

# Novel Architecture Simulation, Emulation, and Evaluation with the CRNCH Rogues Gallery

Jeffrey Young, PhD · Rogues Gallery Director, Sr. Research Scientist ·

Thanks to Aaron Jezghani, PACE Scheduler Architect, Sr. Research Scientist

Co-PIs: Tom Conte, Ada Gavrilovska, Jennifer Hasler, Rich Vuduc

SP: Alex Daglis

Technical Support: Sterling Peet, Will Powell

Funded by NSF Rogues Gallery (CNS-2016701)



# The Rogues Gallery



Rogues Gallery is an NSF funded post-Moore testbed for CISE researchers and the community.

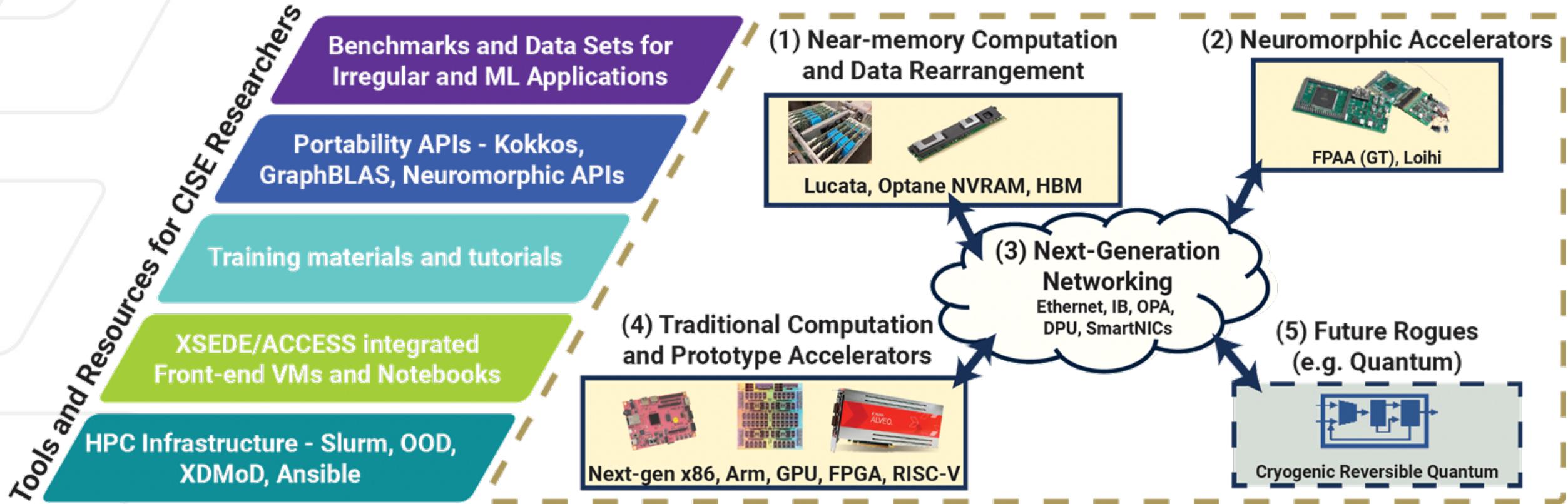
- The Gallery contains 40 servers and 40+ development boards – Intel CLX, SKL, ICX; AMD/NVIDIA GPUs; Arm; RISC-V; Xilinx
- ***Extreme heterogeneity*** with GPU, FPGAs, FPAs, Optane Memory, InfiniBand, OmniPath, and Ethernet networking



Introduction



# The Rogues Gallery



Introduction

# Rogues Gallery Team

## *Co-PIs*



Tom Conte



Jennifer Hasler



Ada Gavrilovska



Rich Vuduc

## *Senior Personnel*



Alex Daglis



Sterling Peet



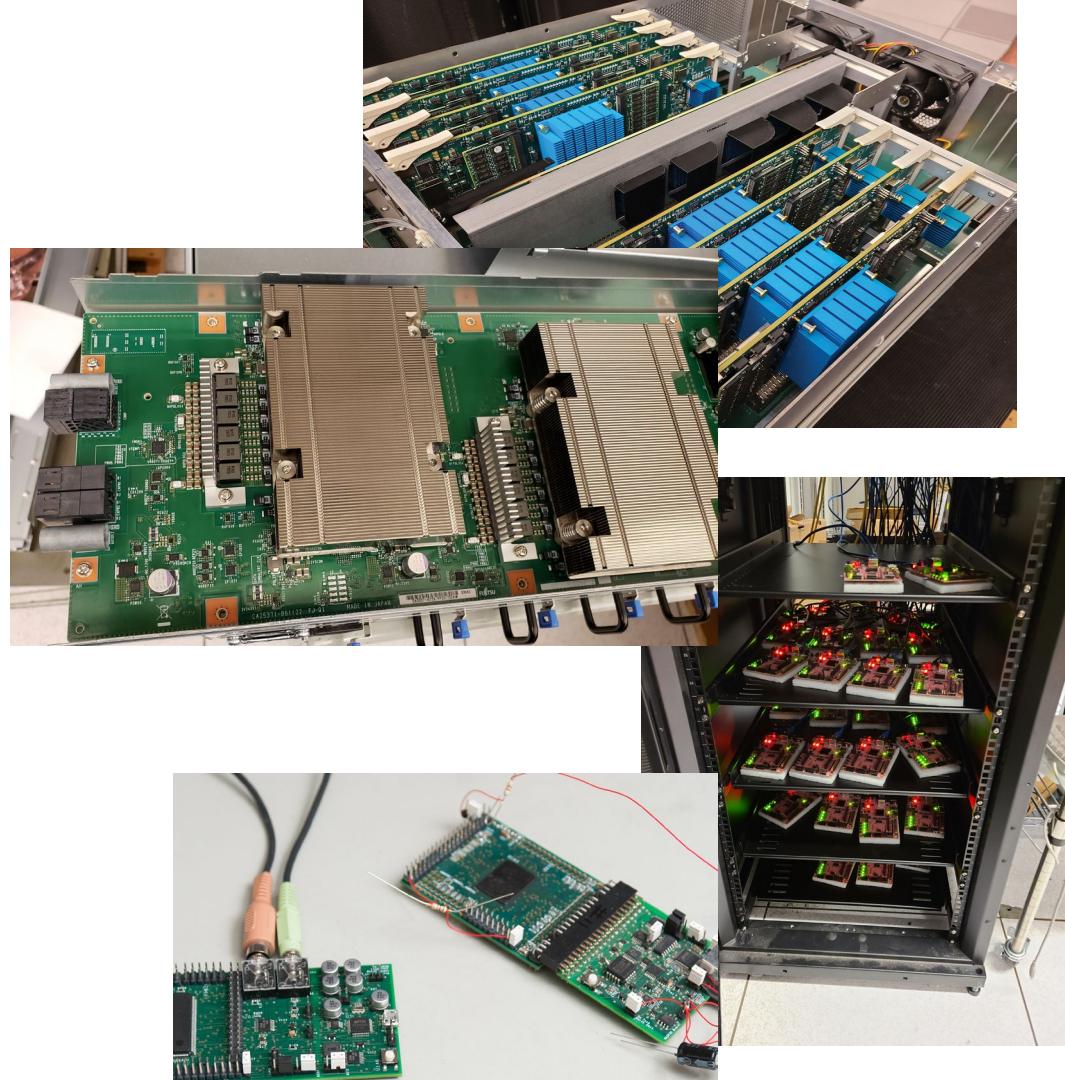
Will Powell



Aaron Jezghani (PACE)

# Rogues Gallery Highlights

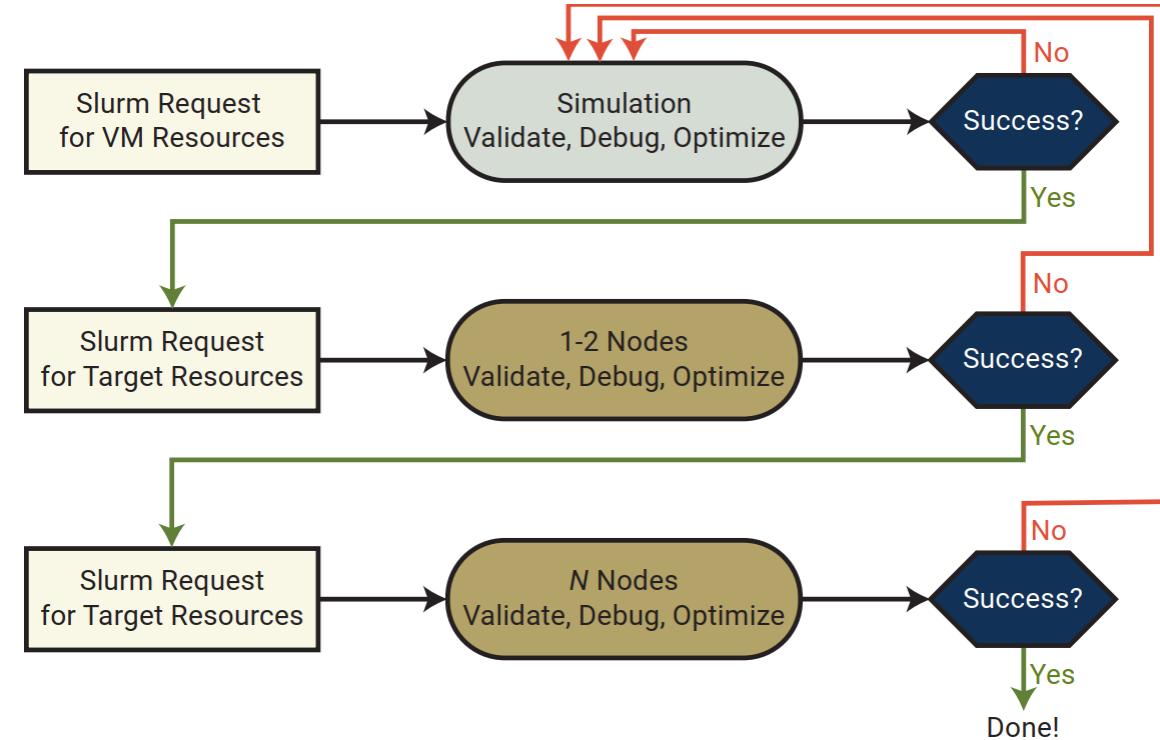
- First deployment of A64FX with Open OnDemand, later adopted by RIKEN for Fugaku
- Largest public instance of Lucata Pathfinder - #211 on 2021 Graph500 and #46 on GreenGraph500 rankings
- Support for over 180 researchers and 60-70 external users across multiple areas
- Support for 80-130 students each year with Pynq boards and similar infrastructure



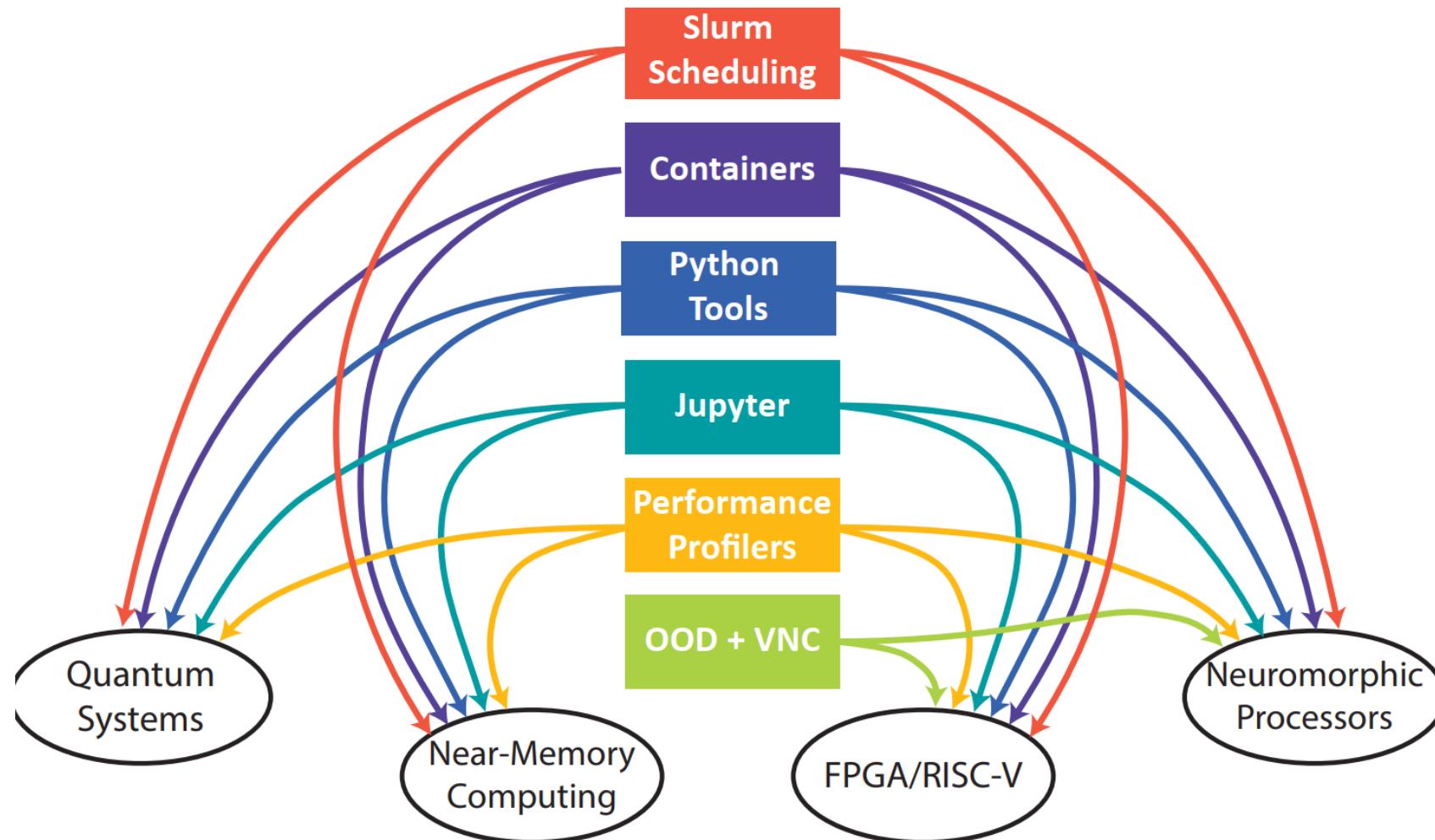
Introduction

# Rogues Gallery Workflows

Novel architecture evaluation follows a classic simulation, emulation, and evaluation pathway, even for atypical architectures!



# Enabling Tools and Technologies



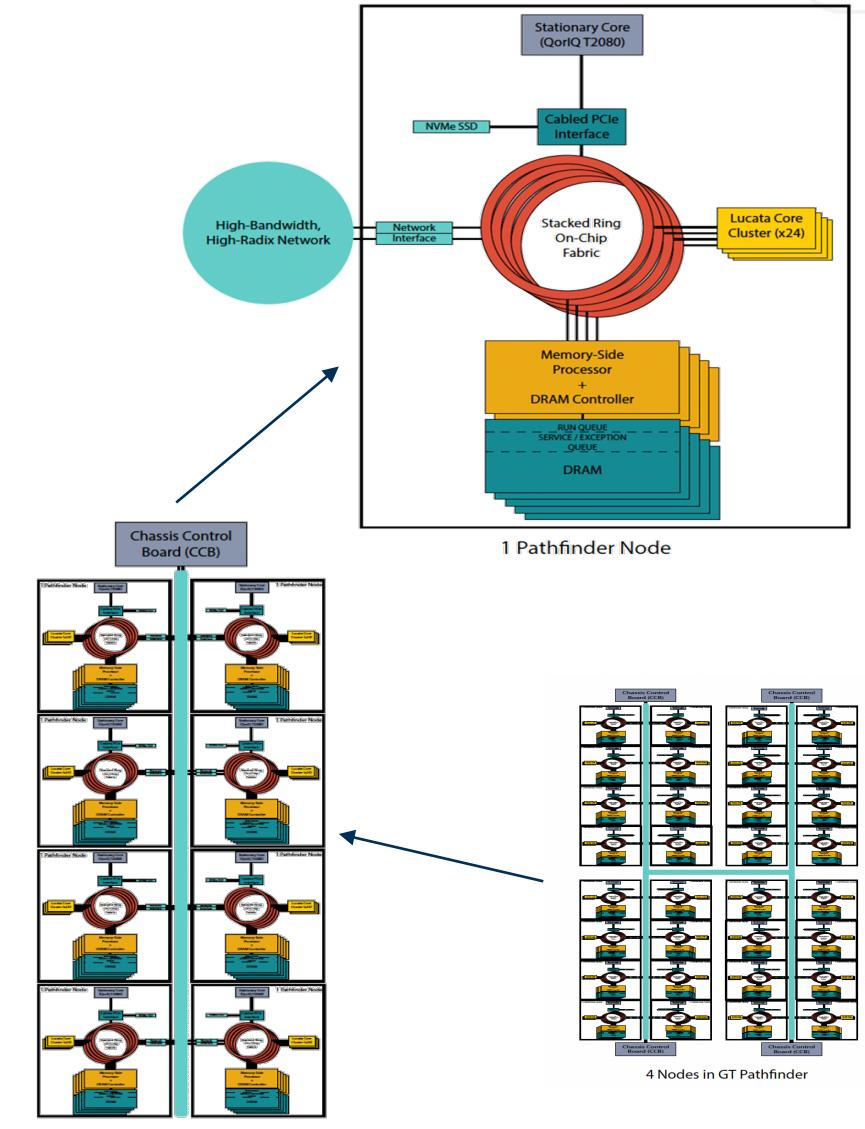
# Example: The Lucata Pathfinder System



- Next iteration of Emu Chick prototype
- Design motivated by scaling issues for problems hindered by data access patterns
  - E.g. graphs, sparse matrix algebra
- Rather than migrating large swaths of data across network, hardware threads can move to the memory
  - Demonstrated scaling efficiency with benchmarks such as BFS, hpcg vs. x86

# Reconfigurable Hardware

- Each chassis contains:
  - 1 Chassis Control Board (PowerPC E6500) for node/network management
  - 8 compute nodes, which include:
    - 1 stationary compute core (PowerPC E6500)
    - 24 Lucata compute cores (Proprietary IP)
    - Stacked ring-on-chip fabric to interface to network/DRAM/compute
    - High-bandwidth, high-radix network for inter-node/chassis communication
- Systems can be configured as single or multi-chassis
  - Defines network and available Lucata compute cores for running applications
  - Scale hardware according to the problem size
- GT Pathfinder consists of 4 chassis



Introduction

The Pathfinder

User Workflows

Education and Training

Summary

# High-Level Libraries/APIs for Code

- Lucata-provided software environment
  - Custom adaptation of Yocto Linux kernel and base utilities
  - Configuration tools to dynamically modify chassis, network
  - Performance monitoring libraries to check efficiency, energy
  - Custom API for thread management and task execution
  - Robust simulation capabilities to validate before deployment
- Open-Cilk for parallel programming
  - Open-source platform for task-parallel code development
  - Leverages lightweight backend ABIs for target dispatch
  - Minimal effort to port code from x86 to Pathfinder



Introduction

The Pathfinder

User Workflows

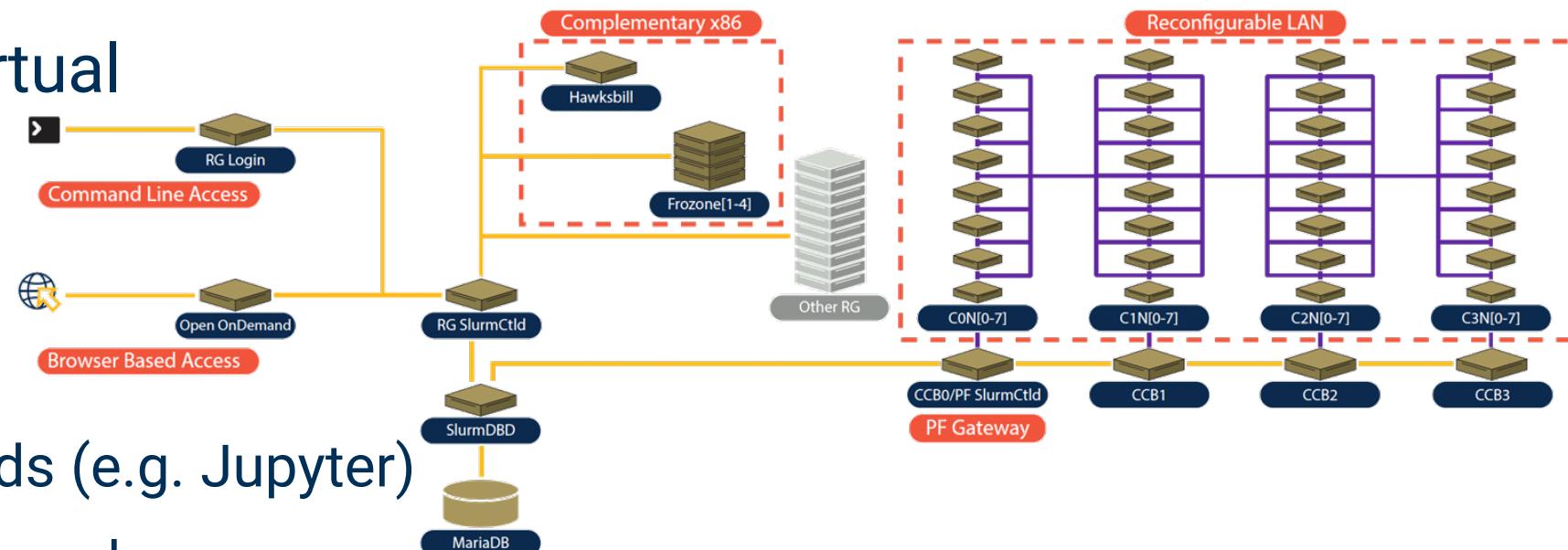
Education and Training

Summary

**GT** Georgia Tech

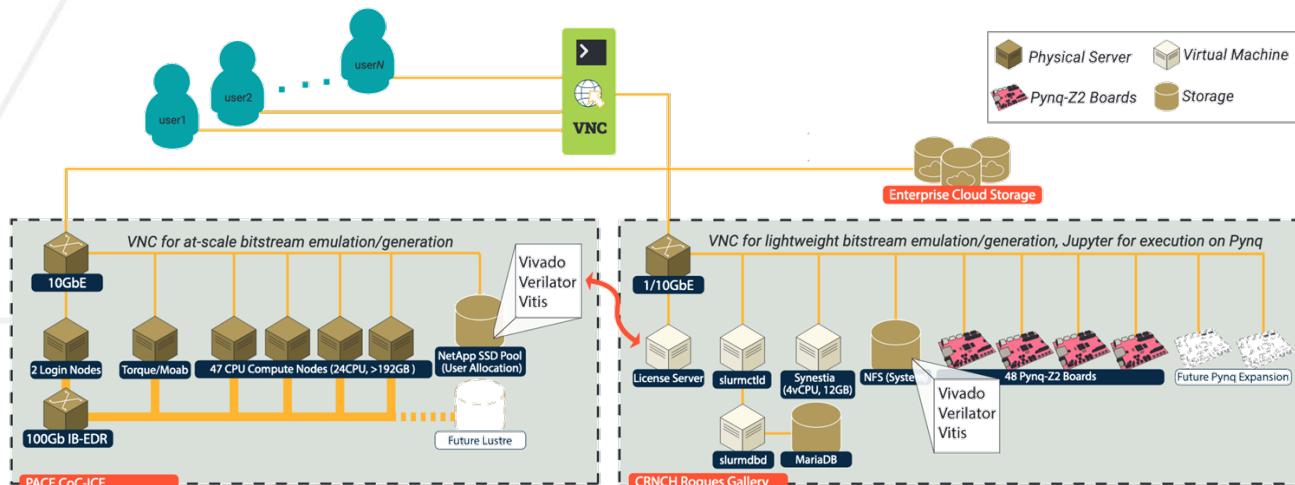
# Accessing the Pathfinder and Related Systems

- Pathfinder access managed via Slurm
  - Federated cluster alongside other Rogues and infrastructure nodes
  - Single entry point, single job database
- x86 physical and virtual machines for development and simulation
  - Host visual frontends (e.g. Jupyter)
  - Comparative benchmarks



# Slurm and Open OnDemand

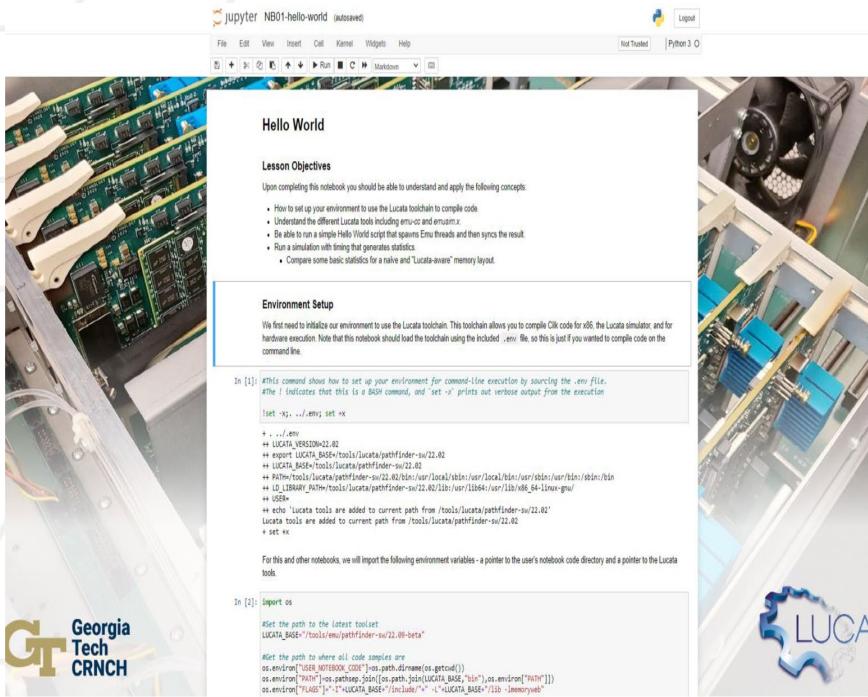
- Slurm scheduling across most resources
  - Interesting challenges with “edge device” integration
  - Running 23.11 with 10 different builds (x86, Arm, NVIDIA, RISC-V, etc.)
- Open OnDemand support to ease entry



Jezghani A, Manalo K, Powell W, Valdez J, Young, J. Onboarding Users to A64FX via Open OnDemand. International Conference on High Performance Computing in Asia-Pacific Region Workshops. 2022; 78-83. <https://doi.org/10.1145/3503470.3503479>

The screenshot shows the Georgia Tech CRNCH OnDemand interface. The top navigation bar includes Georgia Tech, Files, Jobs, Clusters, Interactive Apps, My Interactive Sessions, and Dev. The main area features the CRNCH logo and a "ROGUES GALLERY" section displaying various hexagonal icons representing different hardware components. Below this, a message states: "OnDemand provides an integrated, single access point for all of your HPC resources." A "Message of the Day" section follows, and a "Message of the Day" section below it lists various job submission options. The bottom of the page is powered by OPEN OnDemand.

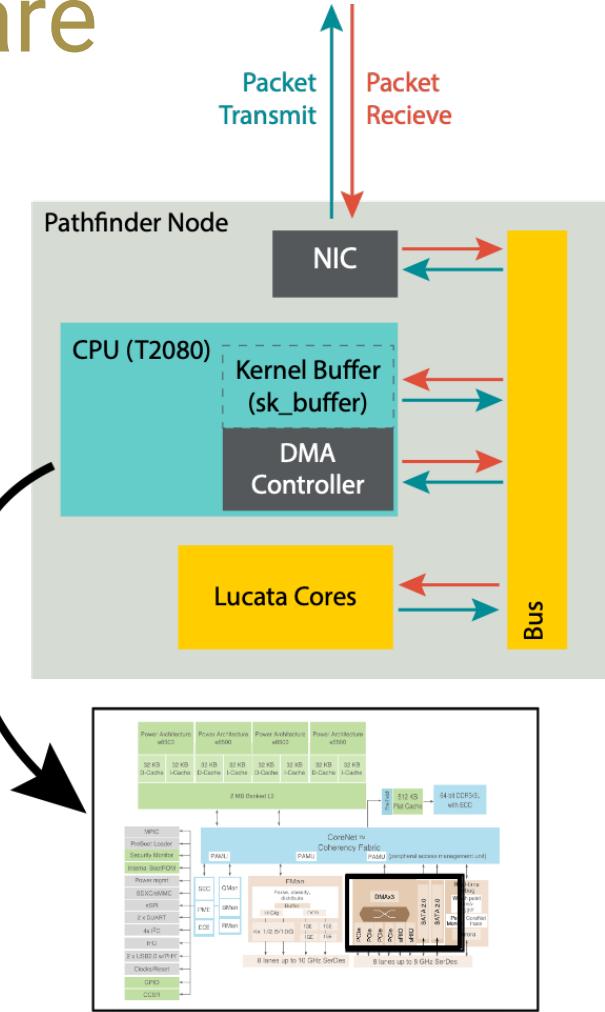
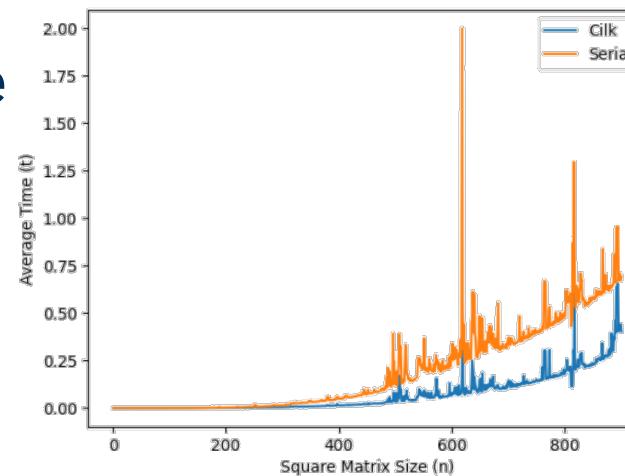
# Community Tutorials to Engage Researchers



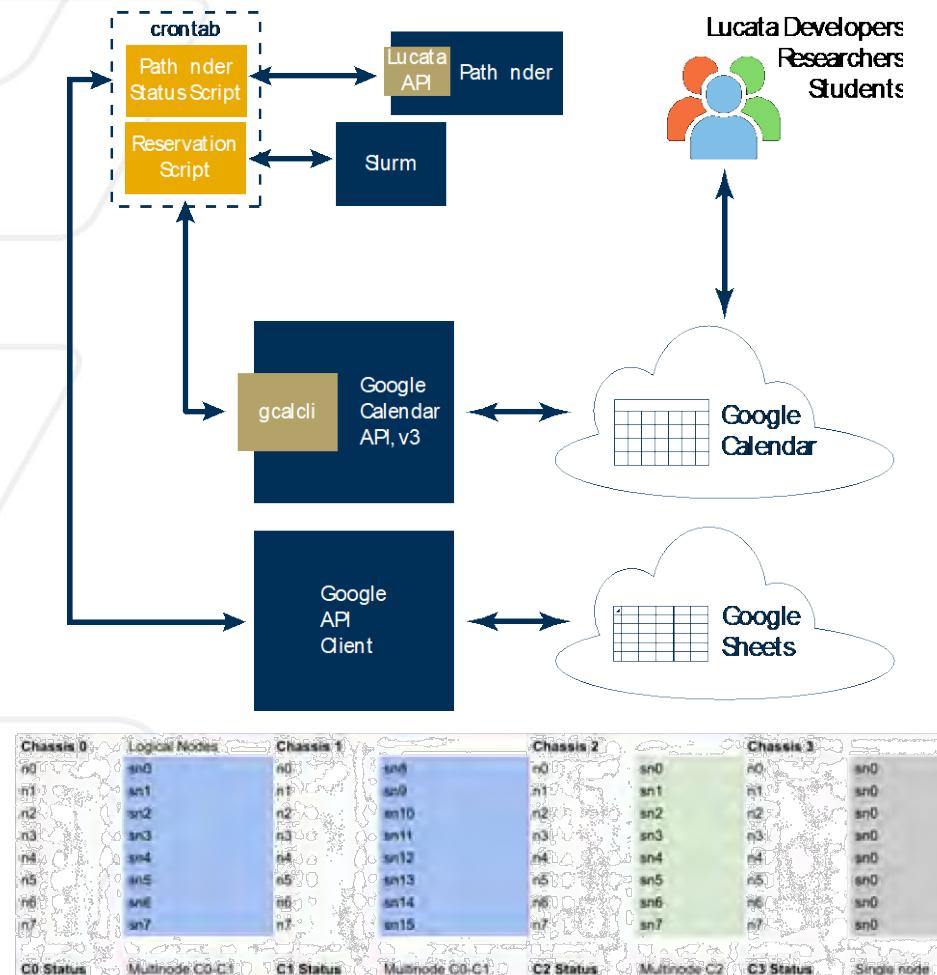
- Tutorials covering Emu Chick/Lucata Pathfinder presented at numerous conferences
  - General RG overview at ASPLOS19 and PEARC19
  - Dedicated tutorial at PEARC21 and HPEC22
  - Iteratively developed content and workflows, such as Jupyter-based approach for HPEC22
- Lessons learned for the platform and tutorials in general:
  - User engagement depends on user interface and prerequisite knowledge
  - Conferences with fewer cross-listed and hybrid options increased attendance and participation
  - Attendees may have HPC experience, but still need more focused examples for new architectures

# Extending CS Curriculum to Novel Hardware

- Future Computing with the Rogues Gallery VIP
  - Near-memory subteam focuses on the Pathfinder architecture for parallel and distributed computing
- Explore benchmarks on x86 and Pathfinder systems
  - hpcg to assess scaling performance for SpMV
  - BFS to demonstrate graph capabilities
  - Pointer Chase (pChase) to compare memory latency
- DMA driver prototype for Lucata compute cores
  - Project option for CS3210: *Design of Operating Systems*



# Pathways for Career Development



Student employees tasked with development of administrative utilities

- Google Calendar to manage system access between students, researchers, and Lucata engineers
- Published system configuration to aid users in running workflows
- Introduces students to various levels of software and utilities in a near-production environment
  - Python and Google APIs to facilitate access to system reservations/configurations
  - Slurm scheduler and Lucata APIs to interface with and configure Pathfinder system
  - Crontab and Ansible to automate and manage deployment

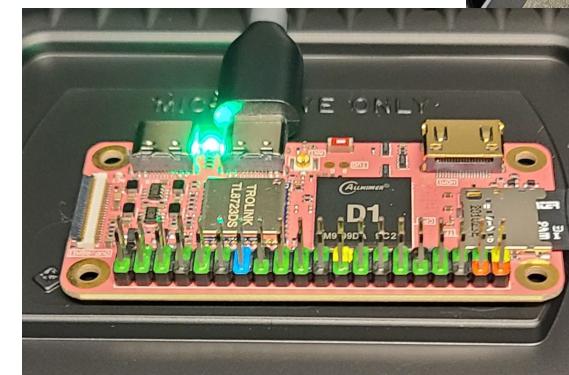
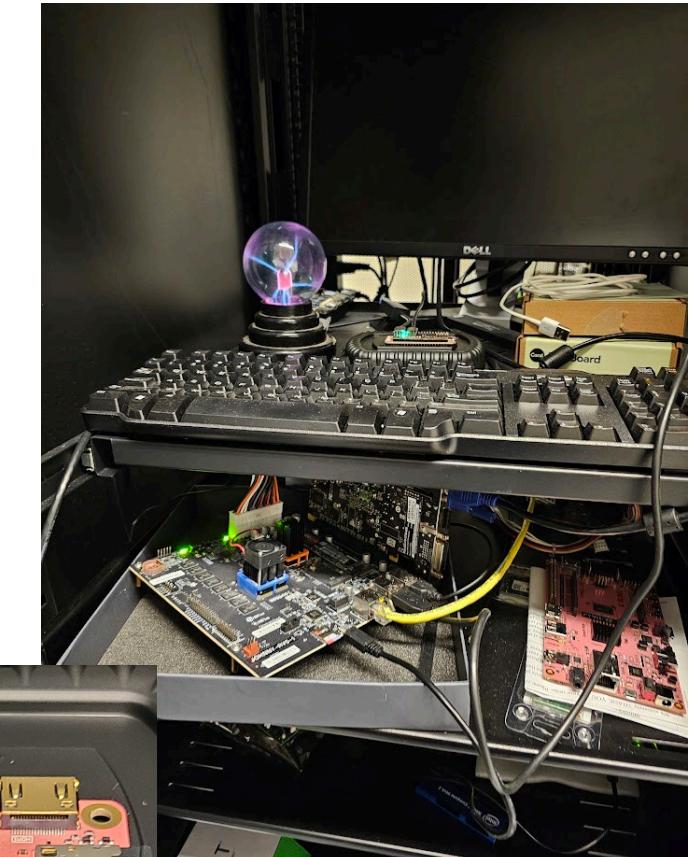
# Challenges Exposed

Getting to this level requires an enormous investment in backend resources, training, and infrastructure

It's hard to imagine exerting the same level of effort for every new platform!

e.g., *Currently, RG uses 10 different builds of Slurm and 3 schedulers*

However, recognizing the patterns for typical novel architecture workflows and mapping them to common infrastructure can provide a solid basis for research

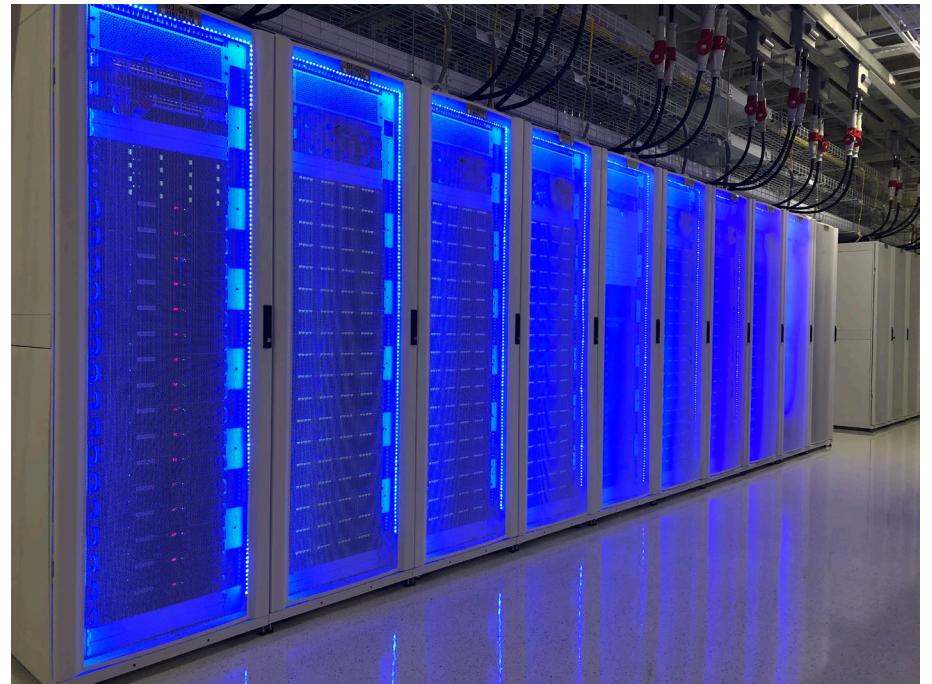


# Follow-on Effects – Evaluating the Testbed

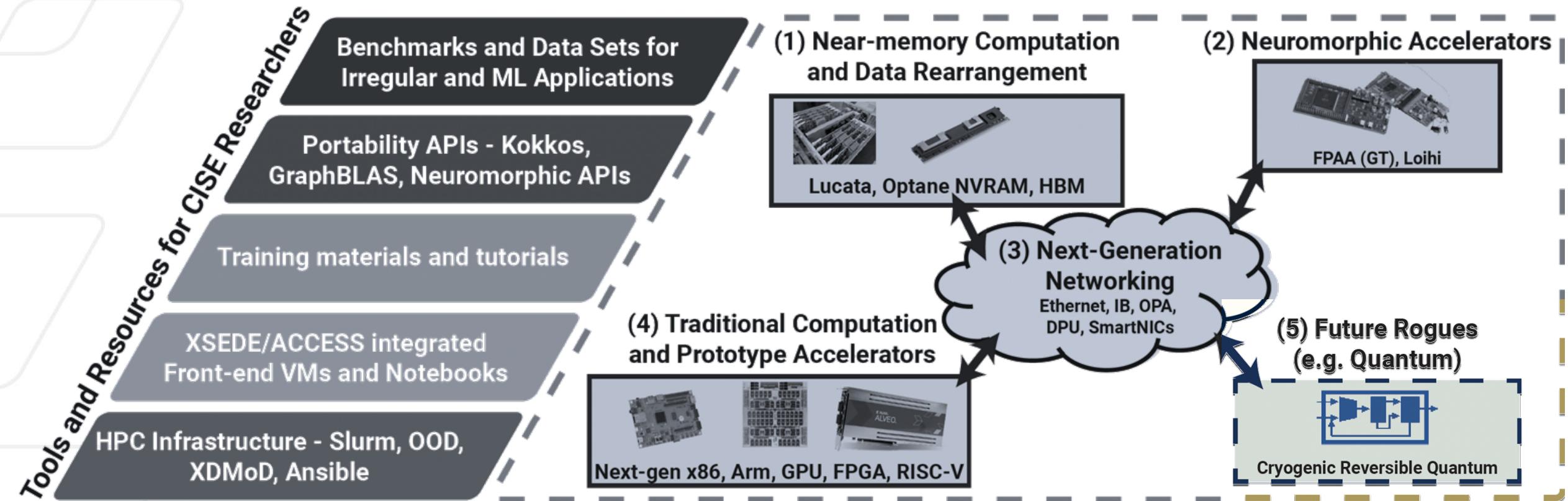
When we think about tomorrow's supercomputers,  
they are **highly heterogeneous**.

Testbeds like the Rogues Gallery allow us not only  
to evaluate far-off technologies but they are  
prototypes for near-term production systems

- Supporting increased heterogeneity in software, hardware, and operating systems
- Slurm features like node helpers and NSS support for launching cloud, container jobs
- Further adoption of user interfaces like Open OnDemand that make access easy so researchers can focus on the challenging systems and co-design tasks.



# New Frontiers



Summary

# New Frontiers

Frontiers

Benchmarks and Data Sets for

(1) Near-memory Computation

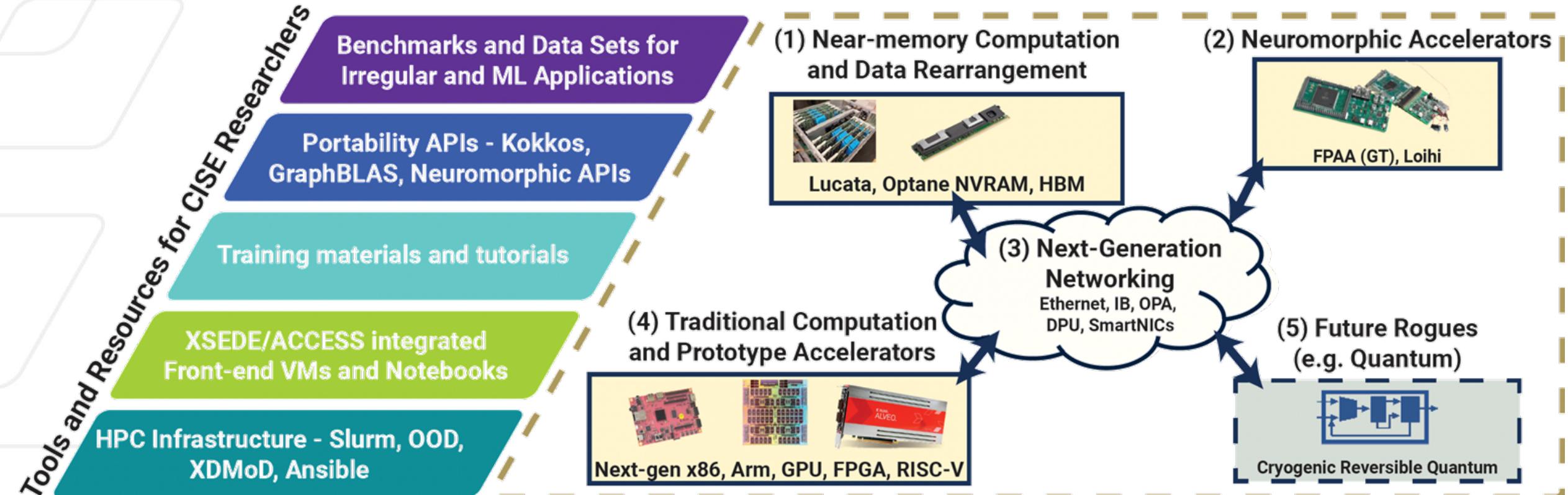
(2) Neuromorphic Accelerators

The near-term Rogues Gallery is likely to include:

- The evolution of smart networking and smart data movement
  - Additional disaggregated memories and accelerators
  - Support of additional libraries (GraphBLAS on DPU!) and benchmarking tools
  - Multi-chip neuromorphic platforms
  - Quantum compilers and larger, more efficient quantum simulations
- 
- More focus on containerized test environments that make better use of emulation resources
  - Better support for AI workflows that use novel architectures in tandem (CPU+GPU+?)
  - And much more!

Summary

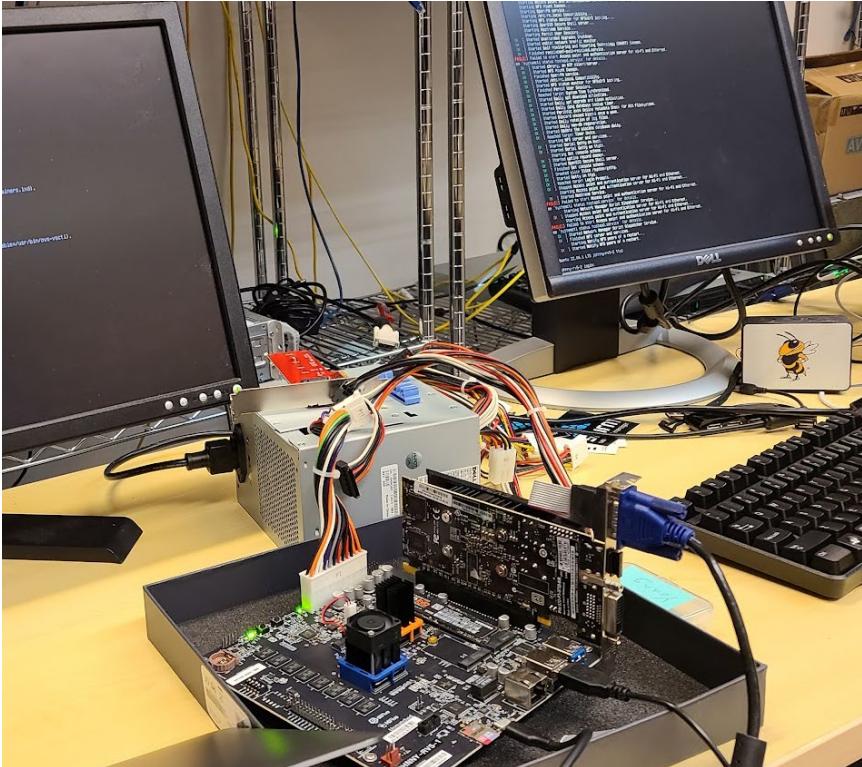
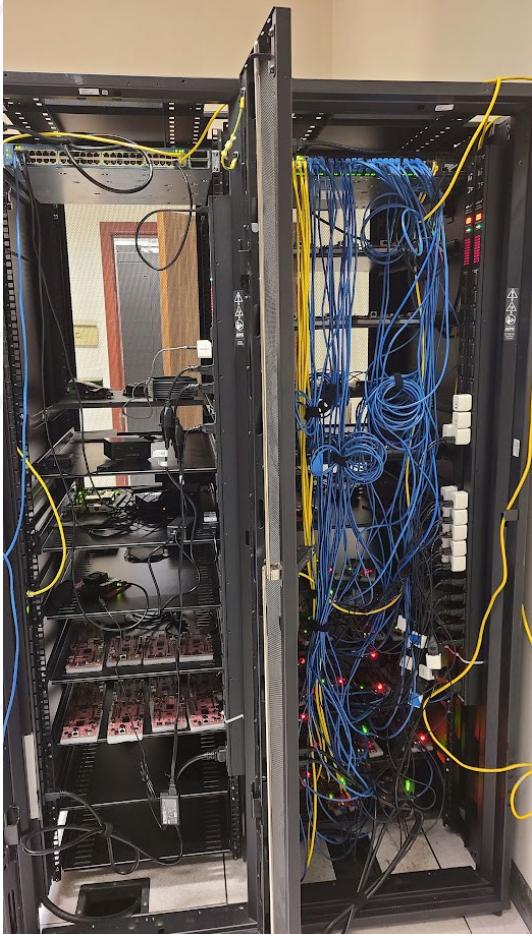
# New Frontiers



Extreme heterogeneity requires an extreme evolution of how we simulate and evaluate systems!

Summary

# Questions?



**Project site:** [crnch-rg.cc.gatech.edu](http://crnch-rg.cc.gatech.edu)  
**GH Presence:** <https://github.com/gt-crnch-rg>