



Programming Novel AI Accelerators for Scientific Computing

Murali Emani

Argonne Leadership Computing Facility

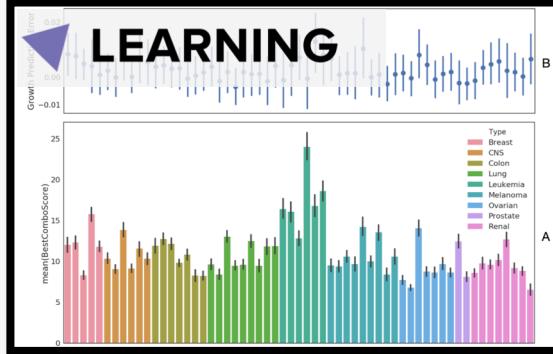
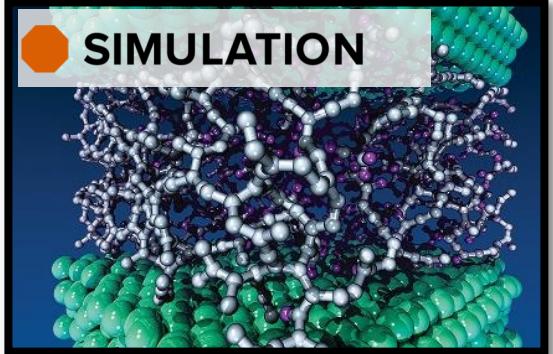
memani@anl.gov

Argonne Leadership Computing Facility



The Argonne Leadership Computing Facility provides world-class computing resources to the scientific community.

- Users pursue scientific challenges
- In-house experts to help maximize results
- Resources fully dedicated to open science



ALCF offers different pipelines based on your computational readiness. Apply to the allocation program that fits your needs.



Architecture supports three types of computing

- § Large-scale Simulation (PDEs, traditional HPC)
- § Data Intensive Applications (scalable science pipelines)
- § Deep Learning and Emerging Science AI (training and inferencing)

ALCF AI Testbed

<https://www.alcf.anl.gov/alcf-ai-testbed>



Cerebras CS-2



SambaNova DataScale SN30



Graphcore
Bow Pod64



Habana
Gaudi1



GroqRack

- Infrastructure of next-generation machines with AI hardware accelerators
- Provide a platform to evaluate usability and performance of AI4S applications
- Understand how to integrate AI systems with supercomputers to accelerate science

Recent ALCF AI Testbed Updates

ALCF AI Testbed Systems are in production and available for allocations to the research community

<https://www.alcf.anl.gov/science/directors-discretionary-allocation-program>



SambaNova SN30

SambaNova upgraded to latest 2nd generation SN30 accelerators and scaled to 8 nodes with 64 AI accelerators



Graphcore BowPod64

Graphcore upgraded to latest Bow generation accelerators and scaled to a Pod-64 configuration with 64 accelerators



Cerebras CS-2

Cerebras CS-2 upgraded to an appliance mode to include Memory-X and Swarm-X technologies to enable larger models and scaled to two CS-2 engines



GroqRack

Groq system has been recently upgraded to a GroqRack with nine nodes, each consisting of eight GroqChip Tensor streaming processors accelerators

Argonne Leadership Computing Facility

[ALCF Resources](#)[Science](#)[Community and Partnerships](#)[About](#)[Support Center](#)

ALCF User Guides

[Home](#)[Account and Project Management](#)[Data Management](#)[Services](#)[Running Jobs with PBS at the ALCF](#)[Polaris](#)[Theta](#)[ThetaGPU](#)[AI Testbed](#)[Getting Started](#)[Cerebras](#)[Graphcore](#)[Groq](#)[SambaNova](#)[Data Management](#)[Cooley](#)[Aurora/Sunspot](#)[Facility Policies](#)

ALCF AI Testbed



The [ALCF AI Testbed](#) houses some of the most advanced AI accelerators for scientific research.

The goal of the testbed is to enable explorations into next-generation machine learning applications and workloads, enabling the ALCF and its user community to help define the role of AI accelerators in scientific computing and how to best integrate such technologies with supercomputing resources.

Table of contents

[How to Get Access](#)[Getting Started](#)[How to Contribute to Documentation](#)

Tutorial Agenda

<https://github.com/argonne-lcf/Alaccelerators-SC23-tutorial>

| Time (MST) | Topic |
|----------------------|---|
| 08.30 - 8.35 | <u>Introduction to AI Testbed at ALCF (ANL)</u> |
| 08.35 - 8.50 | <u>Claire Zhang (Cerebras Systems)</u> |
| 08.50 - 9.05 | <u>Petro Junior Milan (SambaNova Systems)</u> |
| 09.05 - 9.20 | <u>Alex Tsyplikhin (Graphcore)</u> |
| 09.20 - 9.35 | <u>Sanjiv Shanmugavelu (Groq)</u> |
| 09.35 - 9.50 | <u>Leon Tran (Intel Habana)</u> |
| 10.00 - 10.30 | Break |
| 10.30 - 12.00 | Hands session on the AI Testbed (ANL) |

How to use ALCF AI Testbed

<https://github.com/argonne-lcf/AIaccelerators-SC23-tutorial>

Request Account on AI Testbeds At ALCF

- Request an [ALCF Computer User Account](#) if you do not currently have one
- If you have an ALCF Account that is currently inactive, submit an [account reactivation](#) request*.
- If you have an active ALCF account, click [Join Project](#) to submit a membership request. Specify the following in your request: Project Name: `aitestbed_tutorial`

Contact accounts@alcf.anl.gov M-F 9am to 5pm CT. Reach out to us on slack channel `#help-accounts` on [ALCF-AIAccelerator-tutorials](#) Slack.

| SC23 Tutorial allocation will stay active till end of November 2023.



Getting Started on ALCF AI Testbed:

Apply for a Director's Discretionary (DD) Allocation Award

Director's Discretionary (DD) awards support various project objectives from scaling code to preparing for future computing competition to production scientific computing in support of strategic partnerships.

Cerebras CS-2, SambaNova SN30, Graphcore Bow Pod64, and GroqRack are available for allocations

Allocation Request Form

<https://www.alcf.anl.gov/science/directors-discretionary-allocation-program>

AI Testbed User Guide

<https://www.alcf.anl.gov/alcf-ai-testbed>

Thank You

- This research was funded in part and used resources of the Argonne Leadership Computing Facility (ALCF), a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.
- Venkatram Vishwanath, Michael Papka, William Arnold, Varuni Sastry, Sid Raskar, Zhen Xie, Rajeev Thakur, Bruce Wilson, Anthony Avarca, Arvind Ramanathan, Alex Brace, Zhengchun Liu, Hyunseung (Harry) Yoo, Corey Adams, Ryan Aydelott, Kyle Felker, Craig Stacey, Tom Brettin, Rick Stevens, and many others have contributed to this material.

Please reach out for further details

Venkat Vishwanath, Venkat@anl.gov

Murali Emani, memani@anl.gov

