The Architecture of Orbital Data Centers: Evaluating FPGAs, GPUs, and TPUs for Edge Computing in Low Earth Orbit

daniel@herrington.ai

The Terrestrial Bottleneck and the Emergence of the Near-Earth Cloud

The exponential proliferation of artificial intelligence, particularly the ubiquitous integration of large language models, multimodal generative algorithms, and massive-scale data analytics, has catalyzed an unprecedented demand for computational infrastructure.[1] This demand is rapidly colliding with the physical, energetic, and regulatory limits of terrestrial data centers. Power interconnection queues in major global technology hubs, such as Northern Virginia, currently extend up to seven years.[2]

Concurrently, the thermal management requirements of modern silicon architectures heavily strain local water resources, elevating operational and reputational risks.[2] Forward-looking projections indicate that by the year 2030, data centers within the European Union alone will account for approximately four percent of the region's total electricity demand, consuming an estimated 108 terawatt-hours.[2] To contextualize this volume, this energy consumption is equivalent to the continuous output of more than ten nuclear reactors operating at 1,200 megawatts electric each, surpassing the entire annual electricity consumption of nations such as the Netherlands.[2]

Faced with these severe terrestrial bottlenecks, the aerospace and hyperscale technology sectors are actively pivoting toward an infrastructural paradigm that was once relegated to the realm of science fiction: the deployment of orbital data centers within Low Earth Orbit (LEO).[2] The realization of this "near-Earth cloud" is driven by a convergence of three distinct technological and economic trends. First, the commercialization of space flight, largely driven by the development of ultra-heavy, reusable launch vehicles such as SpaceX's Starship, is projected to drive launch costs down to approximately $200 per kilogram by the mid-2030s.[1] Second, certain orbital regimes, particularly dawn-dusk Sun-synchronous orbits at altitudes between 1,200 and 1,400 kilometers, offer near-continuous solar exposure.[2] In the vacuum of

space, unattenuated solar panels can generate up to eight times more power than identical panels subjected to atmospheric interference on Earth.[3] Third, the deployment of highly capable orbital computing infrastructure allows for true edge computing in space, fundamentally resolving the severe downlink bottlenecks associated with transmitting raw satellite imagery and telemetry back to terrestrial ground stations.[2]

As this infrastructure transitions from conceptual study to operational reality, a foundational engineering debate has emerged regarding which silicon architectures are best suited for the unique and exceedingly harsh constraints of space. Graphics Processing Units (GPUs), Tensor Processing Units (TPUs), and Field-Programmable Gate Arrays (FPGAs) each offer distinct advantages and severe drawbacks.[6] While there is a common theoretical misconception that the vastness of space renders the physical size and power consumption of microelectronics irrelevant, the immutable laws of thermodynamics and orbital mechanics dictate otherwise. In reality, thermal dissipation, radiation tolerance, and strict specific power-to-mass ratios are the absolute arbiters of hardware viability.[2] Within this context, evaluating the survivability of TPUs and GPUs, while exploring the deployment of large-scale FPGA arrays, is critical to understanding the future topology of space-based data centers.

The Physical Constraints of the Space Data Center Environment

To rigorously evaluate the processing hardware proposed for the near-Earth cloud, one must first analyze the environmental constraints of Low Earth Orbit. The assumption that power and physical footprint are negligible concerns in space represents a fundamental misunderstanding of spacecraft engineering. While real estate in orbit is technically infinite, the mass that can be economically lifted into that orbit is strictly finite, and the physics of cooling high-performance microelectronics in a vacuum present formidable engineering challenges.[7]

Thermodynamics, Thermal Dissipation, and the Vacuum Bottleneck

Unlike terrestrial data halls, where ambient air and chilled water systems provide highly efficient convective and conductive cooling, LEO deployments exist in a vacuum where convective cooling is physically impossible.[7] Heat dissipation in space relies entirely on thermal radiation, a process governed by the Stefan-Boltzmann law.[8] This principle dictates that the thermal energy emitted by a black-surfaced plate as infrared radiation scales proportionally with the surface area of the radiator and the fourth power of its absolute temperature.[8]

Because radiated power scales linearly with surface area, any increase in the electrical power consumed by computing components demands a proportional and massive increase in radiator size.[7] For example, the International Space Station utilizes this exact thermodynamic approach, requiring 477 square meters of complex, ammonia-filled radiators merely to

dissipate 75 kilowatts of waste heat.[8] Scaling this technology to support a theoretical 100-kilowatt orbital data center would necessitate approximately 600 square meters of highly specialized, lightweight radiator surface area.[8] To support a 1-megawatt orbital facility, engineers would need to deploy an immense 6,000 square meters of radiator structures.[8] These structures must be highly reliable, requiring absolutely no maintenance access, and must deploy autonomously using complex, origami-inspired folding mechanisms.[8]

The constraints of launchable mass directly limit these thermal management structures, which in turn strictly constrains computing density.[7] For instance, the Aetherflux constellation design plans to host approximately ten interconnected GPUs per satellite.[7] To power and cool this relatively small compute cluster, each satellite requires a 93-square-meter solar panel array (roughly the size of eight terrestrial parking spaces) and a massive 46-square-meter thermal radiator.[7] Furthermore, LEO satellites experience extreme thermal cycling. As satellites pass into the Earth's shadow during eclipse events, they face temperature fluctuations ranging drastically from +120 degrees Celsius in direct sunlight down to -250 degrees Celsius.[2] Managing this requires passive thermal environments, heavy metallic heat pipes, and highly conservative power densities.[2] Consequently, performance per watt is not merely an economic metric of electricity cost, as it is on Earth, but the primary structural determinant of satellite size, mass, and operational viability.

The Radiation Hazard Landscape

Beyond thermal limitations, the second absolute constraint on orbital computing is cosmic radiation. Computers operating in Low Earth Orbit are subjected to a continuous, high-energy bombardment of charged particles originating from the Sun, as well as galactic cosmic rays from deep space.[9] The space environment exposes vehicles to radiation levels that can reach up to approximately 90 krad(Si) per year at the upper limits, particularly when passing through concentrated zones of trapped radiation such as the South Atlantic Anomaly.[10]

These high-energy particles interact destructively with silicon microelectronics, resulting in several distinct, highly disruptive failure modes that necessitate unique hardware architectures[9]:

Radiation Fault Type	Physical Mechanism	Impact on Compute Hardware
Total Ionizing Dose (TID)	The cumulative, long-term degradation of atomic and molecular structures within	Induces threshold voltage shifts in transistor gate oxides, charge buildup,

	silicon lattices due to continuous exposure to ionizing radiation.	increased leakage current, and the formation of trapped holes, eventually leading to permanent processor failure.[9]
Single Event Upset (SEU)	A transient, non-destructive error caused when a single high-energy stray particle strikes a microcircuit and alters its state, flipping a binary bit from 0 to 1, or vice versa.	Results in silent data corruption, software glitches, memory irregularities, and highly inaccurate AI inference outputs if left unmitigated.[9]
Single Event Functional Interrupt (SEFI)	A disruption in the control circuitry of a processor caused by a particle strike, often resetting the device, placing it into an anomalous test mode, or halting its operation entirely.	Requires an immediate soft or hard reboot of the system, temporarily knocking the computing node offline and interrupting data streams.[11]
Single Event Latch-up (SEL)	A severe and potentially destructive condition where a particle triggers a parasitic short circuit within the silicon substrate, causing a massive internal surge in electrical current.	Will physically melt and destroy the microchip unless power is instantly severed by dedicated external hardware monitoring circuits.[11]

Standard Commercial-Off-The-Shelf (COTS) processors are highly susceptible to these anomalies. To operate safely, orbital data centers must employ varying degrees of passive shielding and highly complex active mitigation techniques, fundamentally altering the architecture and selection of the hardware deployed.[11]

Architectural Evaluations: TPUs and GPUs in Orbit

The aerospace and technology industries have taken distinct, divergent approaches to deploying high-performance silicon in space, primarily by testing architectures that are native to terrestrial machine learning workloads. The user's query highlights a specific concern regarding the survivability of GPUs versus TPUs in this radioactive environment, prompting a

detailed examination of both paradigms.

Tensor Processing Units: Project Suncatcher and the Monolithic Approach

Google's Project Suncatcher represents a concerted, high-investment effort to migrate massive, batch-processing machine learning workloads to space.[1] The architectural philosophy relies on treating the vacuum of space as an infinite passive cooling medium while harnessing unattenuated solar power to drive fleets of satellites equipped with Google's proprietary Tensor Processing Units (TPUs).[1]

Addressing the hypothesis that TPUs are capable of surviving the orbital radiation environment, Google engineers conducted extensive terrestrial testing.[3] They subjected the Trillium architecture, specifically Google's v6e Cloud TPU, to a 67MeV proton beam at the UC Davis particle accelerator to simulate the effects of cosmic radiation, focusing on both Total Ionizing Dose and Single Event Effects.[3] The empirical results demonstrated that the TPUs could survive radiation levels equivalent to a five- or six-year mission lifespan without suffering permanent hardware destruction.[1] However, the testing revealed that the High Bandwidth Memory (HBM) subsystems coupled with the TPUs were the most sensitive components, beginning to exhibit data irregularities and bit-flip errors after a cumulative dose of 2 krad(Si).[1]

While the silicon can survive, the TPU approach is highly specialized and somewhat rigid.[6] TPUs are meticulously optimized for massive-scale deep learning training and dense matrix multiplications, primarily using transformer models.[1] However, they are fundamentally designed to operate within the highly controlled, infinitely connected environment of a terrestrial Google Cloud data hall.[6] Because failed TPUs cannot be manually replaced by technicians in orbit, Google's proposed orbital topology utilizes highly redundant provisioning and tightly clustered satellite formations to simulate a continuous data center.[16]

Specifically, the design envisions an 81-satellite cluster operating in extremely close proximity, restricted to a mere 1-kilometer radius.[1] This formation is approximately 120 times tighter than the spacing of standard Starlink satellites.[3] To facilitate the massive data bandwidth required by TPUs, these satellites must maintain highly precise, machine-learning-controlled optical communication links running at 1.6 terabits per second via free-space optics.[1] This architecture is effectively a monolithic supercomputer distributed across independent, co-orbiting nodes. While theoretically highly efficient for latency-tolerant batch computations and massive model training, TPUs lack the peripheral adaptability required for direct sensor interfacing, rapid radio frequency modulation, or dynamic signal routing.[2]

Graphics Processing Units: The NVIDIA Jetson and ScOSA Framework

Graphics Processing Units remain the dominant force in the broader technology ecosystem for

parallel processing, high-throughput tasks, and flexible artificial intelligence training.[6]
Recognizing the vast software maturity of the GPU ecosystem, commercial space ventures and national space agencies have actively sought to harden NVIDIA hardware for orbital deployment.[7]
Addressing the user's concern that GPUs might not survive space radiation: GPUs are indeed highly vulnerable in their native, commercial state.[18] However, they can be made viable through the implementation of extensive external hardware scaffolding and rigorous software redundancy. The Scalable On-board Computing for Space Avionics (ScOSA) framework, developed for advanced robotic space missions, demonstrates precisely how a high-performance COTS GPU, such as the NVIDIA Jetson AGX Orin Industrial module, can function reliably in Low Earth Orbit.[11]
The Jetson AGX Orin represents a massive leap in orbital compute potential. Built on the Ampere architecture, the 64GB variant features a 2048-core GPU alongside 64 dedicated Tensor cores and a 12-core ARM v8.2 CPU.[9] It delivers up to 275 Trillion Operations Per Second (TOPS) of AI performance, offering up to 5.3 FP32 teraflops of CUDA compute.[20] Crucially for space deployment, it utilizes 64 GB of LPDDR5 RAM equipped with Error-Correcting Code (ECC) to natively mitigate memory bit-flips.[9]
Because the GPU itself is not intrinsically radiation-hardened at the silicon level against catastrophic events, the ScOSA architecture pairs the Orin module with a highly robust, radiation-hardened Co-CPU, specifically the GR740.[11] This Co-CPU acts as an active, invincible supervisor. If a cosmic ray induces a Single Event Functional Interrupt (SEFI) causing the GPU to hang or output corrupted data, the Co-CPU detects the failure via continuous heartbeat monitoring and active watchdogs.[11] The supervisor can actively reboot the node or completely re-flash the GPU's Non-Volatile Memory express (NVMe) storage directly from a protected, radiation-immune Programmable Read-Only Memory (PROM) module.[11] Furthermore, the system employs software-level Triple Modular Redundancy (TMR) and checkpoint-restore services, ensuring that AI tasks are replicated across multiple nodes and voted upon before any inference is transmitted back to Earth.[11]
While this hybrid approach proves that GPUs can survive, their thermal power envelopes remain aggressive and highly problematic.[22] The Jetson AGX Orin operates with a configurable power draw between 15 watts and 60 watts.[19] In the strictly thermally constrained vacuum environment, running a GPU at a sustained 60-watt load requires substantial custom aluminum conduction paths.[11] To prevent the silicon from overheating during intense AI inference bursts, engineers must employ Phase Change Material (PCM) buffers containing Hard Fischer–Tropsch wax.[11] As the GPU generates heat, the wax melts, absorbing a massive 240,000 Joules per kilogram of latent heat of fusion, thereby extending the operational run-time before thermal throttling occurs.[11] Thus, while GPUs provide immense, flexible throughput, their integration requires heavy, complex thermal and radiation mitigation infrastructure.

The Architecture and Inherent Value of FPGAs in

Space

Given the raw throughput of GPUs and the massive batch-processing capabilities of TPUs, evaluating the unique value proposition of Field-Programmable Gate Arrays (FPGAs) in a space-based data center requires a deep understanding of their architectural differences.[6] While GPUs excel at high-throughput parallel processing and TPUs at dense matrix multiplication at scale, FPGAs provide extreme power efficiency, deterministic low latency, deep hardware reconfigurability, and inherent radiation tolerance at the silicon substrate level.[6]

Space-Grade FPGA Architectures: The AMD Versal XQR Series

The benchmark for modern space-grade FPGAs is the AMD (formerly Xilinx) Versal XQR series.[24] This hardware represents the aerospace industry's first 7-nanometer radiation-tolerant Adaptable Compute Acceleration Platform (ACAP).[15] Unlike legacy FPGAs, which contained solely a sea of programmable logic gates, modern space-grade FPGAs are highly complex, heterogeneous System-on-Chips (SoCs).[15]

The Versal architecture is segmented into highly specialized processing engines, all seamlessly integrated via a programmable Network-on-Chip (NoC) [15]:

1. **Scalar Engines:** The platform includes a hardened dual-core ARM Cortex-A72 application processing unit, complete with ECC-protected L1 and L2 caches.[15] This provides the general-purpose compute required for running Linux-based operating systems and Kubernetes orchestration algorithms.[27] It is paired with a dual-core ARM Cortex-R5F real-time processing unit designed for deterministic, safety-critical subsystem management.[15]
2. **Adaptable Engines:** This section contains the traditional programmable logic associated with FPGAs. Devices like the Versal AI Core series contain up to 1.9 million system logic cells and nearly 900,000 Look-Up Tables (LUTs).[15] This allows aerospace engineers to create custom, application-specific data pathways and highly localized memory hierarchies directly at the hardware level, bypassing the latency associated with general-purpose CPU architectures.[15]
3. **Intelligent Engines:** To compete with GPUs and TPUs in artificial intelligence workloads, the Versal integrates massive arrays of Very-Long Instruction Word (VLIW) processors equipped with Single Instruction Multiple Data (SIMD) vector units.[15] In the AI Core variant (e.g., the XQRVC1902), 400 AI Engine tiles provide massive acceleration for complex matrix multiplications.[14] In the AI Edge variant (e.g., the XQRVE2302), 34 AIE-ML tiles are explicitly optimized for neural network data types, supported by hundreds of specialized Digital Signal Processing (DSP) engines.[15]

Inherent Radiation Resilience and Mission Longevity

Unlike COTS GPUs that require external watchdogs, thick aluminum shielding, and rad-hard companion processors to survive, space-grade FPGAs are manufactured with intrinsic mitigation strategies designed into the silicon.[11] The Versal XQR devices undergo rigorous Class B manufacturing and qualification processes, strictly adhering to military and aerospace standards (such as MIL-PRF-38535) and supporting extreme operational temperature ranges from -55°C to +125°C.[15]

Crucially, the silicon is engineered to withstand Total Ionizing Dose and destructive Single Event Latch-ups natively.[15] For the management of transient Single Event Upsets (SEUs), the platform employs fully integrated, self-sufficient mitigation frameworks, notably XilSEM.[15] This technology continuously scrubs the configuration memory of the FPGA, instantly detecting and correcting bit-flips caused by cosmic rays without requiring any external hardware scrubber or halting the primary computational workload.[15] This self-sufficiency drastically reduces the overall component count on the satellite bus, subsequently decreasing launch mass and eliminating numerous potential single points of failure that plague hybrid GPU designs. The overall design is optimized for Space 2.0 applications, offering certified support for mission durations spanning up to seven years.[15]

The Economics of Power Efficiency and Thermal Management

In the strictly thermally constrained vacuum of Low Earth Orbit, performance per watt is paramount. While the user hypothesized that power does not matter in space, the reality of radiator mass proves that the most energy-efficient processor is ultimately the most scalable processor.[7] FPGAs excel fundamentally in running low-latency, real-time inference tasks at absolute minimal power.[6]

When benchmarked against competing silicon architectures using a standard image classification neural network (ResNet50 v1.5), the Versal AI Core series demonstrated massive efficiency advantages.[27] The Versal AI Core VC1902 achieved 66 frames-per-second per watt (FPS/W), vastly outperforming rival FPGA baselines, such as the Intel Agilex, which achieved only 24 FPS/W.[27] The VC1902 device delivers 6,471 frames per second while consuming an estimated total system power of only 87 Watts, providing 133 INT8 TOPS of peak inference performance.[27]

While a GPU may require a highly fluctuating 60W power budget and thick phase-change thermal spreaders to manage unpredictable inference spikes [11], FPGAs process continuous inference streams with significantly lower, deterministic energy draws.[27] In an FPGA architecture, data does not need to constantly travel back and forth to external DDR memory over a power-hungry peripheral bus. Instead, the adaptable logic paths channel data directly from one customized processing block to the next via internal distributed RAM and UltraRAM.[27] This streamlined, custom memory hierarchy makes FPGAs the hardware of choice for

"instant-on" performance and highly efficient continuous edge AI applications.[6]

Silicon Architecture	Primary Orbital Use Case	Power Efficiency Profile	Radiation Mitigation Strategy	Core Strengths in LEO
GPU (e.g., Jetson AGX Orin)	Flexible, high-throughput AI inference and training.[6]	High power draw (15W-60W); requires heavy thermal buffering (PCM).[11]	Requires external rad-hard Co-CPU, NVMe re-flashing, and TMR.[11]	Mature software ecosystem; massive parallel throughput.[20]
TPU (e.g., Trillium v6e)	Monolithic, large-scale deep learning and batch processing.[6]	Highly efficient for dense matrix multiplication, but requires continuous operation.[6]	Relies on redundant clustering (81-satellite arrays) to bypass failed nodes.[16]	Unmatched scale for transformer models in tightly coupled clusters.[1]
FPGA (e.g., AMD Versal XQR)	Deterministic low-latency edge inference, SDRs, and signal routing.[6]	Exceptionally high performance per watt (up to 66 FPS/W) for specific tasks.[27]	Inherent Class B rad-tolerance; integrated memory scrubbing (XilSEM).[15]	Deep hardware reconfigurability; minimal external component dependency.[15]

Edge Computing: Overcoming the Downlink Bottleneck

The primary operational use case driving the deployment of orbital data centers is edge computing for Earth Observation (EO) and remote sensing.[2] Modern satellites equipped with high-resolution optical cameras, Hyperspectral Imaging (HSI) sensors, and Synthetic Aperture Radar (SAR) generate vast quantities of raw telemetry, easily reaching terabytes of imagery per day.[32] This immense data volume quickly outstrips the physical capacity of the satellite's RF downlink to ground stations.[5] The economic and strategic value of the satellite constellation is consequently throttled; critical intelligence regarding rapidly unfolding natural disasters,

unauthorized maritime movements, or dynamic military deployments can be delayed by hours or even days while waiting for the spacecraft to achieve line-of-sight to a terrestrial antenna.[32] FPGAs are perfectly suited to shatter this downlink bottleneck by processing complex sensor data directly on-orbit, transforming raw physics data into highly compressed, actionable data streams prior to transmission.[18]

Real-Time Synthetic Aperture Radar (SAR) Focusing

Synthetic Aperture Radar is an active remote sensing technology that utilizes microwave pulses to penetrate cloud cover and darkness, offering all-weather visibility. However, the raw data captured by a SAR sensor consists of highly fragmented, unfocused wave reflections. Focusing this raw data into a geometrically accurate, viewable image is an extremely computationally intensive process involving massive Fast Fourier Transforms (FFTs) and complex matrix multiplications.[34]

When implemented on general-purpose CPUs, SAR focusing is prohibitively slow, making real-time analysis impossible. While GPUs offer the thousands of parallel cores necessary to accelerate this mathematical workload, their high power budgets often exceed the stringent 15-watt power limits available on smaller, lightweight drone or CubeSat platforms.[35] By contrast, FPGAs can execute these specific algorithms with deterministic, real-time precision while adhering strictly to low power constraints. Utilizing an AMD Xilinx Alveo-class FPGA architecture equipped with High-Bandwidth Memory (HBM), the highly complex SAR focusing kernel can process an entire azimuth block—representing up to 375 square kilometers of the Earth's surface—in just 1.48 seconds.[34] Because the physical satellite sensor requires 2.21 seconds to acquire that same amount of raw data, the FPGA processes the radar imagery significantly faster than it is collected.[34] Operating at a highly efficient 130 MHz clock rate and utilizing 12.5 gigabytes of on-device HBM, the FPGA achieves true real-time, low-latency radar focusing directly at the edge.[34]

Hyperspectral Image Processing and AI-Driven Anomaly Detection

Hyperspectral cameras represent another massive data challenge, capturing light across hundreds of narrow spectral bands to create massive, multi-dimensional data cubes utilized in agriculture, climate monitoring, and defense.[33] The European Space Agency’s HYPSO-1 cube-satellite demonstrates the absolute necessity of onboard processing pipelines to manage this data. By utilizing onboard FPGA logic to capture, bin, and compress the imagery immediately after capture and prior to transmission, the satellite avoids completely overwhelming its limited communication hardware.[33]

Furthermore, advanced deep learning models can be highly optimized for direct FPGA deployment to analyze this imagery autonomously. Historically, deploying AI onto FPGAs

represented a significant knowledge barrier, requiring expertise in both data science and hardware description languages.[32] Today, specialized low-code software development kits (SDKs) streamline this process. Data scientists utilize Quantization-Aware Training (QAT) to convert massive, floating-point neural networks into highly efficient, integer-based approximations (such as 8-bit or 4-bit weights).[32] Advanced frameworks, such as the FINN pipeline, then transform these quantized models (exported in QONNX format) into streaming hardware dataflows, automatically generating distinct, physical Intellectual Property (IP) logic blocks for each neural network layer.[32]

These custom AI models run directly on the FPGA fabric, drawing minimal power while performing complex vision tasks. Current orbital applications include utilizing U-Net models for precise water body segmentation, and employing YOLOv3-tiny models for real-time ship detection and classification.[32] By executing these models onboard, the satellite can transmit a 5-kilobyte text alert confirming that a specific class of vessel has been detected at a precise GPS coordinate, rather than downlinking a 5-gigabyte raw image cube for terrestrial analysts to search.[32] This paradigm maximizes the tactical value of the satellite pass, enabling autonomous, time-critical decision-making.[32]

Inter-Satellite Communications and the Starlink Synergy

The user query correctly posits the vast potential of integrating orbital data centers with established communications constellations similar to SpaceX's Starlink.[4] To achieve a genuinely functional "near-Earth cloud," independent compute nodes cannot operate in isolation. They must be intricately stitched together via a highly dynamic, immensely high-bandwidth communication fabric consisting of thousands of Inter-Satellite Links (ISLs).[37]

This networking requirement represents perhaps the greatest, most irreplaceable unique value of FPGAs in the space deployment architecture. While TPUs and GPUs act primarily as endpoint processors consuming and generating data, FPGAs act as the foundational hardware routers and signal processors—the central circulatory system routing data across the entire constellation.[39]

Software Defined Radio (SDR) and Digital Beamforming

Mega-constellations like Starlink and OneWeb rely heavily on Ka-band and Ku-band active phased-array antennas and highly advanced Software Defined Radios (SDRs).[36] As the orbital RF spectrum becomes increasingly crowded and highly contested, space architectures necessitate waveforms that can dynamically shift frequencies, rapidly employ anti-jamming techniques, and utilize complex low probability of intercept (LPI) signatures for secure military communications.[38]

FPGAs are the primary hardware capable of executing the intense, low-latency digital signal processing algorithms at the heart of these SDRs.[30] An FPGA can handle the immediate,

real-time digitization of analog RF signals at the sensor edge, instantly executing highly complex error-correcting codes (such as Forward Error Correction and LDPC decoding) while performing digital beamforming to precisely direct radio signals toward moving ground stations.[31]

Because the logic gates within an FPGA are completely reconfigurable, an orbital data center equipped with FPGA-based SDRs—such as the commercially available CesiumAstro SDR-1001 or the IQ Spacecom XLink systems—can radically alter its communication protocols on the fly.[43] If a new, highly efficient LTE-grade waveform or an advanced DVB-S2X modulation scheme is developed by terrestrial engineers, a new bitstream can simply be uploaded to the satellite.[43] This software upload physically rewires the FPGA's internal hardware gates, allowing the satellite to natively support the new protocol without requiring the launch of replacement hardware.[43]

Laser Cross-Links and Dynamic Optical Routing

To move terabits of data between orbital data center modules without suffering from RF interference or spectrum licensing restrictions, advanced constellations are aggressively transitioning to Optical Inter-Satellite Links (OISL)—specifically, high-speed laser communications.[38] Initiatives such as NASA's CLICK (CubeSat Laser Infrared CrosslinK) mission have successfully demonstrated the viability of establishing gigabit-per-second data rates between rapidly moving spacecraft.[45] These systems utilize precise 1550-nanometer near-infrared lasers directed by highly complex fine steering mirrors capable of maintaining 70-microradian wide beams over vast distances.[45]

However, routing packetized data across a rapidly moving web of thousands of LEO satellites requires specialized, high-performance hardware routers. Traditional, static Internet routing protocols fail entirely in space because the physical topology of the network changes by the second as satellites orbit the Earth at velocities exceeding 7.5 kilometers per second.[46] Systems like the ASCoT routing architecture leverage the predictability of satellite trajectories to effect dynamic, position-based data routing.[46]

The SPACE!Box project illustrates precisely how FPGAs are uniquely utilized to build these high-performance, low-latency space routers.[27] Interfacing directly with standard SpaceWire protocols or high-speed optical links, the FPGA acts as the central network switch.[39] The custom logic fabric intercepts high-speed packet traffic at Layer 1 and Layer 2, dynamically references rapidly updating forwarding tables, manages multi-gigabit ingress and egress memory queues, and implements critical traffic shaping algorithms to mitigate network overloads across heterogeneous communication links.[27] Standard GPUs and highly rigid TPUs are fundamentally incapable of performing this type of continuous, low-level, physical-layer packet switching with the energy efficiency required by spacecraft.[17]

The Strategic Power of On-Orbit Reconfigurability

The fundamental weakness of launching inflexible Application-Specific Integrated Circuits (ASICs) or highly rigid TPUs into the vacuum of space is that the hardware architecture cannot be physically upgraded for the entire duration of its operational lifespan, which typically spans five to ten years.[47] Artificial intelligence and telecommunications are aggressively fast-moving fields; state-of-the-art neural network architectures and data quantization methodologies undergo complete generational shifts every twelve to eighteen months.[2] Consequently, a space data center equipped exclusively with fixed-architecture silicon risks rapid, unavoidable technological obsolescence long before its solar panels degrade.

FPGAs offer a profound strategic advantage known as "true unlimited on-orbit reconfiguration," often termed the "Upgrade-on-the-Fly" capability.[15] If a major breakthrough in AI quantization occurs, or a fundamentally new, highly efficient mathematical mechanism for hyperspectral analysis is invented, the FPGA can be completely reprogrammed remotely via a secure software upload.[15] This capability extends far beyond merely updating standard software applications; the uploaded bitstream literally alters the physical hardware data paths, restructures the cache memory hierarchies, and reconnects the logic unit structures to perfectly match the precise mathematical requirements of the new algorithm.[15]

The European Space Agency's OPS-SAT mission successfully and dramatically demonstrated this operational paradigm. The mission featured a highly flexible processing platform explicitly designed to allow third-party researchers and software developers to upload entirely new hardware configurations to the satellite's onboard FPGA.[32] This unprecedented access enabled the execution of advanced anomaly detection algorithms, adaptive software-defined radio protocols, and dynamic debugging capabilities in a live, hostile operational setting.[48] This extreme adaptability transforms the satellite from a single-purpose, rapidly aging tool into a continuously evolving, orbiting computational laboratory that remains at the bleeding edge of technology for its entire lifespan.

Strategic Synergies: Designing Hybrid Computing Constellations

Ultimately, evaluating the utility of GPUs, TPUs, and FPGAs for space-based data centers should not be viewed as a zero-sum, mutually exclusive competition. Rather, it represents a complex exercise in architectural synergy. The most effective, resilient, and economically viable orbital infrastructure will undoubtedly rely on highly integrated Hybrid Computing platforms—combining dissimilar computing technologies to harness their collective, complementary advantages while mitigating their individual weaknesses.[18]

A fully matured Low Earth Orbit data center architecture, deeply integrated with a mega-constellation communications network similar to Starlink, would feature highly specialized, interoperating nodes:

1. **The FPGA Routing and Intelligent Edge Node:** These highly efficient, radiation-tolerant

chips would be placed on every single satellite within the constellation.[32] Operating continuously, the FPGAs manage the gigabit laser cross-links, execute complex digital beamforming for dynamic RF downlinks, and run highly optimized, low-latency AI inference models on live sensor streams.[27] They act as the primary filter, instantly discarding useless data, detecting time-critical anomalies, and routing the valuable intelligence through the network.[32]

2. **The TPU and GPU Heavy Compute Cluster:** Deployed as dedicated, massive satellite modules flying in tightly controlled, close-proximity formations, these nodes represent the heavy computational core of the orbital cloud.[3] Utilizing 1.6 Tbps optical links, they receive the pre-filtered, aggregated data routed to them by the FPGA edge nodes.[1] Harnessing their immense, parallel matrix-multiplication throughput and drawing upon the massive, continuous solar power generated in dawn-dusk orbits, these TPU and GPU clusters perform the latency-tolerant batch compute tasks.[2] They execute deep learning training runs, refine global climate models, and process massive cryptographic workloads that are fundamentally unsuited for the smaller edge nodes.[2]
3. **Radiation-Hardened Executive Supervisors:** Beneath these high-performance commercial accelerators sit the classic, highly resilient, purpose-built space CPUs (such as the ARM Cortex-R5F cores or the GR740).[11] These processors ensure the ultimate survival of the constellation by managing critical thermal load shedding, handling precise orbital station-keeping, actively repairing logic faults in the COTS accelerators, and maintaining absolute control during passage through severe radiation zones.[11]

Conclusion

The proposed deployment of massive data centers into Low Earth Orbit represents a monumental, paradigm-shifting evolution in global computational infrastructure. By elevating processing hardware into the vacuum of space, the technology sector seeks to bypass the crippling terrestrial bottlenecks of power grid interconnection delays, massive water consumption, and strict regulatory oversight, instead tapping into the endless, unattenuated solar energy of the cosmos. However, the extreme physical constraints of specific launch mass, absolute susceptibility to cosmic radiation, and the severe complexities of vacuum thermodynamics impose strict, unyielding limitations on hardware selection and architectural design.

While initiatives such as Google's Project Suncatcher successfully prove that custom architectures like TPUs can survive LEO radiation to provide immense, distributed scale for batch processing, their rigid nature requires immense redundant clustering. Similarly, while the ScOSA framework proves that standard commercial GPUs can be coerced into operating safely in space to provide unparalleled parallel processing throughput, they demand heavy, highly complex external thermal buffering and radiation scaffolding to survive.

Within this extreme environment, Field-Programmable Gate Arrays offer a highly specialized,

uniquely powerful, and utterly irreplaceable utility. FPGAs resolve the fundamental downlink bottleneck by providing the deterministic latency and extreme performance-per-watt necessary to process immense streams of raw radar and hyperspectral imagery directly at the sensor edge. Furthermore, FPGAs act as the foundational telecommunications backbone of the near-Earth cloud, executing the software-defined radio protocols and dynamic optical packet routing necessary to stitch thousands of independent satellites into a cohesive, high-speed network. Crucially, their inherent capacity for infinite, deep hardware reconfigurability ensures that an orbital data center can evolve continuously alongside rapid terrestrial advancements in artificial intelligence and networking. By acting as the highly adaptable, intelligent sensory and routing fabric of the orbital infrastructure, FPGAs ensure that the massive raw data generated in space is successfully refined, rapidly routed, and securely delivered to Earth as actionable intelligence.

Works cited

1. Towards a future space-based, highly scalable AI ... - Google, accessed June 23, 2026, https://services.google.com/fh/files/misc/suncatcher_paper.pdf
2. Data centers go orbital | Arthur D. Little, accessed June 23, 2026, https://www.adlittle.com/en/insights/viewpoints/data-centers-go-orbital
3. Google Plans to Launch Data Centers Into Space, and the Math Actually Works - Medium, accessed June 23, 2026, https://medium.com/@cognidownunder/google-just-announced-theyre-putting-data-centers-in-space-37934fe7ea84
4. SpaceX FCC filing: 1 Million Satellites For Orbital Data Center Push - Reddit, accessed June 23, 2026, https://www.reddit.com/r/spacex/comments/1qrxosl/spacex_fcc_filing_1_million_satellites_for/
5. Radiation-tolerant FPGAs solve the satellite signal processing bottleneck, accessed June 23, 2026, https://militaryembedded.com/radar-ew/signal-processing/radiation-tolerant-signal-processing-bottleneck
6. GPU vs. TPU vs. FPGA: Choosing the Right Accelerator for AI | Shieldbase, accessed June 23, 2026, https://shieldbase.ai/blog/gpu-vs-tpu-vs-fpga-choosing-the-right-accelerator-for-ai
7. Data Centers in space, is it feasible? - Sener, accessed June 23, 2026, https://www.group.sener/en/insights/data-centers-in-space-is-it-feasible/
8. LEO Data Centers | Space-Based Computing Infrastructure 2025, accessed June 23, 2026, https://leodatacenters.com/
9. S-A2300 - NVIDIA Jetson AGX Orin-based Supercomputer ..., accessed June 23, 2026, https://emcomo.com/ai-supercomputer-survives-extreme-radiation-in-low-earth-orbit-leo-simulation
10. (PDF) Total Ionizing Dose Radiation Testing of NVIDIA Jetson Orin NX System on Module, accessed June 23, 2026,

https://www.researchgate.net/publication/387116286_Total_Ionizing_Dose_Radiation_Testing_of_NVIDIA_Jetson_Orin_NX_System_on_Module
11. Feasibility of a COTS GPU-Based On-board AI ... - electronic library -, accessed June 23, 2026, https://elib.dlr.de/218458/1/2025_IAC_spAIce.release.pdf
12. Exploring a space-based, scalable AI infrastructure system design - Google Research, accessed June 23, 2026, https://research.google/blog/exploring-a-space-based-scalable-ai-infrastructure-system-design/
13. Sources of Single Event Effects in the NVIDIA Xavier SoC Family under Proton Irradiation - UPCommons, accessed June 23, 2026, https://upcommons.upc.edu/bitstreams/80d67a8c-34a3-408b-aaf1-5bc5bc48b1c1/download
14. Contains the AMD Versal™ AI Core XQR Series specifications including radiation parameters, key features, DC and AC switching characteristics, packaging and pinout, and ordering information. - DS946, accessed June 23, 2026, https://docs.amd.com/r/en-US/ds946-xqr-versal-ai-core
15. XQR Versal™ for Space 2.0 Applications - AMD, accessed June 23, 2026, https://www.xilinx.com/content/dam/xilinx/publications/product-briefs/xilinx-xqr-versal-product-brief.pdf
16. Towards a future space-based, highly scalable AI infrastructure system design - arXiv, accessed June 23, 2026, https://arxiv.org/html/2511.19468v1
17. FPGA or GPU? Analyzing comparative research for application-specific guidance - arXiv, accessed June 23, 2026, https://arxiv.org/html/2511.06565v1
18. Onboard Processing With Hybrid and Reconfigurable Computing on Small Satellites | Pitt Space, accessed June 23, 2026, https://space.pitt.edu/sites/default/files/2024-10/Onboard-Processing-with-Hybrid-and-Reconfigurable-Computing-on-Small-Satellites.pdf
19. NVIDIA Jetson AGX Orin | Datasheet - OpenZeka, accessed June 23, 2026, https://openzeka.com/wp-content/uploads/2022/03/jetson-agx-orin-datasheet.pdf
20. Jetson AGX Orin for Next-Gen Robotics - NVIDIA, accessed June 23, 2026, https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-orin/
21. NVIDIA Jetson AGX Orin Series, accessed June 23, 2026, https://www.nvidia.com/content/dam/en-zz/Solutions/gtcf21/jetson-orin/nvidia-jetson-agx-orin-technical-brief.pdf
22. FPGA vs GPU vs TPU : Architectural, Performance & Feature - WatElectronics.com -, accessed June 23, 2026, https://www.watelectronics.com/fpga-vs-gpu-vs-tpu/
23. FPGA vs. GPU for Deep Learning Applications - IBM, accessed June 23, 2026, https://www.ibm.com/think/topics/fpga-vs-gpu
24. AMD Versal™ XQR Adaptive SoCs, accessed June 23, 2026, https://www.amd.com/en/products/adaptive-socs-and-fpgas/versal/space-grade.html
25. High-performance on-board instrument computer and SDR for small satellites - Technologies portfolio - Institut d'Estudis Espacials de Catalunya, accessed June

23, 2026, https://www.ieec.cat/en/project/50/high-performance-on-board-instrument-computer-and-sdr-for-small-satellites/
26. AMD VERSAL™ XQR PORTFOLIO FOR SPACE APPLICATIONS, accessed June 23, 2026, https://www.amd.com/content/dam/amd/en/documents/solutions/aerospace-and-defense/amd-xqr-versal-product-brief.pdf
27. AI Inference with Versal™ AI Core Series - AMD, accessed June 23, 2026, https://www.xilinx.com/content/dam/xilinx/publications/solution-briefs/xilinx-versal-ai-compute-solution-brief.pdf
28. AI Engine - 2025.2 English - UG1273, accessed June 23, 2026, https://docs.amd.com/r/en-US/ug1273-versal-acap-design/AI-Engine
29. Radiation Tolerant Versal AI Edge Series: Data Sheet - Arrow Electronics, accessed June 23, 2026, https://static6.arrow.com/aropdfconversion/18b9ecd747c83422045d2b49b658ab64f94297a7/ds955xqrversalaiedge.pdf
30. Satellite Communications | Lattice FPGA Applications, accessed June 23, 2026, https://www.latticesemi.com/en/Solutions/Solutions/SolutionsDetails02/satellite-communications
31. Space - AMD, accessed June 23, 2026, https://www.amd.com/en/solutions/aerospace-and-defense/space.html
32. On-orbit data processing utilizing FPGA-accelerated deep learning ..., accessed June 23, 2026, https://indico.esa.int/event/552/contributions/10992/attachments/7074/13262/On-orbit%20data%20processing%20utilizing%20FPGA-accelerated%20deep%20learning%20models.pdf
33. Robust and Reconfigurable On-Board Processing for a ... - MDPI, accessed June 23, 2026, https://www.mdpi.com/2072-4292/15/15/3756
34. Real Time Floating Point SAR Focusing on FPGA, accessed June 23, 2026, https://elib.dlr.de/204048/1/p60-mandapati.pdf
35. Improved Real-Time SPGA Algorithm and Hardware Processing Architecture for Small UAVs, accessed June 23, 2026, https://www.mdpi.com/2072-4292/17/13/2232
36. The Truth About SpaceX's "Orbital Datacenters" : r/SpaceXLounge - Reddit, accessed June 23, 2026, https://www.reddit.com/r/SpaceXLounge/comments/1t6zt1c/the_truth_about_spacexs_orbital_datacenters/
37. Deep Tech to Space: Space Data Centers and AI Revolution at the Edge - arXiv, accessed June 23, 2026, https://arxiv.org/html/2605.19892v1
38. 15 laser communications systems for inter-satellite links (ISLs) in LEO, that you can buy in May 2025 - Blog, accessed June 23, 2026, https://blog.satsearch.co/2025-05-01-15-laser-communications-systems-for-inter-satellite-links-isls-in-leo-that-you-can-buy-in-may-2025
39. Parallel Eclipse-Aware Routing on FPGA for SpaceWire-Based OBC in LEO Satellite Networks - MDPI, accessed June 23, 2026,

https://www.mdpi.com/2224-2708/14/4/73
40. SPACE!Box: On-Board Data Router for Future NTN and OBDH, accessed June 23, 2026, https://indico.esa.int/event/552/contributions/11111/attachments/7051/12939/EDHPC_SpaceBox_Presentation.pdf
41. Radio Frequency Payloads Tutorial EDHPC 2025.pdf - Indico at ESA / ESTEC, accessed June 23, 2026, https://indico.esa.int/event/552/contributions/11462/attachments/7144/13559/Radio%20Frequency%20Payloads%20Tutorial%20EDHPC%202025.pdf
42. Satellite Communications Prototyping and Deployment with NI COTS Components - National Instruments, accessed June 23, 2026, https://www.ni.com/content/dam/web/pdfs/satcom_prototyping_solution_brochure_ltr_wr.pdf
43. Software Defined Radios (SDRs) for space and satellite applications ..., accessed June 23, 2026, https://blog.satsearch.co/2022-06-27-software-defined-radios-sdrs-for-space-and-satellite-applications
44. Q&A on SDR for small satellites: preconfigured vs. customised solutions - Alén Space, accessed June 23, 2026, https://alen.space/sdr-for-small-satellites-preconfigured-vs-customised-solutions/
45. The CubeSat Laser Intersatellite CrosslinK (CLICK) - NASA TechPort - Project, accessed June 23, 2026, https://techport.nasa.gov/projects/94065
46. Data Centric, Position-Based Routing In Space Networks, accessed June 23, 2026, https://geometry.stanford.edu/lgl_2024/papers/gpbg-dcprsn-05/gpbg-dcprsn-05.pdf
47. GPUs under constant pressure in AI data centers -- 9% annualized failure rate -- Without radiation. Space data centers is such a dumb idea. : r/RealTesla - Reddit, accessed June 23, 2026, https://www.reddit.com/r/RealTesla/comments/1s09vlj/gpus_under_constant_pressure_in_ai_data_centers_9/
48. Using a Reconfigurable FPGA for Emerging Space Applications - TU Graz Repository, accessed June 23, 2026, https://repository.tugraz.at/records/8bvh6-py284