A scalable algorithm for User Grouping in LEO-MEO-HEO Hybrid Very High Throughput Satellite Constellations

Nils Pachler, Edward F. Crawley, Bruce G. Cameron

Abstract

Satellite communications are transforming the *when* and *how* people can access broadband Internet, enabling connectivity in markets unreachable by terrestrial networks, such as isolated regions or *connectivity on-the-go*. Part of the new architectures' success relies on intricate hybrid constellation designs that combine multiple orbital shells at different altitudes, such as the SpaceX 4-shell LEO and the Boeing 10-shell LEO-MEO-HEO constellations. Under the new systems, satellite operators will need automated and scalable mechanisms able to efficiently group and distribute individual customers across satellites (the *User Grouping* problem) in order to maximize satellite utilization and achieve increased constellation capacity. While previous studies propose methods for single-altitude designs, algorithms for hybrid systems are yet to be developed. This work aims to breach this gap by 1) formulating the User Grouping problem for hybrid constellations as a Mixed Integer Linear problem, and 2) developing a scalable methodology tailored for high-dimensional scenarios. By using the SpaceX and the Boeing constellations as examples, this work demonstrates that the proposed approach can provide high quality solutions in feasible time for scenarios with up to 100,000 customers, which represents realistic operational conditions, with a minimum 77% reduction in maximum satellite load compared to methods for single-altitude designs.

Index Terms

Satellite Communications, Resource Management, Integer Programming, Beam steering

Corresponding author: Nils Pachler (pachler@mit.edu)

All authors are with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 02139 Cambridge MA, USA (e-mail: pachler@mit.edu, crawley@mit.edu, bcameron@mit.edu)

I. Introduction

The upcoming generation of satellite communication constellations will revolutionize the accessibility and timing of broadband Internet for societies. Notably, the innovative system designs put forth by established industry players (e.g., SES and ViaSat) and emerging entrants (e.g., SpaceX and Amazon) will expand the boundaries of satellite network capacity [1], reaching the order of tens of terabits per second (Tbps). These advancements in satellite communication offer substantial potential to address connectivity requirements in various markets, including remote and rural areas with insufficient terrestrial infrastructure [2], as well as meet on-the-go connectivity needs in sectors such as aviation and maritime domains [3].

To attain their expected performance, novel systems rely on two key technological advancements: 1) improved spacecraft payload, which enables better utilization of limited resources, made possible through upgrades to satellite components (such as phased array antennas, or adaptive modulation and coding) and 2) a substantially larger satellite segment distributed across different altitudes, which expands the pool of available resources, made possible through reduced manufacturing and launching costs.

The augmented capabilities of novel designs, however, entail heightened operational complexity. To effectively utilize the new flexibilities, operators require automated systems that can leverage the innovative payload and constellation capabilities to optimize performance. In this line, recent research has focused on maximizing satellite performance through the development of autonomous techniques that can proficiently manage payload flexibilities [4]. Nevertheless, under the novel constellation configurations, satellites do no exist as individual entities, but as a network of interconnected resources. How to effectively leverage the capabilities of multiple satellites, located at varying altitudes, to optimize capacity remains an open challenge [5]. Addressing this issue can lead to improved communication capabilities, enhanced connectivity in underserved regions, and greater accessibility to broadband Internet, ultimately bridging the digital divide and fostering socio-economic development.

To that end, this study aims to develop a novel formulation that maximizes capacity by effectively leveraging the interconnected resources of multiple satellites located at varying altitudes. The proposed approach encompasses both an optimal method, which utilizes the complete formulation, and a scalable approach that iteratively addresses smaller portions of the problem. The results demonstrate that the scalable approach can deliver high quality solutions in feasible time for scenarios with up to 100,000 customers, which represents realistic operational conditions, with a minimum 77% reduction in maximum satellite load compared to current methods. Even more, a comparison under low-dimensionality conditions reveals both the optimal and scalable methods yield solutions with comparable performance.

The remainder of the paper is organized as follows: section II provides a comprehensive literature review, exploring the flexibilities of hybrid constellations and establishing the contributions of this work; section III defines a novel formulation to maximize the capacity on hybrid constellations; section IV details the two approaches presented in this work for low and high dimensional scenarios; section V presents the experimental results of this work, which include validation, tradespace, and performance analyses; section VI discusses the validity of the assumptions the formulation is based on; and section VII concludes this study by summarizing the main findings and their significance.

II. Background and Related work

One of the complexities of hybrid constellations compared with single-altitude designs is that the footprint of the beam (i.e., the intersection between the beam projected from the satellite and the Earth) might differ depending on the altitude (see Figure 1). This implies that a user (i.e., a ground terminal that request service) who could be covered by a beam from a higher altitude satellite may no longer be within the footprint if served by a lower altitude satellite. To address this challenge, one common approach is to adjust the aperture angle of the beam (also referred to as beamwidth) based on the altitude to achieve similar footprints [6]. To establish a viable user-to-beam mapping, known as the User Grouping problem, maintaining similar footprints allows for simple, but efficient solutions. The industry-standard grid-like approach [7] achieves this by dividing the Earth surface into a grid of beams. Combining this method with Beam Hopping, which serves a subset of beams concurrently, enables operators to concentrate constellation resources on high-demand areas while streamlining operations, ultimately enhancing overall performance [8].

In addition to grid-like methods, alternative user-centric approaches have been proposed in the literature. These methods focus on identifying a reduced set of beams while distributing user demand across them effectively [9]–[12]. These works aim to optimize the user-to-beam mapping using various optimization approaches (k-means [9], genetic algorithm [10], quantum annealing [11], and convex-concave procedure [12]). [13] demonstrates that grid-like and user-

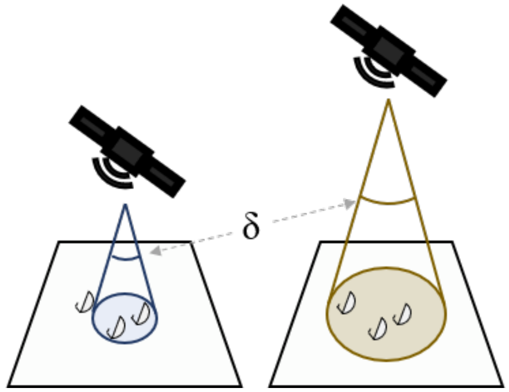


Fig. 1. Footprint of satellites at different altitudes for the same beam shape

centric approaches yield comparable performance when assuming a uniform footprint. However, in practical scenarios, modern satellite antennas allow only a limited variation in the aperture angle during operations [14]. Consequently, a uniform footprint is feasible primarily for constellations with satellites at similar altitudes, where the variation in aperture angle remains minimal, thereby reducing the requirements on payload technology. In the case of hybrid constellations combining LEO (low Earth orbit), MEO (medium Earth orbit), and HEO (high Earth orbit), where satellites may be at significantly different altitudes, utilizing different shapes becomes a necessity. Furthermore, we argue that leveraging the flexibility offered by satellites at varying altitudes by enabling distinct footprints does not necessarily compromise operational efficiency and can unlock substantial capacity gains.

Recent literature has explored the possibility of including Beam Shaping capabilities, which involve changing the shape of the beam, into the User Grouping problem. In this context, the footprint of the beam is included as an additional flexibility, leading to increased performance. Similar to scenarios with a uniform footprint, the literature is categorized into grid-like and user-centric approaches. Grid-like methods aim to cover specific areas with beams, adjusting their shape to achieve balanced demand. For instance, Honnaiah et al. [15] utilizes Voronoi maps for adaptive beam layout plans, generalized in [16] with a broader optimization framework considering inter-beam interference. Regarding user-centric approaches, techniques include clique cover with stochastic optimization [17], k-means clustering [18], mixed-integer linear programming with k-means [19], p-center clustering [20], and hierarchical clustering [21]. Notably, these approaches assume that the footprint is a controllable variable, overlooking the influence of the altitude on this parameter. Additionally, the controllable range of the aperture angle is relatively small [14]. Consequently, existing methodologies are ill-suited for hybrid constellations, which necessitate accounting for the relationship between satellite altitudes and beams while mapping users and shapes to beams. Unfortunately, to the best of the authors' knowledge, no previous studies have effectively addressed the User Grouping problem considering the influence of variable footprints resulting from satellites positioned at significantly different altitudes.

This research aims to address this gap in the literature by proposing a novel approach that considers the interdependence between satellite altitudes and beams while solving the User Grouping problem with variable beam shapes. In particular, the contributions of this paper are as follows:

- A novel multi-objective formulation for the User Grouping problem that connects users, shapes, and altitudes with beams, thus being able to leverage the flexibilities of hybrid constellations.
- Two mixed integer linear programming approaches to address the complexity of this formulation and obtain optimized solutions, one optimal for low dimensionality, and one scalable for high dimensionality.
- Validation and performance analyses on the SpaceX and Boeing constellation designs to verify and assess the impact of the proposed methodology to solve the User Grouping problem.
- A trade-space analysis to determine the trade-off between the different objectives and their impact on performance.

III. Problem definition

This section describes the User Grouping problem for hybrid constellations. To that end, a novel linear multi-objective formulation, which includes the necessary variables, constraints, and objectives, is described.

A. Problem description

The User Grouping problem is a crucial challenge within satellite communication systems, involving the mapping of users to beams to maximize capacity. In this context, users refer to ground terminals that request services from a satellite operator, while beams represent the electromagnetic signals emitted by satellites to fulfill users' needs. Successful communication hinges on two key requirements: 1) each user must be consistently associated with a beam,

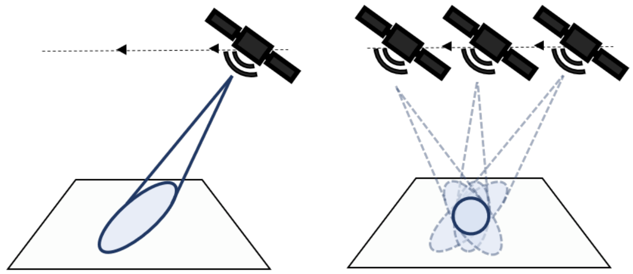


Fig. 2. Difference between footprint and footprint contour. A footprint corresponds to the intersection of a beam with the Earth's surface at a specific point in time, while a footprint contour corresponds to the intersection of all footprints with the same beam center at any point in time.

and 2) each user must fall within the footprint covered by the corresponding beam. Although the User Grouping process does not directly impact spectrum or power usage, previous studies [13] have demonstrated that reducing the number of beams can significantly enhance capacity, provided the demand of each beam remains within the bounds dictated by the available spectrum and power resources. This improvement stems from the ability to group users into a smaller set of beams, thereby optimizing spatial separation and frequency spectrum utilization.

In the context of hybrid constellations, one potential approach to reduce the number of beams is by leveraging the largest available footprint, typically associated with the satellite at the highest altitude. However, when served by satellites at lower altitudes, due to the smaller footprints of such satellites, there is a risk of users falling outside the coverage area (refer to Figure 1). To address this concern, all users would need to be associated with the highest altitude satellites, while satellites at lower altitudes would remain unused. Consequently, such an arrangement diminishes the overall capacity of the constellation. Therefore, maximizing capacity in hybrid constellations requires striking a balance between minimizing the number of beams and appropriately distributing them across different altitudes.

B. Problem set-up

This study considers a generic LEO-MEO-HEO constellation, distributing satellites across altitudes denoted by $\mathcal{A} = \{a_1, a_2, ..., a_A\}$. It assumes that the mean altitude for each satellite remains fixed, and the eccentricity is negligible, thereby causing minimal variations in footprint size. Additionally, satellites associated with a specific altitude are assumed to: 1) possess the same capacity c_a, 2) continuously cover a designated portion of the Earth surface, and 3) generate beams of shape ϕ_a. While ϕ_a may be consistent across altitudes, variations in the generated footprints can arise due to geometric considerations. The assumption includes satellites equipped with a programmable payload capable of adjusting beam direction over time based on ground commands. These assumptions align with contemporary satellite constellation design proposals [22]–[24]. Notably, although permitted in the following formulation, satellites are not required to possess beam shaping capabilities.

C. Problem formulation

The User Grouping problem involves assigning a set of users, denoted by $\mathcal{U} = \{u_1, u_2, ..., u_U\}$, to a set of beams, denoted by $\mathcal{B} = \{b_1, b_2, ..., b_B\}$. Each user u is characterized by its position p_u (assumed to be static, see discussion in section VI-B) and demand d_u, while each beam b is defined by its center position p_b, shape ϕ_b, and associated altitude a_b. It is assumed that the shape of each beam corresponds to a static circular shape (see discussion in section VI-C). It is assumed that each beam can only be served by satellites located at altitude a_b. For the purpose of this work, the user-beam mapping must remain fixed over an extended period of time, and users associated with a beam must always fall within the footprint of the beam. Note that for non-geostationary orbits (NGSO), the footprint of the beam changes over time.

To eliminate the time dependency, we introduce a *footprint contour* Φ_b for each beam, representing the intersection of all footprints at any given time (see Figure 2). Notably, the footprint contour is geometrically defined by the beam center p_b, the shape ϕ_b, and the set of visible and attainable positions by a satellite at altitude a_b. The enclosed surface within the footprint contour corresponds to positions that consistently fall under the beam coverage at all times. Conversely, a point outside the footprint contour may experience occasional coverage but cannot be covered continuously. By associating each beam with a specific footprint contour, we can ensure continuous coverage if a user is assigned to beam b and its position p_u lies within the footprint contour Φ_b. The number of unique footprint contours remains relatively low, primarily determined by the specific shape ϕ_b of the beam and the altitude a_b of the corresponding satellite, both of which typically exhibit limited variability.

To determine the mapping of users to beams, we introduce two new binary variables: $x_{u,b}$ and y_b. The variable $x_{u,b}$ indicates whether a user u is mapped to a beam b. In cases where certain users cannot be associated with specific beams due to power or antenna constraints, or other factors (e.g., a small handheld device being associated with a geostationary satellite), we enforce $x_{u,b} = 0$. To ensure that each user is mapped to exactly one beam, we impose the constraint $\sum_b x_{u,b} = 1$. Moreover, $x_{u,b}$ can only take the value 1 if the user's position falls within the footprint contour of the corresponding beam, which we express using the inequality $x_{u,b} \leq \mathbb{1}_{p_u \in \Phi_b}$. The variable y_b corresponds to an activation variable, set to 1 when at least one user is mapped to beam b, which can be encoded using the following inequality:

$$\sum_u \mathbb{1}_{p_u \in \Phi_b} y_b \geq \sum_u x_{u,b} \quad \forall \mathcal{B} \tag{1}$$

To distribute the load evenly across all altitudes, we calculate the demand of each beam b as the sum of the demand of all the users mapped to that beam, which is expressed as $\sum_u d_u x_{u,b}$. However, prior to computing the load at a specific altitude, it is important to consider that the capacity is highly localized. Specifically, the capacity of the satellite constellation is dependent on the satellites visible within a given region. To account for this, we introduce a new variable $\nu_{b,p,a}$, which we call the *impact* of beam b to the point p at altitude a. The point p is defined using a combination of longitude and latitude coordinates, and $\mathcal{P}$ denotes the set containing all points p. The value of $\nu_{b,p,a}$ is determined by three factors:

1) If the beam is not associated with the altitude a, the impact is zero.
2) If the beam is not visible from a satellite at position p and altitude a, the impact is zero.
3) If neither of the preceding conditions is true, then the impact is defined as $\frac{1}{N_{b,p,a}}$, where $N_{b,p,a}$ denotes the minimum number of satellites at altitude a that are visible to beam b.

In summary, the impact of a beam b to a point p at altitude a quantifies the expected portion load of the beam that will be absorbed by the a satellite at point p, given the localized capacity of the constellation. Note that the impact of a beam depends only on the geometry of the problem, and is not an optimization variable. Now, we proceed to calculate the load at a specific point for a given altitude, represented by the auxiliary variable $\eta_{p,a}$ as follows:

$$\eta_{p,a} = \sum_b \nu_{b,p,a} \sum_u d_u x_{u,b} \quad \forall p \in \mathcal{P}_a, \forall a \tag{2}$$

To determine if a particular altitude can serve a specific point, we define $\mathcal{P}_a$ as the set of points for which there exists at least one beam with a non-zero impact at that altitude. Utilizing this set, we can compute the maximum (μ_p^+) and minimum (μ_p^-) load density over all altitudes as follows:

$$\begin{aligned}\mu_p^+ &\geq \frac{\eta_{p,a}}{c_a} \quad \forall p \in \mathcal{P}_a, \forall a \in \mathcal{A} \\ \mu_p^- &\leq \frac{\eta_{p,a}}{c_a} \quad \forall p \in \mathcal{P}_a, \forall a \in \mathcal{A}\end{aligned} \tag{3}$$

Note that μ_p^+ and μ_p^- are auxiliary variables with a value at least as high as the most loaded altitude, and at most as high as the least loaded altitude, respectively. To account for the fact that different altitudes may have satellites with varying capacities, we normalize the load of each altitude by the reference capacity of the satellites at that altitude. Thus, we can evaluate the quality of the load distribution across altitudes by calculating the difference between the most and least loaded altitudes over all points, denoted with the auxiliary variable γ:

$$\gamma \geq \mu_p^+ - \mu_p^- \quad \forall p \tag{4}$$

It should be noted that this metric considers potential disparities in the load across different regions, as the objective is to minimize the maximum difference across all regions rather than optimizing for the highest value.

Building upon the rationale presented in Section III-A, optimizing capacity within the User Grouping problem for hybrid constellations hinges on finding a equilibrium between reducing the number of beams and distributing them evenly across altitudes. Minimizing the number of beams is relevant to increase spectrum efficiency [13], while ensuring load balance across satellites enhances overall satellite utilization [25]. Both metrics have demonstrated significance in maximizing performance within the domain of satellite constellations, and at this point, it remains uncertain which factor has a more pronounced impact on the User Grouping problem. In this formulation, minimizing the number of beams can be achieved by minimizing the sum of y_b across all beams $\mathcal{B}$. Balancing the load across altitudes can be achieved by minimizing the load difference between the most and least loaded altitudes, represented by γ. By employing a weighted sum of these two objectives, the complete formulation is as follows:

$$\begin{aligned}
x_{u,b}, y_b, \eta_{p,a}, \mu_p^+, \mu_p^-, \gamma = \\
\text{argmin} \quad & w_l \gamma + w_d \sum_b y_b \\
s.t. \quad & \gamma \geq \mu_p^+ - \mu_p^- && \forall p \in \mathcal{P} \\
& \mu_p^+ \geq \frac{\eta_{p,a}}{c_a} && \forall p \in \mathcal{P}_a, \forall a \in \mathcal{A} \\
& \mu_p^- \leq \frac{\eta_{p,a}}{c_a} && \forall p \in \mathcal{P}_a, \forall a \in \mathcal{A} \\
& \eta_{p,a} = \sum_b \nu_{b,p,a} \sum_u d_u x_{u,b} && \forall p \in \mathcal{P}, \forall a \in \mathcal{A} \\
& x_{u,b} \leq \mathbb{1}_{p_u \in \Phi_b} y_b && \forall\, u \in \mathcal{U}, b \in \mathcal{B} \\
& \sum_b x_{u,b} = 1 && \forall\, u \in \mathcal{U} \\
& x_{u,b} \in \{0, 1\} && \forall\, u \in \mathcal{U}, b \in \mathcal{B} \\
& y_b \in \{0, 1\} && \forall\, b \in \mathcal{B}
\end{aligned} \tag{5}$$

Here, we introduce the weight parameters w_l and w_d to accommodate the two different objectives of balancing the load across altitudes and reducing the number of beams, respectively. Increasing the value of w_l encourages a more balanced load distribution across the different altitudes, thereby enhancing the overall efficiency of the constellation through more effective utilization of the satellites. However, this comes at the expense of managing a larger number of beams. Conversely, increasing the value of w_d may lead to suboptimal satellite utilization but can result in a reduction in the number of beams. The selection of appropriate values for w_l and w_d will be discussed in section V-C. This formulation corresponds to a NP-Hard formulation of the User Grouping problem, where the associated decision problem is NP-Complete (see Appendix A).

IV. Proposed approaches

This section describes two approaches to resolve the proposed formulation: one optimal for low dimensionality, and one scalable for high dimensionality. In addition, the complexity analysis of each formulation is detailed.

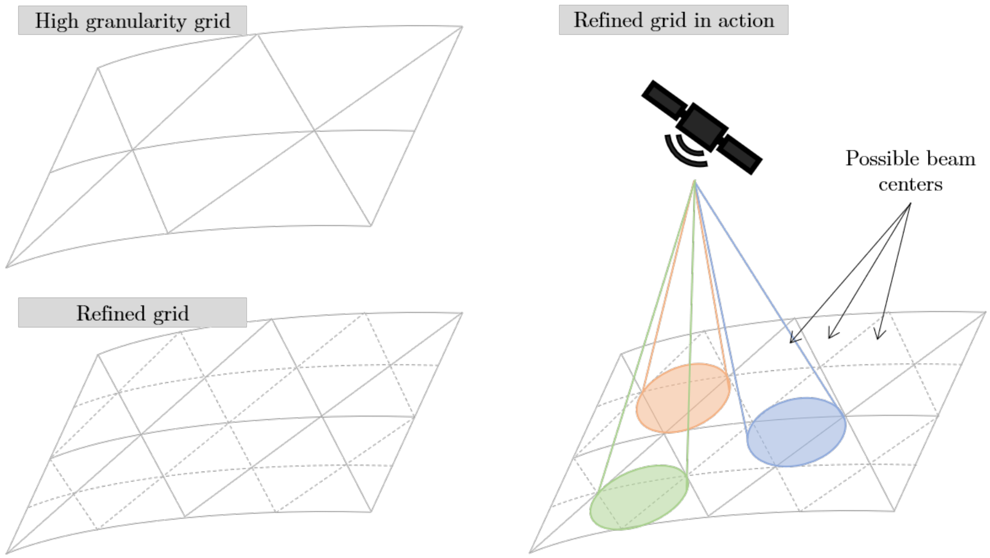


Fig. 3. 3D example of a grid construction using an icosahedral tessellation

A. Listing the possible beams

To address the problem at hand, the initial step is to obtain a representative list of possible beams. However, due to the infinite number of possible beams resulting from the continuous nature of beam centers, enumerating all beams becomes computationally intractable. Previous studies [17], [20], [26] have tackled this issue by identifying the minimum number of beams required to cover all users. Additionally, further research [13] has demonstrated the advantages of reducing the number of beams in terms of resource consumption, highlighting that a grid-like approach can be as effective as more complex optimization frameworks. Building upon these findings, we propose utilizing a grid-based approach for determining the possible beam centers.

To generate a high-quality grid with the desired granularity, we employ the concept of spherical tessellation, as detaled in [27]. This concept is equivalent to the notion of hexagonal grid, commonly used by satellite operators [28]. Initially, we cover the Earth's surface using a known grid composed of triangles with high granularity. Each initial triangle is then recursively subdivided into smaller triangles until the sub-triangle size is sufficiently small to fit within a specified footprint contour. The center of each final triangle defines a potential beam center. Figure 3 illustrates the construction of such a representation, using a simplified example. As the specific beam centers depend on the choice of the initial grid, we generate N_{tess} randomized lists of initial positions to obtain different grids. The set of possible beams $\mathcal{B}$ corresponds to obtaining the refined grid for each footprint contour size a total of N_{tess} times. It is important to note that the number of possible beams is directly related to the number of altitudes, the number of footprint contours considered per altitude, the granularity of the grid, and the number of different initializations. As a reference, a constellation with four altitudes, one contour per altitude, a grid cell area of 400 km^2, and 100 random initializations results in approximately 2×10^8 possible beams. This estimation is applicable to constellations like SpaceX, providing an order of magnitude for the number of possible beams.

However, it is worth mentioning that many beams will not cover any users, particularly those over oceanic areas, while some beams will cover the same set of users from the same altitude and footprint, with slight variations in beam center. To reduce the number of possible beams, we eliminate beams if we can find another beam in the set that covers an equal or larger number of users at the same altitude using the same shape. Additionally, we exclude beams that cannot be served continuously, meaning beams for which the minimum number of visible satellites $N_{b,p,a}$ is zero.

B. *Optimal approach*

The proposed formulation in this study presents a mixed-integer linear formulation of the User Grouping problem, which encompasses a combination of continuous and discrete variables subjected to linear constraints and objectives. Due to the inclusion of integer variables, the problem exhibits inherent complexity and falls into the realm of NP-hard problems. Nonetheless, for scenarios with a reduced number of variables, commercially available solvers can be utilized to identify an optimal solution. In this research, we refer to the direct utilization of a commercial solver on the proposed formulation as the *optimal approach*.

To mitigate the computational burden associated with solving the problem, we employ a two-step approach. Initially, the algorithm is executed using a smaller set $\mathcal{P}$ of the problem, thereby yielding an solution for a low quality grid. Subsequently, the algorithm is re-executed using an enlarged set $\mathcal{P}$ with the warm-start obtained from the previous step. This strategic approach significantly diminishes the computational time required while still attaining the optimal solution for the problem.

C. Scalable approach

In high-dimensional scenarios, applying commercial solvers directly on the proposed mixed-integer formulation may be computationally intractable within a reasonable amount of time. To tackle this issue, we propose a scalable sub-optimal approach that leverages the flexibility of hybrid constellations while still achieving high-quality solutions within practical time limits. The proposed approach involves an iterative process that fixes some of the variables at each iteration and solves for the remaining variables. In particular, at each iteration, we undertake the following steps:

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 the set $\mathcal{U}^*$.
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 the improvement in the objective value between iterations falls below a threshold I_{thres} for N_{conv} consecutive iterations. I_{thres} is defined as a threshold below which two solutions are considered to perform equally. This iterative approach enables the discovery of near-optimal solutions for the User Grouping problem, even in high-dimensional scenarios where the direct utilization of commercial solvers is computationally infeasible. We refer to this approach as the *scalable approach*. The pseudo-code for this approach is summarized in Algorithm 1. Similar to the optimal approach, we initially obtain a warm-start solution using a smaller set $\mathcal{P}$, which we subsequently refine using a larger set $\mathcal{P}$.

D. Complexity analysis

To assess the complexity of the proposed approaches, we introduce the parameter M, representing the maximum number of beams covering a user. We compute M as $M = \max_u \sum_b \mathbb{1}_{p_u \in \Phi_b}$. The problem formulation comprises two types of variables: decision variables ($x_{u,b}$), which scale

Algorithm 1 Iterative, scalable approach the mapping between users and beams

Input: $\mathcal{B}, \mathcal{U}$
Input: $N_b, N_u, N_{conv}, I_{thres}$
Output: $x_{u,b} \forall u \in \mathcal{U}, b \in \mathcal{B}$

$x_{u,b}$ = Random Initialization or Warm-Start
i=0, Prev=Inf
converged=False
while not converged **do**
  $\mathcal{B}^* = \mathcal{B}^i = \mathcal{U}^* = \mathcal{U}^i = \emptyset$
  $\mathcal{U}^* = \mathcal{U}^* \cup \{Rand(\mathcal{U})\}$
  while $|\mathcal{B}^i| < N_b$ and $|\mathcal{U}^i| < N_u$ and $\mathcal{U}^* \neq \emptyset$ **do** ▷ Terminate when
    ▷ enough users/beams are selected
    $u = Rand(\mathcal{U}^*)$
    for $b \in \mathcal{B}$ **do** ▷ Add all associated beams
      ▷ to the possible beams set
      if $p_u \in \Phi_b$ and $b \notin \mathcal{B}^i$ **then**
        $\mathcal{B}^* = \mathcal{B}^* \cup \{b\}$
    $\mathcal{U}^* = \mathcal{U}^* \setminus \{u\}$
    $\mathcal{U}^i = \mathcal{U}^i \cup \{u\}$ ▷ Include the selected user
    for $b \in \mathcal{B}^*$ **do**
      for $u \in \mathcal{U}$ **do** ▷ For each possible beam, add all associated
        ▷ users to the possible users set
        if $p_u \in \Phi_b$ and $u \notin \mathcal{U}^i$ **then**
          $\mathcal{U}^* = \mathcal{U}^* \cup \{u\}$
    $\mathcal{B}^i = \mathcal{B}^i \cup \mathcal{B}^*$ ▷ Include the selected beams
    $\mathcal{B}^* = \emptyset$
  Solve Equation 5 using commercial solvers and assuming $\mathcal{U} \setminus \mathcal{U}^i$, and $\mathcal{B} \setminus \mathcal{B}^i$ are fixed
  if Prev - Current Objective $> I_{thres}$ **then** ▷ Check convergence
    i=0
  else
    i = i + 1
    if i = N_{conv} **then**
      converged=True
  Prev = Current Objective

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 $|\mathcal{P}|$. Thus, the overall memory requirement of the problem is $\mathcal{O}(M|\mathcal{U}| + |\mathcal{B}| + 2|\mathcal{P}| + 1)$. Notably, once all $x_{u,b}$ values are computed, the auxiliary decision variables remain fixed. As a result, the problem's search space exhibits exponential growth in $M|\mathcal{U}|$ due to the binary nature of the decision variables. In particular, since the problem is an NP-hard integer linear problem, and each user needs to select one beam, the worst-case computation time scales as $\mathcal{O}(M^{|\mathcal{U}|})$. This is the expected complexity of the optimal approach.

Conversely, the scalable method reduces complexity by sacrificing optimality and allowing for a smaller number of free variables. Specifically, the number of varying decision variables is now $\mathcal{O}(MN_u)$, while the number of auxiliary variables is $\mathcal{O}(N_b + M + 2|\mathcal{P}| + 1)$. Consequently, the memory load of the scalable approach is $\mathcal{O}(MN_u + N_b + M + 2|\mathcal{P}| + 1)$. Similar to the optimal approach, once the decision variables are solved, the values of auxiliary variables become fixed,

TABLE I
SUMMARY OF THE ORBIT CHARACTERISTICS OF THE PROPOSED CONSTELLATIONS. *A* REPRESENTS SHELLS APPROVED BY THE FCC AND *P* REPRESENTS PENDING CHANGES. * DENOTES ORBITAL PLANES WITH NON-ZERO ECCENTRICITY. TABLE ADAPTED FROM [29] WITH PERMISSION OF THE AUTHORS.

System	Altitude (km)	Inclination (°)	Planes	Satellites per plane	State	Number of satellites
SpaceX	540	53.2	72	22	A	4,408
	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
	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	2	A	
	35,786*	63.4	5	1	A	

reducing the search space to $\mathcal{O}(M^{N_u})$ in the worst case. This scalability improvement, however, comes at the expense of potential optimality loss. By adjusting hyperparameters N_u and N_b, we can regulate the complexity of the problem solved by the mathematical solver.

V. RESULTS

This section details the simulation results when applying the proposed methodology under realistic operational scenarios. In particular, this section studies the validity of the proposed framework, explores the trade-off between the different metrics, and evaluates the performance of the solution against existing methods in cases with up to 100,000 users.

A. Experimental set-up

To assess the effectiveness of the proposed algorithms, it is necessary to establish the characteristics of the satellite constellation and user distribution. In this study, we based our analysis on the latest public filings for two existing constellation designs: SpaceX Starlink [22], [30]–[33] and The Boeing Company [23], [34]. The orbital configurations of these constellations are summarized in Table I (adapted from [29] with permission of the authors). The SpaceX constellation was chosen for validation purposes due to the similar orbital configurations exhibited by satellites at different altitudes. Additionally, the Boeing constellation was selected to evaluate the

TABLE II
SUMMARY OF THE PARAMETERS OF THE SIMULATION. * THE TIME LIMIT IS ONLY APPLICABLE FOR THE SCALABLE APPROACH.

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

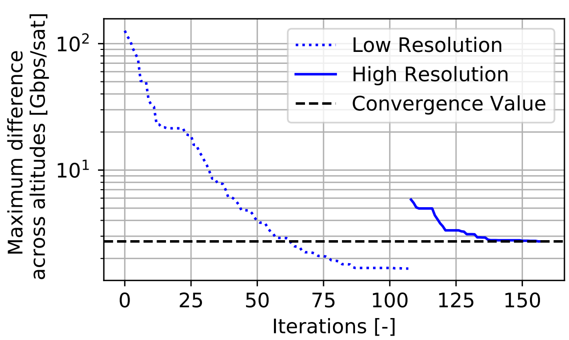
algorithms' performance against designs featuring multiple altitudes in LEO, MEO, and HEO. It is assumed that all satellites have payload proportional to the requirements of the altitude, such that the capacity for all satellites can assumed to be the same ($c_a = c \ \forall a \in \mathcal{A}$). Circular beams with a 2° aperture angle and a minimum elevation angle of 25° are considered for both constellations , which are aligned with the public specifications of modern constellations [22], [28]. The values for $N_{b,p,a}$ were obtained through constellation simulations. The initial position of the satellites are directly extracted from the respective public filings, and continually updated during simulation using orbital mechanics principles assuming ideal elliptical orbits. It is assumed that all users can be connected to any altitude. See section VI-D for a discussion about this assumption.

To evaluate the proposed algorithms, user demand and distribution are simulated using real-world data sources. The geographic distribution of users is sampled from the World Population distribution [35], which provides an estimation of the global human population density. The probability of a user being located in a given area is directly proportional to the population in that cell. Notably, this user model focuses on residential customers, disregarding other user types, and assumes uniform service penetration worldwide. User demand is assumed to be 100 Mbps for all users. The number of users varies depending on the specific experiment, and the details will be provided for each analysis.

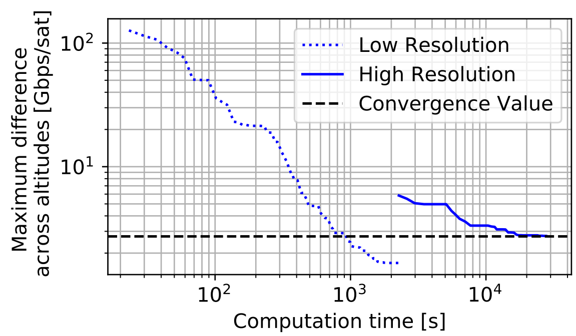
The mathematical solver employed in this study is Gurobi [36], version 9.1.2. Gurobi is a widely-used commercial solver that offers high-performance capabilities and is freely available for academic use. The hyperparameters used in the proposed method, including N_u, N_b, I_{thres}, and N_{conv}, along with the Gurobi parameters, are summarized in Table II. Furthermore, to evaluate the performance of the proposed method, a series of experiments were conducted, as

TABLE III
SUMMARY OF THE ANALYSIS EXECUTED IN THIS WORK

Experiment	Constellation	Number of users	Objective weight
Validation	SpaceX	20,000	$w_l = 1$, $w_d = 0$
Tradespace	SpaceX & Boeing	5,000	Variable
Performance	Boeing	Up to 100,000	$w_l = \hat{w}_l$, $w_d = \hat{w}_d$



(a) Maximum difference across altitudes vs. Iterations



(b) Maximum difference across altitudes vs. Computation time

Fig. 4. Evolution of the objective value with iterations and time. The dotted line represents time spent on the warm start. The y-axis for both figures and the x-axis on the lower figure are in logarithmic scale.

outlined in Table III.

B. Validation analysis

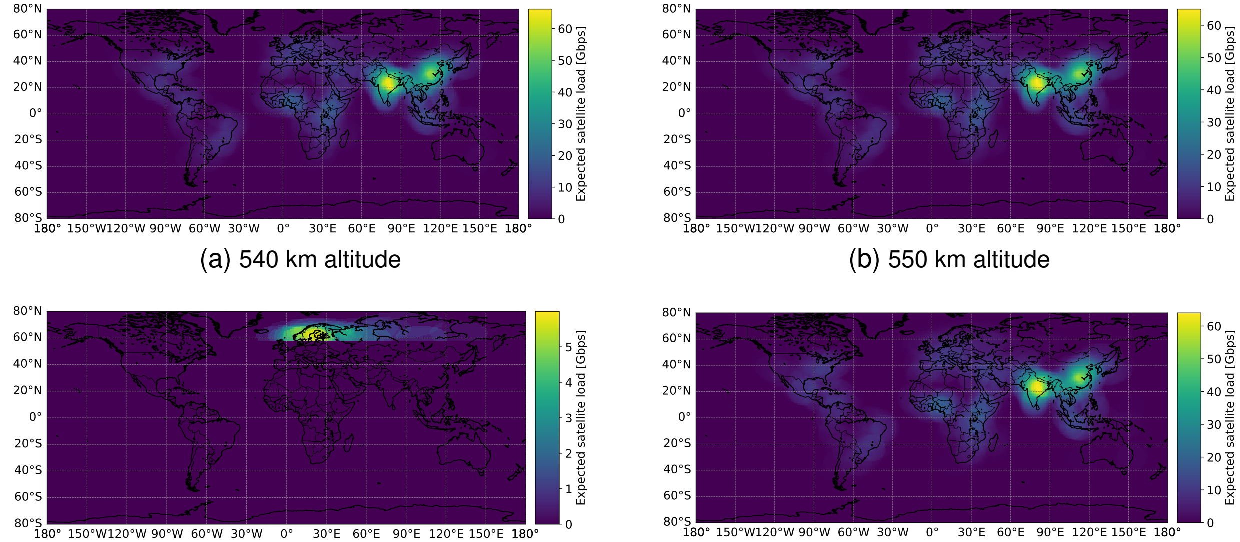
To ensure the effectiveness of the proposed methodology, it is essential to validate that the scalable algorithm functions as intended. This validation involves verifying two key aspects: 1) confirming that the algorithm achieves convergence when the solution stabilizes, and 2) verifying that the optimized solution, considering load balancing, achieves an equitable load distribution across different altitudes. To address these concerns, a validation experiment was conducted for the scalable approach, utilizing the SpaceX constellation design with 20,000 users. In this experiment, a weight of 1 was assigned to w_l (load balancing) and 0 to w_d (number of beams).

The results of the validation experiment are presented in Figures 4a and 4b, illustrating the evolution of the objective value over iterations and time, respectively. Several noteworthy observations can be derived from these figures. Firstly, the warm-start approach, utilizing a small set $\mathcal{P}$, facilitates the algorithm in reaching a reasonable solution rapidly. However, it tends to converge to local optima, hindering the retrieval of a near-optimal value. This behavior is evident during the transition between low and high resolutions, where the warm-start solution proves to be non-optimal in the high-resolution grid. On the other hand, the high-resolution algorithm successfully achieves convergence, yielding a solution that closely approaches optimality when compared to the warm-start approach. It is important to note that the high-resolution method demands a longer computational time (approximately 8 hours in this case) to accomplish this improved performance.

To verify that the algorithm behaves as expected, we computed the expected satellite load over the Earth at each latitude and longitude position. Figure 5 presents the results of this analysis, demonstrating an equitable load distribution across altitudes. In particular, we observe that satellites at different altitudes were expected to observe similar load values when passing over the same position, thanks to the utilization of the *impact* parameter in the formulation. Notably, even with varying numbers of satellites, altitudes 540 km, 550 km, and 570 km exhibit similar expected satellite load values. It is worth mentioning that the 560 km altitude consists solely of polar satellites unable to continuously serve latitudes between +/- 58°, which is appropriately accounted for in the beam selection process, ensuring that no users within those latitudes are associated with the 560 km altitude. These results validate the proposed scalable algorithm and provide evidence that it behaves as intended.

C. *Tradespace analysis*

To optimize the capacity of a satellite constellation, it is important to strike a balance between minimizing the number of beams and distributing the load across different altitudes, as discussed in section III-A. Although the aforementioned objectives were integrated into a single metric using weights, the trade-off between the different objectives remains to be studied. To that end, this work examines the performance of the scalable algorithm using different weight values and compares the results with two methods developed for single-altitude constellations: a grid-like [7] a user-centric approach [26]. Notably, when using these methods, it is necessary to assume the most restrictive footprint for all beams since there is no information available regarding which satellite will serve a given beam. In the case of [7], a redundancy factor of 2.5 ($\Delta = 2.5$) is assumed, and users are mapped to the closest beam. As emphasized in Section



(a) 540 km altitude

(b) 550 km altitude

(c) 560 km altitude (Polar)

(d) 570 km altitude

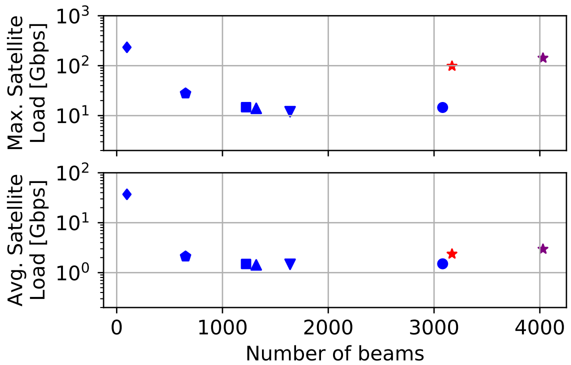
Fig. 5. Expected satellite load over each latitude and longitude at the specified altitude when using the proposed solution for the User Grouping problem for the SpaceX constellation with 20,000 users. This visualization does not take into account the capacity limits imposed by the satellite payload. Therefore, the expected load might be higher than the achievable load.

TABLE IV
SUMMARY OF THE WEIGHTS TESTED ON THE TRADESPACE ANALYSIS

w_l (Boeing)	w_l (SpaceX)	w_d
1		0
$10\hat{w}_l$	$1000\hat{w}_l$	$\hat{w}_d$
$\hat{w}_l$	$100\hat{w}_l$	$\hat{w}_d$
$0.1\hat{w}_l$	$10\hat{w}_l$	$\hat{w}_d$
$0.01\hat{w}_l$	$\hat{w}_l$	$\hat{w}_d$
0		1

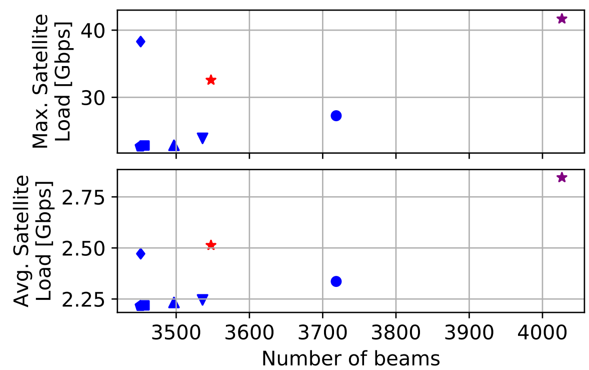
II, methods incorporating adaptive beam shaping are unsuitable for hybrid constellations due to the predominant influence of altitude on footprint size, with onboard control effects being relatively minor.

To enable a fair comparison of the scalable approach performance with different weight values, it is essential to normalize the objectives of the formulation presented in Equation 5. To that end, we first obtain the solution corresponding to $w_l = 1$ and $w_d = 0$, and compute the load balance γ_l and the number of beams N_l for this solution. We then define the two normalization values as: $\hat{w}_l = \frac{1}{\gamma_l}$ and $\hat{w}_d = \frac{1}{N_l}$. By utilizing these normalization values, the trade-off between different objectives can be analyzed using the weights outlined in Table IV. It should be noted that this



	Objective Weight	Num. beams	Max. Load [Gbps]	Avg. Load [Gbps]
●	w_d=0; w_l=1	3078	14.54	1.502
▼	w_d=$\hat{w}_d$; w_l=10$\hat{w}_l$	1639	11.96	1.468
▲	w_d=$\hat{w}_d$; w_l=1$\hat{w}_l$	1319	14.0	1.43
■	w_d=$\hat{w}_d$; w_l=0.1$\hat{w}_l$	1217	14.74	1.492
⬟	w_d=$\hat{w}_d$; w_l=0.01$\hat{w}_l$	650	28.47	2.155
♦	w_d=1; w_l=0	86	233.2	37.02
★	User-centric [26]	3166	100.1	2.415
★	Grid-like [7]	4026	145.3	3.022

(a) Boeing



	Objective Weight	Num. beams	Max. Load [Gbps]	Avg. Load [Gbps]
●	w_d=0; w_l=1	3718	27.26	2.336
▼	w_d=$\hat{w}_d$; w_l=1000$\hat{w}_l$	3536	23.89	2.245
▲	w_d=$\hat{w}_d$; w_l=100$\hat{w}_l$	3497	22.86	2.233
■	w_d=$\hat{w}_d$; w_l=10$\hat{w}_l$	3456	22.82	2.219
⬟	w_d=$\hat{w}_d$; w_l=1$\hat{w}_l$	3451	22.75	2.218
♦	w_d=1; w_l=0	3450	38.3	2.472
★	User-centric [26]	3547	32.64	2.515
★	Grid-like [7]	4026	41.76	2.847

(b) SpaceX

Fig. 6. Maximum and average satellite load for the User Grouping problem using different weight factors. The red and purple stars represent the solution obtained with the existing user-centric and grid-like methods, respectively. 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.

normalization strategy ensures unbiased results and enables the different weights to have similar orders of magnitude.

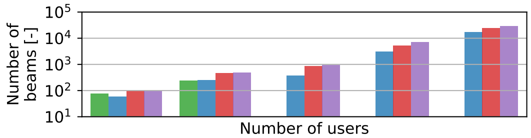
Before evaluating the performance of different weights, it is important to acknowledge that while reducing the number of beams has proven effective in maximizing system capacity for single-altitude constellations [13], this may not be directly applicable to hybrid constellations. Consequently, evaluating based solely on this metric could lead to biased outcomes. To provide a more comprehensive evaluation, we compute the real system capacity using two metrics: the average satellite load across used satellites (i.e., ignoring unused satellites) and the maximum satellite load across the constellation. A lower average and maximum satellite load indicate better resource utilization and a higher constellation capacity. To compute these metrics, after obtaining the User Grouping solution, the constellation is simulated by assigning each beam to the satellite with the highest elevation angle at the corresponding altitude. The average and maximum loads are then computed by aggregating the loads of all beams for each satellite.

Figure 6 presents the results for the maximum and average satellite loads for the Boeing and SpaceX constellations with 5,000 users using the scalable algorithm with multiple weight factors, as well as the user-centric and grid-like mechanisms presented in [7], [26], which focus on minimizing the number of beams. Several observations can be made from the results. Firstly, solely focusing on one objective does not maximize capacity: when only prioritizing load balancing ($w_d = 0, w_l = 1$, blue circle), the approach generates solutions with a large number of beams, indicating less effective use of spacecraft resources. Conversely, when solely minimizing the number of beams ($w_d = 1, w_l = 0$, blue diamond), the approach emphasizes grouping users as much as possible, resulting in an oversaturation of the highest altitude satellites. This effect becomes more pronounced as the constellation expands across various altitudes (e.g., Boeing), but is still noticeable even when the number of altitudes and altitude differences are small (e.g., SpaceX). Irrespective of the constellation, the best results to achieve both the lowest average and maximum satellite load are obtained when $w_l = \hat{w}_l$ and $w_d = \hat{w}_d$. Additionally, it can be observed that using single-altitude approaches leads to underutilization of the constellation. Even if it allows for a higher flexibility during operations, neglecting the flexibilities offered by hybrid constellations in the problem formulation results in a less effective distribution of beams across satellites.

D. Performance analysis

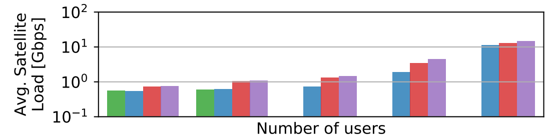
Finally, this study evaluates the scalability and performance of the proposed methods under various operational scenarios. Specifically, the optimal and scalable approaches are assessed using the Boeing constellation with different numbers of users: 100, 500, 1,000, 10,000, and 100,000. To facilitate comparison, the method developed for single-altitude constellations [26] is included. For this evaluation, the weight values for each objective are set to $w_l = \hat{w}_l$ and $w_d = \hat{w}_d$, as they have shown to strike a reasonable compromise for the average and maximum satellite load. The execution time for each test is limited to 100,000 seconds (approximately 28 hours).

Figure 7 presents the results of the performance evaluation for different numbers of users, including the number of beams, average and maximum satellite load, and total computation time. It is important to note that, consistent with the findings from the previous section, the proposed methods outperform the prior methods designed for the User Grouping problem in single-altitude constellations. Particularly, the scalable approach consistently achieves solutions with 30% to 57% fewer beams, 13% to 45% less average satellite load, and 77% to 91% less maximum load compared to the user-centric method. Furthermore, it achieves solutions with 39%



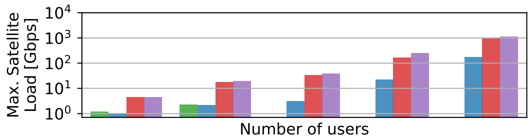
	100	500	1000	10000	100000
Hybrid - Optimal	76.0	243.0	-	-	-
Hybrid - Scalable	60.0	255.0	372.0	3091.0	17150.0
User-centric [26]	99.0	471.0	868.0	5265.0	24280.0
Grid-like [7]	99.0	495.0	950.0	7030.0	29620.0

(a) Number of beams



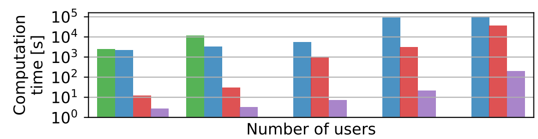
	100	500	1000	10000	100000
Hybrid - Optimal	0.5528	0.5995	-	-	-
Hybrid - Scalable	0.5346	0.6121	0.7291	1.89	11.13
User-centric [26]	0.7302	1.044	1.32	3.381	12.72
Grid-like [7]	0.7412	1.089	1.438	4.383	14.71

(b) Average satellite load



	100	500	1000	10000	100000
Hybrid - Optimal	1.185	2.2	-	-	-
Hybrid - Scalable	1.0	2.18	3.065	22.16	168.6
User-centric [26]	4.42	17.76	32.67	164.4	936.3
Grid-like [7]	4.435	18.94	38.56	253.0	1128.0

(c) Maximum satellite load



	100	500	1000	10000	100000
Hybrid - Optimal	2545.0	11820.0	-	-	-
Hybrid - Scalable	2155.0	3378.0	5531.0	89650.0	94700.0
User-centric [26]	11.92	30.34	994.9	3143.0	36060.0
Grid-like [7]	2.811	3.263	7.296	20.97	204.6

(d) Computation time

Fig. 7. Evolution of the number of beams, maximum satellite load, average satellite load and computation time with the number of users for the proposed method and existing user-centric and grid-like methods. The y-axis is in logarithmic scale. Executions are limited to 100,000s (around 28h). The optimal method does not provide a solution in less than the maximum allowed time for more than 500 users.

to 61% fewer beams, 24% to 57% less average satellite load, and 77% to 92% less maximum load compared to the grid-like method. This highlights the critical role of incorporating the flexibilities offered by hybrid constellations in optimizing system capacity.

Moreover, the proposed method demonstrates its capability to obtain solutions within a practical computational time, given that the User Grouping problem only needs to be computed once at the initiation of operations and whenever there are changes in the user distribution during operations. Importantly, in operational scenarios, the upcoming user distribution is often known well in advance, providing ample time for computation. Additionally, the algorithm can leverage the previous solution as a warm start, leading to a considerable reduction in computing time for subsequent iterations. This combination of features ensures the feasibility and efficiency of the method in real-world satellite communication systems.

It is worth noting that both the optimal and scalable approaches yield similar solutions in terms of the number of beams and average and maximum satellite load for low-dimensional scenarios (up to 500 users). However, the optimal method fails to scale beyond this point, underscoring the scalability advantage of the proposed scalable algorithm in fully exploiting the flexibilities of hybrid constellations. This enables the algorithm to achieve close-to-optimal solutions for the

User Grouping problem within a reasonable time.

VI. Discussion

This section discusses the main assumptions of this work, and provides a brief overview on how the formulation could be adapted to relax them.

A. Implementation in real operations

Beyond its heightened performance, the proposed formulation lends itself to implementation in the next generation of systems with minimal adjustments. The methodology operates effectively within constellations characterized by low eccentricities and equipped with programmable payloads, aligning with contemporary practices as highlighted in Section III-B. Additionally, although versatile enough to accommodate various beam shapes, the implementation does not mandate satellites to possess Beam Shaping capabilities.

Furthermore, the user-beam mapping is designed for long-term validity, obviating the need for frequent updates and reducing telemetry requirements. Modifications to the beam layout are infrequent, typically occurring in rare instances such as the introduction of a new user not covered by an existing beam or during a satellite contingency. In such cases, the iterative nature of the approach allows the prior plan to serve as a warm start, significantly reducing computation time.

Importantly, in contrast to prevailing standards, the proposed method allows for multiple beams to coexist on the same Earth segment. While this potential facilitates increased spectrum utilization under careful management, it introduces additional complexities in addressing Frequency Assignment. Nevertheless, prior works [37] have developed scalable methods adept at handling the intricacies of existing constellations. Finally, the formulation exhibits adaptability to heterogeneous operational requirements, as discussed in the subsequent sections.

B. Fixed vs. Mobile users

One of the main assumptions of the formulated problem is that the positions of users are fixed and known. While this assumption is valid for most broadband customers, such as residential and backhauling users, satellite networks have the potential to serve moving users whose positions change over time. Additionally, the uncertainty in user positions can pose challenges when addressing the User Grouping problem. Nevertheless, it should be noted that attempting to group multiple moving users into individual beams for prolonged periods might not be feasible or practical. Instead, assigning one beam per user is a more practical approach. Depending on the certainty of user positions, two mechanisms can be considered: 1) If the position is known, tracking techniques (e.g., [38]) can be utilized to dynamically adjust the position of the beam center over time, ensuring continuous coverage, and 2) If there is high uncertainty or a significant number of mobile users, a grid-like approach can be employed to ensure continuous coverage over the Earth. To incorporate these considerations and leverage the flexibility of hybrid orbits, the formulation could include estimations of the impact and demand for each beam ($\nu_{b,p,a}$ and d_b, respectively) to account for moving users. However, addressing the challenges and opportunities associated with these mechanisms is beyond the scope of this paper and is left as potential future work.

C. Using different footprints

While the results presented in this study showcase a fixed circular beam shape, the formulated problem accommodates arbitrary footprints, provided that the contour of the footprint can be computed and remains constant over time. This flexibility enables the consideration of various beam shapes that better align with the desired coverage and user distribution. However, it is worth mentioning that the current study focuses on the specific case of a fixed circular beam shape, and the exploration of different beam footprints and their implications on system performance and optimization is left for future investigations.

D. Mapping between users and altitudes

In the context of simulations, a fundamental assumption of this study pertains to the unrestricted assignment of users to any altitude. However, in practical scenarios, distinct altitudes may impose specific requirements on ground terminals owing to elevated propagation losses. Consequently, certain users may encounter limitations in connecting to all available altitudes. To address this consideration, the formulation can be readily adapted by constraining a user's connectivity options exclusively to beams associated with valid altitudes, thereby accommodating the inherent restrictions imposed by altitude-specific characteristics. The specific analysis on how these constraints affect the performance of the proposed approaches is beyond the scope of this work and left as a possible avenue of future research.

E. Beams without continuous coverage

This study assumes that users require continuous coverage for service provision. However, in reality, certain user types may not have the same level of coverage requirement. For instance, some users may only need intermittent connectivity, such as uploading data to an external server once a day for a limited duration, without the need for continuous connection. To accommodate such users, the formulation allows for the inclusion of beams that are not served continuously, with the condition that the duration of time during which a beam has at least one visible satellite exceeds the time required by the customers for service. In other words, users cannot be associated with beams whose coverage time is shorter than the agreed-upon service duration. This consideration expands the flexibility of the approach to cater to diverse user requirements. The specific treatment of intermittent users and the optimization of User Grouping for such cases are beyond the scope of this work and present opportunities for future research.

F. Alternative objective functions

In contrast to most literature in satellite communications, this study does not utilize power or spectrum usage as the primary capacity reference. This choice is motivated by two factors:

1) The complex nature of thousands of satellites with distinct resource pools makes it uncertain whether power and spectrum usage can adequately capture the intricacies of capacity allocation. Given the focus of this work on distributing beams and users across satellites, a more appropriate measure of capacity is to assess the utilization of different resource pools rather than evaluating the impact on each individual pool.
2) The formulation in this study does not assume any prior knowledge of frequency spectrum or power allocation for each beam, as it does not address the Frequency Assignment and Power Allocation problems explicitly. Therefore, the computation of spectrum and power usage is not feasible without this information. While existing literature offers approaches to solve these problems [39]–[41], addressing them falls beyond the scope of this paper and serves as potential future research.

Moreover, quality-based metrics such as latency or quality of service are not incorporated into the current formulation. This omission is justified under the assumption of homogeneous users, where the primary objective is to serve as many users as possible. Although the formulation allows for the inclusion of quality-based metrics by constraining the feasible altitudes for each user, exploring this aspect lies beyond the scope of this paper and is left as an avenue for future investigation.

VII. Conclusions

This paper introduces a novel methodology for addressing the User Grouping problem, specifically tailored to hybrid constellations, with a focus on enhancing capacity. First, the shortcomings of existing methods in accommodating multiple satellites at varying altitudes are highlighted. To bridge this gap, a single-objective formulation of the User Grouping problem is presented, simultaneously considering users, shapes, and beams. Two distinct methods are proposed to address the complexity associated with this formulation, catering to low and high dimensionality scenarios. The effectiveness of the proposed methods is assessed through three comprehensive analyses conducted on the SpaceX and Boeing constellation designs: convergence, tradespace, and performance evaluations. Beyond the scenarios studied in this work, the flexibility and applicability of the proposed methodology is discussed across heterogeneous operational requirements. Key conclusions derived from this work are as follows:

- Traditional methods for the User Grouping problem overlook the potential advantages offered by hybrid constellations, resulting in suboptimal satellite utilization within modern designs like SpaceX and Boeing constellations.
- In hybrid constellations, both the number of beams and the distribution of load across different altitudes play critical roles, in contrast to single-altitude designs.
- The methodology proposed in this work effectively harnesses the flexibilities intrinsic to hybrid constellations, achieving a favorable balance between the aforementioned objectives and maximizing satellite utilization.
- Particularly, the scalable approach consistently yields notable improvements, achieving a minimum reduction of 13% in average satellite load and 77% in maximum satellite load when compared to existing techniques across all examined scenarios, which raise up to 57% and 92% in particular instances, respectively. In low dimensionality settings, both the scalable and optimal approaches exhibit comparable performance.
- The scalable approach demonstrates its practical applicability by successfully addressing realistic operational scenarios encompassing up to 100,000 users, whereas the optimal approach proves viable only within feasible time limits for cases involving up to 500 users.

These findings collectively establish the efficacy of the proposed methodology in leveraging the advantages of hybrid constellations and optimizing the User Grouping problem, ultimately facilitating enhanced satellite utilization and capacity within modern satellite designs.

ACKNOWLEDGMENTS

This work was supported by SES. The authors would like to thank SES for their input to this paper and their financial support. The project that gave rise to these results also received the support of a fellowship from ”la Caixa” Foundation (ID 100010434). The fellowship code is LCF/BQ/EU21/11890113.

APPENDIX A: PROOF OF NP-COMPLETENESS

The NP-completeness of the User Grouping decision problem can be demonstrated by its combination of the set cover problem and a generalized multiway number partitioning problem, indicating NP-hardness. Additionally, the existence of a polynomial-time verifier confirms that the problem belongs to NP.

To establish NP-hardness, we consider three cases based on the objective weights: $\omega_l = 0$, $\omega_d = 0$, and $\omega_l \neq 0, \omega_d \neq 0$.

When $\omega_l = 0$, the problem aims to find the minimum set of beams y_b that collectively cover all users, with each beam able to cover only a subset of users. The coverage of user u by beam b is determined by the condition $p_u \in \Phi_b$. We define $\mathcal{U}_b$ as the set of users covered by beam b, and the objective 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 [42].

When $\omega_d = 0$, the problem aims to find a partition of users that minimizes the difference between altitudes. To prove NP-hardness, we make the following assumptions: $\mathcal{P}$ 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 ($d_u \in \mathbb{N}, \nu_{b,p,a} \in \mathbb{N}$), and the capacity of all altitudes are equal ($c_a = c \; \forall a \in \mathcal{A}$). Given these conditions, we seek 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 [43].

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, we provide a polynomial-time verifier that, given $x_{u,b}\ \forall u \in \mathcal{U}, b \in \mathcal{B}$, verifies if $\omega_l\gamma + \omega_d \sum_b y_b \leq k$ for a given $k \in \mathbb{R}_+$. The verifier efficiently resolves the auxiliary variables based on the formulation described in III-C, resulting in linear-time verification with respect to the input size. This proof conclusively establishes the User Grouping decision decision problem's membership in NP, demonstrating its NP-completeness.

REFERENCES

[1] 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.

[2] 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," Grand View Research, Tech. Rep., 2022, available at: https://www.researchandmarkets.com/reports/4700903/video-streaming-market-size-share-and-trends.

[3] N. S. Research, "Aeronautical Satcom Markets, 10th Edition," Northern Sky Research, Tech. Rep., 2022, available at: https://www.nsr.com/?research=aeronautical-satcom-markets-10th-edition.

[4] 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.

[5] B. Al Homssi, A. Al-Hourani, K. Wang, P. Conder, S. Kandeepan, J. Choi, B. Allen, and B. Moores, "Next generation mega satellite networks for access equality: Opportunities, challenges, and performance," *IEEE Communications Magazine*, vol. 60, no. 4, pp. 18–24, 2022.

[6] 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.

[7] 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.

[8] 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)*. IEEE, 2021, pp. 1345–1351.

[9] 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)*. IEEE, 2018, pp. 1042–1046.

[10] N. Pachler, E. F. Crawley, and B. G. Cameron, "A genetic algorithm for beam placement in high-throughput satellite constellations," in *IEEE Cognitive Communications for Aerospace Applications Workshop*, 2021.

[11] 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.

[12] D. Wu, C. Qin, Y. Cui, P. He, and R. Wang, "Multi-group multicast beamforming in leo satellite communications," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022, pp. 1564–1569.

[13] N. Pachler, J. J. G. Luis, E. F. Crawley, and B. G. Cameron, "A Unified Resource Allocation Framework and Impact Evaluation for NGSO satellite constellations," *Aerospace Science and Technology (Under review)*, 2024.

[14] J. A. Vásquez-Peralvo, J. Querol, F. Ortíz, J. L. G. Rios, E. Lagunas, V. M. Baeza, G. Fontanesi, L. M. Garcés-Socorrás, J. C. M. Duncan, and S. Chatzinotas, "Flexible beamforming for direct radiating arrays in satellite communications," *IEEE Access*, 2023.

[15] 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.

[16] H. Chaker, H. Chougrani, W. A. Martins, S. Chatzinotas, and J. Grotz, "Matching traffic demand in geo multibeam satellites: The joint use of dynamic beamforming and precoding under practical constraints," *IEEE Transactions on Broadcasting*, vol. 68, no. 4, pp. 819–833, 2022.

[17] 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.

[18] 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*. IEEE, 2020, pp. 1–6.

[19] 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.

[20] 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. 57 578–57 588, 2021.

[21] 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.

[22] Space Exploration Holdings, LLC, "SAT-MOD-20230215-00036," http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2023021500036, 2023, accessed: 2023-03-10.

[23] The Boeing Company, "SAT-MOD-20211104-00148," http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2021110400148, 2021, accessed: 2023-03-09.

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

[25] 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.

[26] N. Pachler, M. Guerster, I. del Portillo, E. F. Crawley, and B. G. Cameron, "Static beam placement and frequency plan algorithms for LEO constellations," *International Journal of Satellite Communications and Networking*, 2020.

[27] R. Williams, *The geometrical foundation of natural structure: A source book of design*. Dover Publications, 1979.

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

[29] 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 on Satellite Mega-Constellations in the 6G Era*, 2021.

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

[31] ——, "SAT-MOD-20181108-00083," http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2018110800083, 2018, accessed: 2020-12-09.

[32] ——, "SAT-MOD-20190830-00087," http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2019083000087, 2019, accessed: 2020-12-09.

[33] ——, "SAT-MOD-20200417-00037," http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATMOD2020041700037, 2020, accessed: 2020-12-09.

[34] The Boeing Company, "SAT-LOA-20170301-00028," http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATLOA2017030100028, 2017, accessed: 2023-03-09.

[35] 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, Palisades, NY, 2018.

[36] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2023. [Online]. Available: https://www.gurobi.com

[37] 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.

[38] 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*. Springer, 2017, pp. 415–422.

[39] O. Y. Kolawole, S. Vuppala, M. Sellathurai, and T. Ratnarajah, "On the performance of cognitive satellite-terrestrial networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 668–683, 2017.

[40] Z. Qin, X. Zhou, L. Zhang, Y. Gao, Y.-C. Liang, and G. Y. Li, "20 years of evolution from cognitive to intelligent communications," *IEEE transactions on cognitive communications and networking*, vol. 6, no. 1, pp. 6–20, 2019.

[41] O. Simeone, "A very brief introduction to machine learning with applications to communication systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 4, pp. 648–664, 2018.

[42] R. M. Karp, *Reducibility among combinatorial problems*. Springer, 2010.

[43] 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.

Biography Section

Nils Pachler is a PhD candidate in the System Architecture Group/Engineering Systems Lab of the AeroAstro department at MIT. He works on the application of general optimization algorithms to address complex real-world problems, such as Dynamic Resource Management in satellite communications. In 2019, he completed a double BS degree in Informatics Technology and Aerospace Engineering under the CFIS (Interdisciplinary Higher Education Centre) program from Universitat Politècnica de Catalunya. Nils carried out his Bachelor Thesis at MIT on dynamic allocation of resources in a multibeam satellite system. In 2021 he was awarded a La Caixa fellowship. In 2022, he received his Master's degree in Aerospace Engineering from MIT. His research interests include autonomous systems, optimization algorithms, and real-world decision-making.

Prof. Edward F. Crawley received an Sc.D. in Aerospace Structures from MIT in 1981. His early research interests centered on structural dynamics, aeroelasticity, and the development of actively controlled and intelligent structures. Recently, Dr. Crawley's research has focused on the domain of the architecture and design of complex systems. From 1996 to 2003 he served as the Department Head of Aeronautics and Astronautics at MIT, leading the strategic realignment of the department. Dr. Crawley is a Fellow of the AIAA and the Royal Aeronautical Society (UK), and is a member of three national academies of engineering. He is the author of numerous journal publications in the AIAA Journal, the ASME Journal, the Journal of Composite Materials, and Acta Astronautica. He received the NASA Public Service Medal. Recently, Prof Crawley was one of the ten members of the presidential committee led by Norman Augustine to study the future of human spaceflight in the US.

Dr. Bruce G. Cameron is the Director of the System Architecture Group at MIT. His research interests include technology strategy, system architecture, and the management of product platforms. Previously, Dr. Cameron ran the MIT Commonality study, a 30-firm investigation of platforming returns, which concluded that firms face systemic downward pressure on commonality, partially resulting from challenges capturing the costs of variety. Dr. Cameron has supervised over 50 graduate students, and has directed research projects for Amazon, BP, Sikorsky, Nokia, Caterpillar, AMGEN, Verizon, and NASA. Current research efforts include: Platform management in large R&D organizations, System architecture of complex systems, Switching costs and retention incentives in ride hailing firms, and Satellite mega-constellations in Lower Earth Orbit.