LLMSPACE: CARBON FOOTPRINT MODELING FOR LLM INFERENCE ON LEO SATELLITES

Anonymous authors
Paper under double-blind review

ABSTRACT

Large language model (LLM) inference imposes growing energy and carbon costs. Solar-powered low Earth orbit (LEO) platforms can reduce reliance on grid electricity, but manufacturing, launch, peripheral, and radiation-hardening requirements introduce substantial embodied carbon. We present *LLMSpace*, a component-level framework for modeling the carbon and performance of LEO-based LLM inference, including compute hardware, solar arrays, batteries, radiative cooling, launch emissions, and radiation-hardening overheads. Across the evaluated scenarios, launch carbon intensity is the primary determinant of the orbital–terrestrial carbon crossover, while radiation-hardening power overhead is an important secondary factor. Workload and GPU evaluations further show that prompt and output lengths, batch size, request rate, and hardware throughput substantially affect inference energy and amortized embodied carbon per token. These results show that LEO inference is not inherently lower-carbon; its carbon benefit depends on launch accounting, mission lifetime, terrestrial grid intensity, hardware efficiency, workload characteristics, and system utilization.

1 INTRODUCTION

The widespread deployment of LLMs has introduced substantial energy and environmental costs. A single short GPT-4o query consumes ~ 0.43 Wh, translating to an estimated 392–463 GWh of electricity for GPT-4o inference in 2025 (Jegham et al., 2025), comparable to the annual consumption of 35K average U.S. households. As LLMs become increasingly integrated into ubiquitous applications, from web browsers to operating systems, their inference workloads will continue to grow. Concurrently, global data-center electricity demand is projected to $> 2\times$ by 2030 (Pilz et al., 2025), underscoring the need to reduce the energy and carbon footprint of large-scale LLM inference.

To mitigate these impacts, industry and academia have increasingly explored solar-powered LEO infrastructure for connectivity and orbital computing, including efforts from Microsoft (2025), StarCloud (2025), Google (2025), Amazon (2025), and related academic studies (Wang et al., 2024; Bleier et al., 2023; Aili et al., 2025). Such platforms could host AI accelerators, including GPUs (Slater et al., 2020) and TPUs (Casey et al., 2023), to execute LLM inference using solar energy and reduce dependence on terrestrial power grids. Operating at altitudes of several hundred to $\sim 2{,}000$ km, these satellites are launched by vehicles such as Falcon 9 (Ohs et al., 2025) and use ground and inter-satellite links to provide services.

Although solar power can reduce the operational carbon of LEO-based LLM inference, embodied emissions from satellite manufacturing, launch, and radiation-hardened hardware may offset these benefits. However, no existing work jointly models the operational carbon, embodied carbon, and inference throughput of LEO inference systems. Prior work characterizes LLM training (Faiz et al., 2024) and inference (Fu et al., 2025; Jegham et al., 2025) in terrestrial data centers, while existing LEO-computing carbon models (Ohs et al., 2025; Aili et al., 2025) have three key limitations:

- **Incomplete peripheral modeling.** Existing models either omit critical peripherals, such as cooling systems (Ohs et al., 2025) or independent solar arrays (Aili et al., 2025). They also cover limited design choices, excluding alternatives such as multi-junction solar arrays and radiative cooling panels with integrated heat pipes.
- **No radiation-hardened hardware modeling.** Existing works (Ohs et al., 2025; Aili et al., 2025) assume commercial off-the-shelf (COTS) CMOS hardware and neglect space-specific radiation

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

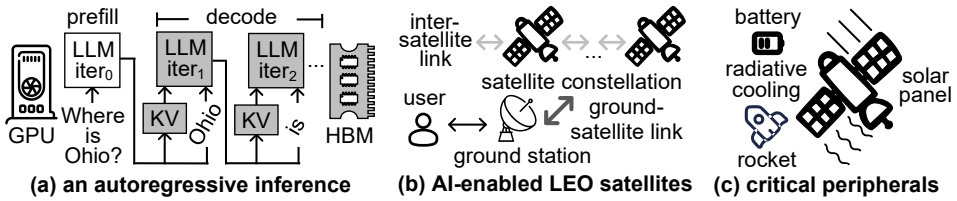

Figure 1: LEO satellite platforms for large-scale LLM inference.

effects, including total ionizing dose (Slater et al., 2020), displacement damage (Srour & Palko, 2013), and single-event effects (Yue et al., 2015). COTS GPUs may survive only ~ 2 years in orbit (Slater et al., 2020), while COTS DRAM may fail within months (Lee et al., 2023), motivating radiation-hardened designs whose additional embodied carbon remains unmodeled.

- **No LLM-specific workload modeling.** Prior LEO-computing carbon models (Ohs et al., 2025; Aili et al., 2025) focus on conventional workloads, such as image processing (Bleier et al., 2023), and do not capture important LLM-inference characteristics such as prompt length, generated-token count, batch size, and request rate. Consequently, they cannot characterize how diverse LLM workloads affect inference throughput, energy, and amortized embodied carbon.

This paper presents *LLMSpace*, a carbon footprint modeling framework for LEO LLM inference satellites. Our contributions are:

- **Comprehensive peripheral modeling.** *LLMSpace* quantifies the embodied carbon of key peripherals, including solar arrays, batteries, and radiative cooling panels, and supports multiple peripheral configurations across satellite classes with different operational lifetimes.
- **Radiation-hardened hardware modeling.** *LLMSpace* models the embodied carbon of AI accelerators fabricated using radiation-hardened process technologies, such as fully depleted silicon-on-insulator (Paillet et al., 2006). It separates process-node scaling from radiation-hardening-by-design overheads, accounts for architectural mitigation techniques, and captures radiation-hardened HBM designs for LLM inference based on MRAM technologies (Marinella, 2021).
- **LLM workload-aware carbon modeling.** *LLMSpace* captures workload-dependent LLM inference by accounting for prompt and generated-token lengths, batch size, request rate, and hardware characteristics when estimating inference throughput, energy, and amortized embodied carbon. It further propagates compute-system power into the provisioning and embodied carbon of solar arrays, batteries, and radiative cooling.

2 Background

LLM Inference. As shown in Figure 1(a), LLM inference generates tokens autoregressively in two phases. During *prefill*, all prompt tokens are processed in parallel to generate the first output token, while attention states are stored in the key–value (KV) cache. During *decode*, subsequent tokens are generated sequentially using the cached KV states. Prefill is typically compute-bound, whereas decoding is memory-bound due to repeated HBM accesses and often dominates inference energy consumption. For example, a short GPT-4o request consumes ~ 0.43 Wh, corresponding to an estimated 392–463 GWh of annual electricity demand in 2025 (Jegham et al., 2025).

AI-Enabled Satellites for LLM Inference. Recent studies (Wang et al., 2024; Bleier et al., 2023; Aili et al., 2025) propose solar-powered LEO platforms to reduce the terrestrial energy demand of large-scale computing. These satellites could host AI accelerators such as GPUs or TPUs and are deployed using launch vehicles. Due to resource and radiation constraints, onboard LLM training remains impractical, making inference the primary target workload (Wang et al., 2024). As shown in Figure 1(b), user requests are transmitted through ground stations and ground–satellite links for onboard inference. Powered by solar energy (Bleier et al., 2023), such LEO platforms can shift LLM inference away from terrestrial power grids.

Embodied Carbon of AI-Enabled LEO Satellites. *The carbon footprint of LEO-based LLM inference comprises operational and embodied emissions.* Operational emissions arise from electricity consumed during computation (Fu et al., 2025); for solar-powered satellites, in-orbit operational emissions can be substantially reduced. Embodied emissions arise from vehicle launching, satellite manufacturing, and the fabrication of computing, communication, power, and thermal-management hardware. Consequently, compared with terrestrial data centers powered by renewable energy, LEO computing can incur substantially greater embodied carbon (Bleier et al., 2023; Aili et al., 2025). As

shown in Figure 1(c), satellite computing additionally requires solar arrays, batteries, and radiative cooling systems, each contributing to embodied emissions (Aili et al., 2025).

Hardware Reliability in Space. Spaceborne CMOS devices are exposed to ionizing radiation and energetic particles that degrade reliability and service lifetime (Slater et al., 2020). Three major mechanisms are total ionizing dose (TID) (Slater et al., 2020), displacement damage (DD) (Srour & Palko, 2013), and single-event effects (SEE) (Yue et al., 2015). TID causes cumulative charge trapping, inducing threshold-voltage shifts and increased leakage. DD creates lattice defects through non-ionizing energy loss, degrading carrier transport. SEE results from individual energetic-particle strikes and can induce transient faults, data corruption, latch-up, or destructive failure. Together, these mechanisms motivate radiation-hardened computing hardware for sustained LEO operation.

Radiation-Hardening Techniques. Radiation tolerance is achieved through complementary process- and architecture-level techniques (Kannaujiya & Shah, 2024). Process-level approaches, including radiation-aware CMOS layouts and fully depleted silicon-on-insulator (FD-SOI) technologies (Paillet et al., 2006), reduce charge collection and improve TID tolerance, while material engineering can enhance DD resistance (Srour et al., 2003). Such technologies are often implemented at mature process nodes (e.g., 7 nm) (Maillard et al., 2024), potentially increasing die area and embodied carbon relative to advanced-node COTS hardware. Architecture-level techniques, including error-correcting codes (ECC), triple modular redundancy (TMR), and latch-up protection (Jeon et al., 2010), further improve resilience at the cost of additional hardware redundancy, area, and embodied carbon.

3 Related Work and Motivation

The large-scale deployment of LLMs has raised substantial environmental concerns (Jegham et al., 2025). Prior studies (denoted by TERR) have quantified the carbon footprint of LLM training (Faiz et al., 2024) and inference (Fu et al., 2025; Jegham et al., 2025) in terrestrial data centers. Recent LEO computing proposals (Slater et al., 2020; Casey et al., 2023) explore solar-powered LEO platforms for AI inference; however, these systems additionally require launch vehicles, satellite platforms, power and thermal subsystems, and space-reliable computing hardware. Thus, while solar power reduces in-orbit operational emissions, embodied carbon from launch, satellite manufacturing, and radiation-hardened hardware may offset these benefits.

Table 1: Comparing *LLMSpace* against prior work.

scheme	space-based	peripheral breakdown	rad-hard hardware	LLM-specific
TERR	✗	✗	✗	✓
EIR	✓	✓	✗	✗
NE	✓	✗	✗	✗
LLMSpace	✓	✓	✓	✓

As summarized in Table 1, existing LEO-computing carbon models (Ohs et al., 2025; Aili et al., 2025) have three key limitations. First, they either omit critical peripherals, such as radiative cooling panels (Ohs et al., 2025) or independent solar arrays (Aili et al., 2025), with limited support for alternative configurations. Second, they assume COTS CMOS hardware and neglect the embodied-carbon and reliability implications of radiation-hardened computing. Third, they target conventional workloads, such as image processing (Bleier et al., 2023), without capturing critical LLM-specific characteristics, including prompt length, generated-token count, batch size, and request rate. To address these limitations, we propose *LLMSpace*, a carbon modeling framework that jointly captures peripheral configurations, radiation-hardened hardware, and LLM-specific characteristics for LEO-based inference satellites.

$$C_{\mathrm{L}} = I_{\mathrm{L}} \cdot m \quad (1) \qquad m_{\mathrm{so}} = a_{\mathrm{so}} \cdot A_{\mathrm{so}} \quad (2) \qquad C_{\mathrm{so}} = I_{\mathrm{so}} \cdot A_{\mathrm{so}} \quad (3) \qquad P_{\mathrm{so}} = \mathrm{ISI} \cdot A_{\mathrm{so}} \cdot \eta \quad (4)$$

$$m_{\mathrm{ba}} = a_{\mathrm{ba}} \cdot \mathrm{cap}_{\mathrm{ba}} \quad (5) \qquad C_{\mathrm{ba}} = I_{\mathrm{ba}} \cdot \mathrm{cap}_{\mathrm{ba}} \quad (6) \qquad Q = \epsilon \cdot \sigma \cdot A_{\mathrm{ra}} \cdot (T_{\mathrm{ra}}^4 - T_{\mathrm{bg}}^4) \quad (7)$$

4 LLMSpace

4.1 Embodied Carbon Modeling

We model the embodied carbon of LEO-based inference satellites by accounting for launch emissions, critical power and thermal peripherals, and radiation-hardened computing hardware.

Launch. The launch emissions attributable to a payload (C_{L}) are modeled using Equation 1 (Aili et al., 2025), where m is the payload mass and I_{L} is the payload-normalized launch carbon intensity. For example, a Falcon 9 launch emits $\sim 3.3 \times 10^5$ $\mathrm{kgCO_2e}$ and supports a payload of up to 22.8 ton (Barker et al., 2024), yielding $I_{\mathrm{L}} \approx 14.5$ $\mathrm{kgCO_2e/kg}$. Estimates vary substantially with the accounting boundary. Including re-entry emissions increases the reported intensity to

Table 2: Solar array technologies (Louwen et al., 2016; Chanoine et al., 2015) (the unit of $a_{\mathrm{so}} \mid I_{\mathrm{so}}$ is $\mathrm{kg/m^2} \mid \mathrm{kgCO_2e/m^2}$).

parameter	Si panel	GaAs array	multi-junction
a_{so}	1	1.1	1.2
η	0.17	0.3	0.34
I_{so}	120	150	160

Table 3: Radiative cooling panel technologies (Gilmore, 2002) (the unit of $a_{\mathrm{ra}} \mid I_{\mathrm{ra}}$ is $\mathrm{kg/m^2} \mid \mathrm{kgCO_2e/kg}$).

parameter	passive body-mount	aluminum honeycomb	advanced heat-pipe
a_{ra}	2.7	2.8	3
ϵ	0.8	0.95	0.98
I_{ra}	12	15	19

Table 4: Battery technologies (Knap et al., 2020; Hasan et al., 2025) (the unit of $a_{\mathrm{ba}} \mid I_{\mathrm{ba}}$ is $\mathrm{kg/kWh} \mid \mathrm{kgCO_2e/kWh}$).

parameter	LFP	NMC	rad-hard
a_{ba}	8.3	4.5	5.5
I_{ba}	60	80	100

Table 5: CPA (Faiz et al., 2024) of COTS CMOS technologies (unit: $\mathrm{kgCO_2e/cm^2}$).

process (nm)	28	14	7	4
COTS	0.9	1.2	2.15	3

$51.7\ \mathrm{kgCO_2e/kg}$ (Ohs et al., 2025), while still larger estimates may include launch-vehicle manufacturing (Calabuig et al., 2024). We therefore treat I_{L} as an uncertain parameter. Appendix A discusses the underlying estimates and accounting boundaries, and Section 6 evaluates their effects through sensitivity analysis.

Critical Peripherals. Different LEO satellites differ in service lifetime, power demand, and thermal constraints, resulting in distinct peripheral designs and embodied carbon footprints. We therefore model the embodied carbon of key peripherals, including solar arrays, batteries, and radiative cooling panels, across different satellites.

- **Solar Arrays.** A LEO satellite completes an orbit roughly every 90–100 minutes (Prol et al., 2022), corresponding to $N_{\mathrm{cyc}} \approx 15$–$16$ orbital cycles per day. Each orbit typically includes 55–65 minutes of sunlight for power generation and battery charging, followed by 30–35 minutes of eclipse during which stored energy powers onboard LLM inference. The solar-array mass is computed using Equation 2, where A_{so} denotes the total array area and a_{so} its areal mass density. Table 2 summarizes a_{so} for Si panels, GaAs arrays, and multi-junction arrays. Manufacturing embodied carbon is calculated using Equation 3, where I_{so} denotes the technology-specific embodied carbon intensity. Array output power P_{so} is computed using Equation 4 from the incident solar irradiance $\mathrm{ISI} = 1367\ \mathrm{W/m^2}$ and conversion efficiency η. These technologies exhibit distinct mass, carbon, and efficiency trade-offs. For example, Si panels (Louwen et al., 2016) have lower conversion efficiency than advanced alternatives but offer lower areal mass density a_{so} and manufacturing carbon intensity I_{so} among the considered technologies.
- **Batteries.** Batteries supply power during eclipse periods and undergo frequent charge–discharge cycles. Small satellites typically use lithium iron phosphate (LFP) batteries with capacities of tens to hundreds of watt-hours, whereas larger platforms (Bleier et al., 2023; Amazon, 2025) employ kilowatt-hour-scale nickel–manganese–cobalt (NMC) batteries with enhanced thermal management and radiation tolerance (Hasan et al., 2025). Battery mass is computed using Equation 5, where a_{ba} is the specific mass and $\mathrm{cap}_{\mathrm{ba}}$ is the required capacity. Manufacturing embodied carbon is calculated using Equation 6 with battery-specific carbon intensity I_{ba}. As summarized in Table 4, battery technologies differ in a_{ba} and I_{ba}. Larger satellites generally require greater battery capacity and more robust designs, increasing their embodied carbon.
- **Radiative Cooling Panels.** In vacuum, spacecraft reject heat primarily through thermal radiation. The required radiator area follows the Stefan–Boltzmann relation (Gilmore, 2002), given in Equation 7, where Q is the heat load, ϵ the surface emissivity, σ the Stefan–Boltzmann constant, A_{ra} the effective radiator area, and T_{ra} and T_{bg} the radiator and background temperatures, respectively. Small satellites typically use passive body-mounted radiators for modest thermal loads, whereas larger platforms employ honeycomb or heat-pipe radiators for multi-kilowatt dissipation (Bleier et al., 2023; Amazon, 2025). Radiator mass m_{ra} is computed using Equation 8 from its areal density a_{ra}, while manufacturing embodied carbon is calculated using Equation 9 with carbon intensity I_{ra}. Table 3 summarizes the technology-dependent a_{ra}, ϵ, and I_{ra}. As onboard compute power increases, the resulting thermal load requires larger radiators, making thermal management an increasingly important contributor to embodied carbon.

Computing and Communication Hardware. We model the embodied carbon of hardware required for LEO-based LLM inference, including processors, accelerators, memory, storage, and communication subsystems.

$$m_{\mathrm{ra}} = a_{\mathrm{ra}} \cdot A_{\mathrm{ra}} \quad (8) \qquad C_{\mathrm{ra}} = I_{\mathrm{ra}} \cdot m_{\mathrm{ra}} \quad (9) \qquad C_{\mathrm{lo}} = \mathrm{CPA} \cdot A_{\mathrm{lo}} \quad (10)$$

$$C_{\mathrm{me}} = \mathrm{CPC} \cdot \mathrm{cap}_{\mathrm{me}} \quad (11) \qquad C_{\mathrm{RH}} = C_{\mathrm{lo}} \cdot S_{\mathrm{RHA}} \quad (12) \qquad P_{\mathrm{RH}} = P_{\mathrm{sy}} \cdot S_{\mathrm{RHP}} \quad (13)$$

$$C_{\mathrm{t}} = \dot{C}_{\mathrm{em}} T_{\mathrm{cy}} / B / N_{\mathrm{t}} \quad (14) \quad O_{\mathrm{so}} = P_{\mathrm{so}} \cdot N_{\mathrm{cyc}} \cdot T_{\mathrm{lit}} \quad (15) \qquad O_{\mathrm{ba}} = N_{\mathrm{cyc}} \cdot \mathrm{cap}_{\mathrm{ba}} \cdot \mathrm{DoD} \quad (16)$$

- **CPU and GPU.** Processor-logic manufacturing carbon is computed using Equation 10, where A_{lo} denotes die area and CPA the fabrication carbon intensity per unit area (Faiz et al., 2024). Table 5 summarizes CPA for COTS CMOS technologies. Process-level radiation hardening introduces additional fabrication steps and can be implemented at only mature nodes (e.g., 7 nm) (Maillard et al., 2024), potentially increasing fabrication carbon intensity. Architecture-level techniques, including ECC, TMR, and latch-up protection, can further increase chip area and power consumption. However, no publicly characterized H100-class accelerator provides matched conventional and radiation-hardened measurements of fabrication intensity, die area, and system power. We therefore treat these overheads as independent exploratory variables rather than assuming fixed hardening factors. Equation 12 introduces S_{RHA}, which captures the combined area penalty from process-node migration and radiation-hardening fabrication and by-design, while Equation 13 introduces S_{RHP} to capture the corresponding power overhead. Appendix B defines the exploratory parameter space for S_{RHA} and S_{RHP}; Section 6 evaluates representative configurations and reports the resulting sensitivity and crossover analyses.
- **DDR and HBM Memory.** Memory embodied carbon is computed using Equation 11, where $\mathrm{cap}_{\mathrm{me}}$ denotes capacity and CPC is the carbon intensity per unit capacity (Faiz et al., 2024). For COTS DRAM, we use $\mathrm{CPC} = 0.4\ \mathrm{kgCO_2e/GB}$ for 18 nm DDR and $0.85\ \mathrm{kgCO_2e/GB}$ for HBM. COTS DRAM is suitable for approximately two-year LEO missions (Slater et al., 2020), while longer ($\sim$ 10-year) missions require radiation-tolerant alternatives. We model MRAM-based DDR- and HBM-compatible memory, which improves resilience to TID, DD, and SEE (Lakys et al., 2011), with respective carbon intensities of 0.6 and $1.7\ \mathrm{kgCO_2e/GB}$ due to additional processing and magnetic materials (Bayram et al., 2016).
- **NAND SSD.** The manufacturing embodied carbon of COTS NAND flash SSDs is also computed using Equation 11. For triple-level cell 3D-NAND flash, $\mathrm{CPC} = 22\ \mathrm{kgCO_2e/TB}$ (Weppe et al., 2025). Radiation-tolerant configurations—including single-level operation, enhanced ECCs, periodic scrubbing, and shielding—extend lifetime to 10 years (Mercury Systems, 2026), approximately increasing embodied carbon to $\mathrm{CPC} = 66\ \mathrm{kgCO_2e/TB}$.
- **Computing Subsystems.** Besides logic and memory devices, a computing server equipped with one or more GPUs (NVIDIA Corporation, 2025b) includes electromechanical modules, discrete electronic components (e.g., capacitors, resistors, diodes, and inductors), printed circuit boards (PCBs), interconnects, and structural elements. High-performance platforms, such as the NVIDIA DGX A100, H100, and B200, each weigh about $\sim$ 130 kg (NVIDIA Corporation, 2023).
- **Communication Subsystems and Satellites.** As LEO satellites are primarily communication-oriented, we jointly model the embodied carbon and mass of the satellite platform and its communication subsystem. CubeSat platforms typically employ RF-based ground-to-satellite links (GSLs) operating in UHF/S/X bands, supporting data rates from Kbps to several hundred Mbps with transmit powers of 1–20 W (Palo et al., 2014). These communication payloads generally weigh 0.5–5 kg, corresponding to $\sim$ 5–50 $\mathrm{kgCO_2e}$ embodied carbon (Aili et al., 2025). In contrast, larger platforms (e.g., Starlink V2) adopt laser inter-satellite links (ISLs) capable of up to 200 Gbps at roughly 1 W optical power (Chaudhry & Yanikomeroglu, 2021). Although energy efficient during operation, laser terminals incur substantial embodied carbon (150 $\mathrm{kgCO_2e}$) and add about 28 kg per terminal to payload mass (Gregory et al., 2024). High-capacity GSL systems further target downlink/uplink rates exceeding 96/32 Gbps (De Gaudenzi, 2025), with transmit equivalent isotropically radiated power reaching 1.75 kW (Aili et al., 2025). Under such configurations, the combined mass and manufacturing embodied carbon of the satellite and communication subsystem can be substantial; for example, a Starlink satellite weighs about 191 kg and corresponds to $\sim$2940 $\mathrm{kgCO_2e}$ embodied carbon (Aili et al., 2025).

Embodied Carbon Amortization. Because embodied carbon is independent of utilization, its per-token allocation depends on request rate. Let λ denote the request arrival rate, B the batch size, and $T(B)$ the batch inference latency; the batch cycle time is $T_{\mathrm{cy}} = \max(T(B), B/\lambda)$. Let $\dot{C}\mathrm{em}$

Table 7: Controlled analytical comparison of embodied carbon for Starlink-V1 LEO satellites equipped with a DGX H100 node and launched by Falcon 9 (unit: $\mathrm{tCO_2e}$). The constructed reference is a component-sum inventory, not an independent measurement.

component	COTS								rad-hard			
	Ref.		EIR		NE		LLMSpace		Ref.		LLMSpace	
	manu	launch	manu	launch	manu	launch	manu	launch	manu	launch	manu	launch
solar array	8.04	0.88	8.18	0.99	-	-	8.18	0.99	16.08	1.75	16.35	1.98
battery	0.54	0.42	0.58	0.48	0.58	0.48	0.58	0.48	1.08	0.84	1.17	0.95
cooling panel	1.18	0.82	-	-	0.96	0.93	0.96	0.93	2.36	1.65	1.93	1.86
comp. HW	2.17	1.68	2.17	1.89	2.17	1.89	2.17	1.89	7.92	1.68	7.92	1.89
net+satellite	1.63	2.28	1.63	2.41	1.63	2.41	1.63	2.41	1.63	2.28	1.63	2.41
total	19.64		18.33 [-6.69%]		11.06 [-43.7%]		20.22 [+2.95%]		37.26		38.08 [+2.18%]	

denote embodied carbon amortized over the system lifetime and $N\mathrm{t}$ the average number of input prompt and generated tokens per request. Equation 14 gives the embodied carbon per token, C_{t}. At low request rates, C_{t} increases as λ decreases; at high rates, throughput is bounded by $B/T(B)$, and additional demand requires additional hardware.

4.2 Operational Carbon Modeling

The daily energy generated by the solar array, O_{so}, is computed using Equation 15, and the daily energy processed by the battery is given by Equation 16, where $\mathrm{cap}_{\mathrm{ba}}$ denotes the battery capacity, and DoD indicates the depth of discharge. Solar generation must satisfy two requirements: (i) powering the computing and communication subsystems during sunlit periods and (ii) recharging the battery within the available sunlight duration T_{lit}. Accordingly, the battery must provide sufficient usable capacity to sustain subsystem operation throughout each eclipse period.

5 Evidence and Evaluation Scope

This section evaluates two limited forms of evidence for *LLMSpace*. First, component-level checks compare selected model inputs and intermediate predictions with independently reported values. These checks assess parameter plausibility, but do not validate subsystem aggregation or a complete satellite inventory. Second, a controlled comparison with EIR (Ohs et al., 2025) and NE (Aili et al., 2025) evaluates component coverage under a common COTS Starlink-V1 + DGX-H100 configuration. Because the constructed reference and compared models share some literature-derived inputs, this comparison assesses coverage and arithmetic consistency rather than predictive accuracy. Neither form of evidence validates launch intensity, and radiation-hardening overheads. Public data do not provide matched COTS and radiation-tolerant H100-class measurements of fabrication intensity, logic area, and system power. Section 6 therefore treats these quantities as exploratory variables and reports the conditions under which orbital and terrestrial carbon rankings change.

Table 6: Component-level parameter plausibility checks. Differences ([diff]) are relative to the external values.

Quantity	External	**LLMSpace**
Falcon 9 I_L ($\mathrm{kgCO_2e/kg}$)	12.85	14.5 [+12.8%]
Si I_{sol} ($\mathrm{kgCO_2e/m^2}$)	118	120 [+1.7%]
aluminum I_{rad} ($\mathrm{kgCO_2e/kg}$)	18.4	15.0 [−18.5%]
NMC I_{bat} ($\mathrm{kgCO_2e/kWh}$)	74	80 [+8.1%]

Component-Level Plausibility Checks. Table 6 compares selected *LLMSpace* coefficients with independent external values. For launch, the Falcon 9 payload-normalized intensity I_L is evaluated using published emissions and payload-capacity data (293 $\mathrm{tCO_2e}$ for a 22.8 t payload) (Osoro et al., 2026). For Si solar arrays, Clemente et al. (2024) reports an embodied carbon intensity of $118\ \mathrm{kgCO_2e/m^2}$, compared with $120\ \mathrm{kgCO_2e/m^2}$ used in Equation 3. The radiator coefficient is compared with an independent primary-aluminum carbon intensity (Paraskevas et al., 2016). Finally, the NMC battery coefficient is compared with a reported median cradle-to-gate carbon intensity (Peiseler et al., 2024). These comparisons assess the plausibility of individual model coefficients and do not constitute full-system validation.

Controlled Comparison with Prior Work. We compare *LLMSpace* with EIR (Ohs et al., 2025) and NE (Aili et al., 2025) under a common COTS Starlink-V1 configuration comprising a DGX

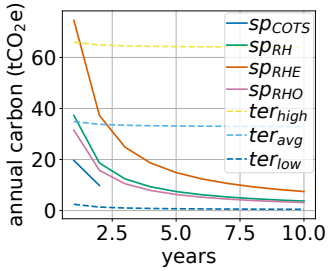

Figure 2: The comparison of average annual carbon between orbital and terrestrial inferences.

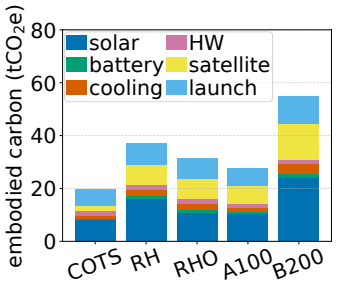

Figure 3: The embodied carbon breakdown of various hardware configurations.

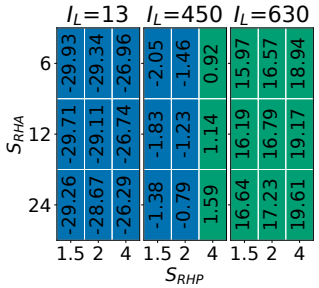

Figure 4: ΔC sensitivity study of S_{RHA}, S_{RHP} and I_L (blue: $\Delta C < 0$; green: $\Delta C > 0$).

H100 node, Si solar array, NMC battery, and aluminum-honeycomb radiator. We assume a 10 kW compute-power budget (Newkirk et al., 2025) and a two-year service life. The DGX H100 contains eight 4 nm H100 GPUs, and its manufacturing embodied carbon is derived in Appendix C. To isolate differences in model coverage, all three frameworks use identical component quantities, carbon-intensity coefficients, launch intensity, hardware assumptions, and mission lifetime. As shown in Table 7, EIR omits the radiator and NE omits the solar array, whereas *LLMSpace* includes both. Relative to the constructed component-sum reference, EIR, NE, and *LLMSpace* differ by -6.69%, -43.7%, and $+2.95\%$, respectively. The smaller absolute deviation of *LLMSpace* reflects broader component coverage under controlled inputs, not greater empirical accuracy. We also evaluate an illustrative radiation-hardened (rad-hard) configuration for a ten-year service lifetime using a 7 nm radiation-hardening technology (Maillard et al., 2024). We set $S_{\mathrm{RHA}} = 12$ in Equation 12 and $S_{\mathrm{RHP}} = 2$ in Equation 13, treating both as scenario assumptions rather than empirically validated overheads. Accordingly, the compute-power budget increases from 10 to 20 kW, which increases the required solar-array, battery, and radiator capacities. Appendix D derives the constructed external reference inventory.

Scope of Claims. Results support only the consistency of selected component inputs with independent references and the broader component coverage of *LLMSpace* relative to prior work under controlled COTS assumptions. It does not validate absolute carbon estimates, a complete LEO LLM-inference system, radiation-hardening overheads, or hypothetical hardware feasibility and qualification.

6 Scenario Analysis and Use Cases

We use *LLMSpace* to address: (1) under what combinations of radiation-hardening manufacturing-carbon and power overheads does an orbital data center achieve lower embodied carbon than an equivalent terrestrial data center? (2) how do LLM workload characteristics and GPU choice affect inference throughput, operational energy, and amortized embodied carbon per token?

6.1 Comparing Orbital and Terrestrial Data Centers

Orbital vs. Terrestrial Data Centers. Figure 2 compares the annual carbon of orbital and terrestrial data centers. For terrestrial computing, we calculate

$$C_{\mathrm{annual,terr}} = \mathrm{PUE} \times \mathrm{CI} \times E_{\mathrm{hard,op}} + C_{\mathrm{hard,em}}, \tag{17}$$

where PUE is power usage effectiveness, CI is grid carbon intensity, $E_{\mathrm{hard,op}}$ is annual compute-system energy consumption, and $C_{\mathrm{hard,em}}$ is annualized hardware manufacturing embodied carbon. We assume $\mathrm{PUE} = 1.1$ and a 10 kW compute-power budget. Based on Google's 2026 Environmental Report (Google, 2026), the low-, average-, and high-carbon terrestrial scenarios, denoted by $\mathrm{ter}_{\mathrm{low}}$, $\mathrm{ter}_{\mathrm{avg}}$, and $\mathrm{ter}_{\mathrm{high}}$, use grid intensities of 3, 348, and 679 $\mathrm{gCO_2/kWh}$, respectively. The low-carbon scenario approximates a data center powered almost entirely by carbon-free electricity, such as Google's Stockholm region. Grid carbon intensity therefore dominates the variation among the terrestrial scenarios. For orbital computing, we evaluate COTS ($\mathrm{sp}_{\mathrm{COTS}}$) and radiation-hardened ($\mathrm{sp}_{\mathrm{RH}}$) configurations launched by Falcon 9. To illustrate peripheral optimization, we replace the Si solar array in $\mathrm{sp}_{\mathrm{RH}}$ with a higher-efficiency multi-junction array. The resulting configuration, denoted by $\mathrm{sp}_{\mathrm{RHO}}$, requires less array area and mass, thereby reducing both manufacturing embodied carbon and launch emissions (Figure 3). We also define an extended stress-test configuration, $\mathrm{sp}_{\mathrm{RHE}}$, with $S_{\mathrm{RHA}} = 24$ and $S_{\mathrm{RHP}} = 4$, twice the values assumed for $\mathrm{sp}_{\mathrm{RH}}$. These factors are sensitivity

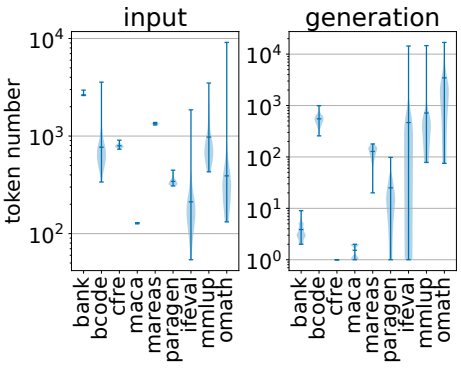

Figure 5: Prompt lengths and Generated tokens of selected LLM workloads.

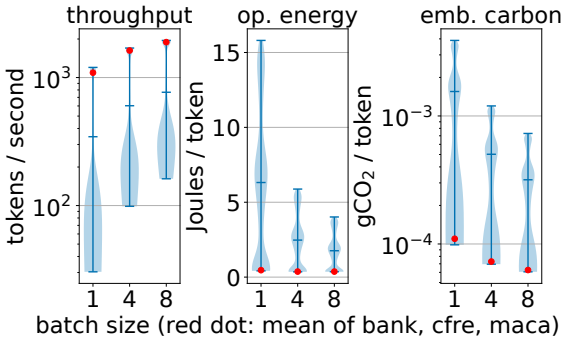

Figure 6: Inference throughput, operational energy, and embodied carbon of various batch sizes on an H100 GPU.

settings rather than empirical bounds on radiation-hardened hardware. Despite its two-year service lifetime, $\mathrm{sp}_{\mathrm{COTS}}$ has lower average annual total carbon than $\mathrm{ter}_{\mathrm{avg}}$. For $\mathrm{sp}_{\mathrm{RH}}$, annualized embodied carbon decreases with service lifetime, bringing its average annual total carbon below $\mathrm{ter}_{\mathrm{avg}}$ and closer to $\mathrm{ter}_{\mathrm{low}}$. Nevertheless, $\mathrm{ter}_{\mathrm{low}}$ remains the lowest-carbon scenario, showing that terrestrial computing powered by nearly carbon-free electricity can outperform orbital computing. Even under the more conservative $\mathrm{sp}_{\mathrm{RHE}}$ assumptions, orbital computing falls below $\mathrm{ter}_{\mathrm{avg}}$ after three years, indicating that its carbon advantage can persist under substantially larger assumed radiation-hardening manufacturing and power overheads.

Uncertain-Parameter Crossover Analysis. We consider a Starlink-V1 platform carrying a hypothetical radiation-hardened (RH) DGX H100 system and launched by Falcon 9. The principal uncertainties are the launch carbon intensity I_{L}, radiation-hardening manufacturing-carbon factor S_{RHA}, and compute-system power factor S_{RHP}; the solar-array, battery, and radiator parameters are comparatively well constrained by component-level data. Because launch emissions constitute a substantial fraction of embodied carbon (Figure 3), we jointly vary $S_{\mathrm{RHA}} \in \{6, 12, 24\}$, $S_{\mathrm{RHP}} \in \{1.5, 2, 4\}$, and $I_{\mathrm{L}} \in \{13, 450, 630\}$ $\mathrm{kgCO_2e/kg}$. Appendix A defines the low, intermediate, and high launch-intensity scenarios, which span direct propellant-combustion emissions to broader process-based accounting. Appendix B defines the hardening factors and their sensitivity ranges; $S_{\mathrm{RHA}} = 12$ and $S_{\mathrm{RHP}} = 2$ are the nominal assumptions for migrating from a 4 nm COTS H100 GPU to a hypothetical 7 nm RH implementation, while the remaining values represent lower- and higher-overhead scenarios. Figure 4 compares the 10-year average annual total carbon of $\mathrm{sp}_{\mathrm{RH}}$, denoted by C_{sp}, with that of $\mathrm{ter}_{\mathrm{avg}}$ at CI $= 348$ $\mathrm{gCO_2e/kWh}$, denoted by C_{terr}. We define $\Delta C = C_{\mathrm{sp}} - C_{\mathrm{terr}}$, where $\Delta C < 0$ favors orbital computing and $\Delta C > 0$ favors terrestrial computing. Among the evaluated parameters, S_{RHA} has the smallest effect because GPU manufacturing contributes a limited fraction of total carbon. The effect of S_{RHP} is larger because higher compute power increases the required solar-array, battery, and radiator capacities, thereby increasing both peripheral manufacturing and launch emissions. Nevertheless, I_{L} dominates the crossover. At $I_{\mathrm{L}} = 13$ $\mathrm{kgCO_2e/kg}$, all evaluated hardening configurations yield $\Delta C < 0$; at $I_{\mathrm{L}} = 630$ $\mathrm{kgCO_2e/kg}$, all yield $\Delta C > 0$. At the intermediate value of 450 $\mathrm{kgCO_2e/kg}$, the outcome depends on the hardening overheads: lower S_{RHA} and S_{RHP} preserve the orbital advantage, whereas higher values favor terrestrial computing. Thus, I_{L} is the dominant crossover uncertainty, followed by S_{RHP}.

6.2 LLM Workload and Hardware Tradeoffs

Workload-Specific Throughput and Carbon. We evaluate nine HELM tasks: banking77 (*bank*), bigcodebench (*bcode*), cleva:coreference_resolution (*cfre*), cleva:mathematical_calculation (*maca*), cleva:mathematical_reasoning (*mareas*), cleva:paraphrase_generation (*paragen*), ifeval (*ifeval*), mmlu_pro (*mmlup*), and omni_math (*omath*) (Liang et al., 2022). Figure 5 reports the minimum, mean, and maximum prompt and output lengths. We run Qwen2.5-32B-Instruct-GPTQ-Int8 on a single NVIDIA A100 (Falk et al., 2026), H100 (NVIDIA Corporation, 2025b), or B200 (NVIDIA Corporation, 2025a) GPU with batch sizes of 1–8. Throughput and energy are measured using `nvidia-smi` at batch sizes 1, 4, and 8. Each experiment is repeated three times, and the mean is reported. The measurements calibrate Vidur-Energy (Özcan et al., 2025) to estimate unmeasured configurations. Figure 6 reports H100 throughput, operational energy, and amortized embodied carbon. Prefill is primarily compute-bound, whereas decoding is memory-bandwidth-bound and

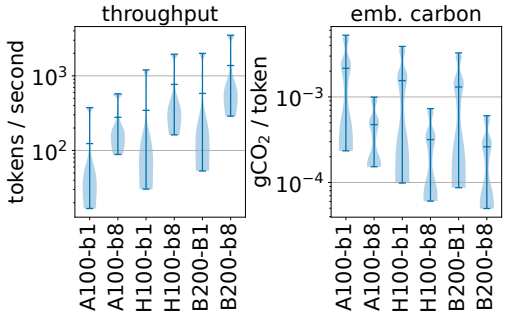

Figure 7: Inference throughput and embodied carbon comparison of various batch sizes between A100, H100, and B200.

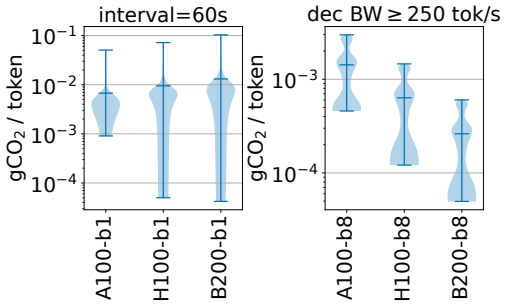

Figure 8: Inference embodied carbon under various performance requirements on different GPUs including A100, H100, and B200.

dominates long-output workloads. Because *bank*, *cfre*, and *maca* generate fewer tokens, they achieve higher throughput and lower energy per token. For most workloads, throughput remains below $500\ \mathrm{tokens/s}$ and energy below $8\ \mathrm{J/token}$. Communication consumes $\sim 0.5\ \mu\mathrm{J/bit}$ (Ohs et al., 2025) and is negligible relative to inference; computation therefore dominates operational energy. Larger batches improve utilization and reduce energy per token. We amortize embodied carbon over its 10-year lifetime and compute C_{t} using Equation 14. Under sustained demand, lower batch latency and higher token throughput reduce the amortized embodied carbon per token.

Throughput–Embodied-Carbon Tradeoff. We examine the effect of GPU selection by replacing the radiation-hardened (RH) DGX H100 node with hypothetical RH DGX A100 (Falk et al., 2026) and RH DGX B200 (NVIDIA Corporation, 2025a) nodes. All configurations assume a 10-year service lifetime, $S_{\mathrm{RHA}} = 12$, and $S_{\mathrm{RHP}} = 2$. As shown in Figure 3, the RH A100 system consumes $\sim 13\ \mathrm{kW}$, permitting smaller power and thermal-control subsystems and reducing total embodied carbon by 27.68% relative to the RH H100 configuration. In contrast, the RH B200 system consumes $\sim 30\ \mathrm{kW}$, requiring larger power and thermal-control subsystems and increasing embodied carbon by 43.37%. Because LEO inference is powered entirely by the solar arrays, this analysis focuses on inference throughput and embodied carbon rather than operational carbon. Figure 7 compares the inference throughput and embodied carbon per token of the A100, H100, and B200 configurations across batch sizes. Each successive GPU generation substantially improves inference throughput. Although newer GPUs require greater system-level embodied carbon, their higher throughput can reduce the embodied carbon attributed to each token under sustained utilization. This advantage assumes that sufficient requests are continuously available to utilize the GPU and that system demand can benefit from its maximum throughput. Under intermittent or constrained workloads, the relative advantage changes. When only one request arrives within each 60-s interval, as shown in Figure 8, the B200 has the highest embodied carbon per token, followed by the H100, because the time-amortized embodied carbon accrues regardless of GPU utilization (Equation 14). Conversely, when a minimum decoding throughput of $250\ \mathrm{tokens/s}$ is required, three A100 GPUs or two H100 GPUs are needed to satisfy the constraint, whereas one B200 GPU is sufficient. The replication required by the older GPUs increases their system-level embodied carbon per token; consequently, the three-A100 configuration incurs the highest embodied carbon among the evaluated alternatives.

7 Conclusion

This paper presented *LLMSpace*, a component-level framework for evaluating the carbon and performance of LLM inference on AI-enabled LEO satellites. *LLMSpace* integrates launch emissions, power and thermal subsystems, radiation-hardened hardware, and workload-dependent inference behavior. Across the evaluated scenarios, launch carbon intensity is the primary determinant of the orbital–terrestrial crossover, with radiation-hardening power overhead providing an important secondary effect. GPU and workload studies further show that inference throughput, batching, request rate, and workload characteristics influence energy and the amortization of embodied carbon per token. Higher throughput can reduce per-token embodied carbon under sustained utilization, while this benefit diminishes under sparse demand or throughput constraints. Overall, LEO inference is not inherently lower-carbon; its carbon benefit depends on launch accounting, mission lifetime, terrestrial grid intensity, hardware efficiency, workload characteristics, and system utilization.

AI USE DISCLOSURE

We used generative AI tools only to assist with the presentation of this work. Specifically, we used generative AI to improve wording, grammar, readability, and organization of portions of the manuscript, and to help rewrite and format latex code for text, equations, tables, and related manuscript elements. We did not use generative AI to generate experimental data, develop the underlying *LLMSpace* methodology, formulate the main research hypotheses, perform the carbon-footprint calculations, conduct the experiments, or determine the reported results and conclusions. All AI-assisted text and latex modifications were reviewed and edited by the authors for technical correctness and consistency with the underlying analysis. The authors take full responsibility for the final content of the paper, including all text and artifacts produced with the assistance of AI.

ETHICS STATEMENT

This work studies the energy and carbon implications of LLM inference on LEO satellite platforms and does not involve human subjects, personal data, or user studies. Our analysis relies on published hardware specifications, carbon data, and benchmark workloads. The radiation-hardened hardware configurations considered in this work are hypothetical design scenarios rather than representations of currently deployed systems. Because carbon estimates depend on modeling boundaries and uncertain parameters, including launch carbon intensity, hardware manufacturing, mission lifetime, and terrestrial grid carbon intensity, we explicitly report these assumptions and evaluate their sensitivity. The purpose of *LLMSpace* is to support transparent comparison of the environmental trade-offs of terrestrial and space-based LLM inference, rather than to advocate deployment in either setting. We are not aware of additional concerns related to privacy, security, discrimination, or other human-subject risks arising from this work.

REPRODUCIBILITY STATEMENT

We have taken several steps to support reproducibility of the results in this work. The main text specifies the carbon-accounting methodology, system assumptions, workload definitions, and experimental setup used in *LLMSpace*, while the appendices provide additional derivations, parameter choices, component inventories, and sensitivity-analysis settings. We also provide the source code used to implement the *LLMSpace* model and reproduce the numerical analyses, tables, and figures as supplementary material. The supplementary code includes the model parameters and calculation procedures used for component-level embodied carbon, launch emissions, power and thermal provisioning, radiation-hardening scenarios, and workload-dependent inference evaluation. Together, the paper, appendices, and supplementary code are intended to enable readers to reproduce the reported results and evaluate alternative assumptions and system configurations.

REFERENCES

Ablimit Aili, Jihwan Choi, Yew Soon Ong, and Yonggang Wen. The Development of Carbon-Neutral Data Centres in Space. *Nature Electronics*, pp. 1–11, 2025.

Amazon. Amazon Leo: Low Earth Orbit (LEO) Satellite Network. `https://www.aboutamazon.com/what-we-do/devices-services/project-kuiper`, 2025.

AMD. Zen 5. `https://en.wikipedia.org/wiki/Zen_5`, May 2026.

Connor R Barker, Eloise A Marais, and Jonathan C McDowell. Global 3D Rocket Launch and Re-entry Air Pollutant and CO2 Emissions at the Onset of the Megaconstellation Era. *Scientific data*, 11(1):1079, 2024.

Ismail Bayram, Enes Eken, Donald Kline, Nikolas Parshook, Yiran Chen, and Alex K Jones. Modeling STT-RAM Fabrication Cost and Impacts in NVSim. In *IEEE Seventh International Green and Sustainable Computing Conference*, pp. 1–8, 2016.

Nathaniel Bleier, Muhammad Husnain Mubarik, Gary R Swenson, and Rakesh Kumar. Space Microdatacenters. In *IEEE/ACM International Symposium on Microarchitecture*, pp. 900–915, 2023.

Tyler F. M. Brown, Michele T. Bannister, Laura E. Revell, Timofei Sukhodolov, and Eugene Rozanov. Worldwide Rocket Launch Emissions 2019: An Inventory for Use in Global Models. *Earth and Space Science*, 11(10):e2024EA003668, 2024.

Guillermo J Dominguez Calabuig, Andrew Wilson, Sifeng Bi, Massimiliano Vasile, Martin Sippel, and Martin Tajmar. Environmental life cycle assessment of reusable launch vehicle fleets: Large climate impact driven by rocket exhaust emissions. *Acta Astronautica*, 221:1–11, 2024.

Megan C Casey, Justin S Goodwill, Edward J Wyrwas, Rebekah A Austin, Christopher M Wilson, Scott D Stansberry, Nicolas Gorius, and Shahid Aslam. Single-Event Effects on Commercial-Off-The-Shelf Edge-Processing Artificial Intelligence ASICs. *IEEE Transactions on Nuclear Science*, 70(8):1716–1723, 2023.

Augustin Chanoine, Y Le Guern, F Witte, J Huesing, T Soares, and L Innocenti. Integrating Sustainability in the Design of Space Activities: Development of Eco-Design Tools for Space Projects. In *5th CEAS Air & Space Conference*, pp. 1–11. Delft University of Technology, 2015.

Aizaz U Chaudhry and Halim Yanikomeroglu. Laser Intersatellite Links in a Starlink Constellation: A Classification and Analysis. *IEEE vehicular technology magazine*, 16(2):48–56, 2021.

Maurizio Clemente, Luuk Van Sundert, Mauro Salazar, and Theo Hofman. A framework to estimate life cycle emissions for vehicle-integrated photovoltaic systems. In *IEEE Vehicle Power and Propulsion Conference (VPPC)*, pp. 1–6, 2024.

Riccardo De Gaudenzi. Satellite networks: Past present and future challenges-part 2: Broadband systems. *IEEE Aerospace and Electronic Systems Magazine*, 2025.

Ahmad Faiz, Sotaro Kaneda, Ruhan Wang, Rita Chukwunyere Osi, Prateek Sharma, Fan Chen, and Lei Jiang. LLMCarbon: Modeling the End-to-End Carbon Footprint of Large Language Models. In *The Twelfth International Conference on Learning Representations*, 2024.

Sophia Falk, David Ekchajzer, Thibault Pirson, Etienne Lees-Perasso, Augustin Wattiez, Lisa Biber-Freudenberger, Sasha Luccioni, and Aimee van Wynsberghe. More than carbon: Cradle-to-grave environmental impacts of genai training on the nvidia a100 gpu. *Environmental Impact Assessment Review*, 121:108525, 2026.

Zhenxiao Fu, Fan Chen, Shan Zhou, Haitong Li, and Lei Jiang. LLMCO2: Advancing accurate carbon footprint prediction for LLM inferences. *ACM SIGENERGY Energy Informatics Review*, 5 (2):63–68, 2025.

David G. Gilmore. *Spacecraft Thermal Control Handbook, Volume I: Fundamental Technologies*. The Aerospace Press, El Segundo, CA, 2002. ISBN 978-1884989033.

Google. Exploring a Space-Based, Scalable AI Infrastructure System Design: Project Suncatcher. https://research.google/blog/exploring-a-space-based-scalable-ai-infrastructure-system-design/, 2025. Google Research Blog.

Google. 2026 environmental report. Technical report, Google LLC, June 2026. URL https://sustainability.google/files/google-2026-environmental-report.pdf.

Mark Gregory, Frank Heine, Andrej Brzoska, Klaus Oestreich, Robert Mahn, Patricia Martin Pimentel, and Herwig Zech. Status on Laser Communication Activities at Tesat-Spacecom. In *Free-Space Laser Communications XXXVI*, volume 12877, pp. 1287704. SPIE, 2024.

M.M Hasan, Rezwanul Haque, M.I Jahirul, Mohammad G Rasul, I.M.R Fattah, N.M.S Hassan, and M Mofijur. Advancing Energy Storage: The Future Trajectory of Lithium-Ion Battery Technologies. *Journal of Energy Storage*, 120:116511, 2025.

Nidhal Jegham, Marwan Abdelatti, Chan Young Koh, Lassad Elmoubarki, and Abdeltawab Hendawi. How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference. *arXiv preprint arXiv:2505.09598*, 2025.

Seungjune Jeon, BVK Vijaya Kumar, Euiseok Hwang, and Michael K Cheng. Evaluation of error correcting codes for radiation-tolerant memory. *The Interplanetary Network Progress Report*, 42: 181, 2010.

Aryan Kannaujiya and Ambika Prasad Shah. Radiation Effects in VLSI Circuits–Part I: Historical Perspective. *IETE Technical Review*, 41(6):716–735, 2024.

Vaclav Knap, Lars Kjeldgaard Vestergaard, and Daniel-Ioan Stroe. A Review of Battery Technology in Cubesats and Small Satellite Solutions. *Energies*, 13(16):4097, 2020.

Y Lakys, WS Zhao, JO Klein, and C Chappert. Hardening Techniques for MRAM-based Non-volatile Storage Cells and Logic. In *IEEE European Conference on Radiation and Its Effects on Components and Systems*, pp. 669–674, 2011.

Gyeongyeop Lee, Minki Suh, Minsang Ryu, Yunjong Lee, Jin-Woo Han, and Jungsik Kim. Investigation into the Degradation of DDR4 DRAM Owing to Total Ionizing Dose Effects. *IEEE Access*, 11:97456–97465, 2023.

Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, et al. Holistic Evaluation of Language Models. *arXiv preprint arXiv:2211.09110*, 2022.

Atse Louwen, Wilfried Van Sark, Ruud Schropp, and André Faaij. A Cost Roadmap for Silicon Heterojunction Solar Cells. *Solar Energy Materials and Solar Cells*, 147:295–314, 2016.

Pierre Maillard, Abhijitt Dhavlle, Yanran Paula Chen, Nicholas Fraser, Yushan Chen, Nicholas Vacirca, Jeff Barton, Martin L. Voogel, and Minal Sawant. Protons evaluation of 7nm versal ai engine (aie) based radiation tolerant platform for deep learning applications. In *IEEE Radiation Effects Data Workshop*, pp. 1–8, 2024.

Matthew J Marinella. Radiation Effects in Advanced and Emerging Nonvolatile Memories. *IEEE Transactions on Nuclear Science*, 68(5):546–572, 2021.

Mercury Systems. RH3480: 480GB 3U VPX Rad-Tolerant SSDs. https://www.mrcy.com/products/data-management/rad-tolerant-storage-memory/model-rh3480, 2026.

Microsoft. New azure space products enable digital resiliency and empower the industry. https://azure.microsoft.com/en-us/blog/new-azure-space-products-enable-digital-resiliency-and-empower-the-industry/, 2025.

Alex C Newkirk, Jared Fernandez, Jonathan Koomey, Imran Latif, Emma Strubell, Arman Shehabi, and Constantine Samaras. Empirically-Calibrated H100 Node Power Models for Accurate AI Training Energy Estimation. *Environmental Research: Energy*, 2(4):045016, 2025.

Chenxu Niu, Wei Zhang, Yongjian Zhao, and Yong Chen. Energy efficient or exhaustive? benchmarking power consumption of llm inference engines. *ACM SIGENERGY Energy Informatics Review*, 5(2):56–62, 2025.

NVIDIA Corporation. NVIDIA DGX H100/H200 User Guide. https://docs.nvidia.com/dgx/dgxh100-user-guide/introduction-to-dgxh100.html, 2023.

NVIDIA Corporation. Product Carbon Footprint (PCF) Summary for NVIDIA HGX B200. https://images.nvidia.com/aem-dam/Solutions/documents/HGX-B200-PCF-Summary.pdf, July 2025a.

NVIDIA Corporation. Product Carbon Footprint (PCF) Summary for NVIDIA HGX H100. https://images.nvidia.com/aem-dam/Solutions/documents/HGX-H100-PCF-Summary.pdf, July 2025b.

Robin Ohs, Gregory F Stock, Andreas Schmidt, Juan A Fraire, and Holger Hermanns. Dirty Bits in Low-Earth Orbit (LEO): The Carbon Footprint of Launching Computers. *ACM SIGENERGY Energy Informatics Review*, 5(2):26–33, 2025.

O. B. Osoro, E. J. Oughton, A. R. Wilson, and A. Rao. Emissions Assessment of Low Earth Orbit Broadband Megaconstellations: Starlink, OneWeb and Kuiper. *Advances in Space Research*, 77 (11):11001–11013, 2026.

Miray Özcan, Philipp Wiesner, Philipp Weiß, and Odej Kao. Quantifying the Energy Consumption and Carbon Emissions of LLM Inference via Simulations. *arXiv preprint arXiv:2507.11417*, 2025.

P Paillet, M Gaillardin, V Ferlet-Cavrois, A Torres, O Faynot, C Jahan, L Tosti, and S Cristoloveanu. Total Ionizing Dose Effects on Deca-Nanometer Fully Depleted SOI Devices. *IEEE transactions on nuclear science*, 52(6):2345–2352, 2006.

Scott Palo, Darren O'Connor, Elizabeth DeVito, Rick Kohnert, Scott H Schaire, Steve Bundick, Gary Crum, Serhat Altunc, and Thomas Winkert. Expanding CubeSat Capabilities with A Low Cost Transceiver. In *Annual AIAA/USU Conference on Small Satellites*, 2014.

D. Paraskevas, K. Kellens, A. Van de Voorde, W. Dewulf, and J. R. Duflou. Environmental Impact Analysis of Primary Aluminium Production at Country Level. In *Procedia CIRP*, volume 40, pp. 209–213, 2016.

L. Peiseler et al. Carbon Footprint Distributions of Lithium-Ion Batteries and Their Materials. *Nature Communications*, 15:10301, 2024.

Konstantin F Pilz, James Sanders, Robi Rahman, and Lennart Heim. Trends in AI Supercomputers. *arXiv preprint arXiv:2504.16026*, 2025.

Fabricio S Prol, R Morales Ferre, Zainab Saleem, Petri Välisuo, Christina Pinell, Elena Simona Lohan, Mahmoud Elsanhoury, Mohammed Elmusrati, Saiful Islam, and Kaan Çelikbilek. Position, Navigation, and Timing (PNT) through Low Earth Orbit (LEO) Satellites: A Survey on Current Status, Challenges, and Opportunities. *IEEE Access*, 10:83971–84002, 2022.

K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer cmos circuits. *Proceedings of the IEEE*, 91(2): 305–327, 2003.

Windy S Slater, Nayana P Tiwari, Tyler M Lovelly, and Jesse K Mee. Total Ionizing Dose Radiation Testing of NVIDIA Jetson Nano GPUs. In *IEEE High Performance Extreme Computing Conference*, pp. 1–3, 2020.

J.R. Srour and J.W. Palko. Displacement Damage Effects in Irradiated Semiconductor Devices. *IEEE Transactions on Nuclear Science*, 60(3):1740–1766, 2013.

J.R. Srour, C.J. Marshall, and P.W. Marshall. Review of Displacement Damage Effects in Silicon Devices. *IEEE Transactions on Nuclear Science*, 50(3):653–670, 2003.

StarCloud. How Starcloud Is Bringing Data Centers to Outer Space. https://blogs.nvidia.com/blog/starcloud/, 2025. NVIDIA Blog.

Meiqi Wang, Han Qiu, Longnv Xu, Di Wang, Yuanjie Li, Tianwei Zhang, Jun Liu, and Hewu Li. A Case for Application-Aware Space Radiation Tolerance in Orbital Computing. *arXiv preprint arXiv:2407.11853*, 2024.

Olivier Weppe, Thibaut Marty, Sébastien Toussaint, Nicolas Brusselmans, Jean-Christophe Prévotet, Jean-Pierre Raskin, and Maxime Pelcat. Embodied carbon footprint of 3d nand memories. In *ACM International Conference on Computing Frontiers: Workshops and Special Sessions*, pp. 108–116, 2025.

Geoffrey Yeap, SS Lin, YM Chen, HL Shang, PW Wang, HC Lin, YC Peng, JY Sheu, M Wang, X Chen, et al. 5nm cmos production technology platform featuring full-fledged euv, and high mobility channel finfets with densest 0.021 μm 2 sram cells for mobile soc and high performance computing applications. In *IEEE International Electron Devices Meeting*, pp. 36–7, 2019.

Suge Yue, Xiaolin Zhang, Yuanfu Zhao, Lin Liu, and Hanning Wang. Modeling and Simulation of Single-Event Effect in CMOS Circuit. *Journal of Semiconductors*, 36(11):111002, 2015.

A Launch Carbon Intensity and Sensitivity Scenarios

Published launch-carbon estimates vary substantially with launch vehicle, payload utilization, and accounting boundary. For consistency, we define the capacity-normalized launch carbon intensity as

$$I_{\mathrm{L}} = \frac{C_{\mathrm{launch}}}{M_{\mathrm{LEO}}}, \tag{18}$$

where C_{launch} is the carbon attributed to one launch under a specified accounting boundary and M_{LEO} is the vehicle's maximum payload capacity to low Earth orbit (LEO). Equation 18 assumes full utilization of the stated payload capacity; consequently, it represents a lower per-kilogram allocation for the corresponding launch and accounting boundary. For a partially utilized launch, the actual manifested payload should replace M_{LEO}. Depending on the stated boundary, I_{L} may represent either direct propellant-combustion emissions or a broader process-based estimate.

Direct launch-event emissions. We first estimate C_{launch} using representative propellant inventories (Osoro et al., 2026) and direct CO_2 emission factors (Brown et al., 2024): $0.6\ \mathrm{kgCO_2/kg}$ for RP-1/LOX propellant, $0.3\ \mathrm{kgCO_2/kg}$ for HTPB-based solid propellant, $0.2\ \mathrm{kgCO_2/kg}$ for hypergolic propellant, and zero direct CO_2 for LH_2/LOX propellant. These factors quantify only the CO_2 released during launch and exclude upstream propellant-production emissions and the non-CO_2 climate effects of rocket exhaust. Because CO_2 has a characterization factor of one, these direct CO_2 values are numerically equivalent to their CO_2-equivalent contributions.

- **Falcon 9.** A Falcon 9 consumes approximately $488{,}370\ \mathrm{kg}$ of RP-1/LOX propellant per launch (Osoro et al., 2026). Its estimated direct launch emissions are

$$C_{\mathrm{launch}}^{\mathrm{F9}} = 488{,}370 \times 0.6 = 293{,}022\ \mathrm{kgCO_2}. \tag{19}$$

  With a maximum LEO payload capacity of $22{,}800\ \mathrm{kg}$, the corresponding capacity-normalized intensity is

$$I_{\mathrm{L}}^{\mathrm{F9}} = \frac{293{,}022}{22{,}800} \approx 12.85\ \mathrm{kgCO_2/kg}. \tag{20}$$

- **Soyuz-FG.** A Soyuz-FG consumes approximately $218{,}150\ \mathrm{kg}$ of RP-1/LOX propellant and $7{,}360\ \mathrm{kg}$ of hypergolic propellant per launch (Osoro et al., 2026). Its estimated direct emissions are

$$C_{\mathrm{launch}}^{\mathrm{Soyuz}} = 218{,}150 \times 0.6 + 7{,}360 \times 0.2 \approx 132{,}362\ \mathrm{kgCO_2}. \tag{21}$$

  Using a maximum LEO payload capacity of approximately $7{,}100\ \mathrm{kg}$ gives

$$I_{\mathrm{L}}^{\mathrm{Soyuz}} = \frac{132{,}362}{7{,}100} \approx 18.64\ \mathrm{kgCO_2/kg}. \tag{22}$$

- **Ariane 5.** An Ariane 5 consumes approximately $480{,}000\ \mathrm{kg}$ of solid propellant, $184{,}900\ \mathrm{kg}$ of LH_2/LOX propellant, and $10{,}000\ \mathrm{kg}$ of hypergolic propellant per launch (Osoro et al., 2026). Its estimated direct emissions are

$$C_{\mathrm{launch}}^{\mathrm{A5}} = 480{,}000 \times 0.3 + 10{,}000 \times 0.2 = 146{,}000\ \mathrm{kgCO_2}. \tag{23}$$

  Using a maximum LEO payload capacity of approximately $21{,}000\ \mathrm{kg}$ gives

$$I_{\mathrm{L}}^{\mathrm{A5}} = \frac{146{,}000}{21{,}000} \approx 6.95\ \mathrm{kgCO_2/kg}. \tag{24}$$

The capacity-normalized direct-emission estimates for these vehicles range from approximately 6.95 to $18.64\ \mathrm{kgCO_2/kg}$. These estimates exclude propellant production, launch-vehicle manufacturing, transportation, ground infrastructure and operations, vehicle recovery and refurbishment, and re-entry. Rocket exhaust also contains H_2O, black carbon, NO_x, alumina, and other species with potential climate effects (Brown et al., 2024). These non-CO_2 effects are not converted into CO_2-equivalent emissions in the present direct-emission estimates.

Broader process-based estimate. Calabuig et al. (2024) report a GWP100 impact of $4.71\ \mathrm{ktCO_2e}$ per launch for a reusable LH_2/LOX launch vehicle capable of delivering $7.5\ \mathrm{ton}$ to orbit. Normalizing the reported impact by payload capacity yields

$$I_{\mathrm{L}}^{\mathrm{process}} = \frac{4.71 \times 10^6}{7{,}500} = 628\ \mathrm{kgCO_2e/kg}. \tag{25}$$

Table 8: Launch-carbon intensity scenarios used in the sensitivity analysis. The values represent different accounting assumptions and should not be interpreted as universal empirical bounds.

Scenario	I_{L} ($\mathrm{kgCO_2e/kg}$)	Accounting basis	Role in analysis
Reference	12.85 (≈ 13)	Falcon 9 direct emissions	Reference launch assumption
Intermediate	450	Exploratory sensitivity value	Broader-boundary stress test
High	628 (≈ 630)	Process-based GWP100 estimate	Broadest evaluated boundary

Unlike the preceding combustion-only estimates, this process-based value includes launch-vehicle and propellant production, transportation, ground and launch operations, recovery, and refurbishment. The difference between this estimate and the direct-emission values therefore reflects not only vehicle characteristics but also a substantially broader accounting boundary.

Sensitivity scenarios. Because the available estimates differ in both magnitude and system boundary, we treat I_{L} as an uncertain scenario parameter rather than a universal vehicle constant. Table 8 summarizes the three values evaluated in the sensitivity analysis.

The reference scenario adopts the Falcon 9 direct-emission estimate because Falcon 9 is the launch vehicle assumed in our system evaluation; this value is not the minimum among all evaluated vehicles. The intermediate value of $450\ \mathrm{kgCO_2e/kg}$ is an exploratory setting between the reference and process-based scenarios. It is included to examine the transition between these accounting regimes and is not attributed to a particular vehicle or empirical study. The high scenario adopts the process-based estimate of $630\ \mathrm{kgCO_2e/kg}$ reported by Calabuig et al. (2024). Collectively, these scenarios assess whether the orbital–terrestrial total-carbon comparison remains robust as the assumed launch boundary expands from direct launch emissions to broader launch-system contributions.

B Radiation-Hardening Sensitivity Scenarios

No publicly characterized H100-class GPU provides matched COTS and radiation-hardened measurements of fabrication carbon intensity, logic area, and compute-system power. We therefore treat these quantities as independent exploratory variables. The resulting configurations are conditional scenarios rather than predictions, empirical confidence intervals, or physical bounds for a future radiation-hardened H100-class GPU.

B.1 Fabrication Carbon Intensity and Area

Let a COTS GPU fabricated at process node p_{COTS} have die area A_{COTS}, and let its hypothetical radiation-hardened counterpart fabricated at node p_{RH} have area A_{RH}. We define

$$S_{\mathrm{area}} = \frac{A_{\mathrm{RH}}}{A_{\mathrm{COTS}}}, \qquad S_{\mathrm{CPA}} = \frac{\mathrm{CPA}_{\mathrm{RH}}(p_{\mathrm{RH}})}{\mathrm{CPA}_{\mathrm{COTS}}(p_{\mathrm{COTS}})}. \tag{26}$$

Here, S_{area} captures the combined area overhead from process-node migration and radiation-hardening-by-design, while S_{CPA} captures the change in fabrication carbon intensity per unit area, including both node-dependent fabrication intensity and additional radiation-hardening process overhead. The corresponding logic-manufacturing emissions are

$$C_{\mathrm{lo}} = A_{\mathrm{COTS}} \cdot \mathrm{CPA}_{\mathrm{COTS}}(p_{\mathrm{COTS}}), C_{\mathrm{RH}} = A_{\mathrm{COTS}} \cdot S_{\mathrm{area}} \cdot \mathrm{CPA}_{\mathrm{COTS}}(p_{\mathrm{COTS}}) \cdot S_{\mathrm{CPA}}. \tag{27}$$

Thus, Equation 12 can be expressed as

$$C_{\mathrm{RH}} = C_{\mathrm{lo}} \cdot S_{\mathrm{RHA}}, \qquad S_{\mathrm{RHA}} = S_{\mathrm{area}} \cdot S_{\mathrm{CPA}}. \tag{28}$$

For our H100 scenario, the COTS GPU uses 4 nm technology (NVIDIA Corporation, 2025b), whereas the hypothetical radiation-hardened implementation uses a 7 nm radiation-tolerant process (Maillard et al., 2024). Table 5 gives conventional fabrication intensities of $3.0\ \mathrm{kgCO_2e/cm^2}$ at 4 nm and $2.15\ \mathrm{kgCO_2e/cm^2}$ at 7 nm. We explore an additional $1.5\times$—$3\times$ fabrication-intensity overhead for radiation hardening at 7 nm. Therefore,

$$S_{\mathrm{CPA}} = \frac{(1.5\text{—}3) \times 2.15}{3.0} \approx 1.08\text{—}2.15. \tag{29}$$

For area scaling, ideal geometric node scaling from 4 nm to 7 nm gives the conservative upper-end assumption

$$S_{\text{area}} = \left(\frac{7}{4}\right)^2 \approx 3.06. \tag{30}$$

This geometric factor likely overstates practical density scaling because process-node labels do not translate directly into proportional whole-chip density. As an empirical reference, the 4 nm H100 contains ~ 80 billion transistors in an $814\ \text{mm}^2$ die, whereas the 7 nm A100 contains 54.2 billion transistors in $826\ \text{mm}^2$. These specifications correspond to approximately 0.0102 and $0.0152\ \mu\text{m}^2$ per transistor, respectively, or a 50% increase in area per transistor. Nevertheless, we use $S_{\text{area}} = 3.06$ as a conservative exploratory assumption that can additionally accommodate radiation-hardening-by-design overhead. Combining the assumed area and fabrication-intensity factors gives

$$S_{\text{RHA}} = S_{\text{area}} S_{\text{CPA}} \approx 3.3\text{—-}6.6. \tag{31}$$

Because these values are derived from heterogeneous process data rather than a matched COTS/rad-hard H100 implementation, they should not be interpreted as empirical bounds. More conservative values, such as $S_{\text{RHA}} = 12$ used in our illustrative rad-hard configuration, are therefore evaluated separately as sensitivity scenarios rather than justified by this coefficient-derived range.

B.2 Power

Let a COTS GPU fabricated at process node p_{COTS} consume power P_{COTS}, and let its hypothetical radiation-hardened counterpart fabricated at p_{RH} consume P_{RH}. We define the compute-power scaling factor in Equation 13 as

$$P_{\text{RH}} = P_{\text{COTS}} S_{\text{RHP}}, \qquad S_{\text{RHP}} = \frac{P_{\text{RH}}}{P_{\text{COTS}}}. \tag{32}$$

GPU power comprises dynamic and leakage components,

$$P_{\text{GPU}} = P_{\text{dyn}} + P_{\text{leak}}, \tag{33}$$

where

$$P_{\text{dyn}} \propto \alpha C_{\text{eff}} V_{\text{DD}}^2 f, \qquad P_{\text{leak}} \approx V_{\text{DD}} I_{\text{leak}}. \tag{34}$$

Here, α is the switching activity, C_{eff} the effective switched capacitance, V_{DD} the supply voltage, and f the operating frequency. Leakage current depends strongly on threshold voltage, temperature, device dimensions, process variation, gate leakage, and threshold-voltage-cell composition (Roy et al., 2003). Therefore, process-node labels alone are insufficient to predict the power of a hypothetical 7 nm H100-class GPU.

We express the overall power factor as

$$S_{\text{RHP}} = (1 - r_{\text{leak}}) S_{\text{dyn}} + r_{\text{leak}} S_{\text{leak}}, \tag{35}$$

where r_{leak} denotes the leakage fraction of the COTS GPU, and S_{dyn} and S_{leak} denote the dynamic- and leakage-power scaling factors, respectively. For LLM inference workloads, we use $r_{\text{leak}} \approx 0.2$ (Niu et al., 2025). Considering only the supply-voltage contribution when migrating from 4 nm to 7 nm, the dynamic-power factor is

$$S_{\text{dyn},V} = \left(\frac{V_{\text{DD,7nm}}}{V_{\text{DD,4nm}}}\right)^2. \tag{36}$$

Using $V_{\text{DD,4nm}} = 0.70$ V and $V_{\text{DD,7nm}} = 0.75$ V gives

$$S_{\text{dyn},V} = \left(\frac{0.75}{0.70}\right)^2 \approx 1.15, \tag{37}$$

corresponding to an approximately 15% voltage-induced increase in dynamic power. The complete dynamic-power factor additionally depends on changes in C_{eff} and f.

Similarly, considering only the explicit voltage term in leakage power gives

$$S_{\text{leak},V} = \frac{V_{\text{DD,7nm}}}{V_{\text{DD,4nm}}} = \frac{0.75}{0.70} \approx 1.07. \tag{38}$$

Table 9: Exploratory radiation-hardening scenario grid. The settings are sensitivity values rather than empirical bounds or confidence intervals.

Parameter	Interpretation	Scenario settings
S_{RHA}	Logic-manufacturing carbon factor relative to COTS	$6, 12, 24$
S_{RHP}	Compute-system power factor relative to COTS	$1.5, 2, 4$

Table 10: Modeled component-production inventory of the DGX H100-based compute node. Thermal hardware, assembly, transportation, and unquantified system-level components are excluded.

Component	Carbon	Source	Modeling basis
GPUs	0.3320 $\mathrm{tCO_2e}$	(NVIDIA Corporation, 2025b)	H100, network processors, and other ICs
HBM	0.5460 $\mathrm{tCO_2e}$	(NVIDIA Corporation, 2025b)	640 GB HBM, $0.853\ \mathrm{kgCO_2e/GB}$
Host CPUs	0.0390 $\mathrm{tCO_2e}$	Eq. 10	two 4 nm CPUs; $13\ \mathrm{cm}^2$, $3\ \mathrm{kgCO_2e/cm}^2$
System DRAM	0.8192 $\mathrm{tCO_2e}$	Eq. 11	2048 GB DDR; $0.4\ \mathrm{kgCO_2e/GB}$
NAND flash	0.3520 $\mathrm{tCO_2e}$	Eq. 11	16 TB of 3D NAND; $22\ \mathrm{kgCO_2e/TB}$
Others	0.0838 $\mathrm{tCO_2e}$	(NVIDIA Corporation, 2025b)	PCBs, interconnects, and common components
Modeled total	**2.172** $\mathrm{tCO_2e}$		included component-production contributions

This estimate excludes changes in I_{leak}, which can vary substantially across process technologies. Using these voltage-only factors in Equation 35 gives

$$S_{\mathrm{RHP},V} = 0.8 \times 1.15 + 0.2 \times 1.07 \approx 1.13. \tag{39}$$

Thus, supply-voltage differences alone imply an approximately 13% increase in GPU power, but this value should not be interpreted as the complete node-migration power overhead. As an independent technology-level reference, TSMC reports that its 5 nm process provides $\sim 30\%$ lower power than 7 nm at comparable performance (Yeap et al., 2019). Reversing this relationship gives

$$\frac{P_{7\mathrm{nm}}}{P_{5\mathrm{nm}}} \approx \frac{1}{0.70} = 1.43. \tag{40}$$

Because NVIDIA 4N is a customized process and this foundry-level comparison is not based on an H100-class GPU, we use $S_{\mathrm{RHP}} = 1.5$ only as a nominal exploratory scenario. A more conservative value, $S_{\mathrm{RHP}} = 2$, is used for the illustrative rad-hard configuration and evaluated separately in the sensitivity analysis rather than treated as an empirically validated power overhead.

B.3 Combined Uncertain Parameter Sensitivity Scenarios

Table 9 defines the exploratory radiation-hardening design space. We vary S_{RHA} and S_{RHP} independently because no matched H100-class measurements establish a reliable correlation between manufacturing-carbon and power overheads. Increasing S_{RHA} directly increases logic-manufacturing carbon according to Equation 28. Increasing S_{RHP} raises compute-system power and consequently increases the required solar-array, battery, and radiator capacities. These effects are propagated through Equations 4, 5–6, and 7–9. The resulting peripheral masses additionally contribute launch carbon through Equation 1. The logic-manufacturing factor is applied only to accelerator logic and does not scale the mass of the complete DGX system or satellite platform. The named configurations in Section 6 correspond to selected points in this design space rather than nominal predictions. For each combination of S_{RHA} and S_{RHP}, we evaluate whether and under what conditions the embodied carbon ordering between orbital and terrestrial inference changes. A carbon-favorable orbital configuration must additionally satisfy the modeled power, battery-capacity, radiator-area, payload-mass, memory, and service-life constraints. Satisfying these modeled constraints does not establish that the corresponding radiation-hardened accelerator is manufacturable, reliable, or space-qualified.

C DGX H100-Based Component Inventory

Table 10 summarizes the modeled component-production inventory of the DGX H100-based compute node. The inventory combines vendor-reported contributions for the HGX H100 baseboard with

equation-based estimates for the host CPUs, system DRAM, and NAND storage. It therefore represents a partial component inventory rather than a vendor-reported product carbon footprint for a complete DGX H100 system.

The HGX H100 contribution is derived from NVIDIA's product carbon-footprint assessment of the HGX H100 baseboard (NVIDIA Corporation, 2025b). We retain the reported contributions from HBM, GPUs, integrated circuits, and other non-thermal baseboard components. The last category comprises printed circuit boards, interconnects, mechanical and electromechanical components, and common electronic components. We exclude the vendor-reported thermal hardware because the terrestrial cooling subsystem is not retained in the modeled orbital configuration; spacecraft thermal control is modeled separately. Product assembly and transportation are also excluded to maintain a consistent component-production boundary.

The host-CPU contribution is estimated using Equation 10. We assume two AMD EPYC server CPUs (AMD, 2026), each fabricated using a 4 nm process and having a die area of $6.5\ \mathrm{cm}^2$. Applying a logic-fabrication carbon intensity of $3\ \mathrm{kgCO_2e/cm^2}$ yields

$$C_{\mathrm{CPU}} = 2 \times 6.5 \times 3 = 39\ \mathrm{kgCO_2e} = 0.039\ \mathrm{tCO_2e}. \tag{41}$$

The system-DRAM contribution is estimated for 2048 GB of DDR memory using the capacity-based model in Equation 11. With a DRAM fabrication intensity of $0.40\ \mathrm{kgCO_2e/GB}$, its embodied carbon is

$$C_{\mathrm{DRAM}} = 2048 \times 0.40 = 819.2\ \mathrm{kgCO_2e} = 0.8192\ \mathrm{tCO_2e}. \tag{42}$$

For the study-specific 16 TB storage configuration, we estimate the triple-level-cell 3D-NAND contribution using the average manufacturing intensity of $22\ \mathrm{kgCO_2e/TB}$ reported by Weppe et al. (2025):

$$C_{\mathrm{NAND}} = 16 \times 22 = 352\ \mathrm{kgCO_2e} = 0.3520\ \mathrm{tCO_2e}. \tag{43}$$

Summing the included contributions yields a modeled embodied-carbon total of $2.172\ \mathrm{tCO_2e}$. This value remains a partial component-production estimate rather than a complete DGX H100 product footprint. Specifically, the model does not quantify the DGX motherboard, CPU packaging, complete DDR DIMMs, and enclosures, power-supply units, server chassis, system-level interconnects, or final product integration.

D Construction of the Component-Sum Reference for Table 7

We construct the external reference in Table 7 using independently selected coefficients under a common Starlink-V1 + DGX-H100 configuration comprising a DGX H100 node, Si solar array, NMC battery, and aluminum-honeycomb radiator. The DGX H100 contains eight 4 nm H100 GPUs. For the COTS configuration, we assume $P_{\mathrm{sys}} = 10\ \mathrm{kW}$ (Newkirk et al., 2025) and a two-year service life. A LEO satellite completes an orbit roughly every 90–100 minutes (Prol et al., 2022), corresponding to $N_{\mathrm{cyc}} \approx 15$–16 orbital cycles per day. Each orbit typically includes 55–65 minutes of sunlight for power generation and battery charging, followed by 30–35 minutes of eclipse during which stored energy powers onboard LLM inference. We use a nominal 95-minute orbit comprising 60 minutes of sunlight and 35 minutes of eclipse, corresponding to approximately 15.2 cycles per day. The component-sum reference is constructed as follows.

- **Falcon 9 Launch Intensity.** Osoro et al. (2026) reports the launch of a Falcon 9 vehicle costs $293\ \mathrm{tCO_2e}$. The vehicle can support a 22.8 t payload.

$$I_{\mathrm{L}} = \frac{293,000\ \mathrm{kgCO_2e}}{22,800\ \mathrm{kg}} = 12.85\ \mathrm{kgCO_2e/kg} \tag{44}$$

- **Si Solar Array.** The solar array must both power the system during sunlight and replenish the energy consumed during eclipse. For continuous operation, the required array area is

$$A_{\mathrm{sol}} = \frac{P_{\mathrm{sys}} \cdot (T_{\mathrm{sun}} + T_{\mathrm{ecl}})}{\mathrm{ISI} \cdot \eta \cdot T_{\mathrm{sun}}}. \tag{45}$$

  Using $P_{\mathrm{sys}} = 10$ kW, $T_{\mathrm{sun}} = 1$ h, $T_{\mathrm{ecl}} = 35/60 = 0.583$ h, ISI $= 1367$ W/m^2, and $\eta = 0.17$ gives

$$A_{\mathrm{sol}} = \frac{10{,}000(1 + 0.583)}{1367 \times 0.17} \approx 68.1\ \mathrm{m}^2. \tag{46}$$

This idealized value assumes lossless battery charging and power conversion and excludes degradation and design margin; including these effects would increase the required area. Equations 2 and 3 yield $8.04\ \mathrm{tCO_2e}$ of manufacturing carbon. Applying I_{L} in Equation 1 adds $0.88\ \mathrm{tCO_2e}$ from launch.

NMC Battery. For an eclipse duration $T_{\mathrm{ecl}} = 0.583$ h and $\mathrm{DoD} = 0.8$, the required capacity is

$$\mathrm{cap} = \frac{P_{\mathrm{sys}} \cdot T_{\mathrm{ecl}}}{\mathrm{DoD}} = 7.29\ \mathrm{kWh}. \tag{47}$$

Using $a_{\mathrm{bat}} = 4.5\ \mathrm{kg/kWh}$ and $I_{\mathrm{bat}} = 74\ \mathrm{kgCO_2e/kWh}$ (Peiseler et al., 2024) yields $0.54\ \mathrm{tCO_2e}$ of manufacturing and $0.42\ \mathrm{tCO_2e}$ of launch carbon.

Aluminum-Honeycomb Radiator. We assume a thermal load $Q = P_{\mathrm{sys}} = 10\ \mathrm{kW}$. Equation 7, with $\epsilon = 0.95$, $T = 300$ K, and $T_{\mathrm{bg}} = 3$ K, gives $A_{\mathrm{rad}} = 22.9\ \mathrm{m^2}$. With $a_{\mathrm{rad}} = 2.8\ \mathrm{kg/m^2}$ and $I_{\mathrm{rad}} = 18.4\ \mathrm{kgCO_2e/kg}$ (Paraskevas et al., 2016), the radiator contributes $1.18\ \mathrm{tCO_2e}$ from manufacturing and $0.82\ \mathrm{tCO_2e}$ from launch.

DGX H100 Node. Appendix C details the component-level carbon calculation for the DGX H100 node, whose modeled manufacturing embodied carbon is $2.172\ \mathrm{tCO_2e}$. Given a system mass of $130.45\ \mathrm{kg}$ (NVIDIA Corporation, 2023), the baseline launch intensity of $I_{\mathrm{L}} = 12.85\ \mathrm{kgCO_2e/kg}$ attributes an additional $1.68\ \mathrm{tCO_2e}$ of launch emissions to the node.

Starlink-V1. The communication and satellite-platform inventory contributes $1.63\ \mathrm{tCO_2e}$ from manufacturing (Aili et al., 2025) and $2.28\ \mathrm{tCO_2e}$ (Osoro et al., 2026) from launch.

Summing all components yields a COTS reference of $19.64\ \mathrm{tCO_2e}$.

Radiation-Hardened Configuration. For the radiation-hardened (rad-hard) configuration, we assume that the H100-class GPUs and host CPUs migrate from the 4 nm COTS process to a 7 nm radiation-tolerant process (Maillard et al., 2024) and support a 10-year service life. We set $S_{\mathrm{RHA}} = 12$ in Equation 12 and $S_{\mathrm{RHP}} = 2$ in Equation 13. These values are exploratory scenario parameters rather than empirically measured overheads for a rad-hard H100 system. The assumed power factor increases the compute-system power from 10 to 20 kW:

$$P_{\mathrm{sys,rad}} = S_{\mathrm{RHP}} \cdot P_{\mathrm{sys}} = 20\ \mathrm{kW}. \tag{48}$$

This higher power demand increases the required solar-array, battery, and radiator capacities according to the corresponding subsystem models. We scale the manufacturing embodied carbon of the GPU and CPU logic by $S_{\mathrm{RHA}} = 12$ to represent the combined effects of process-node migration and radiation-hardening-by-design, including architectural redundancy such as error-correcting codes (ECC) and triple modular redundancy (TMR). We additionally replace the COTS HBM, DDR DRAM, and NAND flash with modeled radiation-tolerant alternatives. Specifically, we assume MRAM-based DDR and HBM with carbon intensities per unit capacity of $0.6\ \mathrm{kgCO_2e/GB}$ and $1.7\ \mathrm{kgCO_2e/GB}$, respectively, and radiation-tolerant single-level-cell NAND flash with an intensity of $66\ \mathrm{kgCO_2e/TB}$. Under these assumptions, the manufacturing embodied carbon of the rad-hard DGX H100 hardware is $7.91\ \mathrm{tCO_2e}$. Combining this hardware estimate with the separately modeled satellite, peripheral, and launch contributions yields a rad-hard component-sum reference of $37.26\ \mathrm{tCO_2e}$.

These calculations provide a reproducible component-sum reference for controlled model comparison rather than an independent system-level measurement. In particular, the assumed process migration, radiation-hardening factors, and memory technologies should be interpreted as exploratory scenario parameters, not as empirically validated characteristics of an existing rad-hard DGX H100 system.