



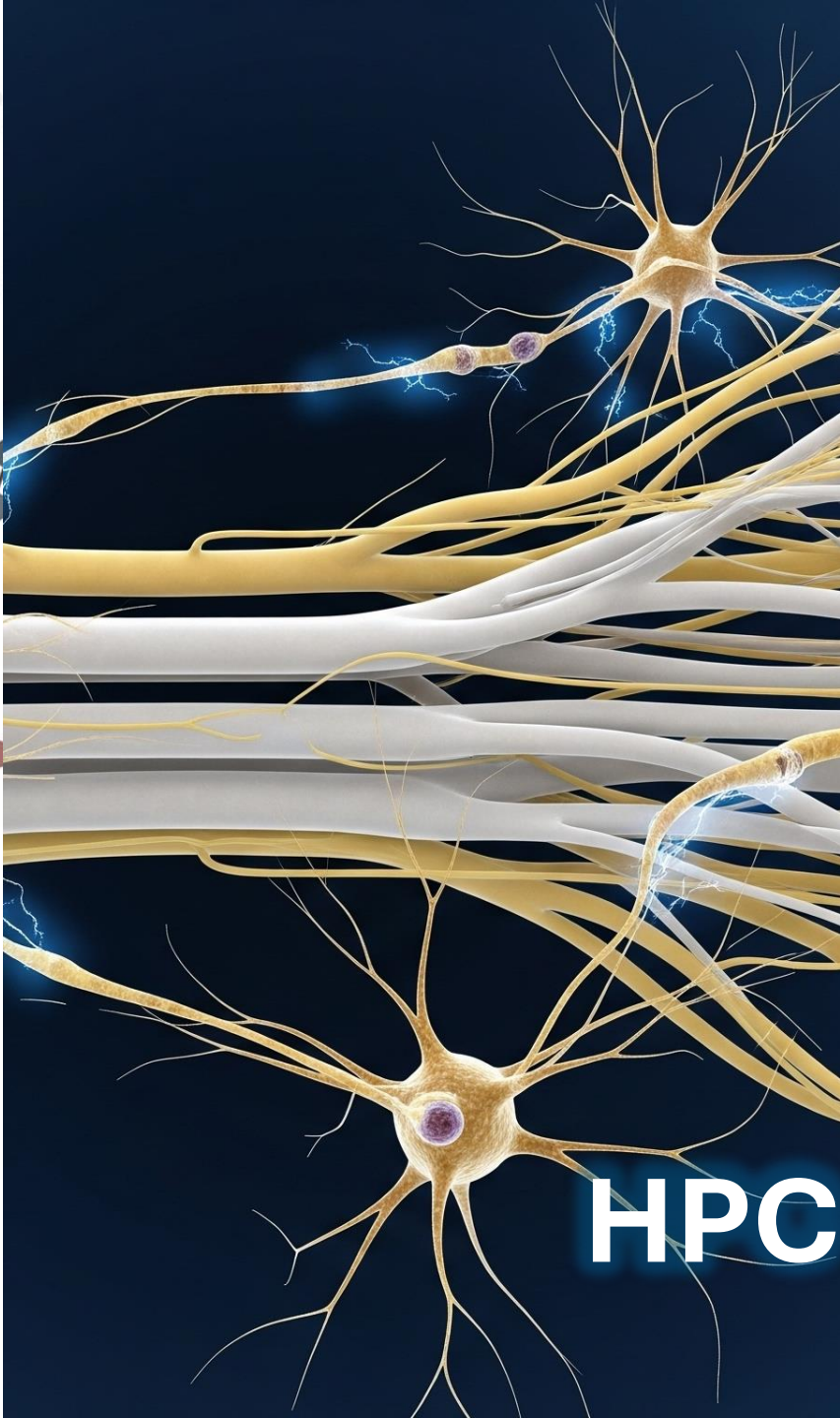
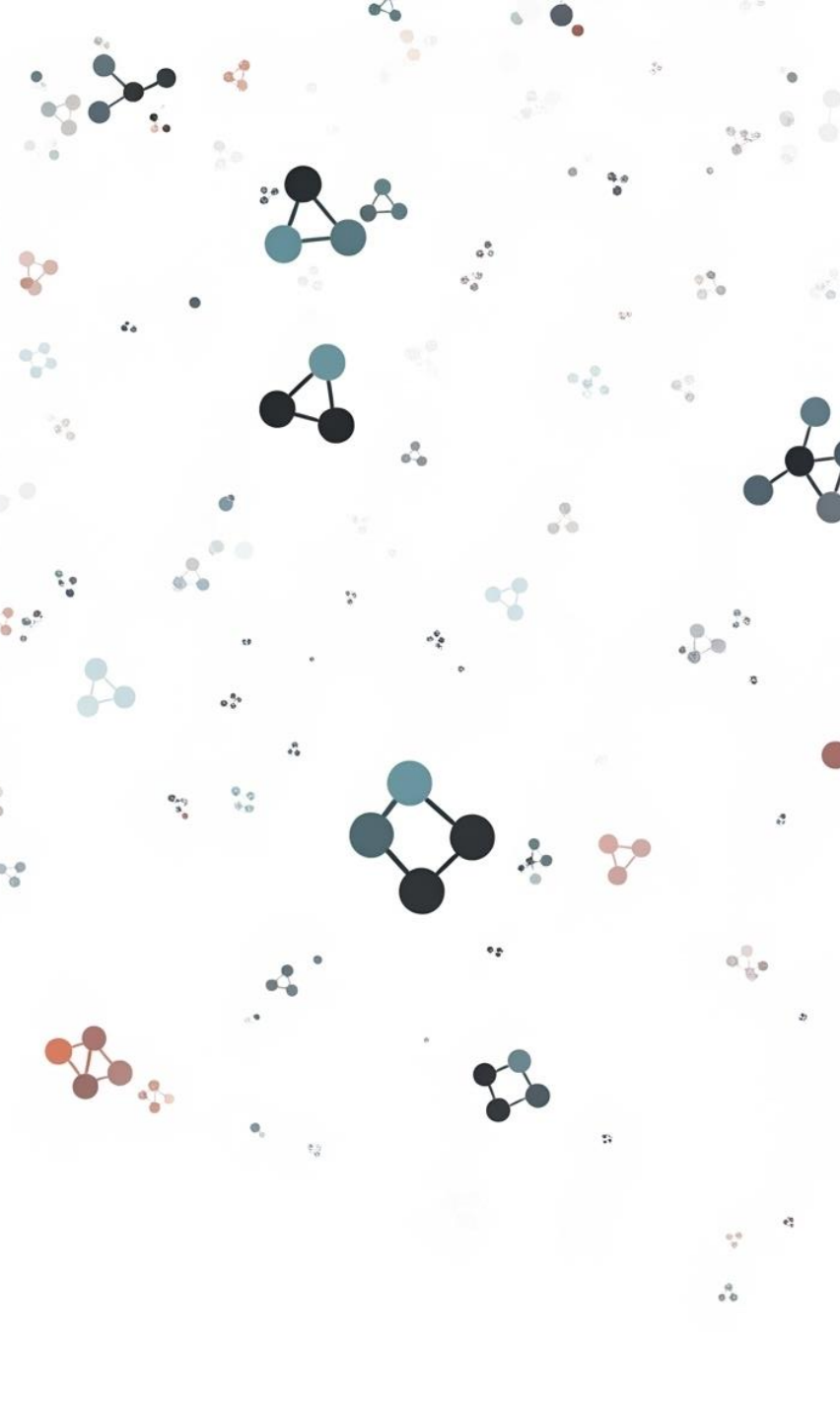
HPC Everyday & Everywhere

Kevin A. Brown, PhD

Assistant Computer Scientist
Argonne National Laboratory
kabrown@anl.gov

HPC Everyday





HPC Everywhere

Kevin A. Brown

Researcher?

*Research is my medium,
not my passion.*

Where my passion lies:

- Event Planning
- Program Development
- Community Building



Department of Entertainment (DOE)
Fun Opportunity Announcement (FOA)

Halloween Costume Party!

FOA Number: HW2024.10.25-B240

Application Deadline (party date):
October 25, 2024 at 4:30pm - 6:00pm

*Follow-on fun (after party) is available
starting 6:30pm @ ReNEW Westmont*

Point of Contact (POC): Kevin B.

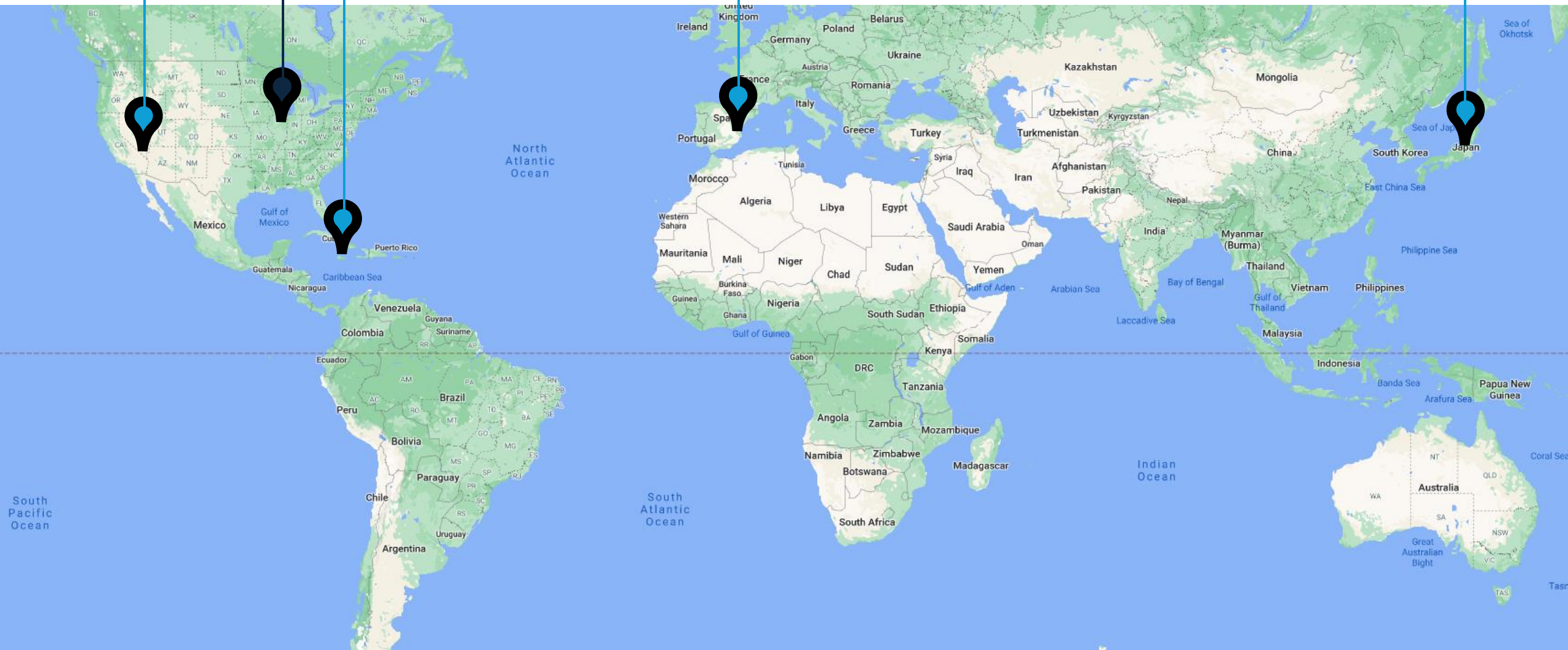
**Argonne
2019**

**Lawrence
Livermore
National Lab**

**Digicel, Jamaica
[mobile network co.]**

**Barcelona
Supercomputing Center
(BSC)**

**Tokyo Institute of Technology
(Tokyo Tech)**







My Background

No Job Offer – PNNL, USA | English Teacher, Japan | Others

MANY Rejected Papers – Various conferences

Not Selected – Prime Minister Youth Award, Jamaica

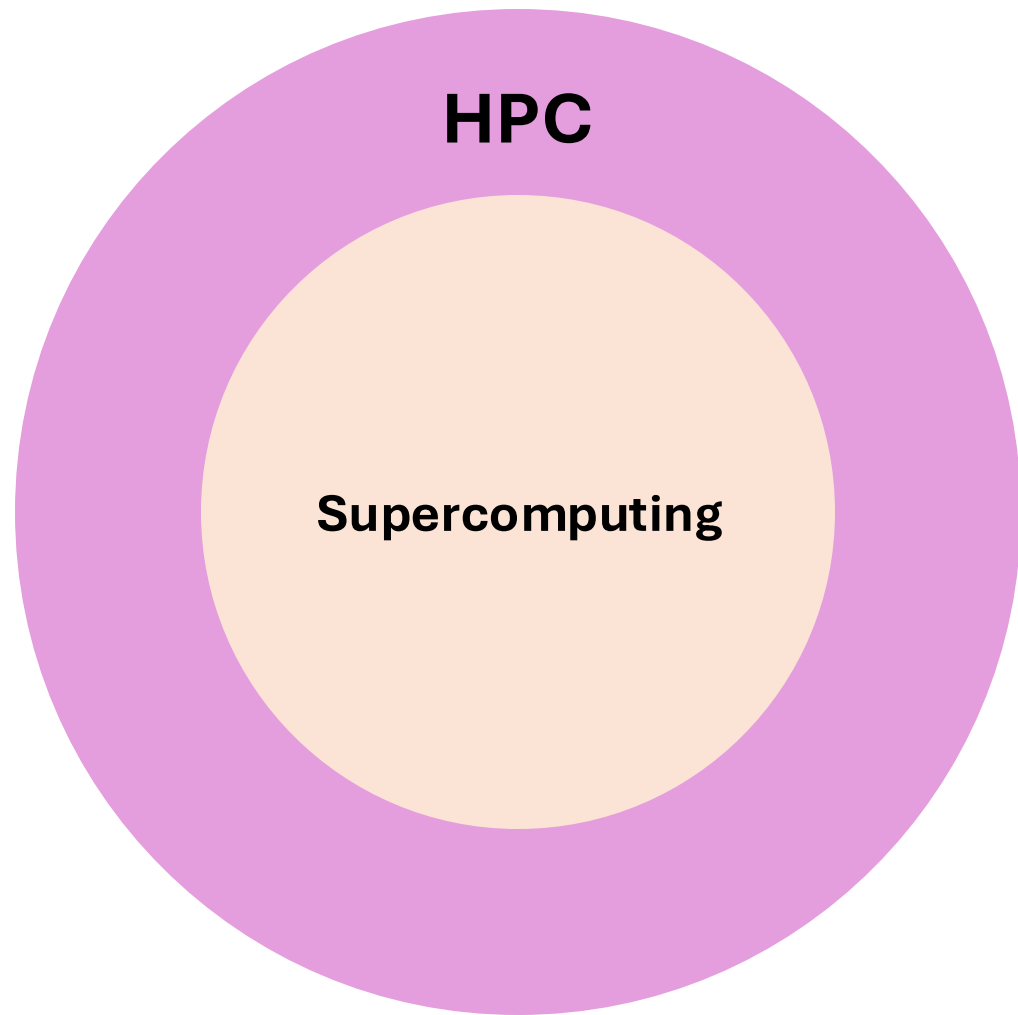
No Scholarship for 1st Year Undergrad – Jamaica

Lost Scholarship in Final Year Undergrad – Jamaica

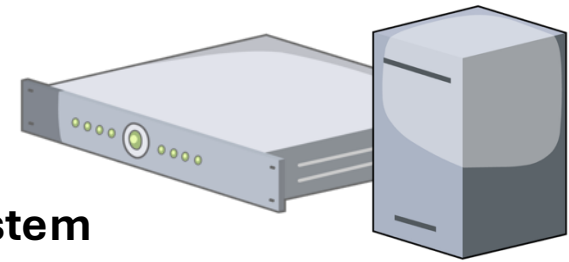
Lost Friendships & Relationships – Jamaica, USA, Japan

Lost Election – Student Union President, University of Technology, Jamaica

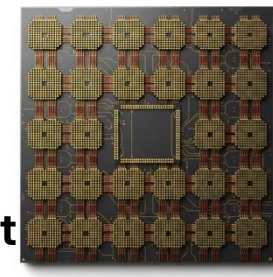
Major Research Mistakes – Too many to count



Facility



System



**Sub-
component**

Infrastructure



Application

```
# modsim

# Fluid Events - Changing rates
import matplotlib.pyplot as plt
from matplotlib import style

plt.style.use('default')

#tdf = tdf[tdf['x'] > 0]

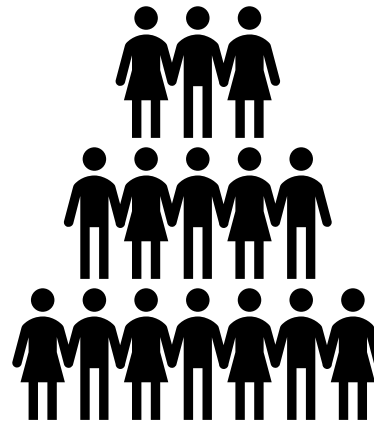
x0 = tdf[tdf['x'] > 0]
y0 = tdf[tdf['x'] > 0]

x1 = tdf[tdf['x'] > 0]
y1 = tdf[tdf['x'] > 0]

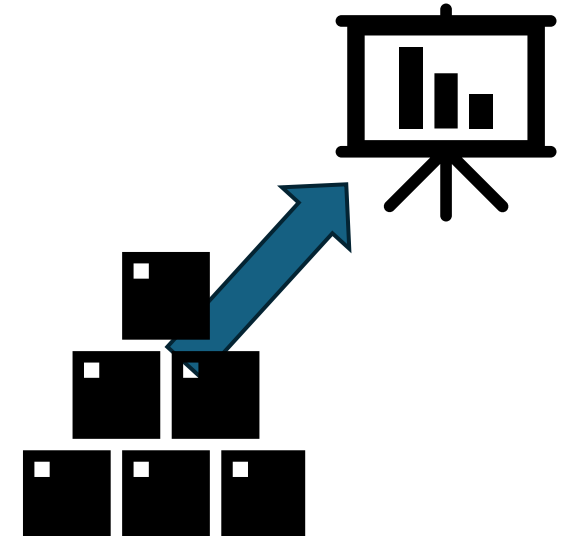
x2 = tdf[tdf['x'] > 0]
y2 = tdf[tdf['x'] > 0]

flows = {0: [], 1: [], 2: []}
for key in change_schedule.keys():
    for k, v in change_schedule[key].items():
        if v == 0:
            v = 0.001
            tmp = (v/25)*100
            myrate = base_rate * (100/tmp)
            flows[k].append((key, round(myrate, 6)))

print(flows)
```



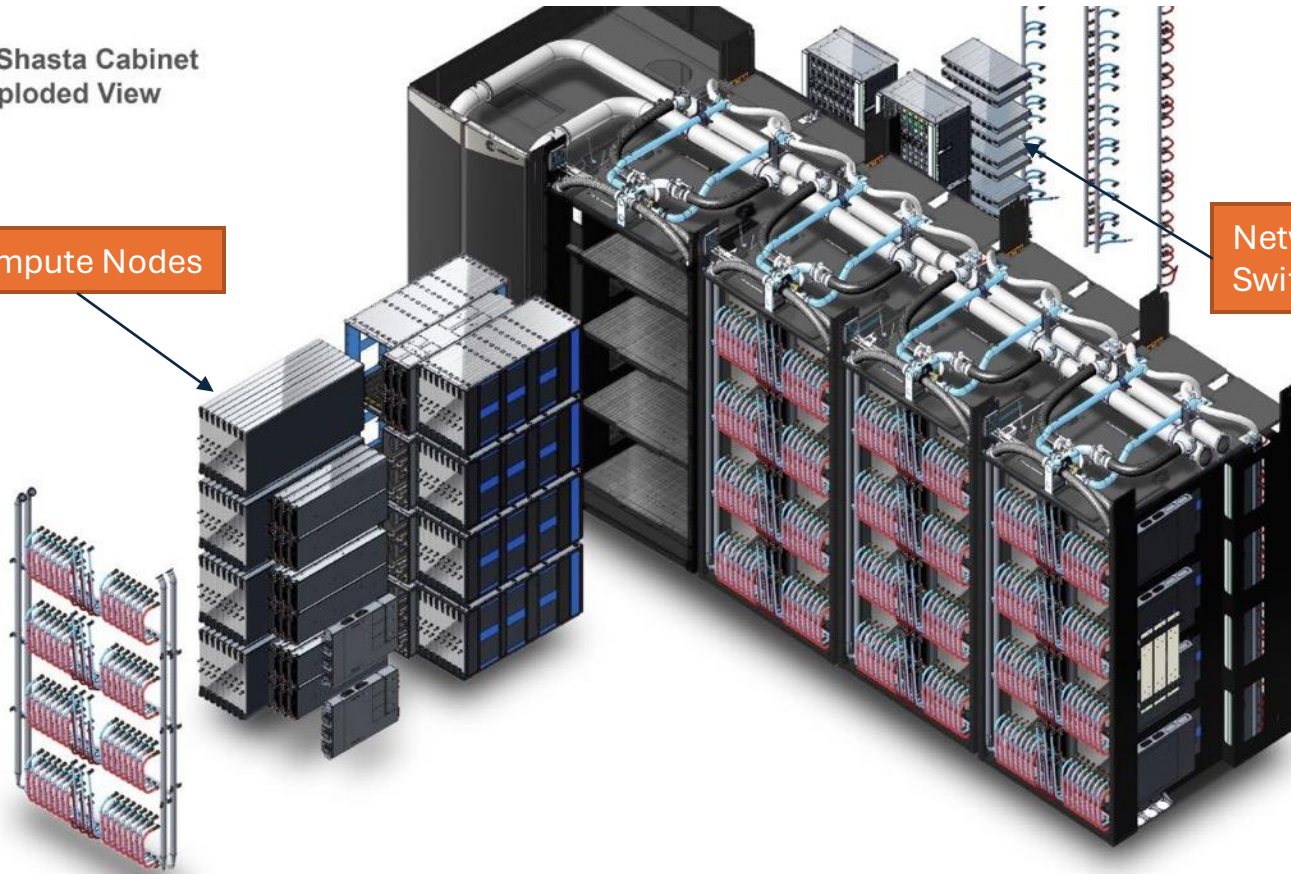
Goal



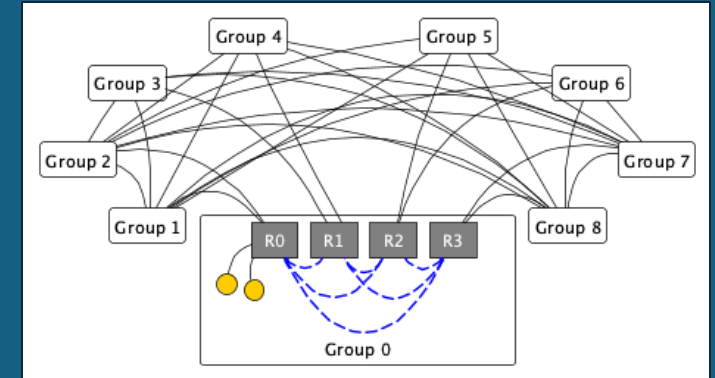
Cray Shasta Cabinet Exploded View

Compute Nodes

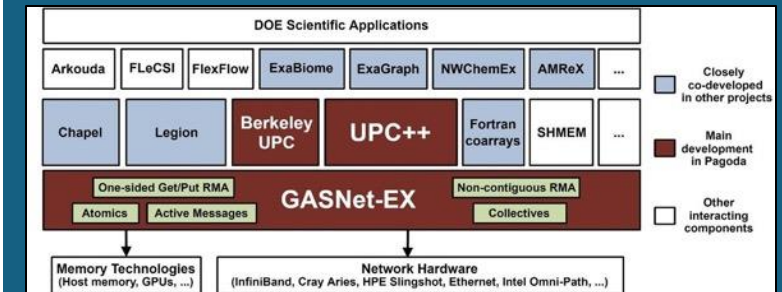
Network Switches



Network Topology View



Software Stack View

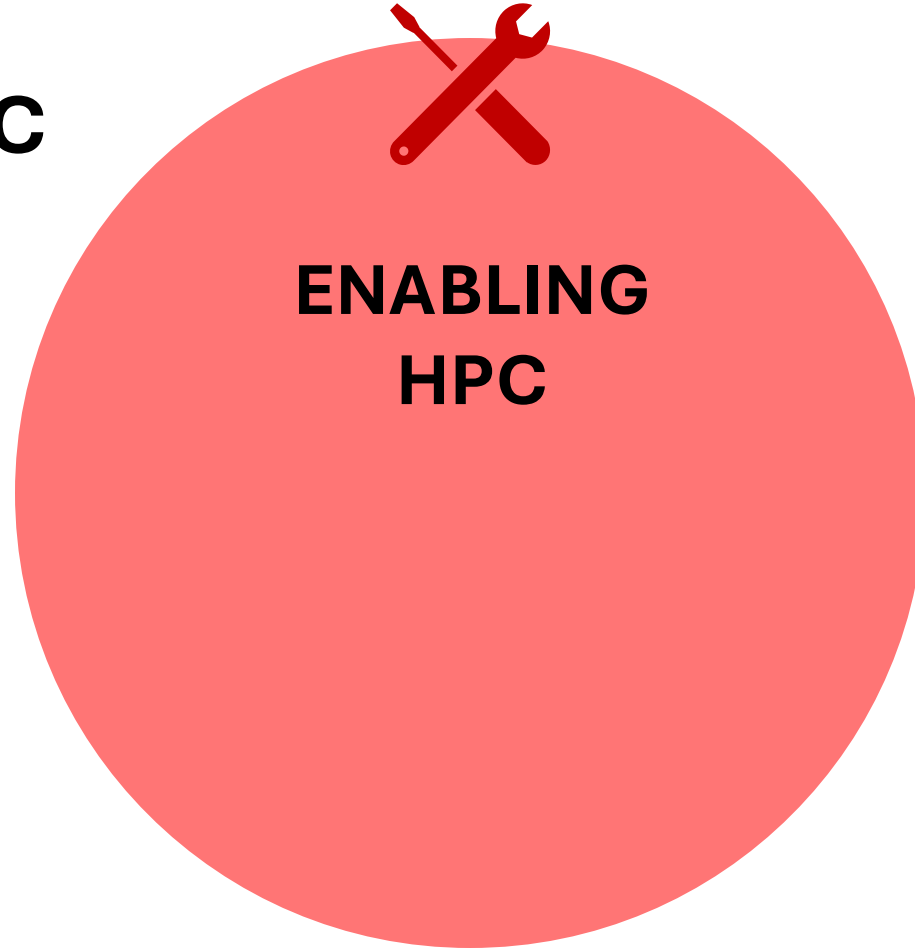


...

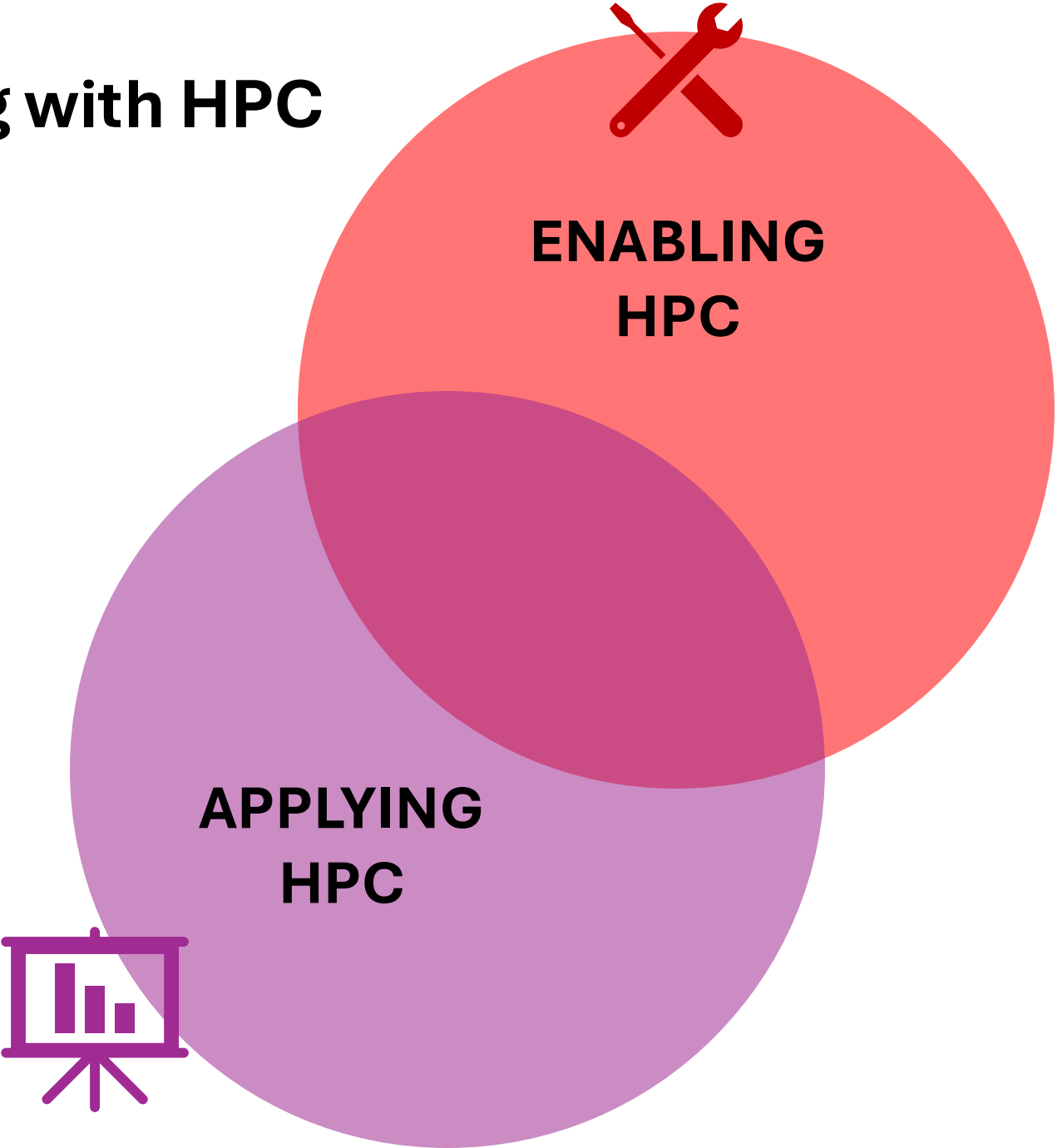
Engaging with HPC



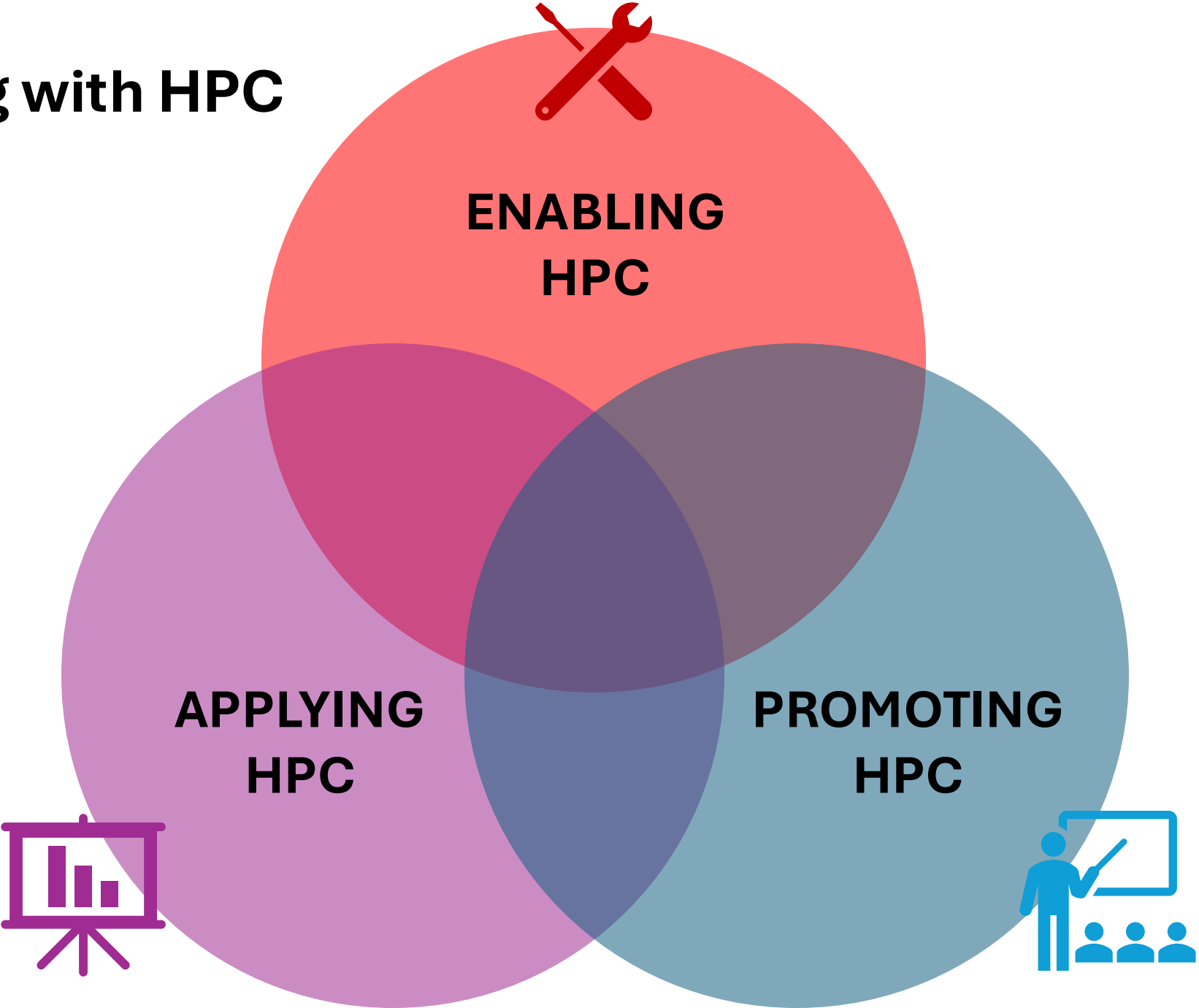
**ENABLING
HPC**



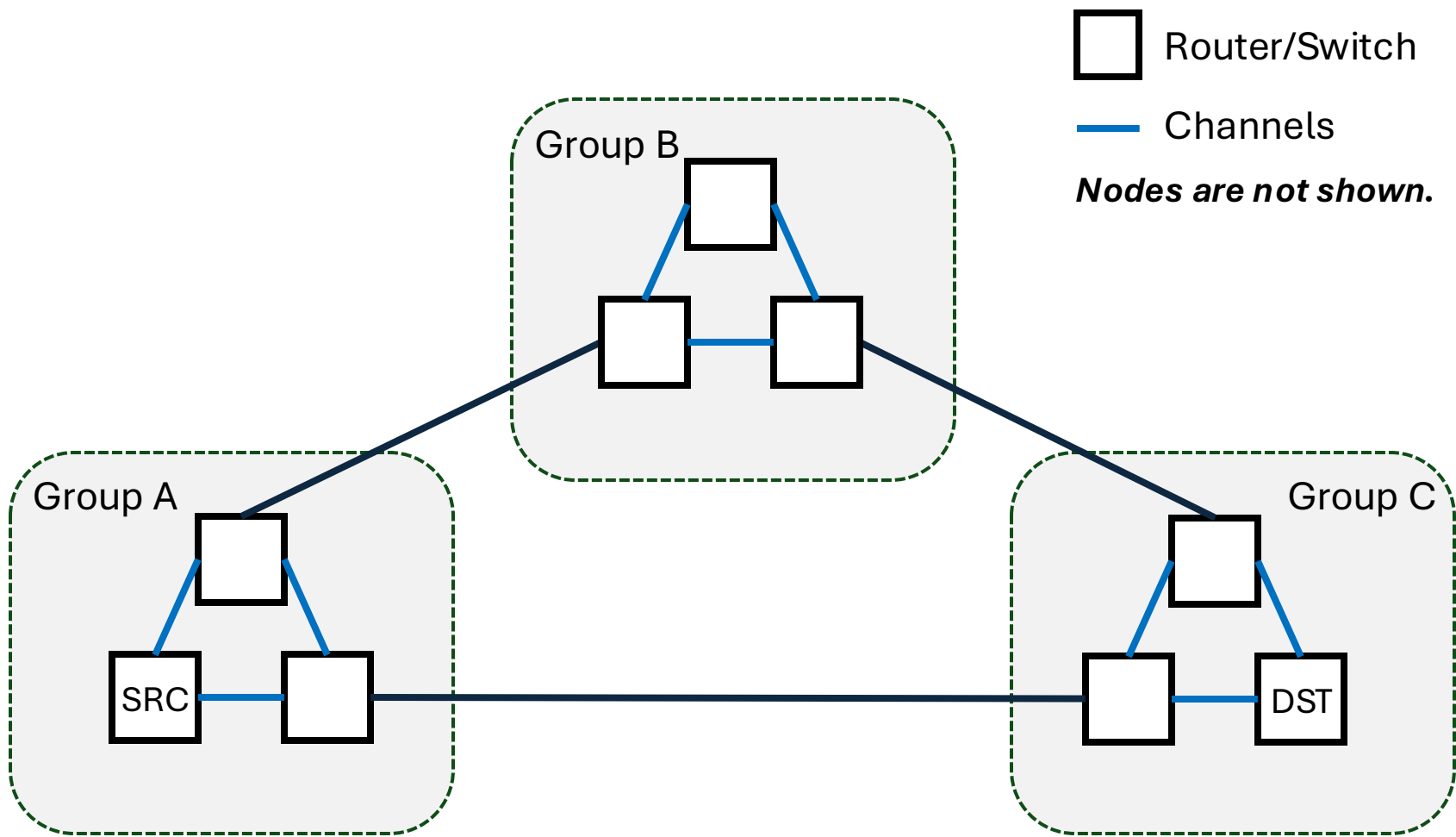
Engaging with HPC



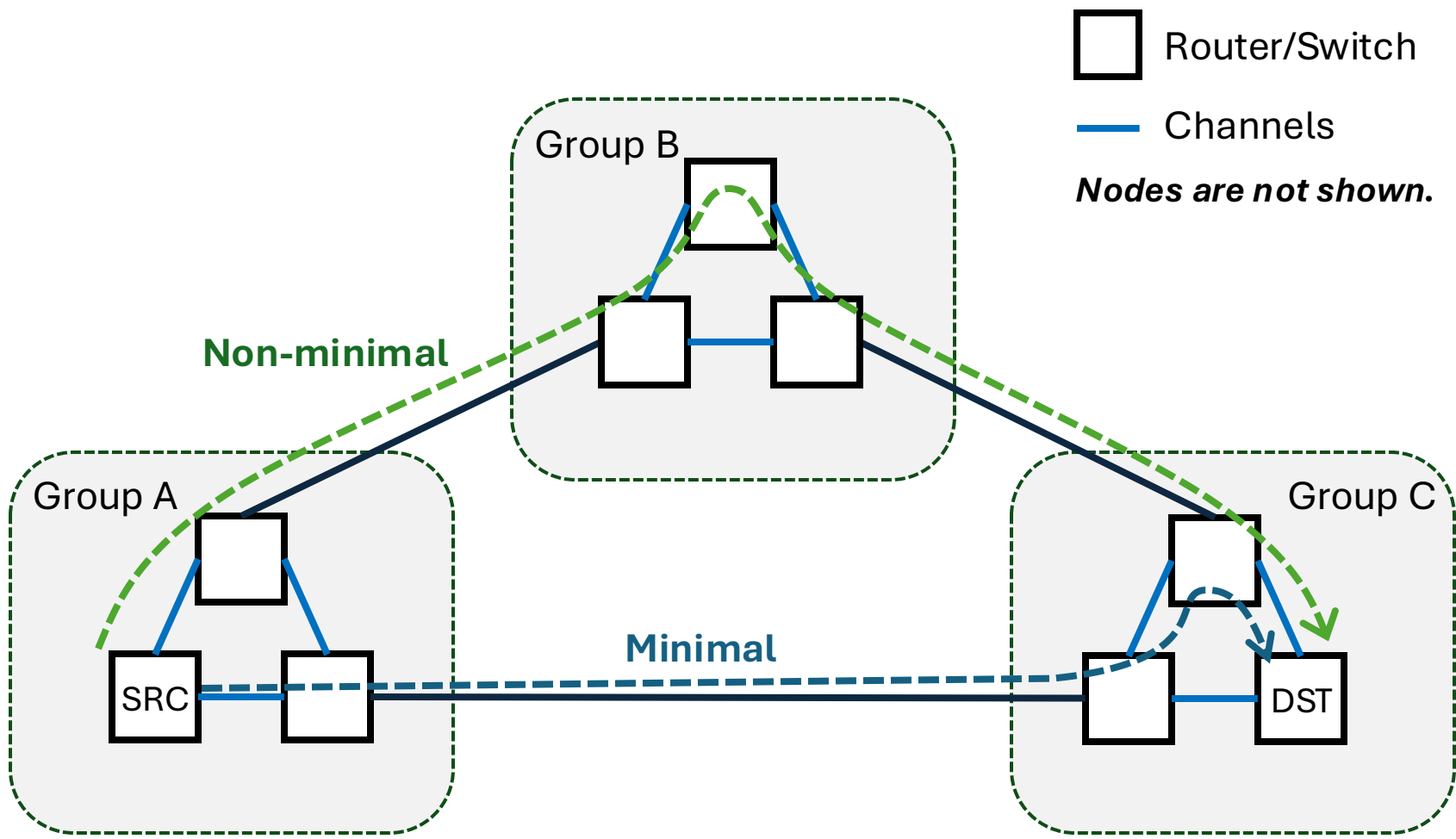
Engaging with HPC



Network Topology

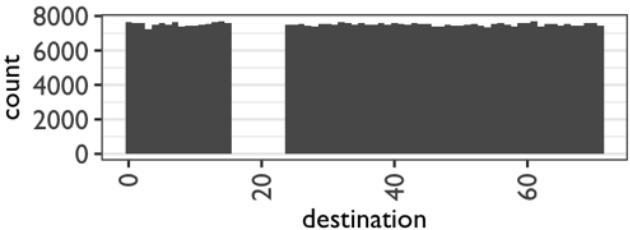


Network Topology



Performance Analysis

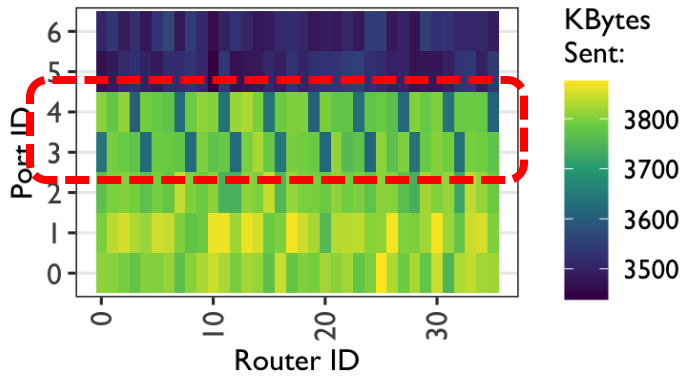
Count Messages Sent



Balanced distribution of packets



Total Traffic Per Switch Port



Unexpected traffic pattern across global ports



Understanding HPC Network Behavior Using Low-level Metrics

Kevin A. Brown and Robert B. Ross
Mathematics and Computer Science Division, Argonne National Laboratory
kabrown@anl.gov

ABSTRACT

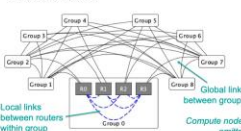
Supercomputer systems are complex interconnections of nodes supported by advanced networking. However, creating accurate simulation models is an error-prone process that requires accurately modeling real-world capabilities without full, complex technical re-implementations.

To better evaluate network model activities, we use a performance refinement process to drill down into routing anomalies present in the dragonfly network model of the CODES simulation toolkit [1]. This refinement process starts with viewing high-level performance data and then iteratively uses lower-level, component-specific data to locate the source of the anomaly.

We identify new low-level routing metrics that can expose the behavior of network model and attach these metrics to packets flowing through the system. Using these per-packet, per-hop routing metrics, we effectively attributed anomalies in system-wide traffic distribution to the routing algorithm's interaction with network connectivity configuration.

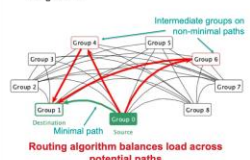
DRAGONFLY NETWORKS

- The 1D Dragonfly [2] is hierarchical network topology used by supercomputers
 - Each router is connected to a set of compute nodes
 - Routers are interconnected within logical groups via "local" links
 - Groups are then interconnected via "global" links
- There are multiple paths between source and destination nodes



ADAPTIVE ROUTING

- Progressive adaptive routing (PAR) [3] attempts to use the fastest path between source and destination routers
 - The congestion at potential paths is evaluated at each stop along the packet's journey
 - The packet is re-routed along alternative non-minimal paths when the original paths are more congested.



EXTENDED ANALYSIS OF ROUTING BEHAVIOR ON DRAGONFLY NETWORK

Environment Setup	72-nodes, 9-group, 36-routers 1D Dragonfly network 2 nodes/router 4 routers/group 25Gb/s link bandwidth
Traffic	Uniform Global Random Traffic 100% injection load, all nodes involved
Simulation	Simulated by the CODES/ROSS simulation toolkit [1] Metrics are recorded every 25 microsecond interval

Typical High-level Analysis

- We use a uniform global random traffic pattern to evaluate PAR routing in CODES
 - Traffic from each node is sent uniformly to all global destinations

Count Packets Sent by Nodes in Group 2

Count of Packets Using Each Intermediate Group [Source Group = 2]

Incorporating Component-specific Analysis

- The PAR algorithm decides which port is used by a packet leaving each router
 - Ports may become blocked due to congestion and cannot send data
- Online port-specific metrics show how well traffic load is balancing across ports

Total Traffic Per Router Port

Global Port Traffic For Router in Group 2

Incorporating Per-packet Analysis and Visualization

- The intermediate group taken by a packet is recorded in each packet along its journey
 - This marks the non-min. path used
- The load on non-minimal paths is confirmed to be unbalanced by viewing the path distributions

Illustration of Network Connectivity Highlighting Select Non-minimal and Minimal Paths

FINDINGS

- Port- and packet-specific metrics expose that a non-minimal path is rarely taken when it shares the source router with a minimal path
 - The basic PAR algorithm in CODES compares only the loads on min. and non-min. paths
- The 2x minimal path bias will cause the non-min. routing to appear more loaded

CONCLUSION

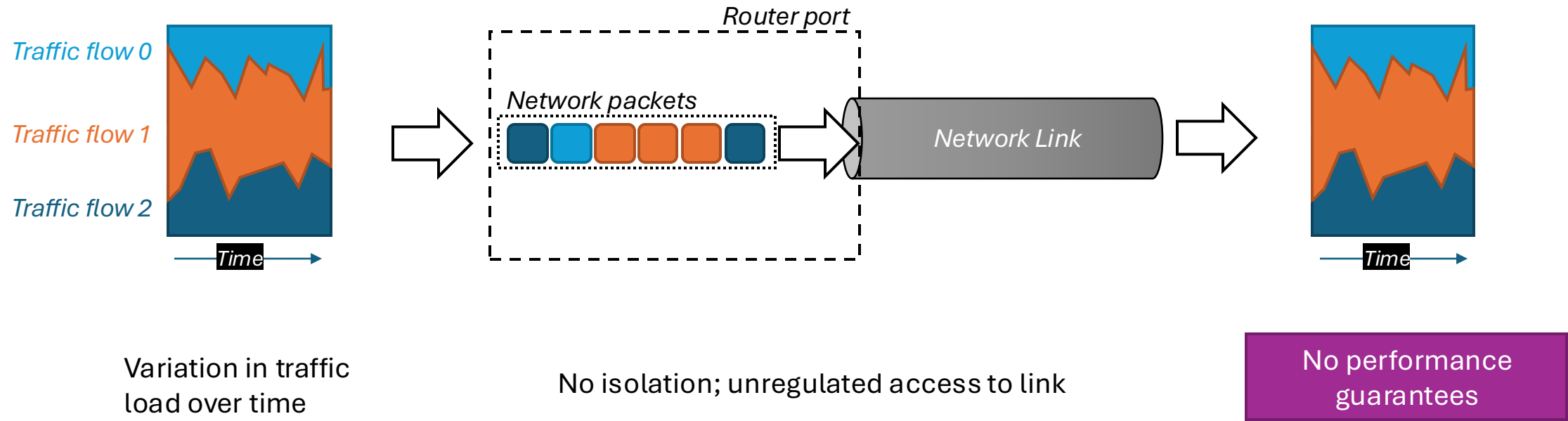
- Taken alone, high-level metrics may not show anomalies in routing behavior
- Iteratively refining the analysis approach can systematically expose anomalies
 - Component- and packet-specific measurements should be incorporated in the analysis
- Low-metric metrics requires storing large volumes of data ~1ms of activity can generate GBs of data

NEXT STEPS

- Evaluate larger scale network behaviors
- Incorporate in-situ analytics framework for online streaming of data from the simulation
- Optimize the adaptive routing algorithm in CODES

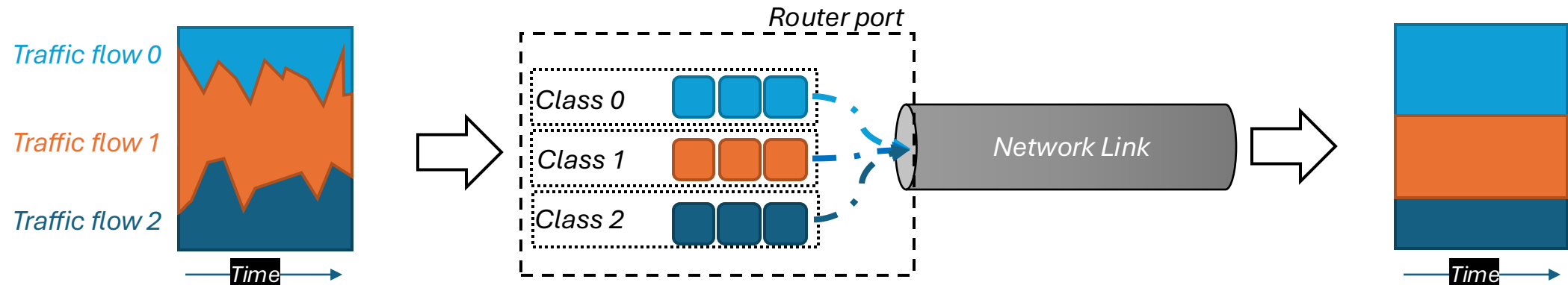
Quality of Service

Network without QoS



Quality of Service

Network with QoS

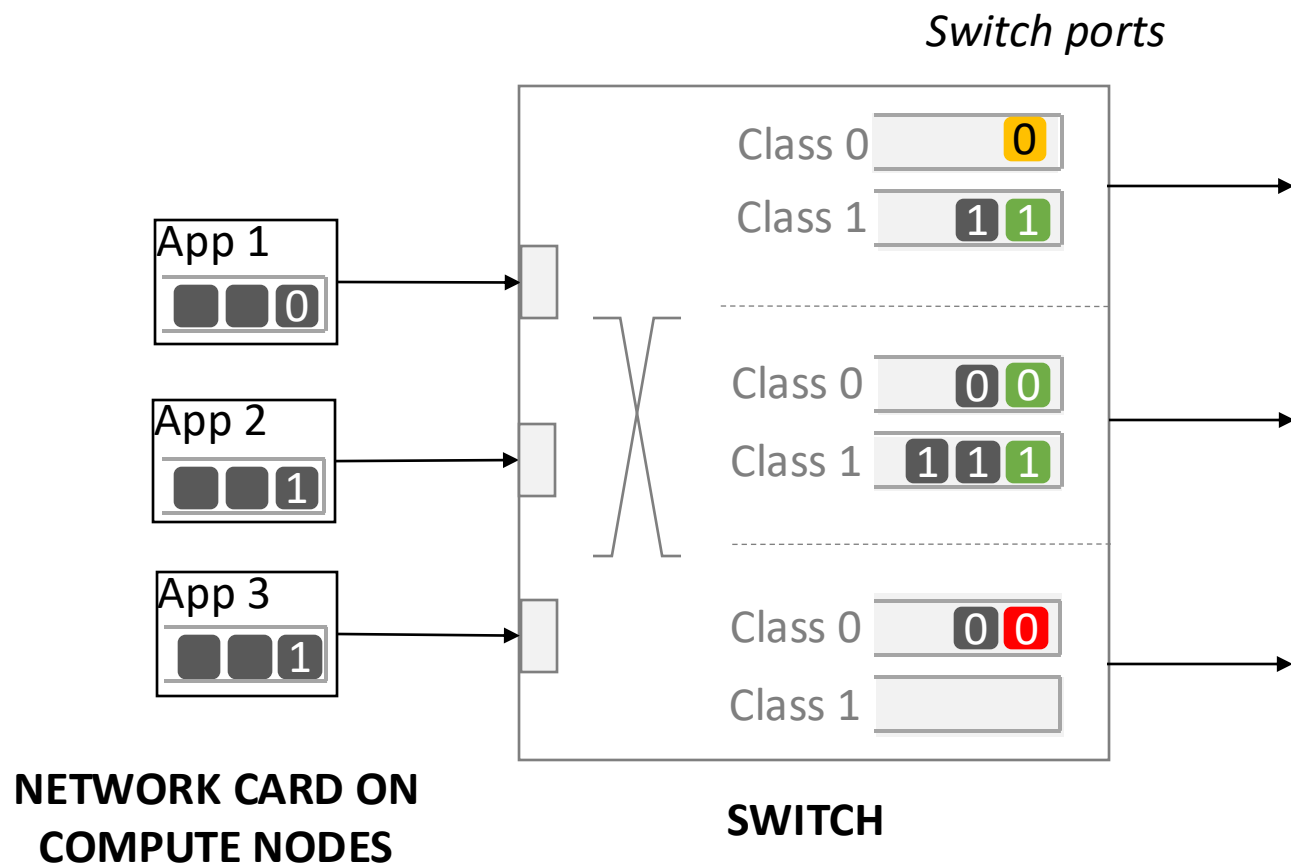


Variation in traffic load over time

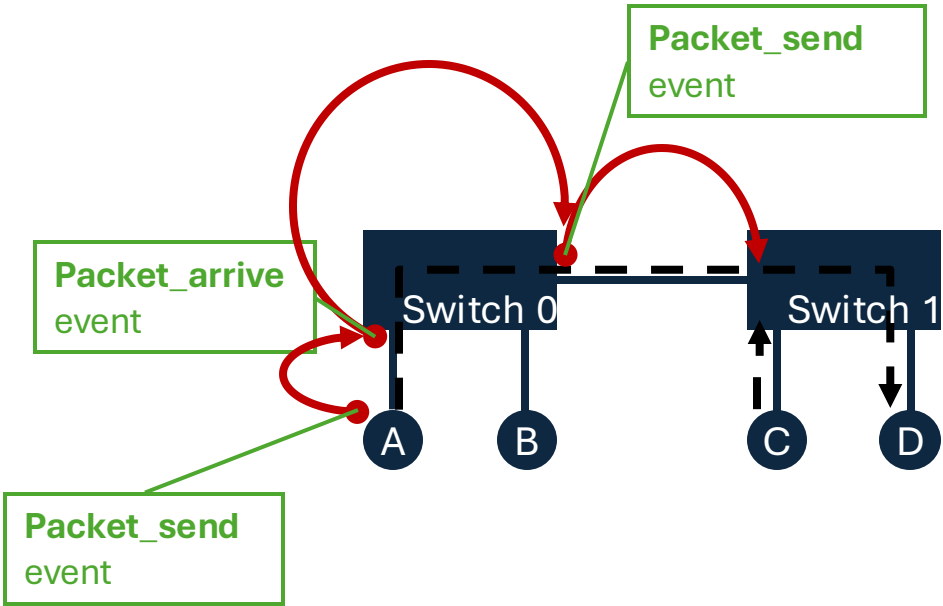
Isolation across multiple traffic classes;
QoS policy regulates access to link

Consistent performance guaranteed

Quality of Service



Simulating Supercomputer Networks

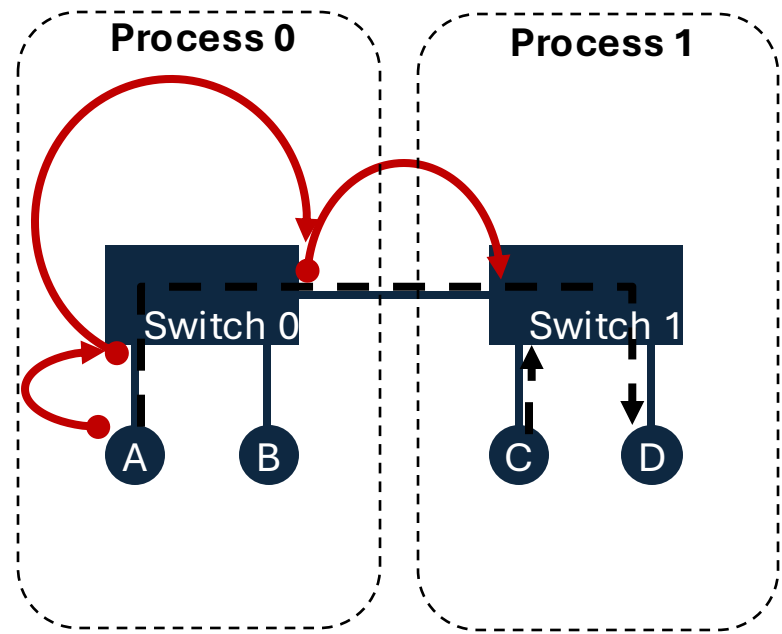


Network model

Time	Switch 0 events	Switch 1 events
5	Packet(A)Arrive*	-
10	Packet(A)Send	-
15	-	Packet(A)Arrive
16	-	Packet(C)Arrive*
20	-	Packet(A)Send
21	-	Packet(C)Send

Events in the network model

Simulating Supercomputer Networks



Network model

Process 0	
Time	Switch 0 events
5	Packet(A)Arrive*
10	Packet(A)Send

Process 1	
Time	Switch 1 events
15	Packet(A)Arrive
16	Packet(C)Arrive*
20	Packet(A)Send
21	Packet(C)Send

Events in the network model

Teaching and Supporting HPC



SC24 Early Career Program Chair

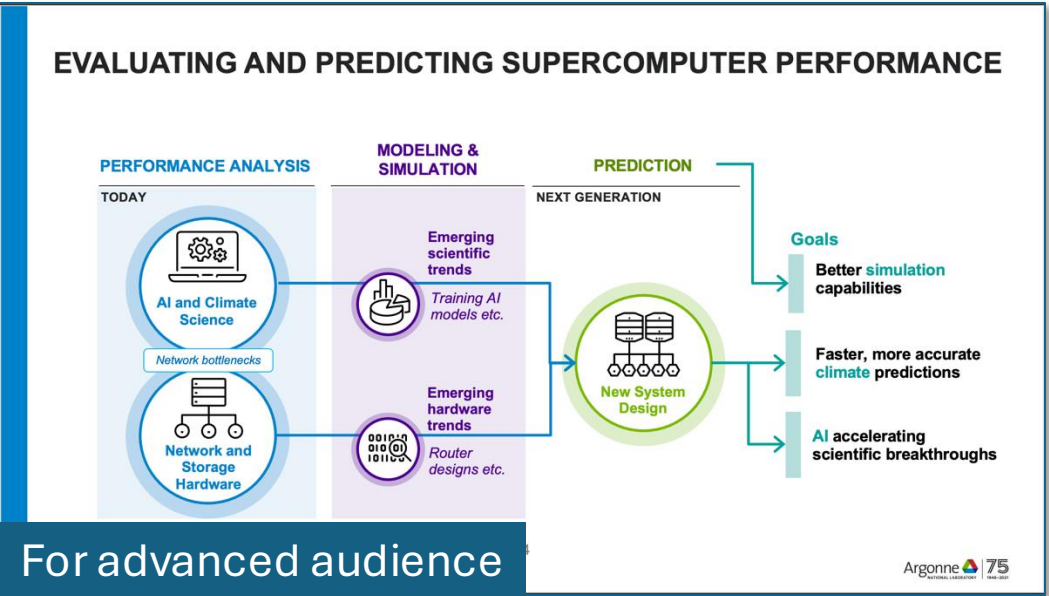
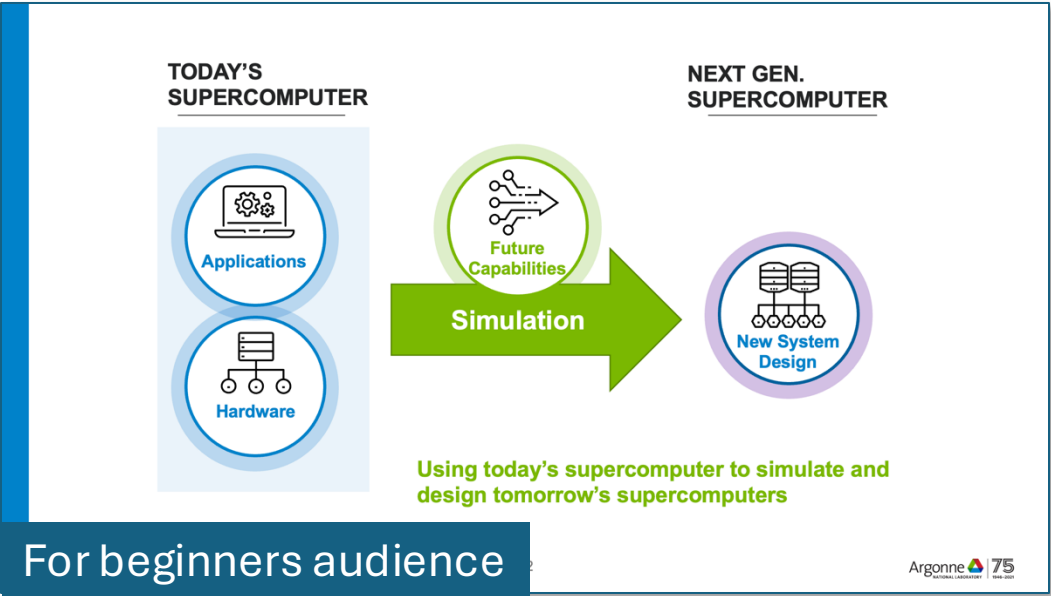
- Proposing Work
- Reporting Work
- Managing Work
- Career Paths Panel



CARLA 2025 General Co-chair

- Keynotes (NASA, Argonne Lab, etc.)
- HPC/AI workshops, tutorials, et al.
- Exhibitor booth
- Networking events

Promoting HPC



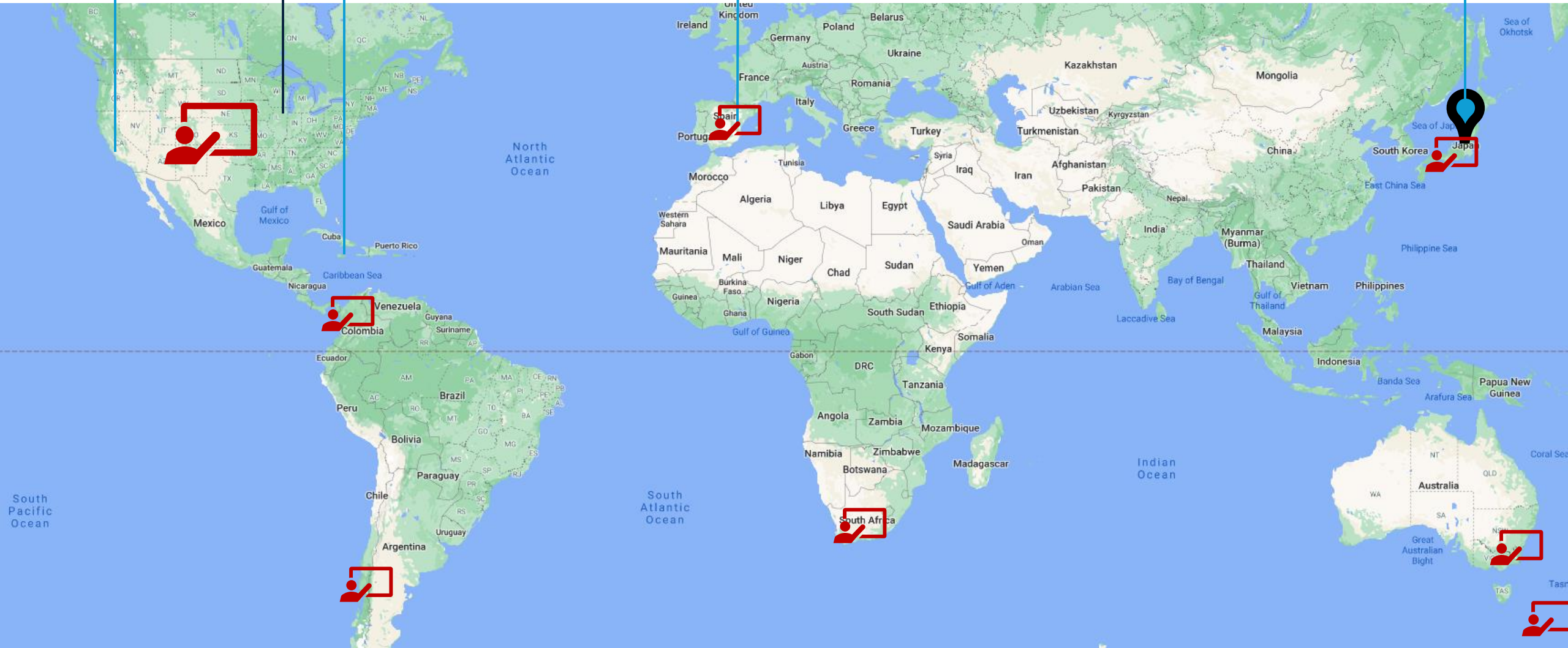
Argonne 2019

**Lawrence
Livermore
National Lab**

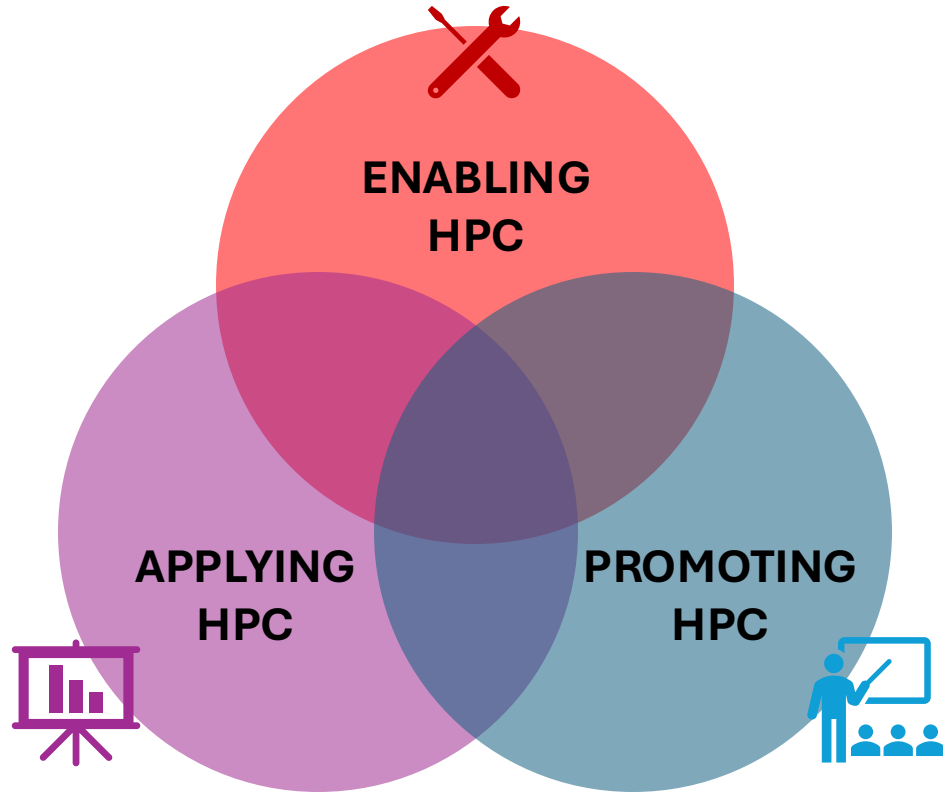
Digicel, Jamaica
[mobile network co.]

Barcelona Supercomputing Center (BSC)

**Tokyo Institute of Technology
(Tokyo Tech)**



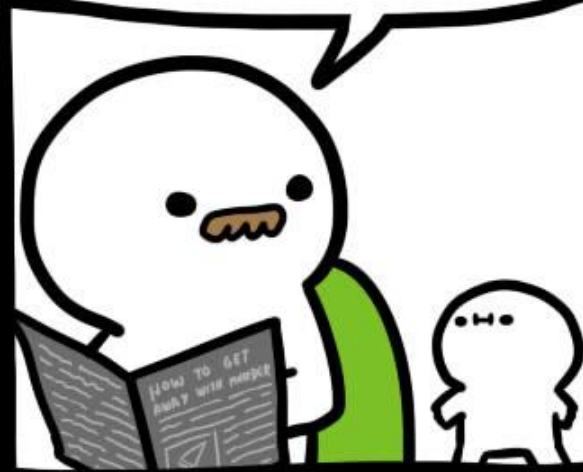
Engaging with HPC



The Journey

1. It's OK to not know what you want to do
 - It may not exist yet
2. Your path to HPC may not be linear
3. Your work with HPC may not be well bounded

BILLY
HOW WAS SCHOOL?



HEARD YOU JOINED
THE SOCCER TEAM



HOW W... **BILLY?!**



HOLY ~~SHIT~~...
HE IS GOOD

