SharP

Towards Programming Hierarchical-Heterogeneous Memory based Extreme-Scale Systems

Manjunath Gorentla Venkata and Ferrol Aderholdt

OAK RIDGE National Laboratory

Goals

- ❑ Simple, usable, and portable abstraction for hierarchical-heterogeneous memory
- ❑ Unified programming constructs for Big-Compute and Big-Data applications
- ❑ Native support for data-centric abstractions
- ❑ Portability across diverse (GPU-based and Xeon-Phi) memory hierarchy systems

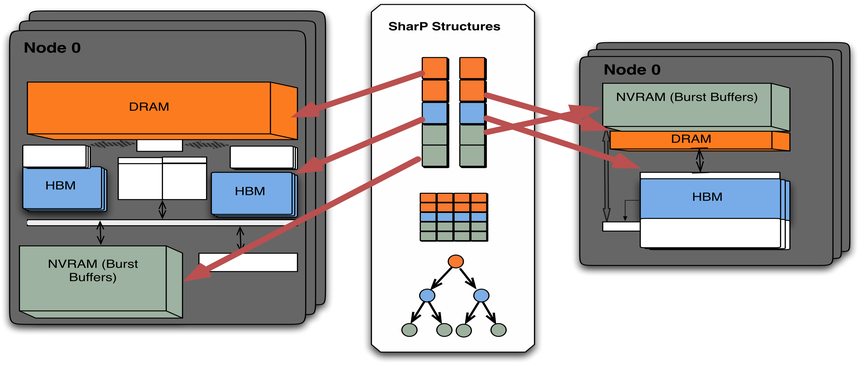
Approach

❑ Data-structure based and data-centric programming construct to support

- Various views of the data, including global and local view
- Data resiliency, sharing, locality and affinity
- Uniform interface across hierarchical and heterogenous memory

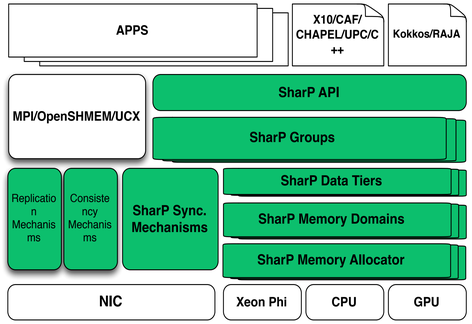
❑ Unified memory abstraction that manages data across various hierarchical and heterogenous memory

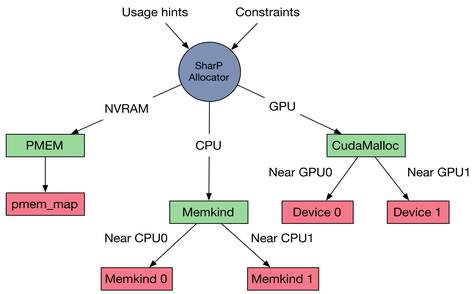
- User-friendly posix-like memory management interfaces and semantics



Mapping of distributed data-structures onto different architectures

SharP Implementation

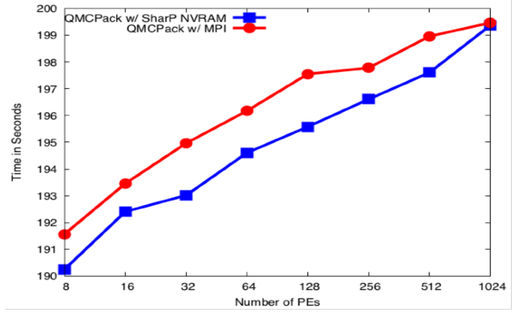


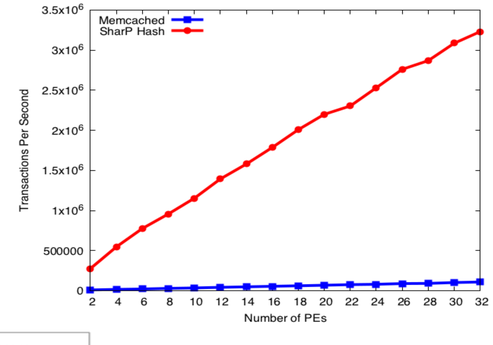


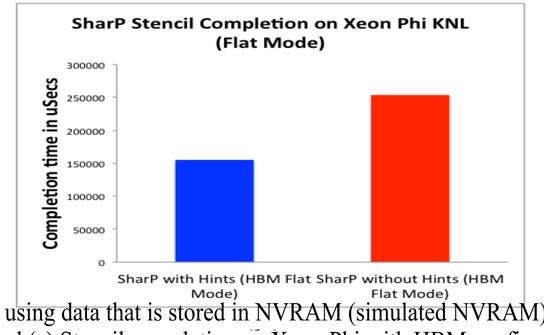
❑ Implementation of SharP

- Library-based API and implementation
- Supports locality-aware memory management on various memories types including DDRAM, HBM, and Persistent Memory
- Supports SharP Arrays and Hash tables
- Interoperable with MPI, OpenSHMEM, and OpenMP

Evaluation: Big Compute and Big Data







(a) QMCPack using data that is stored in NVRAM (simulated NVRAM) (b) Memaslap evaluation of SharP Hash and Memcached (c) Stencil completion on Xeon-Phi with HBM configured in Flat mode