Quantum Leap in Accelerated Computing:

The Quest of the Missing Links between Quantum Annealer and HPC

Liwen Shih, Ph.D.
Computer Engineering
University of Houston – Clear Lake
Houston, Texas, USA
shih@uhcl.edu

***Abstract*—As the arrival of Adiabatic Quantum Annealing Computers (QACs) era, there are ample opportunities to search for the missing links between QACs and HPC (High-Performance Computing). QACs such as D-Wave 2000Q Systems are analog Quantum Annealers capable of instantly zooming in to optimal solutions. We are optimistically aiming at broadening the perspective and impact of QAC by harvesting QAC progress to potentially benefit most every HPC application by optimizing the software/hardware mapping. The narrowed-down, fittest processor schedules found through quantum (or hybrid classical simulated) annealing search enhancement, can then be further compared and fine-tuned to run computation more efficiently in production mode on target HPC systems. With our novel perspective of linking QAC and HPC for a broader application impact, we hope to encourage more and various developments of emerging quantum computer endeavors to eventually make the most of manual tweaking of various problem solving, including parallel programming, unnecessary.**

Keywords—Quantum annealing; adaptive parallel software/hardware mapping; algorithm-specific processor scheduling; topology-aware network scheduling; optimization

I. BROADEN QUANTUM ANNEALING IMPACT

With increased parallel resource, HPC is encountering greater challenges than ever pushing from petascale toward exascale in power, energy, cost and space. These increasing unmet HPC needs may well be answered by several emerging technologies in quantum, nano, optical, and brain-inspired design. Every HPC user, facilitator, and system software developer is facing constant complex optimization tasks from work-load partition, resource allocation, to network fragmentations and hopping minimization.

As the arrival of Adiabatic Quantum Annealing Computers (QACs) era, there are ample opportunities to search for the missing link between QACs and HPC. In this paper, we will describe the topology-aware parallel processor network scheduling, and propose preliminary novel examples to test the feasibility in applying QAC toward HPC optimization challenges.

II. PARALLEL TASK/PROCESSOR MAPPING

The prime challenge in parallel computing is how best it is to map the parallel software to parallel hardware adaptively (Fig. 1) as per the various inter-processor communication latencies, so that the best performance can be achieved, in terms of speedup, cost efficiency, memory space, power saving, reliability, etc.

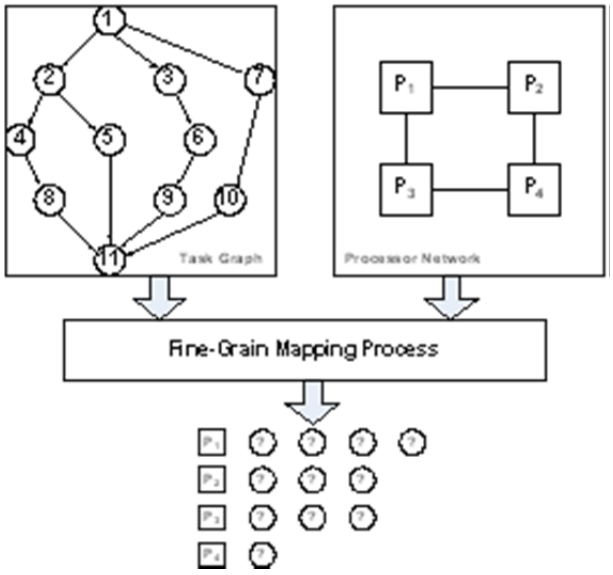

Fig. 1. Quantum optimized processor scheduler.

For example, the performance of using multi-cores within a single cluster blade tends to be faster than adding additional cores across blade boundaries in TACC's massive parallel Sun Constellation Ranger cluster, due to the much longer inter-blade communication than within one blade, see Fig. 2.

In a large example of processor subnet allocation of 4,096 Cray XE6 processor nodes, our 2013 manual CyberShake topology-aware tuning mapped a prism-shaped earthquake wave propagation software computation to a matching hardware prism subnet (see yellow boxes in Fig. 3), achieved a 35% speedup of the fastest seismic hazard wave propagation code on Blue Waters Cray torus cluster [2]. Cray XE6 nodes are carefully assigned avoiding the randomly located non-compute nodes while considering the varied inter-node speed and bandwidth along x, y and z dimensions.

Since processor scheduling problem is a well-known NP-hard problem, we applied a heuristic, such as A* parallel fine-grain match, to generalize software/hardware mapping [1].

Among the example topology-aware software/hardware mappings tested for a given robot elbow controller task graph of irregular shape, our automatic adaptive resource mapper wisely scheduled mostly 0- or 1-hop data transfers on 9 various example network topologies as shown in Fig. 4, with more 1-hop data exchange in lower-diameter topologies for better

parallel speedup. Expensive inter-processor data exchanges with 2 to 6 hops are kept lowest under 10% of total transfers.

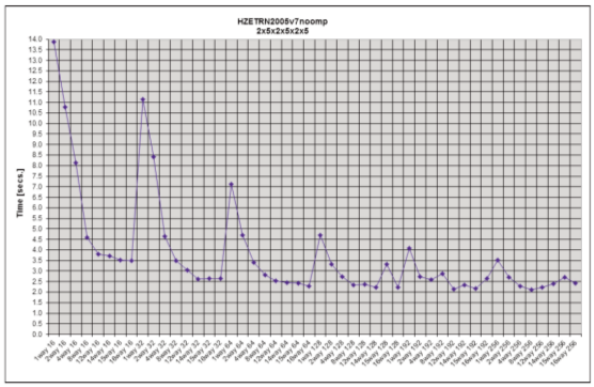

Fig. 2. TACC Ranger Cluster run-time wave, rising at every 16-core inter-blade interval (NASA HZETRN Space Radiation Flux Estimation run by Frank Christiny).

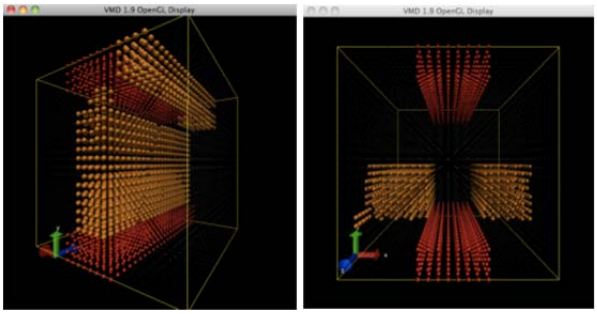

Fig. 3. NCSA cray blue waters cluster processor network topology tuning visualization to minimize network frangmatation [2].

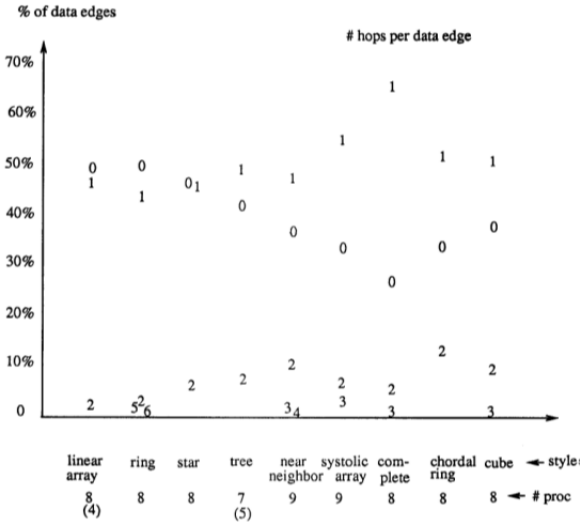

Fig. 4. Inter-PE hops for robot elbow manipulator computation task [1].

Beyond manual and heuristic topology tuning, Quantum Annealing can be incorporated in HPC schedulers, such as SLURM, to automate application-specific cluster subnet topology request and allocation of the HPC cluster resource management system to further minimize latency caused by excessive switch hops due to network fragmentation to expedite HPC performance.

III. Toward Quantum Processor Scheduler

Adiabatic Quantum Computers, made available by D-Wave chips, solve discrete combinatoric optimization problems efficiently by evaluating and comparing many solutions (superposition) within given constraints (entanglement) simultaneously, such as Traveling Salesman Problem (TSP) and Map Coloring problem [3].

We can apply QACs to partition computation workload for heterogeneous HPC processor cores similar to NASA's more complex 10-satellite network cameras scheduling problem with slew constraints to optimize the global image coverage of a given set of targets currently solved on D-Wave system.

We can also request and allocate application-specific processor subnet to a computation workload using D-Wave web browser qubit lattice interface to map the entire target HPC cluster network topology in the qubit lattice with weight values and coupler strengths representing pair-wise processor selection correlation strength.

In conclusion, to broaden the perspective of QACs' impact, we are working on demonstrating the feasibility that QACs possess powerful potential to optimize the software/hardware mapping in general to greatly benefit a wide range of classical HPC computation applications in terms of performance, energy/space efficiency, reliability and scalability.

Acknowledgment

The author sincerely appreciate the collaboration and team support of NASA LaRC, NSF, XSEDE, SDSC, NCSA, TACC, SCEC, CyberShake, Cray and D-Wave System.

References

[1] Shih, Liwen: Adaptive latency-aware parallel resource mapping: task graph scheduling onto heterogeneous network topology. XSEDE 2013: 52 ACM 2013 Article.

[2] Cui, Yifeng, Efecan Poyraz, Jun Zhou, Scott Callaghan, Phillip Maechling, Thomas Jordan, Liwen Shih, Po Chen: Accelerating CyberShake Calculations on the XE6/XK7 Platform of Blue Waters. XScale 2013.

[3] D-Wave: On Demand Webinar: Intro to Quantum Computing and D-Wave, June 17, 2014.