



Contribution ID: 170

Type: not specified

## KEYNOTE 5: QA<sup>A</sup>HPC – The Quest for Quantum/AI Optimized HPC Workflow

*Thursday, 7 December 2023 09:00 (45 minutes)*

As the emergence of ExaFLOPS Top500 Systems like ORNL Frontier HPC cluster in June 2022, we plan to apply our innovative fine-grained topology-aware software-hardware ATMapper to improve benchmark performance toward ExaFLOPS system's peak performance. Due to application challenges in data movement, limited degree of parallelism, sparse matrix and/or irregular workflow, the sustained benchmark performance like HPCG can only reach ~1% of system peak performance (14PF/1685 PF), compared to world's best HPCG Benchmark of ~3% peak performance (16PF/537PF) by Riken Fugaku cluster in November 2022. Comparing two software-hardware graph-mapping approaches for workflow partitioning/assignment/scheduling in our previous 2021 DoE VFP project, we tested Dr. Butko's load-balanced LBNL TIGER mapper using D-Wave's Quantum/SimulatedAnnealer, and our Dr. Shih's self-organizing load-imbalanceATMapper using AI A\* search. We are optimistic about designing a better future Q/AI TIGER/ATMapper hybrid to help most any complex, irregular HPC applications finding the best topology-aware processor assignment (or application-custom network topology synthesis) given their computation workflow dependence constraints. Dr. Shih's ATMapper is a self-organizing load-imbalanced static workload assignment/scheduler, capable of an average 0.5 data hop on 90% of data movement (0 hop: reusing same processor node as possible, or 1 hop: transferring data if necessary to immediate neighbor node), comparing to the typical 3 hops data movement among switches on ORNL Frontier Dragonfly topology enhanced by dynamic HPE Cray Slingshot Interconnect. We hope that our static algorithm-specific topology-aware ATMapper workload scheduling will complement HPE's Slingshot Interconnect dynamic run-time load-balanced traffic routing optimization to increase HPCG software benchmark performance (currently <3%) moving closer toward full system peak performance. With QAs' negligible cost/power/space requirement, QA<sup>A</sup>HPC software-hardware co-design optimization is a green game-changer toward computation cost efficiency and sustainability for both HPC application users and data center providers.

### Student or Postdoc?

**Primary author:** Prof. SHIH, Liwen

**Presenter:** Prof. SHIH, Liwen

**Session Classification:** Keynote