



[www.chameleoncloud.org](http://www.chameleoncloud.org)

# THE CHAMELEON TESTBED AND SYSTEMS RESEARCH

**Kate Keahey**

Mathematics and CS Division, Argonne National Laboratory

CS CASE, University of Chicago

*keahey@anl.gov*



# CHAMELEON: AN EXPLORATORY TESTBED



- ▶ From large to small – diversity and scale in hardware:
  - ▶ **Supercomputing datacenters** (UC, TACC) over 100G network – to **edge devices**
  - ▶ **Diverse:** FPGAs, GPUs, NVMe, NVDIMMs, Corsix switches, edge devices via CHI@Edge, etc.
  - ▶ **Distributed: CHI-in-a-Box** sites at **IIT, NCAR, Northwestern, and UIC** – more coming
- ▶ Chameleons like to change – testbed that adapts to your experimental needs
  - ▶ **From bare metal reconfigurability/isolation** -- KVM cloud – to containers for edge
  - ▶ Capabilities: power on/off, reboot, custom kernel boot, serial console access, etc.
- ▶ Based on mainstream open source – proud to be cheap!
  - ▶ 50% leveraging and influencing **OpenStack** + 50% “special sauce”
- ▶ Promoting disruption in digital artifact sharing
  - ▶ Integration with **Jupyter** for non-transactional experiment packaging
  - ▶ **Trovi** for experiment sharing and discovery, **Chameleon Daypass** for access sharing
  - ▶ Reproducibility and education: digital sharing killer apps!



# TESTBED BY THE NUMBERS

600+  
Papers  
published

95  
Countries

350+  
Institutions

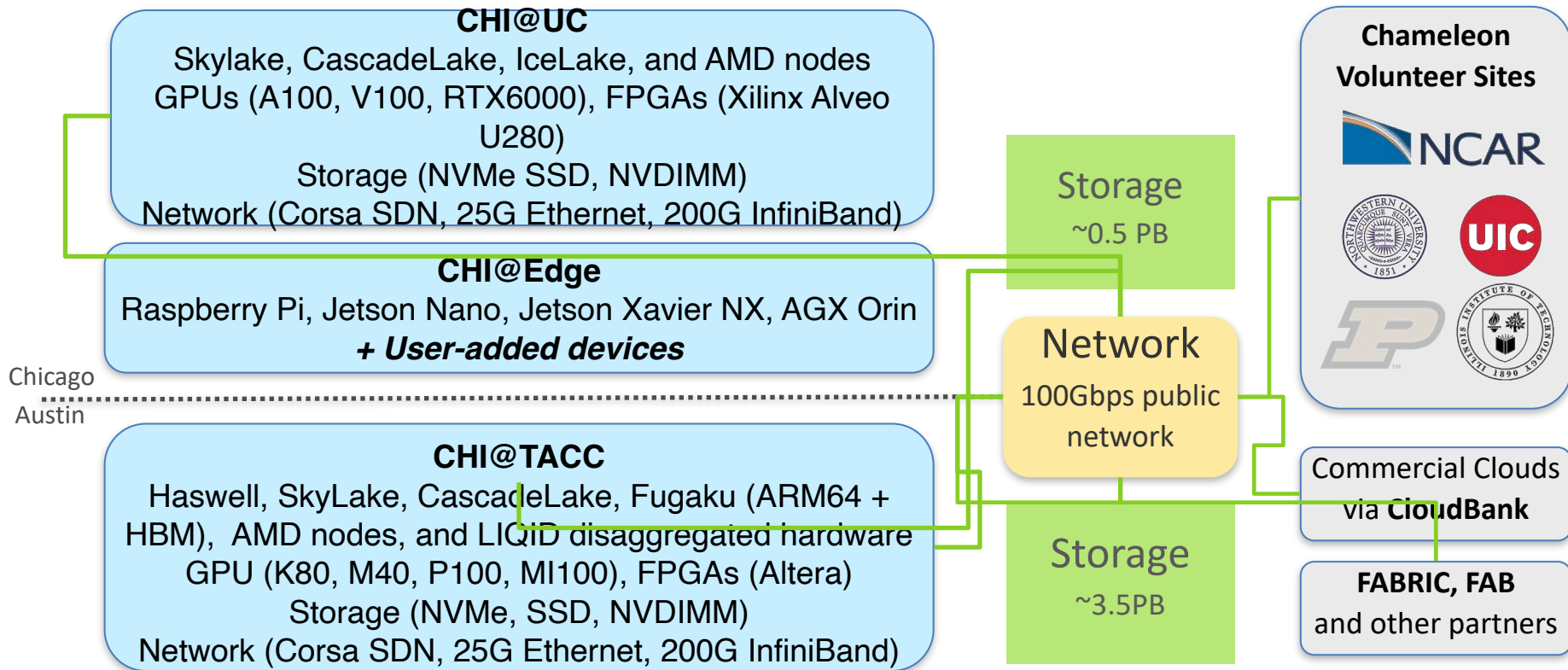
8,000  
+  
Users

1,000  
+  
Unique  
projects

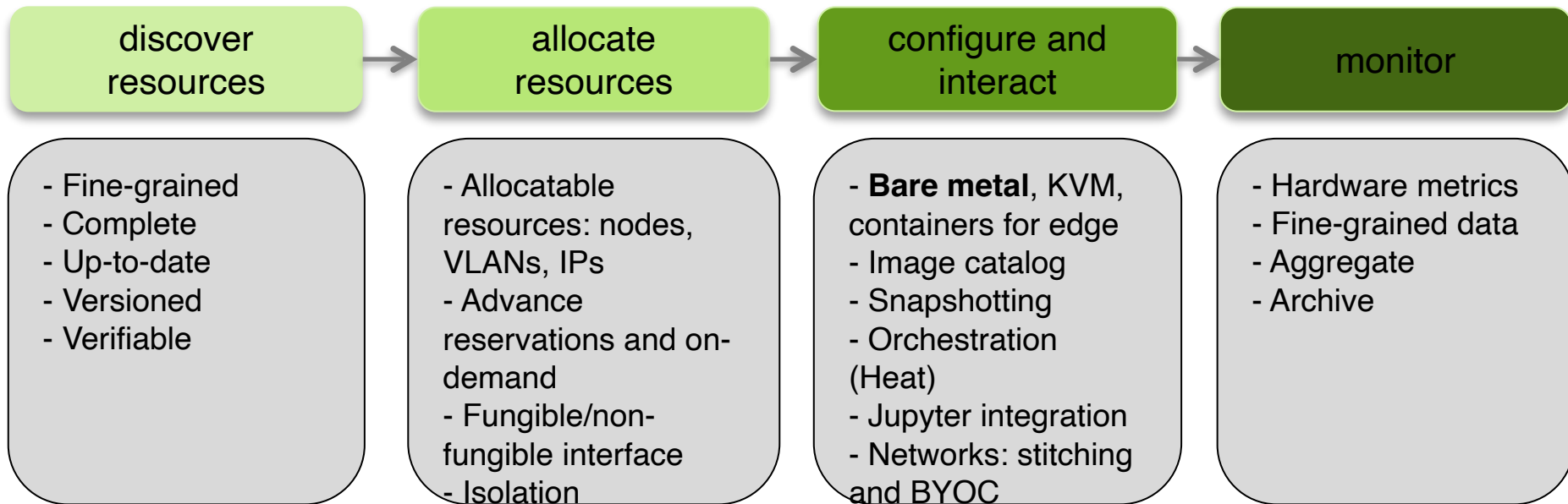


# CHAMELEON HARDWARE

*Notable still to come:  
disaggregated hardware (GigaIO)*



# CHI EXPERIMENTAL WORKFLOW



*Authentication via federated identity, accessed via GUI, CLI and **python/Jupyter***

*Paper: "Lessons Learned from the Chameleon Testbed", USENIX ATC 2020*

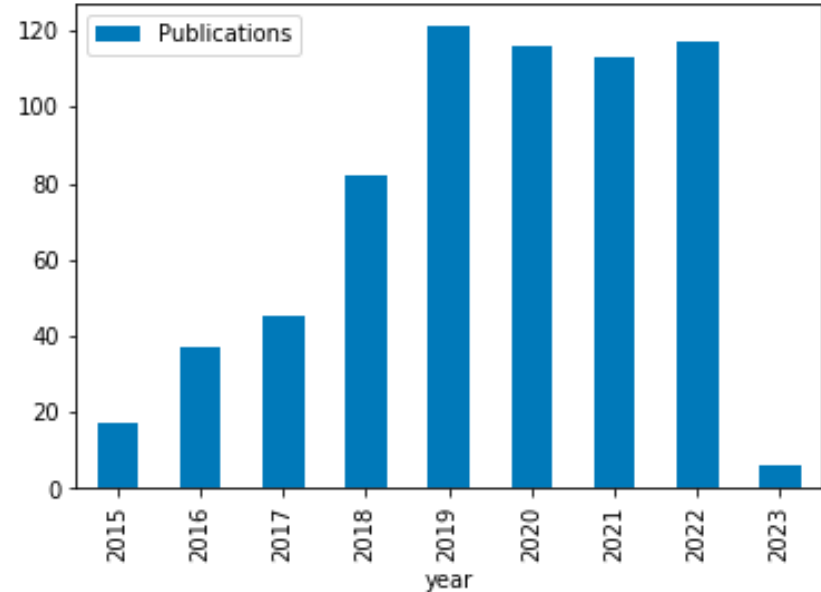
# NOT JUST A TESTBED, A COMMUNITY



Supporting research projects in architecture, operating systems design, virtualization, power management, real-time analysis, security, storage systems, databases, networking, machine learning, neural networks, and many others.

# RESEARCH IMPACTS

- ▶ Of users, projects, and projections
- ▶ Publications: 613 and counting
- ▶ User publications on our web page
- ▶ Sources
  - ▶ user-reported
  - ▶ google scholar
  - ▶ scopus
  - ▶ semantic scholar
- ▶ References versus acknowledgements
- ▶ Close to ~9,000 citations of research



# HIGHLIGHT: CHI-IN-A-BOX



- ▶ CHI-in-a-box: packaging of CHameleon Infrastructure (CHI)
  - ▶ Internal packaging of a commodity-based testbed
  - ▶ Packages the system as well as the operations model
  - ▶ Hub and spoke management, version-controlled site configuration management as code, containerization, monitoring, detection, and remediation tools
  - ▶ Support for Bring Your Own Device (BYOD) model: Doni allows administrators to dynamically enroll resources, define availability windows, and streamline operations
- ▶ Deployment
  - ▶ Deployed Associate/Volunteer Sites: IIT, NCAR, Northwestern, Purdue, and UIC
  - ▶ Independent testbed: ARA
  - ▶ In conversation/progress: SDSC, OCT/U Mass, FIU, ORNL, KTH (edge/wireless only), NUS, and others

*Paper: "CHI-in-a-Box: Reducing Operational Costs of Research Testbeds", PEAR*



# HIGHLIGHT: CHI@EDGE



A lot like a cloud!  
All the features we know  
and love – but for edge!  
“Edge to cloud from one  
Jupyter notebook.”

Not at all like a cloud!  
Location, location, location!  
IoT: cameras, actuators,  
SDRs!  
Not server-class!  
And many other challenges!



## ► CHI@Edge: all the features you love in CHI, plus:

- Reconfiguration through non-prescriptive **container deployment** via OpenStack interfaces (using K3 under the covers)
- Support for “standard” **IoT peripherals** (camera, GPIO, serial, etc.) + easy for you to add support for your own peripherals
- **Bring Your Own Device (BYOD): Mixed ownership** model via an SDK with devices, virtual site, and **restricted sharing** – building on OpenBalena



*Paper: “Chameleon@Edge Community Workshop Report”, 2021*

# HIGHLIGHT: PRACTICAL REPRODUCIBILITY

*Practical reproducibility == cost-effective enough to be mainstream*

- ▶ End-to-end packaging with literate programming
  - ▶ Credential integrated JupyterLab environment: convenience of notebook + power of testbed
  - ▶ Imperative, non-transactional, annotated
- ▶ Trovi: an experiment sharing repository
  - ▶ Portal to present, browse, filter, and find
  - ▶ Integrated with Jupyter/Chameleon, Swift, Zenodo, and github – working with others
  - ▶ Open APIs: integration with FABRIC, Jetstream2, and other testbeds
- ▶ Chameleon daypass



# ENCOURAGING EDUCATION

- ▶ Educational projects over the last 2 years
  - ▶ '21/'22: 76 projects total / 30 teaching a class/course
  - ▶ '22/'23: 91 projects total / 32 teaching a class/course
  - ▶ Educational projects include: exploration/developing educational materials, use in classes, targeted training/one-off education events (e.g., summer schools), tutorials, hackathons, student competitions
- ▶ IndySCC: a non-course educational use example
  - ▶ IndySCC: major International student competition held annually at the Supercomputing conference, combined with instruction and skill development activities over the leading up period
  - ▶ Objective: achieve the best performance within a power budget over a 49 hour competition
  - ▶ Hosted on Chameleon since 2021: 5 teams (2021); 11 (out of 28) teams (2022), 15 teams (out of 34) teams (2023)
- ▶ Blog: <https://chameleoncloud.org/blog/category/education/>
- ▶ User videos: <https://chameleoncloud.org/chameleon-cloud-users-meeting/user-meeting-2023/>



*IndySCC SC'22 competition results*



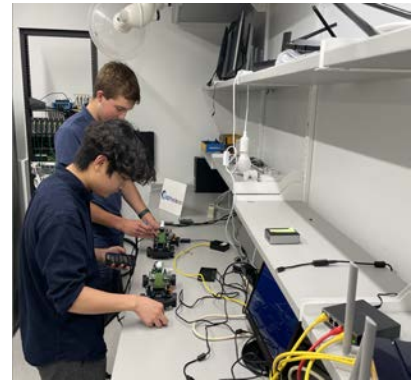
*Georgia Tech team showing off their HPL run at SC'22*

# AUTOLEARN: A CASE STUDY

**A collection of courselets exploring concepts in autonomous driving**

- **Contains three types of courselet layers:**
  - ◆ Data collection (actual car versus simulator)
  - ◆ Machine Learning courselets training models
  - ◆ Verification via self-driving (actual car versus simulator)
- **Supports different emphasis and different pathways through the curriculum:**
  - ◆ Introduction to engineering might emphasize driving the actual car
  - ◆ Machine learning focus might use the simulator
- **Contain suggestions for exercises and individual exploration:**

*Paper: “AutoLearn: Learning in the Edge to Cloud Continuum”, EduHPC 23*



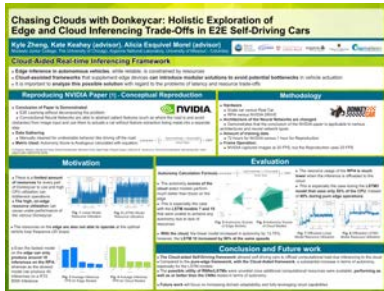
REU 2023 students working on hardware setup for autonomous vehicles

# REU STUDENT POSTERS - SC'23



## Road to Reliability: Optimizing Self-Driving Consistency with Real-Time Speed Data [2]

- ▶ How do we improve speed consistency in autonomous vehicles?
- ▶ Use real-time optical encoder speed data in self-driving model training



## Chasing Clouds with Donkeycar: Holistic Exploration of Edge and Cloud Inferencing Trade-Offs in E2E Self-Driving Cars [3]

- ▶ What are the trade-offs or offloading self-driving inferences to proximal cloud sources?
- ▶ Analyzing autonomy score and resource utilization when offloading to the NVIDIA triton server hosted on Chameleon cloud



[2] Road to Reliability: Optimizing Self-Driving Consistency with Real-Time Speed Data. William Fowler, Kate Keahey and Alicia Esquivel Morel. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC'23 Poster). IEEE Press. November 2023

[3] Chasing Clouds with Donkeycar: Holistic Exploration of Edge and Cloud Inferencing Trade-Offs in E2E Self-Driving Cars. Kyle Zheng, Kate Keahey and Alicia Esquivel Morel. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC'23 Poster). IEEE Press. November 2023

# SUMMARY

- ▶ Chameleon is a cloud that evolves as the scientific frontier advances
  - ▶ Bare metal reconfiguration, direct resource access
  - ▶ Diverse range of hardware including innovative offerings in disaggregated hardware
- ▶ Chameleon is an open platform for reproducible systems research
  - ▶ Support for practical reproducibility including **programmatic interface to the testbed via Jupyter**, an **experiment sharing hub** (Trove), and **Chameleon daypass** for access supporting reproducibility
- ▶ Not just a testbed, a community of users
  - ▶ 8,000+ users, 1,000+ projects, 623 publications (lower bound) and counting
  - ▶ New research constitutes as the most significant use of the testbed
  - ▶ New focus on education, pioneering new ways of digital content use



[www.chameleoncloud.org](http://www.chameleoncloud.org)  
g