



U.S. ARMY COMBAT CAPABILITIES DEVELOPMENT COMMAND
ARMY RESEARCH LABORATORY
 DOD SUPERCOMPUTING RESOURCE CENTER

Summer 2024
 Newsletter

bits

Good To Know!

Need Information for ARL
 DSRC User Services?
dsrchelp@arl.army.mil

The SCOUT LSF
 Scheduler Guide is
 available at:
[www.arl.hpc.mil/docs/
 scoutLsfGuide.html](http://www.arl.hpc.mil/docs/scoutLsfGuide.html)

Users are reminded to
 update their current
 and future HPC
 requirements in the
 HPCMP pIE database.

Questions? Contact:
outreach@arl.hpc.mil



DOD HPC
 MODERNIZATION PROGRAM

“Fran” TI-24 System Announced

The U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) DoD Supercomputing Resource Center (DSRC) is pleased to announce its newest addition, an HPE Cray EX System. The HPCMP announced the TI-24 award on 21 June 2024. It will be named Fran for Frances “Fran” Bilas, the final honoree among the original six ENIAC programmers. It is a single, large unclassified HPC emphasizing standard compute nodes. It will be the first system in our new A212 Facility and is scheduled for delivery in December 2024.

Fran will consist of 173,184 total compute cores (AMD EPYC 9654 “Genoa” 2.4 GHz 96 core processors) contained in 792 standard compute nodes, 32 AI/ML compute nodes with 4 NVIDIA H200 NVL Tensor Core GPUs per Node (128 GPUs total), 30 Persistent Service Framework nodes, 24 large memory compute nodes, and 24 visualization compute nodes. It will have the Cray Slingshot-11 Dragonfly Interconnect, ClustorStor E1000 storage with Lustre, Slurm job scheduler, and Singularity.

Fran comes with several “firsts”. The first system to go into our A212 Facility. The first TI system for which the vendor is not responsible for providing system administration; the ARL system administration team will provide those services. Fran is the first



ARL TI system with a 7-year maintenance period. Fran has the first NVIDIA H200 Tensor Core GPUs at ARL, specifically designed for generative AI, large language models, and HPC workloads. Fran is (almost) the first TI system to be delivered with Persistent Services Framework nodes supporting virtual machines and container encapsulated workloads. (They were added to Centennial and to SCOUT, and Jean was contracted to have them but taken out of service.)

With the arrival of Fran at the end of Calendar 2024, the ARL DSRC will have platforms on the floor celebrating all six original ENIAC programmers: Frances Elizabeth “Betty” (Snyder) Holberton, Betty Jean “Jean” (Jennings) Bartik, Kathleen “Kay” (McNulty) Mauchly Antonelli, Ruth (Lichterman) Teitelbaum, Marlyn (Wescoff) Meltzer, and Frances V. “Fran” (Bilas) Spence.

User Group Meeting UGM 2024

The annual HPCMP User Group Meeting was hosted by the Department of the Navy at the Naval Information Warfare Center (NIWC) Atlantic.

See the website for Program update and highlights along with abstracts from all the presentations insights. Stay tuned for information about next years Users Group Meeting.



Naval Information Warfare Center
 (NIWC) Atlantic
 North Charleston, South Carolina,
 September 10 – 12, 2024.

<https://ugm.hpc.mil/index.html>

2024 HIP Intern Symposium



On 25 July 2024, the ARL DSRC hosted the annual HIP Intern Symposium, sponsored by the DoD HPC Modernization Program (HPCMP). The audience of visitors, ARL scientists and engineers, and remote viewers heard 16 project presentations. Live and virtual participants were welcomed by the acting HPCMP Deputy Director, Dr. Leslie Leonard. The acting Director, Dr. Kevin Newmeyer, participated virtually.

Of the 16 presentations, 13 directly supported Army Modernization Priorities: Future Vertical Lift, Next Generation Combat Vehicle, and Long Range Precision Fires. Of the remaining three, two supported the Army Evaluation Center, creating natural language processing tool to help evaluators reduce time, cost, and human error in evaluating programs of record, and one supported the Naval Research Laboratory. We will follow up with the mentors and students to assist them in getting their HIP intern final reports published as journal articles. It was exciting and encouraging to hear these future computational scientists describe their introduction to high performance computing.

Vince Perry and Mariya Occorso, alumni of this Program, related their experience:

Mariya Occorso: “My internship was indescribable. Prior to interning, I had a cloudy idea of what it meant to be a computer scientist and no idea what I wanted to do for a career. Then COVID-19 happened, and I was gifted this internship where I could do something. It was a bit unconventional at first, but as the summer went on, I was able to envision a future where I was able to research and develop AI and machine learning technologies. I felt inspired. The atmosphere and work environment felt like home. I was able to not only work on interesting projects, but I was also able to make my first publication. Once my first internship ended, I couldn’t imagine myself not going back the next summer. The second internship I did through HIP taught me so much more about data science and confirmed that I wanted to come back for full time work. Now that I am a full-time employee, I have been gifted with opportunities that I couldn’t have dreamed of when I was submitting my application for my first internship. For summer 2024, I have been able to mentor the next generation of interns and pass on some of the insights from my own internships. Outside of working with interns this summer, I was also invited to join a panel at the Pentagon to discuss the future workforce and share my experience.”

Vince Perry: “My role as an intern in the summer of 2015 was my first ever internship experience with no concept of how the knowledge I gained in school would translate to a real-world setting. As an intern, I researched scientific visualization technologies and developed a 3D simulation of Army assets that displayed network link up- and down-times among air, ground, and sea assets. The integration of applying my programming knowledge from school, learning about scientific visualization, and working on an Army problem captivated me to pursue a career in the DoD research and development space. While a bit overwhelming at first, there was a synergy between learning new things and experimenting with, testing, and developing what I had learned that I knew was an environment I wanted to work in. Transitioning to a full-time employee has allowed me to continue learning new things every single day on the job, from new computing architectures, hardware, and software paradigms to new advancements in algorithms and modeling and simulation. It is a career that I am very passionate about and consider myself very fortunate, solving Army problems utilizing the Army’s most advanced High Performance Computing systems. Because of my experience, I have become a devoted mentor and co-mentor of 10 HIP interns since 2018 hoping to inspire undergraduates with an internship experience similar to my own, equipping the workforce of the future with the skills and tools to utilize HPC to solve the DoD’s most difficult problems.”

UPDATE: - TI-22

TI-22 passed acceptance testing and government software was installed in August. The system will be in Pioneer mode until the system is stable.

Announced in the fall of 2022 as HPE systems. The unclassified system, Ruth, for Ruth Lichterman, is an EX-4000, has 127,488 AMD EPYC Genoa compute cores and 64 AMD MI-250 GPUs, supported by 335 terabytes of memory and 17 petabytes of storage, including 2 petabytes of NVMe-based solid state storage.

The classified system, Marlyn, for Marlyn Wescoff, is an HPE EX-2500 system with 57,600 AMD EPYC Genoa compute cores supported by 135 terabytes of memory and 13 petabytes of storage, including 2 petabytes of NVMe-based solid state storage.

If your daily or weekly residence in B120 includes walks past Room 106 (where Excalibur and Centennial used to be), you’ve witnessed shrinking stacks of boxes piled on wooden pallets and a new computer rise from the floor.



A2I2

The work continues...

For the past two and a half years you have read a series of installments about the Army Artificial Intelligence Innovation Institute (A2I2) Facility, from groundbreaking to raising of the shell to installation of massive generators and finishing the interior. You've seen beautiful photos and time lapse videos documenting the progress.

You've read a collection of specifications about the Facility: it's Module 1 of four Modules; has 11,000 square feet of floor space; two 2500KW diesel generators for backup power with transformers, automatic transfer switches, electrical distribution system including battery and rotary UPS back-up power, distribution panels, PDUs, and sub-panels; cooling towers, chillers, heat exchanger; 2,100 tons of cooling. So What?

Many of us are familiar with Building 120, our home for the past 10+ years and the former BRL Supersonic Wind Tunnel building. (The supersonic wind tunnel was completed in 1948.) With the decommissioning of classified computers in Buildings 328

(Continued on page 4 "A2I2")



*(top): Welding pipe for cooling.
(middle): Progress in the mechanical room.
(bottom): New computer room where "FRAN" will be installed.*



(“A2I2” from page 3 continued)

and 394, Building 120 became the single remaining building for housing HPCs. It has 20,590 square feet of processing space, unclassified + classified.

Module 1 of A2I2 has 11,000 square feet, plus an additional 23,500 square feet of lobby and hallway, 1st floor mechanical/electrical rooms, and mezzanine mechanical/electrical rooms. Consider other nerd facts. Building 120 has 10 mW power capacity (and cannot be expanded) and 1,800 tons of chilled water cooling capacity. Module 1 of A2I2, when fully built out (not initially), will have 15 mW power capacity (5 mW initially) and 2800 tons of chilled water cooling capacity. Multiply by 4 to see how Building 120 is dwarfed, and it has no expansion capability.

Maybe you’ve taken part in a tour of the A2I2 Facility lately. Some of us did recently and that motivated this update. You can read specs all day long, you can see artist concepts or engineering drawings, you can hear descriptions of what the facility will be used for, but until you walk through that building and observe the scale and magnitude of it, you will never appreciate what the team has accomplished, what the current space will accommodate, or what the A2I2 Facility portends for future computing.

Everything is laid out as HPC infrastructure. The main floor has two very large rooms for future Technology Insertions (TIs). Each room can accommodate two TIs; that takes us rather far into the future. One of the two rooms is constructed to house classified computers. Out back are the backup generators, chiller, heat exchanger, and cooling towers. Across the driveway in the front is the electrical substation for all APG. If demand requires more than it currently supports, land immediately north of it is available for expansion.

TI-24 will be the first system in the A2I2 Facility. It will be a single, large unclassified HPC, possibly the largest in the HPCMP.

What is an A2I2 Ecosystem?

It is an ecosystem of networks, data repositories, tools laboratories, computing resources, & expertise with goals to:

- Advance AI capabilities for autonomous maneuver for multi domain operations.
- Host challenging problems addressing Army multi domain operations.
- Facilitate collaborative research with academia, industry, and government.
- Establish repository of proven AI algorithms and labeled data.

An A2I2 Ecosystem will address complex issues, such as:

- How intelligent systems reason about, interact with, and manipulate the environment to achieve complex military relevant actions.
- How diverse, embodied agents collectively sense, infer, reason, plan, and execute in collaboration with Soldiers and in the face of a peer adversary.

HPC Training:

To see latest schedule and enroll, visit the HPCMP PET Training webpage:

training.hpc.mil

Webinars:

Upcoming events.
(Webcast Enrollment Required)

• HPC New Account Orientation

September 18
2:00 - 3:30PM ET

• Writing Efficient Python for HPC

September 26
2:00 - 3:30PM ET

If you have suggestions, problems, or need training information /assistance

Email:

HPCtraining@hpc.mil



Contact us for more information:

www.arl.hpc.mil

email:

outreach@arl.hpc.mil



Directors Message: Matt Goss

Where did the summer go? Work was progressing nicely on the A2I2 Facility, we were planning the annual HIP Intern Symposium, we were awaiting announcement of TI-24, then suddenly it was all over. The A2I2 Facility will be ready to occupy in December, in time to house Fran, our large unclassified TI-24 system announced in July, and the HIP Intern Symposium was a huge success. I know part of the issue is that I currently have one foot in the Pentagon supporting Dr. Forrest Shull, Principal Director for Advanced Computing and Software, and the other foot in the ARL DSRC. Even with my long legs that’s quite a stretch. Time and change seem to be moving faster than ever.

That is nowhere more apparent than in artificial intelligence (AI) and machine learning (ML), especially in Generative AI capabilities, such as Large Language Models (LLMs). It seemed as if ChatGPT exploded onto the scene from nowhere, we began delving into the strengths and risks of LLMs, and then they were everywhere and performing remarkable tasks in text, code, and image generation just for the asking. Now they are pervasive and Deputy Secretary of Defense, Kathleen H. Hicks, has established the Generative AI and LLM Task Force, Task Force Lima, to focus the Department’s exploration and responsible fielding of generative AI capabilities.

The ARL DSRC was already leaning forward, working with PET to host an LLM on SCOUT and transition it into a user service. We are closely examining LLM hardware requirements and how those should influence future TIs. I want to hear about your experience with Generative AI, issues you’ve experienced, and what you think would make the ARL DSRC more responsive to user needs.