Electronic Design

Hussein Osman, Director of Segment Marketing,
Lattice Semiconductor, https://www.latticesemi.com/

FPGAs: Co-Processors of the Future for Edge AI?

By taking on security, control, and sensor duties, FPGAs are emerging as the glue for increasingly complex edge AI systems.

As AI continues to advance, the distribution of compute and processing workloads is expanding out of traditional data centers and to the edge — everywhere from under the hood of cars to inside robotics systems. This modular approach to processing helps increase system speed and capabilities, but such performance comes at a cost. Distributed processing means a lot more connections, sensors, and data.

Meanwhile, the volume of auxiliary tasks like power sequencing, board control, I/O management, data management, and security is also on the rise. Paired with faster upgrade cycles and diverse system demands, it's becoming clear that no single traditional chip — GPUs, CPUs, XPUs, or otherwise — can manage this complexity on its own.

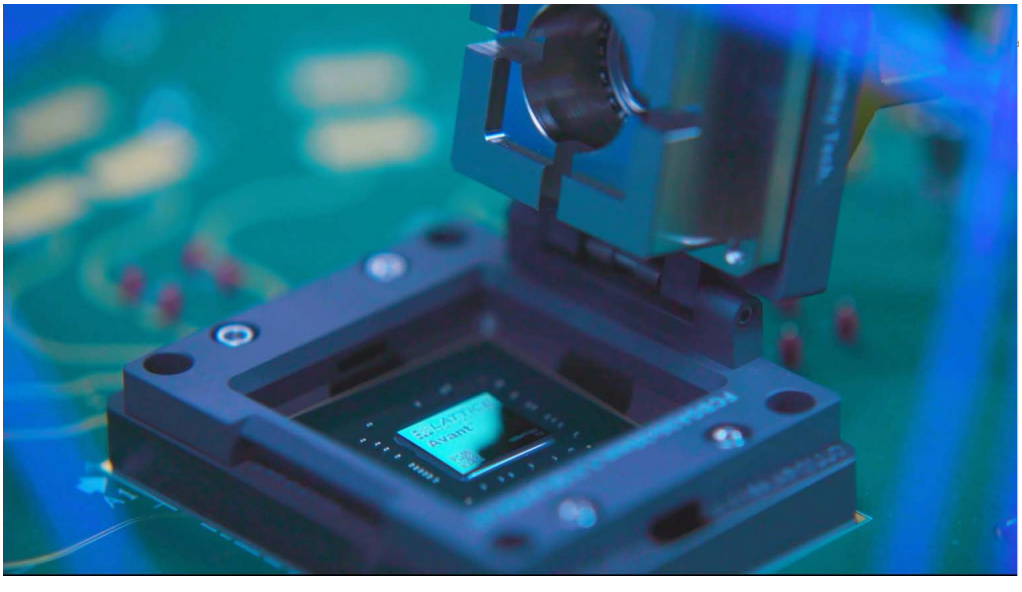
To address this growing challenge, engineers are turning to new architectures that depend on "teams" of processors to advance AI capabilities. The approach uses additional "support silicon" processors to offload high-intensity tasks from centralized chips, delivering a critical performance boost while designers work to refine top-of-stack processors.

That said, top-of-stack improvement can only push performance so far. To avoid bottlenecks that throttle next-gen AI systems, designers will need to devote time and research to optimizing performance at every level, and that starts with silicon.

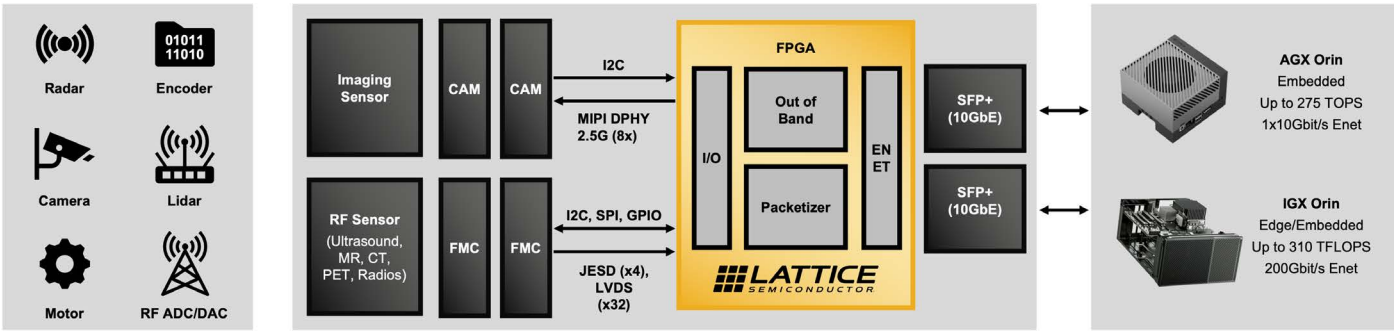
Redefining the Role of Support Silicon

The concept of outsourcing tasks to secondary chips is already common. In data centers, data processing units (DPUs) offload networking and storage from the CPU, while in robots and vehicles, microcontrollers (MCUs) run control loops and enforce safety. But as AI workloads grow, these chips leave gaps, and a new class of companion chips is stepping in to fill them.

Here's where these co-processors are already making a



1. A technician positions semiconductor chips on a testing platform. (Credit: Lattice Semiconductor)



2. In sensor bridges, FPGAs deliver tightly synchronized data streams directly into GPU memory. (Credit: Lattice Semiconductor)

difference:

- **Data centers**, where companion chips work alongside CPUs, GPUs, and DPUs to take on board-level control and management, such as power sequencing, system bring-up, health monitoring, and secure boot, so that the primary processors can stay focused on AI workloads.
- **Industrial robotics**, where companion chips sit between sensors, actuators, and computing hardware, aggregating camera, radar, and other sensor data and handling deterministic I/O and timing so that perception and control loops stay precise as sensor counts grow.
- **Automotive systems**, where companion chips sit between the central system-on-chip (SoC) and safety MCUs to manage interfacing, sensor aggregation, and power sequencing, keeping safety-critical functions responsive in real-time while letting automakers adopt new SoC generations without redesigning the system.

In each case, the primary processor performs best when it can offload non-core work. But not all companion silicon is equal: Reaching the next level of functionality requires processors that bring determinism, parallelism, and programmability to the table. These qualities enhance overall system performance rather than simply sustain it.

The best edge AI architectures tap into the same principle: A team of specialists, each with a defined role, together enabling performance that no single processor could sustain alone *(Fig. 1)*. The co-processor handles control, interfacing, security, power management, and sensor aggregation so that the central processor can devote its resources entirely to AI compute. Instead of subordinates filling gaps, these chips are purpose-built to make the primary processor more effective than it could be on its own.

FPGAs Bring Flexibility to AI Companion Role

While several types of silicon can play the companion role, field-programmable gate arrays (FPGAs) offer a flexible platform for handling tasks that don't justify tying up valuable accelerator resources.

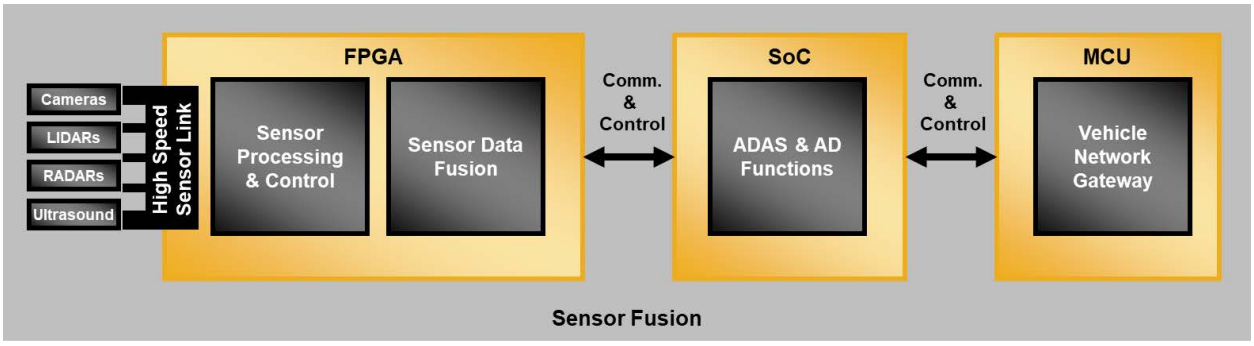
Low-power FPGAs are particularly well-suited to the companion role: Operating at low power envelopes, they handle critical non-core tasks without competing for the power budgets that primary processors require to sustain AI workloads. As a result, they have the potential to serve as "everywhere companion chips," equally compatible with edge, industrial, and automotive system as they are with data center racks.

FPGAs are also well-equipped to support future innovation in edge AI operations, bringing capabilities that general-purpose MCUs can't easily match. FPGAs can execute multiple sensor, control, and I/O tasks in parallel with deterministic, low-latency timing. As a result, they can help maintain performance in demanding, high-risk environments and systems with significant resource constraints.

Furthermore, FPGAs are designed for flexibility, and critically, they don't compete with the primary processors they sit alongside. Because they're processor-agnostic, FPGAs can work closely with popular GPUs, CPUs, NPUs, and SoCs. Moreover, their operations can be tailored to complement the preferred primary chip, making them a low-risk architectural partner rather than a future design lock-in.

As cyberattacks become more advanced, FPGAs are becoming an increasingly important part of the security stack. They can serve as a hardware root of trust (RoT), supporting secure boot, device authentication, and the protection of sensitive data. Some FPGAs support post-quantum cryptographic (PQC) algorithms, including CNSA 2.0 and NIST standards, and since they're reprogrammable, algorithm updates can be deployed without a full system redesign or silicon replacement.

This feature set allows FPGAs to serve as a unifying system layer that reliably offloads complexity, preserves determinism, and provides continuity throughout system lifecycles.



3. In emerging ADAS architectures, FPGAs act as the control center for sensor fusion. (Credit: Lattice Semiconductor)

Where FPGAs Fit in Emerging AI Architectures

FPGAs provide engineers with a way to better connect compute, security, I/O, and sensors wherever AI workloads run, enhancing system operations in each of these applications.

In data centers, FPGAs can handle board-level control and management. While CPUs and newer classes of chips such as DPUs continue to own their specific tasks, FPGAs can take on power sequencing, system bring-up, health monitoring, and orchestration. They also handle higher-value functions such as attestation, secure boot, and PQC-ready security, where deterministic, low-latency response is non-negotiable.

Because the FPGA is first on and last off, it also provides a stable control and security foundation across processor refresh cycles. This tends to be important given the constant evolution of rack infrastructure in AI data centers.

In industrial robotics, too, FPGAs can bridge diverse sensor interfaces such as MIPI and GMSL to high-speed Ethernet, enabling precise, deterministic I/O, timing, and data conditioning. For instance, NVIDIA's Holoscan Sensor Bridge technology uses FPGAs to handle high-bandwidth data streams from cameras, radars, and other sensors that are beyond the scope of MCUs. The FPGA can adapt MIPI CSI-2, GMSL inputs, and other sensor interfaces to high-speed Ethernet for seamless transfer.

The FPGA puts sensor data into packets with low latency and sends it over Ethernet, where the Holoscan technology delivers it directly to GPU memory with minimal host CPU involvement. By delivering tightly synchronized data streams directly into GPU memory, the FPGA enables real-time AI perception pipelines in robotics systems while the primary SoC or other processor remains free to focus on inference *(Fig. 2)*.

In automotive systems, FPGAs can anchor the control, safety, and interface logic under the hood. In one emerging architecture, the FPGA sits between the central SoC and an external, safety-focused MCU: The SoC serves as the central brain for AI workloads, the MCU acts as a firewall against unsafe commands, and the FPGA handles deterministic sensor aggregation, interfacing, and monitoring between them *(Fig. 3)*.

Because the underlying logic is programmable, it helps ensure system longevity even as vehicle functions change. Automakers are able to adopt new SoC generations without redesigning the system architecture, reducing rework and revalidation effort. FPGAs can also serve as a deterministic hub, deciding when to put predictability and stability over raw throughput to preserve real-time responsiveness.

FPGAs are all about flexibility, so positioning them as a specialized companion chip may seem counterintuitive. But it's the same flexibility that allows a single device to be tailored to a specific companion role. FPGAs can be programmed to specific parameters and adjusted as needs change, ensuring specialization and refinement will continue after deployment.

FPGAs: Everywhere Co-Processors for the Edge

Optimization at the top of edge AI deployments will inevitably hit a ceiling, and generic support silicon will not be enough to push technology past that plateau. What designers need instead are true companions to their primary processors. They need complementary components capable of handling a wide range of non-core tasks, eliminating bottlenecks, and actively elevating system performance rather than simply sustaining it.

Low-power FPGAs represent a natural starting point as designers embrace this shift. Their blend of performance, efficiency, determinism, security, and programmability makes them well-suited to serve not as afterthoughts, but as foundational members of a well-considered architecture. As edge AI scales, the designers who treat every chip in the system as a contributor, not just an accessory, will be the ones building the architectures that endure.