OCP FUTURE TECHNOLOGIES SYMPOSIUM

OCP Global Summit

November 8, 2021 | San Jose, CA

Software Defined Memory Group Update

Manoj Wadekar, Meta

SDM Team Charter

- **Identify** key applications driving adoption of Hierarchical/Hybrid memory solutions
- Establish architecture and **nomenclature** for such Systems
- Offer **benchmarks** that enable validation of novel ideas for HW/SW solutions for such systems

Rekha Pitchumani
Samsung

Samir Mittal
Micron

Chagam Anjaneya
Intel

Dennis Hahn
Informa

Clifford Grossman
OCP

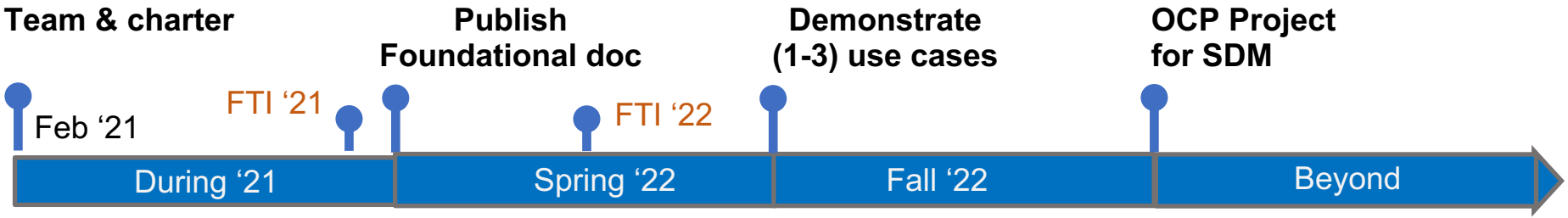
Manoj Wadekar
Meta

SW Defined Memory (SDM) Group

- SDM opportunity is around data
  - The growth in data and its processing has ramped quickly in OCP data centers
  - In-memory data processing allows processing in parallel and real-time for insights
- Challenge is better usage of memory for data workloads
  - Memory used for data processing is being asked to do more (bigger, faster & more resilient)
  - Physical memory options are proliferating allowing use case innovation
- SDM abstractions are expected to release new value for memory use
  - Expected usage: location transparency, protocol abstraction, automated tiering, unified namespace, thin provisioning and fine grained QoS

New memory markets can be enabled by merging software and novel memory technologies

SDM investigation roadmap



- Definition document
- Definition slides
- SDM Survey
- OCP use case inputs
- FTI – Industry presentations

- Work (1-3) use cases
- Expand circle of expertise
- Establish thought leadership

- Facilitate demonstrations
- Evangelize work
- Get industry involved
- Work additional use cases (according to feedback)

- *Fold project back into OCP*
- *Establish incubator project*

Investigation approach:

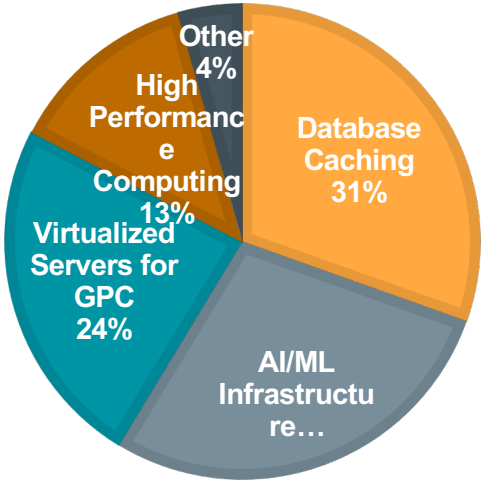
- Identify opportunities in industry to improve memory use
- Understand workload problems and select use cases
- Develop solutions use case architectures
- Offer Benchmarks for work-load specific use cases

Survey

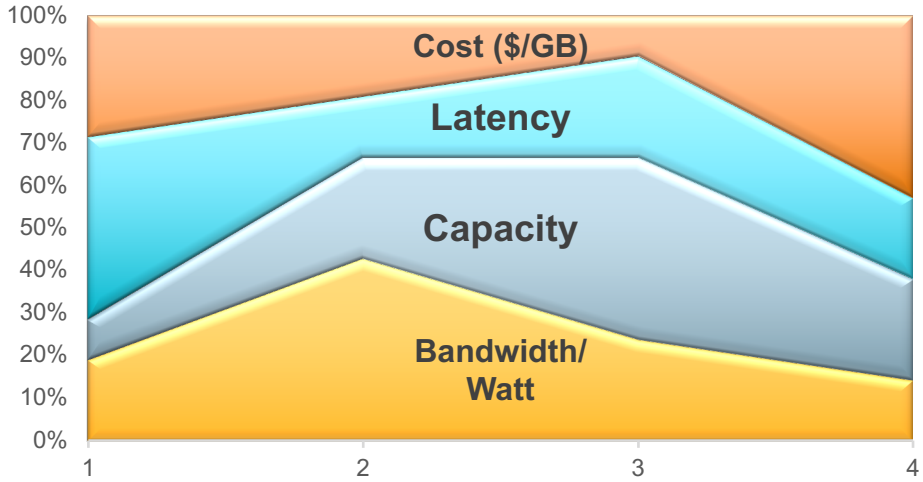
Identify key applications driving adoption of Hierarchical/Hybrid memory solutions

Memory-Bound Applications & Their Needs

Obtained via Industry Survey



Application Currently Memory-Bound!

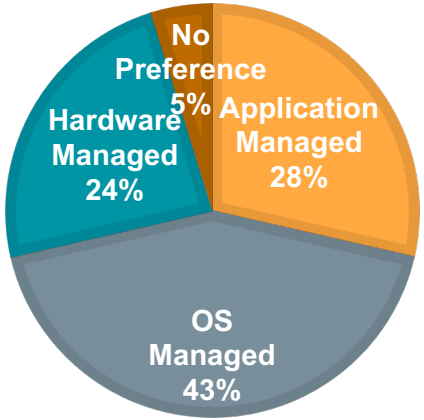


What is most to an application varies by the application

Modern applications demand memory expansion, driving needs vary

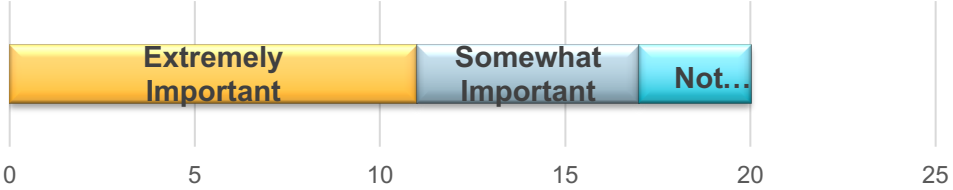
Motivation for Software Defined Memory

Obtained via Industry Survey



Strong interest in Software Managed Memory solution

Is pooling and sharing of memory capacity across multiple servers as a key aspect for your future infrastructure growth?



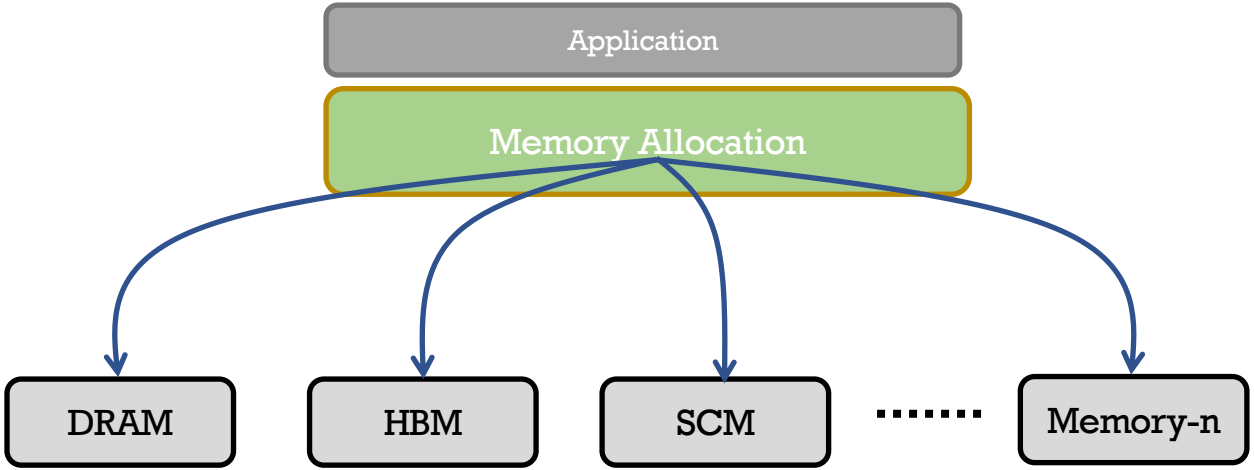
Interconnect Technologies of Interest in the Next 2-3 Years

Memory Expansion needs: Software managed, server-local and pooled

SDM Nomenclature

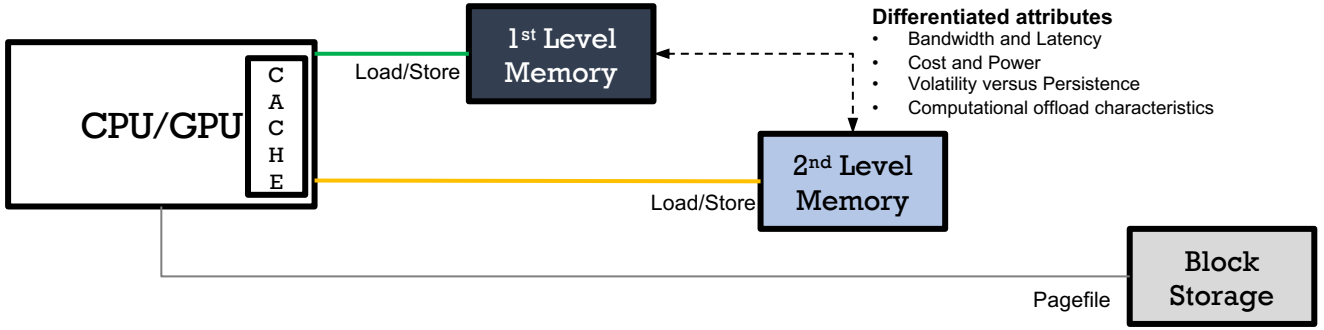
Establish architecture and **nomenclature** for SDM Systems

Software Defined Memory



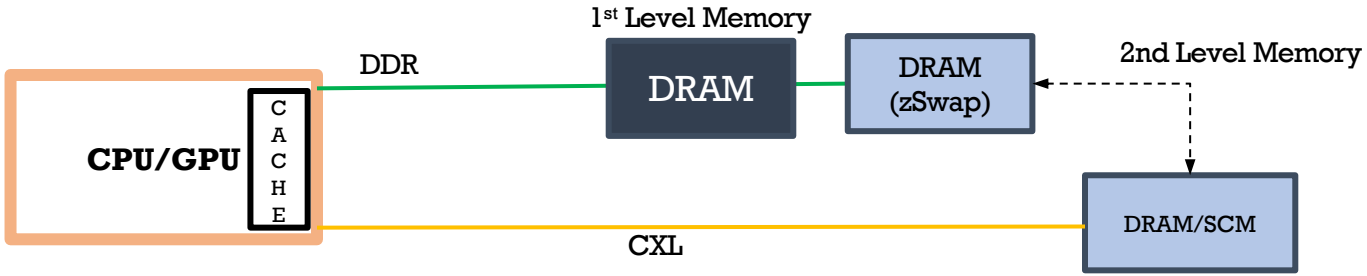
Software-Defined Memory (SDM) is an emerging architecture paradigm that provides software abstraction between applications and underlying memory resources with dynamic memory provisioning to achieve the desired application SLA

SDM Hierarchy – Logical View



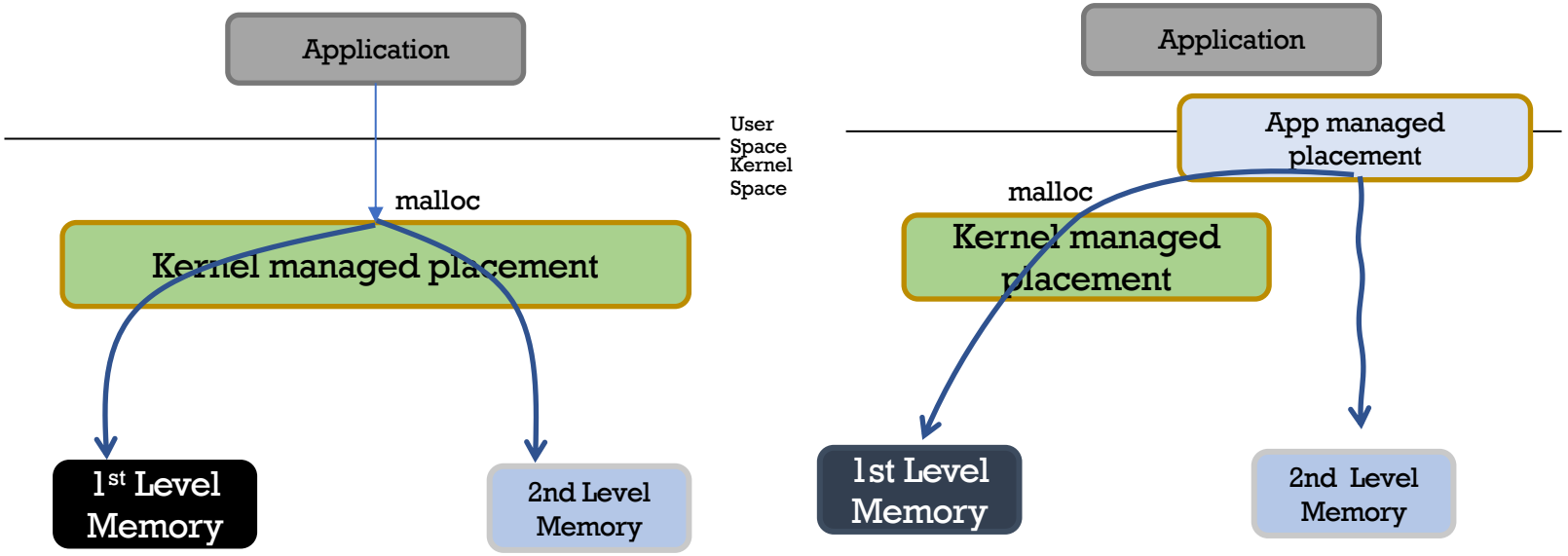
- 1st Level Memory: Preferred for OS and/or application memory allocation
  - Socket local DRAM
  - UPI connected DRAM
  - HBM
- 2nd Level Memory: Option for application memory allocation based on SLA
  - DDR-x connected Storage Class Memory
  - Memory expansion with CXL connected DRAM or SCM
  - Compressed memory pages on block storage
  - CXL fabrics based pooled memory

SDM in General Purpose Compute



- 2nd level memory use cases:
  - Compressed page store in zSwap based memory for efficiency improvements
  - Bandwidth memory expansion for Memcache class applications
  - Capacity memory expansion for database IO cache & user mode memory allocations
  - Application aware Persistent Memory integrations

SDM – Software View



SDM Benchmarks

Offer **benchmarks** that enable validation of novel ideas for HW/SW solutions for such systems

SDM Applications, benchmarks

Category	Workload	Description
Micro benchmarks	**Intel® Memory Latency Checker (MLC)** https://www.intel.com/content/www/us/en/developer/articles/tool/intelr-memory-latency-checker.html	Measure memory latencies and b/w, and how they change with increasing load on the system.
	Stream Triad http://www.cs.virginia.edu/stream/FTP/Code/ http://www.cs.virginia.edu/stream/ref.html	simple, synthetic benchmark designed to measure sustainable memory bandwidth (in MB/s) and a corresponding computation rate for four simple vector kernels: Copy, Scale, Add and Triad
	fio (Flexible I/O Tester) https://github.com/axboe/fio.git	Block, pmem benchmark tool
Caching	**CacheBench** https://cachelib.org/docs/Cache_Library_User_Guides/Cachebench_Overview https://github.com/facebook/CacheLib.git	benchmark and stress testing tool to evaluate cache performance with real cache workloads
	memtier https://github.com/RedisLabs/memtier_benchmark	Redis and Memcache traffic generation and benchmarking tool

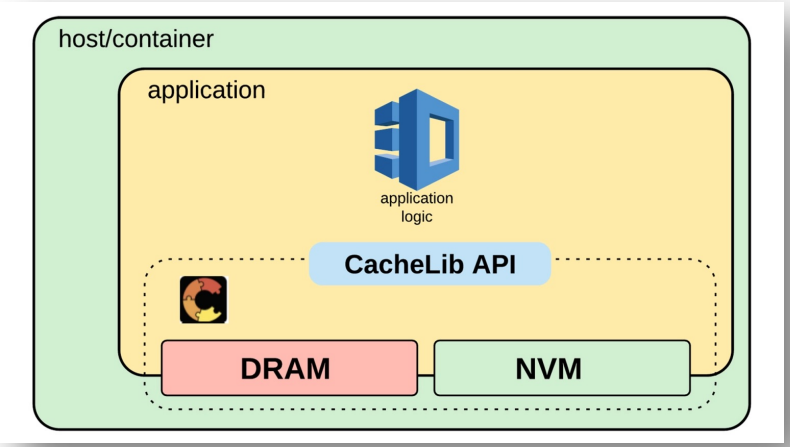
SDM Applications, benchmarks

Category	Workload	Descrit
Databases	**db_bench** https://github.com/facebook/rocksdb.git	benchmark RocksDB's performance
	SysBench https://github.com/akopytov/sysbench	Database benchmarking (e.g. MySQL)
	cassandra-stress https://docs.datastax.com/en/dse/5.1/dse-dev/datastax_enterprise/tools/toolsCStress.html	Java-based stress testing utility for basic benchmarking and load testing a Cassandra cluster.
	HammerDB https://hammerdb.com/	benchmarking and load testing software for databases (Oracle Database, SQL Server, IBM Db2, MySQL, MariaDB and PostgreSQL).
AI	**MLPerf** https://github.com/mlcommons/inference	performance benchmarks that cover a range of leading AI workloads widely in use.
Bigdata	**HiBench** https://github.com/Intel-bigdata/HiBench	big data benchmark suite

Benchmarks - gaps

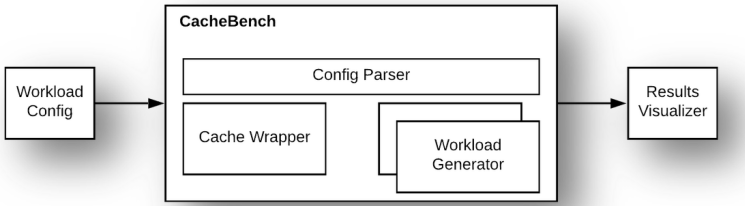
- Transparent Memory use cases can take advantage of existing benchmarks
- We need benchmarks that are focused on tiered memory
  - E.g. Kernel tiering – how to track hot/cold pages - perf impact
- Application changes may be required for taking advantage of App-managed-memory
- Need industry's help

CacheBench



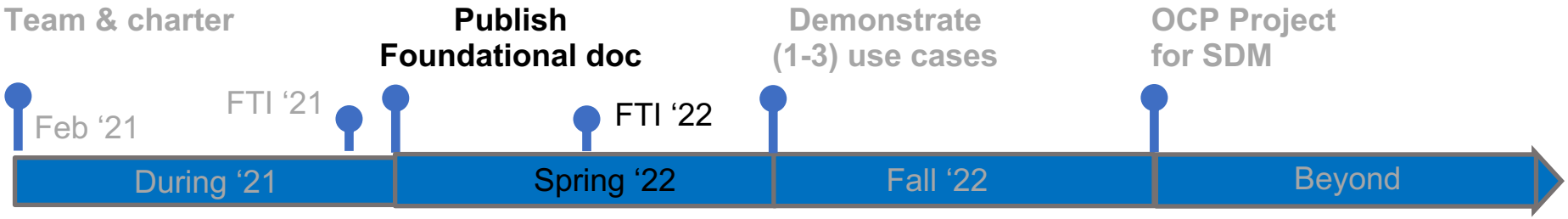
Source: https://engineering.fb.com/2021/09/02/open-source/cachelib/

- **CacheLib** – pluggable **in-process caching engine** to build and scale high-performance services
  - C++ Library
  - Thread-safe API
  - Manages DRAM and Block Caching transparently
- Meta open-source project: https://github.com/facebook/CacheLib
  - *See www.cachelib.org for documentation and more information.*



- **CacheBench** - benchmarking tool for evaluating caching performance

Next Steps



- Definition document
- Definition slides
- SDM Survey
- OCP use case inputs
- FTI – Industry presentations

- Work (1-3) use cases
- Expand circle of expertise
- Establish thought leadership

- Facilitate demonstrations
- Evangelize work
- Get industry involved
- Work additional use cases (according to feedback)

- *Fold project back into OCP*
- *Establish incubator project*

Investigation approach:

- Identify opportunities in industry to improve memory use
- Understand workload problems and select use cases
- Develop solutions use case architectures
- Offer Benchmarks for work-load specific use cases

Call To Action

- Help define finalizing Storage Defined Memory architecture and use cases
- Help define benchmarks for the key use cases using SDM
- Join the team: https://www.opencompute.org/projects/software-defined-memory-workstream

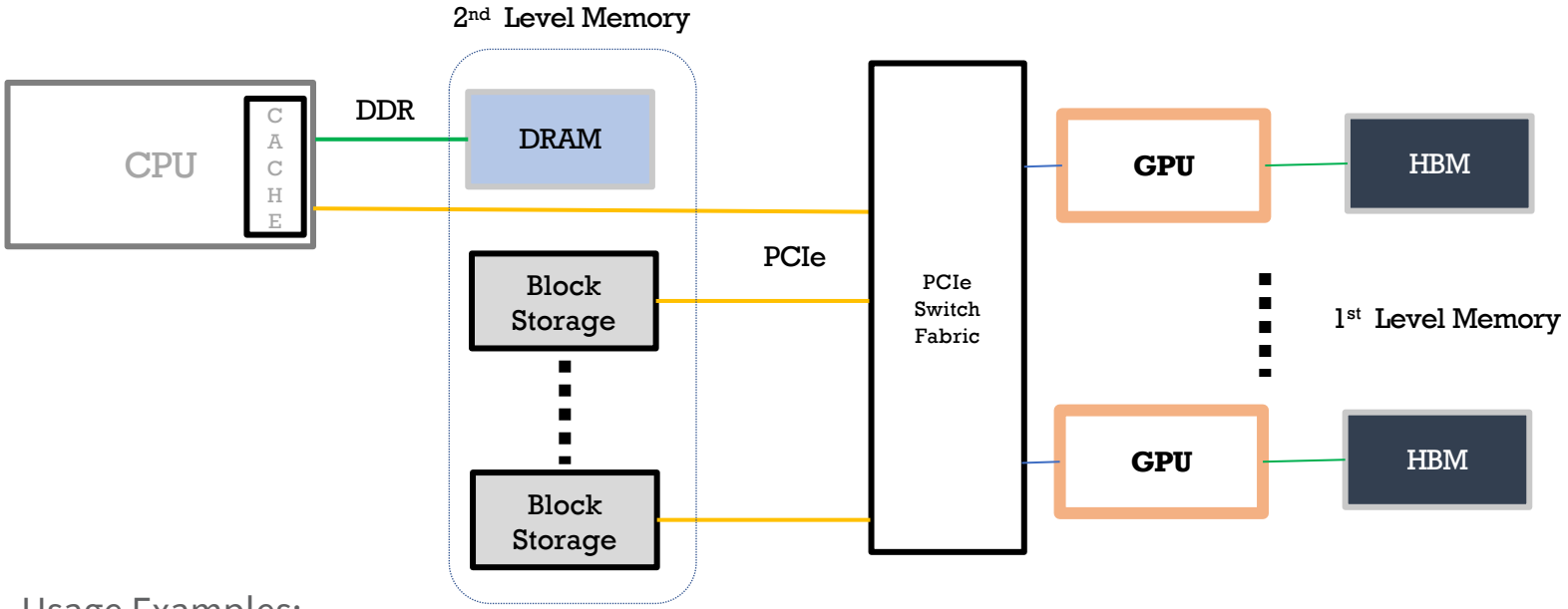
Thank You


OCP
TOTAL SOLUTIONS

OCP FUTURE TECHNOLOGIES SYMPOSIUM

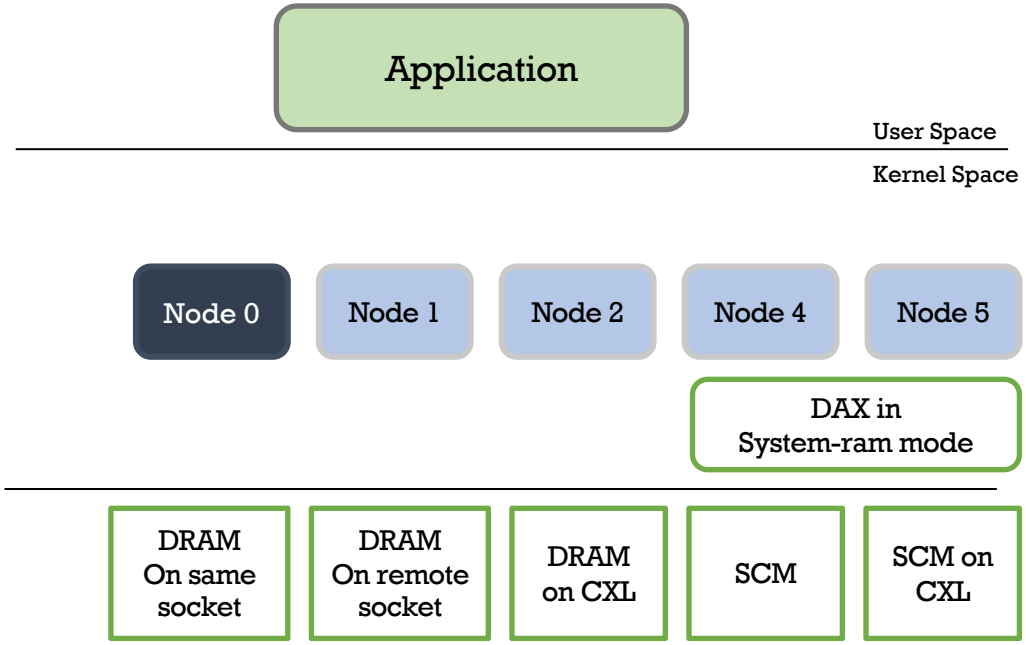
2021 OCP Global Summit | November 8, 2021, San Jose, CA

SDM scenario in GPU training



- Usage Examples:
  - GPUs accessing CPU attached memory through DMA
  - GPU accessing NVMe block storage for training parameters
  - Other current and future interconnects: CXL, NVLink etc.

SDM Hierarchy – SW (NUMA) View



- Kernel can enumerate devices
  - 1LM device showing up on Node 0
  - All other devices on separate Nodes
- SCM can be accessed through DAX in system-ram mode
  - As remote NUMA
- Multiple Namespaces to tabulate devices