Redefining Real-Time AI at the Tactical Edge

James Spriet
EW/ISR SME | Strategy & Comms Manager
jspriet@ampex.com

Data Overload at the Edge

In tactical environments, platforms are flooded with data from an expanding array of advanced sensors, such as Radar, SIGINT, and EO/IR. The sheer volume of this information overwhelms traditional edge computing systems.

These legacy architectures, built on PCIe, create latency bottlenecks by separating CPU and GPU memory. This forces data through slow, serialized pathways, making true real-time sensor fusion impossible. The result is delayed threat recognition, extended sensor-to-shooter timelines, and a critical dependency on remote "reach back" processing, which is an unacceptable liability in contested environments.

A New Architecture for Cognitive AI

Ampex's concept explores a revolutionary edge compute architecture to address this critical capability gap. It's not just an upgrade; it represents a fundamental redesign to eliminate data bottlenecks and unleash the full potential of AI, in place, at the tactical edge.

The Unified Compute Fabric

The heart of our concept is the NVIDIA Grace-Blackwell Superchip, which merges CPU and GPU memory spaces into a single, coherent "zero-copy" fabric using NVLink C2C technology. This would remove the PCIe bottleneck entirely, enabling genuine cognitive processing and AI workloads that were previously impossible at the edge. Data could move directly between the CPU and GPU at up to 900 GB per second. The architecture is envisioned for the most demanding missions, supporting near-terabit data ingestion (~750 Gb/s) and featuring 187 TB+ of local NVMe storage. This would allow platforms to process and store everything at line rate. Our design is meant to unlock PetaFLOP-class compute density, delivering data center performance in a semi rugged, deployable form factor.

Figure 1-NVIDIA BLACKWELL GPUs

Key Capabilities Enabled

For the first time, platforms could ingest and correlate high-fidelity data from Radar, RF, and EO/IR sensors simultaneously, creating a complete operational picture in real-time. Our concept moves beyond simple data processing. It is envisioned to enable genuine cognitive AI, allowing platforms to understand the electromagnetic environment, identify threats, and support autonomous decision-making locally. By eliminating reliance on remote processing, our design aims to help platforms remain fully effective and autonomous, even when communication links are jammed, degraded, or denied.

From Passive Collector to Decisive Asset
Ampex's concept will be engineered from the ground up to survive and thrive in the harshest operational environments. It will be designed to transform a platform from a passive data collector into an active, decisive combat system capable of sensing, understanding, and acting faster than the adversary.

Rugged by Design, Resilient in Operation
The architecture is designed to be packaged in a 4U rugged chassis compliant with MIL-STD-810H and MIL-STD-461G, ensuring reliability under extreme conditions. We plan on ensuring it comes with a well designed hybrid liquid-to-air cooling system. It would be designed to maintain peak performance and hold clocks without throttling, even at ambient temperatures up to +71 °C (+160 °F). Redundant pumps and vapor-chamber heat spreaders would ensure mission continuity. The conceptual system features redundant, isolated power distribution designed for the unique demands of air, ground, and naval platforms, protecting against transients and ensuring stability.

Secure, Intelligent, and Scalable
In this concept, security is built in, not bolted on. Our concept is founded on a Hardware Root of Trust (HRoT) and secure boot processes. It would protect data in use with confidential computing and encrypted memory, ensuring mission integrity from power-on to power-off. The system would leverage a cutting-edge software stack including CUDA, TensorRT, and DOCA to optimize, manage, and deploy AI models efficiently, ensuring deterministic, low-latency performance.

A Platform for the Future, Not Just the Present
The pace of AI development is relentless. The sophisticated models of today, from computer vision to RF machine learning, will be dwarfed by the trillion-parameter foundation models and complex cognitive algorithms of tomorrow.

A critical flaw in many current "AI-ready" edge systems is that they are built to solve yesterday's problem. They lack the memory bandwidth and compute density to handle this evolution, rendering them obsolete upon delivery.

This Grace-Blackwell concept is architected differently. It is not a stopgap solution but an enduring platform. Its PetaFLOP-class performance and, most importantly, its 900 GB/s unified memory fabric are precisely what next-generation AI requires. This design provides a credible, scalable path to deploying future autonomous systems, ensuring warfighters maintain a decisive advantage as new threats and new technologies emerge. It protects the investment by providing a foundation that can grow, not a bottleneck that must be replaced.

The Strategic Impact
This concept represents a potential quantifiable leap in capability, providing the technical foundation for the next generation of EW/ISR. By fusing data and running AI locally, our architecture could dramatically shorten the sensor-to-shooter cycle, creating a decisive time advantage. The architecture is also designed to be scalable. Multiple systems could be clustered to form a "Swarm," enabling distributed inference and collaborative autonomy across multiple platforms.

Our concept provides a credible conceptual path to the autonomous systems of the future, turning massive data streams into a decisive operational advantage.

If you're interested in learning more about this concept and how it could enable your mission, please reach out to our team at sales@ampex.com.

26460 Corporate Avenue, Hayward, CA 94545
(650) 367-2011 | ampex.com