(a) The Gateway Routing problem

(b) Routing graph and paths

Figure 6-2: Gateway Routing illustration and equivalent graph, with examples of constraints. It is assumed that each satellite is connected to only one beam, with the same index as the satellite

Any solution meeting the aforementioned constraints is deemed valid for the Gateway Routing sub-problem as outlined in this study. As an example, Figure 6-2 illustrates the Gateway Routing problem, and provides specific examples regarding the variables and constraints. However, it is important to note that not all solutions hold equal appeal from the operator perspective. Moreover, the objectives of the operator may vary depending on the deployment phase. During operational phases, for instance, operators might prioritize maximizing the overall throughput of the constellation or minimizing user latency. Conversely, during the design phase, operators might focus on exploring trade-offs between the required number of gateways and throughput. To enable this, the subsequent lines delineate various constraints and objectives that could be considered across different deployment phases.

6.2.2 Possible figures of merit

For operators tackling the Gateway Routing sub-problem, operators are interested in three aspects: maximizing the total throughput of the constellation, enhancing the quality of service for users,

and minimizing the required ground infrastructure.

Total Throughput

On a first level, satellite operators are drawn to solutions that maximize the overall capacity of the system. This enables them to accommodate a larger user base, consequently increasing the economic gain of the constellation. Additionally, operators may stipulate a minimum throughput threshold, denoted as t_{min}, which the constellation must attain. Capacities surpassing t_{min} hold greater appeal, while values falling below t_{min} are deemed unacceptable. The surplus capacity above t_{min} is denoted as t_{extra}. To integrate throughput considerations into the formulation, both an objective and a constraint are incorporated:

$$\max_{x_{g,s}, y^k_{b,g}} t_{extra} \tag{6.9}$$

$$\sum_{b,g,k} d_b y^k_{b,g} - t_{min} - t_{extra} = 0 \tag{6.10}$$

Quality of service

Following total throughput, another significant metric of interest is the quality of service experienced by users. This aspect holds particular importance for operators as satisfied users contribute to commercial and economic gains. A crucial determinant of service quality in Internet connectivity provision, alongside total throughput, is reception latency. In this dissertation, overall latency is approximated by the product of each path length (represented as $|p^k_{b,g}|$) and the flow volume traversing the path:

$$\min_{x_{g,s}, y^k_{b,g}} \sum_{b,g,k} d_b y^k_{b,g} (|p^k_{b,g}| - 1) \tag{6.11}$$

Ground Infrastructure

Lastly, operators are concerned with the total ground infrastructure required to sustain constellation operations. By optimizing the number of gateway antennas per ground station, operators can reduce the constellation total cost with minimal impact on overall capacity, leading to a higher economic benefit. Note that while gateway antennas initially entail lower costs compared to establishing entirely new ground stations, each additional terminal introduces heightened operational and

coordination expenses, due to the potential interference across geographically proximate antennas.

Similar to throughput considerations, operators may stipulate a minimum number of gateways for each ground station, denoted as $N_{g,min}$. This value could stem from business-related constraints associated with ground station construction or existing infrastructure from previous constellations. Importantly, this value may vary depending on the specific ground station. The surplus number of gateway antennas required per ground station is denoted as $N_{g,extra}$. Consequently, ground infrastructure considerations are integrated into the formulation through both an objective and a constraint:

$$\min_{x_{g,s}, y^k_{b,g}} \sum_g N_{g,extra} \tag{6.12}$$

$$\sum_{s \in \mathcal{S}} x_{g,s} - N_{g,min} - N_{g,extra} \leq 0 \tag{6.13}$$

Note that this objective is only relevant when studying different network infrastructures, as it will be fixed during operations.

6.2.3 Linear formulation

Based on the three figures of merit, the complete formulation for the Gateway Routing sub-problem is defined as:

$$
\begin{aligned}
\min_{x_{g,s},y^k_{b,g}} \quad & -\omega_T t_{extra} + \omega_L \sum_{b,g,k} d_b y^k_{b,g}(|p^k_{b,g}|-1) + \omega_N \sum_g N_{g,extra} & & \\
s.t. \quad & \sum_{b,g,k} d_b y^k_{b,g} - t_{min} - t_{extra} = 0 & & \\
& \sum_{s\in\mathcal{S}} x_{g,s} - N_{g,min} - N_{g,extra} \leq 0 & \forall g \in \mathcal{G} & \\
& \sum_{g',s'} I_{g,s,g',s'} \frac{f_{g',s'}}{c_{g',s'}} \leq 1 & \forall g \in \mathcal{G}, s \in \mathcal{S} & \\
& \sum_{b,g,k} d_b y^k_{b,g} \mathbb{1}_{\{s_1,s_2\}\in p^k_{b,g}} \leq c_{ISL} & \forall \{s_1,s_2\} \in \mathcal{R}_S & \\
& \sum_{s\in\mathcal{S}} x_{g,s} \leq N_g & \forall g \in \mathcal{G} & (6.14)\\
& \sum_{g\in\mathcal{G}} x_{g,s} \leq N_{ant} & \forall s \in \mathcal{S} & \\
& \sum_{g,k} y^k_{b,g} \leq 1 & \forall b \in \mathcal{B} & \\
& y^k_{b,g} \leq x_{g,s^k_{b,g}} & \forall b \in \mathcal{B}, g \in \mathcal{G}, k & \\
& x_{g,s} \in \{0,1\} & \forall g \in \mathcal{G}, s \in \mathcal{S} & \\
& x_{g,s} \leq 1_{LoS(p_g,p_s,a_s)} & \forall g \in \mathcal{G}, s \in \mathcal{S} & \\
& 0 \leq y^k_{b,g} \leq 1 & \forall b \in \mathcal{B}, g \in \mathcal{G}, k &
\end{aligned}
$$

Where ω_T, ω_L, and ω_N represent the weights of the three different objectives. The presented formulation corresponds to a mixed integer linear formulation of the Gateway Routing sub-problem, which belongs to the family of NP-Hard problems. The main outcome of this problem is the mapping of satellites to gateways encoded in $x_{g,s}$, since the real flow of information, estimated with $y^k_{b,g}$, is dependent on the specific routing protocol implemented, outside the scope of this dissertation.

6.3 Proof of NP-Hardness

While a mixed integer linear formulation is always NP-Hard, at this point it is unclear whether the problem itself, as described in this work is NP-Hard, or whether there exists another formulation that allows for optimal resolution in polynomial time. To that end, the following lines prove that

the Gateway Routing sub-problem is NP-Hard, by proving that the associated decision problem is NP-Complete. Specifically, it is demonstrated that the decision problem is a generalization of the maximum flow problem with disjunctive constraints and fixed charges, proven to be strongly NP-Hard. Using a polynomial time verifyer, it is concluded that the problem is in NP, proving NP-Completeness.

To prove NP-Hardness, three scenarios are assumed, the case where $\omega_N = 0$, the case where $\omega_T = \omega_L = 0$, and the case where $\omega_T \neq 0, \omega_L \neq 0, \omega_N \neq 0$.

For the case where $\omega_{\mathbf{N}} = \mathbf{0}$, all possible paths are initially considered (i.e., where the paths are allowed to have any length). For now, it is also assume that the satellites present no interference ($I_{g,s,g',s'} = 0 \ \forall \{g,s\} \neq \{g',s'\}$). When ignoring the constraints imposed by the number of gateways per ground station ($\sum_{s \in \mathcal{S}} x_{g,s} \leq N_g$), and the number of gateway antennas per satellite ($\sum_{g \in \mathcal{G}} x_{g,s} \leq N_{ant}$), the problem corresponds to a maximum flow, minimum cost problem, which can be shown by using the following construction:

1. Create a source node and a sink node with unlimited capacity.
2. Create one node for each satellite s.
3. Create an edge between the source and satellite s for each beam b that is served by satellite s, with capacity equal to the demand of the beam and no cost.
4. Create an edge between each pair of satellites sharing an ISL connection, with capacity equal to c_{ISL} and cost 1.
5. Create a node for each ground station g.
6. Create an edge between satellite s and ground station g, if the satellite is visible to the ground station, with capacity equal to $c_{g,s}$, and no cost.
7. Create an edge between each ground station and the sink, with unlimited capacity and no cost.

Note that this graph is the same as the one explained in [7]. In this graph, a solution to the original problem, under the given assumptions, can be found by computing the maximum flow, minimum cost solution. Note that the assumption of interference can be added by assuming that,

if two satellites present interference, they share the connection to the ground station, keeping the problem structure, and, therefore, the complexity.

Now, the constraints imposed by the number of gateways per ground station ($\sum_{s\in\mathcal{S}} x_{g,s} \leq N_g$), and the number of gateway antennas per satellite ($\sum_{g\in\mathcal{G}} x_{g,s} \leq N_{ant}$) act as disjunctive constraints, i.e., imposing that only a sub-set of the flows can be active. The problem of maximum flow given binary disjunctive constraints has been proven to be strongly NP-Hard [238]. Furthermore, non-binary disjunctive constraints can be transformed into binary disjunctive constraints [239]. Therefore, when introducing these constraints, the problem becomes NP-Hard.

When assuming paths of bounded length without integrality constraints (i.e., assuming that paths have a maximum length P and the flow is allowed to be non-integral), the problem maintains the complexity of the original problem [240]. Therefore, by combining bounded paths and disjunctive constraints, it is shown that, when $\omega_N = 0$, the problem is NP-Hard.

For the case where $\omega_{\mathbf{T}} = \omega_{\mathbf{L}} = \mathbf{0}$, it is assumed that $t_{min} > 0$. Note that when $\omega_T = \omega_L = t_{min} = 0$, the problem can be trivially solved by assigning $x_{g,s} = 0$. Now, it is initially assumed that each ground station and satellite are not limited to a bounded number of gateways ($N_g = N_{ant} = \infty$). It is also assumed that each ground station has no previous constraints in the number of gateways ($N_{g,min} = 0$). This problem corresponds minimum cost flow problem with fixed charges, i.e., where the edges have a fixed cost per activation, plus variable cost depending on the flow. Specifically, the graph can be constructed using the following approach:

1. Create a source node and a sink node a total flow to transmit equal to t_{min}.
2. Create one node for each satellite s.
3. Create an edge between the source and satellite s for each beam b that is served by satellite s, with capacity equal to the demand of the beam and no fixed or variable cost.
4. Create an edge between each pair of satellites sharing an ISL connection, with capacity equal to c_{ISL} and no fixed or variable cost.
5. Create a node for each ground station g.
6. Create an edge between satellite s and ground station g, if the satellite is visible to the ground station, with capacity equal to $c_{g,s}$, with fixed cost of 1 and no variable cost.

7. Create an edge between each ground station and the sink, with unlimited capacity and no fixed or variable cost.

The minimum cost flow problem with fixed charges has been proven to be NP-Hard [241]. Note that allowing $N_{g,min} \neq 0$ is a strict generalization of the problem, where the fixed cost is now a function of the number of arcs activated. Furthermore, adding disjunctive constraints in the form of N_g and N_{ant} is also a strict generalization of the problem. Therefore, it is shown that, when $\omega_N \neq 0$, the problem is NP-Hard.

For the case where $\omega_{\mathbf{T}} \neq \mathbf{0}, \omega_{\mathbf{L}} \neq \mathbf{0}, \omega_{\mathbf{N}} \neq \mathbf{0}$, it can be observed that the problem is a combination of the two previous cases and, therefore, that the problem is NP-Hard. This is true as long as the condition $\omega_T = \omega_L = t_{min} = 0$, which makes the problem trivially solvable, is false.

The proof for NP-Completeness can be completed by observing that, given $x_{g,s}$ and $y^k_{b,g}$, a polynomial time-veryfier can be obtained by assessing each constraint and objective following the previous formulation. Since the number of constraints scales linearly with the number of ground stations and beams, and quadratically with the number of satellites, it is concluded that the decision problem is in NP, demonstrating that it is NP-Complete.

6.4 Gateway Routing approach

While the previous formulation can help in the study of design of ground and ISL architectures, there are two critical aspects that require further consideration: 1) The computational challenge posed by the high-dimensionality of the problem, and 2) The operators preference for metrics that capture continuous operations rather than discrete time-steps.

6.4.1 Addressing a single time-step

Due to the high number of possibilities to transmit information through the network, especially when ISLs are present, feeding the previous formulation directly into a mathematical solver might be computationally intractable. To address this issue, a series of simplifications that allow for the usage of commercial solvers are introduced.

First, the number of variables can be reduced by recognizing the following: when two beams, denoted as b_1 and b_2, are served by the same satellite, it is possible to replace individual variables

per beam per path ($y^k_{b_1,g}, y^k_{b_2,g}$) with a joined variable ($y^k_{\{b_1,b_2\},g}$) representing the information flow to both beams. This substitution is valid under the condition that either $g_{b_1} = g_{b_2}$ or neither g_{b_1} nor g_{b_2} are specified. This consolidation applies universally to any set of beams serviced by a single satellite, leading to a substantial reduction in the number of variables. Upon aggregation, the total demand equals the sum of the individual beam demands. Once a solution is obtained, the demand for each beam is derived by assuming a demand proportionate to the total flow within the consolidated variable. Note that this simplification does not sacrifice optimality, but reduces the search space significatively.

Next, to reduce the number of possibilities, two mechanisms are employed:

1. Reduction of potential paths: The possible paths are limited to those of at most length $N_{hops} + 1$, where N_{hops} represents the maximum number of hops between satellites within a path. In instances where a beam lacks a nearby ground station within a distance less than $N_{hops} + 1$, only paths originating from the nearest ground station are considered.

2. Limitation of Ground Station Selection: Each satellite potential ground stations are confined to the geographically closest $2N_{ant}$ ground stations. This set of feasible ground stations per satellite is represented as $\mathcal{G}_s$.

These adjustments prioritize computational efficiency over optimality by reducing the total number of potential connections between beams and ground stations. Importantly, these modifications are compatible with the earlier variable grouping strategy, since all variables are affected equally.

6.4.2 Addressing multiple time-steps

The proposed formulation enables operators to compute crucial metrics such as throughput, latency, and required ground infrastructure, along with establishing the mapping between ground stations and satellites, for a single time-step. However, operators prioritize metrics that reflect continuous operations rather than discrete time-steps. Additionally, the weighting or prioritization among these metrics remains undefined. To address these challenges, the following methodology outlines a strategy focused on deriving realistic performance indicators representative of continuous operations by solving the formulation across multiple time-steps.

Note that this methodology aligns with the objective of this chapter, which emphasizes the design aspect of the constellation and exploring trade-offs between various design solutions. Although the

previous formulation is applicable to real-world operations, this study falls out of the scope of this dissertation.

In the context of designing ground and ISL architectures, two scenarios are particularly insightful for guiding operators toward optimal decision-making:

1. Limited supply: In scenarios characterized by unlimited demand, the objective is to identify the decisions that maximize capacity.

2. Limited demand: Conversely, in scenarios where a specific demand distribution is given, the aim is to discern the trade-offs inherent in different decision pathways.

Limited supply

In the scenario of limited supply, the primary objective is to identify the set of decisions that maximizes throughput. This can be achieved by encoding solely the throughput metric ($\omega_L = \omega_N = 0; \omega_T = 1$) in the previous formulation. To derive representative metrics over continuous operations, the proposed approach involves simulating N_{steps} time-steps and aggregating the metrics for each step. Notably, to determine the required ground infrastructure, each ground station necessitates as many gateways as the maximum number observed for that ground station across all time-steps. The calculation of gateways per ground station can be expressed as follows:

$$N_{g,total} = \max_t N_{g,min} + N_{g,extra} \tag{6.15}$$

Limited demand

In the scenario of limited demand, the primary objective is to fulfill all demand and subsequently explore the trade-offs between various decisions. This objective can be encoded using hierarchical objectives:

1. Determine the maximum throughput (t_{max}) at each time-step by solving Equation 6.14 with $\omega_L = \omega_N = 0; \omega_T = 1$.

2. Assign specific weights to ω_L and ω_N, and solve Equation 6.14, with $t_{min} = t_{max}$ to ensure complete demand coverage. Note that, since $t_{min} = t_{max}$, t_{extra} is always 0.

Algorithm 6 Solving the Gateway Routing sub-problem

Input: $p^k_{b',g}$ ▷ Set of paths between group of beams b' and g
Input: $\mathcal{G}_s$ ▷ Possible ground stations for each satellite
Input: $\mathcal{T}$ ▷ Set of time-sets
Input: $\omega_T, \omega_L, \omega_N$ ▷ Set of weights for each objective
Input: $t_{min}, N_{g,min}, c_{g,s}, c_{ISL}, I_{g,s,g',s'}, d_b, N_g, N_{ant}$ ▷ Set of parameters of the problem
Output: $x_{g,s}, y^k_{b',g}$ ▷ Solutions to the formulation

for $t \in \mathcal{T}$ **do**
    Solve Equation 6.14
    for $g \in \mathcal{G}$ **do** ▷ Reset $N_{g,min}$ to account for previous time-steps
        $N_{g,min} = 0$
        for $s \in \mathcal{S}$ **do**
            if $x_{g,s}$ **then**
                $N_{g,min} = N_{g,min} + 1$

Similar to the scenario with limited supply, to derive representative metrics over continuous operations, only N_{steps} time-steps need to be simulated and aggregated. The total number of gateways can be determined using Equation 6.15. However, note that in this case, the different time-steps are not independent. Particularly, with the objective of minimizing the total number of gateways across different time-steps, merely minimizing the total number of gateways at each individual time-step is insufficient. For instance, if a ground station utilizes a total of 5 gateways at a certain point in time, it is likely advantageous to maintain this usage across all time-steps, thereby reducing the requirements on other ground stations.

However, given the complexity of solving a single time-step, addressing all time-steps simultaneously is computationally intractable. To that end, multiple time-steps are disassociated by assigning $N_{j,min}$ at time $t \in \mathcal{T}$, as the number of gateways in ground station j at time $t-1$. Although sacrificing optimality on the complete problem, this simplification allows for a sequential resolution of the individual time-steps, considering them one at a time. The complete logic is summarized in Algorithm 6.

Using this logic allows to reduce the number of gateways on each ground station to only those strictly necessary. With this, the representative metrics can be computed by averaging multiple time-steps and using Equation 6.15 to obtain the total number of gateways on each ground station.

6.5 Gateway Routing complexity analysis

To compute the complexity of the proposed formulation and approach, M_g is defined as the maximum number ground stations visible to a single satellite, computed as $M_g = \max_s \sum_g \mathbb{1}_{LoS(p_g,p_s,a_s)}$.

Furthermore, M_b denotes as the maximum number of ground stations a single beam can connect to, computed as $M_b = \max_b \sum_g \max_k \mathbb{1}_{|p^k_{b,g}| \leq N_{hops}+1}$. Note that, in the original formulation, N_{hops} is assumed to be infinity. Finally, M_p is defined as the maximum number of paths from a single beam to a ground station, computed as $M_p = \max_{b,g} |\mathcal{P}_{b,g}|$.

Regarding the memory requirement, the original formulation has a complexity of $\mathcal{O}(|\mathcal{S}|M_g + |\mathcal{B}|M_bM_p)$, where the first term accounts for the decision variables ($x_{g,s}$) and the second term represents the auxiliary variables ($y^k_{b,g}$, $N_{g,extra}$, and t_{extra}). For the proposed approach, the total necessary memory scales with $O(|\mathcal{S}| + |\mathcal{B}|M_bM_p)$. Note that, compared to the original formulation, there are two main changes: the value of M_g is limited to $2N_{ant}$, which does not depend on the input size, and the value of M_b is greatly reduced as N_{hops} is limited.

Regarding the computation, the complexity of the search space scales with $O(2^{|\mathcal{S}|M_g})$, as each satellite can be associated with multiple ground stations. Note that, once we resolve the decision variables $x_{g,s}$, resolving the rest of the variables can be done in polynomial time. Since the problem is NP-Hard, this is the expected complexity of the resolution. For the proposed approach, M_g becomes bounded by a constant, leading to a complexity of $O(2^{|\mathcal{S}|2N_{ant}})$. It is important to note that the formulations for the proposed approach are also NP-Hard, which can be proven similarly to the original formulation. However, the reduction in complexity is sufficient to be able to be addressed using mathematical solvers, as will be shown in subsequent sections.

6.6 Experimental set-up

The subsequent lines elaborate on the validation and design analysis of the proposed methodology. The experiments aim to achieve four distinct objectives: 1) Verify and validate the proposed methodology, demonstrating that it produces feasible results within reasonable time frames, 2) Evaluate its performance against current, non-optimized methods, 3) Analyze the design of ground infrastructure and its trade-offs on overall performance, and 4) Investigate the design of ISL architectures and their trade-offs on overall performance. To that end, 6 different experiments have been conducted, summarized in Table 6.1.

For executing these experiments, the set of users has been generated using the model outlined in Section 3.2. Two constellation configurations have been utilized:

- SpaceX Constellation: Configured with 4,408 satellites, as detailed in Table 1.1. Operating

Analysis	Objective	Constellation	N_{loc}	$N_{us/loc}$
Validation & Verification	Demonstrate that the proposed methodology produces valid solutions	SpaceX	500	1
Convergence	Demonstrate that the proposed methodology produces high quality feasible solutions in reasonable time	SpaceX	20,000	10
Performance	Analyze the performance with respect to established techniques	SpaceX	20,000	10
Design of ground infrastructure	Study the trade-offs in the design of ground infrastructure on the constellation performance	SpaceX	20,000	10
Design of LEO ISLs	Study the trade-offs of different LEO ISL architectures on the constellation performance	SpaceX	20,000	10
Design of MEO ISLs	Study the impact of different MEO ISL architectures on the constellation performance	SpaceX MEO	20,000	10

Table 6.1: Experimental set-up to study the design of ground infrastructure and ISL architecture.

System	Altitude (km)	Inclination (°)	Planes	Satellites per plane	Number of satellites	ISL
SpaceX MEO	540	53.2	72	22	4,408	Yes
	550	53	72	22		
	560	97.6	6	58		
	560	97.6	4	43		
	9,000	70	36	20		

Table 6.2: Summary of the orbit characteristics of the modified SpaceX constellation.

entirely in LEO, this constellation, owing to its substantial satellite count, is expected to benefit from a large number of distributed gateways and intricate ISL architectures.

- SpaceX MEO Constellation: A modification of the original SpaceX constellation, described in Table 6.2. This variation involves relocating satellites at 570 km altitude to 9,000 km. The objective of this alteration is to explore the impact of different MEO - LEO ISL architectures.

The positions of ground stations have been determined using the model outlined in Section 3.6 and Appendix B. Unless specified otherwise, the experiments utilize a total of 80 ground stations. Additionally, in scenarios where not explicitly stated, it is assumed that each satellite possesses four ISLs with satellites at the same altitude (comprising 2 intra-plane and 2 cross-plane connections). All ISLs are assumed to have a total capacity of 20 Gbps.

Parameter	Value	Parameter	Value
N_{hops}	3	Threads	16
MIP Gap	10^{-4}	MIP Method	3

Table 6.3: Summary of the parameters of the simulation.

ω_N	ω_L
$\omega_l = 0$	$\omega_d = 1$
$\omega_l = 5\hat{\omega}_l$	$\omega_d = \hat{\omega}_d$
$\omega_l = \hat{\omega}_l$	$\omega_d = \hat{\omega}_d$
$\omega_l = 0.2\hat{\omega}_l$	$\omega_d = \hat{\omega}_d$
$\omega_l = 0$	$\omega_d = 1$

Table 6.4: Summary of the weights tested on the design analyses

For standard comparison purposes, it is assumed that only circular beam shapes with an aperture angle of 2° are allowed. The users are grouped into beams follows the algorithm delineated in [168]. Beams are dynamically mapped to the satellite with the highest elevation angle at all times. The frequency allocation for each beam is determined using the algorithm outlined in [168]. The power allocation for each beam adheres to the equations provided in Section 1.2.2, considering the EIRP density limit specified in Table 3.3. Each satellite is equipped with a fixed number of antennas, set to 1 ($N_{ant} = 1$), as per the information presented in Table 3.4. Furthermore, the hyperparameters for all experiments are elaborated in Table 6.3. All simulations are conducted utilizing the commercial solver Gurobi [234] (version 9.1.2) on an Intel(R) Xeon(R) Platinum 8160 CPU @ 2.10GHz, with the capability of running up to 16 simultaneous threads.

Regarding the experiments related to the design studies, two scenarios are considered:

- Limited supply: Aiming to maximize throughput. This scenario is simulated assuming unlimited spectrum availability for the satellite-user connections. Note that the number of users and their demand remains unchanged.

- Limited demand: Aiming to explore the trade-offs between maximizing latency and minimizing necessary infrastructure. This scenario is simulated by employing the frequency bands detailed in Section 3.5 and adopting the Frequency Assignment method described in [168] for the satellite-user connections.

To explore the trade-offs between objectives effectively, it is necessary to normalize them to similar orders of magnitude. The normalization strategy is defined as follows. First, obtain the solution

that maximizes throughput, and compute the total information delay ($l = \sum_{b,g,k} d_b y^k_{b,g}(|p^k_{b,g}| - 1)$) and the number of gateways ($n = \sum_{g,s} x_{g,s}$) for this solution using the formula. With these values, the normalization weights are defined as: $\hat{\omega}_{ISL} = \frac{1}{l}$ and $\hat{\omega}_g = \frac{1}{n}$. Based on these normalization weights, simulations are executed with the weights highlighted in Table 6.4. For experiments with limited supply, to ensure performance is maximized, the methodology is executed again after finding the results for maximum throughput. This time, the weights are set as $\omega_T = 0$, $\omega_{ISL} = \hat{\omega}ISL$, $\omega_g = \hat{\omega}g$, and $tmin = tmax$.

In presenting the simulation results, the following metrics will be provided:

- Total number of gateways: The sum of the maximum number of gateways used on each ground station over all simulated time steps.
- Average number of used gateways: The average number of ground station to satellite connections on each time step.
- Average number of hops: The average number of jumps between satellites for each unit of information, averaged over all time steps.
- Average time on ISL: The average time each unit of information spends in the satellite network. Note that this does not include the time to/from the ground, which is minimally impacted by the Gateway Routing algorithm.
- Average throughput: The average amount of data served over all time steps.
- Average gateway utilization: The percentage of the satellite-gateway link utilized, averaged over all satellite-gateway connections and time steps.
- Average number of saturated links: The number of ISL links with utilization above 90

For readability, most numeric results are located in Appendix C.

6.7 Validation and verification analysis

The objective of the validation and verification analysis is to determine the validity of the results obtained using the proposed methodology. Specifically, the aim is to demonstrate that the solution achieves a consistent flow of information, where the only sources and sinks of data are gateways and

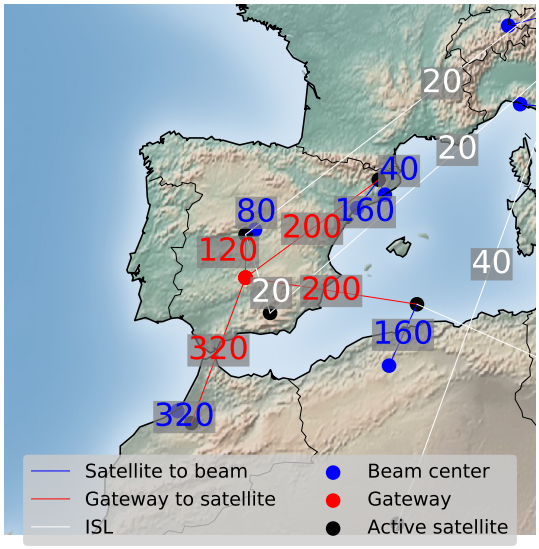


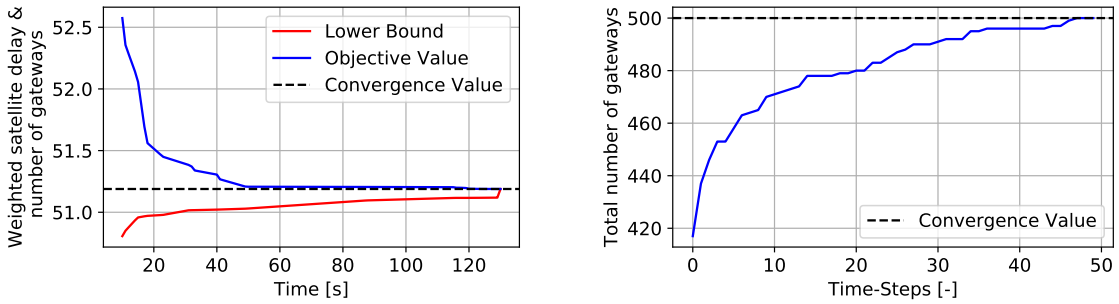
Figure 6-3: Flow of information, in Gbps, through the satellite-terrestrial network with ISLs

beams, respectively. Moreover, it should be ensured that two satellites can exchange information only if they are connected through ISLs. With this purpose, the SpaceX constellation with $N_{loc} = 500$ and $N_{us/loc} = 1$ is simulated.

Figure 6-3 illustrates the flow of information through the satellite-terrestrial network with ISLs over the Iberian Peninsula. It is observed that the flow of information in satellite nodes always balances to zero, indicating that the information entering the node (via gateway connections, depicted in red, or ISLs, in white) equals the information exiting the node (via beam connections, shown in blue, or ISLs, in white). This reaffirms that the solution derived from the proposed methodology is feasible, thereby validating the approach.

6.8 Convergence analysis

The convergence analysis aims to track the evolution of the objective function during a single time-step using the proposed approach and the evolution of the number of gateways per ground station over multiple time-steps. For this analysis, the SpaceX constellation is simulated with $N_{loc} = 20,000$ and $N_{us/loc} = 10$ and executed the proposed methodology for 50 time-steps, with 120 seconds between each time-step.



(a) Single time-step

(b) Multiple time-steps

Figure 6-4: Convergence of the Gateway Routing methodology across a single (left) and multiple (right) time-steps.

Figure 6-4 depicts the convergence of the methodology over a single time-step and multiple time-steps. It can be seen that, within a single time-step, the objective value rapidly progresses towards the optimal value. The optimality of the convergence value has been confirmed using the commercial solver, validating that the methodology reaches the global optima in less than 130 seconds. Regarding multiple time-steps, it is observed that the estimated number of gateways continues to increase over the time-steps. However, this growth tends to converge over time, stabilizing around a value of 500 for 20,000 locations.

6.9 Performance analysis

The objective of this experiment is to evaluate the performance of the proposed methodology against established, non-optimized techniques. As detailed in Section 3.6, current methods for obtaining the satellite-to-gateway mapping do not consider the number of gateways per ground station as a variable to optimize. This section aims to demonstrate that the proposed methodology leads to a reduced necessity for ground stations compared to these techniques.

To achieve this, two approaches have been executed: 1) The methodology explained in Section 6.4, and 2) The technique detailed in [7], where gateways are initially mapped to satellites based on proximity, followed by resolving a Maximum Flow, Minimum Cost problem to obtain the flow through the network. This technique has been modified to include interference by assuming that if two satellites present potential interference, they share the link to the gateway, ensuring fairness

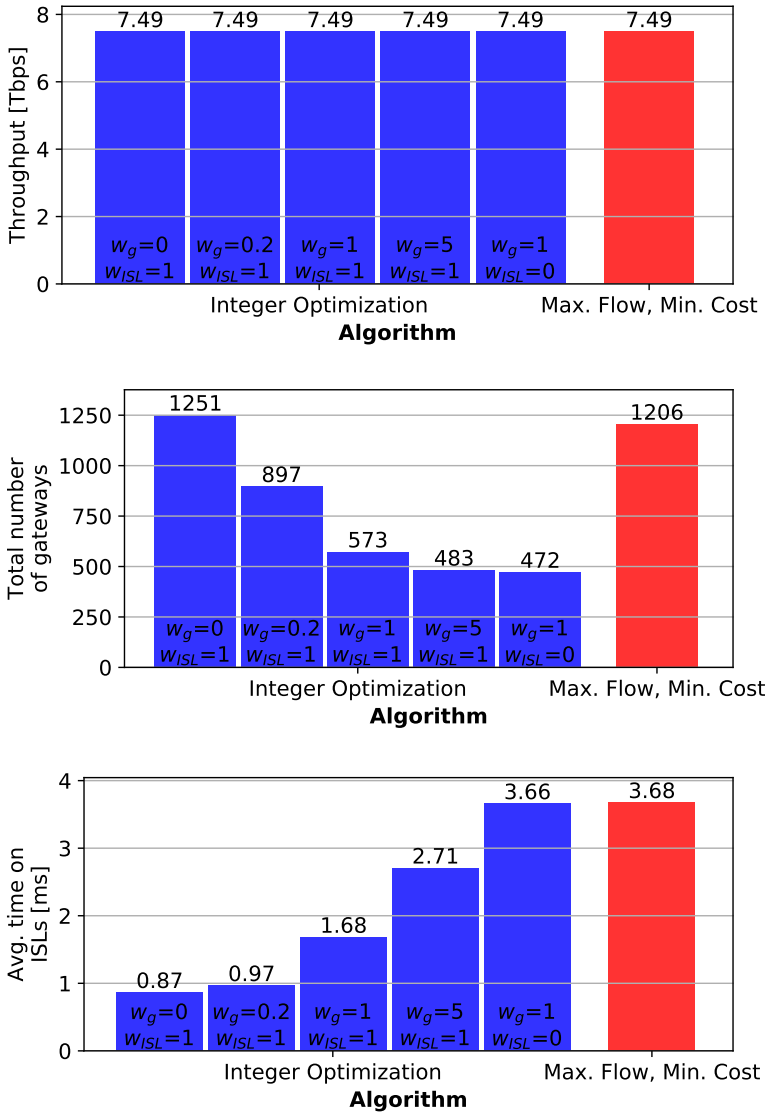


Figure 6-5: Comparison of the proposed Gateway Routing method compared with current literature. Points in blue represent the Integer Optimization (IO) method presented in this chapter, while points in red represent the approach described in [7].

in the comparison between the two algorithms. Both simulations have been conducted using 80 ground stations, with $N_{loc} = 20,000$ and $N_{us/loc} = 10$.

Figure 6-5 and Table 6.5 display the performance of the Gateway Routing method described

Method	w_g	w_{ISL}	Total number of gateways	Average number of used gateways	Average number of hops	Average time on ISL [ms]	Average throughput [Tbps]	Average gateway utilization	Average number of saturated links
IO	0	1	1251	870.18	0.17	0.97	7.49	37.77	5.68
IO	0.2	1	897	685.44	0.20	1.68	7.49	47.92	5.78
IO	1	1	573	435.92	0.38	2.71	7.49	75.21	5.22
IO	5	1	483	349.20	0.60	3.66	7.49	93.38	6.58
IO	1	0	472	343.44	0.81	3.68	7.49	94.50	20.60
Max. Flow, Min. Cost [7]	-	-	1206	849.72	0.14	0.87	7.49	38.96	2.86

Table 6.5: Comparison of the proposed Gateway Routing method compared with current literature. IO stands for the Integer Optimization methods developed in this chapter.

earlier in this chapter compared to the method described in [7]. The results indicate that the proposed approach achieves the same level of throughput while offering better trade-offs in latency and the required number of ground stations. Specifically, when focusing on latency ($w_g = 0, w_{ISL} = 1$), the proposed method requires slightly more gateways for a significant increase in latency (from 3.68ms to 0.87ms). This increase in latency stems from the fact that mapping gateways to satellites based on proximity results in a higher level of interference between satellites, leading to under-utilization of existing links. Conversely, when focusing on ground infrastructure ($w_g = 1, w_{ISL} = 0$), the results show a reduction of 61% in the number of required gateways, while maintaining similar latency. Utilizing other weights allows operators to select a suitable trade-off between latency and ground infrastructure.

6.10 Ground infrastructure design

The aim of this experiment is to evaluate the impact of ground infrastructure design on the performance of a satellite constellation. Specifically, it investigates various performance metrics of the SpaceX constellation, such as the total number of required gateways and average delay, by varying the number of ground stations. Simulations were conducted with 40, 60, 80, and 100 ground stations, utilizing antenna information extracted from the model described in Section 3.6. This experiment provides insights into the performance benefits of deploying an extensive ground infrastructure.

The results of the supply-limited simulations are illustrated in Figure 6-6 and summarized in Table C.1. It is evident that a higher number of ground stations correlates with increased

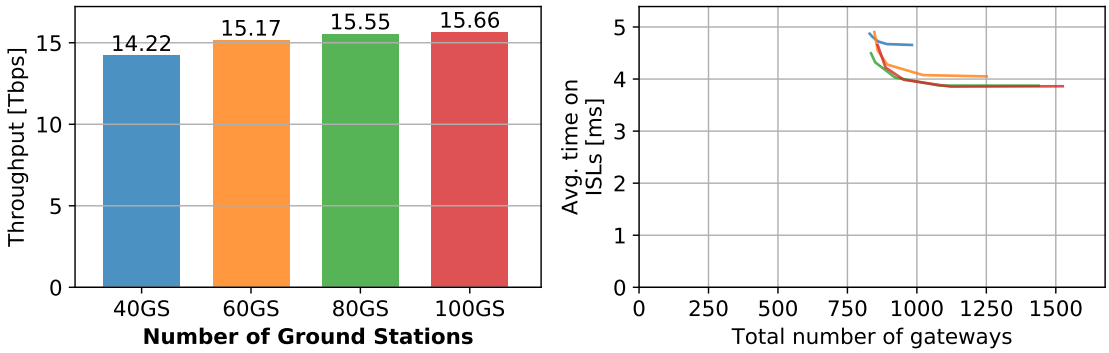


Figure 6-6: Trade-offs of different number of ground stations on supply-limited scenarios

throughput. However, the marginal gain in throughput diminishes with each additional ground station. Specifically, compared to 40 ground stations, employing 60 increases total throughput by 7%, 80 increases it by 3%, and 100 increases it by 1%. Regarding average delay, it is observed that delay tends to increase as the number of ground stations decreases. This can be attributed to the increased distance information must travel to reach a gateway with fewer ground stations. This effect is particularly notable with 40 ground stations but becomes less pronounced with 60 or more.

Furthermore, Table C.1 reveals that average gateway utilization remains nearly constant regardless of the number of ground stations. In other words, when supply is limited, augmenting the number of ground stations does not compromise the efficiency of existing stations; rather, it enhances overall capacity. Note, however, that operating a larger number of ground stations, even with a reduced set of gateway antennas per station, usually comes at a higher cost [199]. Operators should assess the benefit against cost of operating a larger number of ground stations.

The outcomes of the demand-limited simulations are presented in Figure 6-7 and summarized in Table C.2. Interestingly, increasing the number of ground stations does not consistently enhance total throughput. Conversely, the throughput achieved by different configurations exhibits minimal variance, with less than a 3% disparity between the highest (100 GS) and lowest (80 GS) throughput. Regarding delay, a similar trend to the supply-limited scenario is observed: fewer ground stations result in higher delay due to increased information travel distance. This effect is most pronounced with 40 ground stations and diminishes with 60 or more. Comparable to the earlier case, gateway utilization remains steady with the number of ground stations but varies with the total number of

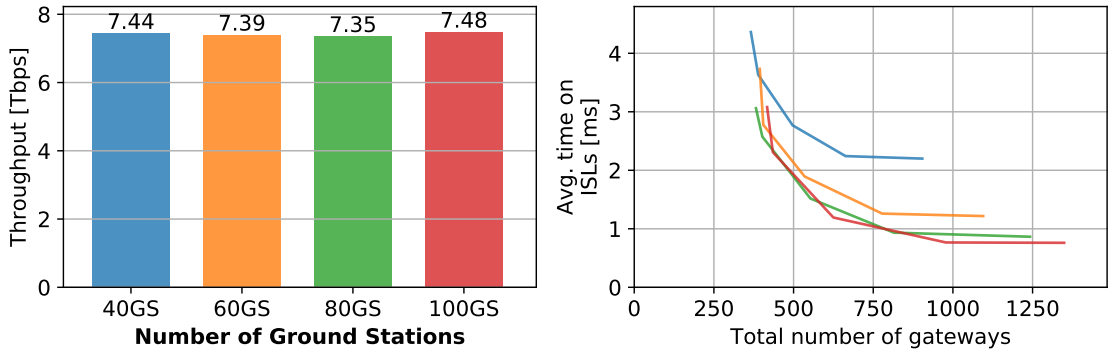


Figure 6-7: Trade-offs of different number of ground stations (GS) on demand-limited scenarios

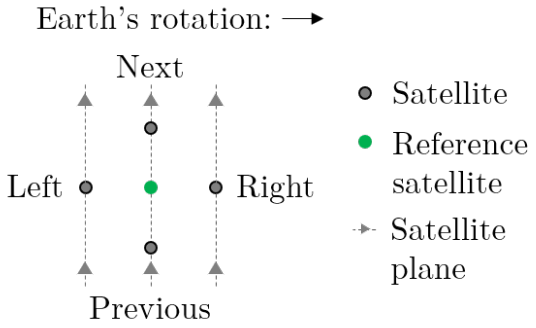


Figure 6-8: ISL notation

gateways employed.

6.11 LEO-to-LEO Inter-satellite link design

The aim of this experiment is to evaluate the trade-offs inherent in designing the ISL architecture and its impact on constellation performance. Specifically, it examines various performance metrics of the SpaceX constellation, such as the total number of required gateways and average delay, while varying the number and configuration of LEO to LEO ISLs. It is assumed in this experiment that satellites can only communicate with those at the same altitude, meaning communication between different altitudes is not feasible. The subsequent section will explore performance when inter-altitude communication is possible.

To evaluate the performance of various ISL architectures, it is necessary to establish the configurations to be examined. For this purpose, the following notation is defined with reference to a satellite, as illustrated in Figure 6-8:

- *Next satellite*: Refers to the closest satellite in the same plane, moving in the direction of orbit.
- *Previous satellite*: Denotes the closest satellite in the same plane but moving in the opposite direction of orbit.
- *Right satellite*: Refers to the closest satellite on the plane located directly to the right of the reference satellite (i.e., in the direction of the Earth's rotation) rotating in the same direction.
- *Left satellite*: Denotes the closest satellite on the plane located directly to the left of the reference satellite (i.e., in the opposite direction of the Earth's rotation) rotating in the same direction.

These definitions can be used in combination: e.g., a next right satellite refers to the next satellite of the right satellite with respect to the reference. Given the consistent orbital speed and orbit evolution, it is assumed that the connection between two satellites remains fixed over time and does not alter with the satellite topology. For this experiment, 11 different configurations have been examined, outlined in Figure 6-9 and Table 6.6. Note that configurations 1 ISL, 2 ISL B, 4 ISL B, 4 ISL C, 6 ISL B, and 6 ISL C are asymmetrical, meaning the connections between satellites may be unidirectional. However, analyzing the impact of directionality on forward-return connectivity is beyond the scope of this dissertation. While the realism of these configurations is subject to future studies, these configurations have been included to study the impact of skipping certain satellites in ISLs in terms of the metrics considered.

The outcomes of the supply-limited simulations are depicted in Figure 6-10 and summarized in Table C.3. As can be observed, the constellation throughput increases as the number of ISLs per satellite rises, regardless of the specific configuration utilized. This augmentation in throughput is attributed to ISLs bolstering virtual capacity in densely populated regions where local gateways may fall short. Transitioning from no ISL to 1 ISL escalates throughput by 32%, from 1 ISL to 2 ISLs it rises between 7% and 16%, from 2 ISLs to 4 ISLs it grows between 15% and 28%, and from 4 ISLs to 6 ISLs it expands between 5% and 13%. Note that the rate of throughput increase

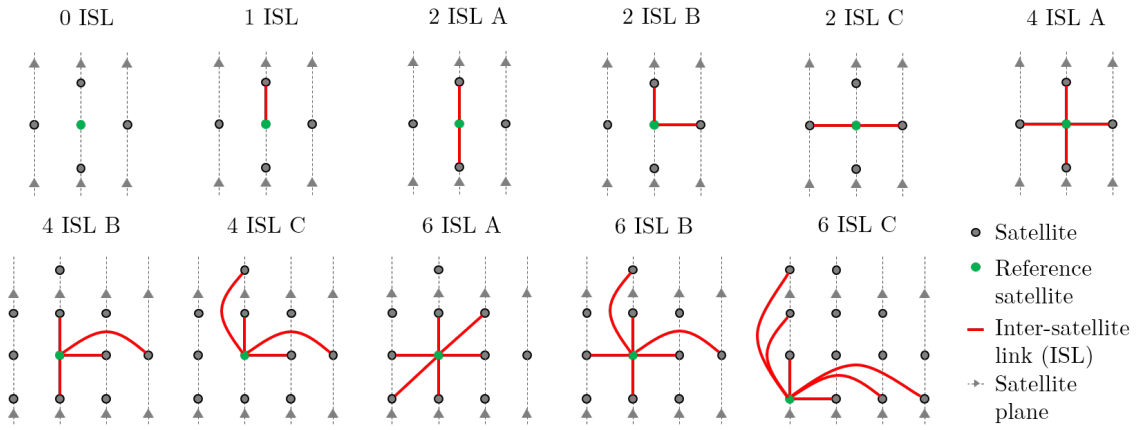


Figure 6-9: Simulated ISL configurations. Each ISL is unidirectional with respect to the reference satellite.

Name	Sat. 1	Sat. 2	Sat. 3	Sat. 4	Sat. 5	Sat. 6
No ISL						
1 ISL	Next					
2 ISL A	Next	Previous				
2 ISL B	Next	Right				
2 ISL C	Right	Left				
4 ISL A	Next	Previous	Right	Left		
4 ISL B	Next	Previous	Right	Right right		
4 ISL C	Next	Next next	Right	Right right		
6 ISL A	Next	Previous	Right	Left	Next right	Previous left
6 ISL B	Next	Next next	Previous	Right	Right right	Left
6 ISL C	Next	Next next	Next next next	Right	Right right	Right right right

Table 6.6: Possible LEO to LEO ISL configurations tested in this dissertation

decelerates with additional ISLs. Particularly noteworthy is the comparison between solutions with no ISL and those with 6 ISLs, revealing approximately double the capacity without any other modifications to payload or software.

Moreover, establishing connections with more distant satellites (options B and C for 4 ISLs and 6 ISLs) further amplifies capacity by facilitating faster load distribution in dense regions. With 4 ISLs, options B and C augment total throughput relative to option A by 1% and 3%, respectively. With 6 ISLs, options B and C enhance total throughput relative to option A by 6% and 5%, respectively. However, this effect is less pronounced than the increase in the number of links. Additionally, due

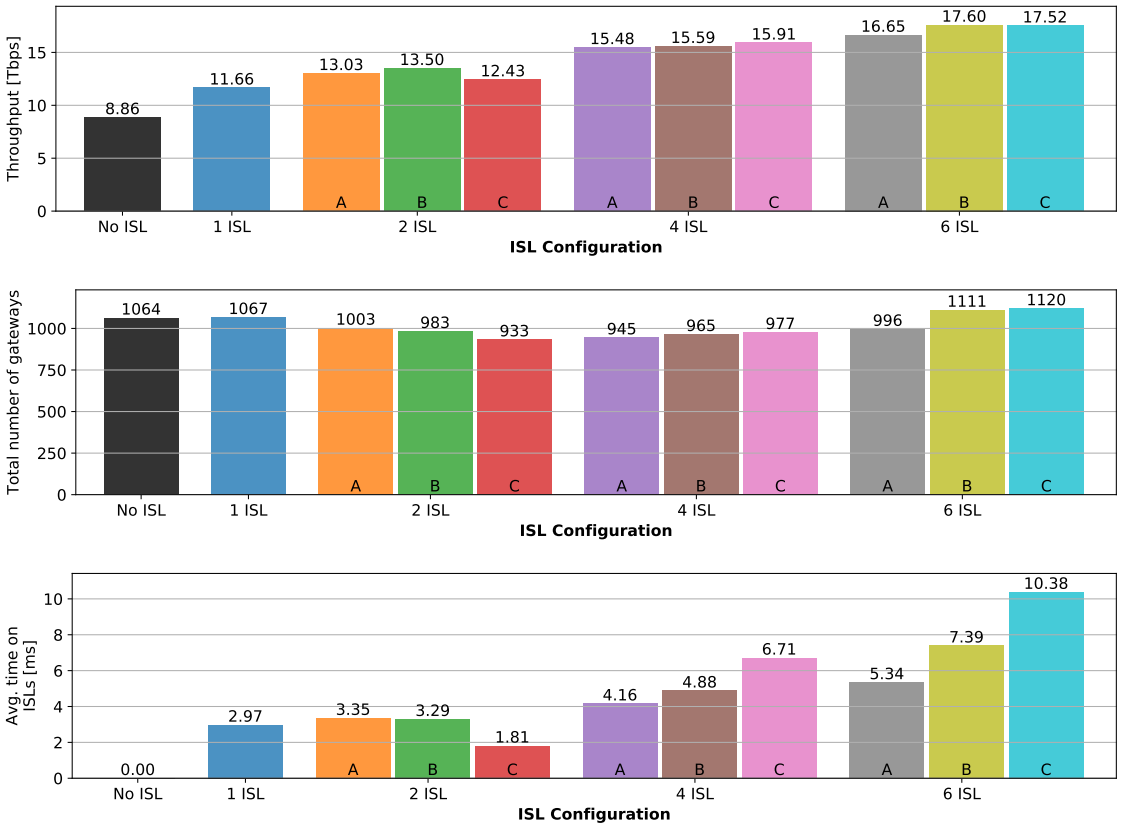


Figure 6-10: Trade-offs of different ISL architectures on supply-limited scenarios

to asymmetry, these options may introduce added complexities when considering the return link.

Surprisingly, a greater number of links do not necessarily correlate with a higher necessity for gateways. Nearly twice the throughput can be achieved by increasing the number of links with minimal modifications to the ground segment. However, it is also important to note that this increase in throughput comes at the expense of heightened connection delay. More distant connections entail longer transmission times via ISLs, resulting in lower average latency.

The outcomes of the demand-limited simulations are illustrated in Figures 6-11, 6-12, 6-13, and 6-14, with corresponding data presented in Tables C.4 and C.5. Beginning with Figure 6-11, which showcases results for ISL configurations featuring 2 ISLs, a notable observation is that all three options yield nearly identical throughput, with less than a 1% difference between the highest and lowest throughput. This uniformity stems from the demand-limited nature of the scenarios,

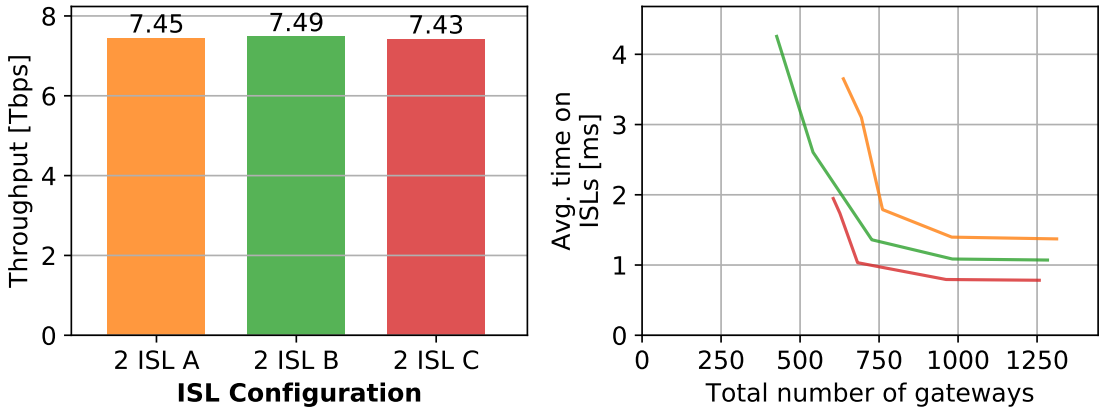


Figure 6-11: Trade-offs of different ISL architectures with 2 ISLs on demand-limited scenarios

where total throughput is constrained by user demand rather than system capabilities. Moreover, it is shown that option A is surpassed by options B and C, as for every point within option A, there exists a corresponding point within options B or C with fewer gateways, lower delay, and higher throughput. Between options B and C, option B can mitigate the need for gateways but at the expense of increased average delay compared to option C. This underscores the notion that utilizing a mix of intra- and cross-plane connections aids in more evenly distributing demand, albeit at the expense of network speed. Additionally, employing the proposed methodology yields diverse solutions with varying trade-offs between required gateways and total delay. While it is possible to achieve the same constellation capacity by reducing the number of gateways, this comes at the cost of increased connection delay, necessitating operators to find an appropriate trade-off to provide the required latency for their customers.

Moving to Figure 6-12, which displays results for ISL configurations featuring 4 ISLs, a similar trend emerges. Again, total throughput is constrained by user demand, resulting in all options delivering nearly identical throughput. In this scenario, option A outperforms both options B and C, as for each point utilizing options B or C, a corresponding point within option A exists with fewer gateways, lower delay, and equal throughput. When demand is the limiting factor, leveraging connections to nearby satellites proves advantageous in terms of quality of service while requiring a comparable number of gateways to other options. Figure 6-13 further emphasizes this conclusion by presenting results for ISL configurations featuring 6 ISLs. Here, option A dominates option B

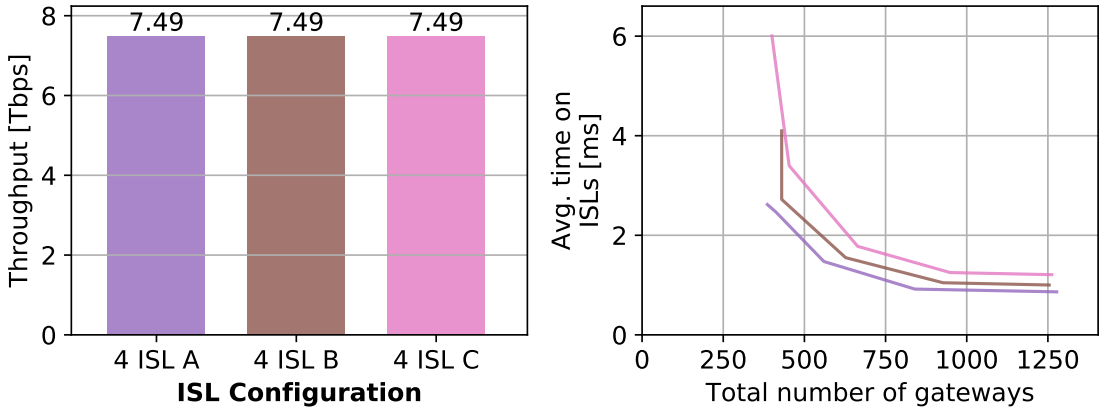


Figure 6-12: Trade-offs of different ISL architectures with 4 ISLs on demand-limited scenarios

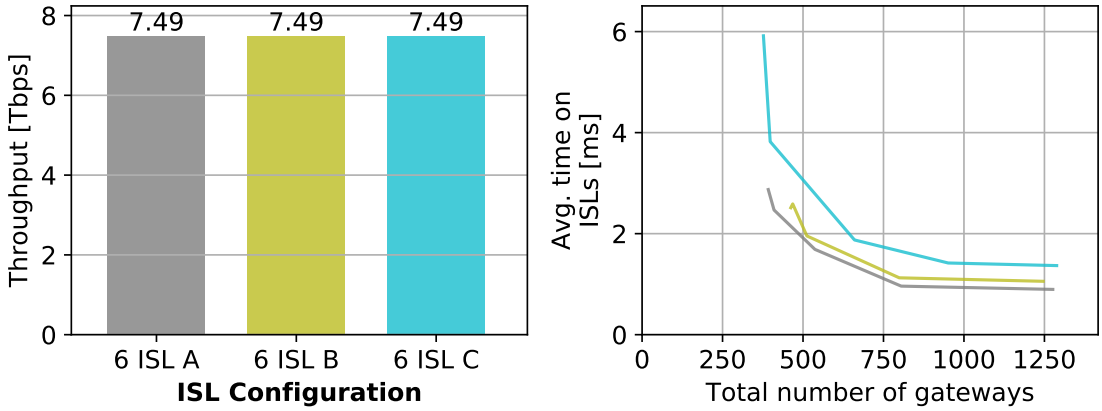


Figure 6-13: Trade-offs of different ISL architectures with 6 ISLs on demand-limited scenarios

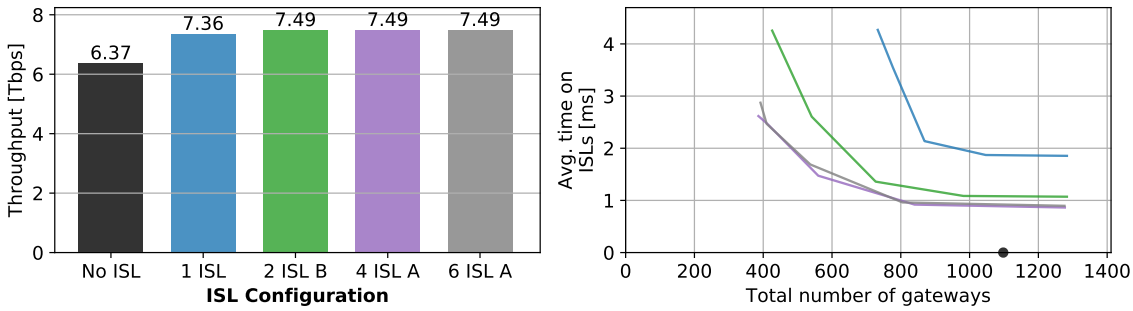


Figure 6-14: Trade-offs of different ISL architectures on demand-limited scenarios

and all points within option C except one, which achieves a slightly reduced necessity for gateways at the expense of a significant increase in delay.

Figure 6-14 presents results for a subset of ISL configurations, revealing that architectures with fewer than 2 ISLs are unable to achieve maximum throughput. This shortfall stems from inadequate local data availability to meet user requirements, and, therefore, ISLs are necessary to supplement it. Contrary to the supply-limited case, increasing the number of ISLs diminishes the required number of gateways and can help alleviate delay. This is because, in scenarios where capacity is sufficient, augmenting ISLs enables information to choose better paths, enhancing service quality and reducing requirements. Consequently, constellations consistently benefit from more ISLs, irrespective of the operational scenario. If demand is lower than capacity, ISLs aid in optimizing routing options; if demand exceeds capacity, ISLs assist in alleviating congestion in dense areas, resulting in higher throughput.

6.12 MEO-to-LEO Inter-satellite link design

The objective of this experiment is to evaluate the trade-offs of designing the ISL architecture on constellation performance. Specifically, the focus is on assessing the benefits of MEO to LEO ISLs. To this end, a modified SpaceX constellation, as shown in Table 6.2, has been simulated. Four different configurations have been explored: 0, 1, 2, and 4 connections between MEO and LEO. MEO satellites are assumed to connect to LEO satellites with the highest elevation angle. For the supply-limited case, three different LEO to LEO ISLs have been tested: no ISL, 2 ISL intra-plane (option 2 ISL A), and 4 ISL (option 4 ISL A). Each configuration has been executed under equal

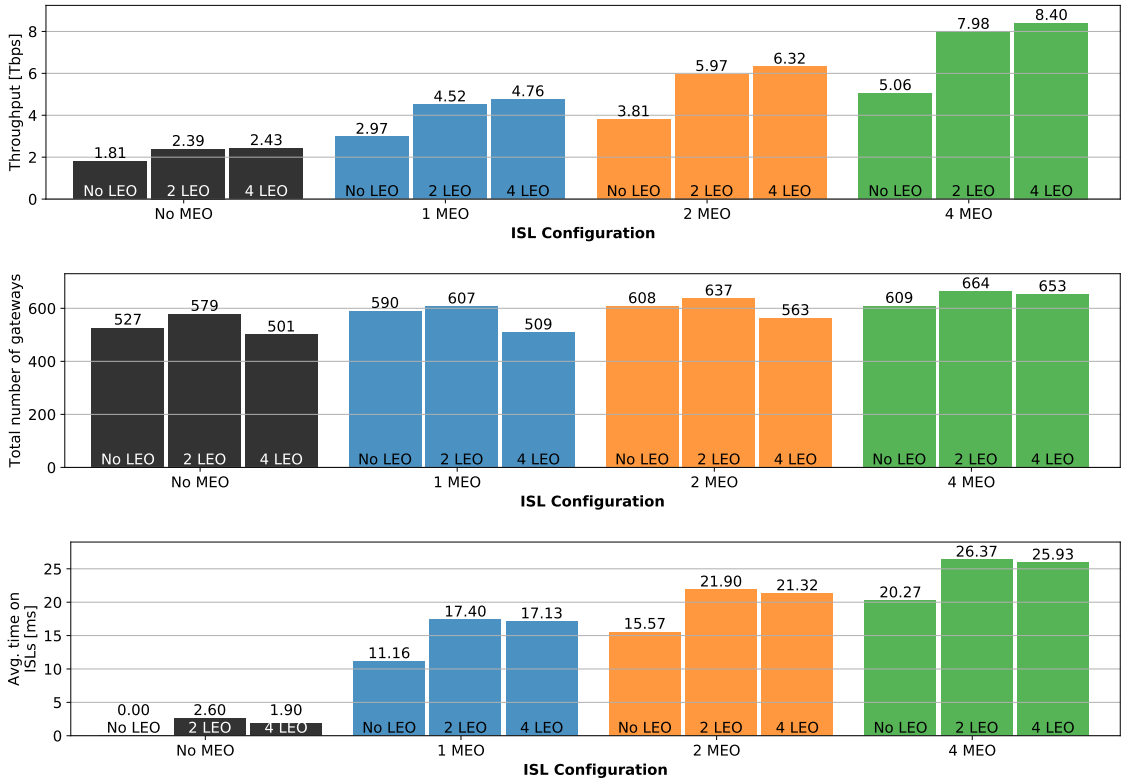


Figure 6-15: Trade-offs of different MEO to LEO ISL architectures on the supply-limited scenario

weights between objectives ($\omega_{ISL} = \hat{\omega}_{ISL}, \omega_g = \hat{\omega}_g$). For the demand-limited case, it is assumed that LEO satellites connect with each other using 4 ISLs (2 intra- and 2 cross-plane, option 4 ISL A). In both supply- and demand-limited cases, MEO satellites cannot connect with each other.

Figure 6-15 and Table C.6 present the results for the supply-limited case using the proposed methodology. As shown, employing MEO to LEO ISLs substantially enhances the constellation throughput. As expected, a greater number of connections between satellites, both between MEO and LEO and among LEO satellites, boosts total capacity. However, remarkably, connectivity to MEO exerts a significant impact on performance: establishing 4 ISLs to MEO with no ISLs between LEO achieves higher throughput than employing a fully connected LEO network with either 0 or 1 link to MEO. This phenomenon arises from the fact that MEO connections bolster throughput in regions with poor local coverage. MEO satellites have broad Earth observation capabilities, reducing the regional requirement for gateways and ground stations compared to LEO.

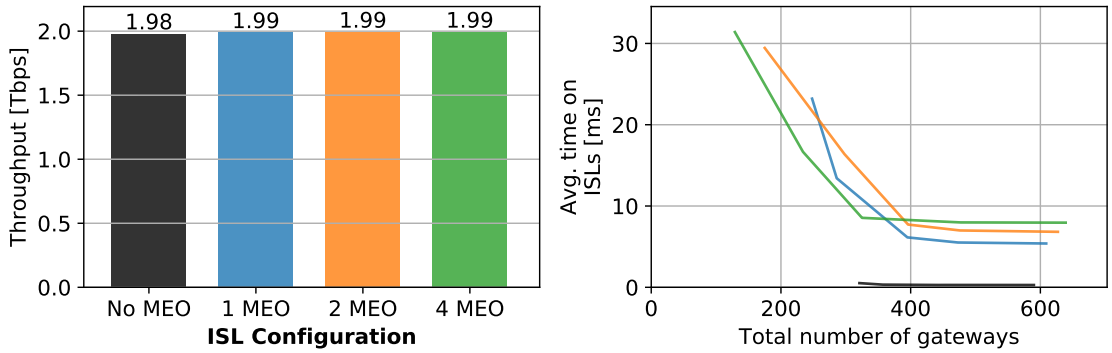


Figure 6-16: Trade-offs of different MEO to LEO ISL architectures on the demand-limited scenario

Additionally, MEO satellites aggregate demand from various regions, rendering them more efficient when utilizing gateway links. Notably, combining a fully connected LEO network with numerous MEO to LEO ISLs achieves the highest throughput, surpassing threefold the throughput achieved without MEO connectivity.

However, this enhancement comes at the expense of increased delay. MEO to LEO links often span significant distances in space, rapidly escalating total delay. Consequently, a greater number of MEO to LEO links directly correlates with higher total delay. Compared to using only LEO ISLs, employing 1 MEO to LEO ISL yields between 7 and 9 times higher delay, employing 2 MEO to LEO ISLs yields between 8 and 11 times higher delay, and employing 4 MEO to LEO ISLs yields between 10 and 14 times higher delay.

Figure 6-16 and Table C.7 illustrate the results for the demand-limited case using the proposed methodology. Here, leveraging MEO to LEO connectivity markedly reduces the necessity for the total number of gateways, albeit at the expense of heightened delay. Specifically, employing 1 MEO to LEO connection can decrease the gateway requirement by 23%, 2 MEO to LEO connections can reduce it by 45%, and 4 MEO to LEO connections can lower it by 60%. Additionally, MEO connectivity slightly enhances coverage, particularly in regions with deficient gateway infrastructure, by utilizing MEO satellites as relays for transmitting information.

However, akin to the previous case, the increase in connectivity and reduced gateway requirements coincide with increased delay. In this scenario, delays can reach up to 42 times those achieved when solely relying on LEO satellites. This arises from the fact that, in scenarios of limited demand,

routing most information through MEO satellites, where it is aggregated toward individual ground stations, reduces the necessary ground infrastructure but necessitates substantial routing time in the satellite segment.

6.13 Chapter summary and conclusions

This chapter has studied the design trade-offs of different ground infrastructure and ISL architecture. To that end, it described a novel methodology to address the Gateway Routing problem, which includes interference considerations and fixed beam-gateway relations, thereby addressing the research gap found in Section 2.2.10.

6.13.1 Chapter summary

The theoretical segment of the chapter initiated with Section 6.1, which described the connection between the Gateway Routing problem and the design of ground infrastructure and ISL architecture. Following this, Section 6.2 outlined a novel formulation for the Gateway Routing problem, emphasizing the maximization of capacity and service quality while minimizing ground infrastructure. Subsequently, Section 6.3 established the NP-Hardness of the new formulation by demonstrating its generalization of the maximum flow problem with disjunctive constraints and fixed charges. To address the complexity of solving this problem in high-dimensional scenarios, Section 6.4 proposed mechanisms to simplify the formulation and provide operationally representative insights. Finally, Section 6.5 concluded the Gateway Routing methodology by evaluating the formulation and approach complexity in terms of memory and computation requirements.

Transitioning to the practical aspect, Section 6.6 delineated the simulation conditions and experimental setups for subsequent sections. Section 6.7 validated the proposed approach by demonstrating its ability to ensure a consistent flow of information through the network. Similarly, Section 6.8 analyzed the convergence of the approach, confirming its capability to yield high-quality solutions within reasonable timeframes. Following this, Section 6.9 compared the proposed Gateway Routing methodology against established non-optimized techniques, highlighting its significant reduction in the necessity for gateways. Additionally, Section 6.10 investigated the impact of ground segment design on overall performance, demonstrating that a low number of ground stations can achieve comparable throughput at the expense of latency. Furthermore, Section 6.11 explored the design

of LEO ISL architectures, revealing that a higher number of ISLs enhances throughput while presenting trade-offs between the number of gateways and latency. Finally, Section 6.12 examined the design of MEO to LEO ISL architectures, underscoring the critical role of MEO to LEO connectivity in maximizing performance.

6.13.2 Response to Research Questions

Research question 6.1

How do the ground infrastructure and ISL architecture affect the constellation performance?

The ground infrastructure affects the supply side of the constellation, dictating the availability and reach of connectivity worldwide. Meanwhile, the ISL architecture shapes how this connectivity traverses the satellite network to various regions on Earth. Both factors significantly influence total throughput by governing the flow of information across the network.

Research question 6.2

Does the capacity scale linearly with the amount of ground station locations?

No. Although expanding the number of ground stations can increase throughput, particularly when existing stations are unable to meet demand, this effect diminishes as the system approaches saturation. Once demand is adequately covered or user link capacity becomes the limiting factor, further increasing ground stations no longer impacts total capacity. However, it can enhance service quality, as information can use more efficient paths to reach their destination.

Research question 6.3

When changing the ISL configuration (i.e., the number of links and the satellite each link connect to), what is the best ISL configuration that maximizes throughput and quality of service, while minimizing the required ground infrastructure?

When developing a LEO ISL architecture aimed at maximizing capacity, denser and further connections can enhance throughput, particularly in densely populated regions where ISLs become a bottleneck. Yet, this increased throughput often comes at the expense of lower service quality, as data traverses the satellite network for longer duration. Conversely, in scenarios with limited

demand, additional connections can mitigate the need for extensive ground infrastructure and improve service quality by enabling more efficient data routing.

In the context of developing a MEO to LEO ISL architecture, establishing connectivity between MEO and LEO layers proves indispensable for maximizing capacity. In certain cases, densely interconnected LEO and MEO layers can boost throughput by a factor of 3, primarily by extending coverage to areas inadequately served by local ground stations. Note that the specific metrics presented in this dissertation only refer to one particular scenario and are meant to showcase the potential of MEO to LEO connectivity. These values may vary according to the scenario. Furthermore, this heightened capacity is counterbalanced by reduced service quality attributed to the latency introduced by MEO-LEO links.

While deeply interconnected ISL architectures can indeed bolster constellation capacity, these benefits often entail increased ground infrastructure requirements and diminished service quality. Consequently, operators must carefully weigh trade-offs between the number of ISL connections, ground infrastructure, service quality, and total throughput.

6.13.3 Specific chapter contributions

The specific contributions of this chapter are as follows:

- Formulated the Gateway Routing sub-problem as a multi-objective mixed integer linear problem, and proved NP-hardness.
- Developed a novel methodology to address the Gateway Routing problem including interference considerations while allowing for fixed beam-gateway mappings.
- Studied the validation and convergence of the proposed methodology, proving that it provides high quality solutions in feasible time.
- Studied the performance of the proposed methodology against current techniques, proving that it is able to achieve better trade-offs between ground infrastructure necessary and service quality.
- Studied the design of ground infrastructure, and showed that a larger number of ground stations allow for higher capacity and improves the service quality.

- Studied the design of LEO ISL architectures, proving that, a larger number of links increases the constellation performance in all cases, either by increasing capacity in dense regions, or allowing data to travel through more efficient paths.
- Studied the design of MEO to LEO ISL architectures, proving that connectivity between MEO and LEO greatly enhances the capacity of the constellation, reducing the necessity for ground infrastructure and increasing the coverage where the local ground station coverage is poor.

Chapter 7

Analyzing constellation performance under realistic operational conditions

The next generation of satellite communications is expected to cater to millions of users spread across the globe [21], enabled by thousands of powerful spacecraft connected to hundreds of ground stations. This high dimensionality, coupled with the integration of software-defined technologies like phased-array antennas and adaptive coding and modulation, adds to the natural complexity of NGSO satellite communications. Consequently, analyzing the performance of these constellations under realistic conditions becomes considerably more intricate compared to previous satellite generations. Performance assessments play a pivotal role for satellite operators and other stakeholders in making informed financial and technical decisions, particularly during the constellation design phase. Given the significance of these evaluations, obtaining precise performance estimations is critical for effective decision-making. However, as outlined in Section 2.4, existing studies primarily offer insights into constellation performance under simplified or idealized conditions [7, 192]. The analysis of next-generation constellations performance under realistic operational conditions remains largely unexplored. This chapter aims to bridge this gap by: 1) introducing an automated resource allocation methodology that addresses critical pre-operational decisions in the RAP, and

2) investigating the performance of modern constellations under realistic operational conditions using the proposed methodology.

The research questions this chapter aims to address are:

Research question 7.1

What is the performance of modern constellations under realistic operational conditions in terms of throughput, required ground infrastructure, and latency? How does this performance compare to ideal operational conditions?

Hypothesis: The objective of this question is exploratory.

Research question 7.2

When splitting users into altitudes, do the benefits of using altitude-appropriate beam shapes overcome the handicaps imposed by restricting the satellite selection to only those at the assigned altitude in terms of throughput?

Hypothesis: On one hand, using the method explained in Chapter 4 to map users to altitudes has proven to give benefits over simple Satellite Routing policies, even with the reduced selection of satellites for each user. On the other hand, Chapter 5 showed that, when optimizing, overseeing a large fleet of satellites allows for a more flexible allocation, leading to increased capacity. It is expected that using altitude-appropriate beam-shapes as defined in Chapter 4 overcomes the handicaps that a smaller selection of satellites entails for the Satellite Routing problem, defined in Chapter 5.

Research question 7.3

How much better do the methods presented in this dissertation perform against existing methods in terms of throughput, number of gateway antennas, and average delay?

Hypothesis: The joint Beam Shaping and User Grouping method and the Satellite Routing approach presented in this dissertation lead to approximately a 100% and 75% increase in capacity at the sub-problem level, respectively. However, since both problems are interconnected, it is unclear what the total gain will be at the system level when using both approaches simultaneously. It is estimated that using both methodologies will lead to at least a 2x increase in total capacity, but could potentially be more. Regarding the Gateway Routing, Chapter 6 showed that the proposed

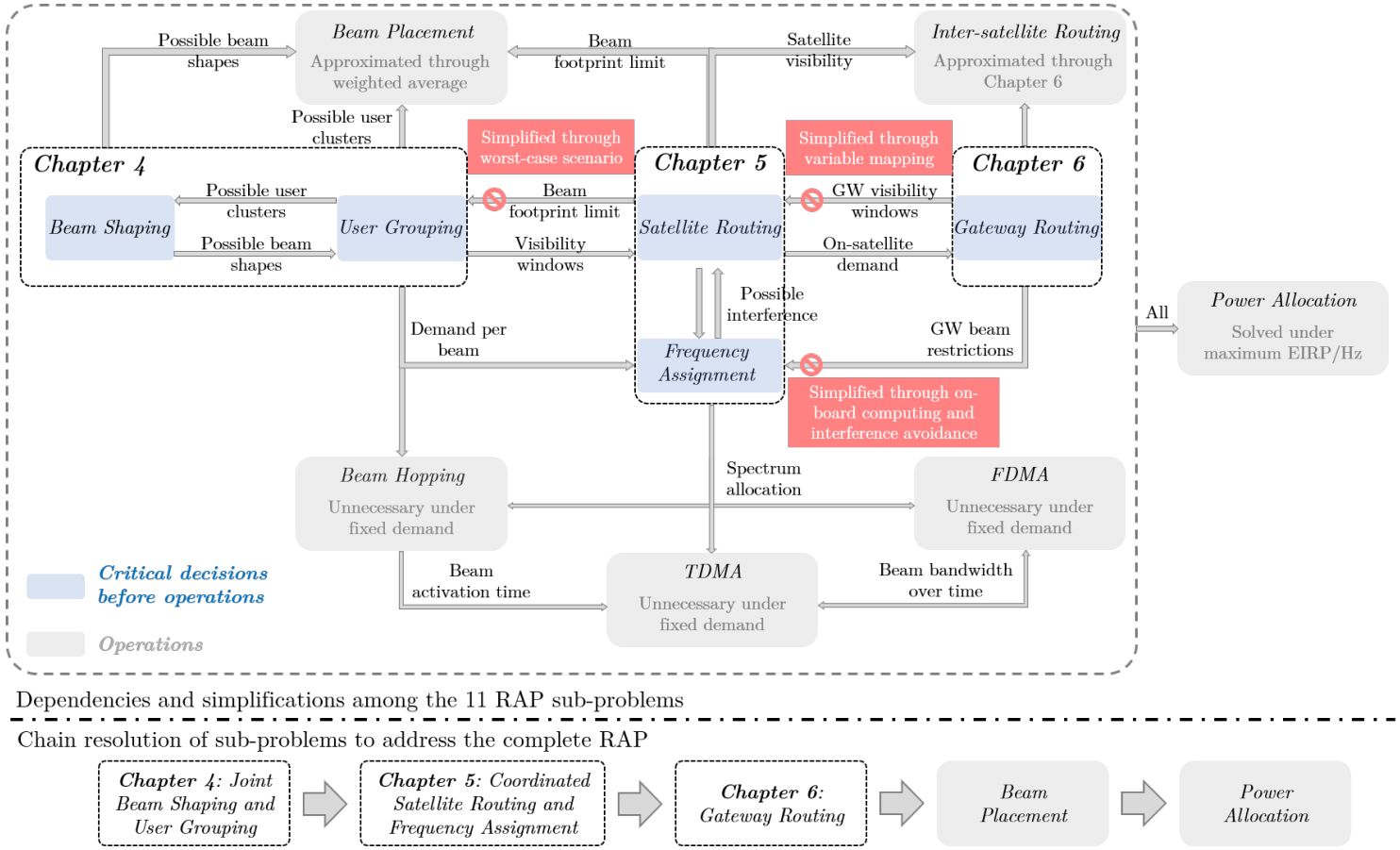
methodology leads to the same throughput as existing methods, but achieves a better trade-off between quality of service and necessary ground infrastructure. It is hypothesized that these trade-offs maintain at the system level.

7.1 Complete RAP Framework

Section 1.2.3 defined the RAP as a compendium of sub-problems that satellite operators must tackle to provide service to users. In conjunction with the models detailed in Chapter 3, it is critical to take appropriate decisions in the RAP to represent realistic operational conditions. The following lines detail a novel methodology to resolve the RAP, encapsulating realistic operational conditions.

The RAP encompasses 11 sub-problems, as detailed in Figure 7-1: Beam Placement, Inter-Satellite Routing, Beam Shaping, User Grouping, Satellite Routing, Gateway Routing, Frequency Assignment, Beam Hopping, FDMA, TDMA, and Power Allocation. Among these, five decisions must be addressed before operations, due to their inter-dependencies: Beam Shaping, User Grouping, Satellite Routing, Gateway Routing, and Frequency Assignment. Note that, while Satellite Routing and Gateway Routing necessitate pre-operational resolution, they also require continuous updates during operations due to the constellation dynamic nature. Notably, appropriate techniques for realistic decision-making on these five sub-problems have been discussed in previous chapters: Chapter 4 addressed the joint Beam Shaping and User Grouping problem, Chapter 5 proposed a coordinated Satellite Routing and Frequency Assignment approach, and Chapter 6 described a novel methodology for the Gateway Routing sub-problem.

To establish a comprehensive framework that addresses the RAP for realistic operational scenarios, two crucial elements require further attention: 1) Resolving the inter-dependencies between sub-problems, and 2) Addressing the sub-problems concerning operations (i.e., Beam Placement, Inter-satellite Routing, Beam Hopping, FDMA, TDMA, and Power Allocation).



Dependencies and simplifications among the 11 RAP sub-problems

Chain resolution of sub-problems to address the complete RAP

Figure 7-1: Dependencies and simplifications among RAP sub-problems (top) and chain resolution of sub-problems (bottom).

7.1.1 Dependencies between sub-problems

The critical sub-problems before operations exhibit inter-dependencies as highlighted in Figure 7-1.

Achieving an optimal solution to the RAP necessitates simultaneously tackling all these sub-problems. However, as discussed in preceding chapters, individually addressing each problem is already computationally challenging, as each one of them is NP-Hard with thousands of variables. As an example, achieving solutions for only the Satellite Routing and Frequency Asssignment coordinated approach took around 40h of computation. It is expected that combining multiple problems only increases this value exponentially. Therefore, concurrent resolution of all sub-problems using current computational tools is computationally intractable. Hence, to address the complete RAP, certain simplifications are necessary. Firstly, as detailed in Chapter 4, the Beam Shaping and User Grouping sub-problems can be addressed simultaneously. Similarly, Chapter 5 describes the coordinated resolution of Satellite Routing and Frequency Assignment. Additionally, the following simplifications are introduced (highlighted in red in Figure 7-1):

- To resolve the User Grouping dependency towards Satellite Routing, a worst-case scenario is adopted, where the footprint contour is selected based on the most pessimistic Satellite Routing solution at each time point. This mirrors the assumption introduced in [12].

- To address the Satellite Routing dependency towards Gateway Routing, it is assumed that beams can be mapped to any ground station, which will be decided upon during operations. This constraint aligns with the services offered by modern satellite operators [39], where the ground station is decided by the operator.

- To tackle the Frequency Assignment dependency towards Gateway Routing, it is assumed that: 1) spacecraft possess on-board computation capable of aggregating data streams from various users, consistent with modern satellite characteristics outlined in Section 3.5, and 2) gateway-satellite links do not interfere with user-satellite links, aligning with the general intent of siting ground stations outside populated regions, as shown in Appendix B.

These simplifications enable a chain resolution to attain a solution for the five critical decisions before operations. Previous chapters have shown how to address each one of these decisions in practical time for high dimensional scenarios with hundreds of thousands of users.

7.1.2 Addressing the remainder of the RAP sub-problems

Each remaining sub-problem is resolved as follows:

- Beam Placement: It is presumed that the center of the beam corresponds to a weighted sum of the users positions. The computation of the beam center position is akin to the methodology delineated in [12] as:

$$\begin{aligned} p_b &= \sum_{u \in \mathcal{V}_b} w_u p_u \\ \tilde{w}_u &= \sum_{v \in \mathcal{V}_b} ||p_u - p_v|| \\ w_u &= \frac{\tilde{w}_u}{\sum_{u \in \mathcal{V}_b} \tilde{w}_u} \end{aligned} \tag{7.1}$$

- Inter-satellite Routing: While the methodology outlined in Chapter 6 does not encompass specific routing protocols, it provides an estimate of the flow between satellites, which will serve as a solution for the Inter-satellite Routing sub-problem.

- Beam Hopping, FDMA, and TDMA: It is assumed that user demand remains constant and equal to the expected demand. This assumption obviates the need to resolve Beam Hopping, FDMA, and TDMA, as their primary advantage lies in addressing variable demand, while being representative of the statistically average scenario.

- Power Allocation: It is assumed that the power for each beam can be computed by using the link budget equation described in Chapter 1.2.2. The radiated power for each beam is upper bounded by the maximum EIRP density multiplied by the total assigned bandwidth to the beam, without any restriction on the total power a satellite can produce.

7.1.3 Addressing the complete RAP

Based on these simplifications, the lower portion of Figure 7-1 displays the block diagram used to solve the complete RAP. In particular, note that the process is sequential, and each sub-problem or combination of sub-problems only need to be addressed once. Note that each block is addressed independently, one at a time. When using optimization, optimization is applied at the sub-problem level and not at the global level. While this is a source of sub-optimality, optimizing all problems at

Constellation	Number of sets $\mathcal{A}$	Specific sets $\mathcal{A}$	Joint Beam Shaping and User Grouping	Satellite Routing
SpaceX	4	One per altitude	Chapter 4	Maximum elevation angle & Chapter 5
SpaceX	1	-	[168]	Maximum elevation angle & Chapter 5
Boeing	10	One per altitude	Chapter 4	Maximum elevation angle & Chapter 5
Boeing	4	Low-LEO (<1,000 km), High-LEO (≥1,000 km, < 2,000 km), MEO, HEO	Chapter 4	Chapter 5
Boeing	3	LEO, MEO, HEO	Chapter 4	Chapter 5
Boeing	1	-	[168]	Maximum elevation angle & Chapter 5

Table 7.1: Summary of the experiments executed to assess the interactions between the joint Beam Shaping and User Grouping and Satellite Routing problems

once is currently intractable due to the complexity of each individual problem and the combinatorial nature of the solutions.

7.2 Interactions between User Grouping and Satellite Routing

Based on the proposed methodology, the joint Beam Shaping and User Grouping and the Satellite Routing problems are solved in sequence. Nevertheless, based on the implementations proposed previously in Chapters 4 and 5, there is one conjunction point that remains unaddressed. On one hand, joint Beam Shaping and User Grouping offers the advantage of tailored beam shapes specific to altitudes, potentially enhancing signal coverage and quality for users within those altitudes. This approach has shown benefits, particularly in scenarios with simpler Satellite Routing policies. On the other hand, the Satellite Routing method has shown to benefit from a larger and more flexible satellite fleet to increase capacity, resulting from a uniform beam shape. It is unclear whether the benefits of the joint Beam Shaping and User Grouping technique overshadow the limitations imposed on the Satellite Routing method. In other words, it is crucial to answer the following question: is it better to split the users and beams into different altitudes and use altitude-appropriate beam shapes, even when this restricts the satellite availability, or is it better to use uniform shapes, and allow for a larger flexibility in the satellite selection?

The objective of this section is to provide an answer to the previous question by analyzing the impact of using both techniques simultaneously. To that end, the following realistic constellations,

Parameter	Value	Parameter	Value	Gurobi Parameter	Value
N_u	1,000	N_b	3,000	Threads	16
I_{thres}	1%	N_{conv}	20	MIPGap	10^{-4}
N_{conv}^{SR}	10	N_{ch}	100	Iteration	300s
N_{conv}^{FA}	15	N_{cutoff}	25	Time Limit	

Table 7.2: Summary of the parameters of the simulation to assess the interactions between the joint Beam Shaping and User Grouping and Satellite Routing problems.

extracted from Table 1.1, have been simulated:

- SpaceX constellation: The objective of testing this constellation is to assess the benefits of splitting the users into altitudes in cases where the altitudes are in close proximity. When using the hybrid algorithm developed in Chapter 4, each set $\mathcal{A}$ is defined as all the satellites with the same mean altitude.

- Boeing constellation: The objective of testing this constellation is to assess the benefits of splitting the users into altitudes in cases where the altitudes are **not** in close proximity. When using the hybrid algorithm developed in Chapter 4, three different configurations have been tested, as highlighted in Table 7.1.

A total of 10,000 user locations (N_{loc}) have been simulated, with 10 users per location ($N_{us/loc}$), following the model detailed in Section 3.2. The set of experiments conducted is summarized in Table 7.1. For comparison, simpler methods for the joint Beam Shaping and User Grouping ([168], briefly explained in Chapter 4), and for the Satellite Routing problem (maximum elevation angle heuristic, as defined in Chapter 5), have also been simulated. For the hybrid method, the weights have been set to $\omega_l = \hat{\omega}_l$ and $\omega_d = \hat{\omega}_d$. Furthermore, the parameters for all experiments are detailed in Table 7.2.

For a standard comparison, only circular beam shapes with aperture angle of 2° are allowed. The footprint contour of each shape and altitude is found by simulating the constellation every second for 24h and finding the contour over all times. All simulations have used the commercial solver Gurobi [234] (version 9.1.2) in an Intel(R) Xeon(R) Platinum 8160 CPU @ 2.10GHz, allowing up to 16 simultaneous threads.

Figure 7-2 and Table 7.3 present the simulation results, yielding insightful observations. As shown in Chapter 4, the developed joint method of Beam Shaping and User Grouping, termed

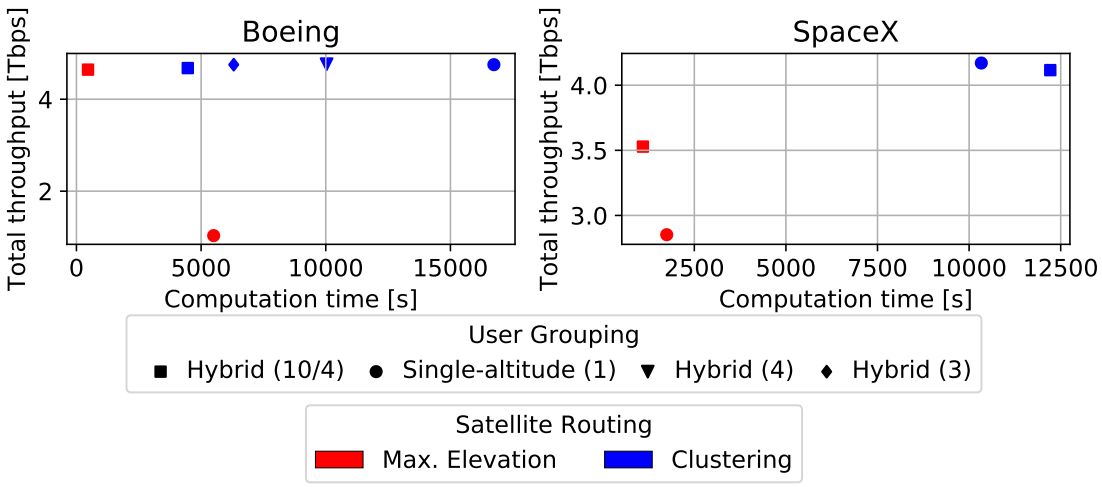


Figure 7-2: Performance when combining the joint Beam Shaping and User Grouping with the Satellite Routing. *Hybrid* refers to the method developed in Chapter 4, while *Single-altitude* refers to the method in [168]. *Clustering* refers to the method developed in Chapter 5, while *Max. Elevation* refers to the maximum elevation heuristic, where beams are mapped to the satellite with the highest elevation angle.

Constellation	Joint Beam Shaping and User Grouping (Num. $\mathcal{A}$)	Satellite Routing	Total throughput [Tbps]	Average active satellites	Total Power [kW]	Average Power on active satellites [W]	Spectrum usage [THz]	Computation Time [s]
Boeing	Hybrid (10)	Clustering	4.678	1116	165.9	148.6	3.941	4436
	Hybrid (10)	Max. Elevation	4.644	974	113.6	116.6	2.255	434
	Single-altitude (1)	Clustering	4.752	1068	282.2	264.2	2.385	16726
	Single-altitude (1)	Max. Elevation	1.033	331	85.36	256.7	0.365	5487
	Hybrid (3)	Clustering	4.761	1135	166.0	146.2	2.960	10032
	Hybrid (4)	Clustering	4.752	1087	176.8	162.5	2.619	6253
SpaceX	Hybrid (10)	Clustering	4.116	998	4.738	4.745	4.929	12195
	Hybrid (10)	Max. Elevation	3.529	798	2.720	3.407	2.223	1076
	Single-altitude (1)	Clustering	4.17	1051	5.705	5.425	3.369	10322
	Single-altitude (1)	Max. Elevation	2.852	698	1.820	2.605	1.662	1735

Table 7.3: Performance when combining the joint Beam Shaping and User Grouping with the Satellite Routing. *Hybrid* refers to the method developed in Chapter 4, while *Single-altitude* refers to the method in [168]. *Clustering* refers to the method developed in Chapter 5, while *Max. Elevation* refers to the maximum elevation heuristic, where beams are mapped to the satellite with the highest elevation angle.

hybrid, notably enhances capacity when coupled with a basic Satellite Routing policy. Specifically, for the Boeing and SpaceX constellations, the improvement stands at 350% and 24%, respectively.

However, with the integration of sophisticated Satellite Routing policies, such as the one detailed in Chapter 5, the hybrid approach proves to be counter-productive, but only marginally, reducing throughput by 2% in both the Boeing and SpaceX constellations. By maintaining a constant footprint and allowing greater flexibility in Satellite Routing, comparable constellation performance is achieved, illustrating that such flexibility need not be embedded within the decision-making process.

Particularly noteworthy in the context of the Boeing constellation is the optimization of satellite grouping based on altitude proximity. Optimal performance is achieved by partitioning satellites into three groups, ensuring that the maximum altitude difference within each group does not exceed 80% of the average altitude. This strategy yields a significant 40% reduction in computation time compared to heuristic approaches. When splitting the satellites into four groups, where the maximum altitude difference within each group does not exceed 50% of the average altitude, only a 0.2% reduction in throughput is observed, while further reducing the computation time by 38% compared to the three-group split, facilitating the resolution of instances with larger user bases. Conversely, for the SpaceX constellation, the adoption of the hybrid technique leads to a 18% increase in computation time compared to the single-altitude method.

Based on these findings, a strategic approach for addressing constellations with varying altitudes is proposed. Constellations featuring satellites at significantly different altitudes will be managed using the hybrid method due to its capacity for comparable throughput with substantially reduced computation time. In particular, altitudes will be split into groups, ensuring that the maximum altitude difference does not exceed 50% of the average altitude. Conversely, constellations with satellites at similar altitudes will adhere to the single-altitude method.

7.3 Performance of the RAP Framework

While various approaches exist for addressing the individual sub-problems comprising the RAP, this dissertation has prioritized the pursuit of *optimized* solutions, aiming to achieve a resource-efficient distribution. These optimization methodologies have demonstrated to yield efficient solutions within the context of individual sub-problems, thereby augmenting the capacity of the satellite system. Nevertheless, the benefit of these algorithms from a system perspective is yet to be determined. The objective of this section is to quantify the impact of employing an optimized allocation by

comparing two strategies:

1. **Optimized**: This entails the utilization of the RAP Framework employing the joint Beam Shaping and User Grouping method described in Chapter 4 for constellations characterized by significant altitude disparities. For constellations featuring satellites at similar altitudes, the method detailed in [168] is adopted, as discussed in the preceding section. Additionally, this strategy incorporates the coordinated Satellite Routing and Frequency Assignment technique explained in Chapter 5 and the Gateway Routing approach outlined in Chapter 6.

2. **Baseline**: In contrast, the baseline approach entails the use of the RAP Framework employing the User Grouping method for single-altitude constellations with fixed beam footprints as described in [168]. This strategy also employs the maximum elevation heuristic for the Satellite Routing sub-problem, as defined in Section 5.10, along with a water-filling-like algorithm for Frequency Assignment as described in [168]. The Gateway Routing approach outlined in [7] is also integrated into this sequence of algorithms. This approach aims to emulate the prevailing state of the industry, where optimization is seldom employed.

Notably, while the baseline strategy only provides one solution to the RAP, the optimized strategy allows for a myriad of points with different trade-offs between throughput, service quality, and number of necessary gateways. To provide insights on the trade-offs offered by the optimized strategy, a total of 14 points have been simulated, using the following logic:

- Initially, the five best points maximizing throughput are determined based on the weights outlined in Table 6.4.

- Subsequently, three additional points are identified, each providing three quarters ($\frac{3}{4}$) of the maximum throughput. This is achieved by employing varying weight configurations: $\omega_l = 0$ & $\omega_d = 1$, $\omega_l = \hat{\omega}_l$ & $\omega_d = \hat{\omega}_d$, and $\omega_l = 0$ & $\omega_d = 1$. The desired throughput level is maintained at three quarters of the maximum through the utilization of the parameter t_{min} in the Gateway Routing formulation.

- Finally, obtain the three best points that provide one half ($\frac{1}{2}$) and one fourth ($\frac{1}{4}$) of the maximum throughput, utilizing the same weight configurations as in the previous simulation.

These simulations are conducted with $N_{loc} = 20,000$ and $N_{us/loc} = 10$. Consistent with the preceding section, only circular beam shapes with a fixed aperture angle of 2° are considered. The

footprint contour for each altitude and shape is determined by simulating the constellation every second over a 24-hour period and extracting the contour across all time instances. All simulations are performed using the commercial solver Gurobi [234] (version 9.1.2) on an Intel(R) Xeon(R) Platinum 8160 CPU @ 2.10GHz, with support for up to 16 simultaneous threads.

To evaluate the results, the following metrics will be presented:

- Average throughput: This metric denotes the average data rate served across all time steps.
- Total number of Gateways: This represents the cumulative count of the maximum number of gateways utilized across all simulated time steps.
- Average latency: This metric quantifies the average time taken by each unit of information to traverse the satellite network, from departure from the gateway antenna to arrival at the user terminal. Unlike the previous chapter, this calculation now incorporates the time spent in transit to/from the ground terminals. This modification accommodates the comparison of different constellation designs at varying altitudes, rather than focusing solely on ISL or ground segment configurations. Notably, the computed time excludes the duration spent in the terrestrial network.
- Total power: This denotes the average transmission power by the entire satellite network on downlinks at each time step. Note that only the forward link is simulated.
- Spectrum Usage: This metric reflects the average spectrum utilization across all satellites in the network on the downlink beams. Again, only the forward link is simulated.
- Spectral efficiency: This metric is the ratio of the total throughput and spectrum usage, and measures how effectively different constellations use the spectrum assigned.
- Gateway Statistics: This includes the average number of active gateways and the average utilization percentage on both total and active links, relative to the ideal values. The utilization is the ratio between the computed throughput and the theoretical maximum throughput under optimal conditions, computed using a 90° elevation angle with no weather attenuation or interference, and assuming full link utilization across all frequency reuses. Note that the utilization can be understood as a *fill rate* regarding the gateway link.

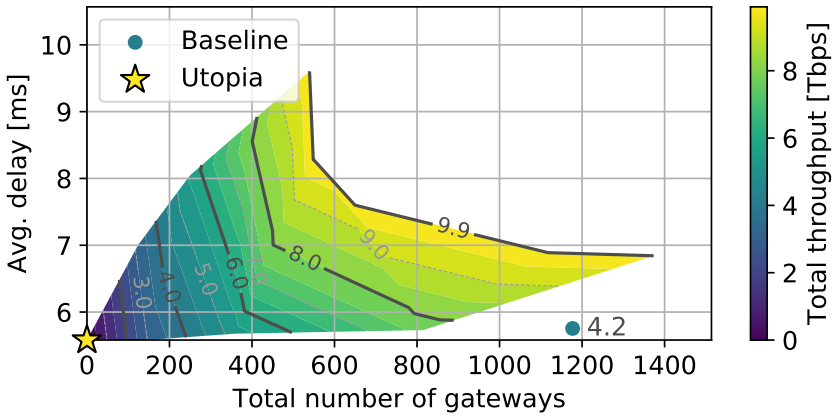


Figure 7-3: Performance comparison of the combination of methodologies developed in this dissertation against standard practice

			Metrics			**Gateway uplink**		
Strategy	w_g	w_{ISL}	Total throughput [Tbps]	Total number of gateways [-]	Average delay [ms]	Average active satellites gateway uplink [-]	Utilization of gateway uplink (total) [%]	Utilization of gateway uplink (active) [%]
Baseline	-	-	4.25	1177	5.75	838.6	3.889	22.43
Optimized	0	1	9.61	1376	6.848	1122	9.062	38.88
	0.2	1		1119	6.894	1018		42.87
	1	1		649	7.608	629.7		69.29
	5	1		549	8.287	546.8		79.40
	1	0		540	8.775	538.1		79.64

			User downlink				
Strategy	w_g	w_{ISL}	Total Power [kW]	Spectrum usage [THz]	Average active satellites user downlink [-]	Utilization of user downlink (total) [%]	Utilization of user downlink (active) [%]
Baseline	-	-	2.816	2.230	944.7	1.921	8.937
Optimized	All	All	13.65	5.262	1449	4.475	13.57

Table 7.4: Performance comparison of the combination of methodologies developed in this dissertation against standard practice. For the optimized strategy, only the points with highest throughput are shown.

- User Statistics: This comprises the average number of satellites hosting active users and the average utilization percentage on both total and active satellites, relative to the ideal values. Similar to the previous case, the utilization is computed under optimal conditions, and it can be understood as a *fill rate* regarding the user link.

Figure 7-3 and Table 7.4 show the results for the simulations. The first aspect to notice is that

the optimized strategy achieves more than twice as much throughput as the baseline strategy (126% more). The primary reason for this is the better distribution of load across satellites, which leads to a significantly higher capacity. The optimized strategy has around 53% more active satellites. Notably, the optimized strategy also allows for a better utilization of each individual satellite, both in terms of user downlink and gateway uplink. The second main aspect to notice is that the optimized strategy obtains a significantly better trade-off between quality of service and necessary amount of ground infrastructure. In particular, the optimized strategy yields solutions with the same delay and a third of the number of gateways for the same the throughput.

The results also highlight the optimized strategy remarkable spectrum utilization, which is more than double that of the baseline strategy (136% increase). This enhanced spectrum utilization not only augments total throughput but also escalates transmission power by a factor of 4.8. The optimized strategy achieves this increase by efficiently distributing load across satellites and optimizing Frequency Assignment.

Figure 7-3 provides valuable insights for estimating the total capacity of the constellation under additional constraints. For instance, if the satellite operator ground segment design were limited to 600 gateways, the constellation could achieve nearly maximum throughput at the expense of reduced quality of service. However, the operator could opt to sacrifice approximately 30% of capacity to reduce total delay in the satellite network by a quarter. Additionally, it is observed that when not utilizing ISLs (i.e., when the total delay is minimized), the maximum capacity of the constellation at 600 gateways diminishes by approximately 35%. This analysis underscores the significance of the proposed methodology in providing insights into the trade-offs affecting constellation design and ancillary infrastructure.

7.4 Constellation performance under realistic operational conditions

Accurate estimations of constellation performance play a crucial role in guiding stakeholders and operators towards informed decisions regarding system design and operations. However, existing literature often relies on simplified or ideal conditions, as elucidated in Section 2.4 [7,192]. In contrast, the preceding section demonstrated the pragmatic estimations achievable through the proposed RAP framework by simulating realistic operational conditions. Building upon this foundation, the

objective of this section is to assess and compare the performance of current megaconstellation designs under realistic operational scenarios, leveraging the proposed RAP framework.

To accomplish this, all nine megaconstellation designs outlined in Table 1.1 have been simulated. Note that these designs correspond to the values obtained from the filings, and may be altered at the discretion of the operator upon deployment. A total of 500,000 users have been distributed across 25,000 locations ($N_{loc} = 25,000$ and $N_{us/loc} = 20$), resulting in a maximum throughput of 50 Tbps, assuming 100 Mbps per user. Each configuration has been executed using the *optimized* strategy detailed in the previous section, employing consistent execution parameters. The nine designs have been categorized into four groups: 1) Low-LEO, 2) High-LEO, 3) MEO, and 4) Hybrid.

7.4.1 Low-LEO megaconstellations

This category comprises the two megaconstellations completely located at low-LEO altitudes (i.e., between 500 km and 1000 km): SpaceX [28] and Amazon [4]. Results are shown in Figure 7-4.

7.4.2 High-LEO megaconstellations

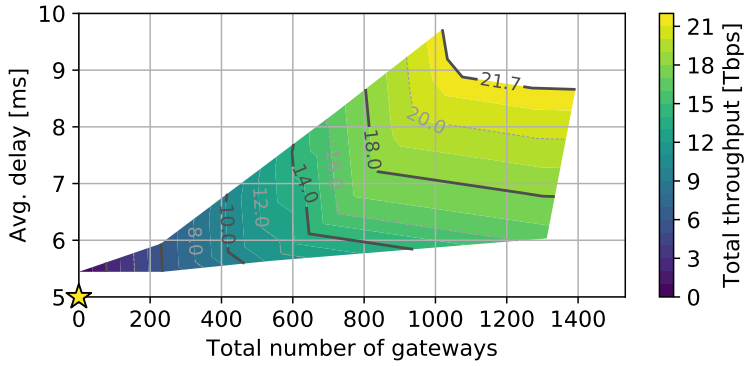
This category comprises the three megaconstellations completely located at high-LEO altitudes (i.e., between 1000 km and 2000 km): OneWeb [24], Telesat [32], and ViaSat [6]. Results are shown in Figure 7-5. For the purpose of comparison, it is assumed that the OneWeb system uses ISLs.

7.4.3 MEO megaconstellations

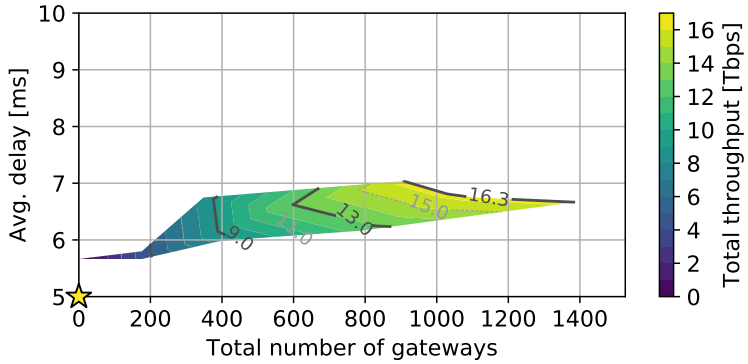
This category comprises the two megaconstellations completely located at MEO altitudes (i.e., between 8000 km and 10000 km): SES-O3b [35] and Intelsat [38]. Note that, while SES-O3b constellation has some satellites in LEO, their number is not sufficient to cover continuously any significant region of the world. Therefore, it is considered that the constellation operates in mostly a MEO regime. Results are shown in Figure 7-6.

7.4.4 Hybrid megaconstellations

This category comprises the two megaconstellations expanding across more than one altitude group: Boeing [30] and CASC [37]. Results are shown in Figure 7-7.



(a) SpaceX



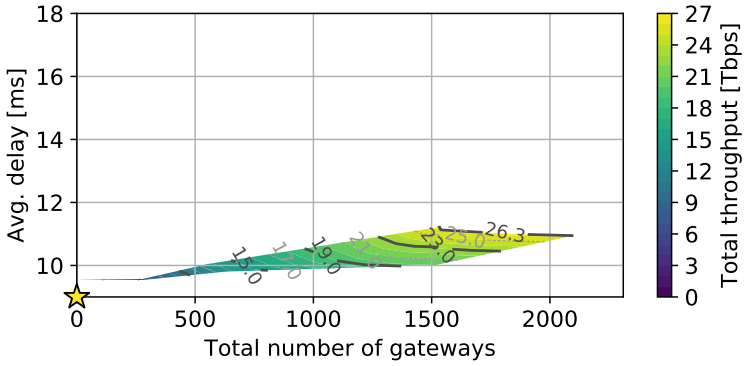
(b) Amazon

Figure 7-4: SpaceX and Amazon megaconstellation performance with 500,000 users

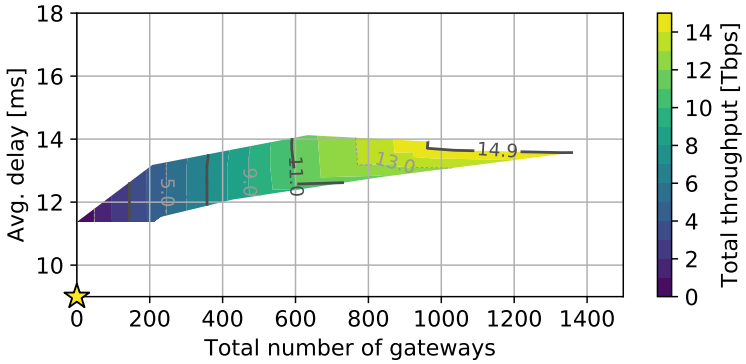
7.4.5 Summary and discussion of results

Figures 7-4, 7-5, 7-6, and 7-7 provide insights into the throughput, number of gateways, and average delay associated with the nine existing megaconstellation designs. Additionally, for comparison purposes, Figure 7-8 compares the maximum throughput and number of satellites, as well as the number of gateways and delay under maximum throughput.

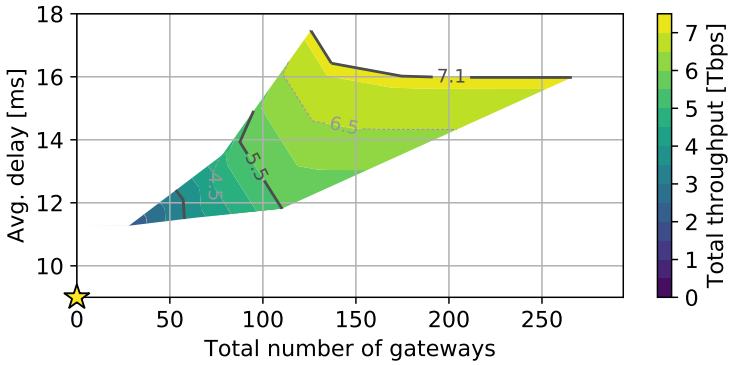
The first aspect to notice is that a higher number of satellites generally correlates with higher throughput, albeit not in a strictly linear manner. It is also evident that no constellation is capable of serving the entire user demand, indicating bottlenecks either in user or gateway links. As expected, constellations with fewer satellites achieve lower throughputs, the lowest being SES-O3b (4.5 Tbps), followed by ViaSat (7.2 Tbps) and Intelsat (7.9 Tbps). Interestingly, despite Intelsat having fewer



(a) OneWeb

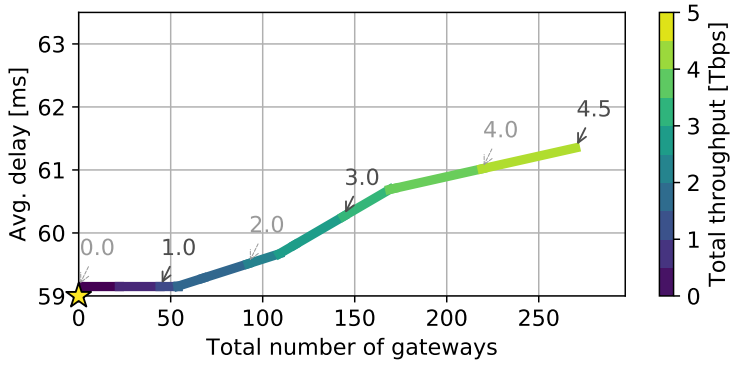


(b) Telesat

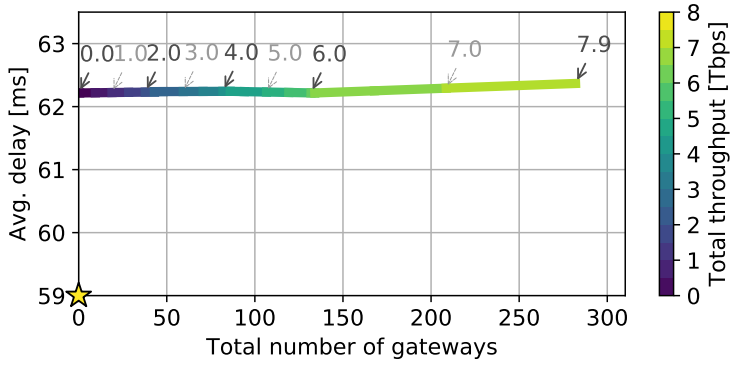


(c) ViaSat

Figure 7-5: OneWeb, Telesat, and ViaSat megaconstellation performance with 500,000 users



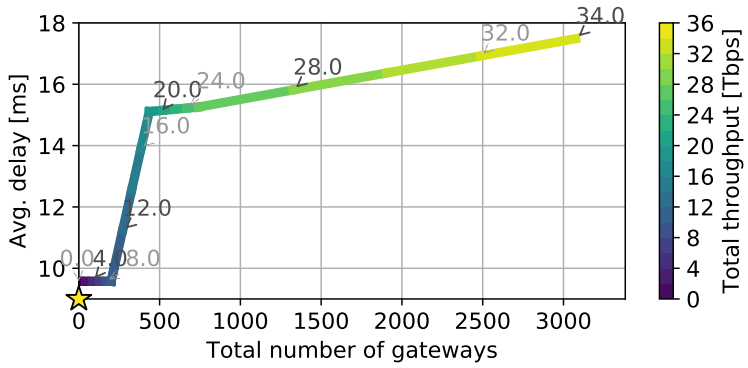
(a) SES-O3b



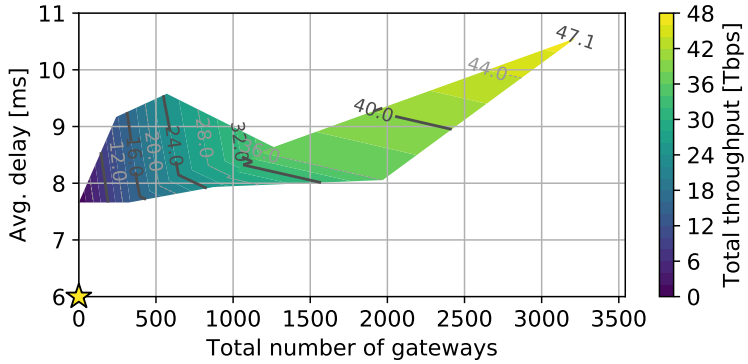
(b) Intelsat

Figure 7-6: SES-O3b and Intelsat megaconstellation performance with 500,000 users

satellites compared to ViaSat (216 vs. 288), its throughput is not adversely affected due to its MEO satellites ability to observe larger portions of the Earth. Telesat achieves nearly twice the throughput of Intelsat at 14.9 Tbps, albeit with five times the number of satellites. This is attributed to the marginal returns of adding satellites, which primarily enhance capacity in previously uncovered dense regions. Similar observations hold for Amazon (16.4 Tbps), SpaceX (21.8 Tbps), and OneWeb (26.3 Tbps), where the latter throughput is capped around 20 Tbps without using ISLs. Notably, Boeing achieves higher capacity than OneWeb with fewer satellites (34.0 Tbps), owing to lower altitude satellites, which enhance the link quality, and a larger number of visible satellites, resulting from a combination of LEO, MEO, and HEO altitudes. CASC outperforms all other constellations with a throughput of 47.1 Tbps, facilitated by its extensive satellite count at lower altitudes, coupled



(a) Boeing



(b) CASC

Figure 7-7: Boeing and CASC megaconstellation performance with 500,000 users

with a wide frequency spectrum reaching up to 8.55 GHz on user downlink—a significant advantage compared to other megaconstellations like OneWeb (2 GHz), SpaceX (2 GHz), Telesat (1.8 GHz), Amazon (2 GHz), or Boeing (2 GHz).

It is important to note that a higher number of satellites also correlates with a greater number of gateway antennas. Constellations with a lower number of satellites (ViaSat, Intelsat, and SES-O3b) require a maximum of 282 gateway antennas to achieve full capacity. This number escalates to thousands for the other systems. Interestingly, both Amazon and SpaceX manage to provide higher throughput than Telesat with the same number of gateways. This disparity is attributed to Telesat lower elevation angle, which diminishes the capacity of communication links. Amazon and SpaceX, on the other hand, leverage higher elevation angles for gateway connectivity, enabling them to

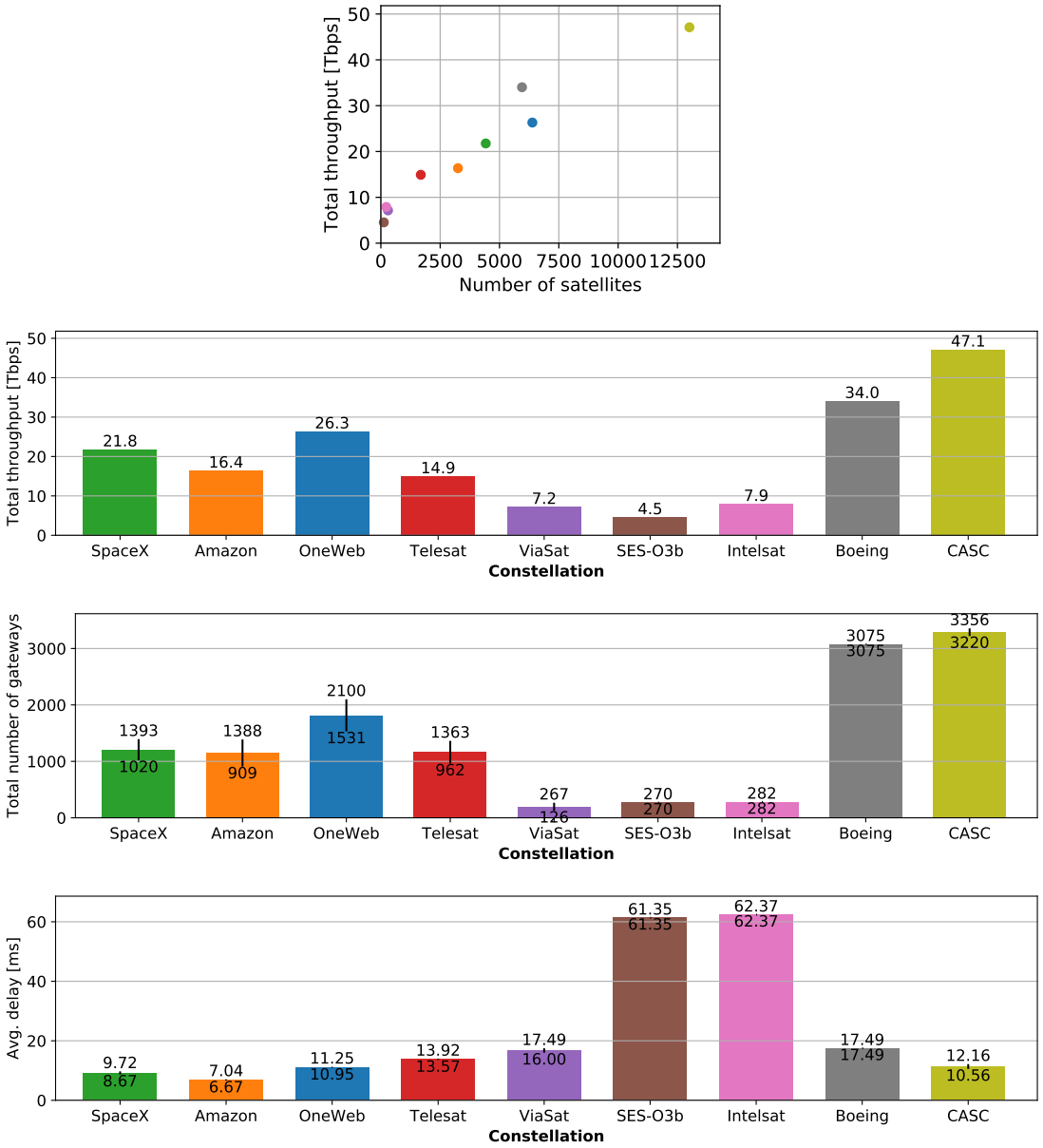


Figure 7-8: Performance comparison on the nine megaconstellation designs

achieve higher capacities in these links and thus reducing the need for gateways. To achieve higher throughput, OneWeb, Boeing, and CASC necessitate a substantial number of gateways—1,531, 3,075, and 3,220 respectively, representing the minimum required to attain maximum throughput

Constellation	Total throughput [Tbps]	Total number of gateways [-]	Average latency [ms]	Spectrum utilization [THz]	Spectral efficiency [bps/Hz]	User downlink utilization [%]	Gateway uplink utilization [%]
SpaceX	21.8	1020 - 1393	8.67 - 9.72	6.5	3.35	6.5	19.9
Amazon	16.4	909 - 1388	6.67 - 7.04	3.9	4.21	8.8	10.6
OneWeb	26.3	1531 - 2100	10.95 - 11.25	8.8	2.99	5.4	17.3
Telesat	14.9	962 - 1363	13.57 - 13.92	4.2	3.55	13.0	22.3
ViaSat	7.2	126 - 267	16.00 - 17.49	2.1	3.43	13.0	3.7
SES-O3b	4.5	270	61.35	1.1	4.09	19.3	17.7
Intelsat	7.9	282	62.37	3.4	2.32	7.1	25.6
Boeing	34.0	3075	17.49	12.4	2.74	3.0	1.2
CASC	47.1	3220 - 3356	10.56 - 12.16	22.7	2.07	1.1	2.7

Table 7.5: Performance comparison on the nine megaconstellation designs

for each constellation.

Regarding average delay from gateway to user, it is primarily influenced by the altitude of the constellation. SpaceX and Amazon achieve the lowest delays, at less than 10 ms, followed by CASC, OneWeb, Telesat, ViaSat, and Boeing, at less than 20 ms. Notably, CASC achieves delays between those of low-LEO constellations (SpaceX and Amazon) and high-LEO constellations (OneWeb, Telesat, and ViaSat) due to its satellites being positioned between these two altitudes. Furthermore, despite employing MEO and HEO satellites, Boeing achieves low average delays thanks to its significantly larger satellite network at LEO altitudes. Conversely, MEO constellations like Intelsat and SES-O3b can only offer around 60 ms delay due to altitude constraints. It is important to note that for MEO constellations, no ISLs are assumed, meaning that throughput, number of gateways, and delay are related through a single curve rather than a surface, as observed in other cases. This implies that any change in one metric will affect the other two. Despite employing ISLs, Boeing also behaves like a unique curve due to its MEO and HEO connections.

Table 7.5 presents additional metrics including spectrum and link utilization for each constellation. Notably, throughput scales in correlation with spectrum utilization. The disparity in spectrum utilization across constellation designs arises from four key factors: 1) The number of visible satellites. Since a larger number of visible satellites facilitates better frequency reuse, 2) Spectrum availability, which varies depending on the constellation filings, 3) Payload technology, which dictates how efficiently a single satellite can reuse the frequency pool, and 4) Link quality, since lower link quality requires more spectrum to achieve the same throughput. Based on this, the constella-

tions with lower spectrum utilization are the one with lowest throughput (SES-O3b, ViaSat, and Intelsat). Notice how Intelsat requires more than 50% of the spectrum than ViaSat to achieve only slightly higher throughput, despite having less satellites. This disparity is attributed to the fact that Intelsat satellites are in MEO, while ViaSat satellites are in LEO. MEO satellites are Intelsat to observe larger portions of the Earth, which enables them to serve more users with less satellites, at the cost of a higher free space loss, leading to lower link quality. They compensate the lower quality by using more spectrum, leading to low spectral efficiencies. On the other side, ViaSat observes significant lower free space loss, leading to higher quality links and not requiring large spectrum usage. However, as their satellites observe smaller portions of the Earth, they are able to cover less users overall. Note that SES-O3b has also satellites in MEO, but achieves a significantly higher spectral efficiency compared to Intelsat. The reason for this is that Intelsat allows users to use the Q-band, while SES only contemplates the Ku- and Ka-bands. To achieve lower transmission power, Intelsat users use a larger spectrum, leading to low spectral efficiency, while SES-O3b users use higher power.

In terms of larger constellations, Amazon achieves higher throughput than Telesat using less spectrum due to the lower link quality of Telesat satellites, stemming from their higher altitude and lower elevation angle. Boeing achieves higher throughput than OneWeb due to higher spectrum utilization, stemming from better links and a larger number of visible satellites. CASC boasts the largest spectrum utilization owing to its extensive satellite network and ample spectrum availability. Note that most systems achieve a spectral efficiency between 3 bps/Hz and 4.25 bps/Hz. The only exceptions are Intelsat, for the reasons mentioned before, and the hybrid systems of Boeing and CASC. Similar to Intelsat, Boeing and CASC allow users to use the Q-band, which allows to reduce the transmission power by using larger portions of the spectrum, thereby achieving lower spectral efficiency.

Regarding user and gateway utilization, none surpass 20% and 30% respectively. In other words, up to 80% of user downlink and 70% of gateway uplink capacity remain unused, primarily due to the large number of satellites over oceanic regions. As a secondary factor, interference between geographically close users, especially in MEO or very large constellations ($> 2,000$ satellites), reduces effective frequency reuse factor and thus lowers utilization. This is the reason while, despite observing larger portions of the Earth, MEO satellites exhibit similar utilization compared to LEO. Notably, Boeing and CASC exhibit significantly lower utilization across all constellations. This is

Constellation	[8]			This dissertation		
	Total throughput [Tbps]	Total number of gateways [-]	Utilization of gateway uplink (total) [%]	Total throughput [Tbps]	Total number of gateways [-]	Utilization of gateway uplink (total) [%]
SpaceX	27.2	∼2500	31.3	21.8	1020 - 1393	19.9
Amazon	53.4	>4000	32.5	16.4	909 - 1388	10.6
OneWeb	30.3	∼2500	24.2	26.3	1531 - 2100	17.3
Telesat	25.4	∼3000	44.3	14.9	962 - 1363	22.3

Table 7.6: Comparison of the analysis performed in this dissertation with prior studies

attributed to their assumed highly capable satellites, which allow for higher throughput but are limited by interference within and between satellites, resulting in unused capacity. This underscores the importance of designing satellites with realistic operational capabilities to avoid adding cost without explicit benefits to the operator.

7.5 Performance comparison against existing analyses

Some of the megaconstellation designs analyzed in the previous section have been previously studied in literature [8] using ideal models to avoid the complexity of the RAP. The objective of this section is to contrast ideal and realistic approaches when analyzing the performance of megaconstellations to provide insights on the advantages and disadvantages of each method.

Table 7.6 compares the outcomes of the analysis performed in this dissertation against those of [8]. The first aspect to notice is that the throughput obtained in this dissertation using realistic methods is lower than that obtained in [8] through ideal methods. This disparity is particularly pronounced in constellations with 2 gateway antennas per satellite (Telesat and Amazon). The reason for this difference lies in the assumption made in [8] that the systems are supply-limited, with significantly higher demand (up to a maximum of 240 Tbps). Under supply-limited conditions, satellites are expected to connect to as many gateways as possible to address the high data demand, thus requiring a high number of gateways. Conversely, this dissertation studies both supply and demand simultaneously, without explicitly limiting them. It is important to note, however, that none of the four systems analyzed in both studies are able to satisfy all the demand. This implies that saturation occurs in some areas of the systems. Since [8] proves that the supply link can achieve much higher limits, it is concluded that the bottleneck occurs in the user downlink, owing

to the possible interference between geographically close users, which limits the capabilities of the satellites. Note that interference is not included in the model provided in [8], but it is included in this dissertation. Based on the new information provided in this dissertation, some conclusions of [8] need to be revisited:

- In supply-limited scenarios, both the number of gateway antennas per satellite and the number of satellites are the primary drivers of capacity. However, in demand-limited conditions, the number of satellites, quality of the link, and the available spectrum become the main factors influencing capacity. This work demonstrates that, when assuming a population-like distribution with 500,000 users, megaconstellations operate under a demand-limited regime. It is crucial to recognize that under these conditions, adding additional gateway antennas only offers minimal value to total capacity, while potentially incurring significant cost implications. As highlighted previously, designing satellites with realistic operational capabilities is essential, as excessive flexibilities may inflate costs without delivering tangible benefits to the operator.

- Under realistic scenarios, the number of necessary gateways is substantially lower than originally estimated. On one hand, this reduction stems from the reduced throughput of the constellations in demand-limited conditions. On the other hand, the rest of the reduction is due to the ground segment optimization, which reduces the necessity for gateway antennas. On the SpaceX constellation, the reduction in throughput accounts for 20% of the reduction in gateway antennas, while the other 30% is accounted by optimizing the ground segment, leading to a total reduction of around 50% compared to prior studies.

- Due to demand-limited conditions, satellite utilization is notably lower than anticipated. Most constellations fail to achieve 20% utilization, with Telesat being the only exception at 22.3%, owing to its lower elevation angle and reduced number of satellites.

7.6 Chapter summary and conclusions

This chapter has presented a complete framework to address the RAP in modern megaconstellations, which includes a study on the interactions between specific sub-problems included in this

dissertation. Then, this framework is used to analyze the performance of existing megaconstellation designs under realistic operational conditions, which provides new insights on the expected performance of the novel systems.

7.6.1 Chapter summary

The chapter initiates with Section 7.1, which integrated the previously discussed components from preceding chapters with existing solutions for the remaining RAP sub-problems, delineating a comprehensive framework to address the complete RAP. Following this, Section 7.2 delved into the interactions between the joint Beam Shaping and User Grouping technique and the Satellite Routing approach, determining the optimal application of each algorithm based on constellation configuration. Section 7.3 conducted a comparative analysis between the performance of the optimized RAP framework and heuristic approaches from literature, revealing significant enhancements in throughput, required ground infrastructure, and service quality achieved by the methods developed in this dissertation. Subsequently, in Section 7.4, the optimized framework was applied to existing megaconstellation designs, illustrating that the anticipated performance of current systems primarily hinges on the number of satellites, link quality, and spectrum availability. Finally, Section 7.5 compared the findings of this dissertation with existing literature, challenging the prevailing notion that current megaconstellation designs, when subjected to realistic operational conditions, would be supply-limited, as it was observed that the user downlink serves as the bottleneck link.

7.6.2 Response to Research Questions

Research question 7.1

What is the performance of modern constellations under realistic operational conditions in terms of throughput, required ground infrastructure, and latency? How does this performance compare to ideal operational conditions?

Current megaconstellations are expected to achieve throughputs on the scale of Terabits per second (Tbps). Smaller megaconstellations, such as SES-O3b, ViaSat, and Intelsat, are forecasted to reach between 4.5 Tbps (SES-O3b) and 7.9 Tbps (Intelsat). In contrast, larger systems like OneWeb, SpaceX, Telesat, Amazon, Boeing, and CASC are projected to expand between 14.9 Tbps (Telesat) and 47.1 Tbps (CASC). The primary drivers of throughput are the number of satellites deployed,

quality of the link (driven by the distance to the ground), and the available frequency spectrum. The total number of gateways varies in proportion to the size of the constellations, ranging from 126 to 282 gateways for smaller systems to approximately 3,000 for the largest systems. Notably, SpaceX, Amazon, and Telesat demonstrate a reduced requirement, needing only around 1,000 gateways. In terms of delay, it is predominantly related to the altitude of the constellation, with SpaceX and Amazon exhibiting lower latency (less than 10 ms from gateway antenna to user terminal), CASC, OneWeb, Telesat, ViaSat, and Boeing in the mid-range (less than 20 ms), and MEO systems such as SES-O3b and Intelsat on the higher end (exceeding 60 ms). Spectrum utilization of each system scales proportionally with throughput, with utilization generally remaining low across all systems, primarily due to empty regions combined with interference among satellites.

Contrary to ideal operational conditions, this dissertation demonstrates that under realistic operational conditions, megaconstellations operate in a demand-limited regime rather than a supply-limited one. This implies that the number of satellites significantly impacts total throughput by facilitating more efficient spectrum utilization on the user link. Notably, certain systems with substantial supply capabilities, such as Telesat and Amazon, have been notably overestimated in terms of throughput, with actual capacities ranging between 42% and 70% lower than initially anticipated, respectively. Nonetheless, this study also reveals that ground station requirements are considerably lower than expected. For instance, SpaceX requires approximately half the number of gateway antennas to deliver around 20% less throughput.

Research question 7.2

When splitting users into altitudes, do the benefits of using altitude-appropriate beam shapes overcome the handicaps imposed by restricting the satellite selection to only those at the assigned altitude in terms of throughput?

The impact of using altitude-appropriate beam shapes on throughput depends on the constellation and the Satellite Routing technique employed. For simpler Satellite Routing algorithms, employing altitude-appropriate shapes consistently enhances throughput. However, with optimized Satellite Routing algorithms, the optimal approach varies based on the constellation altitude distribution. For constellations like SpaceX, where satellites are positioned at similar altitudes, employing a uniform shape for all beams and granting greater flexibility to the Satellite Routing algorithm is advantageous. This strategy maximizes throughput while reducing computation time. Con-

versely, for constellations with satellites at diverse altitudes, such as Boeing, clustering altitudes into groups (e.g., low-LEO, high-LEO, MEO, and HEO) and utilizing altitude-appropriate shapes for these groups benefits performance, although restricting the flexibility of the Satellite Routing algorithm. This approach yields the highest throughput at a reduced computation time.

Research question 7.3

How much better do the methods presented in this dissertation perform against existing methods in terms of throughput, number of gateway antennas, and average delay?

By implementing the optimized methods developed in this dissertation, a twofold increase in total capacity is achievable without requiring additional hardware modifications. Moreover, the results demonstrate a significantly improved trade-off between the number of gateway antennas and delay. Employing optimized methods allows operators to either reduce the number of gateway antennas to a third for the same throughput and delay or double the throughput with the same number of antennas while increasing delay by approximately 20%. Additionally, it is shown that optimized methods enhance spectrum utilization by a factor of two.

7.6.3 Specific chapter contributions

The specific contributions of this chapter are as follows:

- Developed a novel complete RAP framework by combining the methods described in this dissertation with existing methods for the rest of the sub-problems. Notably, this is the first framework capable of addressing from an optimization perspective the flexibilities of novel megaconstellations.
- Studied the interactions between the joint Beam Shaping and User Grouping problem with the Satellite Routing problem, determining the applicability conditions for each technique.
- Studied the performance of the proposed RAP framework against current standard, proving that it provides significant benefits to the operators in terms of throughput, necessary ground infrastructure, and quality of service.
- Analyzed the performance of current megaconstellation designs under realistic operational conditions, concluding that these systems operate in a demand-limited scenario, with the

user downlink being the bottleneck.

- Compared the results of this dissertation with prior studies, showing that the ideal methods used in existing literature tend to overestimate the capabilities of existing designs, especially those with high supply capabilities.

Chapter 8

Analyzing constellation design and user characterization

Hybrid satellite constellations, characterized by satellites at various altitudes, present a compelling option for satellite operators aiming to shape the next generation of satellite networks. These systems offer the potential to amalgamate the advantages inherent in LEO systems, characterized by high-quality links and minimal signal delay, with those of MEO satellites, enabling broader coverage and a diminished reliance on gateway antennas. The synthesis of these features promises enhanced capacity and quality of service, all within a cost-efficient framework. However, designing hybrid constellations introduces an additional layer of complexity compared to their single-altitude counterparts. Beyond the inherent design parameters associated with each altitude, operators must navigate nuanced decisions, such as determining the proportion of satellites distributed across altitudes and designing multiple satellites tailored to specific altitudes. Despite the allure of these hybrid configurations, Section 2.4 underscores the gap in the existing body of literature regarding the design of these systems, attributable largely to their novelty and intricacy. This chapter aims to breach this gap by 1) proposing an innovative methodology that evaluates diverse constellation designs by leveraging a combination of surrogate and realistic techniques, and 2) investigating crucial decisions integral to the design of hybrid constellations, with a particular emphasis on LEO and MEO systems.

The research questions this chapter aims to address are:

Research question 8.1

Given specific cost and capacity ratios between LEO and MEO satellites, what is the optimal proportion between LEO and MEO satellites to maximize throughput?

Hypothesis: If the cost and/or capacity ratios strongly favor one altitude over the other, deploying all satellites at that particular altitude is expected to yield optimal performance. For example, in scenarios where the cost of producing and deploying LEO satellites markedly outstrips the cost of MEO satellites, it is anticipated that deploying the entire satellite constellation in LEO will maximize performance. Conversely, when neither cost nor capacity ratios exhibit a pronounced favoritism towards a specific altitude, it is anticipated that a combination of both LEO and MEO altitudes will result in the highest capacity. Current estimations suggest that, under such conditions, the optimal design would include a substantial LEO segment (exceeding 1000 satellites) complemented by a smaller MEO segment (fewer than 100 satellites).

Research question 8.2

What demand magnitude is more appropriate for LEO constellations? And for MEO constellations?

Hypothesis: Current believe is that LEO systems are more appropriate for less demanding users (up to 200 Mbps per user), such as residential customers, while MEO systems are more appropriate for higher demanding users (more than 1 Gbps per user), such as oil platforms or maritime/aviation customers.

8.1 A framework to explore the design of hybrid constellations

The objective of this chapter is to delve into the design of hybrid megaconstellations. To that end, it is important to establish a design methodology that can provide insights on the performance of different designs. In this regard, the preceding chapter introduced a complete RAP framework that provides specific operations plans and performance metrics adapted for existing megaconstellations. Nevertheless, it is important to note that, while the outcomes of the method provide stakeholders with accurate and specific information on the throughput, quality of service, and required ground segment, it requires a long computation time to do so. In contrast to operational considerations,

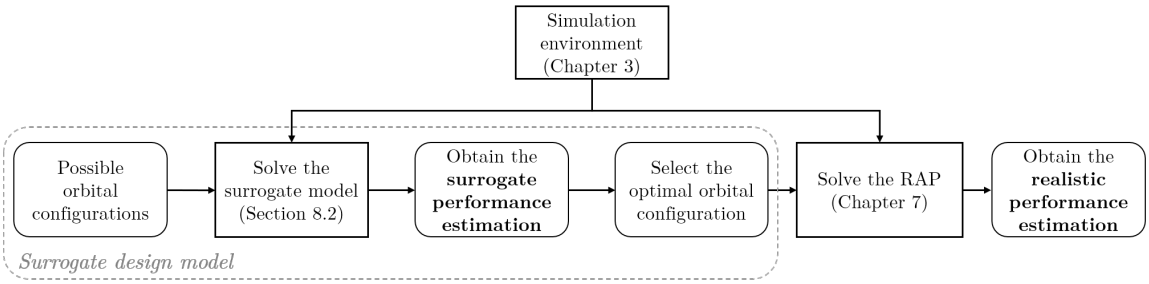


Figure 8-1: Block diagram of the framework to study the design of hybrid constellations

studying the design necessitates a rapid method of performance estimation, where the accuracy and specificity of the solution become less critical, as long as metrics remain consistent across systems.

With this objective in mind, a new framework is outlined, as depicted in Figure 8-1. Within this framework, multiple orbital configurations are swiftly assessed and compared utilizing a surrogate model based on a linear methodology, detailed in Section 8.2. Subsequently, the optimal configuration is chosen and subjected to analysis through the RAP framework introduced in the previous chapter. This approach facilitates expeditious exploration of the design space while still yielding accurate insights into performance metrics for promising designs.

8.2 Surrogate model for design selection

The objective of this section is to establish a novel surrogate model for swiftly evaluating the performance of a satellite constellation by constructing a set of linear equations akin to a graph, building upon the framework outlined in [8]. This proposed method enhances versatility by integrating additional user-satellite flows and considering interference between satellites, rendering it applicable to hybrid systems.

8.2.1 Problem description

The primary objective of a broadband satellite constellation is to establish connections between a set of users $\mathcal{U}$ seeking Internet services and a set of gateways $\mathcal{G}$ providing them. Each user u is defined by a specific position p_u and a demand d_u, while each gateway g is characterized by a position p_g. It is assumed that the characteristics of user and gateway antennas are known, and

gateways have unrestricted Internet access. For the purposes of this formulation, only the forward link is considered.

To facilitate connections between users and gateways, satellite operators have access to a set of satellites $\mathcal{S}$. At any given moment, each satellite s is characterized by a position p_s. It is assumed that the orbital parameters of the satellite, except for the mean anomaly, remain constant. Additionally, for simplicity, each satellite is assumed to have maximum uplink and downlink capacities denoted by c_s^U and c_s^D, respectively. These capacities can be computed using the link budget equations outlined in Section 1.2.2, while considering dual polarization and frequency reuse factors for each direction. Each satellite is permitted to transmit information to a subset of other satellites $\mathcal{S}_s$ through ISLs. It is assumed that ISLs have a maximum capacity of c_{ISL} and that their usage does not interfere with uplink or downlink communication.

Given the geometric considerations of the problem, each satellite at any given time has a subset of visible users and gateways, determined by the line-of-sight definition in Appendix A, denoted as $\mathcal{U}_s$ and $\mathcal{G}_s$ respectively. Symmetrically, each user and gateway have a subset of visible satellites denoted as $\mathcal{S}_u$ and $\mathcal{S}_g$ respectively. For simplicity, it is assumed that each user and gateway can simultaneously transmit data to all visible satellites. Imposing a restriction of only one connection per user or gateway would significantly escalate the problem complexity, as demonstrated in Section 6.3. Each gateway-satellite and satellite-user mapping is assumed to have a limited capacity $c_{g,s}$ and $c_{s,u}$ respectively. These capacities can be computed using the link budget equations described in Section 1.2.2. Additionally, it is assumed that these links cannot leverage multiple frequency reuses due to interference.

8.2.2 Problem formulation

The flow through each gateway-satellite, satellite-satellite, and satellite-user link is represented by $f_{g,s}$, $f_{s,s'}$, and $f_{s,u}$ respectively. Each flow is constrained by the capacity of its respective link. Moreover, the flow towards a user cannot exceed its requested demand, which is encoded as:

$$C1 : \sum_{s \in \mathcal{S}_u} f_{s,u} \leq d_u \quad \forall u \in \mathcal{U} \tag{8.1}$$

Similarly, the flow in the uplink and downlink directions from a satellite cannot exceed the total uplink and downlink capacity respectively, encoded as:

$$\begin{aligned} C2: \sum_{u \in \mathcal{U}_s} f_{s,u} \leq c_s^D \quad \forall s \in \mathcal{S} \\ C3: \sum_{g \in \mathcal{G}_s} f_{g,s} \leq c_s^U \quad \forall s \in \mathcal{S} \end{aligned} \tag{8.2}$$

Furthermore, since satellites do not retain information, the flow towards a satellite must equal the flow exiting the satellite, which is expressed as the following equality constraint:

$$C4: \sum_{g \in \mathcal{G}_s} f_{g,s} + \sum_{s' \in \mathcal{S}_s} f_{s',s} = \sum_{u \in \mathcal{U}_s} f_{s,u} + \sum_{s' \in \mathcal{S}_s} f_{s,s'} \quad \forall s \in \mathcal{S} \tag{8.3}$$

Lastly, to incorporate interference constraints, following the model described in Section 3.4, a binary parameter $I_{u,s,u',s'}$ indicates whether a satellite-user pair (u, s) may experience interference with another satellite-user pair (u', s'). In this formulation, when two connections have potential interference, it is assumed that each satellite can only utilize a portion of the total capacity of the link, and these portions cannot overlap between interfering connections. This concept extends to groups of three or more connections where all pairs share potential interference between them.

To formally define this concept, the set $I^U = \{I_{u,s,u',s'} \forall u \in \mathcal{U}, s \in \mathcal{S}_u, u' \in \mathcal{U}, s' \in \mathcal{S}_{u'}\}$ can be represented as a graph, where each satellite-user connection is a node, and potential interference between two connections is an edge. If there is potential interference between each pair of links, multiple links must share one physical connection. In graph terms, this implies that a subset of nodes must share one physical connection if it represents a fully connected sub-graph, also known as a clique. To introduce the necessary constraints, the set $\mathcal{L}^U$ represents the collection of cliques in the graph encoded in I^U. Then, the interference constraints can be encoded in linear form as:

$$C5: \sum_{(s,u) \in l} \frac{f_{s,u}}{c_{s,u}} \leq 1 \quad \forall l \in \mathcal{L}^U \tag{8.4}$$

A similar formulation can be applied for the gateway-satellite connections:

$$C6: \sum_{(g,s) \in l} \frac{f_{g,s}}{c_{g,s}} \leq 1 \quad \forall l \in \mathcal{L}^G \tag{8.5}$$

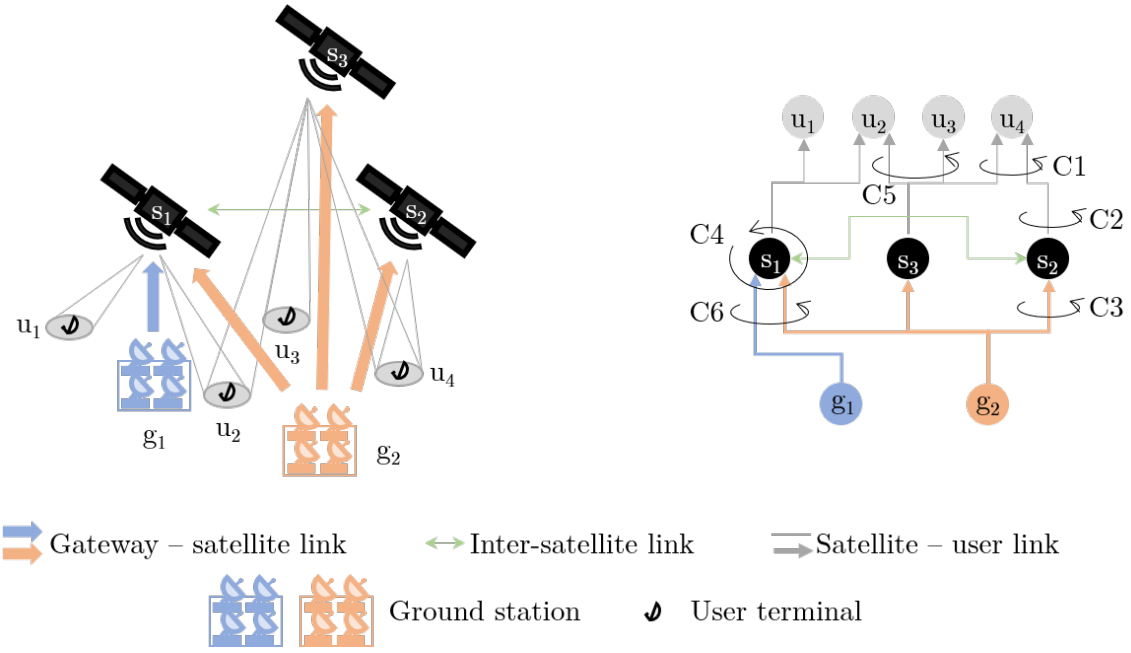


Figure 8-2: Representation of the surrogate model

It is assumed that satellite-user connections do not interfere with gateway-satellite connections. Notably, each link itself forms a clique since $I_{u,s,u,s} = 1$. However, computing the list of cliques in a graph is a known NP-Hard problem [231]. The following subsection details how to address this complexity to enable a quick resolution of the problem. Finally, the primary objective of the formulation is to find the configuration that maximizes the flow towards the users, expressed as:

$$\max \sum_{u \in \mathcal{U}, s \in \mathcal{S}_u} f_{s,u} \tag{8.6}$$

Since there can be multiple solutions to the problem due to the cycles introduced on the ISLs, a secondary objective is to minimize the hops in the satellite network, encoded as:

$$\min \sum_{s \in \mathcal{S}, s' \in \mathcal{S}_s} f_{s,s'} \tag{8.7}$$

The complete formulation can be written as:

$$
\begin{aligned}
\max \quad & \sum_{u \in \mathcal{U}, s \in \mathcal{S}_u} f_{s,u} \\
\min \quad & \sum_{s \in \mathcal{S}, s' \in \mathcal{S}_s} f_{s,s'} \\
s.t. \quad & C1: \sum_{s \in \mathcal{S}_u} f_{s,u} \leq d_u && \forall u \in \mathcal{U} \\
& C2: \sum_{u \in \mathcal{U}_s} f_{s,u} \leq c_s^D && \forall s \in \mathcal{S} \\
& C3: \sum_{g \in \mathcal{G}_s} f_{g,s} \leq c_s^U && \forall s \in \mathcal{S} \\
& C4: \sum_{g \in \mathcal{G}_s} f_{g,s} + \sum_{s' \in \mathcal{S}_s} f_{s',s} = \sum_{u \in \mathcal{U}_s} f_{s,u} + \sum_{s' \in \mathcal{S}_s} f_{s,s'} && \forall s \in \mathcal{S} \\
& C5: \sum_{(s,u) \in l} \frac{f_{s,u}}{c_{s,u}} \leq 1 && \forall l \in \mathcal{L}^U \\
& C6: \sum_{(g,s) \in l} \frac{f_{g,s}}{c_{g,s}} \leq 1 && \forall l \in \mathcal{L}^G \\
& 0 \leq f_{s,s'} \leq c_{ISL} && \forall s \in \mathcal{S}, s' \in \mathcal{S}_s \\
& 0 \leq f_{g,s} \leq c_{g,s} && \forall g \in \mathcal{G}, s \in \mathcal{S}_g \\
& 0 \leq f_{s,u} \leq c_{s,u} && \forall u \in \mathcal{U}, s \in \mathcal{S}_u
\end{aligned}
\tag{8.8}
$$

For reference, Figure 8-2 provides a visual representation of the problem structure and constraint definition. Notably, this formulation solely comprises continuous variables and linear constraints and objectives. Given that there is a prioritization between the objectives, commercial mathematical solvers can swiftly resolve this formulation. Additionally, it is apparent that this formulation resembles a min. cost-max. flow formulation, where users function as sinks, gateways as sources, satellites as intermediate nodes, and links as edges. However, the interference constraints (C5 and C6) deviate from this graph-like representation. This formulation offers a rapid estimation of the throughput of a constellation at a specific point in time, grounded in the theoretical limits of the payload.

8.2.3 Clique generation

As previously mentioned, determining the list of cliques in a graph is an NP-Hard problem. To streamline this, the complexity is mitigated by grouping users. In order to reduce the number of variables and mirror real communication systems, users are grouped based on proximity. These groups can be likened to beams in the actual system. Unlike beams, however, each group can communicate with all visible satellites simultaneously, thus maintaining the characteristics of the proposed formulation. To form user groups, the Earth's surface is partitioned into cells using spherical tesselation [233] to generate a grid of triangles. For the purpose of this dissertation, the grid is refined until the triangle size is smaller than the footprint of the beams employed in the real system. Subsequently, a group encompasses all users within a cell, with each cell constituting a distinct group. If a cell is empty of users, the group is disregarded. The group demand equals the sum of the demands of its users, and its position corresponds to the cell center position. The group antenna characteristics are assumed to be an amalgamation of all users antennas. Based on this simplification, each group operates as a unique user in the original formulation. Importantly, this grid-like definition also confines the potential size of the cliques, as each cell has a fixed number of neighboring cells with which it can potentially interfere. While the clique problem remains NP-Hard, this simplification enables the identification of the list of cliques in practical time through a brute-force algorithm.

8.3 Experimental set-up

By combining the advantages of LEO and MEO satellites, hybrid systems hold the potential to deliver enhanced performance at a comparable cost to single-altitude alternatives. However, understanding the conditions under which these systems excel is crucial. Complementing the described surrogate model, the objective of this section is to outline the experimental setup for studying hybrid constellations.

8.3.1 Parameter definition for a fair comparison

Given the higher altitude, MEO satellites observe higher path loss, necessitating compensatory measures such as higher antenna gains and/or increased power compared to LEO to achieve similar

throughput. Moreover, with the same aperture angle, MEO satellites produce larger footprints, heightening the potential for interference. Additionally, the cost of launching a specific mass is higher for MEO compared to LEO. These observations imply that, even within a hybrid system, LEO and MEO satellites will differ fundamentally in terms of payload and cost. From a satellite operator perspective, these differences may serve as primary factors in selecting one altitude over the other. For example, if manufacturing and deploying LEO satellites prove significantly cheaper than their MEO counterparts, it may be more advantageous to deploy a considerably larger network of LEO satellites, even when they are individually less capable compared to MEO satellites.

To address these differences, two key parameters are defined: the cost ratio and the capacity ratio. The cost ratio represents the cost of a MEO satellite relative to a LEO satellite. While determining the precise costs associated with spacecraft production, deployment, and operation lies beyond the scope of this dissertation, defining a ratio enables a straightforward *equivalence* concept. For example, a cost ratio of 2 signifies that an operator could opt to deploy either 2 LEO satellites or 1 MEO satellite for the same cost. Similarly, the capacity ratio denotes the equivalent capacity of a MEO satellite relative to that of a LEO satellite. In other words, given an ideal distribution of demand, the capacity ratio indicates how much more demand a MEO satellite can serve compared to a LEO satellite at any given point in time. This capacity discrepancy stems from the conventional understanding that since MEO satellites observe larger portions of the Earth, it is beneficial to making them more capable compared to LEO spacecraft. The capacity ratio is solely determined by the specific payload of each spacecraft. Note that both the cost and capacity ratios depend on the LEO and MEO satellite design and are not independent. However, different choices on the specific payload components and overall design can lead to different ratios.

It is important to emphasize that, given specific cost and capacity ratios, a fair comparison can only be made among isocost constellations—those with equivalent costs. Furthermore, it is important to acknowledge that the optimal proportion between LEO and MEO satellites depends on the specific cost and capacity ratios under consideration. For instance, a cost or capacity ratio favoring LEO would emphasize deploying more LEO satellites. Consequently, this dissertation will delineate the cost and capacity ratios for which each proportion is more advantageous.

8.3.2 Generic adaptive constellation

In order to define isocost constellations, a generic adaptive constellation is defined following subsequent rules:

- The constellation can deploy a total of 2,000 equivalent LEO satellites. As an example, for a cost ratio of 2, that means that isocost constellations can range from 2,000 LEO satellites and 0 in MEO, to 0 LEO satellites and 1,000 in MEO.
- LEO satellites are always located at 600 km altitude, with satellite divided in either 25 or 50 planes, and inclinations of 40°, 55°, or 70°.
- LEO satellites FoV has a minimum elevation angle of 25°.
- MEO satellites are always located at 9,000 km altitude, with satellite divided in either 1, 5, or 25 planes, and inclinations of 0°, 30°, 45°, or 60°.
- MEO satellites FoV has a minimum elevation angle of 25° or 40°.
- If the number of MEO satellites is more than 50, 1 MEO plane is not allowed.
- If the number of MEO satellites is 25 or less, 5 MEO planes are not allowed.
- If the number of MEO satellites is 200 or less, 25 MEO planes are not allowed.
- If the number of MEO planes is 1, only the 0° inclination is allowed.
- If the number of MEO planes is not 1, the 0° inclination is not allowed.

Given a specific cost ratio, capacity ratio, and LEO/MEO proportion, this set of rules defines a set of possible constellation designs in which the number of planes, inclination, and elevation angle can attain different values for both LEO and MEO satellites. To obtain the best design within this set, the algorithm defined in Section 8.2 is employed.

8.3.3 Fixed parameters

Irrespective of the chosen design, the satellite payload characteristics, user antenna characteristics, and gateway antenna characteristics are defined in Tables 8.1, 8.2, 8.3. Furthermore, the spectrum availability is described in Table 8.4. Notably, capacity ratio is encoded in the constellation as

Altitude	Transmission				Reception			
	Gain [dB]	EIRP density [dBW/Hz]	Pointing Loss [dB]	Rotation Loss [dB]	Gain [dB]	G/T [dB/K]	Pointing Loss [dB]	Rotation Loss [dB]
LEO	35	-44.03	0.1	0.1	35	10	0.1	0.1
MEO	55	-24.03			55	30		

Table 8.1: Satellite antenna parameters for the generic constellation

additional frequency reuses both in the user and gateway side. It is assumed that the power available scales linearly with the amount of frequency reuses. Only the forward link is simulated.

Parameter	Value	Parameter	Value
Transmission Power	20 dBW	G/T	25 dB
Antenna efficiency	0.65	LNB	2
Pointing loss	1 dB	Feeder loss	1.1 dB
Waveguide loss	0.2 dB	Additional loss	1 dB
Equivalent antenna diameter [m]	1		

Table 8.2: User antenna characteristics for the generic constellation

Diameter [m]	Gain [dB]	EIRP density [dBW/Hz]
4.5	60	-14.0

Table 8.3: Gateway antenna characteristics for the generic constellation

Capacity ratio	General				User	GW
	Num. of LEO GW antennas	LEO Frequency reuse factor	Num. of MEO GW antennas	MEO Frequency reuse factor	Downlink freq. [GHz]	Uplink freq. [GHz]
1	1	4	1	4	10.7 - 12.7	27.5 - 29.1 29.5 - 30
3	1	4	3	12		
10	1	1	†	10		

Table 8.4: Satellite spectrum capabilities for the generic constellation. † Antennas are shared between users and gateways. GW: Gateway

Analysis	Objective
Optimal proportion using the surrogate model	Determine the optimal proportion of LEO and MEO satellites when using the surrogate model
Optimal proportion under optimized RAP	Determine the optimal proportion of LEO and MEO satellites when using the optimized RAP
Impact of the RAP methodology	Determine if the results hold when using different RAP methodologies
Impact of the user distribution	Determine if the results hold across different user distributions
Impact of the constellation sizing	Determine if the results hold across different constellation sizes

Table 8.5: Experimental set-up to determine the optimal proportion of satellites across altitudes under various conditions

8.4 Optimal LEO-MEO proportion experiments and discussion

Based on this adaptive constellation model, this section studies the optimal proportion between LEO and MEO satellites through a set of experiments. Furthermore, it includes a discussion on the conditions for which each altitude is more appealing.

8.4.1 Experiments

To determine the optimal distribution of satellites across different altitudes and assess the generalizability of the findings, five distinct experiments are conducted, as outlined in Table 8.5.

Each experiment entails testing various cost and capacity ratios. Within each combination, six unique scenarios are considered, where a total of 0%, 10%, 25%, 50%, 75%, and 100% of the LEO satellites are exchanged for MEO satellites, at the appropriate cost ratio. In instances where the number of MEO satellites is not an integer, it is rounded down to the nearest whole number. For every scenario, the linear method detailed in Section 8.2 is used to derive the optimal constellation design based on the set of possible designs, followed by the application of the optimized or heuristic RAP methods outlined in Chapter 7. The surrogate model uses $N_{loc} = 25,000$ and $N_{us/loc} = 20$, while the RAP method employs $N_{loc} = 10,000$ and $N_{us/loc} = 10$. Unless explicitly stated, the user model follows the specifications provided in Section 3.2.

For a standard comparison, only circular beam shapes with aperture angle of 2° are allowed. The footprint contour of each shape and altitude is determined by simulating the constellation every

second over a 24-hour period and identifying the contour across all time points. All simulations use the commercial solver Gurobi [234] (version 9.1.2) running on an Intel(R) Xeon(R) Platinum 8160 CPU @ 2.10GHz, with support for up to 16 simultaneous threads. The experimental setup for the RAP method and the associated hyperparameters mirror those detailed in Section 7.3. Further details on the results can be found in Appendix D.

8.4.2 Optimal proportion using the surrogate model

The objective of this experiment is to determine the optimal proportion between LEO and MEO satellites within a hybrid constellation, leveraging the surrogate model developed in this chapter. To achieve this objective, five distinct cost ratios (2, 4, 6, 8, and 10) and three different capacity ratios (1, 3, and 10) are evaluated.

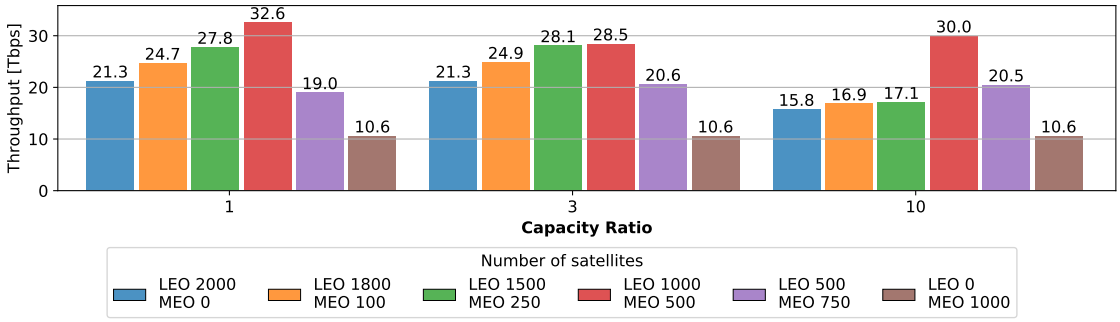


Figure 8-3: Throughput of the optimal design when using the surrogate model under a cost ratio of 2

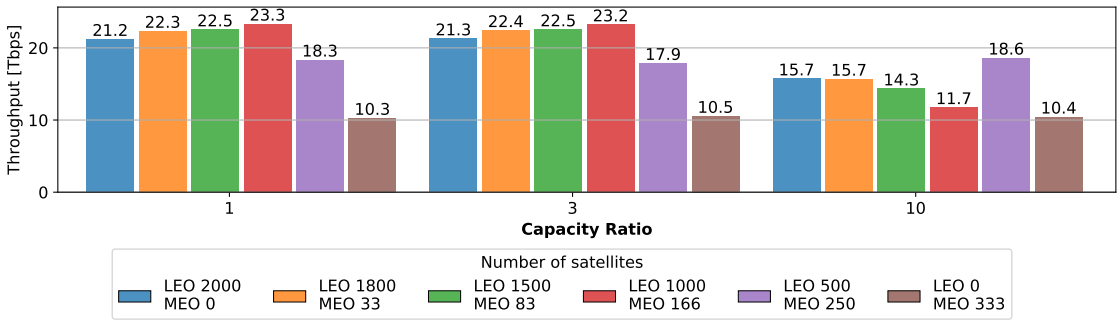


Figure 8-4: Throughput of the optimal design when using the surrogate model under a cost ratio of 6

Cost ratio	Capacity ratio		
	1	3	10
2	50% LEO 50% MEO	50% LEO 50% MEO*	50% LEO 50% MEO
4	50% LEO 50% MEO*	50% LEO 50% MEO	50% LEO 50% MEO
6	50% LEO 50% MEO*	50% LEO 50% MEO*	25% LEO 75% MEO
8	90% LEO 10% MEO**	90% LEO 10% MEO**	25% LEO 75% MEO
10	90% LEO 10% MEO**	90% LEO 10% MEO**	25% LEO 75% MEO

Table 8.6: Optimal budget split across altitudes when using the surrogate model. * and ** indicate that the difference between the best performing and second-to-best performing architectures is less than 5% and 1%, respectively.

Figures 8-3 and 8-4 depict the throughput of the optimal design achieved using the surrogate model for cost ratios of 2 and 6, respectively, across all capacity ratios. Results for other cost ratios are depicted in Appendix D. One notable observation is that for low capacity ratios (1 and 3), the optimal design entails allocating 1,000 satellites in LEO and the remaining in MEO. From a budget perspective, this equates to an equal distribution of costs between LEO and MEO. This trend is consistent even for high capacity ratios (10) paired with low cost ratios (2). However, under high capacity (10) and cost (6) ratios, the optimal design shifts towards MEO, achieved at the configuration comprising 500 LEO satellites and 250 MEO satellites. This implies a budget allocation of 75% to MEO and 25% to LEO.

Expanding on these findings, Table 8.6 illustrates the optimal budget allocation across all cost and capacity ratios according to the surrogate model. It is important to emphasize that to determine the optimal proportion of **satellites** across altitudes, the shown values need to be adjusted according to the cost ratio. As depicted, an equal budget distribution emerges as the optimal choice under low cost ratios. However, as the cost ratio increases, solutions tend to polarize towards the extremes: low capacity ratios steer the optimal proportion towards LEO, whereas high capacity ratios shift it towards MEO. Focusing on the last column (capacity ratio 10), it is observed that as the cost of MEO increases, the constellation tend to shift more towards MEO, which appears counter-intuitive. The reason for that is that low cost ratios allow for a significant number of satellites in MEO. However, having an excessive number of MEO satellites is counter-productive, as the interference between them increases significantly due to the larger footprints. Based on the results, it is observed that

having up to 500 MEO satellites is beneficial, but having more reduces the overall throughput as interference spikes. Therefore, as the cost ratio increases, the optimal design shifts more and more towards MEO, but without surpassing this 500 satellite threshold.

Finally, it is important to acknowledge that the difference between the best performing architectures reduces as the cost ratio goes up. When deploying a large network of LEO satellites, including a small portion of MEO satellites only has minor benefits on the throughput. In other words, as the cost ratio goes up and the preference shifts towards LEO, allocating budget towards MEO satellites only offers minor improvements over single-altitude designs.

8.4.3 Optimal proportion under optimized RAP

The objective of this experiment is to determine the optimal proportion between LEO and MEO satellites in a hybrid constellation when using the optimized RAP developed in this dissertation. Note that, for practicality reasons, only a sub-set of cost-capacity ratio combinations have been tested using the RAP methodology.

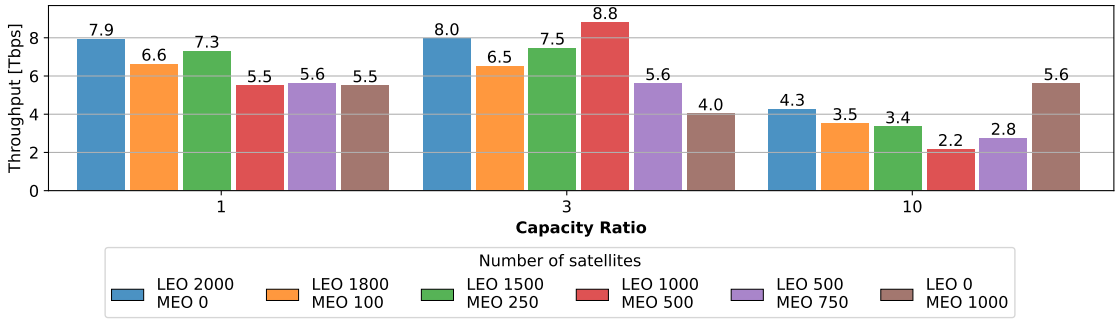


Figure 8-5: Throughput of the optimal design when using the RAP method under a cost ratio of 2

Figures 8-5 and 8-6 depict the throughput of the optimal design utilizing the RAP method with cost ratios of 2 and 6, respectively, across all capacity ratios. Notably, all designs exhibit lower throughput compared to the surrogate model, with hybrid designs experiencing a relatively greater decrease compared to single-altitude constellations. This discrepancy can be attributed to the impact of interference. While the surrogate model incorporates interference in its calculations, it affects only a single time-step at a time. In contrast, the RAP method considers interference across all time-steps. Hybrid systems are more susceptible to interference due to the greater relative

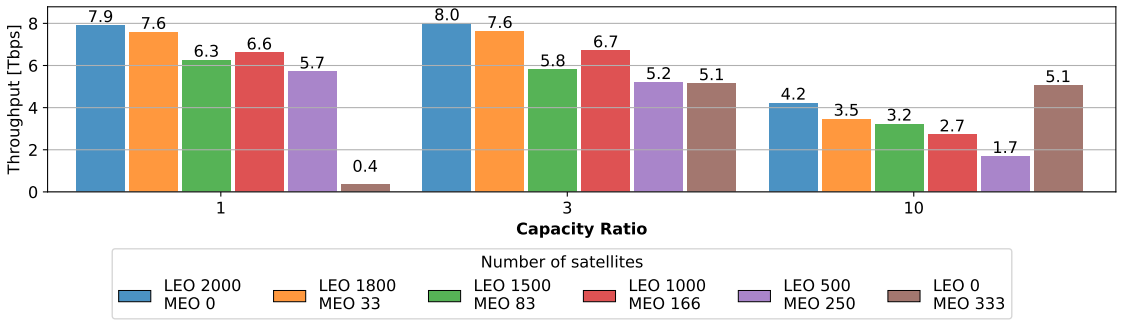


Figure 8-6: Throughput of the optimal design when using the RAP method under a cost ratio of 6

Cost ratio	**Capacity ratio** 1	3	10
2	100% LEO 0% MEO	50% LEO 50% MEO	0% LEO 100% MEO
4	100% LEO 0% MEO	100% LEO 0% MEO*	0% LEO 100% MEO**
6	100% LEO 0% MEO*	100% LEO 0% MEO	0% LEO 100% MEO
8	100% LEO 0% MEO†	100% LEO 0% MEO*	0% LEO 100% MEO†
10	100% LEO 0% MEO†	100% LEO 0% MEO†	0% LEO 100% MEO

Table 8.7: Optimal budget split across altitudes when using the RAP approach. * and ** indicate that the difference between the best performing and second-to-best performing architectures is less than 5% and 1%, respectively. Values marked with † have been extrapolated from the rest of the results.

velocity difference between satellites and the larger footprint of MEO spacecraft. Consequently, hybrid systems are less favorable when evaluating realistic operational scenarios. Nonetheless, there are specific design configurations where MEO-LEO hybrids outperform single-altitude designs. Notably, when the cost ratio is low (2) and MEO satellites possess slightly greater capabilities than LEO satellites (capacity ratio of 3), a balanced budget allocation between MEO and LEO yields superior performance over alternative solutions. Note that this mirrors the solution derived from the surrogate model, albeit with a narrower range of applicability in terms of cost and capacity ratios.

Expanding on these findings, Table 8.7 showcases the optimal budget allocation across all cost and capacity ratios using the RAP approach. Similar to the surrogate model, solutions for higher

Cost ratio	Capacity ratio		
	1	3	10
2	100% LEO 0% MEO	100% LEO 0% MEO	0% LEO 100% MEO**
4	100% LEO 0% MEO	100% LEO 0% MEO	0% LEO 100% MEO
6	100% LEO 0% MEO	100% LEO 0% MEO	0% LEO 100% MEO
8	100% LEO 0% MEO†	100% LEO 0% MEO	0% LEO 100% MEO†
10	100% LEO 0% MEO†	100% LEO 0% MEO†	0% LEO 100% MEO

Table 8.8: Optimal budget split across altitudes when using the heuristic approach. * and ** indicate that the difference between the best performing and second-to-best performing architectures is less than 5% and 1%, respectively. Values marked with † have been extrapolated from the rest of the results.

cost ratios tend to cluster towards the extremes. However, single-altitude designs tend to prevail over hybrid systems. Specifically, under a high cost ratio (4 or higher) and low capacity ratio (3 or lower), a fully LEO satellite system emerges as the dominant solution. Conversely, with a high cost ratio (4 or higher) paired with a high capacity ratio (10), a full MEO constellation design outperforms other configurations. Comparing the surrogate and optimized RAP models, it is seen that surrogate models benefit hybrid constellations, while the RAP model benefits single-altitude constellations. These results indicate that, under simplified conditions, hybrid models offer additional flexibilities over single-altitude constellations, leading to higher throughput. As operational restrictions are implemented, hybrid systems become less and less appealing as they are more complex to operate compared to single-altitude systems. In other words, while hybrid constellations are more appealing in theory, their implementation is more complex and might hinder some or all of the advantages they offer against single-altitude designs.

8.4.4 Impact of the resource allocation method

The objective of this experiment is to evaluate the influence of the resource allocation method on determining the optimal proportion between LEO and MEO satellites within a hybrid constellation. In contrast to the optimized RAP method outlined in this dissertation, this experiment employs the heuristic baseline methodology detailed in Section 7.3. Similar to the preceding scenario, only a subset of cost and capacity ratio combinations have been examined using the heuristic approach.

Cost ratio	Capacity ratio 1	3	10
6	75% LEO 25% MEO	50% LEO 50% MEO	90% LEO 10% MEO
10	100% LEO 0% MEO*	100% LEO 0% MEO	0% LEO 100% MEO

Table 8.9: Optimal budget split across altitudes when using the RAP approach, and users distributed according to the land area. Values marked with * have been extrapolated from the rest of the results.

Table 8.8 illustrates the optimal budget allocation across all cost and capacity ratios based on the heuristic approach. Notably, it becomes evident that there are no conditions wherein hybrid systems outperform single-altitude solutions. Compared to the previous experiment, the solution is identical, except for the cost ratio 2 and capacity ratio 3, where hybrid constellations loose their appeal. Once again, this is attributable to interference: heuristic methods are less capable to address interference, which, as discussed in the previous section, disproportionately impacts hybrid systems compared to single-altitude designs. Consequently, to overcome the limitations posed by interference, hybrid systems necessitate sophisticated resource allocation methods capable of harnessing the inherent flexibilities of these systems. In the absence of such methods, hybrid systems are dominated in terms of throughput by single-altitude designs.

8.4.5 Impact of the user distribution

The objective of this experiment is to assess the impact of user distribution on determining the optimal proportion between LEO and MEO satellites within a hybrid constellation. In contrast to selecting user locations based on population distribution, as outlined in Section 3.2, user locations are chosen proportionally to land area. This approach disperses users across the landscape rather than concentrating them in densely populated regions. The optimized RAP methodology is utilized, and, similar to previous scenarios, only a subset of cost-capacity ratio combinations have been explored using the RAP methodology.

Table 8.9 showcases the optimal budget allocation when employing the RAP method with cost ratios of 6 and 10, respectively, across all capacity ratios, with users distributed according to land area. Compared to the population-based distribution, hybrid systems become more attractive when users are dispersed across the land area. This shift in preference can be attributed, once again,

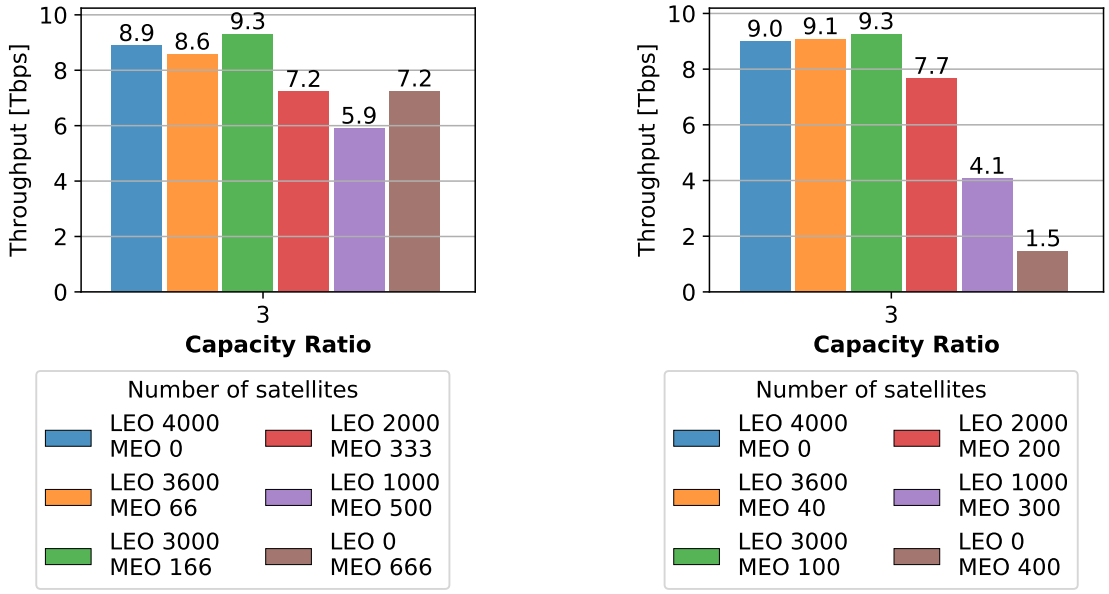
to interference: spreading users across regions reduces interference across beams and satellites, thereby benefiting hybrid systems. Notably, under a cost ratio of 6, hybrid solutions outperform single-altitude systems. The optimal proportion of satellites, however, varies based on the capacity ratio, encompassing a 75/25 budget split under low capacity (1), a 50/50 budget split in slightly higher capacity (3), and a 90/10 budget split at the highest capacity ratio (10). Consistent with previous scenarios, under high cost ratios, solutions tend to cluster towards the extremes, with LEO designs prevailing in low capacity scenarios (up to 3) and MEO designs dominating in high capacity scenarios (10).

Note that, this dissertation simulated two user distributions, one based on population, one based on land area. Under operations, the user distribution is likely to be somewhere in between these two due to two reasons. First, large cities and populated areas are likely to be covered by terrestrial infrastructure, and therefore less dependent on space-based connectivity. Second, connectivity is only required where population resides or works. Therefore, the requirements for space connectivity will be more sparse than population-based distributions, as very dense areas are already connected, and more dense than land-based distributions, as certain regions of the world are completely uninhabited, thereby not requiring connectivity. Furthermore, there will exist regional differences, as, for instance, rural areas in Europe are less dependent on satellite connectivity compared to rural areas in America, Asia, or Africa.

8.4.6 Impact of the constellation magnitude

The objective of this experiment is to evaluate the impact of the number of satellites on determining the optimal proportion between LEO and MEO satellites within a hybrid constellation. Instead of allocating a budget with 2,000 LEO satellites, a budget with 4,000 LEO satellites is considered. Utilizing the optimized RAP methodology, it is anticipated that a larger satellite count will further mitigate interference, resulting in improved performance of hybrid systems. To assess this, two combinations are evaluated, with cost ratios of 6 and 10, respectively, and capacity ratio of 3, while maintaining the user distribution described in Section 3.2.

Figure 8-7 illustrates the throughput of the optimal design achieved using the RAP method with cost ratios of 6 and 10, respectively, for a capacity ratio of 3, with a satellite budget of 4,000 LEO satellites. As expected, increasing the number of satellites enables the resolution framework to mitigate interference more effectively, thereby benefiting hybrid systems. Consequently, under these



(a) Cost ratio of 6

(b) Cost ratio of 10

Figure 8-7: Throughput of the optimal design when using the RAP method under a cost ratio of 6 (left) and 10 (right) with 4000 LEO satellites as budget

conditions, hybrid designs supersede single-altitude systems. Notably, the hybrid design with an equal budget split between LEO and MEO, which previously dominated other solutions with 2,000 satellites, no longer holds dominance under a budget of 4,000 satellites. Instead, the architecture that allocates 75% of the budget to LEO and 25% to MEO emerges as the dominant solution.

8.4.7 Recommendations on hybrid constellation design

In summary, hybrid systems outperform single-altitude solutions when the cost ratio is low and interference is minimal, whether due to dispersed user distribution, optimization methods leveraging system flexibility to mitigate interference, or a larger satellite network. In such scenarios, the prevalent design typically entails an equal budget split between LEO and MEO, although alternative hybrid configurations may dominate under specific circumstances. Conversely, when the cost ratio or interference is high, single-altitude systems take precedence over hybrid designs. Particularly, under low capacity ratios, full LEO designs achieve maximum throughput, whereas under high capacity ratios, full MEO designs deliver optimal performance. These insights are consolidated in

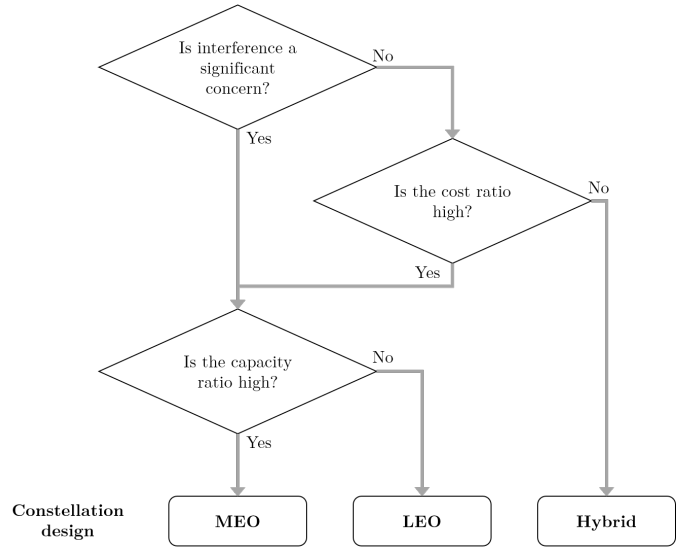


Figure 8-8: Recommendations on the design of satellite communications constellations

Figure 8-8. Note that these insights do not enclose specific numbers regarding the cost and capacity ratios in the decision-making, as previous experiments show that the individual values are heavily dependent on the specific environment. The objective of this chart is to provide operators with insight on the important parameters to take into account when designing satellite constellations. It is envisioned that to provide accurate insight, specific simulations will be required to assess each particular scenario.

The results validate that the hypothesis stated in the research question 8.1 is only partially true. While cost and capacity ratios have an influence on the optimal design, interference proves to be a determining factor when deciding between hybrid and single-altitude constellations. Satellite operators, as well as researchers, must include interference considerations when studying the design of satellite constellations.

8.5 Optimal magnitude of user demand per altitude experiments and discussion

Aside from the optimal proportion of satellites across altitudes, there exists a general belief that LEO systems are better suited towards low demand users (such as residential customers), while

MEO systems are better suited towards higher demand users (such as aerial and maritime). The objective of this section is to assess the validity of this belief by studying the optimal magnitude of user demand per altitude, including a discussion on the impact of the results.

8.5.1 Experiments

To determine the appropriate magnitude of user demand per altitude, two experiments have been conducted, summarized in Table 8.10. The first one focuses on a generic LEO constellation, while the second one focuses on a generic MEO constellation.

Number of LEO satellites	Number of MEO satellites	Frequency reuse	Number of gateway antennas
2,000	0	4	1
0	200	12	3

Table 8.10: Experimental set-up to determine the optimal magnitude of user demand across altitudes

For each experiment, up to 5 different scenarios have been considered, with different number of user locations and user per location, but with the same number of users, summarized in Table 8.11. The objective of these different scenarios is to determine if as hypothesized, LEO constellations are better suited to serve less demanding users, and if MEO constellations are better suited to serve more demanding users.

Figure 8-9 shows the result of this experiment. Regarding the LEO constellation, it is shown that, to obtain the maximum throughput, the optimal magnitude of user demand is around 1 Gbps per user. The reason for this is that a magnitude of around 1 Gbps can be fulfilled with a single satellite-user link, achieving a very high link utilization. Higher magnitudes need to be served by multiple links, increasing interference and achieving a lower overall link utilization, while lower magnitudes do not make use of all available resources. Note that this is significantly higher that current offers, which revolve around 100 Mbps. The reason for this difference lies in the economic gain of these systems: serving 10 users with 100 Mbps each is more economically beneficial than serving 1 user with 1 Gbps, even if the total throughput of the constellation is lower. In alignment with these results, it is important to highlight that 1 Gbps is the current standard for household connectivity in the United States [242].

N_{loc}	$N_{us/loc}$	Total demand per location [Gbps]
20,000	5	0.5
10,000	10	1
5,000	20	2
2,000	50	5
1,000	100	10

Table 8.11: Scenarios considered to determine the optimal magnitude of user demand across altitudes

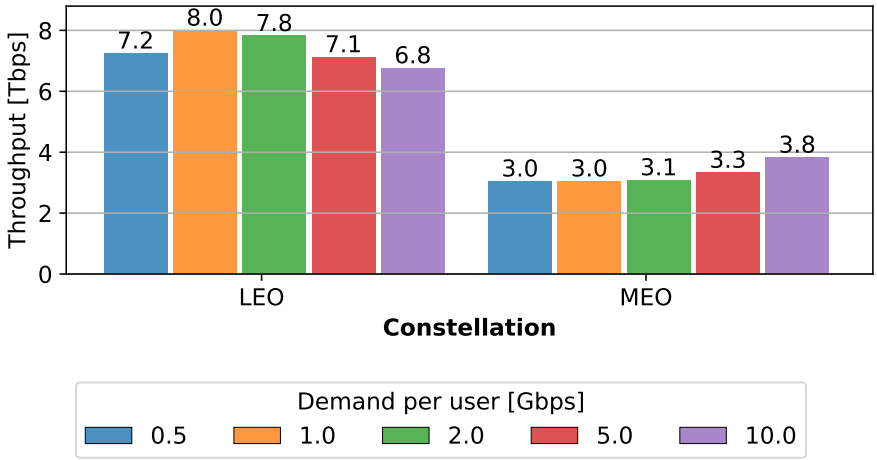


Figure 8-9: Performance on generic LEO and MEO constellations depending on user magnitude of user demand

Regarding the MEO system, the results show that higher magnitudes allow for higher throughput, achieving the optimal at 10 Gbps. Note that higher values are not considered as it goes beyond the link capacity. In conclusion, this experiment provided tangible proof that LEO systems are better suited to serve lower demanding customers, while MEO systems are better suited to serve higher demanding customers, following general belief.

8.6 Chapter summary and conclusions

This chapter has studied the design and performance of hybrid megaconstellations through surrogate and realistic methodologies. Different hybrid megaconstellation designs have been analyzed to establish the optimal proportion of satellites across LEO and MEO altitudes, as well as the optimal

user demand magnitude that maximized throughput for each altitude.

8.6.1 Chapter summary

This chapter commences with Section 8.1, which introduces the design exploration framework employed throughout the chapter. This framework comprises a surrogate model to swiftly filter out dominated designs and a realistic approach to provide accurate performance estimations. Subsequently, Section 8.2 describes the surrogate model estimation, which is grounded in a graph-based approach that expands upon existing literature by accounting for interference between satellites and additional user-satellite flows, thereby making it applicable to hybrid systems. Proceeding, Section 8.3 delineates a generic hybrid constellation utilized in subsequent experiments, along with the concepts of cost and capacity ratios, which define the cost and capacity of a MEO satellite relative to a LEO satellite. Building upon these foundations, Section 8.4 delves into the study of the optimal proportion of satellites across altitudes for a LEO-MEO hybrid system, including various performance estimators, RAP methodologies, user distributions, and constellation sizes. Finally, Section 8.5 investigates the optimal user demand magnitude according to the constellation altitude by examining the total throughput while varying the number of users and the demand magnitude of each user.

8.6.2 Response to Research Questions

Research question 8.1

Given specific cost and capacity ratios between LEO and MEO satellites, what is the optimal proportion between LEO and MEO satellites to maximize throughput?

Alongside the cost and capacity ratios, factors such as the resource allocation methodology, user distribution, and constellation sizing significantly influence the optimal proportion between LEO and MEO satellites, resulting in a non-unique solution. Note that some of these factors, such as the resource allocation methodology, are exclusively determined by decisions taken by the satellite operator, while others also depend on external factors, such as the user distribution. Nevertheless, the findings of this chapter offer valuable insights into general trends. Specifically, hybrid systems outperform single-altitude solutions when the cost ratio is low and interference is minimal. This may occur due to dispersed user distribution across the globe, utilization of optimization methods

leveraging system flexibility to mitigate interference, or deployment of a large satellite network. In such scenarios, the prevalent design often entails an equal budget split between LEO and MEO, although alternative hybrid configurations may dominate under specific circumstances. Conversely, when the cost ratio or interference is high, single-altitude systems outperform hybrid designs. Particularly, under low capacity ratios, full LEO designs achieve maximum throughput, whereas under high capacity ratios, full MEO designs deliver optimal performance.

Research question 8.2

What order of demand is more appropriate for LEO constellations? And for MEO constellations?

The results of this chapter indicate that LEO constellations exhibit higher throughput when serving users with approximately 1 Gbps demand, as this value maximizes single-link utilization. Conversely, MEO constellations demonstrate higher throughput when catering to a smaller subset of highly demanding users (10 Gbps), as interference, which disproportionately impacts MEO satellites compared to LEO, decreases when users are fewer and more widely distributed.

8.6.3 Specific chapter contributions

The specific contributions of this chapter are as follows:

- Extended an existing graph-based methodology for performance estimation in large constellations by including interference considerations and additional user-satellite communications, making it suitable for hybrid constellations.
- Developed an exploration framework that leverages surrogate and realistic methodologies to search through the design space, filtering dominated designs while giving accurate estimations on interesting architectures.
- Studied the optimal proportion across altitudes in a LEO-MEO hybrid satellite constellation under a variety of conditions including different RAP methodologies, user distributions, and constellation sizing.
- Studied the optimal user demand magnitude for LEO and MEO constellations by simulating different configurations regarding number of users and demand per user.

Chapter 9

Conclusions

Satellite communications are becoming a key technology to maintain connectivity in an increasingly information-driven world. To deliver broadband connectivity to a diverse range of users, the next generation of satellite communications relies on highly flexible digital payloads, such as phased array antennas, on-board processing, and adaptive modulation and coding schemes, which allow for a higher degree of control over the limited spacecraft resources. While existing literature explores the novel flexibilities at the spacecraft level, there remains a notable gap in understanding decision-making at the constellation level, which is crucial for operating the next generation of systems.

In this context, the primary objective of this dissertation was to develop optimized decision-making frameworks capable of addressing the flexibilities inherent in modern systems, including hybrid architectures, inclined orbits, and inter-satellite links. Specifically, the resource allocation problem—entailing the efficient distribution of constellation resources—was dissected into various sub-problems, for which novel methodologies were devised to accommodate the flexibilities of modern systems. Moreover, given the lack of insights into the design and performance of megaconstellations, the secondary objective of this dissertation was to employ the proposed frameworks to study the design of hybrid megaconstellations, encompassing both existing and future designs.

This chapter concludes the dissertation by synthesizing the key findings and contributions of this work. Additionally, future avenues for research are outlined, paving the way for further advancements in the field.

9.1 Summary

Chapter 1 started this dissertation by outlining the motivation behind studying satellite constellations, highlighting the recent surge in satellite operators filing for communications constellations comprising thousands of satellites. This surge is facilitated by recent technological advancements enabling increased flexibility onboard spacecraft, thereby enhancing capacity at the expense of complexity. Additionally, the chapter provided an overview of fundamental physical concepts underpinning antenna-based communications and delved into the resource allocation problem in satellite constellations. Subsequently, Chapter 2 conducted a comprehensive literature review, focusing on autonomous resource allocation frameworks for communication satellites. This review identified three main gaps: 1) the absence of methods for hybrid constellations, where satellites may operate at different altitudes, 2) the lack of frameworks for large constellations, where users and gateways might have multiple visible satellites, leading to potential interference, and 3) the lack of insights into the design of hybrid systems. To address these gaps, Chapter 3 laid the groundwork by providing the necessary models to simulate satellite communications operations. These models included the user model, atmospheric model, interference model, satellite payload model, and gateway model, alongside detailing their limitations and assumptions.

Starting with the key contributions of this dissertation, Chapter 4 presented a novel approach to tackle the joint Beam Shaping and User Grouping problem tailored for hybrid constellations. Initially, it formulated the problem as a mixed-integer linear programming (MILP) model, aiming to balance the load across altitudes while employing altitude-appropriate beam shapes. Given the intricacy of this problem, the chapter proposed an iterative approach capable of generating solutions for high-dimensional scenarios. Validation and verification, convergence, tradespace, and performance analyses were conducted to assess the proposed method efficacy, alongside discussing its applicability under various conditions.

Chapter 5 proceeded to address the Satellite Routing and Frequency Assignment problems in a coordinated manner. Initially, it formulated the Satellite Routing problem to distribute the workload across satellites while minimizing interference, proposing a two-step approach based on MILP. Building on this solution, the chapter then tackled the Frequency Assignment problem, offering a ranking-based MILP solution. Verification and validation, convergence, performance, and operational studies were conducted to evaluate the applicability of the proposed coordinated

framework.

Subsequently, Chapter 6 delved into the Gateway Routing problem, formulating it as a MILP model focused on three objectives: maximizing throughput, enhancing quality of service, and minimizing necessary ground infrastructure. Verification, validation, and performance studies were conducted, comparing the proposed method against existing literature. Utilizing this methodology, the chapter investigated the design of ground infrastructure, as well as intra- and cross-altitude inter-satellite links.

Chapter 7 integrated the methodologies from previous chapters, proposing a comprehensive resource allocation framework adept at addressing all modern constellation flexibilities. This framework was compared against existing techniques, representing the current industry standard. Additionally, the chapter employed the proposed method to evaluate and compare existing mega-constellation designs. Finally, Chapter 8 delved into two aspects of hybrid constellation design: 1) determining the optimal proportion of satellites across altitudes, and 2) identifying the optimal user demand magnitude for each altitude. To this end, a rapid graph-based evaluation was proposed, followed by the application of the developed resource allocation framework.

9.2 Contributions

This dissertation has made several contributions regarding the design and operations of the next generation of satellite constellations, which have been outlined at the end of each respective chapter. These contributions, summarized below, can be classified into methodological and domain-specific contributions.

9.2.1 Methodological contributions

- Novel methodology for joint Beam Shaping and User Grouping: Developed a novel formulation to tackle the joint Beam Shaping and User Grouping problem, along with an iterative method to address it effectively in high-dimensional scenarios. This formulation takes into account the flexibility and constraints of hybrid designs by accommodating footprints of varying sizes. Due to its NP-Hard nature, the iterative approach involves breaking down the problem into smaller instances and solving them iteratively, prioritizing practicality over optimality. The method has demonstrated significant improvements over current techniques by achieving

better load balancing across different altitudes and reducing the required number of beams. Furthermore, its effectiveness has been validated in realistic scenarios using the Boeing and SpaceX constellations, including simulations involving up to 100,000 users.

- Novel coordinated Satellite Routing and Frequency Assignment framework Developed a coordinated Satellite Routing and Frequency Assignment framework, leveraging two novel methods developed for each problem. In particular, first the Satellite Routing is addressed, defining a two-step approach to tackle it effectively. The formulation for this problem focuses on load balancing across satellites and interference minimization as its primary objectives. Recognizing the NP-Hardness of the original problem, the formulation is decomposed into two smaller sub-problems, each addressing a specific objective. Second, the Frequency Assignment is addressed, developing a novel method for high-dimensional settings. This approach builds upon an existing formulation, extending it to incorporate a ranking-based system to filter solutions. This coordinated framework effectively distributes the task of addressing interference across the two methods, thereby enhancing capacity. Results demonstrate that the proposed approach achieves significant capacity gains for systems with 200,000 users, reaching up to 75% increased capacity in existing constellation designs such as SpaceX. Moreover, an operational feasibility study suggests that the framework holds promise for real-world applications.

- Novel methodology for Gateway Routing: Developed a novel multi-objective formulation for Gateway Routing, aiming to maximize total throughput, quality of service, and minimize necessary ground infrastructure while considering interference. The proposed approach achieves solutions with a more compelling trade-off between quality of service and required ground infrastructure at equivalent throughput levels compared to existing methods.

- Unique resource allocation problem framework: Developed a complete resource allocation problem (RAP) framework tailored to address the flexibilities of next-generation satellite communications. This framework integrates the methods proposed throughout this dissertation to enable autonomous decision-making for operating satellite constellations. Compared to current practices, the framework doubles the total constellation throughput and provides a more favorable trade-off between quality of service and required ground infrastructure. Additionally, it offers an operational range that can offer valuable insights to satellite operators.

- Novel methodology for the design of megaconstellations: Developed a novel framework for

studying the design of megaconstellations, combining holistic and realistic approaches. The holistic approach relies on a graph-based formulation to swiftly analyze multiple constellation designs, while the realistic technique, based on resolving the RAP, offers specific insights into constellation performance.

9.2.2 Domain-specific contributions

- Study on the design of ancillary infrastructure: Studied the design of ground infrastructure and inter-satellite links (ISLs) in satellite communications, encompassing both intra- and cross-altitude connections. In terms of ground infrastructure, the findings reveal that, in terms of throughput, expanding the number of ground stations proves advantageous only under supply-limited conditions, albeit with diminishing returns. Conversely, in demand-limited scenarios, augmenting the number of stations can enhance the quality of service, as information can utilize more efficient pathways to reach its destination. Regarding intra-altitude ISLs, increasing the number and range of connections can amplify throughput by alleviating demand in densely populated regions. However, expanding the ISL network may compromise quality of service. Conversely, in demand-limited scenarios, extending the ISL network can enhance service quality by facilitating faster data transmission paths. Cross-altitude ISLs are essential for maximizing capacity, particularly as densely interconnected LEO and MEO layers can triple throughput. Nonetheless, this enhancement is accompanied by reduced service quality due to the additional delay introduced by MEO-LEO links.

- Study on the performance of existing megaconstellations: Studied the performance of existing megaconstellation designs in terms of throughput, latency, and necessary ground infrastructure. Results reveal that the throughput and number of gateways are directly dependent on the number of satellites and quality of the link, while the latency is mostly dependent on the altitude. The link quality is mostly affected by the available spectrum and satellite altitude. Furthermore, as opposed to prior studies, the results show that under realistic operational conditions with 500,000 users, current systems are demand-limited.

- Study on the design of future megaconstellations: Studied the optimal proportion of satellites across altitudes in the design of hybrid megaconstellations. Leveraging the proposed holistic-realistic framework, experiments reveal three primary factors influencing this proportion:

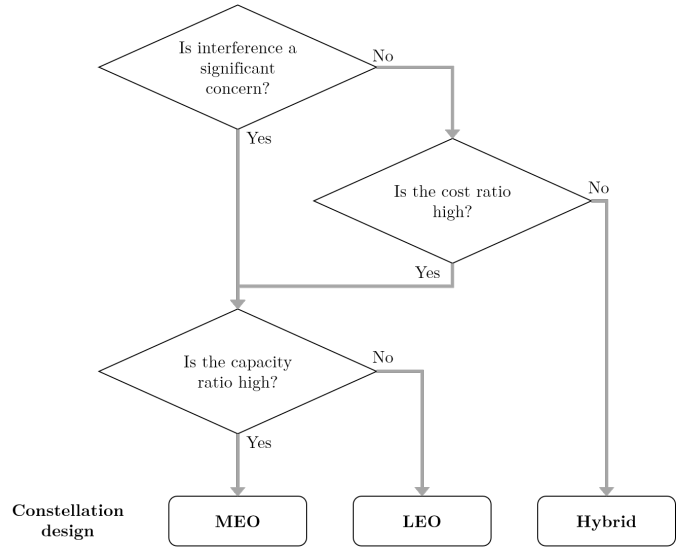


Figure 9-1: Recommendations on the design of satellite communications constellations

interference level, and cost and capacity ratios between MEO and LEO spacecraft. Specifically, under conditions of low interference and cost ratios, hybrid systems, particularly those with an equal budget split between LEO and MEO, outperform alternative solutions. However, in scenarios where the cost ratio or interference are high, LEO constellations dominate under low capacity ratio scenarios, while MEO constellations dominate under high capacity ratio scenarios. For clarity, these insights are summarized in Figure 9-1. Further analysis gives insight on the optimal user demand magnitude across each altitude.

9.3 Future work

The possible avenues for future work are:

- Using a more realistic user distribution: The user distribution utilized in this study adopts a population-like distribution of fixed users, as discussed in Section 3.2. However, this model has inherent limitations, and it might not be representative of the real distribution for two reasons: 1) Major population centers often have access to more cost-effective alternatives to broadband satellite communications, and 2) Segments such as maritime or aviation users are vital for the financial viability of these systems but are not adequately represented. Future

research could explore employing the methodologies proposed in this work under more realistic user distributions. The application and extension of the proposed method to non-fixed users is also left as a possible avenue of future research.

- Enhancing robustness of proposed methodologies: While the proposed methodologies aim to address realistic operational conditions, aspects of robustness remain unaddressed. In the context of the Satellite Routing problem, there is a need to extend the methodology to ensure robustness against satellite failures, thereby ensuring continuous coverage for users. Regarding the Frequency Assignment problem, existing works propose contingency strategies for real-time spectrum management [237], yet their performance under realistic operational conditions requires further investigation. Furthermore, for Gateway Routing, extending the methodology to ensure robustness in satellite-gateway links against gateway failures and weather impairments is essential. Exploring and addressing these aspects would contribute significantly to the resilience and reliability of satellite communication systems.

- Integration of real-time capabilities into the framework: The developed RAP framework assumes constant user demand, simplifying real-time adjustments in spectrum (through Beam Hopping or MF-TDMA approaches) and power (through Power Allocation techniques). Future research could explore integrating real-time management solutions into the framework to adapt to rapid changes in user demand. Combining these capabilities may further enhance the capacity of the constellation.

- Coexistence of multiple constellations: This work operates under the assumption that systems exist in a vacuum, with no overlap on users or spectrum. In reality, satellite systems coexist and compete for similar user bases and spectrum. Future research avenues include developing a realistic model of the satellite competitive environment, considering financial aspects and user distribution. Progress in this area has been made [243]. Additionally, protocols and methodologies need to be developed to address frequency sharing among multiple constellations. Cross-constellation interference mitigation protocols will be crucial to prevent outages and ensure seamless operation.

- Exploration of alternative satellite payloads: The analysis in this study focuses on a specific satellite architecture, including on-board processing, phased-array antennas, adaptive modulation and coding, and ISLs. While representative of current designs, not all systems operate

with the same payload. Future research could investigate and broaden the applicability of the proposed framework under different payload architectures. Understanding how variations in satellite payloads impact system performance and optimization strategies would be valuable for advancing satellite communication technology.

- Extension of problem methodologies: While this dissertation is the first study to present a comprehensive resource allocation framework for modern satellite constellations, it often sacrifices optimality and for practicality to obtain a satisfactory solution. Future research could focus on enhancing optimization methodologies to achieve even greater performance in communication constellations.

- Optimal constellation design: This study initiates the exploration of optimal hybrid mega-constellation design by determining the optimal distribution of satellites across altitudes. However, considerations such as orbital configuration and satellite payload design remain unexplored. Existing works have addressed these aspects for single-altitude constellations, but their applicability to hybrid systems is uncertain. Therefore, future research could delve into determining the optimal orbital configuration and satellite payload design for hybrid megaconstellations to maximize performance and efficiency.

- Extension of the proposed methods to other fields: While the focus of this work is on broadband satellite communications, the developed methodologies have potential applications in other fields. For example, the joint Beam Shaping and User Grouping methodology could be adapted for terrestrial Multiple Input, Multiple Output communication systems. Similarly, the Satellite Routing methodology could find applications in imaging satellites, where services are intermittent and non-stationary. Exploring these avenues could lead to innovative applications and advancements in various domains beyond satellite communications.

Appendices

Appendix A

Geometry of hybrid constellations

This Chapter defines the geometric parameters that are used repeatedly for the formulation of the different problems addressed in this Thesis.

Any constellation is defined by a set of satellites $\mathcal{S}$, which contains orbital information about each satellite, summarized using a reference position $p_s = \{lon_s, lat_s, a_s\}$ and velocity v_s. The mean altitude of each satellite is assumed to be fixed[1]. It is assumed that the satellite constellation can be hybrid: satellites are divided into different orbital planes, which may be at different altitudes. However, it is assumed that the orbital plane of each satellite is fixed (i.e., the orbital configuration of each satellite is static, except for the mean anomaly). Furthermore, I define $\mathcal{S}_a$ as the set of all satellites at altitude a, and a_s as the altitude of satellite s.

Next, I define the set $\mathcal{P}$ as the set of all possible longitudes and latitudes:

$$\mathcal{P} = \left\{ (lon, lat) \in \mathbb{R}^2 \mid lon \in [0, 2\pi], lat \in \left[\frac{-\pi}{2}, \frac{\pi}{2} \right] \right\} \tag{A.1}$$

Now, a satellite is *visible* from a point on the surface of the Earth if the elevation angle of the satellite is higher than the minimum elevation angle (MEA) specified by the operator (see Figure A-1). This is what is known as a satellite being in *line of sight* (LoS) of a terminal, or a terminal being in *field of view* (FoV) of a satellite. I define $\mathcal{P}_{p,a}$ as the set of positions at altitude a that are in LoS of position p located in the Earth's surface:

[1] For orbits with eccentricity 0, this means that the satellites are always at the same altitude

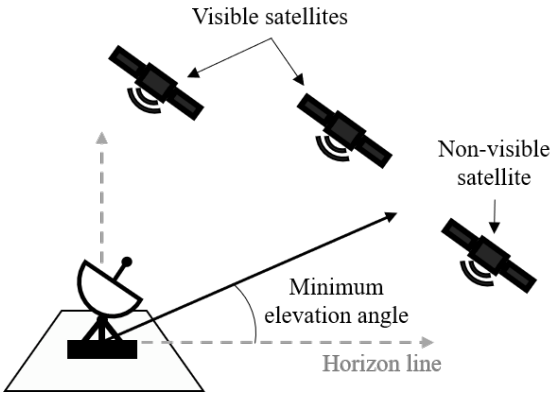


Figure A-1: Visible satellites of a ground terminal

$$\begin{aligned}\mathcal{P}_{p,a} = \{(lon, lat) \in \mathcal{P} |\ & p_c = Cartesian(lon_p, lat_p, R_E), \\ & p_a = Cartesian(lon, lat, a), \\ & \frac{(p_a - p_c) \cdot p_c}{||p_a - p_c||\ ||p_c||} \leq \cos\left(\frac{\pi}{2} - mea\right)\}\end{aligned} \tag{A.2}$$

With this definition, determining if a point p_1 in altitude a is in LoS of point p_2 on the surface of the Earth can be checked by:

$$LoS(p_1, p_2, a) = (lon_{p_1}, lat_{p_1}) \in \mathcal{P}_{p_2,a} \tag{A.3}$$

Note that $LoS(p_s, p, a_s)$ represents if satellite s is in LoS of point p and changes over time as the position of the satellite p_s is time-dependent. Next, I define v_p and $v_{p,a}$ as the minimum number of satellites that are in LoS of point p on the Earth's surface over all times $\mathcal{T}$ using $\mathcal{S}$ or $\mathcal{S}_a$, respectively:

$$\begin{aligned} v_p &= \min_{t \in \mathcal{T}} \sum_{s \in \mathcal{S}} \mathbb{1}_{LoS(p_s, p, a_s)} \\ v_{p,a} &= \min_{t \in \mathcal{T}} \sum_{s \in \mathcal{S}_a} \mathbb{1}_{LoS(p_s, p, a)} \end{aligned} \tag{A.4}$$

Now, $\mathcal{P}_a^E$ is defined as the set of all possible positions over the Earth's surface that see at least one satellite at altitude a at all times:

$$\mathcal{P}_a^E = \{p \in \mathcal{P} \mid v_{p,a} \geq 1\} \tag{A.5}$$

A collection of useful definitions is given in Table A.1.

Variable	Definition
$\mathcal{P}$	Space of all possible longitudes and latitudes
$\mathcal{P}_{p,a}$	Set of all longitudes and latitudes at altitude a that are in line of sight of a point p on the surface of the Earth
$LoS(p_1, p_2, a)$	Function that returns true if the point p_1 at altitude a is in line of sight of point p_2 on the surface of the Earth, false otherwise
v_p	Minimum number of satellites visible at point p at all times given a full constellation
$v_{p,a}$	Minimum number of satellites visible at point p at all times given satellites at altitude a
$\mathcal{P}_a^E$	Set of longitudes and latitudes over the Earth surface with at least one visible satellite at all times at altitude a

Table A.1: Geometric variable definitions in hybrid constellations

Appendix B

Ground Station locations

Ground station name	Latitude [°]	Longitude [°]	Country	Administrator	Region
Alaska Satellite Facility	64.86	-147.85	USA	NEN (NASA)	N. America
Florida Ground Station	29	-81	USA	NEN (NASA)	N. America
McMurdo Ground Station	-77.81	166.69	Antartica	NEN (NASA)	Oceania
Svalbard Ground Station	78.22	15.39	Norway	NEN (NASA)	Europe
Wallops Flight Facility Ground Stations	37.94	-75.49	USA	NEN (NASA)	N. America
Alaska	59.65	-151.54	USA	NEN (NASA)	N. America
Guam	13.42	144.75	USA	NEN (NASA)	N. America
Clewiston	26.73	-82.03	USA	SSC	N. America
Esrange	67.88	21.07	Sweden	SSC	Europe
Fucino	42	13.55	Italy	SSC	Europe
Hartebeesthoek	-25.64	28.08	South Africa	SSC	Africa
Inuvik	68.4	-133.5	Canada	SSC	N. America
O'Higgins	-63.32	-57.9	Antartica	SSC	S. America
Punta Arenas	-53	-71	Argentina	SSC	S. America
Santiago Satellite Station	-33.13	-70.67	Chile	SSC	S. America
USN Western Australia	-29.05	114.9	Australia	SSC	Oceania
Weilheim	47.84	11.14	Germany	SSC	Europe
Hawaii	19.82	-155.47	USA	KSAT	N. America
Tokyo	35.69	139.69	Japan	KSAT	Asia
Singapore	1.35	103.82	Singapore	KSAT	Asia
Trollsat	-72.1	2.32	Antartica	KSAT	Africa
Vardo	70.37	31.1	Norway	KSAT	Europe

Ground station name	Latitude [°]	Longitude [°]	Country	Administrator	Region
Tromso	69.65	18.96	Norway	KSAT	Europe
Grimstad	58.34	8.59	Norway	KSAT	Europe
Puertollano	38.69	-4.11	Spain	KSAT	Europe
Mauritius	-20.35	57.55	Mauritius	KSAT	Africa
Panama	8.54	-80.78	Panama	KSAT	S. America
Central Africa	4.84	10.1	Central Africa	KSAT	Africa
New Zeland	-46.02	167.81	New Zeland	KSAT	Oceania
Kourou	5.16	-52.65	French Guiana	ESA	Europe
Redu	50	5.16	Belgium	ESA	Europe
Cebreros	40.46	-4.46	Spain	ESA	Europe
Villafranca	40.26	-3.57	Spain	ESA	Europe
Maspalomas	27.45	-15.38	Spain	ESA	Europe
Santa Maria	36.59	-25.08	Portugal	ESA	Europe
Malargue	-25.78	-69.4	Argentina	ESA	S. America
Perth	-31.80	115.89	Australia	ESA	Oceania
Delhi	28.55	77.29	India	Viasat	Asia
Chennai	13.13	80.17	India	Viasat	Asia
Sapporo	43.06	141.34	Japan	JAXA	Asia
Accra	5.56	-0.20	Ghana	SES	Africa
Rio de Janeiro	-22.98	-43.35	Brazil	SES	S. America
Addis Ababa	9.01	38.76	Ethiopia	SES	Africa
Adelaide	-34.93	138.6	Australia	SES	Oceania
Lurin	-12.25	-76.88	Peru	SES	S. America
Hortolandia	-22.85	-47.21	Brazil	SES	S. America
Abu Dhabi	24.45	54.38	UAE	SES	Asia
Kowoloon	22.32	114.18	Hong Kong	SES	Asia
Brewster	48.09	-119.78	USA	SES	N. America
Djibouti	11.83	42.59	Djibouti	SES	Africa
Los Angeles	34.05	-118.24	US	SES	N. America
Vernon	34.15	-99.27	USA	SES	N. America
Karachi	24.86	67.1	Pakistan	SES	Asia
Kiev	50.45	30.52	Ukraine	SES	Europe
Dubbo	-32.23	148.63	Australia	SES	Oceania
Thermopylae	38.8032	22.5577	Greece	SES	Europe
Manassas	38.7958	-77.5740	USA	SES	N. America
Seychelles	-4.63	55.45	Seychelles	Laban	Africa
Denver	39.74	-104.99	USA	Intelsat	N. America
Kumsan	35.36	128.41	South Korea	Intelsat	Asia
Napa	38.25	-122.28	USA	Intelsat	N. America
St. John's	47.56	-52.71	Canada	Telesat	N. America
Iqaluit	63.75	-68.52	Canada	Telesat	N. America
Saskatoon	52.13	-106.67	Canada	Telesat	N. America
Honolulu	21.3069	-157.8583	US	Telesat	N. America
Vancouver	49.2827	-123.1207	Canada	Telesat	N. America

Ground station name	Latitude [°]	Longitude [°]	Country	Administrator	Region
Yellowknife	62.454	-114.3718	Canada	Telesat	N. America
Ottawa	45.4215	-75.6972	Canada	Telesat	N. America
Jakarta	-6.1751	106.865	Indonesia	Telesat	Asia
Toronto	43.6532	-79.3832	Canada	Telesat	N. America
Nuuk	64.1825	-51.7354	Denmark	OneWeb	Europe
King Fahd Satellite Telecommunications City	21.3973	39.3831	Arabia	Arab Satellite Communications Organization	Asia
Quito	-0.1807	-78.4678	Ecuador	None	S. America
Kano	12.0022	8.592	Nigeria	None	Africa
Windhoek	-22.5609	17.0658	Namibia	None	Africa
Dakar Senegal	14.8468	-17.2028	Senegal	None	Africa
Cape Verde	14.55	-23.31	Cape Verde	None	Africa
Dublin	53.350140	-6.266155	Ireland	Amazon	Europe
CapeTown	-33.918861	18.423300	South Africa	Amazon	Africa
Gazipur	23.999339	90.389126	Bangladesh	Thales	Asia
Betbunia	22.54757	91.995896	Bangladesh	Thales	Asia
Lachhiwala	30.178015	78.104023	India	Overseas Communication Service	Asia
Maharastra	19.151375	73.957225	India	Overseas Communication Service	Asia
Foggia	41.5082	15.5865	Italy	SpaceX	Europe
Bellingham	48.774	-122.4486	USA	SpaceX	N. America
Isle of Man	54.251186	-4.463196	UK	SpaceX	Europe
Otaru	43.1732	141.2584	Japan	SpaceX	Asia
Anchorage	61.1859	-149.8769	US	SpaceX	N. America
Hawthorne	33.9201	-118.3322	US	SpaceX	N. America
Llano Grande	19.2589	-99.5812	US	SpaceX	N. America
Fort Lauderdale	26.1909	-80.1931	US	SpaceX	N. America
Norcross	33.9550	-84.1980	US	SpaceX	N. America
Noviciado	-33.3927	-70.8833	Chile	SpaceX	S. America
Willemstad	12.0977	-68.9081	Curaçao	SpaceX	S. America
Itaborai	-22.6967	-42.8728	Brasil	SpaceX	S. America
Alfouvar de Cima	38.8685	-9.2822	Portugal	SpaceX	Europe
Chalfont Grove	51.6155	-0.5758	UK	SpaceX	Europe
Lekki	6.4495	3.5877	Nigeria	SpaceX	Africa
Frankfurt	50.3298	8.4708	Germany	SpaceX	Europe
Tea Gardens	-32.5932	152.1042	Australia	SpaceX	Oceania
Muallim	40.7888	29.5094	Turkey	SpaceX	Europe
Robertsdale	30.5670	-87.6460	Panama	SpaceX	S. America
Toa Baja	18.4311	-66.1921	Puerto Rico	SpaceX	N. America
Cleavdon	-36.9897	175.0554	New Zealand	SpaceX	Oceania
Litchfield	41.5450	-73.3540	US	SpaceX	N. America
Angeles	15.1709	120.5057	Philippines	SpaceX	Asia

Ground station name	Latitude [°]	Longitude [°]	Country	Administrator	Region
Hitachinaka	36.3867	140.6137	Japan	SpaceX	Asia
Robbins	38.8750	-121.7071	US	SpaceX	N. America
Kenansville	27.8760	-81.0305	US	SpaceX	N. America
Falda del Carmen	-31.5225	-64.4610	Argentina	SpaceX	S. America
Anakie	-37.9532	144.3282	Australia	SpaceX	Oceania
Cataby	-30.8483	115.6193	Australia	SpaceX	Oceania
Ubon Ratchathani	15.2342	104.8569	Thailand	SpaceX	Asia
Kendari	-3.9774	122.5113	Indonesia	SpaceX	Oceania
Cabon San Lucas	22.9128	-109.9258	US	SpaceX	N. America
Tapachula	14.7862	-92.3671	Mexico	SpaceX	S. America
Kaunas	54.8795	23.8417	Lituania	SpaceX	Asia
Akita	39.6383	140.0647	Japan	SpaceX	Asia
Mossoro	-5.1570	-37.3537	Brazil	SpaceX	S. America
Ikire	7.3875	4.2124	Nigeria	SpaceX	Africa
Suva	-18.1292	178.4676	Tahiti	SpaceX	Oceania
Tumon	13.5028	144.8052	Guam	SpaceX	Oceania
Altay	47.8519	88.1346	China	CNSA	Asia
Chengdu	30.6570	104.0660	China	CNSA	Asia
Xinjiang	39.4747	75.9912	China	CNSA	Asia
Beijing	39.9167	116.3833	China	CNSA	Asia
Changchun	43.866761	125.310742	China	CNSA	Asia
Kashgar	39.467395	75.988195	China	CNSA	Asia
Lingshui	18.508049	110.034506	China	CNSA	Asia
Menghai	21.842867	100.385890	China	CNSA	Asia
Longyan	25.101523	117.034384	China	CNSA	Asia
Nanning	22.817977	108.331523	China	CNSA	Asia
Qingdao	36.156413	120.407392	China	CNSA	Asia
Xiamen	24.487417	118.091945	China	CNSA	Asia
Weinan	34.500436	109.492460	China	CNSA	Asia
Xiangxi	27.954570	109.595659	China	CNSA	Asia
Zhanyi	25.606822	103.817659	China	CNSA	Asia

Table B.1: Ground stations possible locations. Based on the data from [230] in addition to new sites revealed by the operators. The sites regarding the CNSA administration, as well as the sites with an operator marked as None, have been estimated by the author.

Appendix C

Gateway Routing Results

Number of ground stations	ω_g	ω_{ISL}	Total number of gateways	Average number of used gateways	Average number of hops	Average time on ISL [ms]	Average throughput [Tbps]	Average gateway utilization	Average number of saturated links
40	0	1	982	792.32	0.93	4.66	14.22	79.31	341.20
40	0.2	1	892	733.16	0.93	4.67	14.22	85.74	338.42
40	1	1	861	709.58	0.94	4.72	14.22	88.60	340.50
40	5	1	840	696.12	0.96	4.81	14.22	90.29	340.04
40	1	0	830	688.50	0.98	4.87	14.22	91.28	341.96
60	0	1	1251	943.04	0.79	4.05	15.17	70.81	341.72
60	0.2	1	1022	818.78	0.80	4.08	15.17	81.57	337.92
60	1	1	893	720.20	0.84	4.28	15.17	92.76	337.20
60	5	1	859	694.72	0.90	4.55	15.17	96.09	342.14
60	1	0	847	690.00	0.98	4.90	15.17	96.56	364.66
80	0	1	1438	1039.58	0.73	3.87	15.55	65.65	350.96
80	0.2	1	1081	922.64	0.74	3.87	15.55	74.02	349.48
80	1	1	922	792.30	0.77	4.03	15.55	86.34	350.78
80	5	1	850	737.44	0.83	4.32	15.55	92.60	356.92
80	1	0	835	733.90	0.86	4.49	15.55	92.84	369.48
100	0	1	1525	1072.52	0.71	3.86	15.66	63.94	360.14
100	0.2	1	1122	949.44	0.72	3.85	15.66	72.30	355.38
100	1	1	953	822.16	0.75	3.99	15.66	83.61	355.20
100	5	1	887	776.78	0.80	4.22	15.66	88.38	364.94
100	1	0	859	743.88	0.87	4.65	15.66	92.06	380.66

Table C.1: Trade-offs of different number of ground stations on supply-limited scenarios. Linked to Figure 6-6

Number of ground stations	ω_g	ω_{ISL}	Total number of gateways	Average number of used gateways	Average number of hops	Average time on ISL [ms]	Average throughput [Tbps]	Average gateway utilization	Average number of saturated links
40	0	1	904	693.46	0.41	2.20	7.44	47.44	13.50
40	0.2	1	663	611.70	0.43	2.24	7.44	53.75	13.30
40	1	1	497	424.48	0.56	2.77	7.44	77.43	13.88
40	5	1	389	347.32	0.75	3.63	7.44	94.37	16.92
40	1	0	366	338.40	0.88	4.36	7.44	96.20	27.64
60	0	1	1095	789.52	0.25	1.22	7.39	41.30	7.80
60	0.2	1	778	692.72	0.26	1.26	7.39	47.01	8.12
60	1	1	535	437.20	0.42	1.89	7.39	74.41	7.56
60	5	1	405	351.98	0.61	2.78	7.39	91.94	8.80
60	1	0	394	342.78	0.77	3.73	7.39	93.67	20.56
80	0	1	1242	876.78	0.17	0.86	7.35	36.95	6.06
80	0.2	1	815	733.94	0.19	0.93	7.35	44.11	5.52
80	1	1	553	474.50	0.34	1.52	7.35	68.23	5.54
80	5	1	402	354.34	0.58	2.58	7.35	90.91	6.00
80	1	0	382	350.30	0.67	3.06	7.35	91.87	12.78
100	0	1	1349	894.20	0.15	0.76	7.48	36.83	5.90
100	0.2	1	978	806.04	0.16	0.77	7.48	40.84	5.16
100	1	1	625	549.68	0.26	1.19	7.48	60.24	4.94
100	5	1	435	371.70	0.51	2.31	7.48	88.17	5.50
100	1	0	417	361.56	0.65	3.08	7.48	90.48	13.92

Table C.2: Trade-offs of different number of ground stations on demand-limited scenarios. Linked to Figure 6-7

LEO ISL	Total number of gateways	Average number of used gateways	Average number of hops	Average time on ISL [ms]	Average throughput [Tbps]	Average gateway utilization	Average number of saturated links
No ISL	1064	792.02	0.00	0.00	8.86	48.77	0.00
1 ISL	1067	823.56	0.47	2.97	11.66	61.93	138.18
2 ISL (Opt. A)	1003	796.34	0.52	3.35	13.03	71.76	180.80
2 ISL (Opt. B)	983	764.70	0.65	3.29	13.50	77.36	243.64
2 ISL (Opt. C)	933	730.02	0.51	1.81	12.43	74.73	139.42
4 ISL (Opt. A)	945	739.48	0.81	4.16	15.48	91.92	346.92
4 ISL (Opt. B)	965	753.92	0.82	4.88	15.59	90.70	363.10
4 ISL (Opt. C)	977	770.12	0.87	6.71	15.91	90.44	392.24
6 ISL (Opt. A)	996	774.64	0.93	5.34	16.65	94.36	443.28
6 ISL (Opt. B)	1111	807.52	1.01	7.39	17.60	95.59	517.38
6 ISL (Opt. C)	1120	823.30	1.00	10.38	17.52	93.17	509.70

Table C.3: Figures of merit on different LEO ISL configurations when supply-limited. Linked to Figure 6-10

ISL	w_g	w_{ISL}	Total number of gateways	Average number of used gateways	Average number of hops	Average time on ISL [ms]	Average throughput [Tbps]	Average gateway utilization	Average number of saturated links
No ISL	0	1	1097	788.02	0.00	0.00	6.37	35.36	0.00
No ISL	0.2	1	1062	787.68	0.00	0.00	6.37	35.38	0.00
No ISL	1	1	1062	787.70	0.00	0.00	6.37	35.38	0.00
No ISL	5	1	1062	787.72	0.00	0.00	6.37	35.38	0.00
No ISL	1	0	1062	787.70	0.00	0.00	6.37	35.38	0.00
1 ISL	0	1	1283	868.08	0.30	1.85	7.36	37.17	16.40
1 ISL	0.2	1	1047	851.70	0.30	1.87	7.36	37.86	16.42
1 ISL	1	1	869	748.44	0.35	2.14	7.36	43.11	16.52
1 ISL	5	1	777	606.84	0.57	3.55	7.36	53.17	20.66
1 ISL	1	0	733	587.42	0.68	4.27	7.36	54.96	28.34
2 ISL (Opt. A)	0	1	1312	873.84	0.22	1.37	7.45	37.39	10.96
2 ISL (Opt. A)	0.2	1	980	833.04	0.22	1.40	7.45	39.22	11.06
2 ISL (Opt. A)	1	1	761	666.14	0.28	1.79	7.45	49.05	11.42
2 ISL (Opt. A)	5	1	694	531.94	0.49	3.10	7.45	61.46	12.14
2 ISL (Opt. A)	1	0	637	509.78	0.58	3.65	7.45	64.08	14.96
2 ISL (Opt. B)	0	1	1283	866.22	0.21	1.07	7.49	37.90	9.78
2 ISL (Opt. B)	0.2	1	982	825.76	0.21	1.09	7.49	39.74	9.94
2 ISL (Opt. B)	1	1	727	641.60	0.28	1.36	7.49	51.24	9.88
2 ISL (Opt. B)	5	1	541	437.14	0.58	2.61	7.49	74.89	13.18
2 ISL (Opt. B)	1	0	426	389.10	0.89	4.26	7.49	83.72	39.12
2 ISL (Opt. C)	0	1	1257	861.18	0.28	0.78	7.43	37.83	15.22
2 ISL (Opt. C)	0.2	1	962	820.64	0.29	0.79	7.43	39.70	15.20
2 ISL (Opt. C)	1	1	682	613.44	0.37	1.03	7.43	53.19	15.58
2 ISL (Opt. C)	5	1	626	472.82	0.59	1.74	7.43	69.06	17.04
2 ISL (Opt. C)	1	0	605	460.20	0.65	1.95	7.43	70.89	19.64
4 ISL (Opt. A)	0	1	1277	869.06	0.17	0.86	7.49	38.03	5.58
4 ISL (Opt. A)	0.2	1	840	743.76	0.19	0.92	7.49	44.42	5.96
4 ISL (Opt. A)	1	1	560	486.34	0.32	1.47	7.49	67.90	5.40
4 ISL (Opt. A)	5	1	412	364.50	0.55	2.47	7.49	90.14	6.42
4 ISL (Opt. A)	1	0	386	368.22	0.56	2.62	7.49	89.14	7.60
4 ISL (Opt. B)	0	1	1254	875.28	0.17	1.00	7.49	37.75	6.48
4 ISL (Opt. B)	0.2	1	927	775.04	0.18	1.05	7.49	42.61	6.18
4 ISL (Opt. B)	1	1	627	543.40	0.28	1.55	7.49	61.15	5.34
4 ISL (Opt. B)	5	1	430	378.76	0.51	2.72	7.49	86.89	6.30
4 ISL (Opt. B)	1	0	430	359.34	0.74	4.10	7.49	91.28	19.34
4 ISL (Opt. C)	0	1	1262	875.34	0.17	1.21	7.49	37.72	6.24
4 ISL (Opt. C)	0.2	1	949	790.04	0.18	1.25	7.49	41.78	5.58
4 ISL (Opt. C)	1	1	664	572.78	0.26	1.78	7.49	57.97	5.10
4 ISL (Opt. C)	5	1	453	384.76	0.51	3.40	7.49	85.63	5.36
4 ISL (Opt. C)	1	0	400	359.62	0.79	6.00	7.49	91.15	23.42

Table C.4: Figures of merit on different LEO ISL configurations when demand-limited for up to 4 ISL. Linked to Figures 6-11, 6-12, and 6-14

ISL	w_g	w_{ISL}	Total number of gateways	Average number of used gateways	Average number of hops	Average time on ISL [ms]	Average throughput [Tbps]	Average gateway utilization	Average number of saturated links
6 ISL (Opt. A)	0	1	1276	866.48	0.16	0.90	7.49	38.12	4.54
6 ISL (Opt. A)	0.2	1	805	726.82	0.18	0.96	7.49	45.41	4.28
6 ISL (Opt. A)	1	1	537	443.66	0.33	1.69	7.49	74.41	3.58
6 ISL (Opt. A)	5	1	410	362.16	0.48	2.47	7.49	90.77	3.72
6 ISL (Opt. A)	1	0	392	356.16	0.54	2.87	7.49	92.17	5.94
6 ISL (Opt. B)	0	1	1246	872.40	0.16	1.06	7.49	37.73	5.72
6 ISL (Opt. B)	0.2	1	799	719.66	0.18	1.13	7.49	45.70	4.56
6 ISL (Opt. B)	1	1	512	440.80	0.32	1.95	7.49	74.67	3.40
6 ISL (Opt. B)	5	1	468	382.76	0.41	2.59	7.49	85.61	3.62
6 ISL (Opt. B)	1	0	462	396.02	0.40	2.51	7.49	82.75	4.68
6 ISL (Opt. C)	0	1	1288	872.40	0.16	1.37	7.49	37.65	5.22
6 ISL (Opt. C)	0.2	1	951	788.04	0.16	1.42	7.49	41.60	4.54
6 ISL (Opt. C)	1	1	660	591.16	0.23	1.88	7.49	55.55	4.20
6 ISL (Opt. C)	5	1	398	377.20	0.46	3.82	7.49	86.72	4.50
6 ISL (Opt. C)	1	0	377	363.18	0.59	5.92	7.49	89.24	13.14

Table C.5: Figures of merit on different LEO ISL configurations when demand-limited for 6 ISL. Linked to Figures 6-13 and 6-14

LEO ISL	MEO ISL	Total number of gateways	Average number of used gateways	Average number of hops	Average time on ISL [ms]	Average throughput [Tbps]	Average gateway utilization	Average number of saturated links
No	No	527	342.36	0.00	0.00	1.81	42.21	0.00
2	No	579	431.26	0.42	2.60	2.39	38.00	11.12
4	No	501	395.30	0.38	1.90	2.43	44.46	9.96
No	1	590	406.78	0.40	11.16	2.97	51.75	47.42
2	1	607	500.86	1.03	17.40	4.52	56.93	146.46
4	1	509	444.06	1.07	17.13	4.76	70.36	159.72
No	2	608	441.68	0.55	15.57	3.81	58.13	81.50
2	2	637	536.64	1.21	21.90	5.97	68.23	237.80
4	2	563	483.42	1.25	21.32	6.32	81.12	258.32
No	4	609	479.32	0.72	20.27	5.06	68.23	132.34
2	4	664	581.06	1.39	26.37	7.98	81.87	363.30
4	4	653	504.44	1.47	25.93	8.40	94.57	399.72

Table C.6: Figures of merit on different MEO ISL configurations when supply-limited. Linked to Figure 6-15

MEO ISL	w_g	w_{ISL}	Total number of gateways	Average number of used gateways	Average number of hops	Average time on ISL [ms]	Average throughput [Tbps]	Average gateway utilization	Average number of saturated links
No	0	1	590	316.76	0.05	0.28	1.98	28.16	0.00
No	0.2	1	452	314.72	0.05	0.29	1.98	28.35	0.00
No	1	1	443	314.48	0.05	0.29	1.98	28.37	0.00
No	5	1	357	297.94	0.06	0.33	1.98	29.93	0.00
No	1	0	321	279.58	0.11	0.51	1.98	31.86	0.00
1	0	1	609	366.52	0.30	5.39	1.99	32.29	0.74
1	0.2	1	473	356.72	0.30	5.51	1.99	33.43	0.74
1	1	1	395	303.98	0.35	6.14	1.99	40.83	0.78
1	5	1	286	204.60	0.78	13.42	1.99	56.10	4.50
1	1	0	248	157.02	1.52	23.21	1.99	65.87	9.18
2	0	1	627	379.38	0.35	6.82	1.99	34.14	0.70
2	0.2	1	476	359.34	0.35	6.99	1.99	36.95	0.72
2	1	1	396	301.92	0.40	7.72	1.99	46.53	0.82
2	5	1	298	180.96	0.88	16.38	1.99	69.29	6.84
2	1	0	175	112.28	1.84	29.44	1.99	86.59	13.30
4	0	1	639	378.20	0.36	7.95	1.99	37.72	1.02
4	0.2	1	477	351.28	0.36	7.98	1.99	41.61	1.08
4	1	1	325	301.60	0.39	8.55	1.99	51.01	1.24
4	5	1	234	171.10	0.81	16.66	1.99	80.03	7.16
4	1	0	129	91.98	1.82	31.40	1.99	96.19	13.68

Table C.7: Figures of merit on different MEO ISL configurations when demand-limited. Linked to Figure 6-16

Appendix D

Constellation Design Results

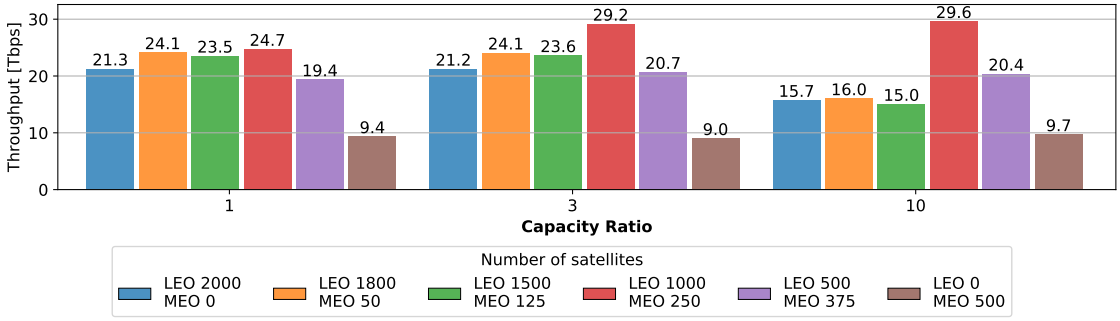


Figure D-1: Throughput of the optimal design when using the surrogate model under a cost ratio of 4

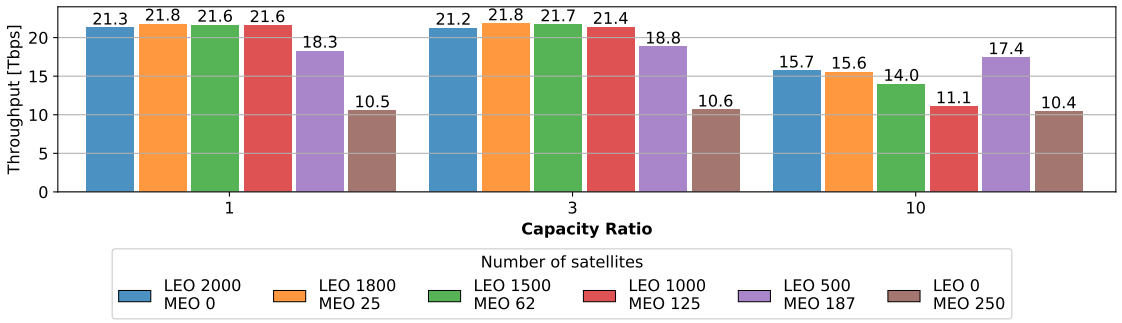


Figure D-2: Throughput of the optimal design when using the surrogate model under a cost ratio of 8

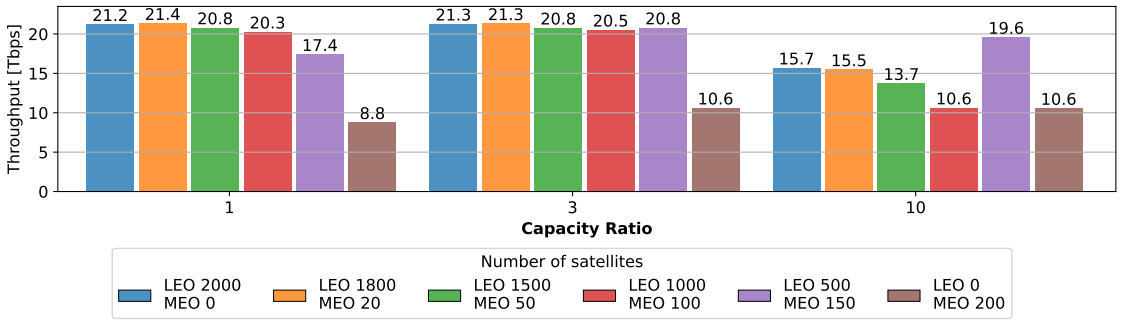


Figure D-3: Throughput of the optimal design when using the surrogate model under a cost ratio of 10

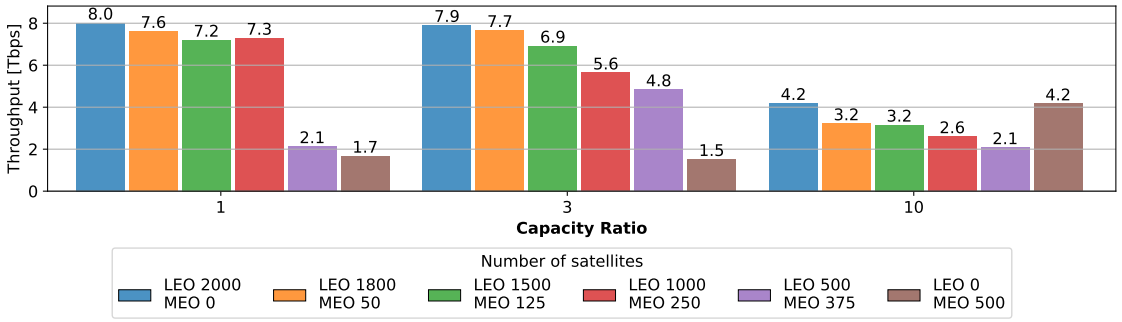
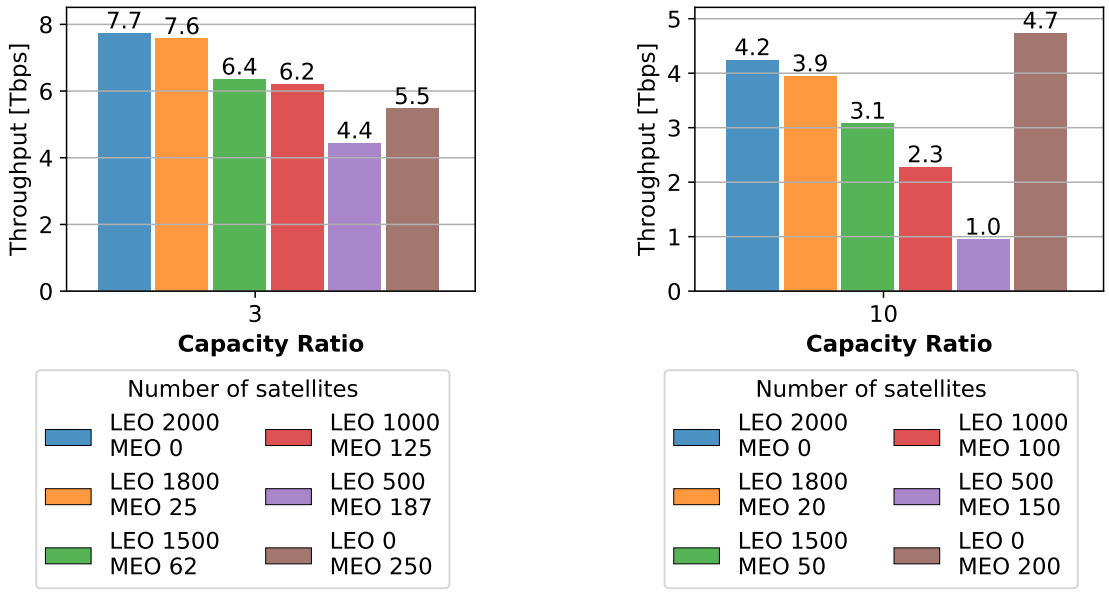


Figure D-4: Throughput of the optimal design when using the RAP method under a cost ratio of 4



(a) Cost ratio of 8

(b) Cost ratio of 10

Figure D-5: Throughput of the optimal design when using the RAP method under a cost ratio of 8 and capacity ratio of 3 (left) and cost ratio of 10 and capacity ratio of 10 (right)

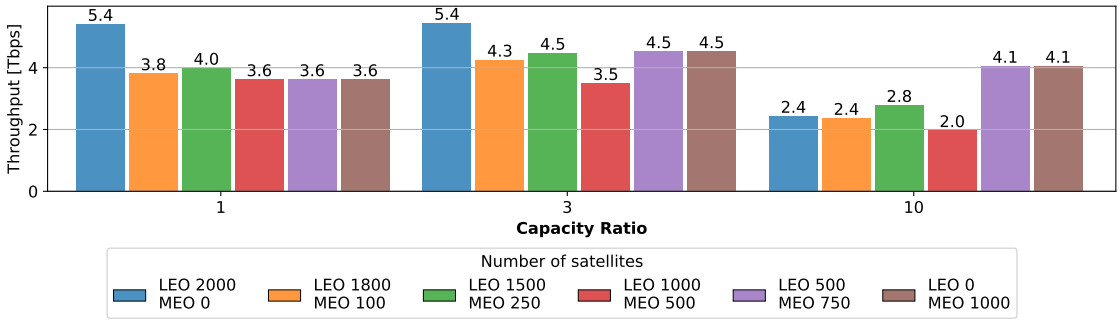


Figure D-6: Throughput of the optimal design when using the heuristic method under a cost ratio of 2

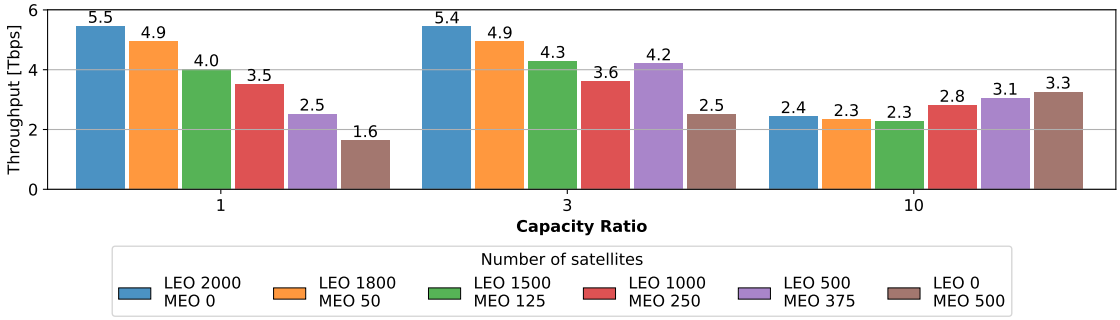


Figure D-7: Throughput of the optimal design when using the heuristic method under a cost ratio of 4

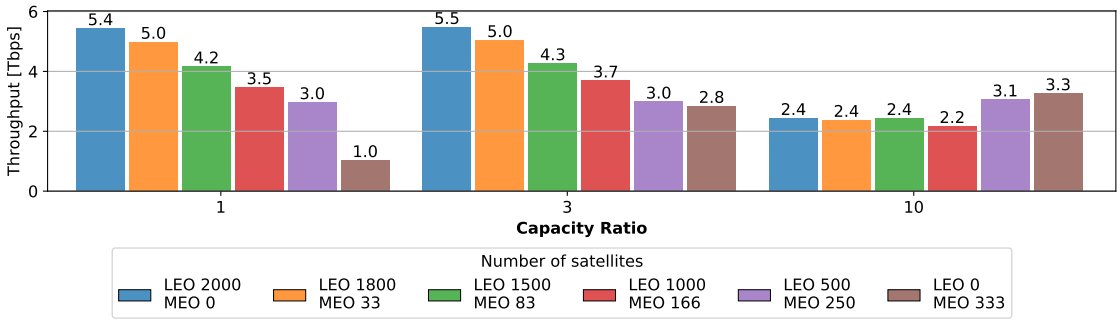
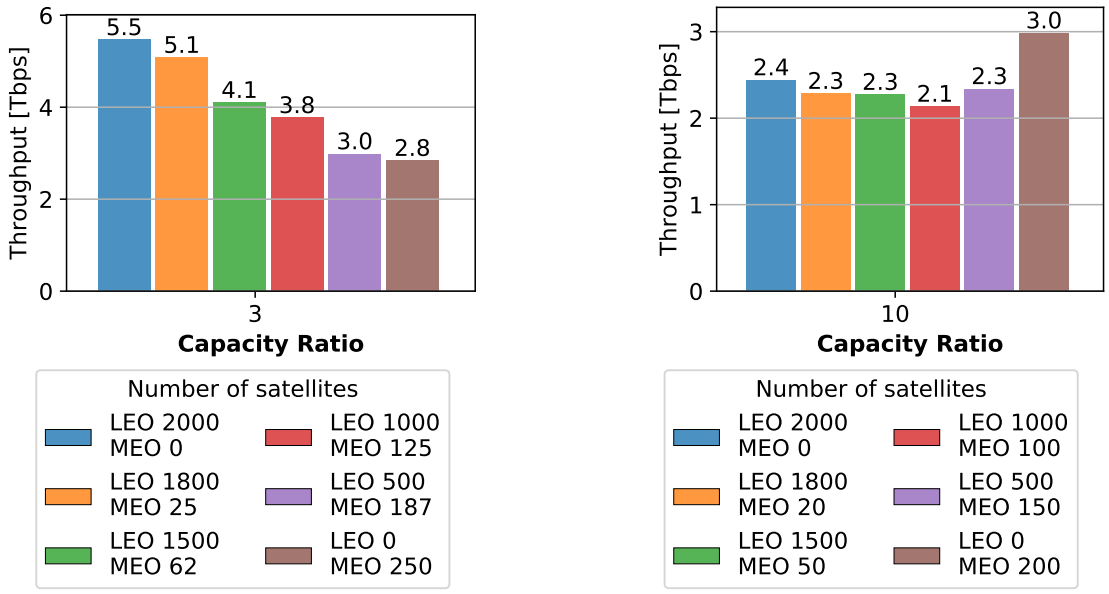


Figure D-8: Throughput of the optimal design when using the heuristic method under a cost ratio of 6



(a) Cost ratio of 8

(b) Cost ratio of 10

Figure D-9: Throughput of the optimal design when using the heuristic method under a cost ratio of 8 and capacity ratio of 3 (left) and cost ratio of 10 and capacity ratio of 10 (right)

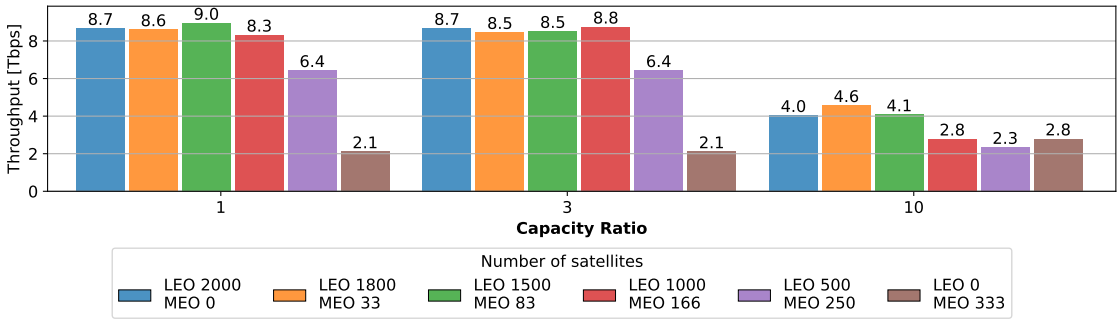


Figure D-10: Throughput of the optimal design when using the RAP method under a cost ratio of 6 using the user distribution proportional to land area

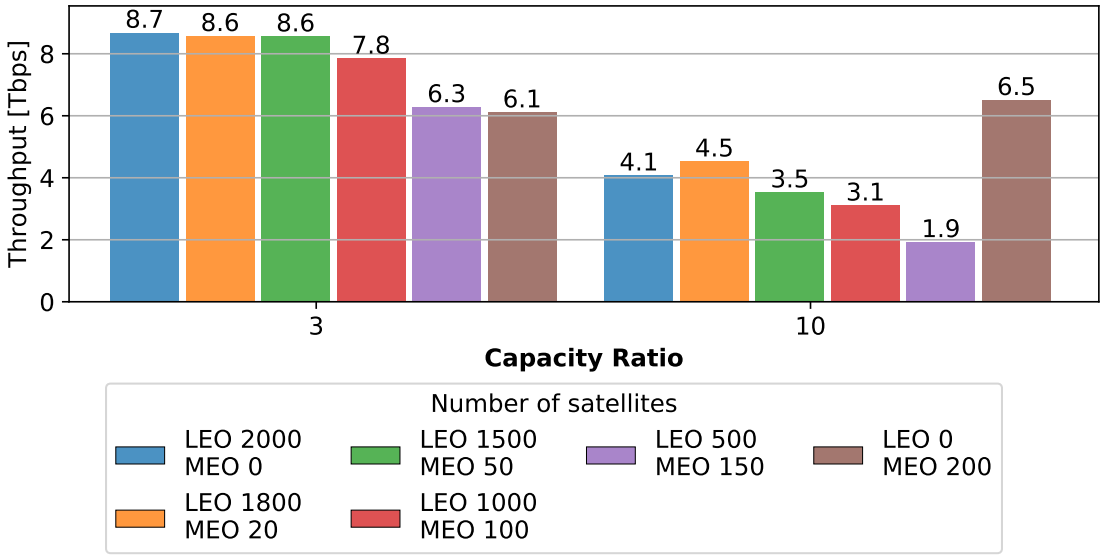


Figure D-11: Throughput of the optimal design when using the RAP method under a cost ratio of 10 using the user distribution proportional to land area

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

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

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

[6] Viasat Inc., "SAT-MPL-20200526-00056." `http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMPL2020052600056`, 2020. Accessed: 2023-03-08.

[7] 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.

[8] 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.

[9] Telesat, "Real-Time Latency: Rethinking Remote Networks." `https://www.telesat.com/wp-content/uploads/2022/11/Real-Time-Latency.pdf`, 2022. Accessed: 2023-03-08.

[10] 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.

[11] J. J. G. Luis, S. Eiskowitz, N. Pachler, E. F. Crawley, and B. G. Cameron, "Towards autonomous satellite communications: An ai-based framework to address system-level challenges," in *AI to Accelerate Science and Engineering Workshop at AAAI Conference on Artificial Intelligence 2022*, 2022.

[12] N. Pachler de la Osa, E. Crawley, and B. Cameron, "A unified resource allocation framework and impact evaluation for ngso satellite constellations (under review)," *Aerospace Science and Technology*.

[13] N. Pachler, "A complete resource allocation framework for flexible high throughput satellite constellations," Master's thesis, Massachusetts Institute of Technology Department of Aeronautics and Astronautics, Cambridge, Massachusetts, May 2022.

[14] 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.

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

[16] NASA, "Communications Satellites: Making the Global Village Possible." `https://history.nasa.gov/satcomhistory.html`, 2010. Accessed: 2023-03-08.

[17] T. Butash, P. Garland, and B. Evans, "Non-geostationary satellite orbit communications satellite constellations history," *International Journal of Satellite Communications and Networking*, vol. 39, no. 1, pp. 1–5, 2021.

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

[19] C. Ravishankar, R. Gopal, N. BenAmmar, G. Zakaria, and X. Huang, "Next-generation global satellite system with mega-constellations," *International journal of satellite communications and networking*, vol. 39, no. 1, pp. 6–28, 2021.

[20] A. Williams, O. Hainaut, A. Otarola, G. H. Tan, and G. Rotola, "Analysing the impact of satellite constellations and eso's role in supporting the astronomy community," *arXiv preprint arXiv:2108.04005*, 2021.

[21] Amanda Yeo, "SpaceX's Starlink announces it now has 1 million users." `https://mashable.com/article/starlink-spacex-1-million-users`, 2022. Accessed: 2022-12-19.

[22] WorldVu Satellites Limited, "SAT-LOI-20160428-00041." `http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATLOI2016042800041`, 2016. Accessed: 2023-03-09.

[23] WorldVu Satellites Limited, "SAT-MPL-20200526-00062." `http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMPL2020052600062`, 2020. Accessed: 2023-03-09.

[24] WorldVu Satellites Limited, "SAT-MPL-20210112-00007." `http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMPL2021011200007`, 2021. Accessed: 2023-03-09.

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

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

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

[28] Space Exploration Holdings, LLC, "SAT-MOD-20230215-00036." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2023021500036, 2023. Accessed: 2023-03-10.

[29] The Boeing Company, "SAT-LOA-20170301-00028." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATLOA2017030100028, 2017. Accessed: 2023-03-09.

[30] The Boeing Company, "SAT-MOD-20211104-00148." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2021110400148, 2021. Accessed: 2023-03-09.

[31] Telesat Canada, "SAT-PDR-20161115-00108." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATPDR2016111500108, 2016. Accessed: 2023-03-09.

[32] Telesat Canada, "SAT-MPL-20200526-00053." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMPL2020052600053, 2020. Accessed: 2023-03-09.

[33] Viasat Inc., "SAT-PDR-20161115-00120." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATPDR2016111500120, 2016. Accessed: 2023-03-08.

[34] O3b Limited, "SAT-MOD-20200526-00058." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2020052600058, 2020. Accessed: 2023-03-09.

[35] O3b Limited, "SAT-AMD-20220509-00048." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATAMD2022050900048, 2020. Accessed: 2023-03-09.

[36] CHINA TELECOM SATELLITE COMMUNICATIONS, "CHN2020-33636." https://www.itu.int/ITU-R/space/asreceived/Publication/DisplayPublication/23708, 2020. Accessed: 2023-03-09.

[37] CHINA TELECOM SATELLITE COMMUNICATIONS, "CHN2020-33634." https://www.itu.int/ITU-R/space/asreceived/Publication/DisplayPublication/23706, 2020. Accessed: 2023-03-09.

[38] Intelsat License LLC, "SAT-PPL-20211104-00146." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATPPL2021110400146, 2021. Accessed: 2023-03-09.

[39] SpaceX, "Capabilities and services." https://www.spacex.com/media/Capabilities&Services.pdf, 2019. Accessed: 2023-03-07.

[40] SpaceX, "Starlink Mission." https://web.archive.org/web/20190515091900/https://www.spacex.com/sites/spacex/files/starlink_press_kit.pdf, 2019. Accessed: 2023-03-07.

[41] Gunter Space, "SES-17." https://space.skyrocket.de/doc_sdat/ses-17.htm, 2023. Accessed: 2023-03-07.

[42] TeleGeography, "Internet Traffic and Capacity remain brisk." https://blog.telegeography.com/internet-traffic-and-capacity-remain-brisk, 2022. Accessed: 2023-03-07.

[43] G. V. Research, "Video streaming market size, share & trends analysis report by streaming type, by solution, by platform, by service, by revenue model, by deployment type, by user and segment forecasts, 2022-2030," tech. rep., Grand View Research, 2022. Available at: https://www.researchandmarkets.com/reports/4700903/video-streaming-market-size-share-and-trends.

[44] N. S. Research, "AERONAUTICAL SATCOM MARKETS, 10TH EDITION," tech. rep., Northern Sky Research, 2022. Available at: https://www.nsr.com/?research=aeronautical-satcom-markets-10th-edition.

[45] Wall Street Journal, "Faster Wi-Fi on Planes Is Taking Off. Profits Aren't." https://www.wsj.com/articles/faster-wi-fi-on-planes-is-taking-off-profits-arent-11666344792, 2022. Accessed: 2023-03-07.

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

[47] 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.

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

[49] 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.

[50] T. Leng, Y. Wang, D. Hu, G. Cui, and W. Wang, "User-level scheduling and resource allocation for multi-beam satellite systems with full frequency reuse," *China Communications*, vol. 19, no. 6, pp. 179–192, 2022.

[51] F. G. Ortiz-Gomez, D. Tarchi, R. Martínez, A. Vanelli-Coralli, M. A. Salas-Natera, and S. Landeros-Ayala, "Cooperative multi-agent deep reinforcement learning for resource management in full flexible vhts systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 1, pp. 335–349, 2021.

[52] 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.

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

[54] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, 1972.

[55] 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," *IEEE Transactions on Wireless Communications*, 2022.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] 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.

[61] 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.

[62] 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.

[63] 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.

[64] 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.

[65] 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.

[66] T. Q. Dinh, S. H. Dau, E. Lagunas, and S. Chatzinotas, "Efficient hamiltonian reduction for quantum annealing on satcom beam placement problem," *arXiv preprint arXiv:2210.06042*, 2022.

[67] 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.

[68] J. J. Sylvester, "A question in the geometry of situation," *Quarterly Journal of Mathematics*, 1857.

[69] 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.

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

[71] 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.

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] J. J. G. Luis, E. F. Crawley, and B. G. Cameron, "Applicability and challenges of deep reinforcement learning for satellite frequency plan design," in *2021 IEEE Aerospace Conference*, 2021.

[78] 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.

[79] 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.

[80] 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.

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

[82] 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.

[83] Y. Li, Y. Fan, S. Liu, L. Liu, and W. Yang, "Overview of beam hopping algorithms in large scale leo satellite constellation," in *2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pp. 1345–1351, IEEE, 2021.

[84] 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.

[85] 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.

[86] 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.

[87] 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.

[88] 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.

[89] 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.

[90] 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.

[91] 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.

[92] 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.

[93] 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.

[94] 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.

[95] 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.

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

[97] 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 Conference on Control Technology and Applications (CCTA)*, pp. 165–170, IEEE, 2018.

[98] 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.

[99] A. Xiao, Z. Chen, S. Wu, S. Jin, and L. Ma, "Collaborative long-short term bandwidth allocation for satellite-terrestrial networks," *IEEE Communications Letters*, vol. 26, no. 5, pp. 1121–1125, 2022.

[100] 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.

[101] 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.

[102] 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.

[103] 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.

[104] 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.

[105] 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.

[106] 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.

[107] 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.

[108] 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.

[109] 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.

[110] 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.

[111] 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.

[112] 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.

[113] 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.

[114] 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.

[115] 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.

[116] 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.

[117] 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.

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

[119] 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.

[120] 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.

[121] 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.

[122] 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.

[123] 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.

[124] 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.

[125] 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.

[126] 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.

[127] 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.

[128] 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.

[129] 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.

[130] 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.

[131] L. Wang, S. Liu, W. Wang, and Z. Fan, "Dynamic uplink transmission scheduling for satellite internet of things applications," *China Communications*, vol. 17, no. 10, pp. 241–248, 2020.

[132] 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.

[133] C.-Q. Dai, J. Luo, S. Fu, J. Wu, and Q. Chen, "Dynamic user association for resilient backhauling in satellite–terrestrial integrated networks," *IEEE Systems Journal*, vol. 14, no. 4, pp. 5025–5036, 2020.

[134] C.-Q. Dai, S. Li, J. Wu, and Q. Chen, "Distributed user association with grouping in satellite–terrestrial integrated networks," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 10244–10256, 2021.

[135] Y. Yin, C. Huang, D.-F. Wu, S. Huang, M. Ashraf, Q. Guo, and L. Zhang, "Deep reinforcement learning-based joint satellite scheduling and resource allocation in satellite-terrestrial integrated networks," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.

[136] H. Jiang, H. Wang, Y. Hu, and J. Wu, "Dynamic user association in scalable ultra-dense leo satellite networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 8, pp. 8891–8905, 2022.

[137] 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.

[138] 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.

[139] C. Liu and J. Kaiser, "Survey of mobile ad hoc network routing protocols," tech. rep., Universität Ulm, 2005.

[140] 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.

[141] 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.

[142] 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.

[143] 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.

[144] 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.

[145] 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.

[146] 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.

[147] 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.

[148] L. Qiao, H. Yan, X. Zhou, Y. Xu, L. Wang, and X. Wen, "Onboard centralized isl-building planning for leo satellite constellation networks," *Electronics*, vol. 12, no. 3, p. 635, 2023.

[149] Z. Lin, H. Li, J. Liu, Z. Lai, and G. Fan, "Inter-networking and function optimization for mega-constellations," in *2022 IFIP Networking Conference (IFIP Networking)*, pp. 1–9, IEEE, 2022.

[150] Y. Lee and J. P. Choi, "Connectivity analysis of mega-constellation satellite networks with optical intersatellite links," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 6, pp. 4213–4226, 2021.

[151] 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.

[152] 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.

[153] 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.

[154] Y. Liu, C. Li, J. Li, and L. Feng, "Joint user scheduling and hybrid beamforming design for massive mimo leo satellite multigroup multicast communication systems," *Sensors*, vol. 22, no. 18, p. 6858, 2022.

[155] 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.

[156] 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.

[157] 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.

[158] K. Y. Zhong, Y. J. Cheng, H. N. Yang, and B. Zheng, "LEO Satellite Multibeam Coverage Area Division and Beamforming Method," *IEEE Antennas and Wireless Propagation Letters*, vol. 20, no. 11, pp. 2115–2119, 2021.

[159] 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.

[160] 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.

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

[162] 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.

[163] 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.

[164] 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.

[165] 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.

[166] 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.

[167] 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.

[168] 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.

[169] 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.

[170] 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.

[171] 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.

[172] 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.

[173] 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.

[174] 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.

[175] 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 multi-beam 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.

[176] 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.

[177] 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.

[178] N. Pachler, J. J. G. Luis, M. Guerster, E. F. Crawley, and B. G. Cameron, "Allocating power and bandwidth in multibeam satellite systems using particle swarm optimization," in *2020 IEEE Aerospace Conference*, 2020.

[179] 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.

[180] 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.

[181] W. Gao, L. Wang, and L. Qu, "Research on joint resource allocation for multibeam satellite based on metaheuristic algorithms," *Entropy*, vol. 24, no. 11, p. 1536, 2022.

[182] 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.

[183] W. Gao, L. Qu, and L. Wang, "Multi-objective optimization of joint resource allocation problem in multi-beam satellite," in *2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, vol. 10, pp. 2331–2338, IEEE, 2022.

[184] 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.

[185] 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.

[186] 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.

[187] 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.

[188] 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.

[189] 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.

[190] 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.

[191] 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.

[192] Z. Lin, H. Li, Y. Li, J. Liu, L. Liu, Q. Zhang, Q. Wu, and Z. Lai, "Systematic utilization analysis of mega-constellation networks," in *2022 International Wireless Communications and Mobile Computing (IWCMC)*, pp. 1317–1322, IEEE, 2022.

[193] R. Deng, B. Di, H. Zhang, and L. Song, "Ultra-dense leo satellite constellation design for global coverage in terrestrial-satellite networks," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2020.

[194] H. Jia, Z. Ni, C. Jiang, L. Kuang, and J. Lu, "Uplink interference and performance analysis for megasatellite constellation," *IEEE Internet of Things Journal*, vol. 9, no. 6, pp. 4318–4329, 2021.

[195] F. Vidal, H. Legay, G. Goussetis, M. Garcia Vigueras, S. Tubau, and J.-D. Gayrard, "A methodology to benchmark flexible payload architectures in a megaconstellation use case," *International Journal of Satellite Communications and Networking*, vol. 39, no. 1, pp. 29–46, 2021.

[196] N. Okati, T. Riihonen, D. Korpi, I. Angervuori, and R. Wichman, "Downlink coverage and rate analysis of low earth orbit satellite constellations using stochastic geometry," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 5120–5134, 2020.

[197] Q. Ouyang, N. Ye, S. Miao, B. Kang, A. Wang, and L. Zhao, "Mega constellation networks are reliable against geographical failure," in *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*, pp. 1–5, IEEE, 2023.

[198] I. del Portillo, S. I. Dolan, B. G. Cameron, and E. F. Crawley, "Architectural decisions for communications satellite constellations to maintain profitability while serving uncovered and underserved communities," *International Journal of Satellite Communications and Networking*, vol. 41, no. 1, pp. 82–97, 2023.

[199] K. T. Li, C. A. Hofmann, H. Reder, and A. Knopp, "A techno-economic assessment and tradespace exploration of low earth orbit mega-constellations," *IEEE Communications Magazine*, 2022.

[200] L. Jia, Y. Zhang, J. Yu, and X. Wang, "Design of mega-constellations for global uniform coverage with inter-satellite links," *Aerospace*, vol. 9, no. 5, p. 234, 2022.

[201] A. H. Sánchez, T. Soares, and A. Wolahan, "Reliability aspects of mega-constellation satellites and their impact on the space debris environment," in *2017 Annual Reliability and Maintainability Symposium (RAMS)*, pp. 1–5, IEEE, 2017.

[202] K. Bhatia, "Starlinked! an analysis of spacex's small satellite mega-constellation under the fourth amendment," *Information & Communications Technology Law*, pp. 1–22, 2022.

[203] 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.

[204] SpaceX, "Starlink specifications." `https://www.starlink.com/legal/documents/DOC-1400-28829-70`, 2023. Accessed: 2023-08-02.

[205] ViaSat, "ViaSat Internet plans." `https://www.viasat.com/satellite-internet/plans/`, 2023. Accessed: 2023-08-02.

[206] Off J Eur Union, "European electronic communications code," tech. rep., 12, 2018.

[207] A. Raju, V. Gonçalves, S. Lindmark, and P. Ballon, "Evaluating impacts of oversubscription on future internet business models," in *International IFIP TC 6 Workshops, ETICS, HetsNets, and CompNets*, pp. 105–112, Springer, 2012.

[208] I. del Portillo, "ITU-Rpy: A python implementation of the ITU-R P. Recommendations to compute atmospheric attenuation in slant and horizontal paths." `https://github.com/iportillo/ITU-Rpy/`, 2017. Accessed: 2019-06-24.

[209] Recommendation ITU-R P.1511 (2019), "Topography for Earth-space propagation modelling," 2019.

[210] Recommendation ITU-R P.1510 (2017), "Mean surface temperature," 2017.

[211] Recommendation ITU-R P.453 (2019), "The radio refractive index: its formula and refractivity data," 2019.

[212] Recommendation ITU-R P.836 (2017), "Water vapour: surface density and total columnar content," 2017.

[213] Recommendation ITU-R P.837 (2017), "Characteristics of precipitation for propagation modelling," 2017.

[214] Recommendation ITU-R P.618 (2017), "Propagation data and prediction methods required for the design of Earth-space telecommunication systems," 2017.

[215] Recommendation ITU-R P.676 (2022), "Attenuation by atmospheric gases and related effects," 2022.

[216] Recommendation ITU-R P.840 (2019), "Attenuation due to clouds and fog," 2019.

[217] Recommendation ITU-R S.465 (2010), "Reference radiation pattern for earth station antennas in the fixed-satellite service for use in coordination and interference assessment in the frequency range from 2 to 31 GHz," 2010.

[218] Recommendation ITU-R S.1528 (2001), "Satellite antenna radiation patterns for non-geostationary orbit satellite antennas operating in the fixed-satellite service below 30 GHz," 2001.

[219] FG-NET2030-Arch, "Network 2030 – architecture framework," tech. rep., ITU-T, 2020.

[220] White Paper, "European vision for the 6g network ecosystem," tech. rep., 5G-PPP, 2021.

[221] 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.

[222] WorldVu Satellites Limited, "SES-LIC-20180727-02075." `http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SESLIC2018072702075`, 2018. Accessed: 2023-08-07.

[223] Space Exploration Holdings, LLC, "SES-LIC-20220514-00525." `http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SESLIC2022051400525`, 2022. Accessed: 2023-08-07.

[224] Space Exploration Holdings, LLC, "SES-LIC-20220514-00534." `http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SESLIC2022051400534`, 2022. Accessed: 2023-08-07.

[225] Kuiper Systems LLC, "SES-LIC-20210409-00634." http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SESLIC2021040900634, 2021. Accessed: 2023-08-07.

[226] ViaSat, Inc, "Viasat's ground network evolution." https://news.viasat.com/blog/scn/viasats-ground-network-evolution, 2023. Accessed: 2023-08-07.

[227] N. Torkzaban, A. Gholami, J. S. Baras, and C. Papagianni, "Joint satellite gateway placement and routing for integrated satellite-terrestrial networks," in *ICC 2020-2020 IEEE international conference on communications (ICC)*, pp. 1–6, IEEE, 2020.

[228] 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.

[229] 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," *International Journal of Satellite Communications and Networking*, vol. 39, no. 1, pp. 47–64, 2021.

[230] 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.

[231] R. M. Karp, *Reducibility among combinatorial problems*. Springer, 2010.

[232] M. R. Garey and D. S. Johnson, "Computers and Intractability. A Guide to the Theory of NP-Completeness," *The Journal of Symbolic Logic*, vol. 48, no. 2, pp. 498–500, 1983.

[233] R. Williams, *The geometrical foundation of natural structure: A source book of design*. Dover Publications, 1979.

[234] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2023.

[235] M. R. Garey, D. S. Johnson, and L. Stockmeyer, "Some simplified np-complete problems," in *Proceedings of the sixth annual ACM symposium on Theory of computing*, pp. 47–63, 1974.

[236] J. Geng, D. Sun, W. Wang, and Y. Liu, "Interference prediction between leo constellations based on a novel joint prediction model of atmospheric attenuation," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 950–955, IEEE, 2022.

[237] G. Casadesus, J. J. G. Luis, N. Pachler, E. F. Crawley, and B. G. Cameron, "Dynamic frequency assignment for mobile users in multibeam satellite constellations," in *International Astronautical Congress*, 2022.

[238] U. Pferschy and J. Schauer, "The maximum flow problem with disjunctive constraints," *Journal of Combinatorial Optimization*, vol. 26, no. 1, pp. 109–119, 2013.

[239] F. Bacchus and P. Van Beek, "On the conversion between non-binary and binary constraint satisfaction problems," in *AAAI/IAAI*, pp. 310–318, 1998.

[240] A. R. Mahjoub and S. T. McCormick, "Max flow and min cut with bounded-length paths: complexity, algorithms, and approximation," *Mathematical programming*, vol. 124, pp. 271–284, 2010.

[241] D. S. Hochbaum and A. Segev, "Analysis of a flow problem with fixed charges," *Networks*, vol. 19, no. 3, pp. 291–312, 1989.

[242] F. C. Commission, "FCC INCREASES BROADBAND SPEED BENCHMARK," tech. rep., Federal Communications Commission, 2024. Available at: https://docs.fcc.gov/public/attachments/DOC-401205A1.pdf.

[243] J. Dingley, P. Belobaba, E. F. Crawley, and B. G. Cameron, "Winning the internet: A tool for simulating the competitive strategies of satellite-based internet providers," in *International Astronautical Congress*, 2022.