

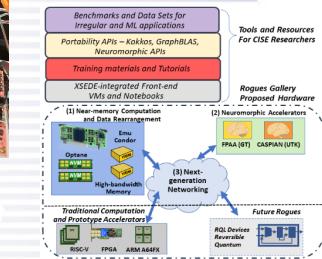
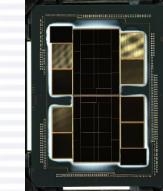


Rogues Gallery 2021 Updates



Presented by: Jeffrey Young
CRNCH Rogues Gallery Director
January 28th, 2021

Rogues Gallery – Timeline



2017

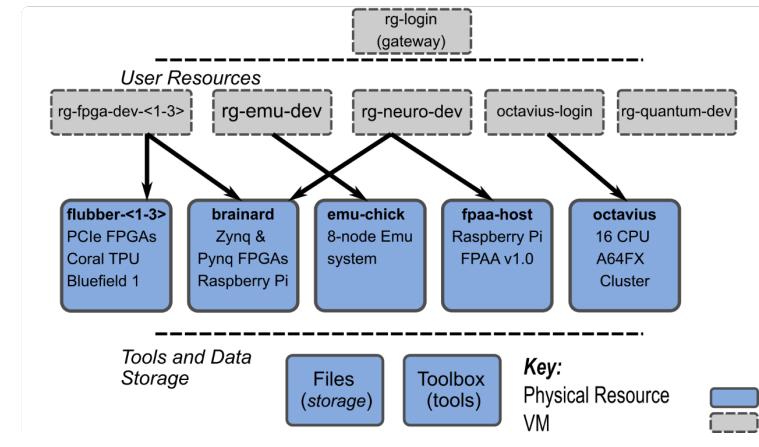
2018

2019

2020

2021

Current Rogues Gallery Stats



The Rogues
Gallery is a
disaggregated
testbed.

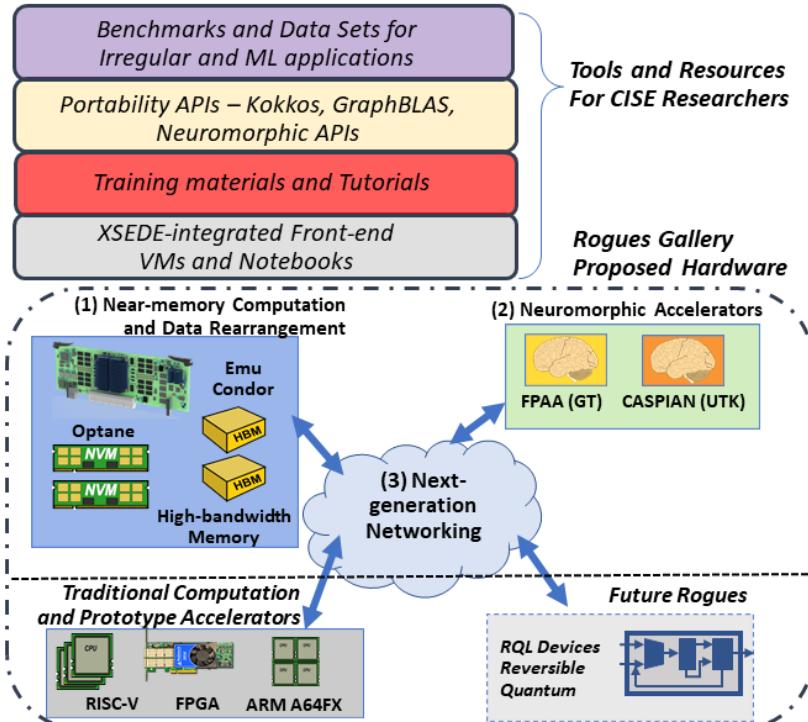
Rogues Gallery Users

- ~10 VMs, 12 servers, numerous boards
- 145 users overall; 26 external users
- 120+ students supported via TechFee associated work in 2020-2021

New Hardware in 2020

- Pynq and FPGA cluster (GT TechFee)
- Arm A64FX (NSF Hive project)
- EDR IB Switch, Bluefield SoC, and InnovaFlex networking

Rogues Gallery: A Community Research Infrastructure for Post-Moore Computing



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

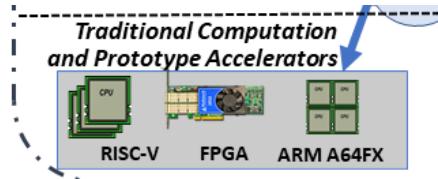
CNS-2016701, \$1.3M over 3 years

Supports deploying:

- Rack-scale Lucata Pathfinder 16 node system
- Neuromorphic accelerators
- Smart networking and 5G equipment
- Backend infrastructure

This grant focuses on **community engagement and post-Moore training**

“More Moore”



Post-Moore computing also includes evolutions to traditional architectures!

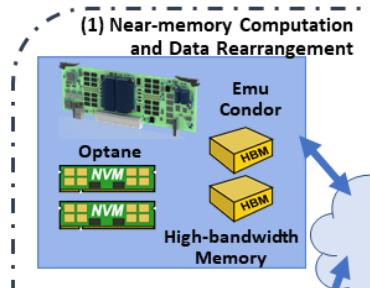
CRNCH recently deployed one of the few open-access Arm A64FX systems as part of the NSF Hive program (OAC-1828187)

- 48 core Arm CPUs with 32 GB of HBM
- Focuses on streaming and low-locality workloads
- First commercial Arm processors with SVE support

FPGAs continue to be important for prototyping novel accelerators

- See the Vortex GPGU as an example of a new RISC-V based FPGA accelerator
vortex.cc.gatech.edu

Near-memory Computing



Related Work:

Emu: Brian Page, Peter Kogge, “Scalability of Streaming on Migrating Threads”, HPEC 2020

Optane: Tony Mason, Thaleia Dimitra Doudali, Margo Seltzer, Ada Gavrilovska, “Unexpected Performance of Intel® Optane™ DC Persistent Memory”, CAL 2020

Sparse data and data movement costs will continue to dominate application concerns for the near future leading to opportunities for near-memory computing

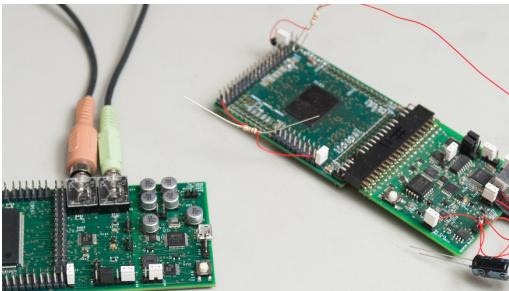
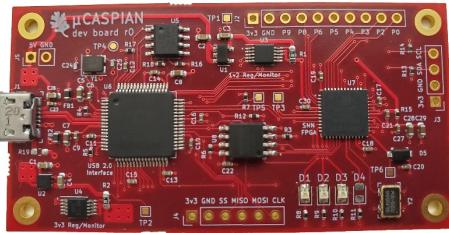
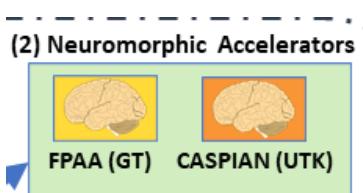
The largest acquisition will be of a Pathfinder-S two chassis, 16 node system

- ~8X the processing elements with faster clocks, improved networking, and software stack
- Bolstered by related NSF projects and efforts like a Kokkos backend for Emu Cilk and GraphBLAS support

Optane NVDIMMs allow for large capacity workloads in traditional server platforms

FPGAs again provide a “near-memory” accelerator platform with on-board HBM

Neuromorphic Computing



Increased energy consumption by new AI algorithms necessitates a shift to more efficient neural-focused architectures

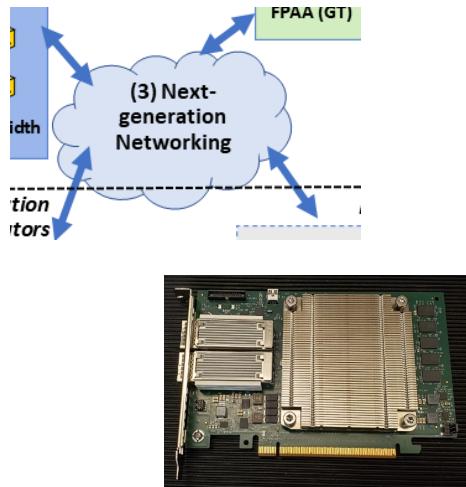
The CCRI will fund the development of a “large-scale” Field Programmable Analog Array (FPAA) based on designs by Dr. Hasler’s group

- The FPAA provides a mixed analog/digital design platform with open-source tooling
- A multi-chip deployment will enable larger SNNs based on hhNeuron-inspired blocks

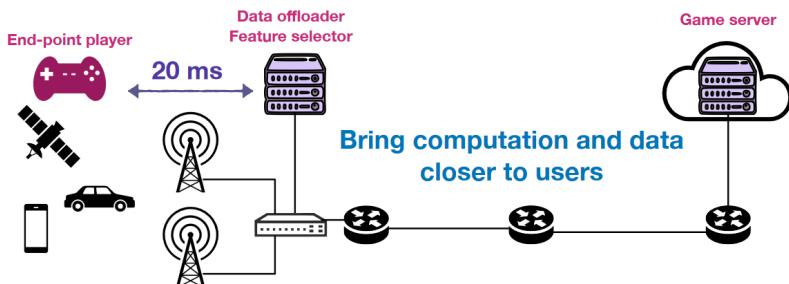
Georgia Tech is also working with ORNL researchers (PI: Parker Mitchell) to deploy their digital neural architecture Caspian on FPGA

- μCaspian fits 256 neurons and 4096 synapses on a tiny FPGA!

Novel Networking



E.g. AR/VR gaming



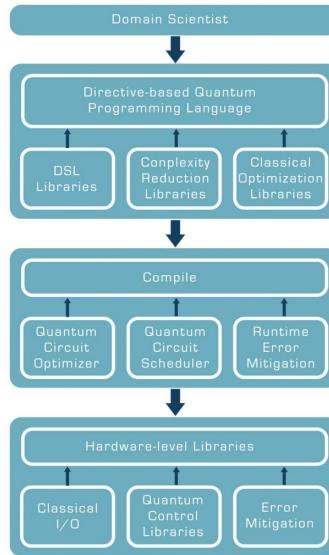
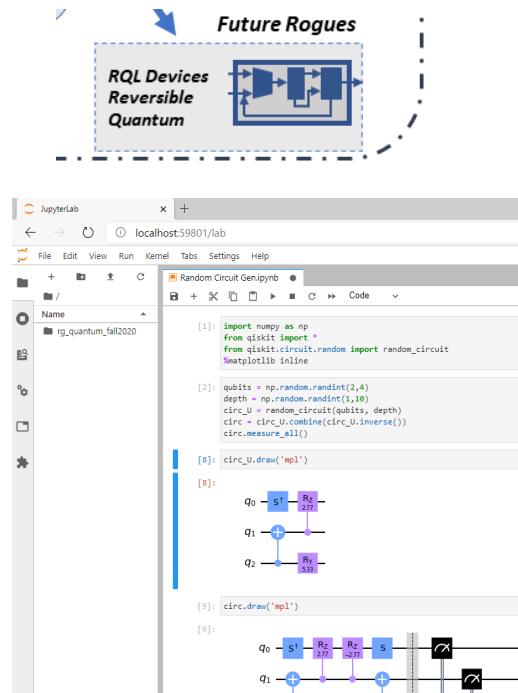
Computation will increasingly move into the network as a means to further reduce data movement. Similarly edge computing will shift processing of data to lower-power devices.

SmartNICs like the NVIDIA Bluefield SoC and QSPF enabled FPGAs enable opportunities for “in-network computing” focused on processing large data sets and data-intensive simulations

- The CCRI enables us to deploy several different SmartNIC accelerators for this purpose

Students working with Drs. Gavrilovska and Bhardwaj are developing next-generation “mobile edge computing” and 5G

Quantum Computing



Quantum computing education is on a robust growth path while hardware remains expensive and limited. However, there are tremendous opportunities for tools, algorithms, and compilers.

CRNCH does not officially support quantum hardware, but we do provide access to common toolkits like QISKit, QCOR, AIDES-QC, and JaqalPaq via VMs

- Jupyter notebooks can be accessed remotely via RG!

Future work is focused on software and scheduling development to support novel testbeds like GTRI's ion trap testbed (currently in use for DARPA ONISQ).

Rogues Gallery: Enabled Research Threads

CISE Enabled Research Pillars

Sparse/Irregular HPC

- Graph analytics
- Scientific Computing
- Database and Big Data Acceleration



HW/SW Codesign

- Polyhedral compilation,
- Design of libraries, runtimes, APIs for novel devices
- Benchmarking and characterization

Machine Learning

- Low-power edge AI
- Autonomous vehicles
- SLAM for robotics
- Dynamic and life-long learning



Next-generation Networking

- 5G software stacks
- Edge computing services
- In-situ and encrypted data analysis
- Data reduction and line-speed DSP

Rogues Gallery Hardware and Software Support

- Emu Pathfinder
- FPGA+HBM
- CASPIAN
- Tensor, Streaming Graph APIs
- ARM A64FX

- Emu Pathfinder
- FPGAs + RISC-V
- Optane
- Kokkos, Habanero-C runtimes

- FPAA and CASPIAN
- EMU Pathfinder
- FPGAs
- RASP/TENNLab SW
- Emu Scikit-learn

- Ettus USRP-2947 and Ettus E-320
- Mellanox Bluefield NICs
- FPGAs
- FPAA and CASPIAN

RG CCRI PIs and Personnel

Co-PIs



Tom Conte



Ada Gavrilovska



Jennifer Hasler



Rich Vuduc

Senior Personnel



Alex Daglis



Semir Sarajlic

Technical Support



Will Powell



Jeff Valdez

Community Outreach

CRNCH RG is focused on supporting post-Moore computing at GT and in the community via:

Minisymposiums and Birds of Feather – RG has supported symposiums at SIAM PP 2020, SIAM CSE 2019, and BoFs at ISC and SC each year

- Popular Arm BoFs at ISC and SC led to the creation of the Arm HPC User's Group (AHUG) with direct involvement from the RG director
- CRNCH also supports the Advanced Architecture TestBed (AATB) BoF at SC each year since 2018



Tutorials – opportunities for post-Moore HW training



Class Support – Techfee and VIP efforts

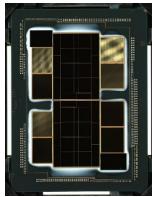
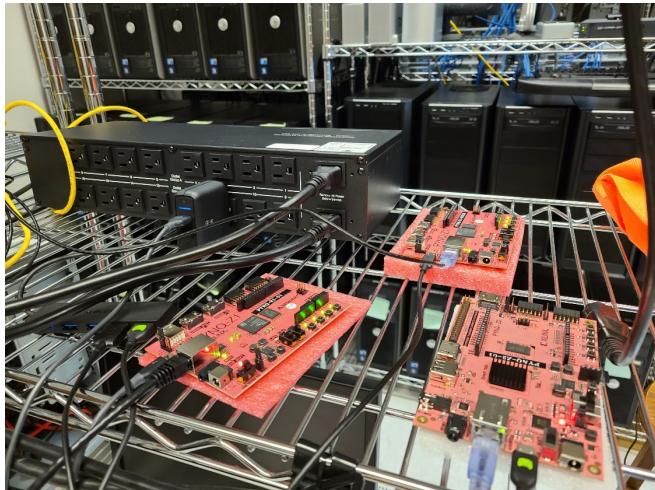
Tutorials

Training is key to engaging students and potential users

- We plan to support 2-3 tutorials per year via the CCRI program
- 2020 will likely be virtual with a target for new Lucata and neuromorphic tutorials
- We also hope to help support Arm's extremely popular SVE hackathons with our A64FX installation.



Rogues Gallery TechFee Support



TechFee is an internal proposal process meant to support students for instructional purposes

- Focus on core classes with increasing focus on our “online” classes like OMSCS
- Provides unprecedented exposure to novel hardware. “Does your school have a DGX, Power9, and A64FX?!”

CRNCH is currently supporting a reconfigurable computing initiative (CS3220, 7290, 8803)

- 2021 will see the deployment of AMD GPU and Intel GPU nodes meant to provide an introductory HPC experience close to that of ANL’s Aurora and ORNL’s Frontier development systems

Vertically Integrated Projects (VIP) Team

- Undergraduate research opportunity for credit; teams are self-directed with guidance from faculty.
- Current projects:
 - NeuroCar – implement sensing and control using SNNs with Nengo FPGA platform; replicate results of the autonomous GT Rally Car with lower power
 - Qubit allocation optimization – evaluate techniques using IBM's Q experience and ORNL's XACC and attempt to build a linear systems algorithm approach to test possible solutions
 - No-history branch prediction – sort the register file on the fly to assist with branch prediction and limit security vulnerabilities



Rogues Gallery – Engagement Opportunities

Request an account on the Rogues Gallery

- <http://crnch.gatech.edu/request-rogues-access>

Corporate sponsorships/partnerships

- CRNCH Rogues Gallery is set up to help test computing hardware for interested external industry partners as part of sponsorship and partnership agreements.

Vertically Integrated Projects (VIP) team

- Suggest project ideas for our undergraduates to work on!
- Learn more at <https://www.vip.gatech.edu/teams/future-computing-rogues-gallery>



Acknowledgments

This work is supported by:

- Hardware donations by Micron, Xilinx, and Intel
- NSF Rogues Gallery (CNS-2016701), NSF “Hive” Project (OAC-1828187), the NSF SuperSTARLU project (OAC-1710371), and IARPA.
- Sandia National Laboratory supports the Rogues Gallery VIP undergraduate research class, and parts of this research have been funded through the Laboratory Directed Research and Development (LDRD) program at Sandia. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA-0003525.