Towards a future space-based, highly scalable AI infrastructure system design

Blaise Agüera y Arcas[*,1], **Travis Beals**[*,1], **Maria Biggs**[*,1], **Jessica V. Bloom**[*,1], **Thomas Fischbacher**[*,1], **Konstantin Gromov**[*,1], **Urs Köster**[*,1], **Rishiraj Pravahan**[*,1] **and James Manyika**[1]
[*]Equal contributions, [1]Google

If AI is a foundational general-purpose technology, we should anticipate that demand for AI compute — and energy — will continue to grow. The Sun is by far the largest energy source in our solar system, and thus it warrants consideration how future AI infrastructure could most efficiently tap into that power. This work explores a scalable compute system for machine learning in space, using fleets of satellites equipped with solar arrays, inter-satellite links using free-space optics, and Google tensor processing unit (TPU) accelerator chips. To facilitate high-bandwidth, low-latency inter-satellite communication, the satellites would be flown in close proximity. We illustrate the basic approach to formation flight via a 81-satellite cluster of 1 km radius, and describe an approach for using high-precision ML-based models to control large-scale constellations. Trillium TPUs are radiation tested. They survive a total ionizing dose equivalent to a 5 year mission life without permanent failures, and are characterized for bit-flip errors. Launch costs are a critical part of overall system cost; a learning curve analysis suggests launch to low-Earth orbit (LEO) may reach $\lesssim$\$200/kg by the mid-2030s.

1. Introduction

1.1. Motivation for ML in space

The development of the Transformer model (1) and the subsequent rise of generative, multimodal AI have led to an explosion of demand for compute capacity. Although dramatic gains have been made in efficiency (e.g. Gemini query energy consumption was reduced 33× over a one year period (2)), AI-based products and services have grown even faster, leading to a rapid increase in data center energy demand. To that end, Google is invested in advancing new forms of power generation (e.g. (3–5)). However, given AI appears to be a foundational general-purpose technology (6) — akin to electricity or the steam engine — we should anticipate that its use will continue to broaden across all aspects of human endeavor from powering the economy, to helping to tackle some of humanity's greatest challenges (7–9), and it can be expected that AI computational needs will grow, as will the energy required to run it.

In this paper and the research "moonshot" it proposes, we look to the future and work back from that. The Sun is by far the largest source of power in our solar system: with an output of 3.86×10^{26} W, the Sun emits more than 100 trillion times humanity's total electricity production. At some point in the future, the best way to power AI will likely thus be to more directly tap into that enormous source of energy. Space-based solar power has long been proposed in various forms (10–12). It has many of the attractive qualities of terrestrial solar power, with the additional advantages that solar panels in certain orbits are exposed to nearly continuous sunshine, and receive up to 8× more solar energy per year than a panel located on Earth at mid-latitude (13). However, getting the generated power back to Earth has been a major challenge for such proposals.

Instead of transmitting power to Earth from space, we propose a future that includes space-based ML "data centers" consisting of many solar-powered satellites networked via free-space optical inter-satellite links. While there are a number of challenges that would need to be addressed to realize this "moonshot," in the long run it may be the most scalable solution, with the additional benefit of minimizing the impact on terrestrial resources such as land and water.

Corresponding author(s): jessicavbloom@google.com

1.2. System design overview

Working backward from an eventual future in which the majority of AI computation happens in space, we identify as an intermediate milestone showing that a space-based system could achieve performance roughly comparable to a terrestrial datacenter. This research initiative is focused on addressing several of the major ingredients required: power generation, high-bandwidth, low-latency communication between chips, radiation-tolerant compute, a thermal management system, and a data link to ground stations. We aim to be as mass-efficient as possible, maximizing compute per kilogram to minimize launch costs, while attending to practical considerations for satellite design, including launch vehicle compatibility, avoidance of space debris, and structural feasibility.

To meet these requirements, we propose working towards a future where we would host the Google tensor processing unit (TPU) accelerator chips on a constellation of solar-powered satellites, with size and number of TPUs per satellite determined by both economic and engineering considerations. We envision launching the satellites into dawn-dusk, sun-synchronous low-Earth orbit (LEO) to maximize power generation while minimizing latency of ground communications and launch cost. To enable ultra-high bandwidth, low-latency data transfer between satellites, they will fly close together and communicate via free-space optics inter-satellite links (FSO ISLs). An ML-based flight control model enables the satellites to maintain close flight proximity while avoiding collisions. Eventually, optical links will also be needed for high-bandwidth communication with the ground, but for a pilot project, radio suffices, avoiding the challenges of overcoming atmospheric interference. Maintaining a dawn-dusk orbit will increase latency to some ground locations, but is advantageous for maximizing power. Cooling would be achieved through a thermal system of heat pipes and radiators while operating at nominal temperatures.

Proposals exist for “monolithic” data centers in space where individual spacecraft significantly exceed the size of any current or planned launch vehicle (14; 15). While such design concepts reduce the need for high-performance inter-satellite links, they involve new challenges: such structures would have to be assembled in space by humans or robots; collision avoidance would be more cumbersome; and structural requirements would add mass and complexity. Our proposed approach would instead rely on arrays of smaller satellites. This more modular design would provide ample opportunity to scale to the terawatts of compute capacity that could fit within the dawn-dusk sun-synchronous low-earth orbital band.

To assess the viability of this concept, this work focuses on several key technological challenges: the required inter-satellite communication bandwidth, the dynamics and control of large, tightly-clustered satellite formations, the radiation tolerance of TPUs, and economic feasibility given expected future launch costs. Other significant challenges such as on-orbit reliability and repair, high-bandwidth ground communications, and thermal management are also discussed in this paper. Our ongoing research towards achieving this moonshot involves refining designs and reducing risks through further analysis, ground-based testing, and in-orbit prototype missions – similar to how Google has approached other ambitious research initiatives.

2. Results

2.1. Inter-satellite links

The networking requirements of large-scale terrestrial machine learning (ML) clusters far exceed the capabilities of current inter-satellite link (ISL) technology. Google’s TPU supercomputers, for instance, utilize a two-tiered networking architecture. A high-speed data center network provides pod-level connectivity (16), while a custom, low-latency optical Inter-Chip Interconnect (ICI) with throughputs on the order of hundreds of gigabits per second per chip facilitates the tightly-coupled communication required for large-scale training workloads (17). In contrast, commercially available optical ISLs offer data rates in the range of 1–100 Gbps.

Our analysis shows that the required aggregate bandwidth per link, on the order of 10 Tbps, is achievable by using Commercial Off-The-Shelf (COTS) Dense Wavelength Division Multiplexing (DWDM) transceiver technology, similar to that used in terrestrial data centers. The primary challenge is that such equipment requires significantly higher received optical power levels, on the order of hundreds of microwatts, compared to the ~1 μW levels typical for traditional long-range ISLs. These power levels can be achieved by drastically reducing the inter-satellite distance. Since for distances larger than the Fresnel limit, received power scales with the inverse square of the distance due to beam divergence, flying the satellites in close formation (hundreds of kilometers, or less) provides ample power to close the link budget for high bandwidth COTS transceivers, as illustrated in Figure 1.

As the distance becomes very short (e.g., ~10km for a 10 cm telescope), spatial multiplexing emerges as a new opportunity for further scaling. The smaller beam spot size at shorter distances allows multiple independent beams to be established between transceiver arrays on different satellites, each carrying a separate DWDM datastream. For example, as illustrated with three examples on the left of Figure 1, a single 10 cm total aperture can be populated with a 2×2 array of independent 5 cm optical systems at a link distance of 2.5 km, or a 4×4 array of 2.5 cm systems at 0.63 km, scaling the total bandwidth with the number of parallel links. This enables further scaling of bandwidth inversely with distance, analogous to achieving high aggregate bandwidth through parallel spatial streams in Google's Palomar Optical Circuit Switch (18).

A bench-scale demonstrator using off-the-shelf components successfully achieved 800 Gbps unidirectional (1.6 Tbps bidirectional) transmission across a short free-space path, validating the potential of this approach.

2.2. Orbital dynamics

Our proposed system design, at scale, will be significantly larger, and entail much closer formation flight (due to inter-satellite communications requirements), than any previous or current satellite constellations. We considered a set of constraints including: maintaining a stable set of nearest neighbors, minimizing latency and maximizing solar exposure and ISL performance. Based on these constraints, Figure 2 shows one possible configuration for an illustrative, planar 81-satellite constellation—-all placed in the orbital plane, at a mean cluster altitude of 650 km. The arrangement here is based on a square rather than hexagonal lattice, mostly to simplify its description. Cluster radius is R=1 km, with distance between next-nearest-neighbor satellites oscillating between (approximately) 100 and 200 m, as is shown in Fig. 3. We note that, of course, evolving constraints could change the optimal architecture for our constellation.

The constellation performs two shape-cycles per full orbit. Peripheral satellite S1 is at apoapsis at $T = 3T_{\text{orbit}}/12$, at altitude $h = a + R/2$ and at periapsis at $T = 9T_{\text{orbit}}/12$, at altitude $h = a - R/2$.

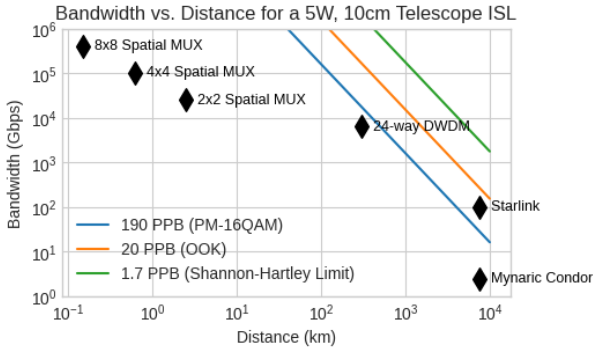

Figure 1 | Existing OISL device specifications vs. proposed design. Lines illustrate the $1/d^2$ relationship between distance and achievable bandwidth for three modulation schemes with different photons-per-bit (PPB) requirements. Commercial systems operate at long ranges, while our proposed system targets much shorter ranges to achieve higher data rates. 24-way dense wavelength-division multiplexing (DWDM) can be achieved up to about 300km distance with a 10cm aperture size. Fitting 2×2, 4×4, and 8×8 spatially multiplexed beams into the same total aperture requires distances of 2.5km, 0.63km and 0.15km (limited by crosstalk rather than power). Modulation schemes shown: Quadrature-Amplitude modulation with 16 symbols (PM-16QAM), on-off keying (OOK), and the Shannon-Hartley limit of channel capacity.

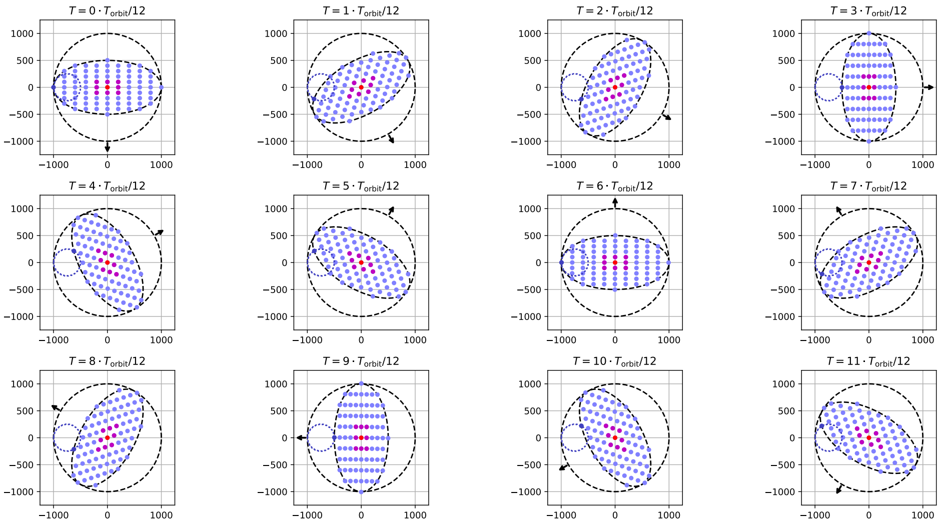

Figure 2 | Evolution of a free-fall (i.e. “no thrust”) constellation subject to Earth’s gravitational attraction plus J2-term (due to Earth’s oblateness) over the course of one orbit, shown at time intervals of 1/12 of a full orbit in a non-rotating coordinate system. Positions are relative to the central reference satellite S0 (red). Horizontal axis is aligned with the negative in-track direction of S0 at t=0, vertical direction correspondingly is “towards zenith at t=0.” Short arrows indicate the “towards center of Earth” direction. Magenta: nearest neighbors (8-neighborhood) of the central satellite S0. Dark blue: the “maximally-distant in in-flight direction at $t = 0$” satellite S1. Dark blue dashed: S1’s cluster-center-relative positions over the course of one orbit. All distances in meters.

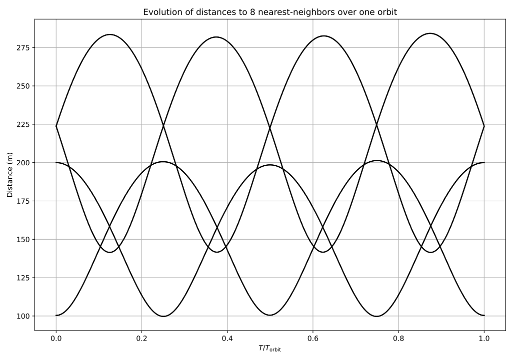

Figure 3 | Evolution of the distance between central "reference" satellite S0 and its (direct and diagonal) nearest neighbors over the course of one orbit under the combined effect of Newtonian Gravity and Earth’s J2-term.

Since the diameter of S1’s orbital ellipse equals that of S0’s, their orbital periods are identical; likewise for other satellites. (Perturbations due to the J2 term and other effects will require slight amendments to this leading-order-effect statement.)

While the constellation remains bounded inside a sphere of radius R, its shape goes through two full cycles over the course of one orbit. At any point in time, all satellites fit into a rotating “±R prograde, ±R/2 in altitude” ellipse. The interior of this rotating ellipse does not perform a rigid rotation: During the course of one orbit, different satellites are closest to the endpoints of its semi-major/minor axes, as one readily observes by following the path of S1.

If satellite motion were perfectly Keplerian (i.e. if Earth’s gravitational field were that of a point mass, and effects such as solar and lunar tides, atmospheric friction, radiation pressure, etc. were absent), this “free fall” constellation would reproduce itself perfectly after a full orbit, at zero delta-v requirement. If maintaining a planar constellation is undesirable, such as for establishing inter-satellite links or to address passive safety concerns, in-plane motion can be superimposed with per-satellite oscillatory out-of-plane motion (one oscillation per orbit). The 2:1 axis ratio of the in-plane bounding ellipsoid is a consequence of orbital dynamics. For a given minimal inter-satellite distance, this design approach leads to the number of satellites scaling quadratically with cluster radius, $N \sim R^2$. Improving the ratio of sunlight capture area over cluster cross-section beyond quadratic scaling would in general, for given minimal inter-satellite distance, require non-Keplerian formation flight approaches—such as electromagnetic formation flight (19)—and would have to pay attention to occlusion of outgoing "rejected heat" IR radiation between satellites.

Taking the leading "oblateness"-related J2-correction to Earth's gravitational field into account, which here would be exploited to keep satellites in a sun-synchronous orbit (i.e. make the orbital plane rotate once per year), the cluster's shape would get deformed slightly over the course of an orbit. This (predictable) drift can be compensated for via a small adjustment to cluster shape. A simplistic numerical calculation establishes that, for an example cluster as described, adjusting the axis-ratio to 2:1.0037 can reduce J2-drift to <3 m/s/year per km of maximal distance from reference orbit. A more in-depth analysis of differential accelerations (e.g. (20)) suggests that formation flight should be feasible with only modest delta-v requirements beyond what would be needed for precise station-keeping of a single satellite.

2.3. TPU radiation testing

Commercial-Off-The-Shelf (COTS) hardware has seen increasing use for space missions (21), such as the Mars Ingenuity helicopter (22). However, high-performance ML accelerators, which are characterized by a cutting-edge process node, large die size, and high floating-point operations per second (FLOPS) capability, represent a new frontier for COTS hardware in space. To address the question of their viability, Google's V6e Trillium Cloud TPU with its associated AMD host server were tested in a 67 MeV proton beam to simulate the operating conditions of sun-synchronous LEO. This work presents the first published radiation-testing results for such a device. For the target sun-synchronous LEO with significant shielding (e.g. 10 mm Al equivalent), the radiation environment is primarily composed of penetrating protons and Galactic Cosmic Rays (GCRs) (23), resulting in an estimated dose of ~150 rad(Si)/year (24). Radiation mainly causes: 1) Total Ionizing Dose (TID) effects, the cumulative build-up of charge in insulating layers leading to device degradation (25), and 2) Single Event Effects (SEEs), which are instantaneous faults caused by a single energetic particle strike generating a dense track of electron-hole pairs (26). Sensitivity depends on the process node and cell libraries used, and has historically decreased, but this trend does not hold beyond the 5nm node (26). For TPUs to be useful for machine learning in space, they must survive the expected 5 year lifetime TID of ~750 rad(Si), and SEEs must be at levels manageable through techniques like memory error correcting codes (ECC). Undetected SEEs are a particular concern for model training, as undetected bit-flips can lead to Silent Data Corruption (SDC). SDC may occur in any operating environment; SEE-induced SDCs are not the only cause (27). The covert nature of these errors requires careful calibration, detection and correction techniques.

Of the different elements of the Trillium TPU, the High Bandwidth Memory (HBM) subsystems exhibited the most sensitivity to TID. In our tests, HBM-specific stress tests began to show irregularities after a cumulative dose of 2 krad(Si), which is almost 3× our minimum requirement for TID survival of 750 rad(Si). All other tests, including end-to-end ML workloads and compute-centric tests, continued to operate correctly up to the maximum tested cumulative dose of 15 krad(Si) on a single chip. No hard failures were attributable to TID up to this level.

HBM was also the most SEE-sensitive component, primarily manifesting as uncorrectable ECC errors (UECCs). For typical inference workloads, the observed HBM UECC rate of approximately one event/50 rad, at an estimated orbital dose of 150 rad(Si)/year, translates to a very low failure probability, on the order of 1 per 10 million inferences. While this error rate is likely acceptable for inference, the impact of SEEs on training jobs, and the efficacy of system-level mitigations, requires further study. Radiation testing of the host system, with a focus on Single Event Functional Interrupts (SEFIs), e.g. system crashes, suggests a rate of one event per 450 rad(Si) for CPU and 400 rad(Si) for RAM.

2.4. Launch cost analysis

High launch costs have historically stymied efforts to harvest solar power in space on large scales (10). The below does not constitute a full economic analysis, but gives proof points demonstrating that there is a feasible path for launch costs to drop sufficiently to no longer be prohibitive. For context, economic feasibility analyses of space-based solar power for Earth use (the most analogous use case to space-based compute) cite \$500/kg to Geostationary Transfer Orbit (GTO) as a viability threshold (10; 11). SpaceX's Falcon Heavy payload capacity to GTO vs LEO is ~40%, so launch costs of \$500/kg to GTO are equivalent to ~\$200/kg to LEO. Further, \$200/kg is often cited, by SpaceX and others, as a threshold beyond which launch could cease to be the limiting cost factor for ambitious programmes (28–30).

We compared the kg/kW solar power (bus and payload) launched ratios for Starlink v2, Starlink v1, OneWeb & Iridium satellites and found that, if LEO launch prices drop to \$200/kg, we can project \$/kW/y launched to LEO ("launched power price") could be ~ \$810/kW/y for a Starlink v2-type constellation (amortized over satellite lifetime), or ~\$810–7,500/kW/y if we include a broader range of satellites with different mass/power ratios to serve varied use cases. For comparison, current power spend for terrestrial data centers in the US is reported to be ~\$570–3,000/kW/y (depending on regional variation in power price and operator power usage effectiveness, or PUE (31–33)). Thus, if launch costs to LEO reach \$200/kg, then the cost of launch amortized over spacecraft lifetime could be roughly comparable to data center energy costs, on a per-kW basis.

SpaceX launch pricing data and mass launched from Falcon 1 to Falcon Heavy [Fig. 4] yields a ~20% learning rate, meaning the price per kg falls by ~20% for every doubling of cumulative mass launched (over all vehicle classes). If the learning rate is sustained—which would require ~180 Starship launches/year—launch prices could fall to <\$200/kg by ~2035. While this would be a substantial achievement for SpaceX (particularly given the technological discontinuity inherent in switching to Starship), it is still far below stated launch targets for Starship. Even if this launch rate is reduced by ~ 70%, prices could drop to \$300/kg in the same timeframe, which would still have a substantial impact on feasibility of large-scale constellations. Further, there is precedent for a sustained ~20% learning rate over multiple decades in other advanced industries leveraging mass-production (notably, solar panels) (34). Given the long lead times required to reach scale for this type of ambitious project, it's strategically beneficial to commence work on early milestones in anticipation of projected price declines.

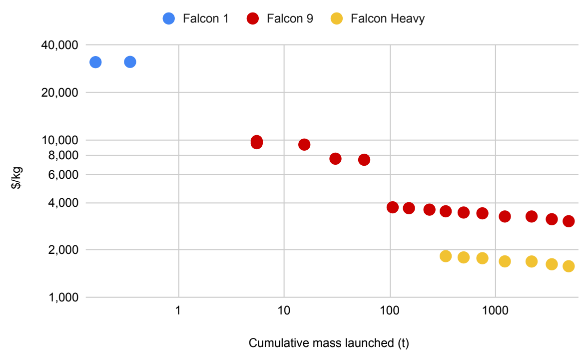

Figure 4 | SpaceX payload mass launched by lowest achieved price, inflation-adjusted, since the first successful Falcon 1 launch, for progressive rocket categories. Note major price discontinuities at the introduction of Falcon 9 and Falcon Heavy (35–37).

Alternatively, our analysis of Starship 4 public specifications and data suggests that SpaceX launch costs to LEO may drop to $\lesssim$ \$60/kg (10× component reuse); if SpaceX's 100× component reuse target were achieved, costs could reach

$\lesssim$ \$15/kg. Assuming 10× reuse and even the highest estimates of current SpaceX margins (up to 75% (38)), launch price to customers would drop to <\$250/kg (although margins are currently supported by SpaceX's near-monopoly and hence likely to decrease with the anticipated entry of competitors such as Blue Origin (39)). Realizing these projected launch costs is of course dependent on SpaceX and other vendors achieving high rates of reuse with large, cost-effective launch vehicles such as Starship.

3. Discussion

The results presented here are a first milestone towards scalable space-based AI; subsequent milestones involve testing aspects of the system in space (other Google/Alphabet research initiatives, such as Waymo and quantum computing, use similar milestone-based approaches). These initial results are an encouraging first step in assessing the feasibility of space-based ML compute at scale, addressing fundamental challenges in inter-satellite communication, orbital formation control, radiation hardness, and launch economics. Future space-based experimental milestones should also involve solutions for thermal management, high-bandwidth ground communications, and on-orbit reliability and repair strategies. In 4.4 we explore further work on orbital dynamics modeling.

Effective thermal management is a critical optimization challenge for power-dense TPUs operating in a vacuum. Advanced thermal interface materials and heat transport mechanisms, preferably passive to maximize reliability, are essential to efficiently move large heat loads from the chips to dedicated radiator surfaces.

Managing potential failures and (relatedly) increasing reliability of compute in space will be critical. Currently, failed TPUs are manually replaced by technicians, which is relatively simple and low-cost on Earth, but obviously impracticable in space. The simplest solution is redundant provisioning. There are also promising research programs aiming to increase system tolerance to faults affecting networking, e.g. via reduced communication (40).

Developing robust optical satellite-ground communications will also be critical for scaled operation, but will necessitate overcoming challenges including atmospheric turbulence, high-speed relative motion errors, and precision beam tracking. This is an active field of development (41; 42) with the current leader, NASA's TeraByte Infrared Delivery (TBIRD) mission, demonstrating 200Gbps ground-LEO communications in 2023 (43).

Looking further ahead, while our proposed constellation design is naturally suited for growing clusters in space, unlocking the full potential of compute in orbit will likely require new design approaches for individual satellites. While we anticipate launch costs continuing to decrease as the industry scales, the floor on fuel price means that there will always be an incentive to minimize mass. Our system design work to this point assumes a relatively conventional, discrete compute payload, satellite bus, thermal radiator, and solar panel design. However, as has been seen in other industries (such as smartphones), massively-scaled production motivates highly integrated designs (such as the system-on-chip, or SoC). Eventually, scaled space-based computing would similarly involve an integrated compute, radiator, and power design based on next-generation architectures, such as computational substrates based on neural cellular automata (44). The TPU-based satellite cluster described here and the preliminary results we describe are the beginning of unlocking that potential. Realizing the full scope of this ambitious vision will require sustained research, iterative refinement of our design, and the achievement of several critical future milestones.

4. Methods

4.1. Orbital dynamics analysis

In order to establish potential feasibility of a given formation flight proposal, it is generally useful to start from some analytically tractable simplifying model, such as the Hill-Clohessy-Wiltshire equations (45), that describe orbital motion of a satellite relative to a circular reference orbit in a Keplerian approximation, to leading order in relative positions and velocities - or more refined generalizations such as the Tschauner-Hempel equations (for arbitrary eccentricity) (46) or Vinti theory (47) (taking the J2 "earth oblateness" contribution into account). If a promising approach can be identified based on such analytic methods, it makes sense to then explore the nature and magnitude of corrections caused by further physical effects that are understood but more difficult to model analytically, in order to see if they are compatible with reasonable mission delta-v budgets. Here, the two main approaches are perturbation theory and numerics. They can be used independently, such as to validate one another's predictions, and often also in conjunction. This leads to advanced numerical methods for formation flight planning.

An overview over the relative importance of non-Keplerian contributions to satellite acceleration can be found in textbooks such as (48). For formation flight, the most important question is how various physical effects affect different satellites in a close-proximity constellation differently. At the envisioned altitude, the by far most important such effect is expected due to the J2-term of the geopotential, and potentially differential atmospheric drag. Effects such as lunar tides will be strongly suppressed by very small factors such as $r_{\text{cluster}}/d_{\text{moon}}$.

For this feasibility study, we determined promising-looking formation flight patterns by starting from the basic Hill-Clohessy-Wiltshire approximation, then mostly handling the adjustments needed to apply them to a sun-synchronous cluster orbit perturbatively, and obtaining results via numerical simulation. Care needs to be taken to retain good control of numerical accuracy. With a model that only involves accelerations with numerically benign behavior, computing orbits to centimeter accuracy vs. orbital diameters of order-of-magnitude 10^7 meters requires results to be correct to at least 9 decimal digits. While the discrepancy between such numerical modeling and real world physics may well be substantially larger than this accuracy target, having good numerical accuracy for models with simplified physics matters for objectives such as estimating the magnitude of some physical effect via a fully-numerical rather than a hybrid numerical-plus-perturbative approach.

Fast and performant (such as GPU-based) floating point arithmetics generally only supports up to `binary64` floating point representations, which give us just short of 16 decimal digits of precision. This strongly suggests use of a high order ODE integration scheme. For this exploration, we use the eighth-order Runge-Kutta `DOP853` method provided by SciPy's `scipy.integrate.solve_ivp` function.

4.2. Inter-satellite link analysis

The feasibility of high-bandwidth, short-range ISLs was assessed through link budget analysis. The achievable data rate of a photon-limited optical link is directly proportional to the received signal power, assuming a constant number of photons-per-bit (PPB) required for a given modulation scheme. In the far-field, it received power scales inversely with the square of the distance ($P_r \propto 1/d^2$) due to beam divergence, for a fixed transmitter power and aperture size. This relationship, and the resulting impact on maximum bandwidth for different modulation schemes, is plotted in Figure 1.

A survey of currently available and announced commercial OISL technologies reveals they are typically designed for link distances of thousands of kilometers, offering maximum data rates ranging from 1 Gbps to 100 Gbps. For example, Starlink's system operates at ~100 Gbps over distances up to ~7500 km, limited by line of sight distances in LEO. These systems (shown in the lower right of Figure 1) do not meet the multi-Terabit per second requirements for tightly coupled ML clusters.

Our approach leverages the enhanced link budget at short distances to employ multi-channel DWDM systems using high-spectral-efficiency coherent transceivers. These are commercially available, e.g. 400G transceivers using PM 16-QAM modulation. Deploying these on a 100GHz ITU frequency grid, a single aperture could support 24 channels (half of the C-band), yielding 9.6 Tbps bidirectional bandwidth. A tighter 75 GHz grid spacing could potentially support 12.8 Tbps per aperture. Such transceivers typically require a received power on the order of -18 dBm per channel, totaling approximately 0.38 mW for a 24-channel system.

In the far field, the received signal power (P_R) is estimated using the Friis transmission formula for free-space optics:

$$P_R = P_T \cdot G_T \cdot G_R \cdot \frac{\lambda}{4\pi d^2} \cdot L_{\text{other}} \tag{1}$$

Where P_T is the transmitted power (assumed to be 5W from a commercial EDFA), G_t and G_r are the transmitter and receiver antenna gain (both 105.1dB), λ is the wavelength (1.55 µm), d is the inter-satellite distance, L_{other} captures other losses (3dB). The antenna gains correspond to a ~10 cm diameter telescope. The beam divergence angle (θ) for a diffraction-limited aperture at 1.55 µm is approximately $\theta \approx 1.22\lambda/D \approx 1.22 \times (1.55 \times 10^{-6}\,\text{m})/0.1\,\text{m} \approx 18.9\,\mu\text{rad}$. To illustrate the power constraints of existing long-range systems, consider a typical LEO-LEO link distance of 5,000 km. At this distance, the beam spot diameter is at least 95 meters, and the best case received power 1.6 µW.

For the short-range, high-bandwidth links central to our spatially multiplexed design, a near-field model provides a useful approximation. In a symmetric confocal system where the beam waist is located midway between the two transceivers, the link distance (L) for a given beam radius (a) at the optics is given by: $L = 2\pi a^2/\lambda$. For example, for a 10 cm diameter beam at the optic (a = 5 cm), and a wavelength of 1.55 µm, the link distance for this system is approximately 10 km under best-case conditions (ignoring beam quality, aperture truncation and pointing jitter). This near-field analysis also shows that at kilometer-scale distances, minimum required aperture size for a single link scales with the square root of the distance. Consequently, more independent links can be packaged into a fixed total area as the distance decreases, causing the total bandwidth available via spatial multiplexing to scale inversely with distance.

The PPB requirements for different schemes shown as lines in Figure 1, illustrating the trade-off between distance and maximum achievable bandwidth, were estimated as follows: On-Off Keying (OOK): ~20 PPB, (PM 16-QAM): ~190 PPB, this coherent schemes requires higher Signal-to-Noise Ratios. This compares to the Shannon-Hartley Limit: ~2.31 PPB (Calculated in the infinite bandwidth, shot-noise limit as $2 \times \ln(2)/0.6$, accounting for a 0.6 quantum efficiency (49)). We assumed the use of standard C-band and L-band wavelengths, consistent with mature DWDM technology.

4.3. TPU Radiation Testing Procedure

Facility and Beam: Testing was conducted at the UC Davis Crocker Nuclear Laboratory, utilizing their 76-inch cyclotron to produce a 67 MeV proton beam. A large 8 cm diameter aperture was used to ensure uniform irradiation of the entire TPU package, including the logic die and HBM stacks. Beam intensities ranged from 2 pA (~2 rad/min) to 1 nA (1 krad/min).

Test Rig and Beam Path: A standard TPU system was tested with heatsinks installed to permit full-power operation. To avoid beam obstruction by the primary topside heatsink, chips were irradiated from the underside. The proton beam traversed the chassis (~1mm Al), the printed circuit board (PCB), and a secondary heatsink assembly before reaching the silicon die, which is directly bonded to the PCB. This secondary heatsink consists of either approximately 3mm of copper or 10mm of aluminum.

Beam Energy Considerations: The range of 67 MeV protons, calculated using the NIST PSTAR database (50), is approximately 18 mm in aluminum and 6.5 mm in copper. Therefore, while most of the proton flux will reach the die (confirmed by imaging through the heatsink assembly with gafchromic film (51)), the energy spectrum incident is not monoenergetic 67 MeV but is lower and spread out due to energy loss and straggling within the heatsink material. Additionally, nuclear interactions within the heatsink materials will generate secondary particles (neutrons, gammas, etc.), contributing to the radiation environment experienced by the device under test.

Software Suite: Subsystems were stressed with a diverse software suite to assess different vulnerabilities. This included memory stress tests targeting the High Bandwidth Memory (HBM) with various read/write patterns, compute-centric tests focusing on the core tensor computation units with minimal HBM traffic, and an end-to-end ML workload running a transformer model to simulate realistic operational stress. The memory tests were configured to log errors and continue, while the compute-centric and end-to-end tests were set to halt on the first significant error to establish dose-to-failure thresholds.

Data analysis: To estimate proton-induced SEE cross-sections, we used the beam characteristics provided by the facility and dose measurements. A dose of 1 rad corresponds to a proton fluence of approximately 9.6×10^6 protons/cm^2. The per-chip cross-section (σ) was calculated as: $\sigma \approx 1.04 \times 10^{-7}/D$ cm^2/chip where D is the dose per event in rad.

HBM was the most SEE-sensitive component, primarily manifesting as uncorrectable ECC errors (UECCs), causing test halts. The characteristic dose for HBM UECC was approximately 44 rad per event (averaged over 203 events). This corresponds to a cross-section of approximately 2×10^{-9}cm^2/chip. Core logic and on-chip SRAM were more robust, with SEE-induced failures in these systems occurring at approximately 150 rad per event, yielding an approximate cross-section of 7×10^{-10}cm^2/chip. System-level crashes, indicative of Single Event Functional Interrupts (SEFIs), were observed on average once per 5 krad of dose per chip. This translates to a SEFI cross-section of approximately 2×10^{-11}cm^2/chip.

It is important to note that HBM correctable error (CECC) counts were not reliably available for this analysis, as single-bit CECCs are only reported by the HBM firmware above a non-configurable, vendor-specific threshold. Thus, a precise in-test CECC to UECC ratio for HBM could not be determined. Instances of data mismatches occurring without a corresponding UECC flag were observed, suggesting Silent Data Corruption (SDC) at a rate of approximately one per 107 rad in one specific test run (averaged over 84 events). This suggests a cross-section of approximately 8.3×10^{-10}cm^2 per HBM assembly for this particular failure mode, though more data is needed to solidify this estimate.

4.4. Launch cost analysis

We projected launch prices via two methods: learning curve projection and analysis of planned Starship 4 specifications and reuse targets. While there is inherent uncertainty (e.g. in future regulatory obstruction, competitive dynamics from new entrants into the launch market and unforeseen technical challenges), both methods support our conclusion that reaching customer prices of $\lesssim$\$200/kg by mid 2030s is plausible under reasonable assumptions for reuse and cumulative mass launched during the time. We stress that the following is not intended as a comprehensive economic feasibility study, but rather a high-level evaluation of the potential for launch costs, specifically, to affect the viability of our proposal.

Sustaining high learning rates over time is obviously challenging, but there are multiple precedents across other advanced industries, e.g. shipbuilding and aerospace (56). In addition, solar panels provide a particularly striking, canonical example of a sustained ~20% learning rate for over 40 years (34).

Sustaining high growth in launched mass rates (and attendant price decline) is unlikely without large-scale, non-SpaceX commercial constellations driving demand (57; 58). Scaling ML in space would be one such use case, providing consistent demand for Starship (or Starship-class) launches.

Our learning curve analysis was based on publicly available historical SpaceX data. When Falcon 1 launched, SpaceX launch prices were $\gtrsim$\$30,000/kg (inflation adjusted) (59), dropping to ~\$1800/kg for Falcon Heavy over the course of <100 successful launches (60), or ~400t cumulative mass launched [Fig. 4]. The introduction of Falcon Heavy yielded a precipitous price decrease compared to Falcon 9, driven by improved economics from a heavier launch. Note that learning curve estimates are highly sensitive to input assumptions (e.g. source for price data, or whether the chosen starting point is the first successful Falcon 1 launch, the Falcon 9 or even the reusable Falcon 9 configuration), but a variety of choices

Table 1 | Launched power prices for a range of LEO satellites

Satellite	Mass (kg)	Power (kW)	Lifespan (y)	Launched power at $3,600/kg ($/kW/y)	Launched power at $200/kg ($/kW/y)
Starlink v2 mini [opt.]	575	28 [est.]	5	$14,700	$810
Starlink v1	260	7 [est.]	5	$26,600	$1,470
OneWeb	150 (52)	0.8 (53)	5	$135,800	$7,500
Iridium	860 (54)	2 (55)	12.5	$124,600	$6,900

all yield results ~18–24%.

Maintaining SpaceX's ~20% learning rate (based on launched mass) through new generations of launch vehicles, reaching <$200/kg by ~2035 would require launching $\sim 370,000$t additional cumulative mass, equivalent to ~1800 Starship launches (assuming 200t capacity (61)). This is an ambitious target, but the required ~180 Starship launches/year (on average, although there will almost certainly be a ramp up period in reality) would fall well below stated launch rate targets (62). Conversely, reaching this target is unlikely without Starship-like launch vehicles, due to Falcon payload volume limitations.

As an alternative method, we calculated costs (to SpaceX) for Starship 4, leveraging recently released reuse targets, as well as plans for payload size, Raptor engine count and rocket dimensions (63). Other inputs, e.g. refurbishment costs as a fraction of vehicle costs (~1% (64)) and failure rates, were extrapolated from existing Falcon 9 data (38; 64; 65). Without assuming any reduction in fuel cost, component reuse drives down projected SpaceX launch costs from ~$460/kg (no reuse) to <~$15/kg (100x reuse of all components). Increasing refurbishment costs to 15%, as a sensitivity analysis, yields costs of $38/kg (100x reuse of all components). Eventually, fuel costs (LOX and liquid methane for Starship), estimated at ~$8/kg (66; 67), will provide a floor on costs. We note that this analysis was based on relatively conservative assumptions, e.g. not including claimed efficiency improvements for Raptor 4 vs Raptor 3 engines (68). We also acknowledge the potential pitfalls of using Falcon 9 data to project trends for Starship, but the intent is to provide an indicative projection based on public data points, which we will update as more information becomes available.

The primary ingredients to run compute are power, infrastructure and chips. To compare annualized cost per unit of power for space and terrestrial data centers, we examine the launch costs for several satellite types, in terms of $/kW launched to LEO ("launched power price"). A strict like-for-like comparison of satellite and terrestrial costs would require more precise costs for satellite design, but we can make an approximate comparison between data center power costs and the launched power price. We do not include infrastructure or building costs, since this is distinct from power, or chip costs, since these will arise for both space and terrestrial configurations.

For the following analysis, we use current launch price $3,600/kg, based on Falcon 9 (reusable configuration), since that is used for the Starlink constellation, (69; 70) and "potential" price $200/kg. Note that we based this analysis purely on SpaceX data since they are by far the leading launch provider. Other incumbent and new entrants into the market (e.g. RocketLab and Blue Origin) are likely only to hasten price declines via competition and erosion of high SpaceX margins (38).

Starlink is the best-in-class proxy, as the largest constellation in orbit. The new v2 mini satellites weigh 575kg (optimized design (71)). Exact specifications have not been publicly released by SpaceX, but photometric and other analyses yield solar panel area ~105m^2 (72; 73). Assuming 22% solar panel efficiency, 1.361kW/m^2 solar insolation and 90% packing area for square cells (74–76), we obtain ~28kW/satellite. Amortized over a 5 year lifespan (77), we obtain a launched power price of \$14,700/kW/y (current prices), falling to \$810/kW/y if launch drops to \$200/kg. We performed similar calculations for Starlink 1, OneWeb, Iridium NEXT and Apex satellites (see [Table 1]), yielding a launched power price range (for launch prices \$200/kg to LEO) of \$810–7,500/kW/y. Note that this extremely large range is at least partially driven by differences in use case and resulting optimization priorities across satellite programs, which in turn affects design decisions and kg/kW ratio.

By comparison, terrestrial power costs for ML-capable data centers in the US are reported to be \$0.06–0.25/kWh (31) and PUE ranges ~1.09–1.4 (32; 33), which yields annual power spend of \$570–3,000/kW/y. That is, if launch costs reach $\lesssim$\$200/kg, annualized cost per unit of power in space could be approximately comparable to terrestrial spend. If, instead, 104,000t are launched (a 72% decrease from the above target), prices could be ~\$300/kg by the mid-2030s. The launched power price for a constellation of Starship v2 mini satellites would then be ~\$1,200/kW/y, still within range of terrestrial annual power prices.

References

[1] Vaswani, A. *et al.* Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 5998–6008 (Curran Associates, Inc., 2017). URL https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[2] Elsworth, C. *et al.* Measuring the environmental impact of delivering ai at google scale (2025). 2508.15734.

[3] Terrell, M. Google signs advanced nuclear clean energy agreement with kairos power. The Keyword (Google Blog) (2024). URL https://blog.google/outreach-initiatives/sustainability/google-kairos-power-nuclear-energy-agreement/.

[4] Commonwealth Fusion Systems. Google and commonwealth fusion systems sign strategic partnership. Commonwealth Fusion Systems Press Release (2025). URL https://cfs.energy/news-and-media/google-and-commonwealth-fusion-systems-sign-strategic-partnership.

[5] Terrell, M. A first-of-its-kind geothermal project is now operational. Google Blog (2023). URL https://blog.google/outreach-initiatives/sustainability/google-fervo-geothermal-energy-partnership/.

[6] Baily, M. N., Byrne, D. M., Kane, A. T. & Soto, P. E. Generative ai at the crossroads: Light bulb, dynamo, or microscope? *Brookings Institution* (2025). URL https://www.brookings.edu/articles/generative-ai-at-the-crossroads-light-bulb-dynamo-or-microscope/. Working Paper from the Center on Regulation and Markets.

[7] Gohr, C. *et al.* Artificial intelligence in sustainable development research. *Nature Sustainability* **8** (2025).

[8] McDuff, D. *et al.* Towards accurate differential diagnosis with large language models. *Nature* **642**, 451–457 (2025).

[9] Jumper, J. *et al.* Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).

[10] NASA Office of Technology, Policy, and Strategy. Space-based solar power. Tech. Rep., National Aeronautics and Space Administration (2024). URL https://www.nasa.gov/wp-content/uploads/2024/01/otps-sbsp-report-final-tagged-approved-1-8-24-tagged-v2.pdf.

[11] Kruft, B. A techno-economic analysis of space-based solar power systems. *Junior Management Science (JUMS)* **8**, 732–771 (2023). URL https://hdl.handle.net/10419/295052.

[12] Asimov, I. Reason. *Astounding Science Fiction* **27** (1941).

[13] Sengupta, M. *et al.* The national solar radiation data base (nsrdb). *Renewable and Sustainable Energy Reviews* **89**, 51–60 (2018). This article describes the data model used for the NSRDB, which produced the image data.

[14] BBC News. India on track for second successful moon landing. BBC News (2024). URL https://www.bbc.com/news/articles/cjewvpkw7weo.

[15] Feilden, E., Oltean, A. & Johnston, P. Why we should train ai in space. White Paper, Star Cloud Inc. (via GitHub Pages) (2025). URL https://starcloudinc.github.io/wp.pdf. Report on Generative AI and orbital data centers.

[16] Singh, A. *et al.* Jupiter rising: A decade of clos topologies and centralized control in google's datacenter network. *ACM SIGCOMM computer communication review* **45**, 183–197 (2015).

[17] Jouppi, N. *et al.* Tpu v4: An optically reconfigurable supercomputer for machine learning with hardware support for embeddings.

In *Proceedings of the 50th annual international symposium on computer architecture*, 1–14 (2023).

[18] Urata, R. *et al.* Mission apollo: Landing optical circuit switching at datacenter scale. In *Proceedings of the ACM SIGCOMM 2022 Conference*, 317–333 (Association for Computing Machinery (ACM), 2022). URL https://doi.org/10.1145/3544216.3544246.

[19] Kong, E. M. C. & Miller, D. W. Electromagnetic formation flight for multisatellite arrays. *Journal of Spacecraft and Rockets* **41**, 659–666 (2004). URL https://arc.aiaa.org/doi/10.2514/1.2172.

[20] D'Amico, S. *Autonomous Formation Flying in Low Earth Orbit*. Ph.D. thesis, Delft University of Technology (2010).

[21] Sinclair, D. & Dyer, J. Radiation effects and cots parts in smallsats. In *SmallSat conference* (2013).

[22] Balaram, B. *et al.* Mars helicopter technology demonstrator. In *2018 AIAA atmospheric flight mechanics conference*, 0023 (AIAA, 2018).

[23] Barth, J. L., Dyer, C. & Stassinopoulos, E. Space, atmospheric, and terrestrial radiation environments. *IEEE Transactions on nuclear science* **50**, 466–482 (2003).

[24] Sinclair, D. & Dyer, J. Radiation effects and cots parts in smallsats (2013).

[25] Schwank, J. R. *et al.* Radiation effects in mos oxides. *IEEE Transactions on Nuclear Science* **55**, 1833–1853 (2008).

[26] Xiong, Y. *et al.* Single-event upset cross-section trends for d-ffs at the 5- and 7-nm bulk finfet technology nodes. *IEEE Transactions on Nuclear Science* **70**, 381–386 (2023).

[27] Dixit, H. D. *et al.* Silent data corruptions at scale (2021). 2102.11245.

[28] Galloway, W. D. *et al.* Developing and testing a low cost electric propulsion system for cubesats. In *31st Annual AIAA/USU Conference on Small Satellites* (Logan, UT, 2017). URL https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=5938&context=smallsat. The full author list is long; citing the lead author and 'many others' is the appropriate citation style for this conference paper.

[29] Mankins, J. C. The cost of reaching orbit: Ground-based launch systems. *Acta Astronautica* **68**, 1726–1738 (2011). URL https://www.sciencedirect.com/science/article/pii/S026596461100035X.

[30] Virtus Solis Technologies. The starship revolution in space. Blog Post (2023). URL https://www.aspistrategist.org.au/the-starship-revolution-in-space/#:~:text=Because%20no%20hardware%20will%20be,a%20year%20from%20this%20location. Article by Virtus Solis Technologies on The ASPI Strategist.

[31] Nlyte Software. Data center rack power costs: A condensed analysis. Blog Post (2024). URL https://www.nlyte.com/blog/data-center-rack-power-costs-a-condensed-analysis/.

[32] Masnadi Shirazi, M., Shehabi, A. & Smith, S. United states data center energy usage report: Trends, projections, and policy. Tech. Rep. LBNL-2024-1000, Lawrence Berkeley National Laboratory (LBNL) (2024). URL https://eta-publications.lbl.gov/sites/default/files/2024-12/lbnl-2024-united-states-data-center-energy-usage-report_1.pdf. Initial Publication Date: December 2024.

[33] Google Cloud Team. Measuring the environmental impact of ai inference. Google Cloud Blog (2024). URL https://cloud.google.com/

blog/products/infrastructure/measuring-the-environmental-impact-of-ai-inference/.

[34] Gomes, P. A. T., Zand, G. & Van Zyl, J.-H. Emissions intensity of generative ai. *Nature Communications* **14**, 6099 (2023). URL https://www.nature.com/articles/s41467-023-41971-7.

[35] Next Spaceflight. Launches. Website (2025). URL https://nextspaceflight.com/launches/.

[36] Mueller, R. E. *et al.* Electric propulsion system testing: Nasa's path to the moon and mars. Tech. Rep. NASA Report ID 20200001093, NASA (2020). URL https://ntrs.nasa.gov/citations/20200001093. AIAA Scitech Forum Presentation, 6–10 January 2020, Orlando, FL.

[37] McDowell, J. Jonathan's space report (jsr). Website (2025). URL https://planet4589.org/space/jsr/jsr.html.

[38] WRAL News Staff. Spacex cost reduction, innovation drive space economy growth. WRAL.com (2024). URL https://www.wral.com/business/technology/spacex-cost-reduction-innovation-october-2024/.

[39] Zastrow, M. Rivals are rising to challenge the dominance of spacex. *MIT Technology Review* (2025). URL https://www.technologyreview.com/2025/04/03/1114198/rivals-are-rising-to-challenge-the-dominance-of-spacex/. Accessed: 2025-10-30.

[40] Douillard, A. *et al.* Diloco: Distributed low-communication training of language models (2023). 2311.08105.

[41] CAILABS Marketing Team. Successful initial tests of the keraunos optical communications satellite. CAILABS News (2024). URL https://www.cailabs.com/news/aerospace-and-defense/successful-initial-tests-of-the-keraunos-optical-communications-satellite/. Official company news release.

[42] MIT News Staff. New communications system achieves fastest laser link in space yet. MIT News (2022). URL https://news.mit.edu/2022/communications-system-achieves-fastest-laser-link-space-yet-1130.

[43] Galloway, W. D., Perez, G. L., Lazo, A. *et al.* Tbird: Optical communication from a small satellite to an optical ground station. In *37th Annual AIAA/USU Conference on Small Satellites* (Logan, UT, 2023). URL https://ntrs.nasa.gov/api/citations/20230007959/downloads/TBIRD-smallsat-2023.pdf. NASA NTRS ID 20230007959; citing lead author and others due to long list of contributors.

[44] Mordvintsev, A., Randazzo, E., Niklasson, E. & Levin, M. Growing neural cellular automata. *Distill* (2020). URL https://distill.pub/2020/growing-ca/.

[45] Clohessy, W. H. & Wiltshire, R. S. Terminal guidance system for satellite rendezvous. *Journal of the Aerospace Sciences* **27**, 653–658 (1960).

[46] Tschauner, J. & Hempel, P. Rendezvous zu einem in elliptischer bahn umlaufenden ziel [rendezvous with a target in an elliptical orbit]. *Astronautica Acta* **11** (1965).

[47] Vinti, J. P. Theory of an accurate intermediary orbit for satellite astronomy. *Journal of Research of the National Bureau of Standards, Section B: Mathematics and Mathematical Physics* **65B**, 169–202 (1961).

[48] Montenbruck, O. & Gill, E. *Satellite Orbits: Models, Methods and Applications* (Springer Berlin Heidelberg, 2013). URL https://link.springer.com/book/10.1007/978-3-642-58351-3.

[49] Kimukin, I. *et al.* Ingaas-based high-performance pin photodiodes. *IEEE Photonics Technology Letters* **14**, 366–368 (2002).

[50] Berger, M. J., Coursey, J. S., Zucker, M. A. & Chang, J. ESTAR, PSTAR, and ASTAR: Computer Programs for Calculating Stopping-Power and Range Tables for Electrons, Protons, and Helium Ions (version 2.0.1). Tech. Rep., National Institute of Standards and Technology, Gaithersburg, MD (2017). URL http://physics.nist.gov/Star. Accessed: 2025-04-08.

[51] Sorriaux, J. *et al.* Evaluation of gafchromic® ebt3 films characteristics in therapy photon, electron and proton beams. *Physica Medica* **29**, 599–606 (2013).

[52] ESA Earth Online. Oneweb. EO Portal Satellite Missions (2025). URL https://www.eoportal.org/satellite-missions/oneweb#overview.

[53] SatNews Staff. Airbus awards rocket lab contract to power next-generation oneweb constellation for eutelsat. SatNews (2025). URL https://news.satnews.com/2025/03/13/airbus-awards-rocket-lab-contract-to-power-next-generation-oneweb-constellation-for-eutelsat/.

[54] ESA Earth Online. Iridium next. EO Portal Satellite Missions (2025). URL https://www.eoportal.org/satellite-missions/iridium-next#eop-quick-facts-section.

[55] Tingley, G. Iridium selects thales alenia to build iridium next constellation. SpaceNews (2025). URL https://spacenews.com/iridium-selects-thales-alenia-build-iridium%E2%80%82next-constellation/.

[56] Anderson, M. B. Learning rate sensitivity model for space system manufacturing and operations. In *International Cost Estimating and Analysis Association (ICEAA) Professional Development & Training Workshop* (2018). URL https://www.iceaaonline.com/wp-content/uploads/2018/07/MM01-Paper-Learning-Rate-Sensitivity-Model-Anderson.pdf. Presented at the ICEAA Workshop.

[57] Hassouni, R. & Orozco-Cabrera, D. The missing rocket: An economic and engineering analysis of the reusability dilemma in the european space sector. *Intereconomics* **60** (2025). URL https://www.intereconomics.eu/contents/year/2025/number/2/article/the-missing-rocket-an-economic-and-engineering-analysis-of-the-reusability-dilemma-in-the-european-space-sector.html.

[58] McKinsey & Company. Space launch: Are we heading for oversupply or a shortfall? McKinsey Insight (2023). URL https://www.mckinsey.com/industries/aerospace-and-defense/our-insights/space-launch-are-we-heading-for-oversupply-or-a-shortfall.

[59] Foust, J. Web entrepreneur eyes small launcher market. SpaceNews (2013). URL https://spacenews.com/web-entrepreneur-eyes-small-launcher-market/.

[60] Moyer, M. Spacex unveils launch of falcon heavy, world's most powerful rocket by 2013. Universe Today (2011). URL https://www.universetoday.com/84638/spacex-unveils-launch-of-falcon-heavy-worlds-most-powerful-rocket-by-2013/#:~:text=Musk%20said%20the%20Falcon%20Heavy,sought%20and%20near%20mythical%20goal.

[61] Wang, B. Spacex version 4 starship will be 30 meters taller in 2027. NextBigFuture.com (2025). URL https://www.nextbigfuture.com/2025/08/spacex-version-4-starship-will-be-30-meters-taller-in-2027.html.

[62] SpaceX. Human spaceflight: Mars. Website (2025). URL https://www.spacex.com/humanspaceflight/mars.

[63] Musk, E. Tweet on launch costs. X (formerly Twitter) (2025). URL https://x.com/elonmusk/status/1960812698037518540.

[64] Marketplace Staff. Spacex engineered cheaper space flight, but startups are entering the market. Marketplace.org (2024). URL https://www.marketplace.org/episode/2024/11/19/spacex-engineered-cheaper-space-flight-but-startups-are-entering-the-market.

[65] CNBC Staff. Full elon musk transcript about spacex falcon 9 block 5 (2018). URL https://www.cnbc.com/2018/05/11/full-elon-musk-transcript-about-spacex-falcon-9-block-5.html. Transcript of a press conference or call.

[66] Cunningham, E. Where does spacex get their rocket fuel? https://primalnebula.com/where-does-spacex-get-their-rocket-fuel/ (2023). Accessed: 21 October 2025.

[67] Defense Logistics Agency. Aerospace standard prices, effective october 1, 2024 (fy2025). Tech. Rep., Defense Logistics Agency (DLA) (2024). URL https://www.dla.mil/Portals/104/Documents/Energy/Standard%20Prices/Aerospace%20Prices/E_2024Oct1AerospaceStandardPrices_240926_3.pdf?ver=0Ssdix0y2Fc4ESxz1aqCDQ%3D%3D. Accessed: 21 October 2025.

[68] YouTube. Elon musk's spacex announced new raptor 4 upgrade...shocked the whole industry! [video]. https://www.youtube.com/watch?v=7OSDPzi42GA&t=4 (2025). Accessed: 21 October 2025.

[69] SpaceX. Falcon 9. Website (2025). URL https://www.spacex.com/vehicles/falcon-9.

[70] Pelham, C. How much does a spacex rocket cost? falcon 9 vs nasa's bill. Technowize (2025). URL https://www.technowize.com/how-much-does-a-spacex-rocket-cost-falcon-9-vs-nasas-bill/.

[71] SpaceX. Starlink progress report. Tech. Rep., SpaceX (2023). URL https://starlink-stories.cdn.prismic.io/starlink-stories/Z3QOWJbqstJ986KD_StarlinkProgress-V11_Low-Res-compressed.pdf. Public report accessible via Prismic CDN.

[72] Mallama, A. *et al.* Starlink generation 2 mini satellites: Photometric characterization (2023). URL https://arxiv.org/abs/2306.06657. 2306.06657.

[73] Rodgers, E. *et al.* Space based solar power. Tech. Rep. NASA Report ID 20230018600, National Aeronautics and Space Administration (NASA) (2024). URL https://www.nasa.gov/wp-content/uploads/2024/01/otps-sbsp-report-final-tagged-approved-1-8-24-tagged-v2.pdf. Published by the Office of Technology, Policy and Strategy (OTPS).

[74] Johnson, L., Carr, J., Fabisinski, L. & Lockett, T. R. Lightweight integrated solar array (lisa): Providing higher power to small spacecraft. Tech. Rep. NASA/TP–2015-218770 / AIAA 2015-3974, NASA George C. Marshall Space Flight Center (2015). URL https://ntrs.nasa.gov/api/citations/20150016538/downloads/20150016538.pdf. Presented at the AIAA Propulsion and Energy Forum, 27–29 July 2015, Orlando, FL.

[75] Sinovoltaics. Packing density of solar cells. Website (2025). URL https://sinovoltaics.com/learning-center/basics/packing-density/#:~:text=Square%20cells%3AThe%20wafers%20of,highest%20packing%20density%20of%2090%25.

[76] Elkelawy, M., Mohamed, A. E. & Seleem, H. E. Optimizing photovoltaic power plant efficiency: A comprehensive study on design, implementation, and sustainability. *Pharos Engineering Science Journal (PESJ)* **2**, 12–21 (2025). URL https://journals.ekb.eg/article_415585.html.

[77] Wall, M. Spacex's starlink satellites: Facts and figures. Space.com (2023). URL https://www.space.com/spacex-starlink-satellites.html.

Acknowledgements

We thank Amaan Pirani for critical contributions to cost modeling and overall feasibility analysis, Marcin Kowalczyk for independent numerical validation calculations, Thomas Zurbuchen for his contributions to the systems and architecture concepts, and Kenny Vassigh and Jerry Chiu for technical input on system and thermal design.

Author Information

These authors contributed equally: B.A.A., T.R.B., M.B., J.V.B., T.F., K.G., U.K., R.P.

B.A.A. conceived of this overall project. T.F. developed the orbital dynamics and formation flight work. U.K. and R.P. designed and performed the radiation testing. J.V.B. conducted the launch cost analysis. J.V.B., U.K. and T.R.B. prepared and proofread the manuscript. K.G. and U.K. conducted the inter-satellite link analysis. T.R.B. and M.B. developed the system design overview. J.M. provided overall guidance and supervision to the project.

Supplementary Information

Orbital dynamics modeling

Upper bounds on mission requirements (such as delta-v requirements) can often be derived from conservative estimates based on simplifying models. When implementing a control model for satellite cluster formation flight, accurate information about satellite position and orientation, precise models of predictable accelerations (due to Earth's known but irregular gravitational potential, solar and lunar tides, etc.), and reasonable models of noisy accelerations (such as due to space weather- dependent atmospheric friction) are indispensable: every in-principle-manageable acceleration that is not included in the control model will manifest as unexpected drift, increasing overall mission delta-v requirements.

One promising approach towards implementing formation flight control is to use backpropagation-based techniques. This in general starts from an objective function whose calculation involves numerical ODE-integration that utilizes the complete motion-state of all satellites in the formation, and whose minimum describes a desirable target-state. This function will in general accumulate (transient) violations of the cluster being in a good configuration. If the control—whose output can be understood as a plan to drive actuators—is implemented in terms of an algorithm with tunable parameters (and may include a learned model), adjoint-state methods can be used to backpropagate objective-function gradients through ODE-integration, and automatic differentiation (typically reverse-mode AD) can then backpropagate gradients into model parameters. Implementing such an approach is greatly simplified by employing a Machine Learning framework such as JAX. This approach can also handle higher-order derivatives via "differentiable programming" approaches.