# Memory Disaggregation Open Challenges in the Era of CXL

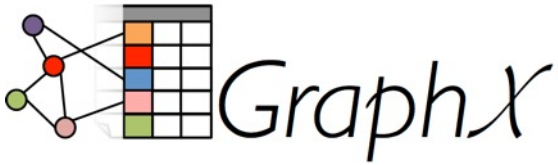
Hasan Al Maruf, Mosharaf Chowdhury



Workshop on Hot Topics in System Infrastructure, Orlando, Florida, 18 June 2023

# Memory is King!

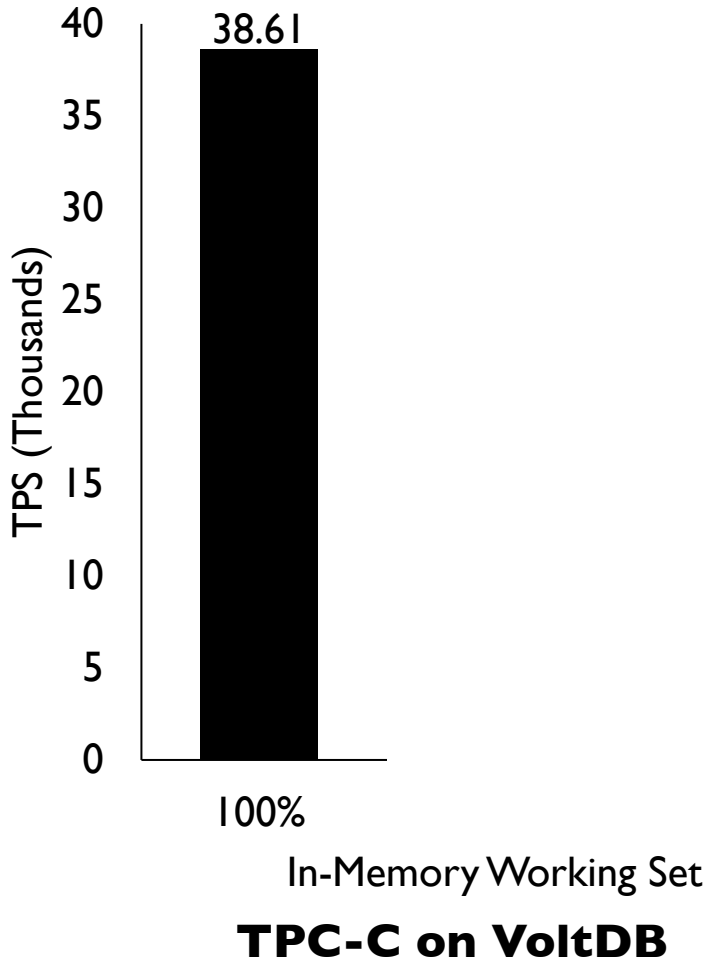

powergraph


redis

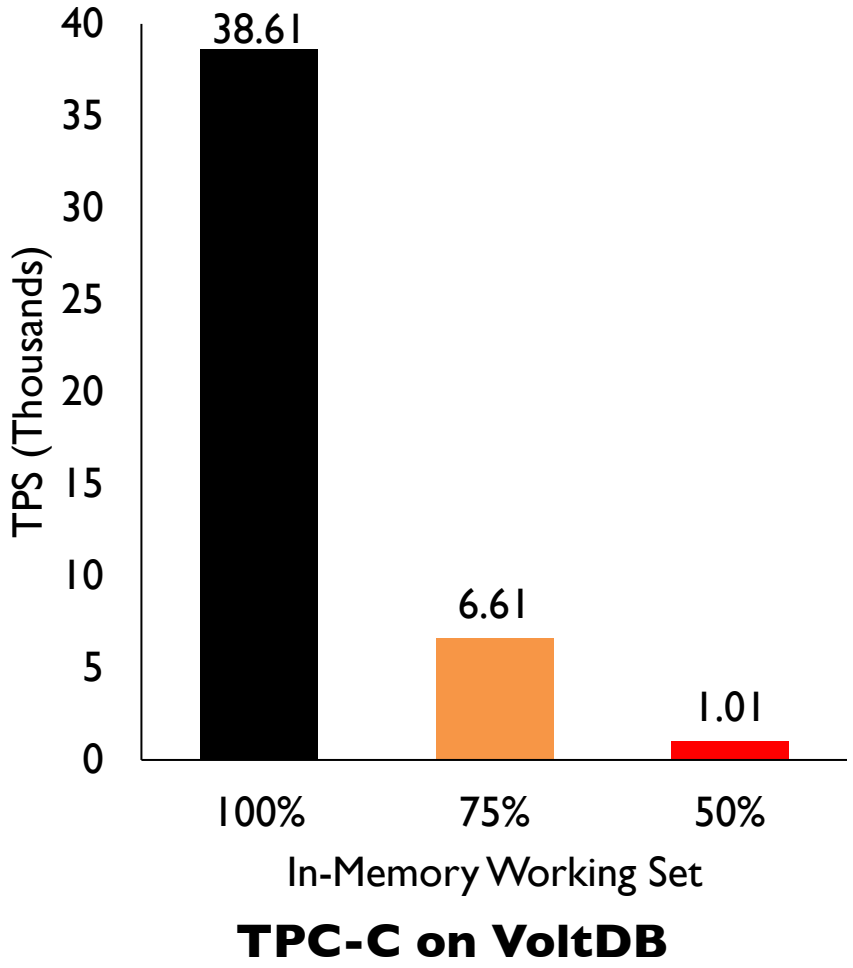
VOLTDB

# In-Memory Applications Perform Great!



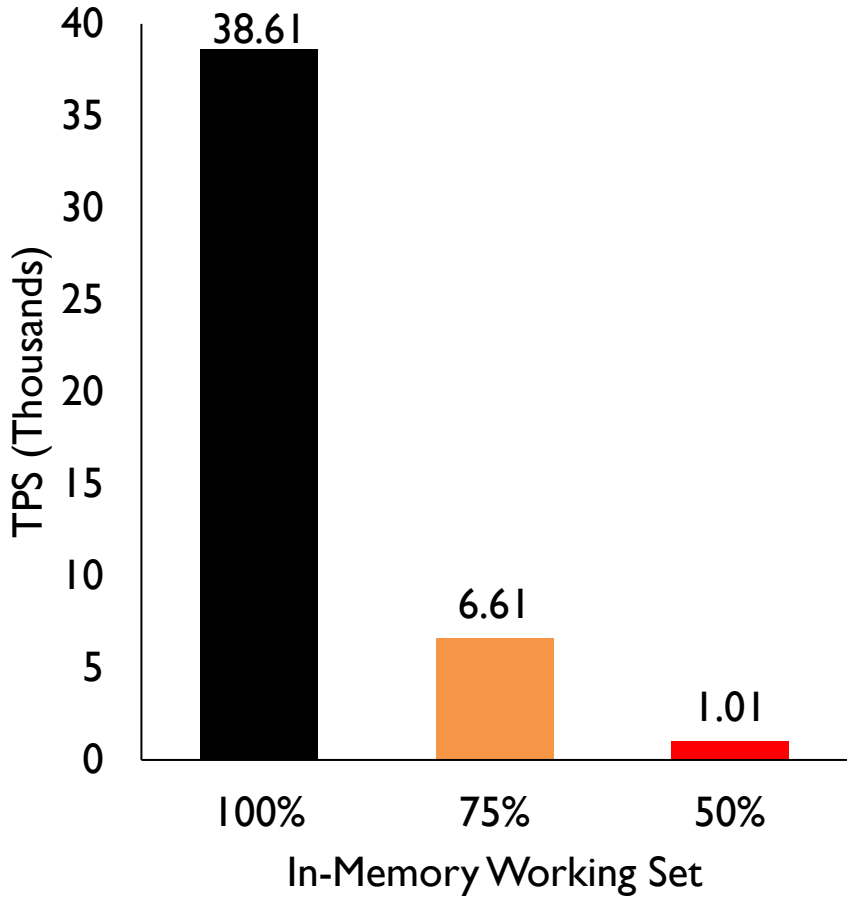

**TPC-C on VoltDB**

# Performs Great Until Memory Runs Out

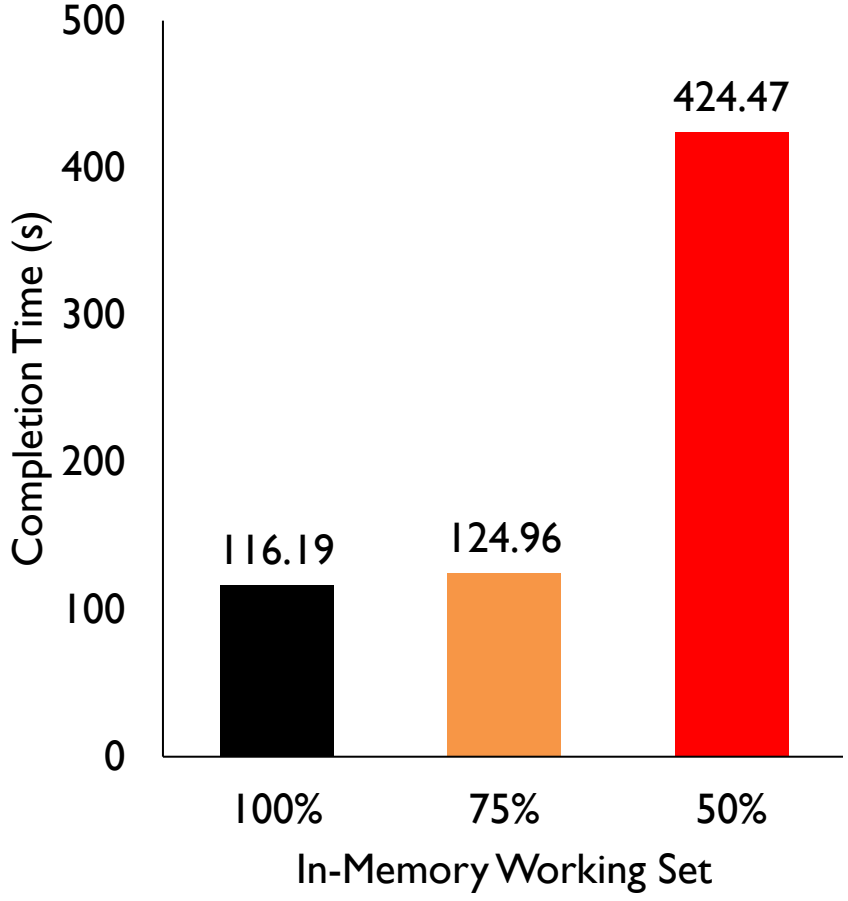



**TPC-C on VoltDB**

# Performs Great Until Memory Runs Out

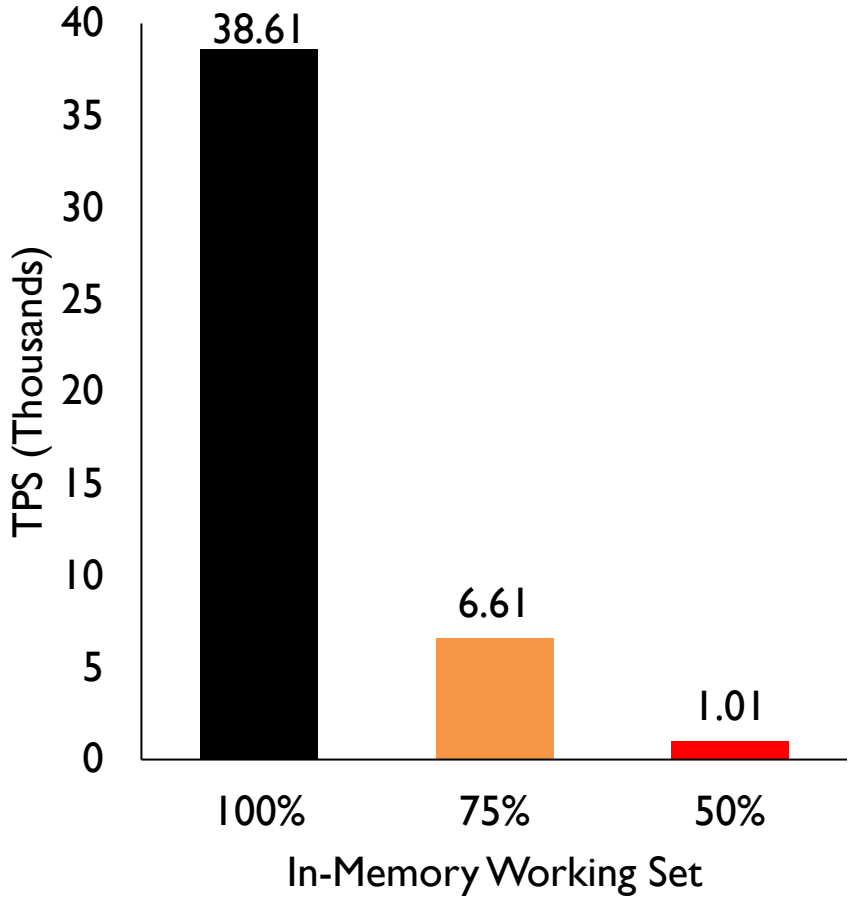



**TPC-C on VoltDB**



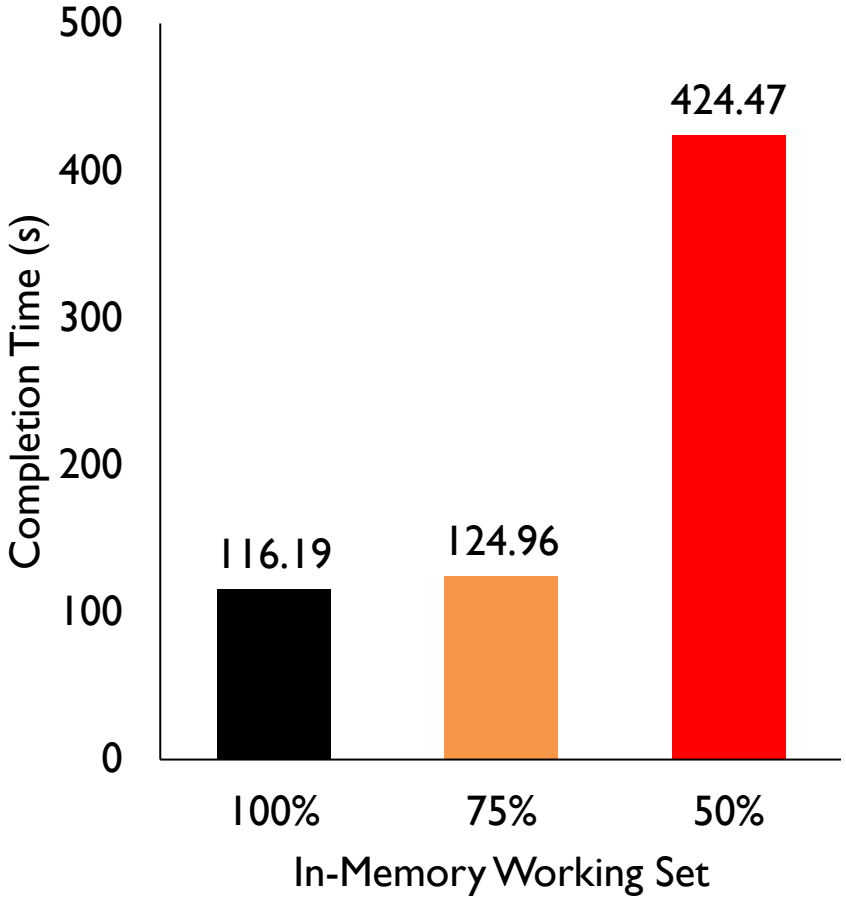

**PageRank on PowerGraph**

# 50% Less Memory Causes Slowdown of …




**TPC-C on VoltDB**

**38X**




**PageRank on PowerGraph**

**4X**

# Between a Rock and a Hard Place

## Underallocation

Leads to severe performance loss

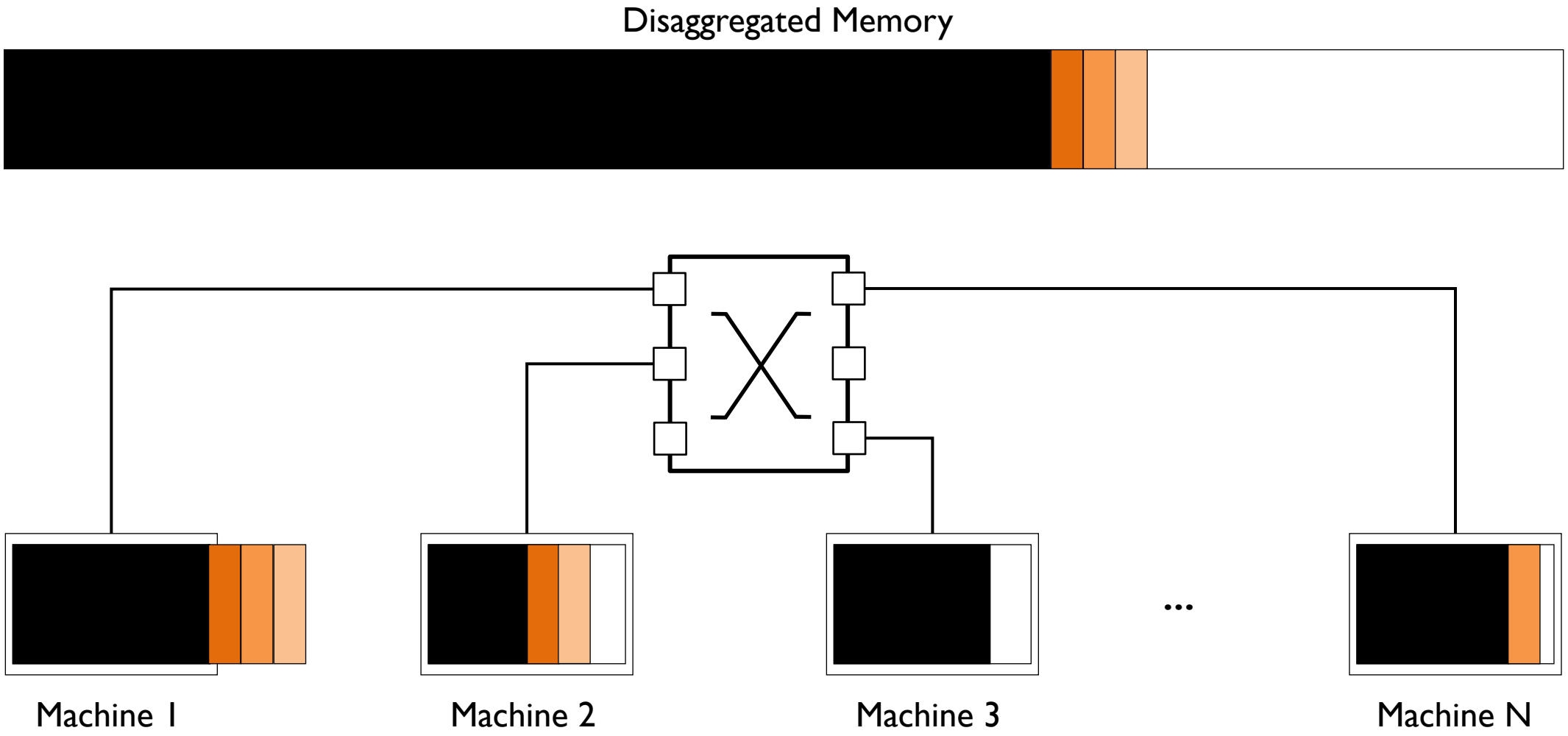
**vs.**

## Overallocation

Leads to underutilization

30-40% in Google, Alibaba, and Meta

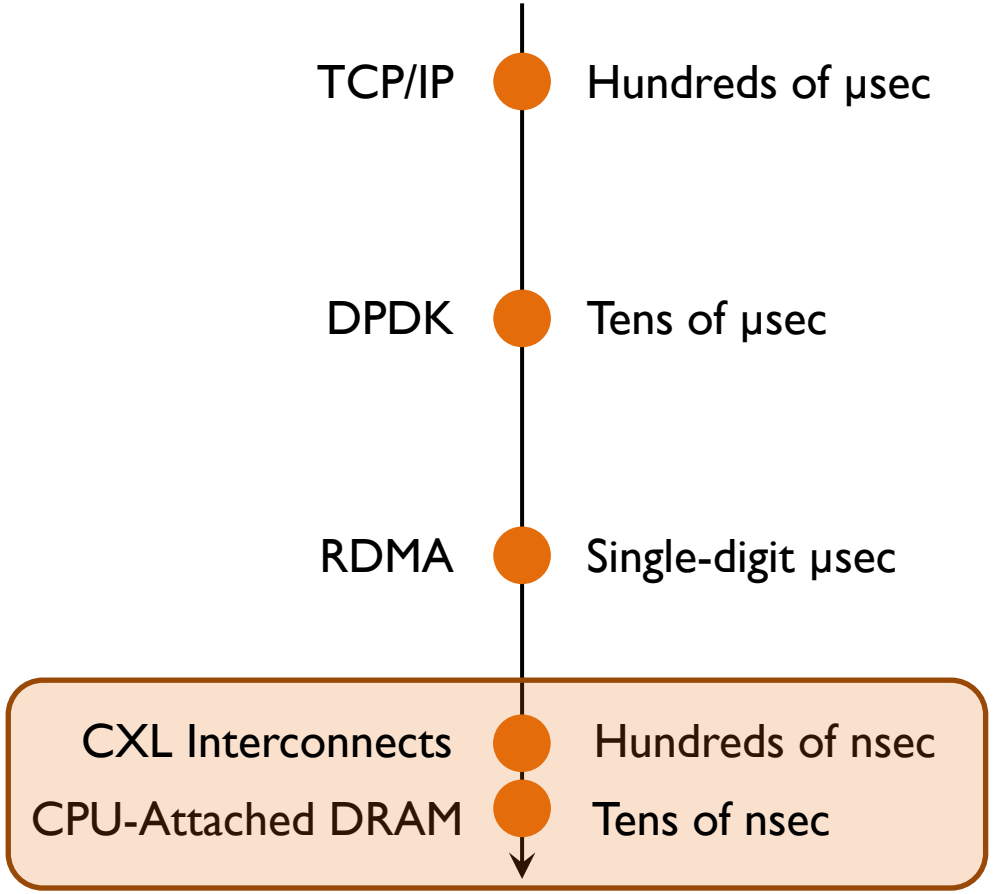
# Memory Disaggregation



# What is Practical Memory Disaggregation?

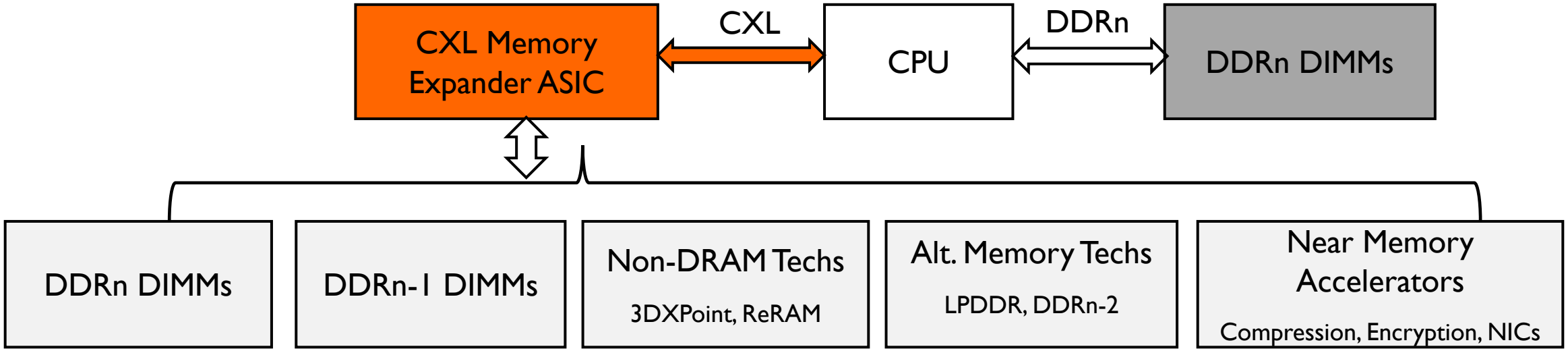

1. Applicability
2. Performance
3. Scalability
4. Resilience
5. Ubiquitous
6. Efficiency
7. Heterogeneity
8. Isolation & QoS
9. …

# Network is Getting Faster!




*time to access a 4KB memory page*

# CXL-based Heterogeneous Memory




## Flexible CPU and memory bus

- different memory capacity to bandwidth ratio
- combine different generation of DIMMs
- use cheaper and low power memory alternatives
- utilize near memory accelerators

# CXL-Memory Characteristics

## Byte addressable in same physical address space

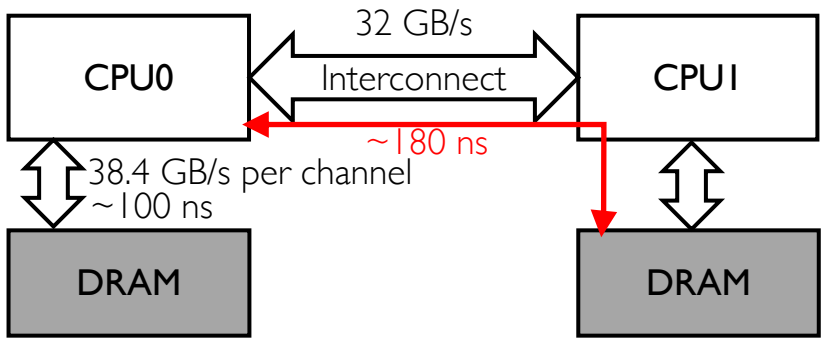
- transparent allocation with cache-line granular access

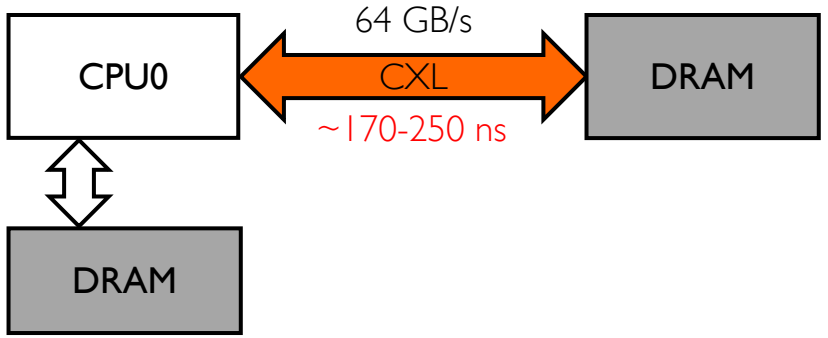
## Memory bandwidth is like DDR channels

- NUMA BW is better than a dual socket system

## Close to NUMA latency on dual socket systems

- adds ~100ns latency over normal DRAM access

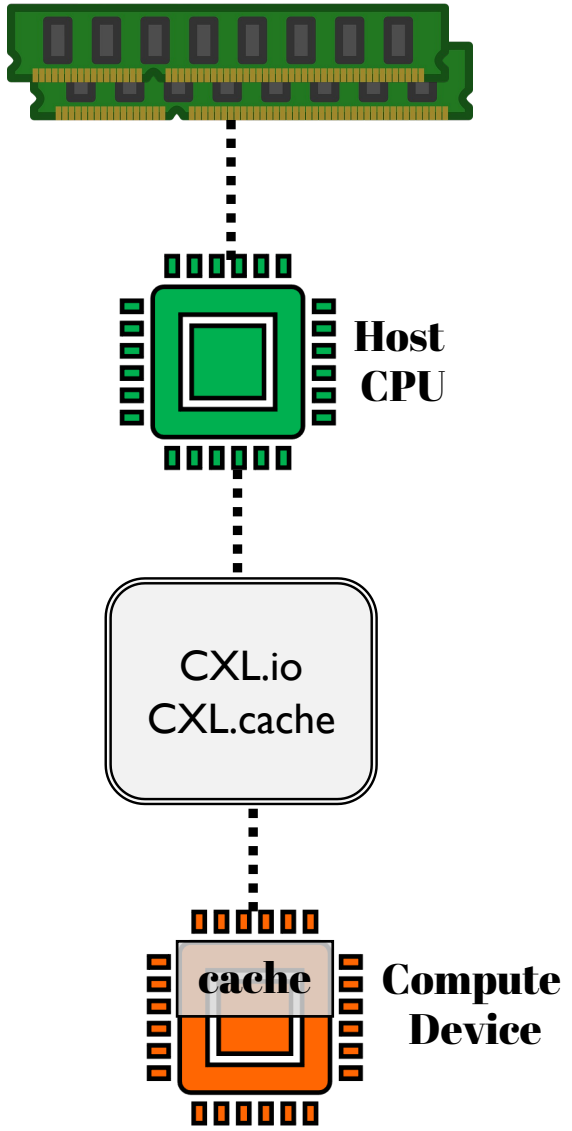


# CXL Devices Type 1

Caching device or accelerators

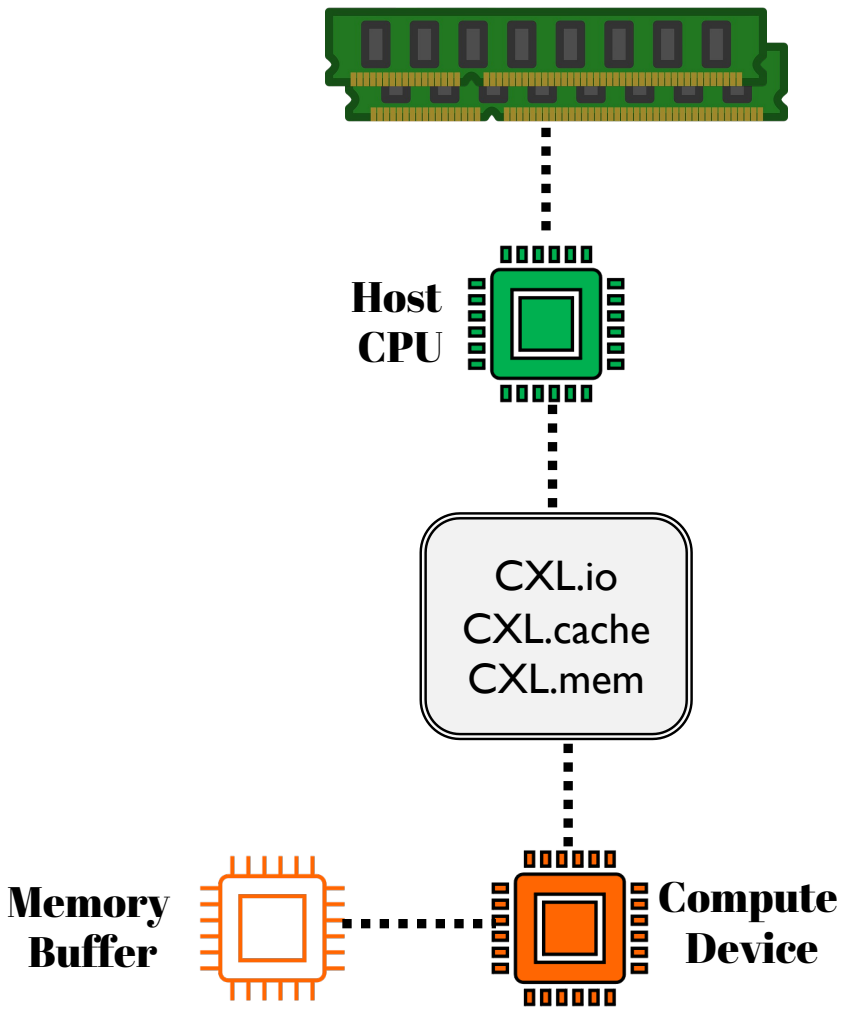
- CPU and CXL devices maintain cache coherency
- CXL device does not share memory
- uses CXL.io and CXL.cache protocols
- use case - SmartNIC



# CXL Devices Type II

## Accelerators with memory

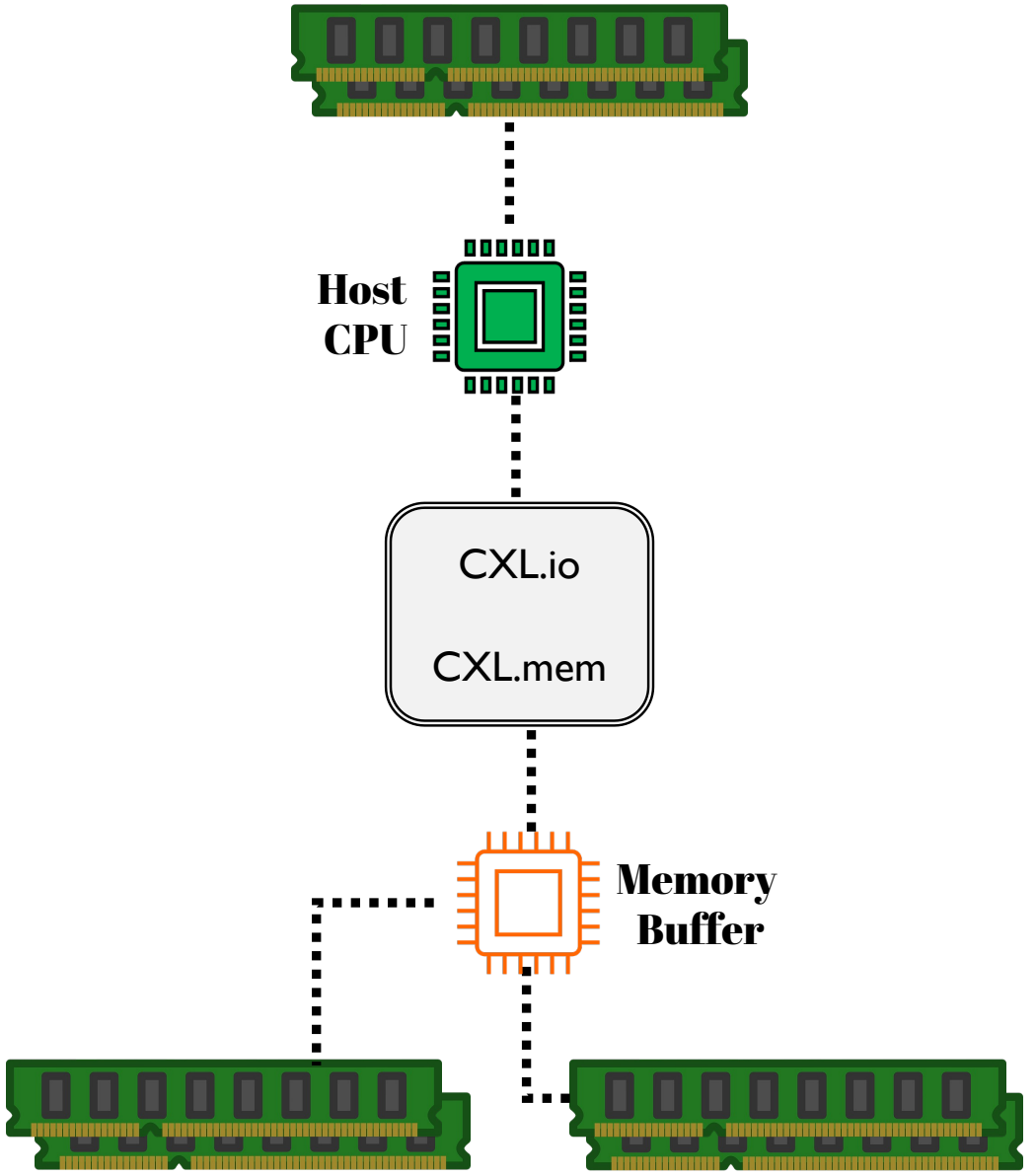
- CPU and CXL devices maintains cache coherency
- uses CXL.io, CXL.cache, and CXL.mem protocols
- host and device memory is available to each other
- use case – GPU/dense computation



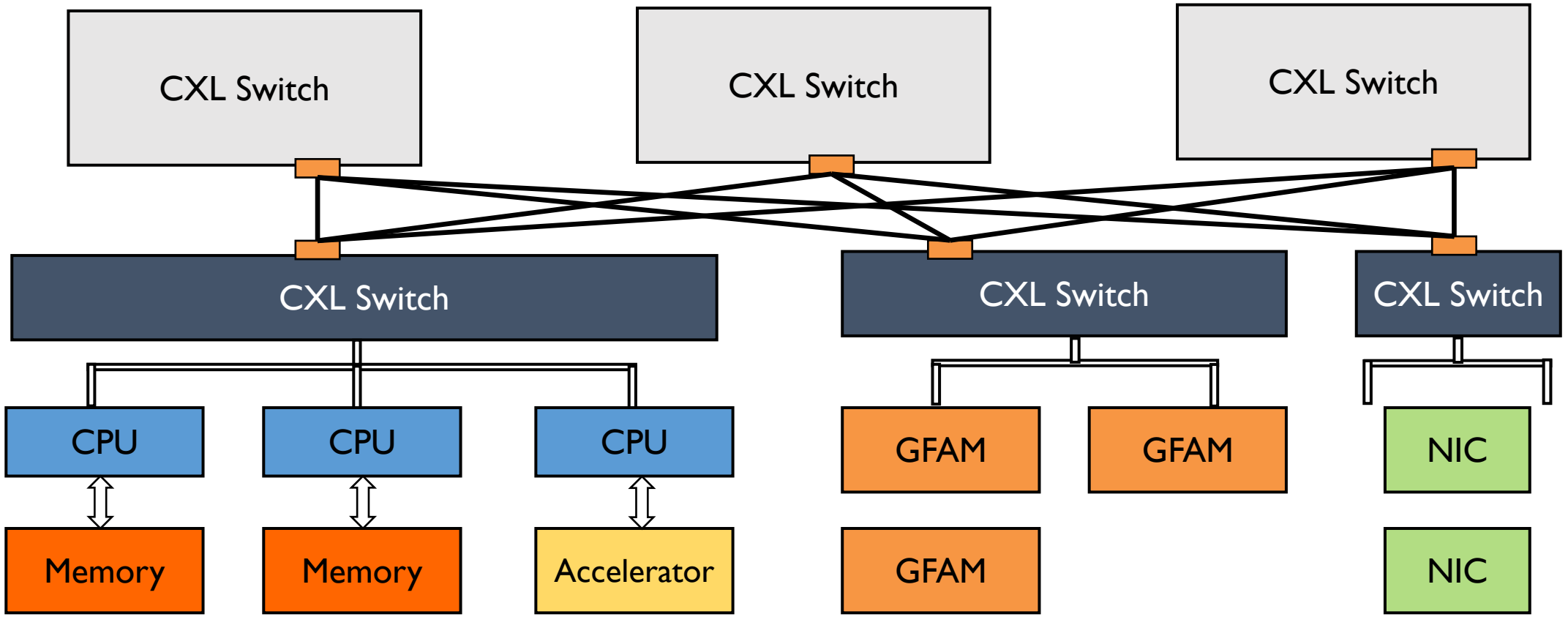
# CXL Devices Type III

Memory pooling or expansion

- CPU and CXL devices don't maintain cache coherency
- uses CXL.io and CXL.mem protocols
- host and device memory is available to each other
- use case – memory capacity or bandwidth expansion



# CXL Protocol




**CXL 3.0**

# Open Challenges - Abstraction

- Memory characteristic-aware fine-grained access
  - cache line, base-page, huge page, object granularity
- Memory allocation across heterogenous NUMA nodes
  - fraction of memory on appropriate tiers
  - promotion/demotion considering QoS and real-time system dynamics
  - allocation policy for memory bandwidth expansion
- Right amount of memory sharing and consistency

# Open Challenges – Rack-scale Objective

- Observability and page temperature determination
  - rack-scale relative hot-warm-cold identification
- Compute offloading and rack-scale scheduling
  - utilizing idle cores for energy efficiency and better performance
- Hardware-software codesign for better eco-system
  - support in CXL switches and/or CPUs for telemetry and better coordination

# What is Practical Memory Disaggregation?

Infiniswap
leap
hydra
memtrade
kona
TPP
aqua

1. Applicability
2. Performance
3. Scalability
4. Resilience
5. Ubiquitous
6. Efficiency
7. Heterogeneity
8. Isolation & QoS
9. ...

source codes available at
https://github.com/SymbioticLab

# Thank You!

for any queries, contact at
hasanal@umich.edu