Enabling Software Defined Satellites with Edge AI

Jim Larson
Sidus Space, Inc.
Cape Canaveral, FL USA
jim.larson@sidusspace.com

***ABSTRACT*—Software-Defined Satellites (SDSs) enable adaptive, software-driven platforms with on-orbit evolution through software updates and reconfiguration. True operational autonomy is achieved through integration of edge Artificial Intelligence (AI) directly onboard, enabling satellites to sense, decide, and act in real-time. This paper examines the role of edge AI within SDS architectures, emphasizing heterogeneous onboard compute comprising CPUs, GPUs, and FPGAs. Drawing on flight-proven experience, we explore architectural considerations, system-level trade-offs, and operational benefits for commercial and government missions. The coordinated use of FPGAs for data verification, GPUs for high-throughput AI inference, and CPUs for system orchestration enables low-latency, low-power, mission-adaptive performance. Edge AI is the critical enabler transforming software-defined systems into intelligence-defined space infrastructure.**

I. INTRODUCTION

Traditional satellite architectures with static configurations offer reliability but suffer from limited flexibility and increased lifecycle costs. Software-Defined Satellites enable post-launch reconfiguration through software abstraction, supporting multi-mission operations and extended lifetimes. However, early SDS concepts remain ground-dependent: without onboard intelligence, reconfigurable satellites rely on operators for analysis and tasking decisions, introducing latency and limiting responsiveness. The logical progression is from software-defined to intelligence-defined space systems, where reconfigurable hardware pairs with edge AI for real-time processing. Edge AI empowers satellites to verify, analyze, prioritize, and act on data on-orbit, enabling genuine operational autonomy beyond ground dependence for critical decisions. Software-defined satellite architectures increasingly align with established software-defined radio concepts and definitions standardized by IEEE [1].

II. EDGE AI IN SDS ARCHITECTURES

Edge AI involves deploying machine learning algorithms directly on spacecraft, co-located with sensors and onboard compute. Within SDS architectures, edge AI enables:

- Real-time data assessment and triage at the source
- Dynamic adjustment of communications waveforms and bandwidth allocation
- Mission task prioritization based on environmental conditions
- Reduction of unnecessary downlinks of low-value data

By moving inference on-orbit, SDS platforms significantly reduce end-to-end latency and enhance operational efficiency. This capability is vital for responsive Earth observation, flexible communications services, and shared satellite platforms supporting multiple stakeholders. Together, these capabilities enable genuinely dynamic mission tasking: SDS platforms can autonomously re-prioritize sensing modes, revisit patterns, and communications resources in response to evolving conditions and user demand, rather than following a static, pre-planned schedule.

III. A HETEROGENEOUS COMPUTE ARCHITECTURE

Heterogeneous onboard compute is most effective when each processing element is assigned tasks aligned with its architectural strengths rather than treated as interchangeable resources. Edge AI workloads must indicate balanced performance with power consumption and operating temperatures. Onboard software architectures for space systems are commonly developed in accordance with established European space software engineering standards [2].

A. Power and Thermal Trade-Offs

Power: Each compute element exhibits distinct power characteristics: CPUs have moderate-to-high power draw but are mitigated by use in general-purpose orchestration tasks; GPUs consume significant power but provide the highest efficiency for parallel AI/ML inference workloads; FPGAs deliver the lowest power consumption, particularly advantageous for complex, pipeline-heavy signal processing and deterministic detection functions. System architecture strategies include model optimization and dynamic scheduling of AI workloads to align with power availability. The coordinated use of these three compute elements—assigning tasks to the most power-efficient processor for each function—enables optimal power management and extended mission endurance.

Thermal: Thermal considerations are equally important when managing resources. Each compute element exhibits distinct thermal characteristics: CPUs generate moderate-to-high thermal density but are mitigated by lighter utilization during normal operations; GPUs produce very high thermal density, especially during sustained inference work, creating acute thermal management challenges; FPGAs run the coolest, dissipating minimal heat even under continuous operation. AI workloads can create highly localized heat densities on Commercial-Off-The-Shelf (COTS) accelerators, especially during burst inference work. SDS architectures must therefore coordinate AI duty cycles with thermal design through packaging, conduction paths, and operational scheduling to avoid performance throttling or premature component aging while still meeting mission responsiveness targets. Strategic assignment of workloads to thermally efficient processors (FPGAs for sustained tasks, GPUs for burst operations) further optimizes thermal margins.

B. Radiation and Reliability

AI-enabled compute systems must be designed with radiation tolerance. This involves hardware selection and software-level resilience strategies. Ensuring predictable behavior under upset conditions is critical for high-reliability missions.

A significant challenge is the performance gap between radiation-hardened processors and COTS AI accelerators. While traditional space-qualified processors offer limited computational power, modern COTS AI accelerators, including GPU-based devices, provide orders of magnitude higher performance. Bridging this gap requires a deliberate system-level hardening strategy.

A common mitigation strategy employs a radiation-tolerant FPGA as a hardware gatekeeper, validating data and system state before inference execution. This pattern enables selective use of high-performance COTS accelerators while preserving predictable system behavior under radiation-induced upset conditions.

C. Role of FPGAs in Spaceborne Edge Computing

FPGAs are reconfigurable hardware devices capable of performing different functions, or multiple functions concurrently, through software rewrites. Unlike sequential CPU execution, FPGAs enable highly parallel, deterministic processing with extremely low latency and power consumption, making them ideal for space applications. On orbit, an FPGA can be reconfigured for tasks such as radar processing, video filtering, encryption, or high-speed routing without physical hardware changes, which is crucial for SDS platforms with long mission lifetimes.

Within a satellite architecture, FPGAs are best suited for:

- Immediate verification and formatting of raw sensor data
- Task-specific sensor integration and preprocessing
- Deterministic, low-latency detection functions

By offloading these functions to hardware, FPGAs reduce the processing burden on higher-level compute elements, ensuring that only validated, well-structured data proceeds through the onboard pipeline.

D. GPUs and CPUs: Complementary Roles in Edge AI

CPU: System Orchestration—The Central Processing Unit (CPU), typically based on low-power Advanced RISC Machines (ARM) architectures, serves as the system manager of the spacecraft. Its responsibilities include:

- Spacecraft command and control
- Operating system and mission software execution
- Communications management and protocol handling Enforcement of mission priorities and operational constraints

CPUs provide deterministic performance and are well-suited for fault-tolerant operations in radiation-aware environments, ensuring safe and predictable spacecraft behavior.

GPU: High-Throughput AI and Analytics—Graphics Processing Units (GPUs) are designed for massively parallel computation, excelling at high-throughput workloads such as AI/ML inference, image and video processing, and multi-sensor data fusion. When deployed at the edge, GPUs enable satellites to perform advanced analytics directly on orbit, including:

- Object and vessel detection
- Pattern recognition and anomaly detection
- Sensor fusion across optical, radar, and RF domains

GPUs dramatically accelerate data analysis and support near-real-time insight generation from onboard sensors by executing thousands of threads simultaneously.

IV. A COORDINATED ONBOARD DATA FLOW

A representative onboard data flow illustrates this coordination:

Sensor → FPGA → CPU → GPU → CPU → Downlink

In this flow, sensors deliver raw data to the FPGA for immediate verification and formatting. The CPU then routes validated data to the GPU for AI-driven analysis. Analytical results, such as detections, tracks, or alerts, are returned to the CPU, which enforces mission rules, manages communications, and downlinks only curated, high-value information. This architecture minimizes latency, reduces power consumption, and significantly lowers downlink requirements compared to architectures that rely on ground-based processing and analysis. In effect, inference and decision-making on orbit reduce response latency, manage scarce downlink capacity more intelligently, and focus ground operations on high-value insight products rather than bulk data transport. Curated telemetry products and downlink data flows are typically formatted to comply with CCSDS space data system standards [3].

V. VERIFICATION, VALIDATION AND TRUSTWORTHY AUTONOMY

For all stakeholders, the adoption of autonomous systems depends on trust. AI-driven decision-making must be verifiable, explainable, and bounded within well-defined operational constraints. From a system design and test perspective, verification and validation (V&V) of onboard AI systems require a combination of simulation, hardware-in-the-loop testing, and on-orbit monitoring. SDS platforms should incorporate mechanisms to assess AI performance over time and to revert to safe operating modes when confidence thresholds are exceeded.

Trustworthy autonomy does not imply unrestricted decision-making. Rather, it reflects a layered approach where AI augments human operators by handling routine optimization and data triage, while preserving human oversight, or "Human-on-the-Loop," for higher-level mission objectives.

In practice, this means that autonomous behaviors, such as on-orbit task reprioritization or anomaly-driven re-tasking, are only enabled once they have been exercised and validated through a full design-and-test loop, and they remain continuously monitored and bounded by mission-level safety constraints. The goal is not maximal autonomy, but safe, assured autonomy whose behavior can be predicted, monitored, and governed throughout the mission lifecycle.

Trustworthy autonomy emphasizes bounded decision authority, where AI augments routine optimization and data triage while mission-level objectives and escalation remain under human oversight.

VI. OPERATIONAL BENEFITS

When edge AI is tightly integrated with software-defined radios, reconfigurable sensors, and heterogeneous onboard compute, Software-Defined Satellites unlock benefits that are operational, not merely architectural. Satellites move from passive collection and ground-dependent interpretation toward time-sensitive on-orbit assessment, prioritization, and action. For operators, the value is straightforward: faster

response, lower downlink burden, more efficient use of spacecraft resources, and mission performance that can evolve after launch.

Representative applications include event-driven Earth observation and maritime anomaly detection, where onboard triage prioritizes exceptions rather than continuous full-resolution downlink.

Several benefits are especially relevant for European and international civil, commercial, and institutional users:

- **Responsive Earth Observation:** Onboard AI can detect events of interest in near real time and prioritize imaging, processing, or transmission accordingly.
- **Flexible and Persistent Communications:** AI-enabled SDS platforms can dynamically adapt link management, bandwidth allocation, and pointing plans, supporting more persistent, service-like connectivity for users even as orbital geometry and network conditions change.
- **Multi-Mission Utility:** Shared spacecraft resources can be balanced across multiple users, payloads, or regions without rigid pre-allocation.
- **Operational Efficiency:** By transmitting curated, high value information instead of full raw datasets, SDS platforms reduce downlink demand and lighten the burden on the ground segments.

VII. On-Orbit Cybersecurity

The software-defined flexibility of SDS platforms, while operationally powerful, also expands their attack surface. Protection against unauthorized command injection, model poisoning, and communications spoofing is critical. For DoD, NASA, ESA, and institutional customers, on-orbit cybersecurity is a procurement requirement. On-orbit command and data protection mechanisms increasingly align with CCSDS security recommendations for space data links [4].

Thus, SDS architecture incorporates cybersecurity as a system-level design requirement. The CPU acts as the spacecraft policy enforcement point, authenticating and integrity-checking all uplinked commands, including model update packages. The FPGA layer provides hardware-level encryption and routing control, preventing inter-component communication interception or manipulation. Secure boot sequences are enforced on all compute elements to prevent unauthorized firmware or model substitution.

At the AI layer, model integrity verification, confirming that an uploaded inference model matches a cryptographically signed reference, prevents adversarial model substitution. Operational boundaries enforced by the CPU's rules-of-engagement layer ensure that even a compromised AI model cannot direct the spacecraft to take actions outside predefined mission parameters. These measures provide defense-in-depth consistent with emerging agency guidance on space system cybersecurity, aligning with the trustworthy autonomy framework.

VIII. Sidus Space Approach: Integrated Flight Proven System

Sidus Space's approach to SDS and edge AI emphasizes flight readiness, system integration, and operational integrity. Rather than treating AI as an add-on, Sidus integrates edge intelligence as a core architectural element, co-designed with radios, sensors, and onboard compute. This system-level perspective enables practical trade-offs between autonomy and resource constraints, ensuring AI capabilities deliver measurable operational value without compromising mission reliability. The result is an SDS platform capable of evolving over time while maintaining predictable and trustworthy behavior.

A. VPX-Based Heterogeneous Compute Stack

Sidus Space implements heterogeneous compute using a standardized VPX backplane architecture, allowing CPUs, GPUs, and FPGAs to operate as a cohesive system within a single form factor. This modular VPX approach enables customers to mix and match compute elements based on mission needs, reducing integration risk while supporting scalable, upgradeable SDS platforms.

These architectural decisions are informed by Sidus Space's ongoing flight programs, where the same VPX-based FeatherEdge stack is integrated into operational spacecraft and exercised as part of real mission campaigns, rather than existing solely as a stand-alone technology demonstrator.

B. Ground Segment Integration and Cloud-Native Operations

Edge AI on SDS platforms must operate in concert with the ground segment. The industry is moving toward cloud-native ground architectures and the convergence of space and terrestrial network management, driven by network standardization, cloudification of mission operations, and optimization of ground terminal scheduling. In this model, onboard inference handles time-sensitive decisions autonomously, while the ground segment manages model updates, mission policy, and higher-order analysis across the constellation.

Sidus Space's Cielo software layer serves as the interface between the onboard compute stack and the ground segment. Cielo provides the command-and-control abstraction necessary for operators to push model updates, adjust inference thresholds, and reconfigure processing pipelines without disrupting core spacecraft operations. This means that the intelligence of an SDS platform is not static at launch but can be continuously refined in response to new mission requirements, environmental conditions, or evolving customer priorities.

For multi-satellite deployments, ground-space software integration enables constellation-level coordination: tasking policies can be distributed across nodes, detection results from one spacecraft can inform the collection priorities of another, and downlink bandwidth can be managed across the fleet. The ability to manage SDS platforms through cloud-native interfaces, rather than bespoke proprietary ground systems, will become a key differentiator for interoperable, scalable space infrastructure.

C. Operational Use Cases

Example 1: European Civil Protection and Climate Resistance: Floods, fires, and infrastructure monitoring with on-orbit triage. Sidus Space can support EU agencies and member-state responders by deploying SDS platforms that monitor coastlines, wildfire-prone regions, and critical infrastructure corridors. Instead of downlinking every

collected image, the spacecraft could identify indicators such as flooding, smoke plumes, damaged transport links, or unusual surface changes directly on orbit, then prioritize only those scenes that merit immediate review. In day-to-day operations, this reduces analyst workload; during a storm, flood, or wildfire event, it helps emergency teams receive relevant alerts and targeted imagery faster, improving response coordination across borders.

Example 2: Global Maritime Safety and Supply Chain Awareness: Detecting anomalous vessel behavior and environmental events. Sidus Space can support European and international maritime stakeholders by using SDS platforms to monitor major shipping routes, offshore energy infrastructure, and environmentally sensitive waters. In this role, the spacecraft would not need to send back continuous full-resolution coverage of the ocean. Instead, onboard analytics could flag unusual vessel behavior, possible oil slicks, or activity near offshore assets, and transmit concise alerts with supporting imagery or metadata. For port operators, insurers, coast guards, and environmental authorities, the service becomes simpler and more useful: not more data, but earlier awareness of exceptions that may require action.

In both cases, the central advantage is the same. The spacecraft does not replace the operator; it reduces the time and bandwidth required to surface what matters. This is where edge AI becomes the operational enabler of the software-defined paradigm, allowing the satellite to support human decision-makers with faster triage, more relevant data products, and greater adaptability across evolving mission demands.

D. Competitive Landscape and Sidus Space Differentiation

The SDS and onboard AI market is maturing, with various commercial and institutional players deploying or developing software-defined, AI-enabled spacecraft. Sidus Space differentiates itself in three key areas:

- Integrated, Flight-Proven System: The FeatherEdge GPU platform is designed as an integrated, flight-proven system, not a collection of separately sourced components. This reduces integration burden and operational risk and ensures validated performance within a single edge compute device.
- 3U VPX Architecture: The VPX architecture enables a modular, mix-and-match compute configuration. This allows adaptation to mission-specific requirements without re-architecting the core platform, such as the command and data handling suite, offering greater configurability than a set of single-purpose payload components.
- Comprehensive System with Operational Support: Sidus's combination of onboard hardware, the Cielo software layer, and extensive spacecraft integration experience means customers acquire a complete system with defined interfaces and operational support, rather than just a compute board requiring independent system integration. These differentiators are particularly relevant for customers requiring traceability, accountability, and a clear path to flight heritage.

IX. CONCLUSION

Software-Defined Satellites represent a critical advancement toward more adaptable and cost-effective space systems. Their full potential is best realized through the deep integration of edge AI and a deliberately architected heterogeneous compute stack comprising CPUs, GPUs, and FPGAs. By intelligently assigning tasks to each element, SDS platforms can operate autonomously, efficiently, and predictably within the resource-constrained space environment.

This architecture transforms satellites from software-defined platforms into intelligence-defined elements of distributed space infrastructure. In that sense, edge AI is what ultimately pushes these systems beyond software-defined toward truly intelligence-defined space infrastructure. This approach strongly supports goals for resilience, interoperability, and long-term sustainability across government, commercial, and institutional domains. As the adoption of SDS accelerates, the co-design of AI, reconfigurable hardware, and onboard compute will be paramount in delivering the next generation of autonomous space systems.

Looking ahead, key open questions include how to standardize assurance and certification processes for AI-enabled SDS architectures and how to share lessons learned on autonomy anomalies across agencies and operators to accelerate the maturation of trusted orbital edge intelligence.

REFERENCES

[1] IEEE Standards Association, “IEEE Standard for Definitions and Concepts for Dynamic Spectrum Access (Software-Defined Radios),” IEEE Std 1900.1-2019, Apr. 2019.

[2] European Cooperation for Space Standardization, “Space Engineering—Software,” ECSS-E-ST-40C Rev. 1, Apr. 2025.

[3] Consultative Committee for Space Data Systems (CCSDS), “Telemetry Channel Coding,” CCSDS 131.0-B-5, Dec. 2024.

[4] Consultative Committee for Space Data Systems (CCSDS), “Space Data Link Security Protocol,” CCSDS 355.0-B-2, Blue Book, Jan. 2023.