THE UNIVERSITY OF TOKYO

東京大学情報基盤センター
INFORMATION TECHNOLOGY CENTER, THE UNIVERSITY OF TOKYO

JCAHPC

HW/SW Optimizations for Emerging HPC Systems: Memory Perspective

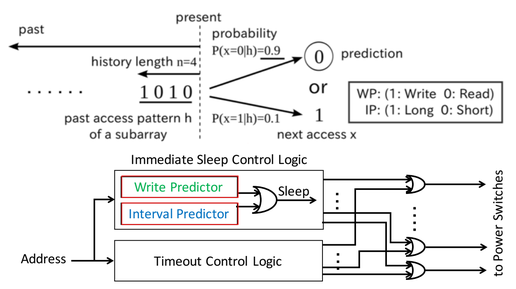
Eishi Arima, arima@cc.u-tokyo.ac.jp

Background

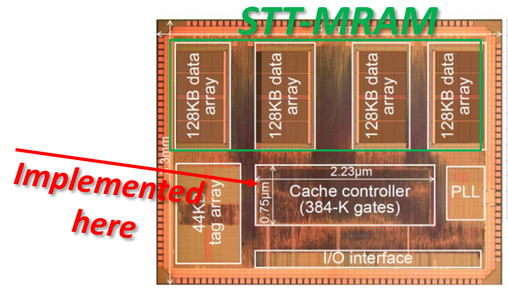
- **Fundamental problem:** Scientific computing is bottlenecked or constrained by memory in terms of bandwidth or capacity
- **Promising direction:** Architecting compute nodes with emerging memory technologies (e.g., 3D stacked DRAMs and NVRAMs) which are scalable in bandwidth or capacity
- **Approaches:** (1) NVRAM based caching; (2) Data transfer optimization; (3) Power management; (4) scheduler optimization

NVRAM Cache Design

- **Goal:** Improving performance or energy efficiency using highly scalable NVRAMs (e.g., STT-MRAM)
- **Problem:** Large energy overhead in peripherals
- **Solution:** Power gating + intelligent sleep control



Sleep control logic @ ICCD'15

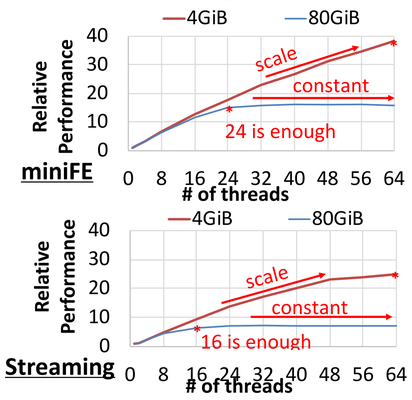


Real chip @ ISSCC'16

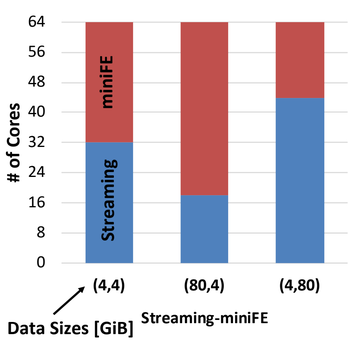
Scheduler Optimization

- **Goal:** Mitigating resource wastes within a node
- **Focus:** Co-scheduling
  - ✓ Co-running multiple jobs on one single node
  - ✓ Mixing CPU/memory intensive jobs is effective
  - ✓ The intensity highly depends on problem size
- **Solution:** Problem size aware co-scheduling
- **Another topic:** I/O staging aware co-scheduling

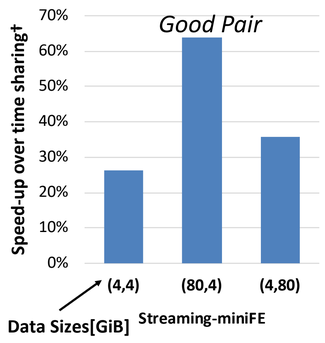
Scalability changes (solo-run)



Optimal core allocations change (co-run)



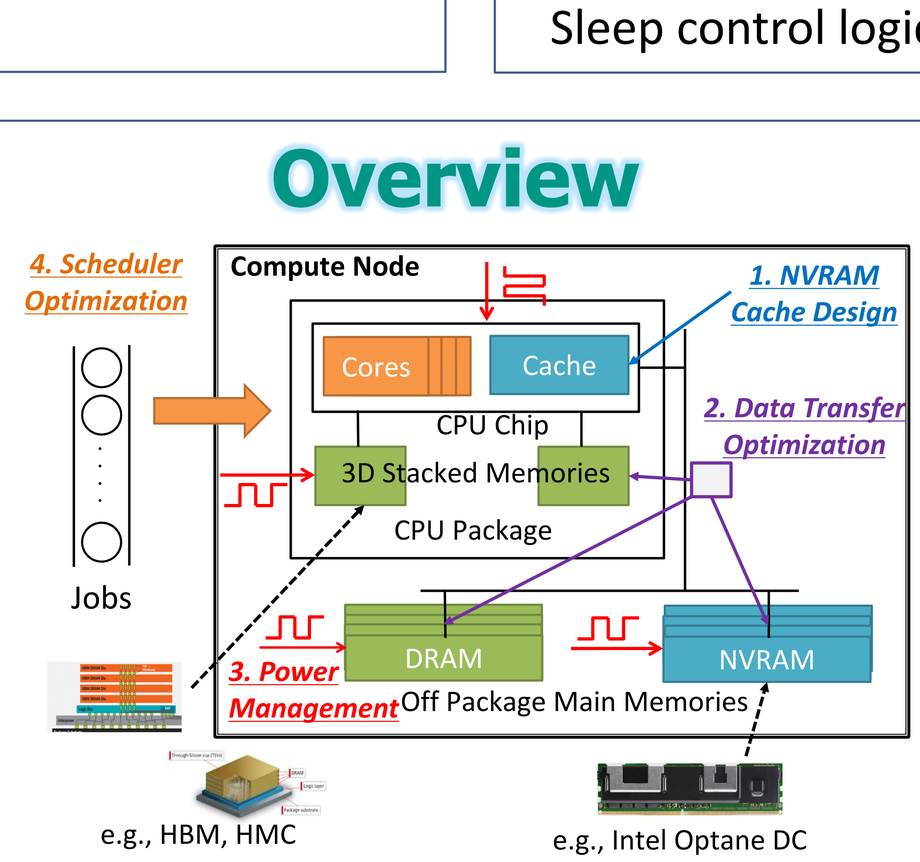
Performance gain changes (co-run)



*Tested on a KNL based system with HW cache mode

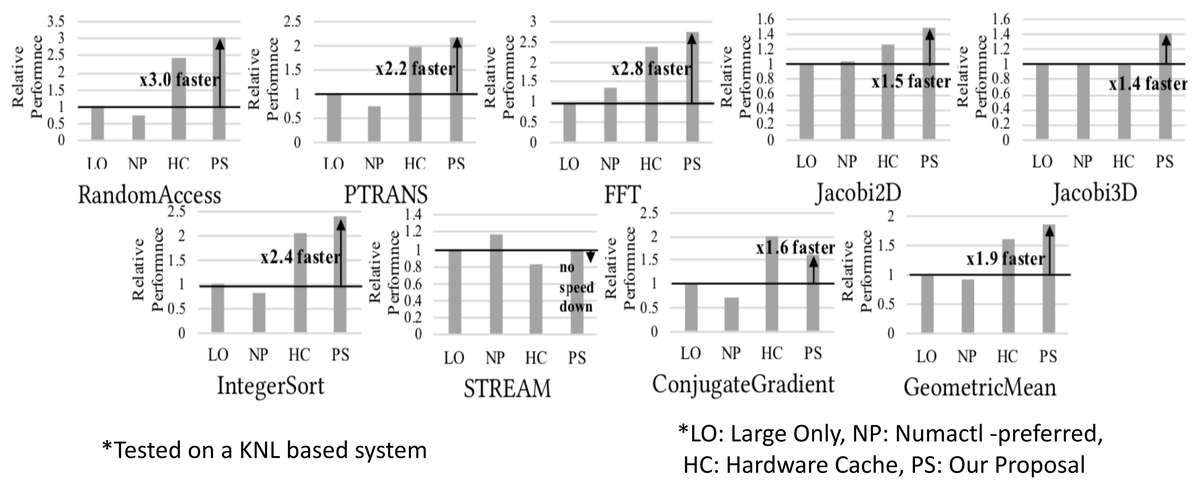
†***Time sharing:*** conventional exclusive solo-runs

Overview



Data Transfer Optimization

- **Goal:** Exploiting both *high bandwidth* and *large capacity* on hybrid main memories
- **Solution:** A compiler-based data transfer optimization
  - ✓ Performance benefit highly depends on access pattern
  - ✓ **Our approach:** *Dynamic* pattern analysis with our *statically* generated code (source-to-source transformation)

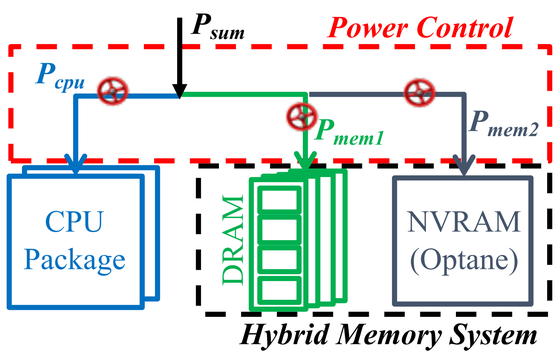


*Tested on a KNL based system

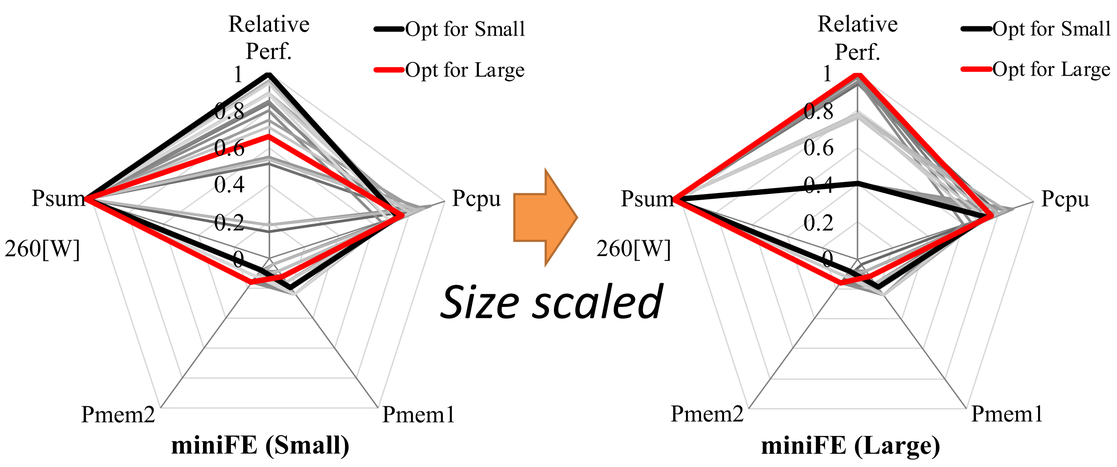
*LO: Large Only, NP: Numactl -preferred, HC: Hardware Cache, PS: Our Proposal

Power Management

- **Goal:** Optimizing power budget allocations to components under a given node power constraint
- **Observation:** Optimal power settings depend on problem size
- **Solution:** Problem size aware power shifting



Hardware setting



Let me know if you are hiring postdocs or scientists related to these topics :)

Contact: arima@cc.u-tokyo.ac.jp