Improving Time-To-Insight for National Security Science with ATS-5, Called Mission

A New Era for ASC Computing

The fifth Advanced Technology System (ATS) in the Advanced Simulation and Computing (ASC) program—ATS-5, now called Mission—will be a critical resource supporting the NNSA stockpile management and modernization programs.

Mission will support current and future simulation codes, future artificial intelligence (AI)–augmented simulations, and AI-augmented workflows to support the design, assessment, and certification of the United States (U.S.) stockpile. By supporting hero-class high-fidelity three-dimensional (3D) weapon simulations, massive ensembles of common model simulations, and frontier-scale AI models on a single platform, Mission will reduce time-to-insight from months to days.

As the first ATS resource to combine multiple architectures to accelerate the full end-to-end workflow, Mission will improve NNSA's ability and agility to manage the stockpile with modeling/simulation and AI.

Codesigned by Los Alamos National Laboratory (LANL) for National Nuclear Security Administration (NNSA) workloads, Mission will be built by Hewlett Packard Enterprise (HPE) and NVIDIA.

What Mission Is

Computing technology is rapidly diversifying to continue to deliver performance improvements in the post–Moore's law era. Mission is the first Advanced Technology System with diverse, purpose-built technologies to support NNSA's most pressing workloads in support of the Weapons Program.

Rather than a single, general-purpose architecture, Mission is composed of three purpose-built architectures: Ranger, Sandstone, and Starlight.

- Ranger will continue LANL's strong support for the highest complexity 3D multi-physics simulations.
- Sandstone will support rapid turnaround of moderate scale 3D simulation suites and provide an optimized environment for large-scale AI training.
- Starlight will provide the first scale-up architecture for frontier AI inference and strong scaling of modeling and simulation (MODSIM).

These architectures will be integrated with a common software environment and shared high-performance storage optimized for MODSIM and AI.

The ASC program has been the catalyst for the most capable codes in the U.S. weapons complex, which has driven sustained technology improvements since its inception in 2005. Over the past decade, however, these improvements have become more difficult to achieve as the ability to continue to shrink transistors has hit physical limits.

In response, in 2014, LANL embarked on a "deep" hardware-software co-design initiative with multiple technology vendors including Broadcom, Cavium, Marvell, NVIDIA, and HPE/Cray. These collaborative efforts spanning over a decade have delivered

- The first new HPC optimized CPU in a generation,
- Microarchitecture that better matches the unique memory-access patterns of our multi-physics codes,
- Advanced Chip-to-Chip (C2C) connectivity of CPU and GPU processors with massive bandwidth and low-latency, and

- Architectures that flexibly compose these technologies to enable optimal performance across a variety modeling/simulation and AI workloads for stockpile modernization,
- All while ensuring broader market alignment of these technologies to leverage economies of scale in the trillion-dollar computing industry.

Future co-design for hardware-software is essential. Maintaining strong, direct partnerships with technology vendors will remain essential in supporting NNSA's national security mission.

Mission will replace Crossroads (ATS-3) as a primary resource to support the stockpile modernization mission and current and planned life extension programs (LEPs). Delivery of the Mission system will begin in 2027 with a planned transition to Secret Restricted Data (SRD) production computing in 2028.

Why Mission Matters

Mission is state of the art—a flagship in high-performance computing with cutting-edge technologies that will deliver major gains in performance and efficiency.

Different from previous systems, Mission is driven by deep workload analysis and designed around NNSA's most pressing application bottlenecks, such as memory-bound sparse accesses and branching found in complex multi-physics codes. Mission is the first system designed to optimize the end-to-end performance of workflows that couple MODSIM, AI, and data analysis.

As resource for the NNSA Tri-labs—Los Alamos National Laboratory, Lawrence Livermore National Laboratory, and Sandia National Laboratories—workload analysis involved subject matter experts spanning all three laboratories, multiple integrated code teams, and weapons science teams. This analysis drove microarchitecture features, technology composition, and the overall system architecture.

Mission will deliver significant performance improvements for our highest-complexity 3D multi-physics simulations on the HPC optimized NVIDIA Vera Vera CPU superchip. NVIDIA's Vera/Rubin CPU/GPU will support GPU-capable workloads spanning MODSIM and AI with the most technologically advanced CPU and GPU processors available.

By concentrating on accelerating time-to-insight of the entire workflow Mission will turn once-heroic, months-long 3D simulation campaigns into routine, fast, and energy-efficient workflows that sustain the U.S. nuclear stockpile confidence into the 2030s and beyond.

Partnerships Make ATS-5 Possible

LANL has concentrated on tailoring hardware technologies in Mission to address simulation characteristics that are not well represented in the broader cloud- and AI-driven market. This focus on the unique characteristics of Laboratory applications has allowed LANL's technology partners to address these unique needs with modest but highly impactful enhancements to their base technology portfolio. By focusing LANL's investments on the unique aspects of Laboratory workloads while leveraging the massive base of investment in cloud and AI for areas of similarity, LANL's investments have an oversized effect for Laboratory applications.

Over a decade of coordinated investment across LANL, the design agencies, ASC, and industry partners has delivered these technologies advancements: ARM-based HPC CPU architectures optimized for sparse, branchy code; advanced chip-to-chip connectivity for flexible and tight multi-processor integration; and GPU technologies tuned for MODSIM and AI needs.

Furthermore, Mission will provide a foundation for integrating future AI capabilities into ASC workflows.

What Is New with Mission

Mission is a heterogeneous computing resource providing purpose-built support for NNSA's most pressing HPC and AI workloads. This comprehensive computing platform is designed to support the diverse mission-critical workloads for national security.

Mission integrates three specialized partitions, each optimized to support the most important components of the complex workloads of the future:

- Ranger: Sparse optimized compute for large-scale high-complexity modeling and

simulation with large memory footprint and sparse memory access requirements.
- Sandstone: Dense optimized compute for artificial intelligence and machine learning (AI/ML) training, including models with 500+ billion parameters, and moderate-scale 3D simulations with dense memory access and high floating-point compute intensity.
- Starlight: AI inference optimized compute for advanced frontier models, agentic AI workflows, and strong-scaling MODSIM with large shared-memory and extreme-bandwidth scale up network requirements.
- All flash high-performance storage connected to each of these partitions with flexible configurations optimized for modeling/simulation and AI/ML needs.

Underpinning these capabilities is Mission's robust infrastructure supporting workflow management, Python notebooks, and planned compute lifecycles to ensure operational continuity and maximize mission impact.

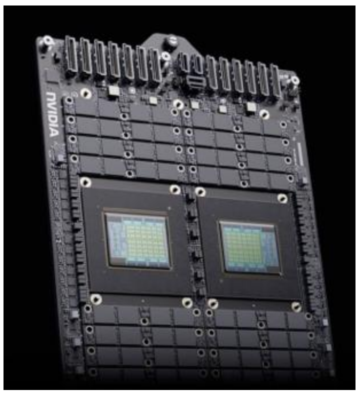

Figure 1. The Vera Vera CPU "Super-Chip" of Ranger (image credit: NVIDIA)

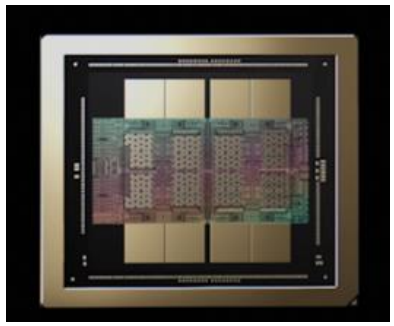

Figure 2. The Rubin GPU Chip of Sandstone and Starlight (image credit: NVIDIA)

Figure 3. The NVL72 rackscale architecture of Starlight (image credit: NVIDIA

Misson's Three Integrated Partitions

Mission implements this heterogeneous approach through three integrated purpose-built partitions to accelerate the most critical components of our workloads:

The Vera Vera CPU Partition—Ranger—delivers high-frequency, high-bandwidth cores with large aggregate and per core memory. Early testing illustrates performance improvements on a single Vera chip compared to two Intel SPR HBM chips (see Figure 4 and Figure 1). Comparing Mission-Ranger to Crossroads node to node performance improvements of over *3x across* the full suite of Tri-lab workloads are expected with little more than a recompilation of the code needed to achieve this performance uplift.

The Vera Rubin NVL4 GPU Partition—Sandstone—provides dense optimized compute and high-bandwidth memory. Processors are integrated via Chip-to-Chip Vera CPU–Rubin GPU coupling and NVLINK interconnecting four GPUs per node (see Figure 2). Table 1 summarizes key metrics across three generations of NVIDIA GPUs.

Figure 4. Early Vera chip testing

Mission-Sandstone is expected to provide 13x performance improvements in Single Physics MODSIM and AI-enabled Single Physics MODSIM when compared to Crossroads. When compared to Venado (Grace Hopper), Mission-Sandstone (Vera Rubin) will provide up to *15x higher AI inference performance and ~5x higher memory bandwidth. When compared to Vera Vera, Rubin will provide ~15x higher memory bandwidth,* enabling significant performance upside for GPU capable multi-physics codes.

Table 1. Vera Rubin will provide a significant performance improvement over Grace Hopper (Venado) for memory bandwidth–bound MODSIM and dense compute–hungry AI workloads.

	Hopper	Blackwell	Rubin
FP8 AI (sparse)	4 PF	10 PF	35 PF
Memory capacity	96GB	186 GB	280 GB
Memory BW	4–5 TB/s	8 TB/s	22 TB/s
NVLink BW	900 GB/s	1.8 TB/s	3.6 TB/s
Inference perf	Baseline	~2.5–3.5×	~10-15×

The Vera Rubin NVL72 Partition—Starlight—provides the same Vera Rubin GPUs as Sandstone but with 72 GPUs interconnected into a scale-up NVLINK fabric for a massive, shared memory footprint with an order of magnitude higher interconnect bandwidth (see Figure 3).

Together, this architecture ensures that Mission can flexibly support the full range of mission-critical computational requirements.

Data and Storage

The Mission storage system is an all-flash, high-bandwidth, high input/output operations per second (IOPs) configuration with large capacity. Mission will provide two storage environments using this hardware. One storage environment will be based on the Lustre parallel filesystem, which has served as a high-performance scratch space for MODSIM on every ATS platform at LANL.

In addition to Lustre, an AI-optimized storage environment based on the Hammerspace pNFS (parallel Network File System) will be available for high-IOP workloads commonly found in AI. This dual-stack approach allows for sharing of resources (capacity and performance) for MODSIM and AI with flexible balancing of storage capacity and performance which was not possible in prior systems.

The Goal: Improving Time-To-Insight

The Tri-labs have daunting MODSIM and AI needs, and each LANL platform has moved the boundary on what was previously possible. Trinity (ATS-1) made major advances in supporting large-scale 3D simulations with a massive memory footprint and massive scale. This allowed the first large-scale high-complexity 3D simulations at realistic resolutions. Crossroads (ATS-3) significantly improved memory bandwidth (more than 5x compared to Trinity) significantly reducing the time-to-solution for these high-complexity simulations.

Mission moves the boundary of what is possible even further than either of these prior platforms.

Mission will drive down time-to-solution for large-scale simulations significantly further than Crossroads while providing much higher memory capacity per core/node. Mission will enable dense compute–bound simulations and AI workloads to achieve multiple factors of speedup. AI training workloads will be accelerated through the most advanced CPU/GPU capability available, large scale, and a robust set of best of breed, industry standard, software tools. Scale-up workloads (frontier AI models and strong-scaled simulations) will have a first of its kind scale-up network on Mission. Taken together, Mission is more than the sum of its partitions—it provides a first of a kind, capability class system that will significantly reduce time-to-insight for our most challenging problems.

LANL staff recently tested NVIDIA's Vera CPU platform in preparation for deployment in upcoming platforms, and the early performance results are impressive. Vera's memory bandwidth and core performance is on track to deliver to our expectations. Similar testing will be conducted on the Rubin GPU followed by the deployment of a number of early testbeds prior to full-system deployment.

Early application readiness activities have already begun through Mission's center of excellence. These activities will lay the early groundwork to ensure that both our codes and Mission are ready to meet the needs of the NNSA on day one.

Mission Project Timeline

- November 2026: Mission Starlight (NVL72) delivery
- January-March 2027: Sortie (Early Access System 1, Early Access System 2, Test and Development System) delivery
- August 2027: Hermes (Application Regression System) delivery
- September 2027: Mission (Ranger and Sandstone) delivery
- **March 2028:** Mission transition to classified computing production for Advanced Technology Computing Campaign