

Introduction to the CRNCH Rogues Gallery

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

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

SP: Alex Daglis

Technical Support: Sterling Peet, Will Powell

Special thanks to Aaron Jezghani, PACE Scheduler Architect, Sr. Research Scientist

Funded by NSF Rogues Gallery (CNS-2016701)



Overview

- Rogues Gallery Overview
 - Achievements and Successes
 - Unique Capabilities
- Promoting Novel Architecture Workflows
 - Example Workflow Built Around the Lucata Pathfinder
- Education and outreach
 - Training efforts
 - Student Engagement
 - Pynq Cluster Use Case
- Integration with PACE
 - Heterogeneous AI workflows
 - Additional Tools and Utilities
- Future Work



Introduction

The Rogues Gallery



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

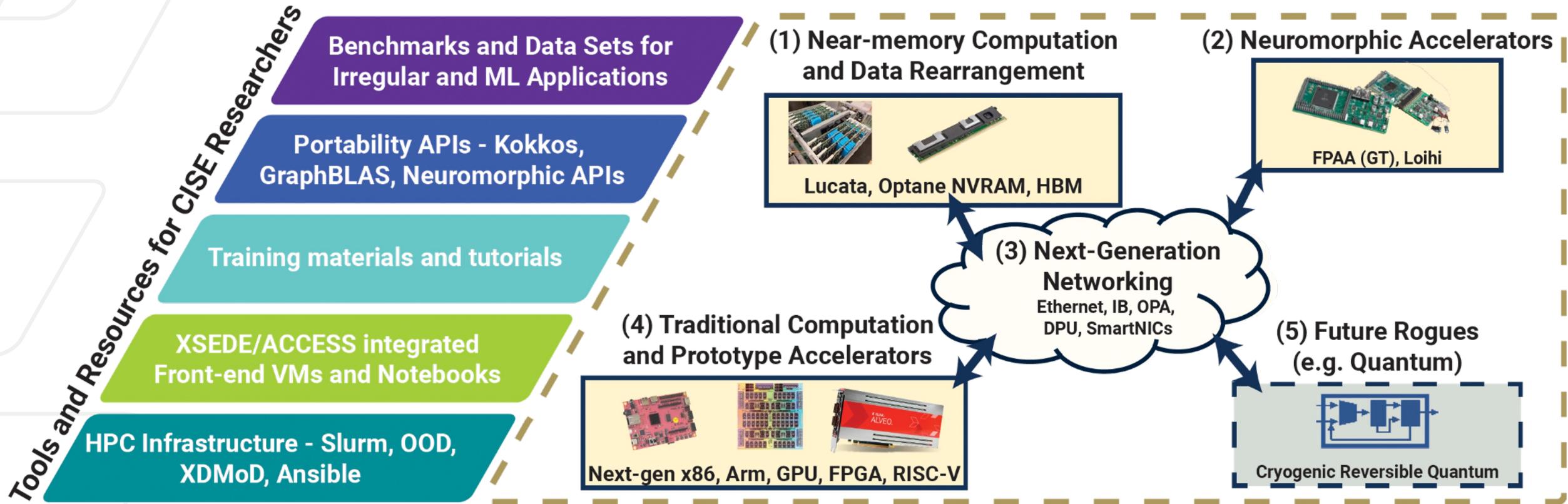
- The Gallery contains 40 servers and 40+ development boards – Intel CLX, SKL, ICX; AMD/NVIDIA GPUs; Arm; RISC-V; Xilinx
- ***Extreme heterogeneity*** with GPU, FPGAs, FPAAs, Optane Memory, InfiniBand, OmniPath, and Ethernet networking



Introduction



The Rogues Gallery



Introduction

Rogues Gallery Team

Co-PIs



Jeff Young



Ada Gavrilovska



Tom Conte



Rich Vuduc

Senior Personnel



Alex Daglis



Sterling Peet



Will Powell

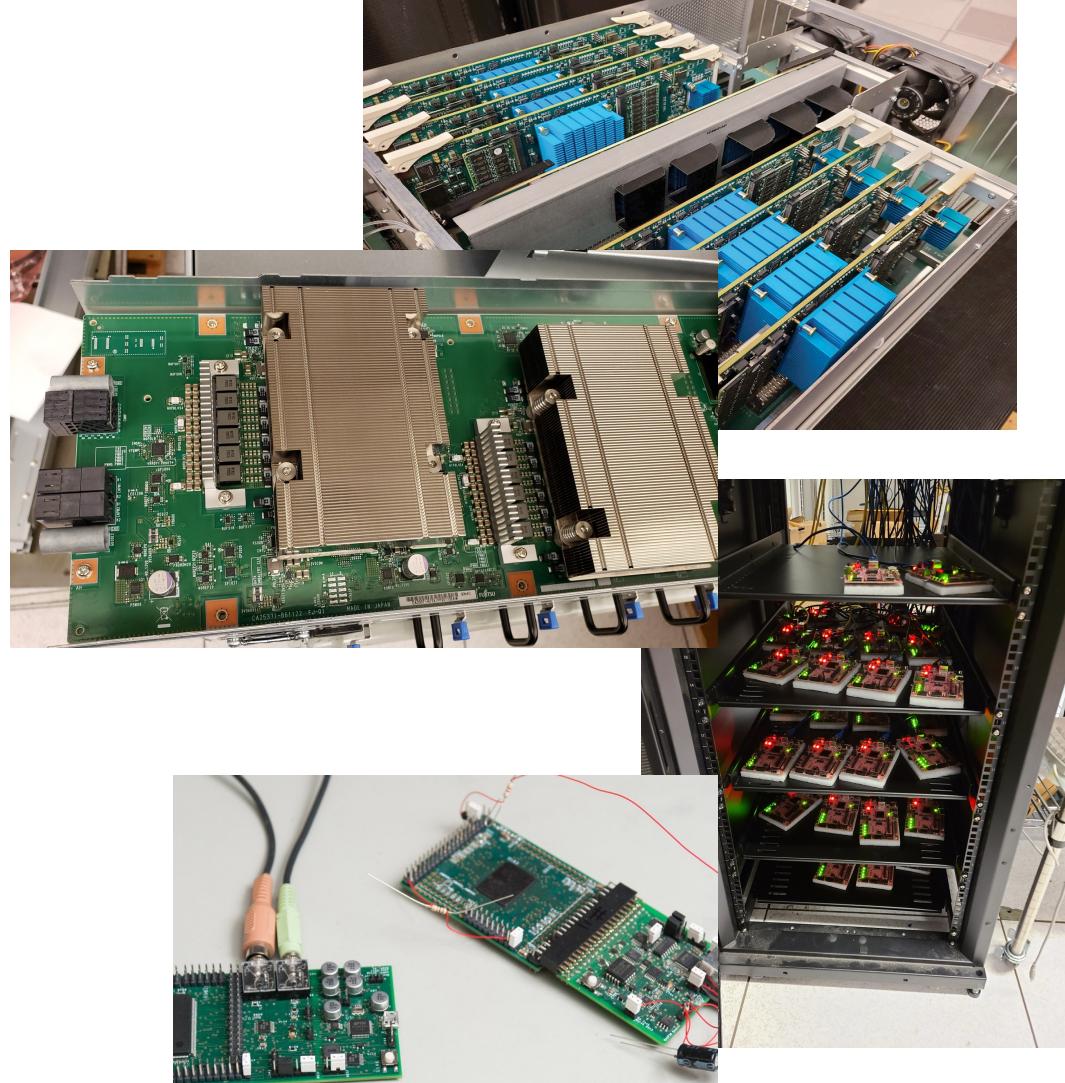


Aaron Jezghani (PACE)

Technical Support

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

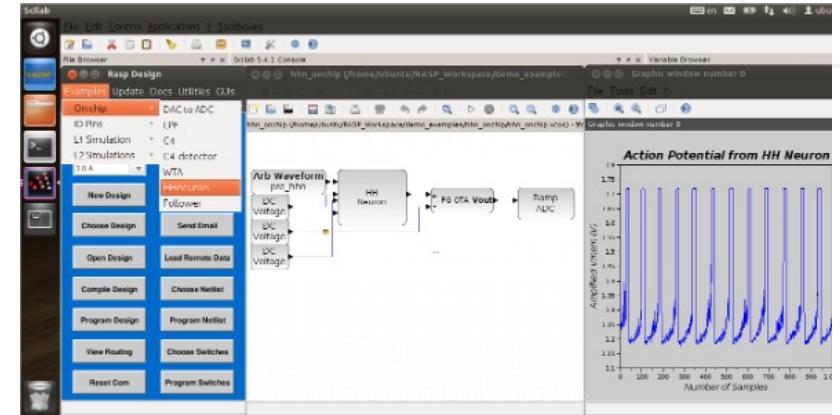
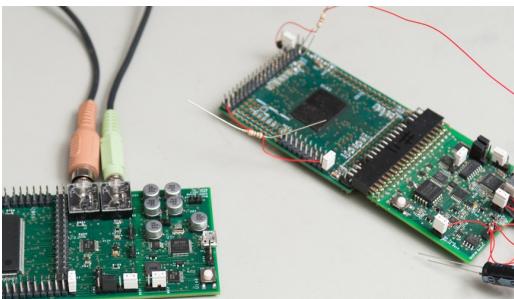


Introduction

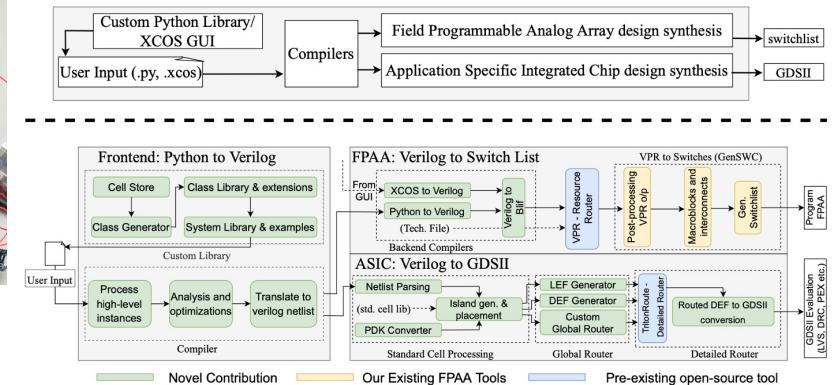


Unique Rogues Gallery Capabilities

- Support for investigating graph analytics accelerators, neuromorphic hardware, and smart networking
- Integration with local educational mission
 - Provides valuable testing of novel architecture
 - Ex: FPAAs Python workflow, PYNQ cluster
- Growing interactions with software development best practices
 - CI/CD, continuous benchmarking



Next generation tools for FPAAs development:
graphical with Xcos interface (above), full open-source analog FPAAs workflow (below)



CI/CD support on the Rogues Gallery

- Part of supporting novel architectures includes supporting **effective software development**
 - CRNCH RG provides Slurm-oriented infrastructure to schedule runners for novel architectures (RISC-V, Arm, Intel GPU, FPGA)
 - Dedicated VM and secure setup process allows for persistent runners



```
# Authentication  
  
✓ Connected to GitHub  
  
# Runner Registration  
  
Enter the name of the runner group to add this runner to: [press Enter for Default]  
  
Enter the name of runner: [press Enter for rg-ci-workflow1] spatter-nvidia-gpu  
  
This runner will have the following labels: 'self-hosted', 'Linux', 'X64'  
Enter any additional labels (ex. label-1,label-2): [press Enter to skip]  
  
✓ Runner successfully added  
✓ Runner connection is good  
  
# Runner settings  
  
Enter name of work folder: [press Enter for _work]  
  
✓ Settings Saved.
```

254 workflow runs			Event ▾	Status ▾	Branch ▾	Actor ▾
Build	Build #215: Scheduled	main	11 hours ago	7m 47s
Build	Build #214: Scheduled	main	yesterday	7m 38s

Learn More: <https://gt-crnch-rg.readthedocs.io/en/main/general/ci-runners.html>

Overview

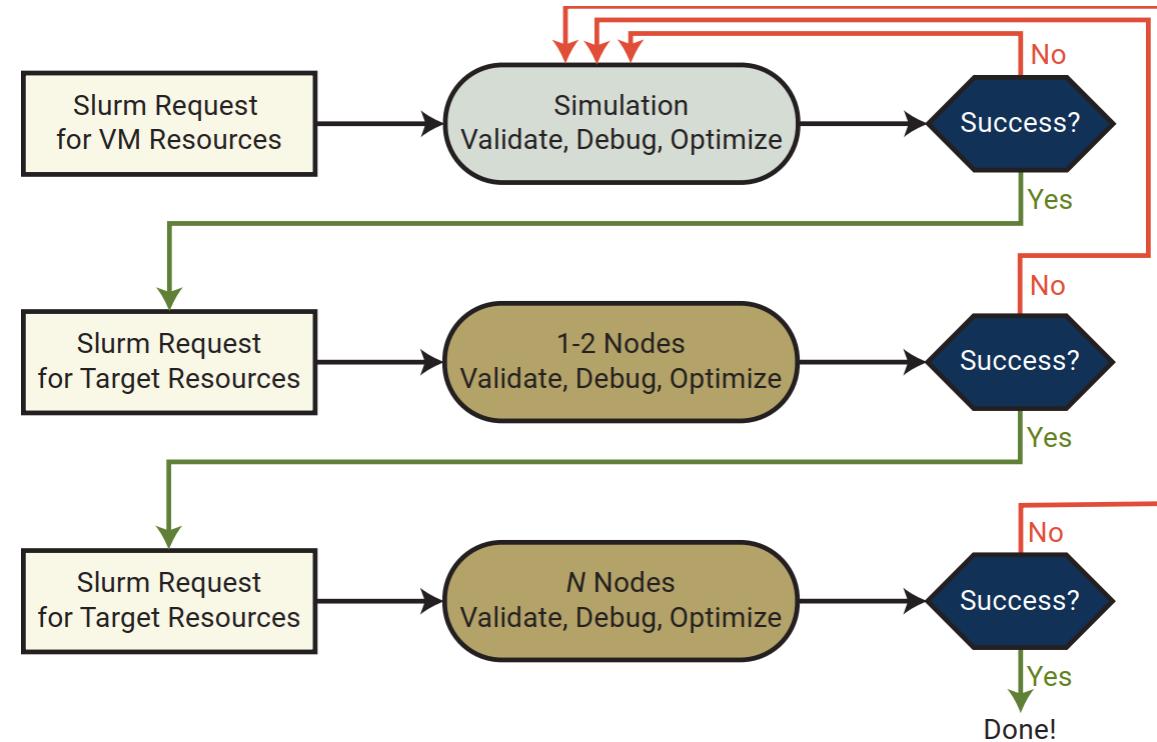
- Rogues Gallery Overview
 - Achievements and Successes
 - Unique Capabilities
- **Promoting Novel Architecture Workflows**
 - Example Workflow Built Around the Lucata Pathfinder
- Education and outreach
 - Training efforts
 - Student Engagement
 - Pynq Cluster Use Case
- Integration with PACE
 - Heterogeneous AI workflows
 - Additional Tools and Utilities
- Lessons Learned and Future Work



RG Workflows

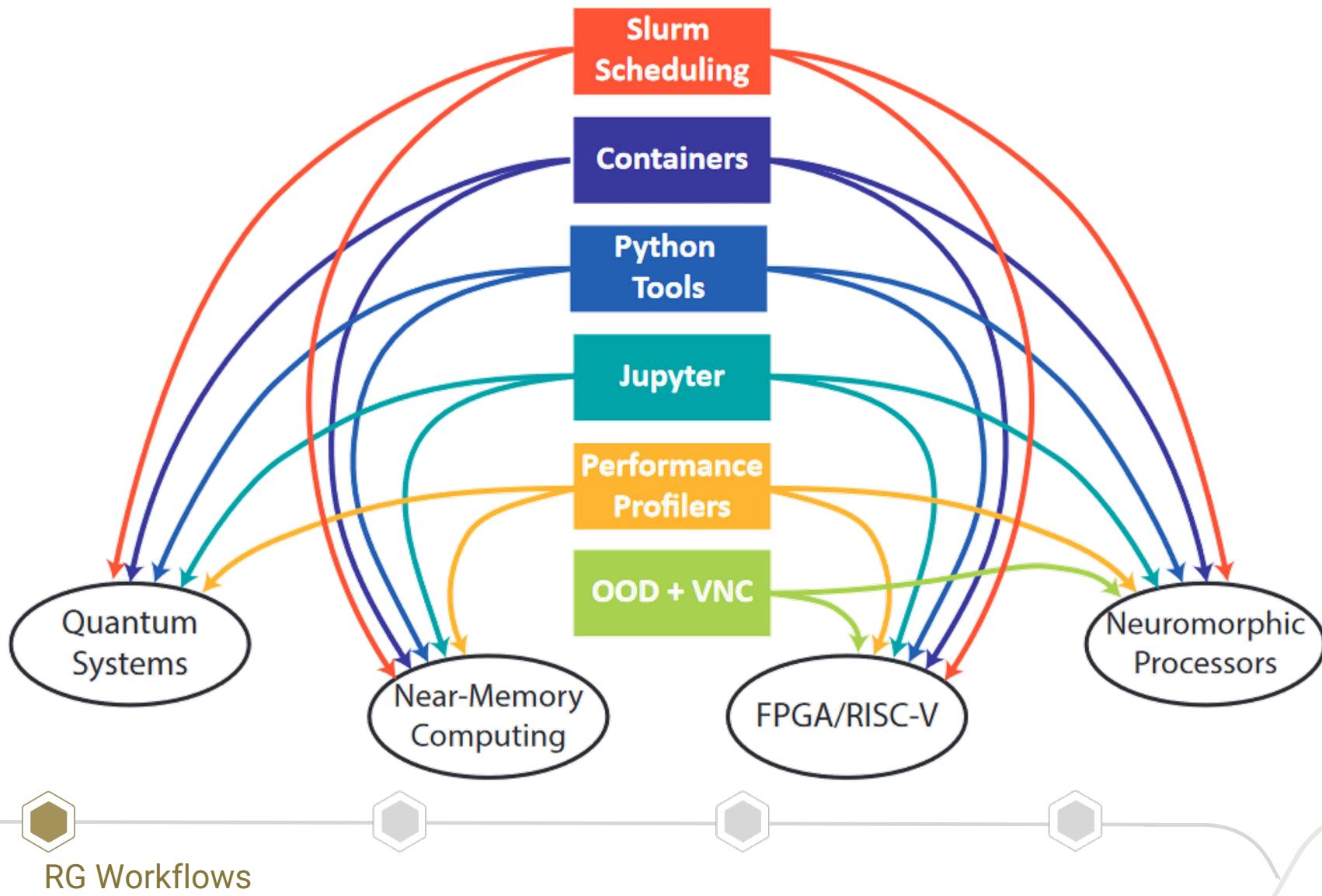
Rogues Gallery Workflows

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



RG Workflows

Enabling Tools and Technologies



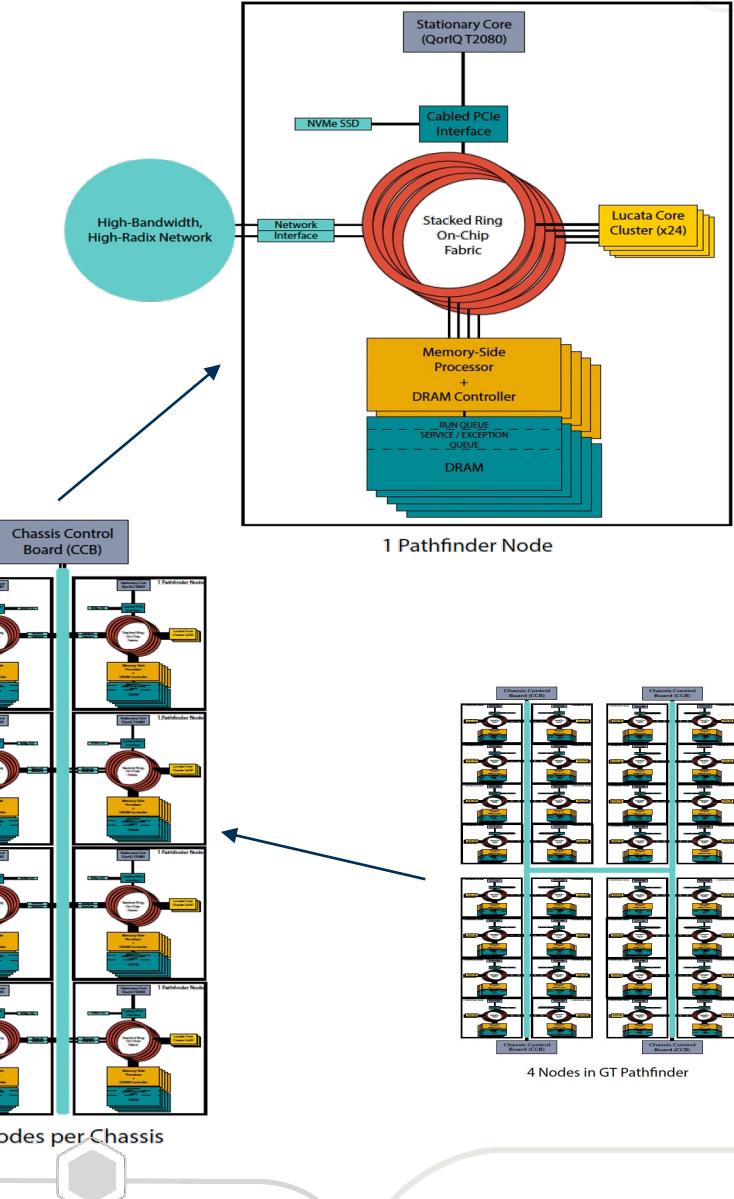
Example: The Lucata Pathfinder System



- 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

Pathfinder - 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



RG Workflows

High-Level Libraries/APIs for Code

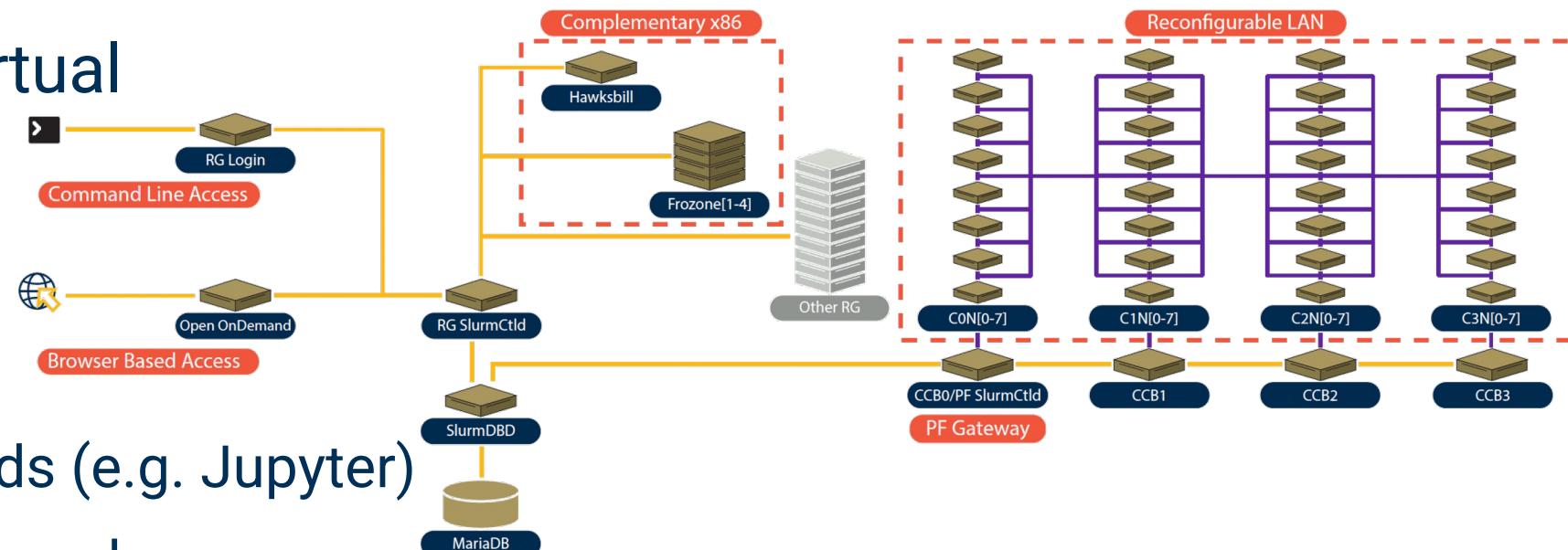
- Lucata-provided software environment
 - Custom adaptation of Yocto Linux kernel and base utilities
 - Configuration tools to dynamically modify chassis, network
 - Performance monitoring libraries to check efficiency, energy
 - Custom API for thread management and task execution
 - Robust simulation 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



RG Workflows

Accessing the Pathfinder and Related Systems

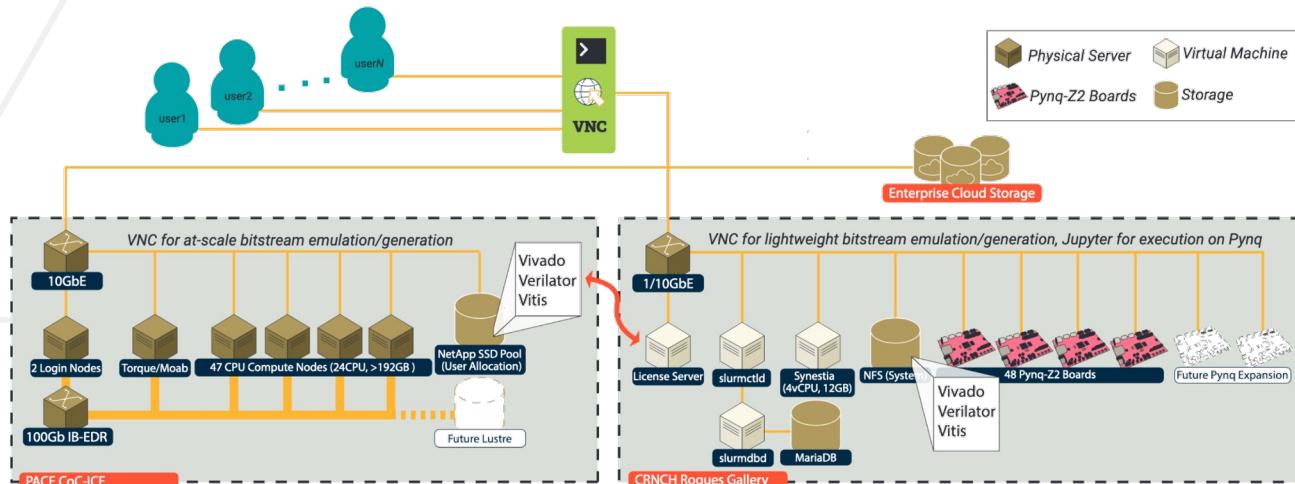
- Pathfinder access managed via Slurm (and calendar schedule)
 - Federated cluster alongside other Rogues and infrastructure nodes
 - Single entry point, single job database
- x86 physical and virtual machines for development and simulation
 - Host visual frontends (e.g. Jupyter)
 - Comparative benchmarks



RG Workflows

Slurm and Open OnDemand

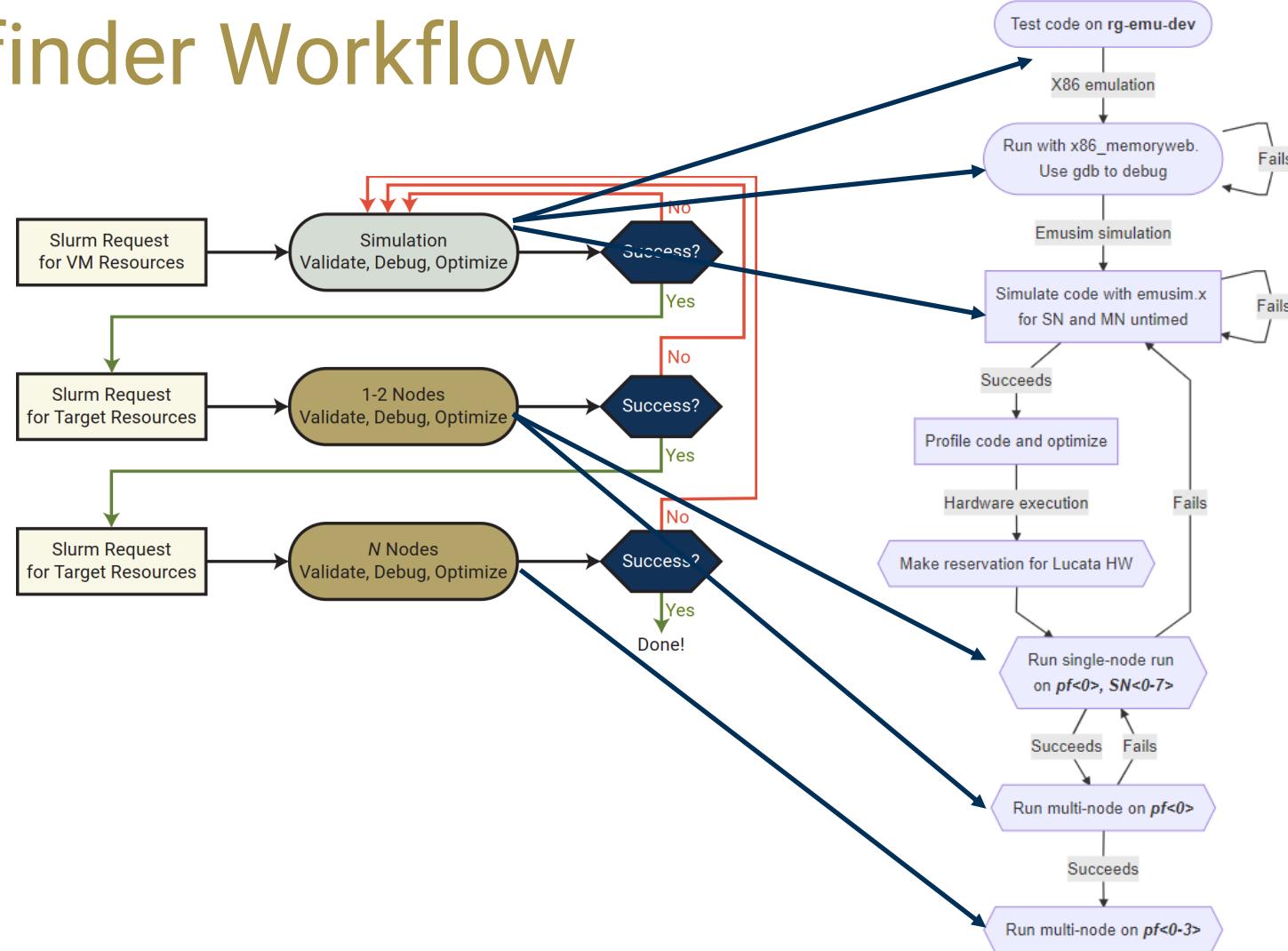
- Slurm scheduling across most resources
 - Interesting challenges with “edge device” integration
 - Running 23.11 with 10 different builds (x86, Arm, NVIDIA, RISC-V, etc.)
- Open OnDemand support to ease entry



The screenshot shows the CRNCH OnDemand interface. The top navigation bar includes Georgia Tech, Files, Jobs, Clusters, Interactive Apps, My Interactive Sessions, and Dev. Below the navigation is the CRNCH logo and the Center for Research into Novel Computing Hierarchies. A "ROGUES GALLERY" section displays a grid of hexagonal icons representing various hardware components. A message below states: "OnDemand provides an integrated, single access point for all of your HPC resources." A "Message of the Day" section follows, and a list of available jobs is provided. The bottom of the page is powered by OPEN OnDemand.

Full Lucata Pathfinder Workflow

Working with the Lucata Pathfinder still requires a complex workflow, but we can break it down into small, concrete steps for users

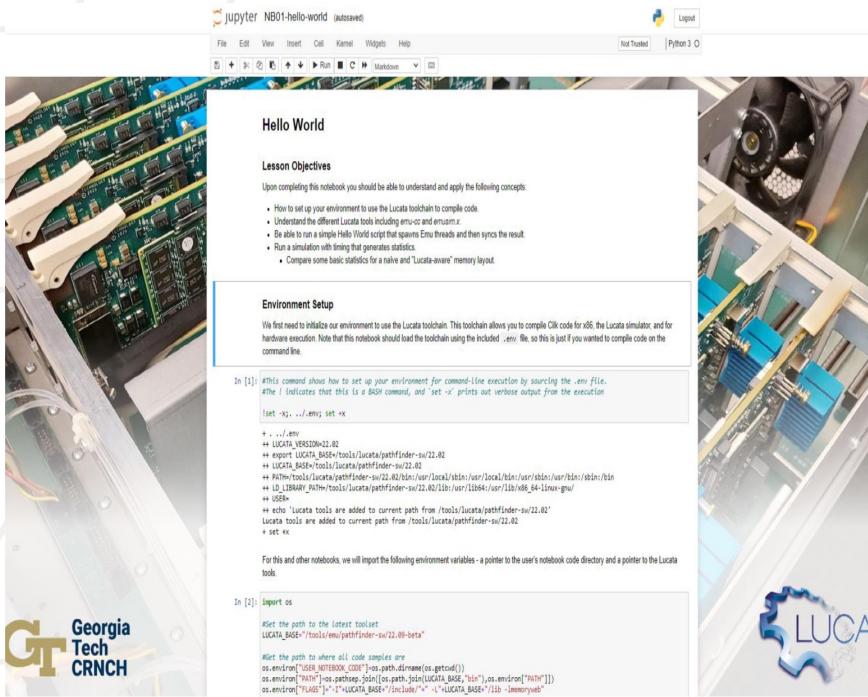


Overview

- Rogues Gallery Overview
 - Achievements and Successes
 - Unique Capabilities
- Promoting Novel Architecture Workflows
 - Example Workflow Built Around the Lucata Pathfinder
- **Education and outreach**
 - Training efforts
 - Student Engagement
 - Pynq Cluster Use Case
- Integration with PACE
 - Heterogeneous AI workflows
 - Additional Tools and Utilities
- Lessons Learned and Future Work

Outreach and Training

Community Tutorials to Engage Researchers



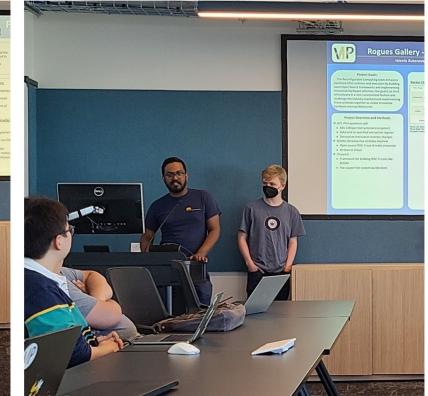
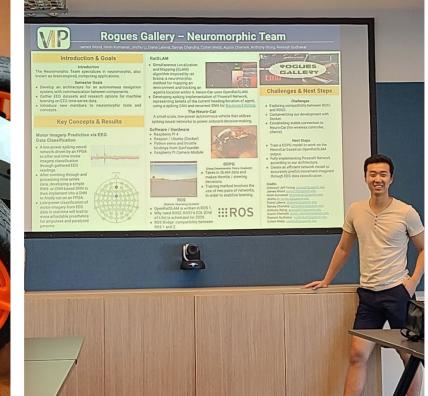
- Tutorials covering Emu Chick/Lucata Pathfinder presented at numerous conferences
 - General RG overview at ASPLOS19 and PEARC19
 - Dedicated tutorial at PEARC21 and HPEC22
 - Iteratively developed content and workflows, such as Jupyter-based approach for HPEC22
- Lessons learned for the platform and 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



Outreach and Training

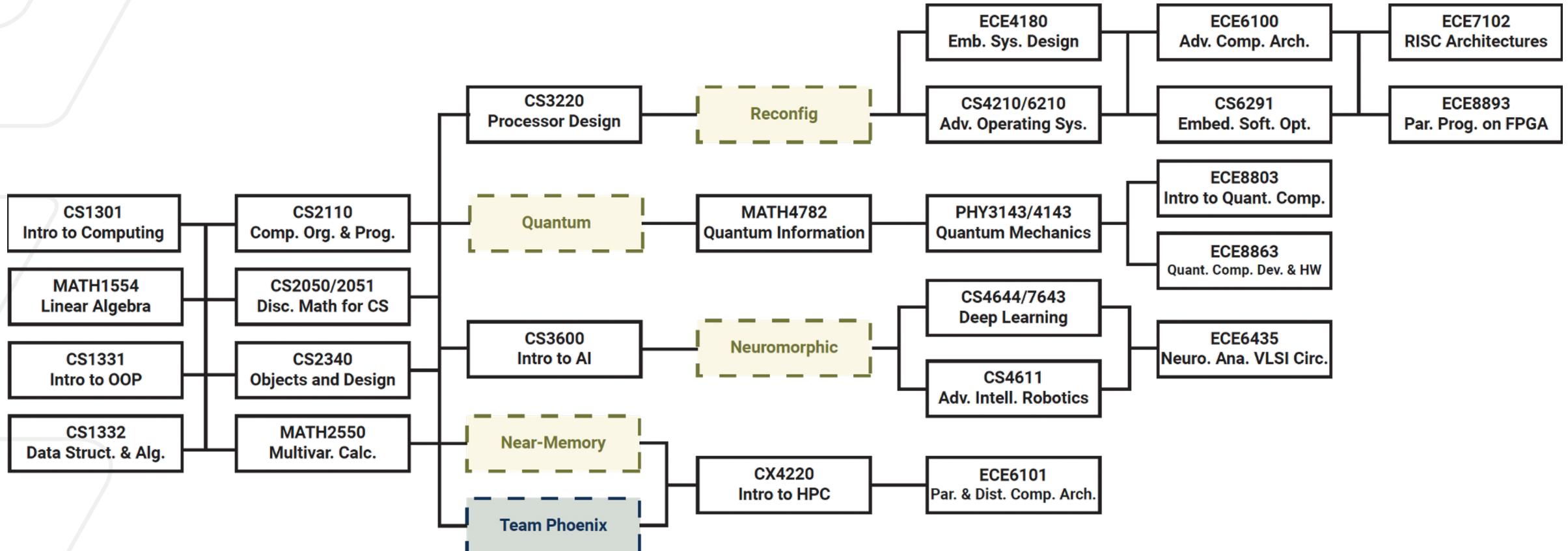
Future Computing with the Rogues Gallery

- The Vertically Integrated Projects (VIP) Program has been active since the 1990s, with a focus on:
 - Multidisciplinary team composition, bringing students from across campus
 - Vertically integrated, including a range of undergrad & graduate ranks
- Georgia Tech hosts 80 VIP teams
 - 40 institutions world-wide are part of the larger VIP consortium
- The Rogues Gallery VIP was started in Fall 2019 semester as a venue for students to explore FPGAs, near-memory compute, and HPC architectures
 - 4 subteams: Near-memory, Reconfigurable, Neuromorphic, Quantum



Outreach and Training

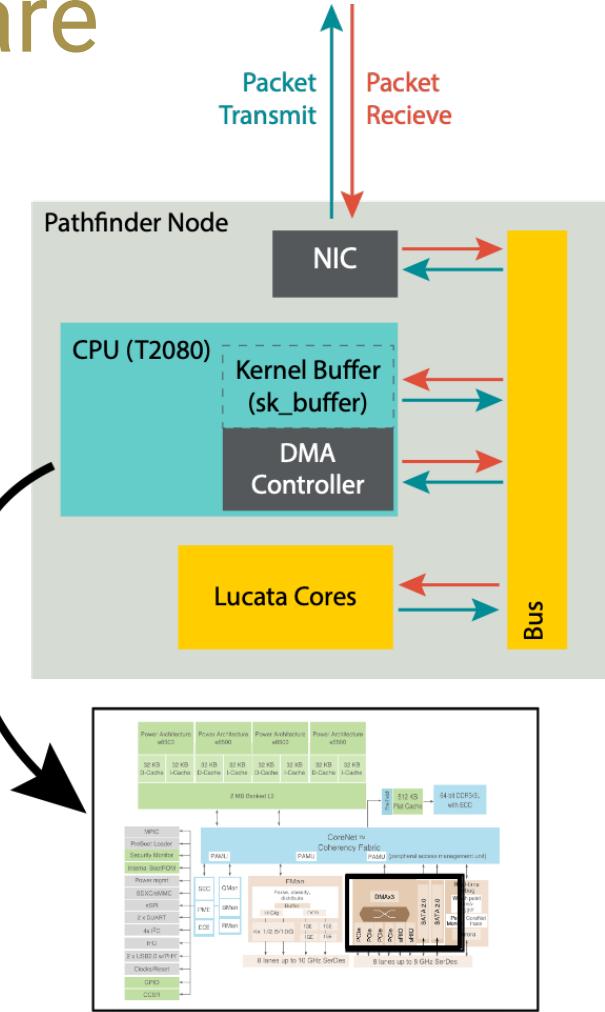
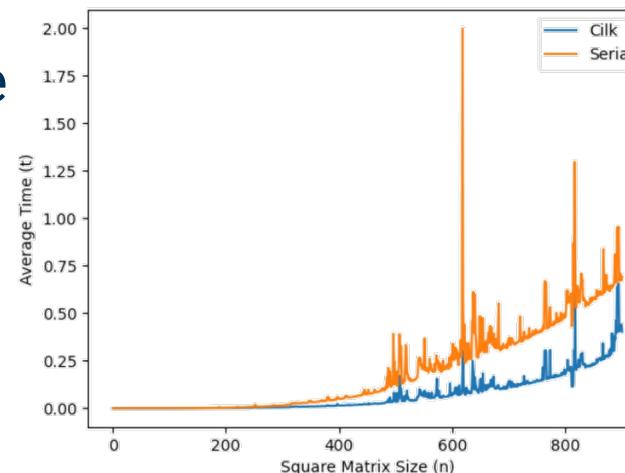
Filling the “Missing Middle”



Check out our paper on the CRNCH RG VIP class in EduPar 2023: <https://bit.ly/3Z3ZbYy>

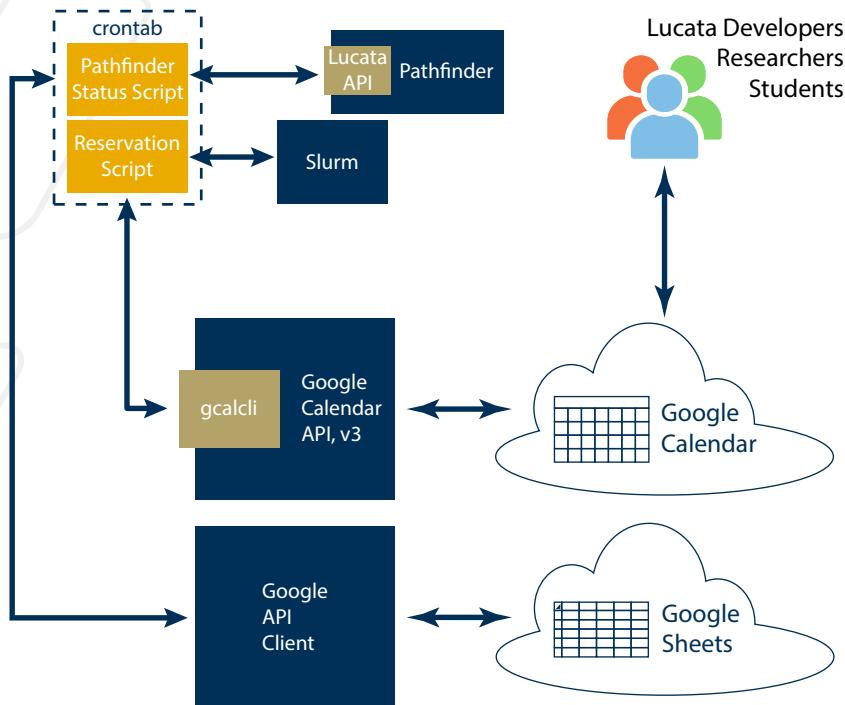
Extending CS Curriculum to Novel Hardware

- 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*



Outreach and Training

Pathways for Career Development



Logical Nodes		Chassis 1		Chassis 2		Chassis 3	
n0	sn0	n0	n1	sn0	n2	n0	sn0
n1	sn1	n1	n2	sn1	n3	n1	sn1
n2	sn2	n2	n3	sn2	n4	n2	sn2
n3	sn3	n3	n4	sn3	n5	n3	sn3
n4	sn4	n4	n5	sn4	n6	n4	sn4
n5	sn5	n5	n6	sn5	n7	n5	sn5
n6	sn6	n6	n7	sn6	n8	n6	sn6
n7	sn7	n7		sn7	n9	n7	sn7

Below the table:
C0 Status Multinode C0-C1 C1 Status Multinode C0-C1 C2 Status Multinode C2 C3 Status Single node

Outreach and Training

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

Overview

- Rogues Gallery Overview
 - Achievements and Successes
 - Unique Capabilities
- Promoting Novel Architecture Workflows
 - Example Workflow Built Around the Lucata Pathfinder
- **Education and outreach**
 - Training efforts
 - Student Engagement
 - **Pynq Cluster Use Case**
- Integration with PACE
 - Heterogeneous AI workflows
 - Additional Tools and Utilities
- Lessons Learned and Future Work

Outreach and Training

Adapting to a Changing Compute Landscape

In 2021, GT was confronted with a need to teach (mostly remote) students courses that normally involve labs that use smaller FPGA boards like the Pynq Z2.

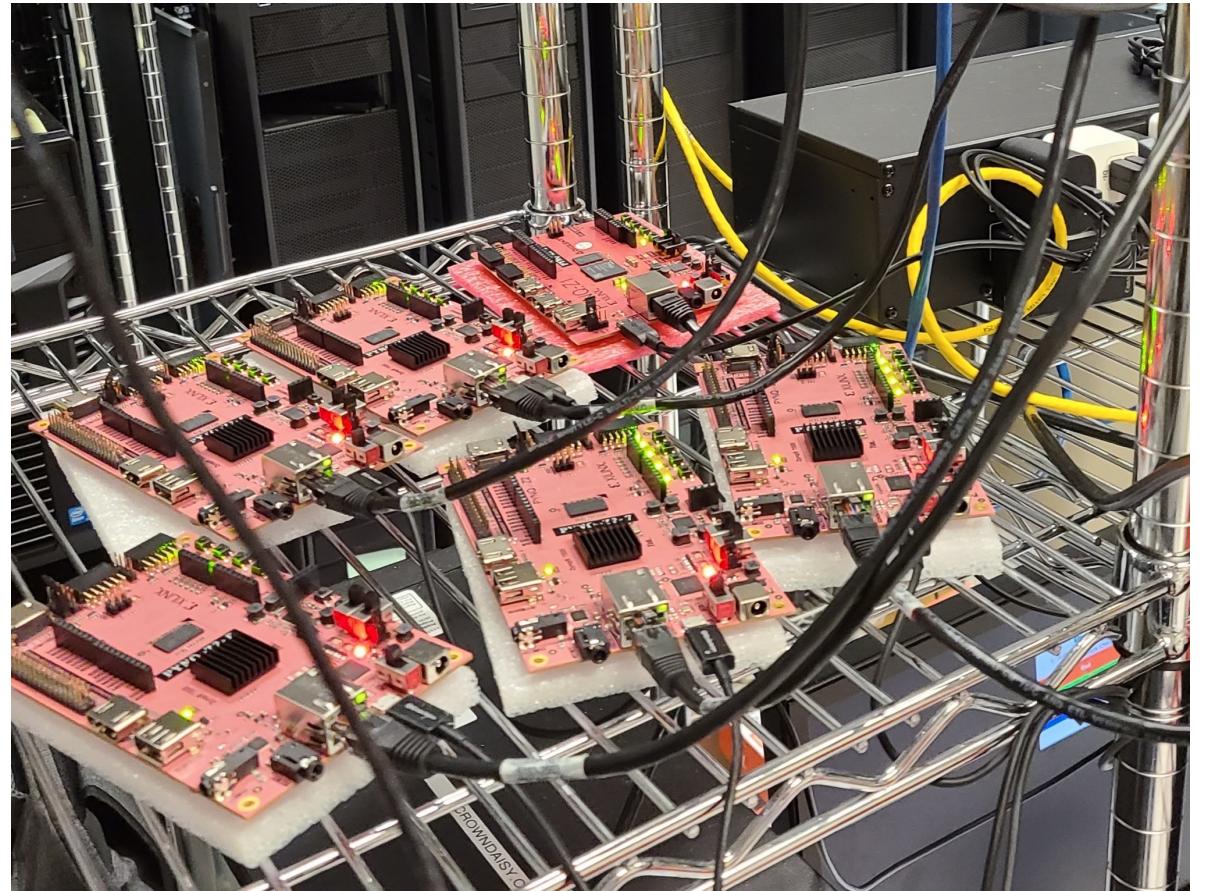
Specifically, we looked at the following questions:

- How can we support larger numbers of students for a ***novel FPGA cluster*** that ties in with an existing HPC system, GT's Instructional Cluster Environment (ICE)?
- How do we support key requirements like ***data separation and privacy for student data*** while allowing for free flow of data between isolated clusters and local access points?
- What types of training would students require to migrate from "hands-on" infrastructure to a remotely scheduled cluster that uses a scheduler like Torque/Moab or Slurm?



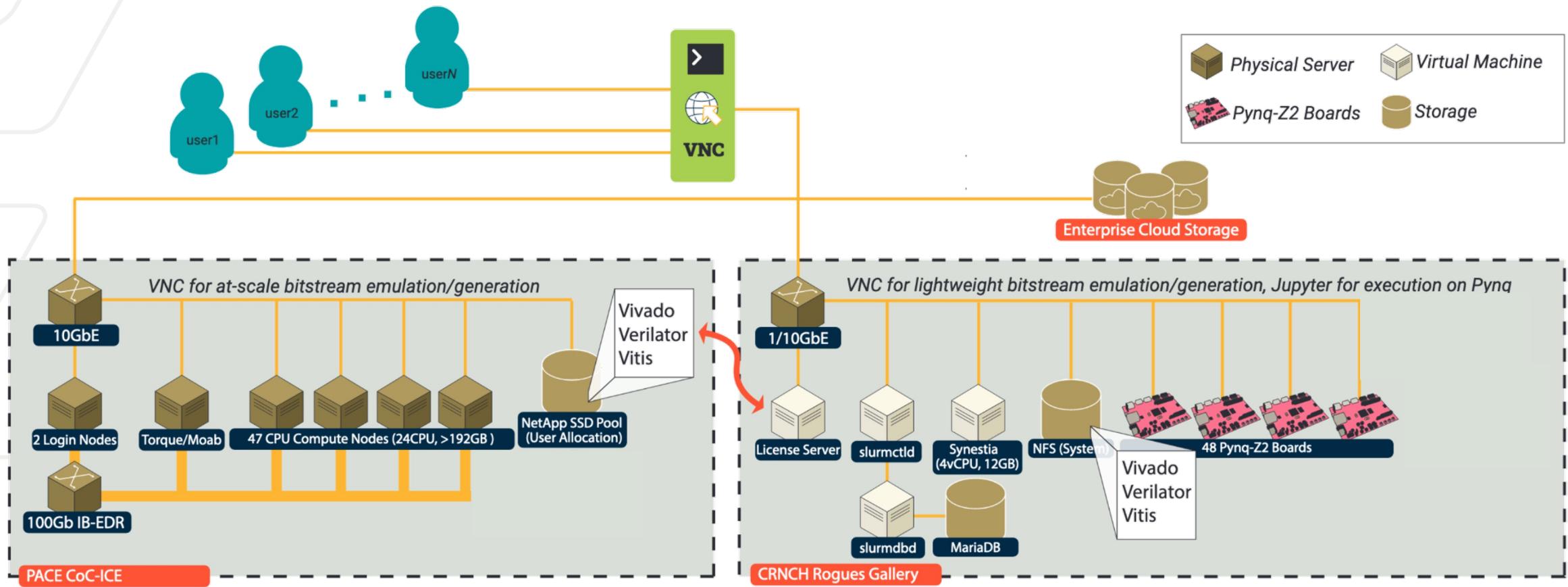
Challenges with Existing Solutions

- Two courses with 90+ (CS 3220 and 40+ (ECE 8803) students
- Typical PYNQ FPGA workflows require:
 - The usage of Vivado and Vitis tools with large memory requirements
 - Hands-on interaction with a board to set up Jupyter webserver interfaces
- Supply chain issues limited the number of boards we could deploy



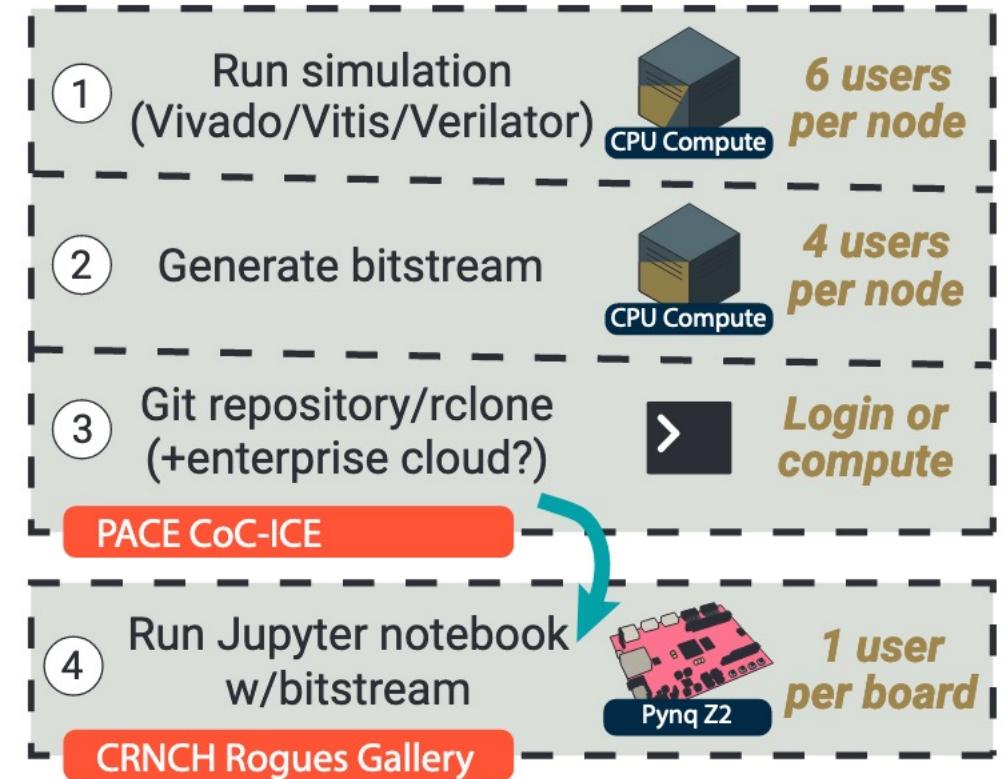
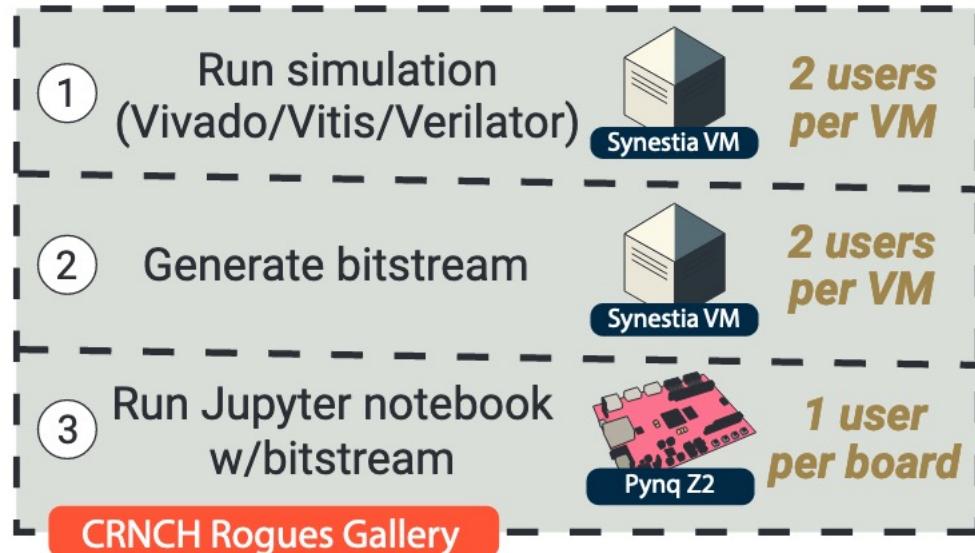
How can we best support remote FPGA development and usage at scale?

CRNCH + PACE Instructional Workflow



See our SEHET22 paper, “Enhancing HPC Education and Workflows with Novel Computing Architectures” at <https://github.com/gt-crnch-rg/rg-publications/>

Two User Workflows



A combined workflow allows for a dramatic reduction in the bottleneck of design and initial testing

- However, there are some issues to consider related to data security
- ICE is set up to manage student data unlike research testbeds like CRNCH RG

Overview

- Rogues Gallery Overview
 - Achievements and Successes
 - Unique Capabilities
- Promoting Novel Architecture Workflows
 - Example Workflow Built Around the Lucata Pathfinder
- Education and outreach
 - Training efforts
 - Student Engagement
 - Pynq Cluster Use Case
- **Integration with PACE**
 - Heterogeneous AI workflows
 - Additional Tools and Utilities
- Lessons Learned and Future Work



Integration

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.



Integration

Integration with Institutional Computing

Application portability across clusters through containerized workflows!



Chisel

Heterogeneous AI Workflows

1 Common Storage

Common storage (CEDAR, IDEaS) across campus allow opportunity to migrate data/models to optimal architectures.

Toolkits for exploring alternative architectures and pipelines to evaluate appropriate hardware before production

2 Simulate and Dispatch

3 Function as a Service

Globus Compute Endpoint (funcX)/CyberShuttle as a means to target different accelerators seamlessly within a single application.

Overview

- Rogues Gallery Overview
 - Achievements and Successes
 - Unique Capabilities
- Promoting Novel Architecture Workflows
 - Example Workflow Built Around the Lucata Pathfinder
 - Support for Software Engineering
- Education and outreach
 - Training efforts
 - Student Engagement
- Integration with PACE
 - Heterogeneous AI workflows
 - Additional Tools and Utilities
- **Lessons Learned and Future Work**

Future Work

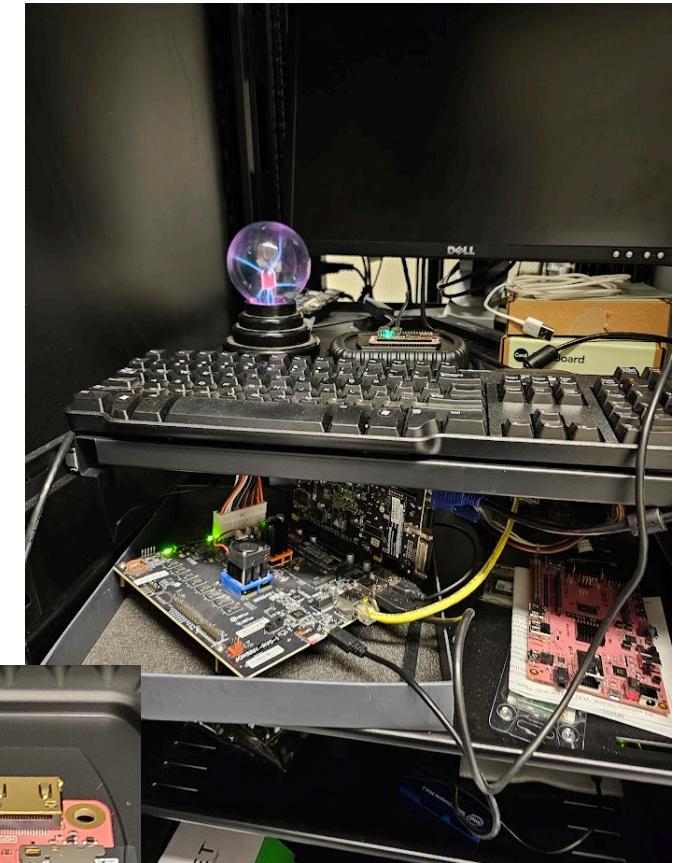
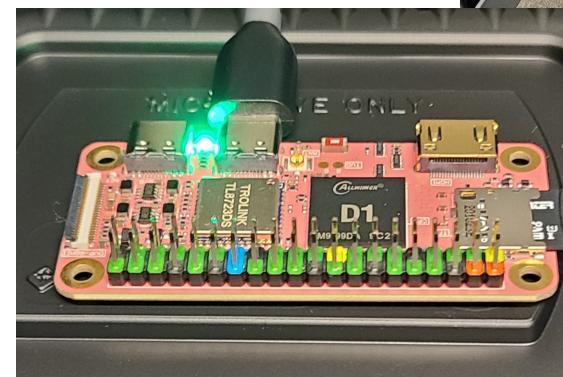
Challenges in CRNCH RG Deployment

Supporting one class of novel architecture requires an enormous investment in backend resources, training, and infrastructure

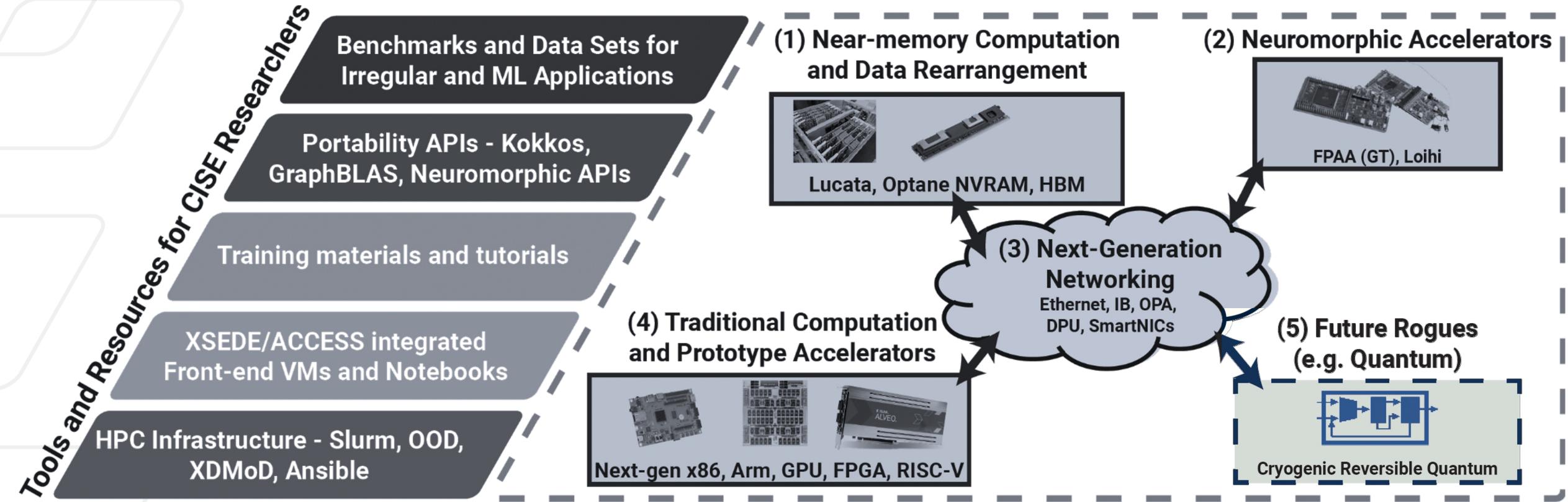
It's hard to imagine exerting the same level of effort for every new platform!

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



New Frontiers



New Frontiers

Frontiers

Benchmarks and Data Sets for

(1) Near-memory Computation

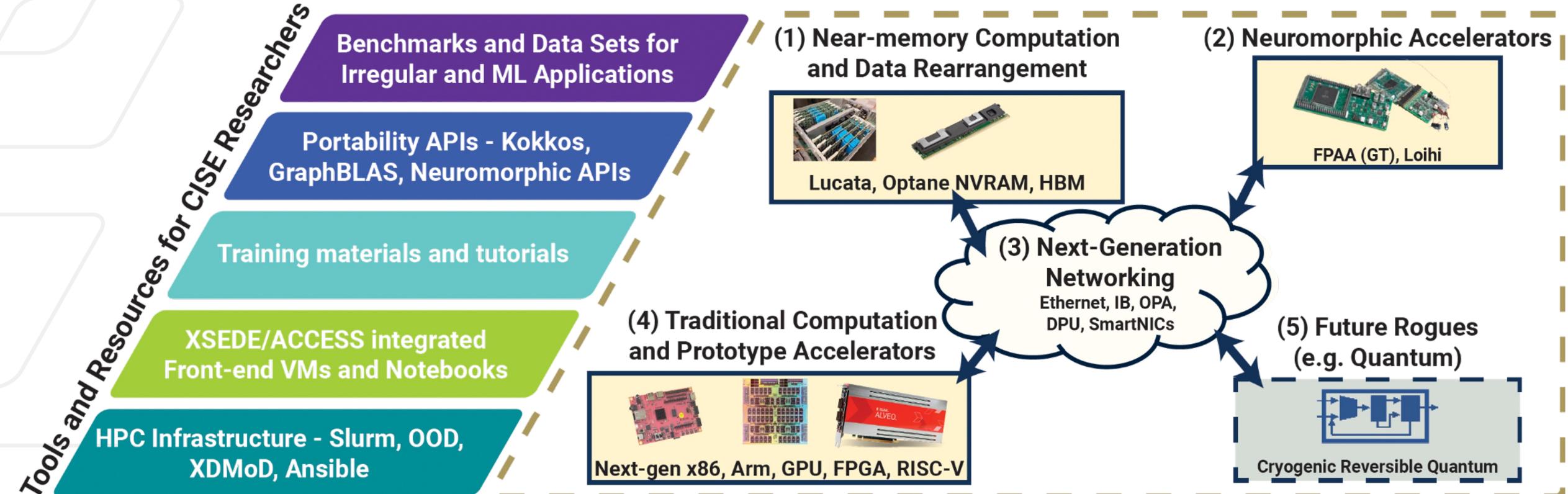
(2) Neuromorphic Accelerators

The near-term Rogues Gallery is not just about focusing on far-off technologies like quantum, reversible, etc. It will also likely 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
- Toolchains for reversible computation simulation and evaluation
- 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

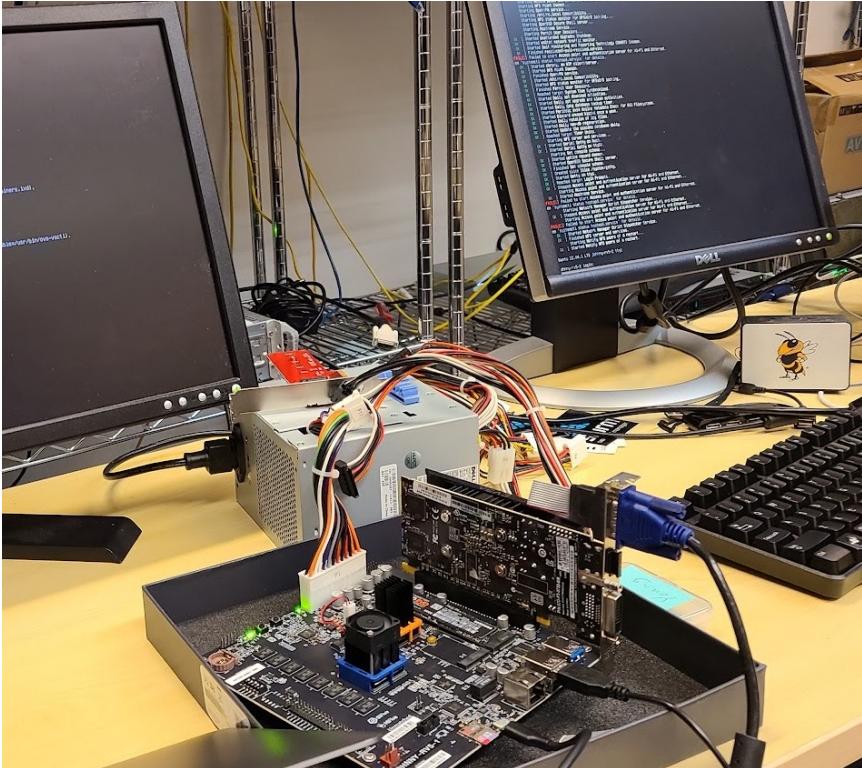
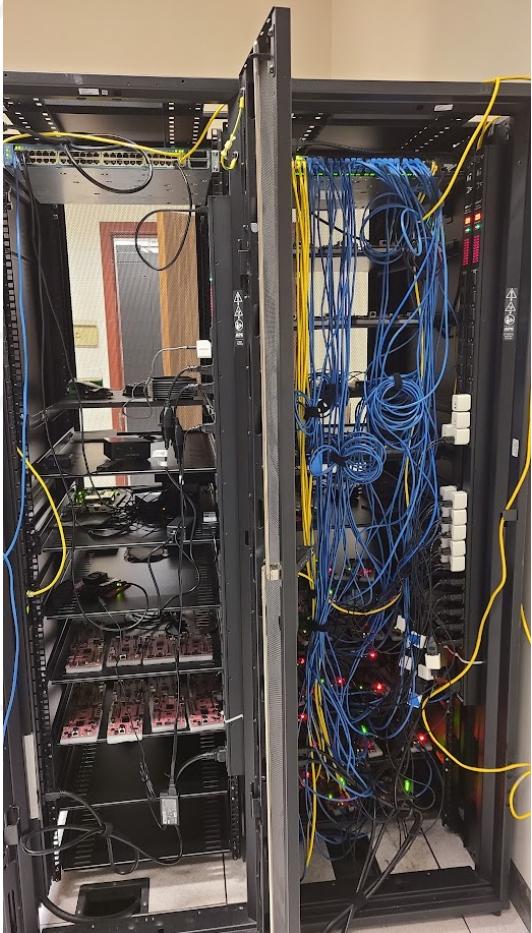
New Frontiers



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

System focus within the testbed will likely change as researchers prioritize different topics, but we should continue to understand similarities and differences and utilize our tools and resources to efficiently tackle complex problems.

Questions?



Project site: crnch-rg.cc.gatech.edu
GH Presence: <https://github.com/gt-crnch-rg>

Apply for access: <https://crnch-rg.cc.gatech.edu/request-rogues-gallery-access/>