

Tutorial on

# **Programming Novel AI Accelerators for Scientific Computing**

SC25  
November 16, 2025

Murali Emani (ANL)  
[memani@anl.gov](mailto:memani@anl.gov)

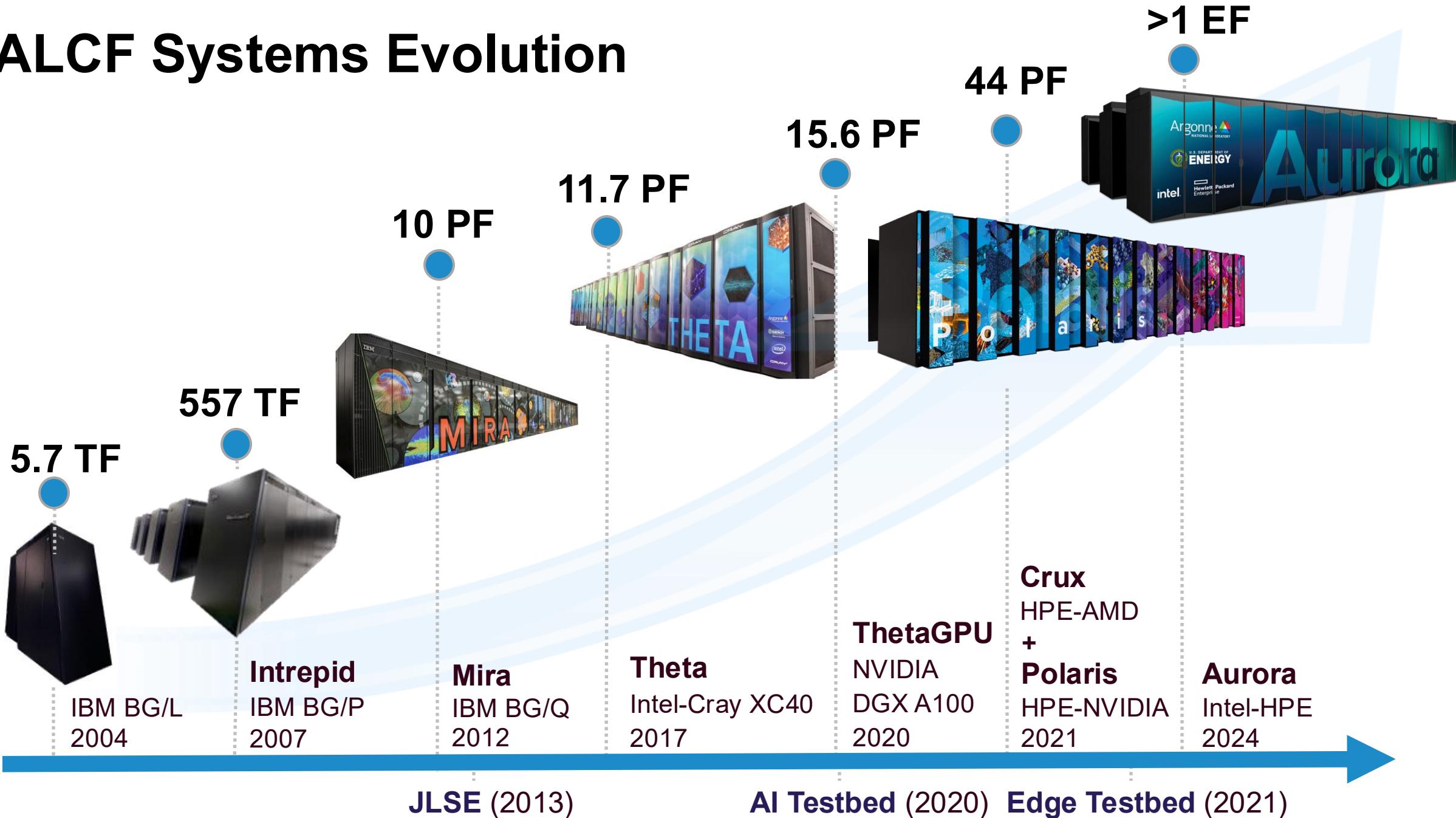
Varuni K Sastry (ANL), Leighton Wilson (Cerebras), Petro Jr. Milan (SambaNova),  
Milind Pandit (Intel), Sanjiv Shanmughavelu (Groq)

# Tutorial Agenda

<https://github.com/argonne-lcf/AIaccelerators-SC25-tutorial/>

| Agenda             |  |
|--------------------|--|
| Time (EST)         | Topic/Speaker  |
| 1:30 PM - 1:40 PM: | Welcome and Overview of the ALCF AI Testbed (Murali) |
| 1:40 PM - 2:20 PM: | SambaNova (Petro) <a href="#">Slides</a>             |
| 2:20 PM - 3:00 PM: | Groq (Sanjif)  |
| 3:00 PM - 3:30 PM: | Coffee Break   |
| 3:30 PM - 4:10 PM: | Cerebras (Leighton)                                  |
| 4:10 PM - 4:50 PM: | Intel (Milind) <a href="#">Slides</a>                |
| 4:50 PM - 5:00 PM: | Q&A and Conclusion (Varuni)                          |

# ALCF Systems Evolution



# ALCF AI Testbed

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



Cerebras (CS-3)



SambaNova SN30/SN40L



Groq



Graphcore



Tenstorrent  
Coming soon!!

- Infrastructure of next-generation machines with hardware accelerators customized for artificial intelligence (AI) applications.
- Provide a platform to evaluate usability and performance of machine learning based HPC applications running on these accelerators.
- The goal is to better understand how to integrate AI accelerators with ALCF's existing and upcoming supercomputers to accelerate science insights

# ALCF AI Testbed

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

## Training

- Cerebras
- Sambanova SN30



SN-30 8 nodes of 8 RDUs



Cerebras CS-3 – 4 WSE

## Inference

- SN40L – Metis
- Groq
- Cerebras
- Tenstorrent



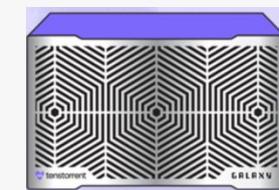
2 nodes of 16 SN40L RDUs



9 Groq nodes,  
8 GroqChip/node (TSPs)



Cerebras CS-3 – 4 WSE



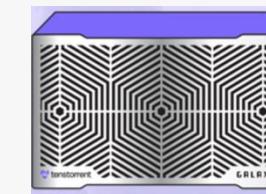
32 Wormhole GU

## HPC

- Cerebras
- Tenstorrent



Cerebras CSL



32 Wormhole GU

# Getting Started on ALCF AI Testbed

Available for Allocations

- Cerebras CS-3,
- SambaNova Datascale SN30,
- GroqRack
- Graphcore Bow Pod64
- Sambanova Inference – Metis SN40L (Available for all ALCF users via Inference Service Endpoints)

[AI Testbed User Guide](#)

## Director's Discretionary (DD) awards

- Scaling code
- Preparing for future computing competition
- Scientific computing in support of strategic partnerships.

[Allocation Request Form](#)

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

## NAIRR Pilot

Aims to connect U.S. researchers and educators to computational, data, and training resources needed to advance AI research and research that employs AI.

<https://nairrpilot.org/>

<https://docs.alcf.anl.gov/ai-testbed/getting-started/>

**ALCF User Guides**

Getting Started

Contribute to User Guides

**User Support**

Submit a Ticket

Get Help &amp; Connect &gt;

**Machines**

Aurora

**AI Testbed**

Cerebras



Graphcore



Groq



SambaNova



Data Management

Crux



Polaris



Sophia

**Running Jobs with PBS**

Example Job Scripts

Machine Reservations

**Data Management**

File System &amp; Storage



Transfer &amp; Sharing

**Services**

Inference Endpoints

JupyterHub

Continuous Integration

**Table of contents**[How to Get Access](#)[Getting Started](#)[How to Contribute to Documentation](#)[Does this doc need an update?](#)[Open an Issue on GitHub](#)

# 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.

The AI accelerators complement the ALCF's current and next-generation supercomputers to provide a state-of-the-art computing environment that supports pioneering research at the intersection of AI, big data, and high performance computing (HPC).

The platforms are equipped with architectural features that support AI and data-centric workloads, making them well suited for research tasks involving the growing deluge of scientific data produced by powerful tools, such as supercomputers, light sources, telescopes, particle accelerators, and sensors. In addition, the testbed will allow researchers to explore novel workflows that combine AI methods with simulation and experimental science to accelerate the pace of discovery.

# Useful Links

## ALCF AI Testbed

- Overview: <https://www.alcf.anl.gov/alcf-ai-testbed>
- Guide: <https://docs.alcf.anl.gov/ai-testbed/getting-started/>
- Training:
  - <https://www.alcf.anl.gov/ai-testbed-training-workshops>
  - <https://github.com/argonne-lcf/Alaccelerators-SC25-tutorial/>
- Allocation Request: [Allocation Request Form](#)
- Support: [support@alcf.anl.gov](mailto:support@alcf.anl.gov)