Feasibility Studies on Next-Generation Supercomputing Infrastructures: System Software and Library Research

Auto-tuning for Quantum-Inspired Annealing

Takahiro Katagiri (RIKEN R-CCS / Information Technology Center, Nagoya University,
E-mail: katagiri@cc.nagoya-u.ac.jp)

Aim of the Project

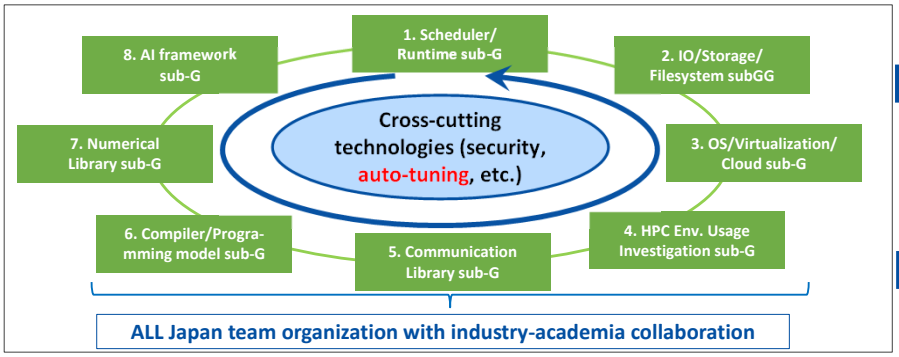
The next-generation computational infrastructure is expected to become a platform for realizing SDGs and Society 5.0 by **providing advanced digital twins** that will bring "**Research DX**" in the science. Aiming to realize a versatile computing infrastructure that can execute entire workflow by making full use of wide range of computational methods, simulation techniques, and BigData at scale, we conduct a holistic investigation on architecture, system software and library technologies through co-design with applications.

As a basic principle of system design, we **practice the "FLOPS to Byte" concept** from architecture development to algorithm or application design to streamline data transfer and computation under power constraints, while taking necessary computing accuracy into consideration. Under the ALL JAPAN team composition, we will investigate system configurations and elementary technologies which improve effective performance of the next-generation computing infrastructure.

System Software and Library Research

- **Investigate technological trend of system software and draw R&D roadmap based on it**
  - Holistic studies on 8 system SW areas, cross-cutting technologies (security, auto-tuning, etc.) and platforming
  - Prioritize SW / tools to develop domestically by considering usage on flagship and 1-tier machines, development cost, proficiency of existing ecosystem, and other factors
- **Study new technological area in system software for industrialization**
  - Investigate new areas of development to encourage use in a wide range of applications

  → Providing a platformed digital twin environment for industrial use, promotion of data utilization, fusion of machine learning & first-principles simulation, advanced large-scale real-time data processing, high security, etc.
- **ALL JAPAN team organization with industry-academia collaboration**
  - Expect future ripple effect on HPCI supercomputer centers



Auto-tuning for Quantum-Inspired Annear and Coherent Ising Machine (CIM)

- **Quantum-inspired computers are being developed.**
  - CMOS Annealing Machine (Hitachi), Digital Annealer (Fujitsu), etc.
- **Parameters to be tuned on CMOS Annealing Machine**

QUBO formula

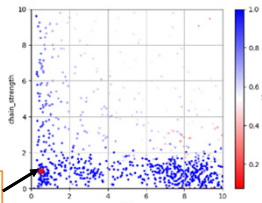
$$H = w_a \sum_{(u,v)\in E} (1-x_u)(1-x_v) + w_b \sum_{v\in V'} x_v$$

(constraint: w_a term; cost: w_b term)

Parameters	Overview
Wa	Coefficient of constraint term
Wb	Coefficient of cost term
chain_strength	Strength of chain

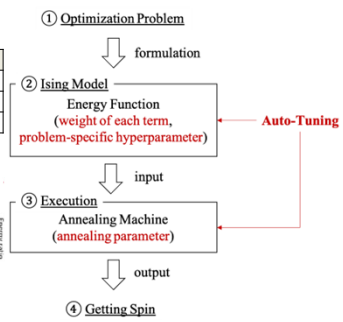
- **Results**

TPE (Tree-Structured Parzen Estimator)



Best Parameter by Optina

- **Procedure of Auto-tuning**



Minimum Vertex Cover Problem
Find V′, which is the vertex covering set (where $|V'|$ is the minimum, Graph $G=(V,E)$, $V' \subseteq V$)

Auto-Tuning (AT) is performed by Optuna.

- **Coherent Ising Machine (CIM)**

CACm (Chaotic Amplitude Control with momentum)

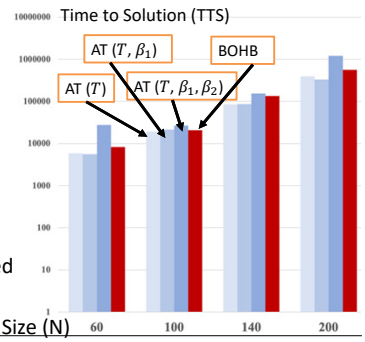
$$x(t+1) = x(t) + \Delta t\left(-\beta(t)x(t) + \alpha e(t) \circ (\Omega\phi(t)) + \gamma(x(t) - x(t-1))\right)$$
$$e(t+1) = e(t) - \zeta e(t) \circ (x(t)^2 - 1)$$

Param	Interpretation	Range
T	Number of time steps	10~1000
β_1	Initial decay rate	0.1~10
β_2	Final decay rate	0.1~10
α	Coupling strength	0.1~10
γ	Momentum term strength	0.1~10
ξ	Rate of change of auxiliary variables	0.001~0.01

Time step size fixed at $\Delta t = 0.5$

Auto-Tuning (AT) is performed by GPTune.

- **Results**



Information Technology Center, Nagoya University
http://www.icts.nagoya-u.ac.jp/en/center/