P3-24: Scalable Systems

Mission-Critical Computing
NSF CENTER FOR SPACE, HIGH-PERFORMANCE, AND RESILIENT COMPUTING (SHREC)

SHREC Annual Workshop (SAW23-24)

University of Pittsburgh
BYU BRIGHAM YOUNG UNIVERSITY
VIRGINIA TECH.
UF UNIVERSITY of FLORIDA

Dr. Alan George
Mickle Chair Professor of ECE
University of Pittsburgh

Calvin Gealy
Jefferson Boothe
Tyler Garrett
David Herr
Yasser Morsy
Thomas Plunkett
Caden Schiffer [UG]
Research Students
University of Pittsburgh

January 17-18, 2024

Number of requested memberships ≥ 5

Goals, Motivations, & Challenges

Goals

- Evaluate **scalability** of apps and libraries across range of compute
- Benchmark realtime network protocols for **space deployment**
- Investigate **software solutions** for fault resilience in GPU systems

Motivations

- Distributed computing enables **exascale** solutions
- Competing **network protocols** have tradeoffs
- GPU accelerators employed by many systems, but increasingly need additional **fault-mitigation** strategies

Challenges

- Computations spread across nodes require **complex orchestration**
- Performance demands and **security requirements** of networked systems are continually increasing
- Benchmarking large apps is time-consuming and **resource-intensive**

Proposed Tasks for 2024

T1 — Hyperscale Processing

Study large-scale apps on distributed CPU and GPU clusters

T2 — Space Network Protocols

Investigate protocols for reliable and efficient space networking

T3 — Fault-Tolerant GPU Computing

Improve reliability of apps deployed to space and HPC systems

T4 — Deep-Learning Metrics and Performance

Characterize DL-model performance using metrics

T1 Hyperscale Processing

Study large-scale apps on distributed CPU and GPU clusters

Task Leader: Jefferson Boothe

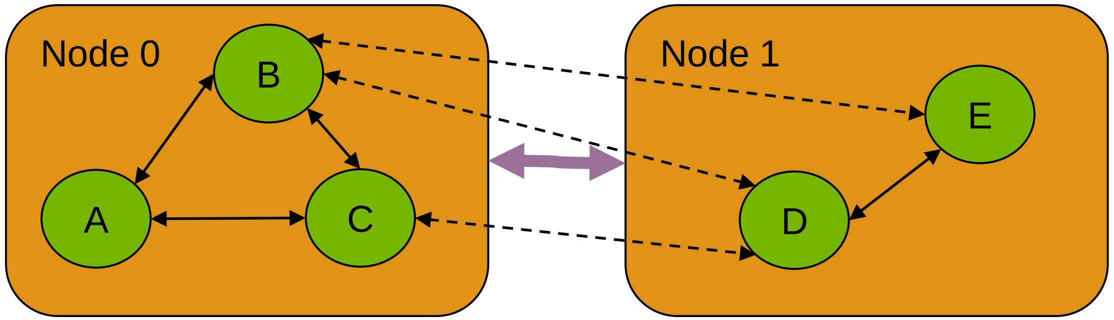
Jefferson Boothe – j.boothe@pitt.edu
Yasser Morsy – yasser.morsy@pitt.edu

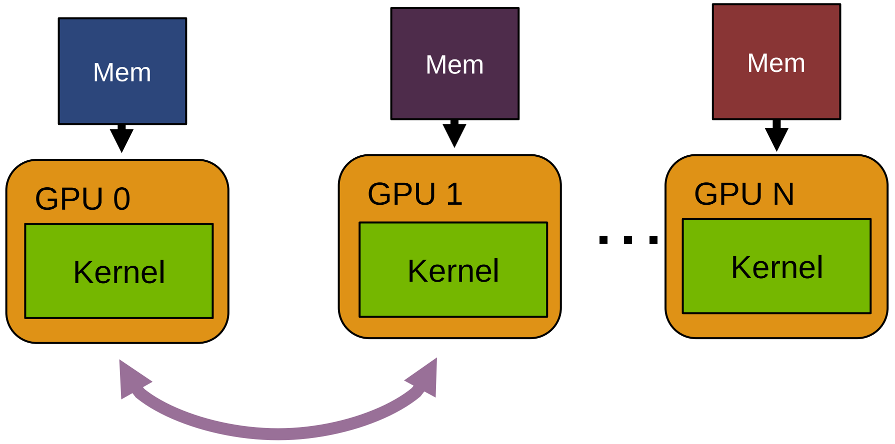
T1: Background

P3

Graph Analysis Apps

- Graphs computationally challenging due to **poor memory locality**
- BFS is used in **Graph500** to test supercomputers on data-intensive apps
- **Irregular communication** patterns can stress parallel programming libraries



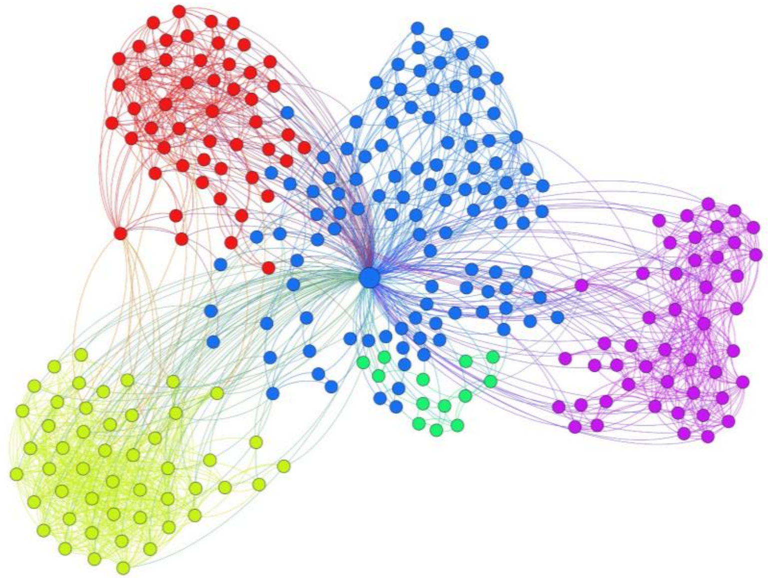


Distributed GPU Computing

- NVSHMEM enables **direct communication** between GPUs using **SHMEM standards**
- UVM P2P enables **direct communication** between GPUs using **non-symmetric memory mapping**

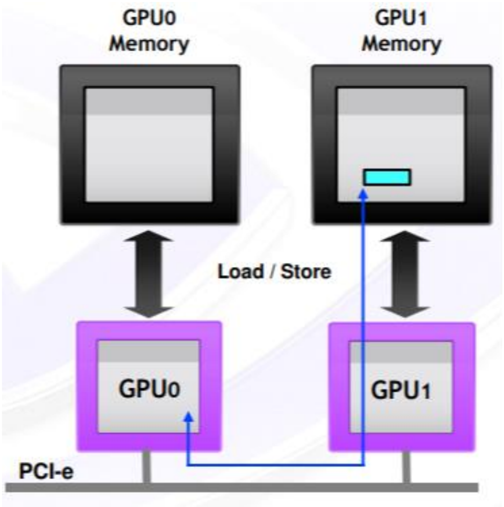
BFS: Breadth-first search
UVM: Unified Virtual Memory
P2P: Peer to Peer

T1: Hyperscale Processing



Graph Analysis with SHMEM

- Extend benchmarking of parallel programming libraries with graph analytic applications on **distributed systems**
- Analyze **scalability** across system configurations
- Utilize Pitt CRC and PSC **Bridges-2** Resources



NVSHMEM and UVM Analysis

- Explore novel **GPU-GPU communication** methods to accelerate numerical linear algebra algorithms
- Characterize behavior on **varying scales** and devices such as NVIDIA HGX A100 at Pitt CRC

Pitt CRC: Center for Research Computing
PSC: Pittsburgh Supercomputing Center
P2P: Peer to Peer
UVM: Unified Virtual Memory

T2 Space Network Protocols

Investigate protocols for reliable and efficient space networking

Task Leader: David Herr

David Herr – david.herr@pitt.edu

T2: Background

Time-Sensitive Networks

- Standard networking protocols are **ineffective** at meeting needs of aerospace apps
- **Time-sensitive protocols** aim to meet these needs of determinism, reliability, and security
- Need **additional comparisons** of these protocols

V&V of Simulation Results

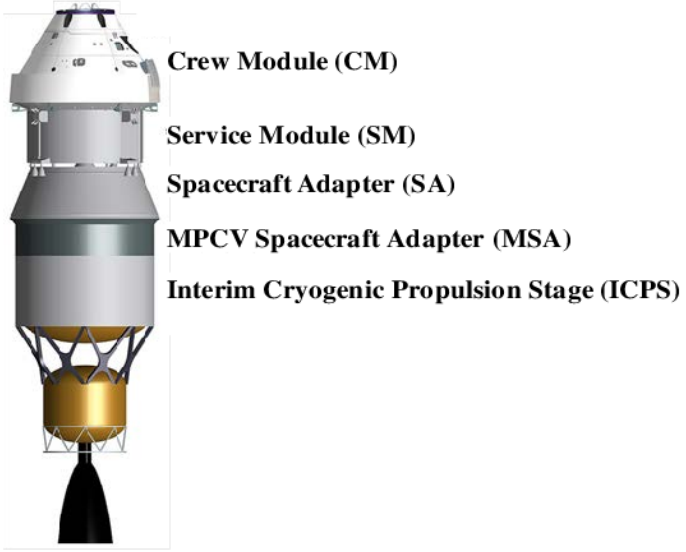
- Simulation of time-sensitive protocols can significantly **reduce costs** and **time** of analysis and comparison
- **Verification and Validation** are crucial to ensuring results are **realistic** to real-world designs
- TTE hardware is **available** for purchase, but open source TSN Aerospace hardware is **still in development**

TSN: Time-Sensitive Networking
TTE: Time-Triggered Ethernet
V&V: Verification and Validation

T2: Space Network Protocols

TSN Protocol Comparison

- Extend simulation benchmarking to include more use cases, such as **Orion (MPCV)** avionics system
- Compare simulation results to equivalent **hardware** setups using TTE testbeds within SHREC Lab



TSN Aerospace V&V

- Investigate **hardware testbed,** like TSN product line from UEI, to compare TSN **simulation results** from OMNeT++ to real hardware
- Compare results from **real hardware** against that of TTE to **deepen comparison** results

TSN: Time-Sensitive Networking
TTE: Time-Triggered Ethernet
MPCV: Mult-Purpose Crew Vehicle
V&V: Verification and Validation
UEI: United Eletronic Industries

T3 Fault-Tolerant GPU Computing

Improve reliability of apps deployed to space and HPC systems

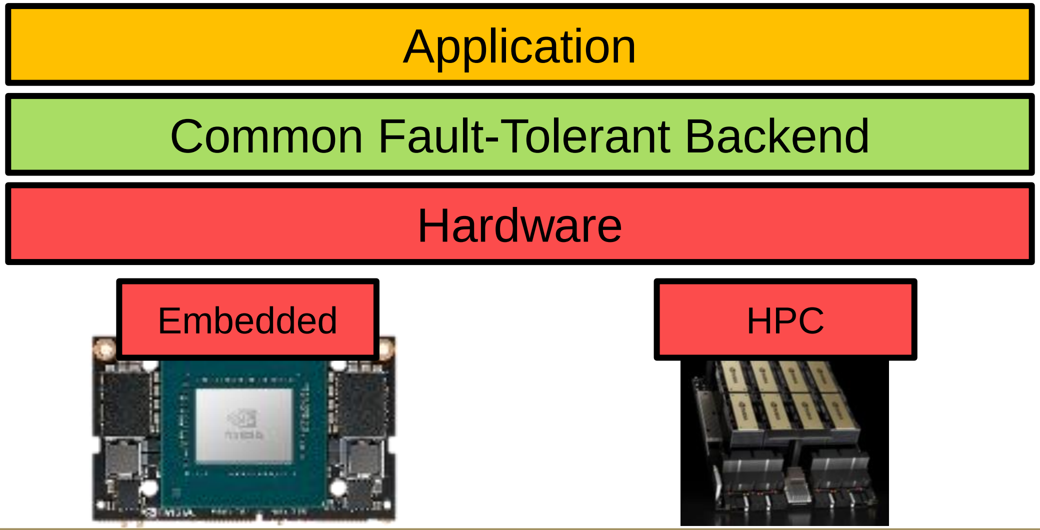
Task Leader: Tyler Garrett

Tyler Garrett – tmg61@pitt.edu
Thomas Plunkett – thp49@pitt.edu

T3: Background

GPU Reliability

- GPU computing enables high-performance apps for both **space** and **HPC**
- Each domain is **vulnerable** to high error rates due to either radiation or wear out
- Cost of **inaccurate** computation can range from high energy consumption to mission failures



Fault Mitigation

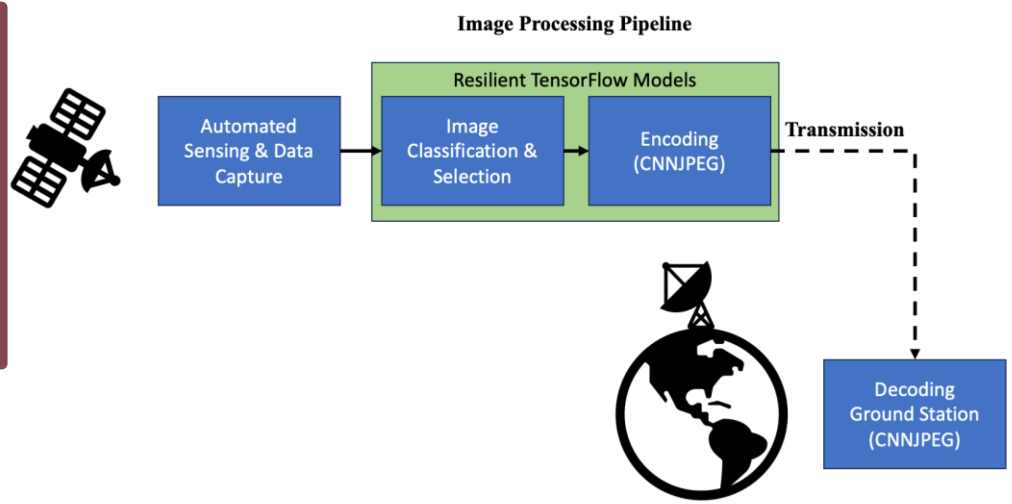
- Similar **fault-mitigation** techniques can apply to embedded systems and supercomputers
- On-chip error detection, isolation, and correction leverage **software-based** solutions
- Using GPU microarchitecture to intelligently map apps enables increased **reliability**

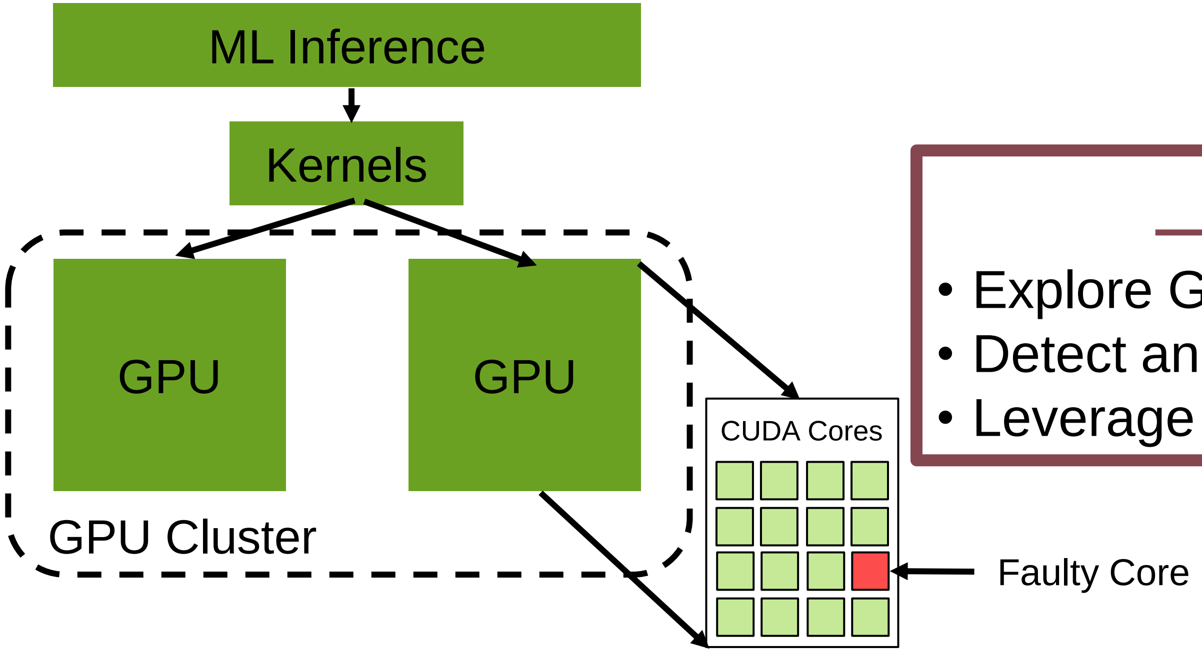
T3: Fault-Tolerant GPU Computing

P3

Reliable ML Apps for Space Missions

- Investigate **soft-error mitigation** on ML models and data collected from STP-H7
- Optimize **GPU-accelerated** ML apps targeting STP-H12 imaging pipeline using RTF





Reliable ML Apps for HPC

- Explore GPU **permanent faults** in HPC ML apps
- Detect and isolate **faulty cores** in HPC systems
- Leverage **custom kernels** to analyze GPU hardware

RTF: Resilient TensorFlow

T4 Deep-Learning Metrics and Performance

Characterize DL-model performance using metrics

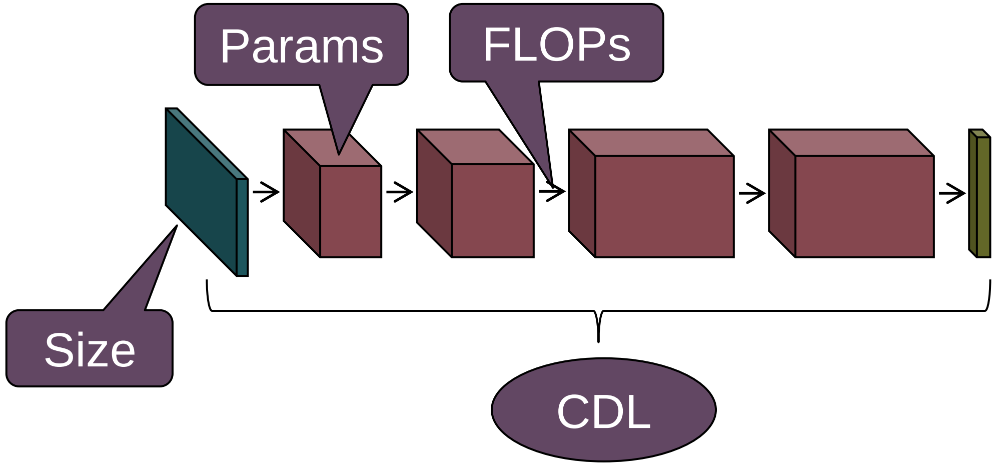
Task Leader: Calvin Gealy

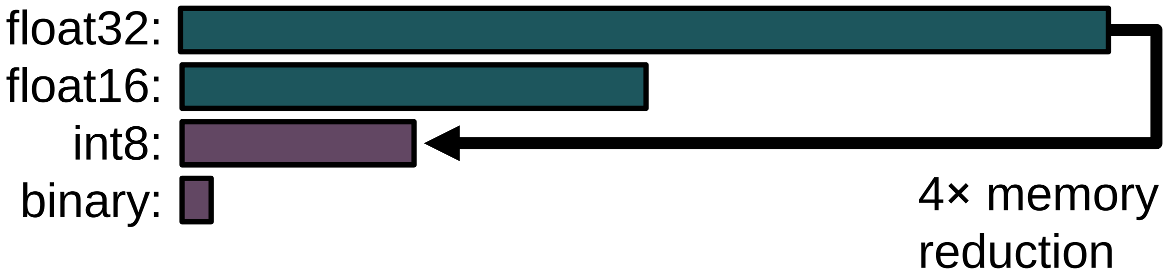
DL: Deep Learning

T4: Background

SPOC Metrics

- SPOC are proposed set of metrics that aim to better represent model performance **on specific hardware**
- Need analysis on how to best make **actionable decision** with these metrics
- Previous SHREC research indicates SPOC helps explain **model performance**





Quantization Metrics

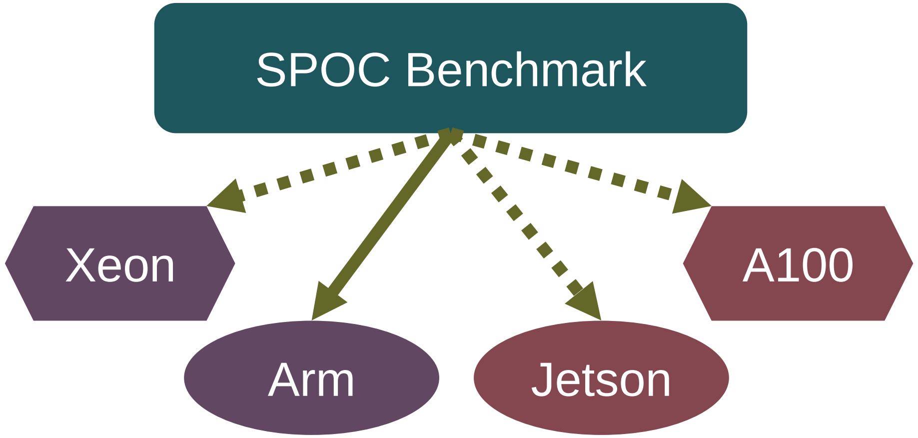
- Quantization often improves inference performance of models by **reducing precision** of operations
- Previous research suggests **varying effects** depending on model type and target device

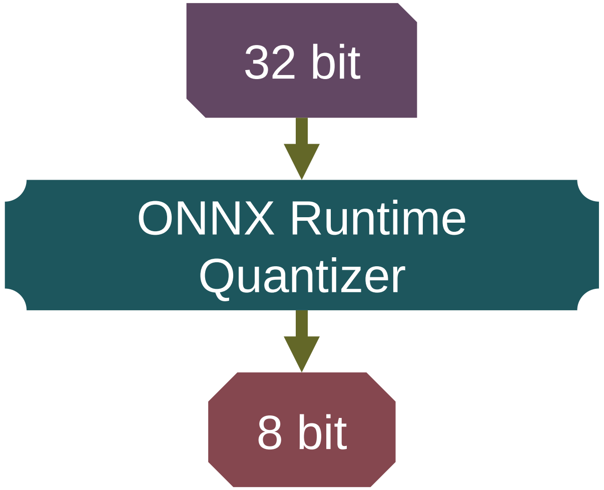
SPOC: Size, Param, Ops, Critical Datapath Length (CDL)

T4: DL Metrics and Performance

Expanding SPOC Metrics to GPUs

- Extend benchmarking of **SPOC metrics** with analysis on embedded and desktop GPU systems
- Characterize metric **scalability** using micro models
- Analyze **latency**, **memory**, and **caching behavior**





Quantization Metrics

- Develop **new metric** to capture effect of quantization on model performance
- Benchmark using previous SPOC testing suite on **embedded** and **desktop** CPU and GPU systems

DL: Deep learning
SPOC: Size, Param, Ops, Critical Datapath Length (CDL)

Milestones, Deliverables, & Budget

MILESTONES

- SMW24 (Jun/Jul 24): Showcase preliminary results on all project tasks
- SAW24-25 (Jan 25): Completion of all project tasks

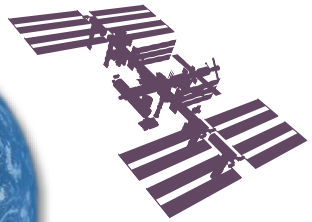
DELIVERABLES

- Monthly progress reports from all projects
- Midyear and end-of-year full reports from all projects
- 4-5 conference or journal publications

BUDGET

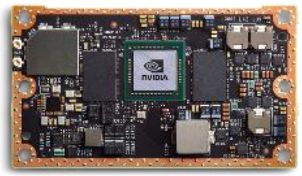
- Minimum recommended: Five (5) memberships

  (250 Votes)



BRIDGES

A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE



Conclusions & Member Benefits

Conclusions

- Analyze scalability of SHMEM and NVSHMEM communication libraries using **high-performance** apps and algorithms
- Investigate simulation of **realtime space protocols** for network performance and potential security challenges
- Evaluate GPU faults for space and HPC systems and investigate software solutions for **fault mitigation**
- Characterize scalability of DL models by using **SPOC metrics** as guidebook to performance

Member Benefits

- Direct influence over processors and frameworks studied
- Direct influence over apps and datasets studied
- Direct benefit from new methods, data, code, models, and insights from metrics, benchmarks, and emulations

DL: Deep learning
NVSHMEM: Nvidia SHMEM
SHMEM: Shared Hierarchical Memory
SPOC: Size, Param, Ops, Critical Datapath Length (CDL)