A feasibility study of quantum annealing for the next-generation computing infrastructure

A Constraint Partition Method for Efficiently Solving Combinatorial Optimization Problems

NUG XXXV 2024

Kazuhiko Komatsu
Tohoku University

13 June, 2024

Agenda

Introduction of feasibility study of quantum computing

- Quantum annealing group by NEC and Tohoku Univ.

Evaluation of annealing machines [QCE23, Komatsu]

- Performance investigations of Quantum and Quantum-inspired annealing machines

A Constraint Partition Method for Combinatorial Optimization Problems [MCSoC23, Onoda]

- Constraint Partion toward large constraint optimization problems

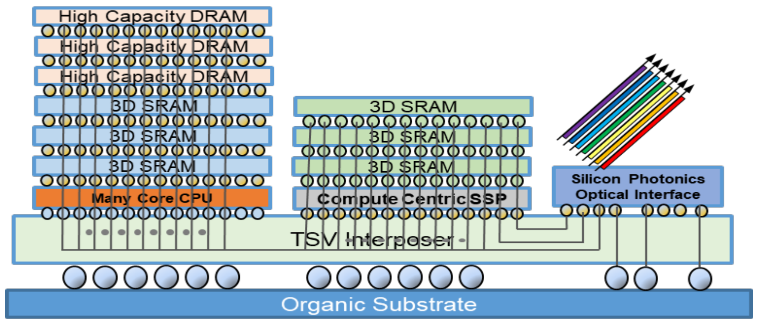
Feasibility studies (2022/08~2025/03)

Overview

- R&D of essential technologies to develop the next-gen. computing infrastructure

System team

- Architecture
- System software
- applications



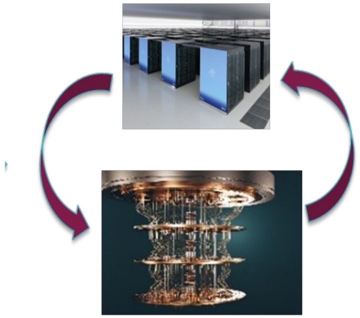
Operation technology

- Operation-related technologies

New computational principles

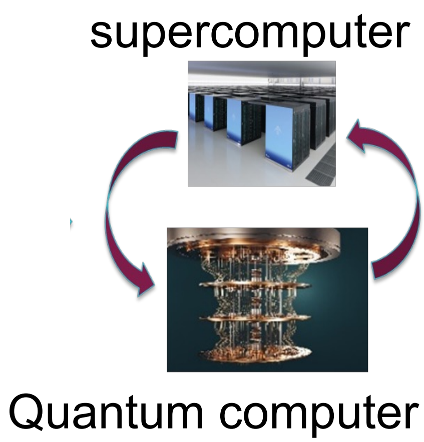
- Quantum supercomputing
  - Hybrid computing by QC, QA, SC

supercomputer



Quantum computer

FS of new computational principals



Overview

- Evaluate the feasibility of "**quantum supercomputing**" by hybrid computing of HPC and quantum computing
  - Study on architecture, system software, and algorithms of quantum supercomputing

Teams

Keio Univ
Architecture for quantum supercomputing

- **Riken R-CCS** — Quantum system software
- **Riken RQC** — Quantum algorithm
- **Fujitsu** — Quantum computer
- **Kyusyu Univ.** — NISQ algorithm
- **Tohoku Univ.** — **Annealing**
- **NEC** — **Annealing**

Activities of Annealing Group

Performance Evaluation of Quantum and Pseudo-Quantum Annealing Machines

- Investigation of Annealing Machines through Performance Evaluation and Analysis
  - Study of Annealing Machines and Their Evaluation Methods
  - Development of Benchmarks for Evaluation
  - Performance Comparison of Various Annealing Machines

Investigation of the Application of Quantum and Pseudo-Quantum Annealing Technologies

- Research and Development Status
- Case Studies of Utilization

Quantum Annealing teams

〇Representative

Tohoku university
Kazuhiko Komatsu

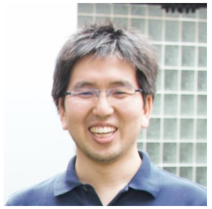
Quantum future society vision
Quantum technology innovation office
Quantum solution

Cyberscience
Center

〇Kazuhiko Komatsu

Mitsuo Yokokawa

GSIS



☆Masayuki Sato

☆Masahito Kumagai
☆Makoto Onoda
☆Kaho Aoyama
☆Huang Chu-Yuan

NEC

〇Shintaro Momose

Hiroshi Chishima

☆ Kotaro Bannai
☆Nakasone

Overview

Introduction of feasibility study of quantum computing

- Quantum annealing group by NEC and Tohoku Univ.

Evaluation of annealing machines [QCE23, Komatsu]

- Performance investigations of Quantum and Quantum-inspired annealing machines

A Constraint Partition Method for Combinatorial Optimization Problems [MCSoC23, Onoda]

- Constraint Partion toward large constraint optimization problems

Varieties of Ising machines

Annealing

- Quantum annealing
  - Analog circuits with quantum effects
    - QA using superconducting quantum circuit by D-Wave Systems, Inc
- Simulated (Quantum-inspired) annealing
  - Use of digital processors such as CPU, GPU, and VE
    - D-wave Neal, Fixstars Amplify Engine, Vector Annealer, and so on
- Dedicated annealing machine
  - Dedicated digital circuits such as CMOS and FPGA
    - Hitachi CMOS Annealer, and so on

Bifurcation

- Bifurcation machines
  - Controlled by the pitchfork bifurcation phenomena
    - Toshiba Simulated bifurcation machine (SBM), NTT Coherent Ising machine (CIM),

Different characteristics and performance

Experimental conditions: Annealing machines

TOHOKU UNIVERSITY

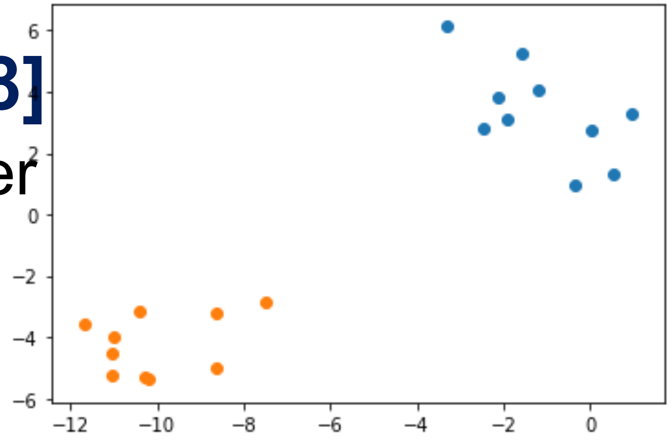
Machines	Hardware	Max # bits	# bits fully	Connectivity	Bit precision	Services
D-wave 2000Q	Quantum circuit QPU	2,048	64	Chimera graph	Analog 5 bits	Cloud
D-wave Advantage	Quantum circuit QPU	5,760	124	Pegasus graph	Analog 5 bits	Cloud
D-wave Advantage2	Quantum circuit QPU	563		Zephyr graph	Analog 5 bits	Cloud
D-wave Leap Hybrid	QPU + Digital circuit	N/A	N/A	N/A	N/A	Cloud
D-wave Neal	CPU	N/A	N/A	Fully	Digital 64 bits	Local
NEC Vector Annealer	VE Type 20B	100,000+	100,000+	Fully	Digital 32 bits	Local
Fixstars Amplify Annealing Engine	Nvidia A100	262,144	131,072	Fully	Digital	Cloud
Hitachi CMOS Annealer	GPU	61,952	176	King graph	Digital 3bits	Cloud
Toshiba SBM	GPUs	10,000,000	10,000,000	Fully	Digital	Cloud

Benchmark: Combinatorial clustering

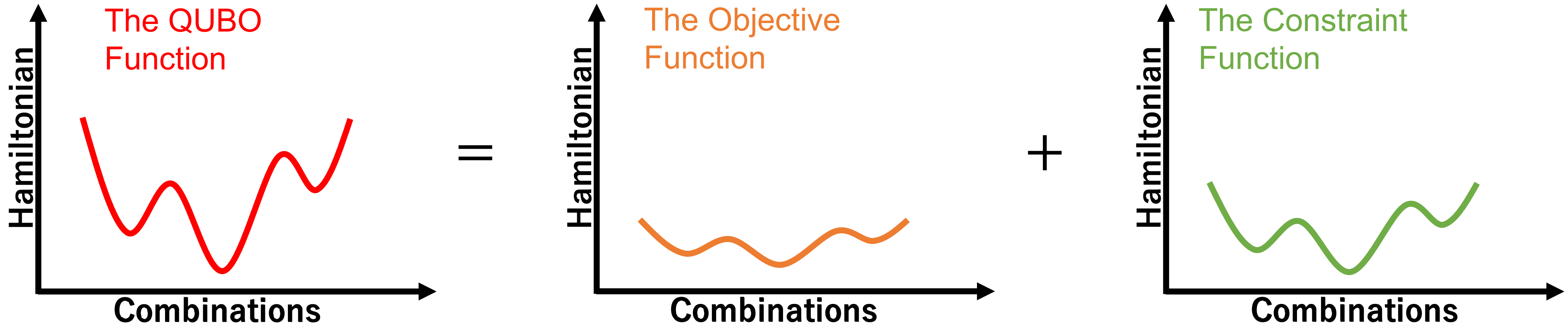
QUBO of Combinatorial Clustering [Kumar, 2018]

- Formulated using the method of the Lagrange multiplier
- $H = \frac{1}{2}\sum_{i,j=1}^{N} d(x_i, x_j) \sum_{a=1}^{K} q_a^i q_a^j + \sum_{i=1}^{N} \lambda_i (\sum_{a=1}^{K} q_a^i - 1)^2$

The Objective Function | The Constraint Function



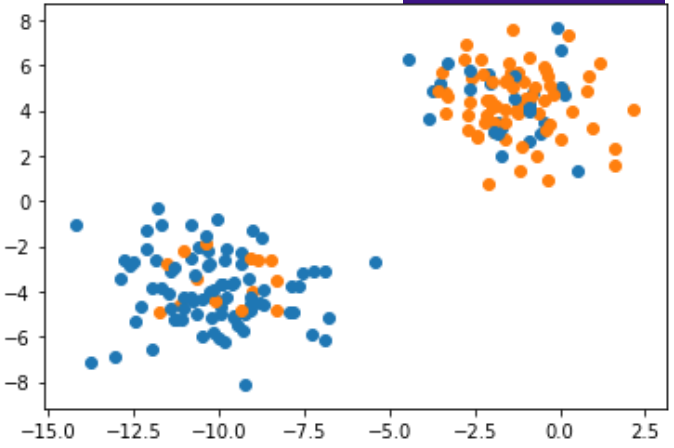
N = 20, One-Hot



Difficulty

Solving problems when size is large

- Bit precision is required for the Lagrange multiplier
- $H = \frac{1}{2}\sum_{i,j=1}^{N} d(x_i, x_j) \sum_{a=1}^{K} q_a^i q_a^j + \sum_{i=1}^{N} \lambda_i \left(\sum_{a=1}^{K} q_a^i - 1\right)^2$

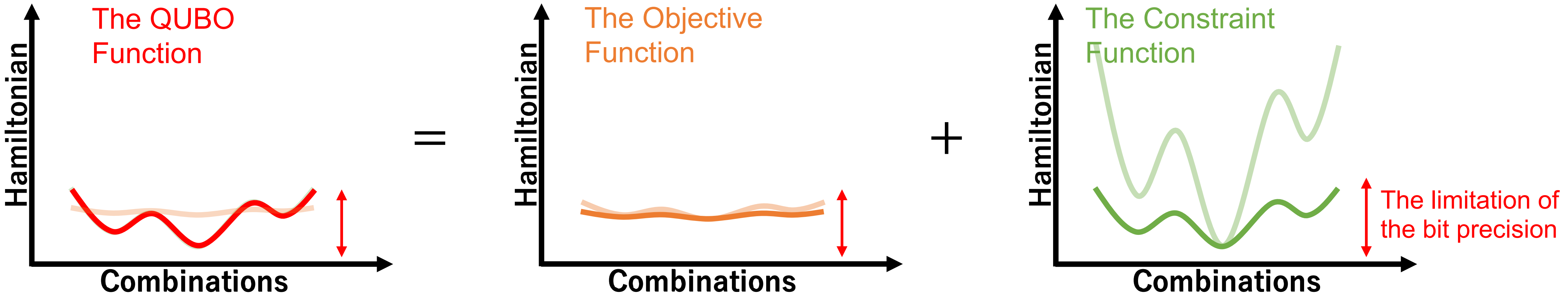


N = 200, One-Hot

The objective function loses its properties

The Objective Function

The Constraint Function



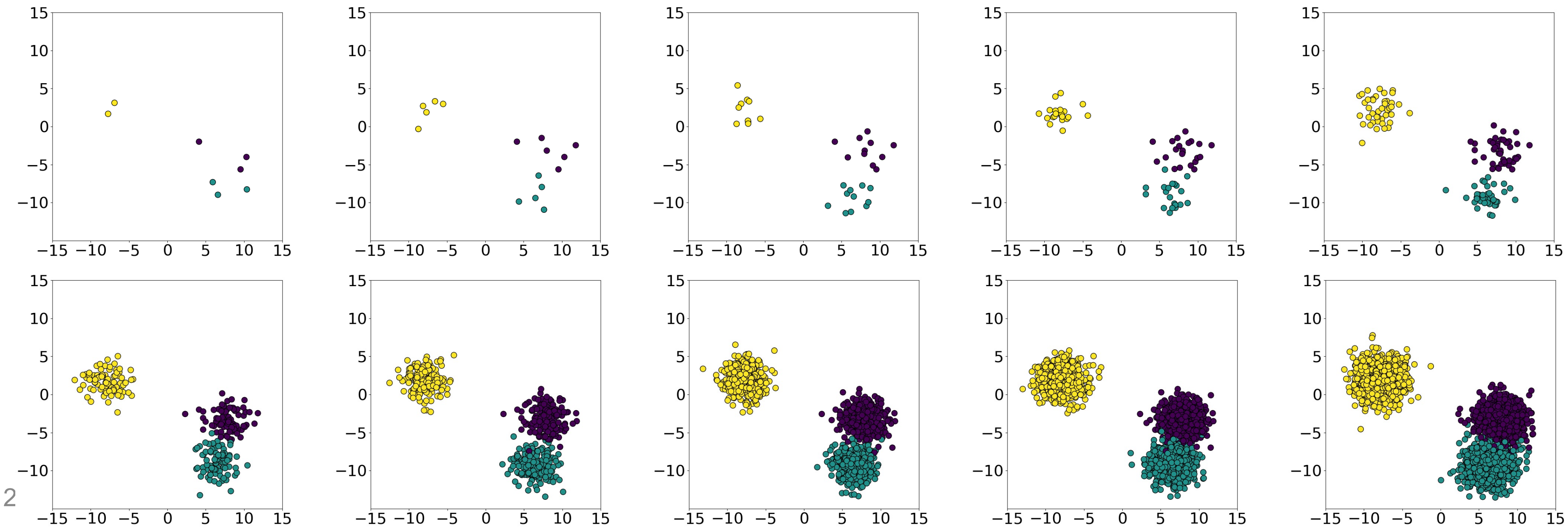
Optimization becomes difficult for problems with a large number of data points due to the large gap between the distance and the Lagrange multiplier (d << λ)

Experimental condition: Dataset

Artificial data

- Number of clusters 3, Number of data 8~4096
- The reference solution: the lowest result obtained among all executions.

Number of trials: 100 for each machine, each data



Evaluation metrics

TTS(Time to solution)

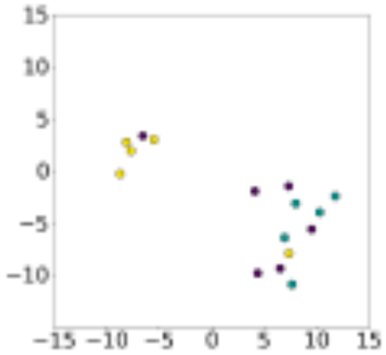
- Execution time to reach the certain precision solution
  - $\mathrm{TTS} = \gamma_{anneal} R + T_{others}$
    - γ_{anneal} : Annealing time
    - R: Annealing times to obtain the reference solution $R = \frac{In(1-p_R)}{In(1-p_{success})}$
    - T_{others}: Time for the other than annealing such as QUBO generation
- The certain precision solution is the answer label

Cost (Accuracy)

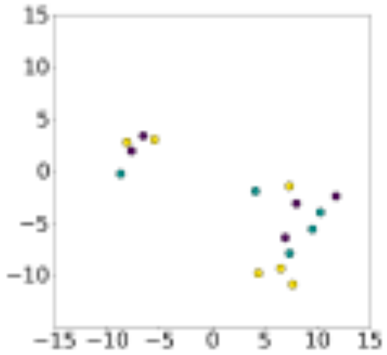
- Sum of distances within the same cluster for all clusters
- The lower the value, the higher the quality of the solution

Execution time

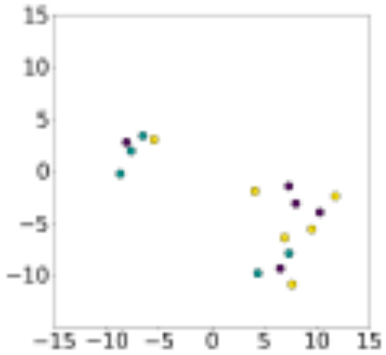
Visualization results (16 data points)



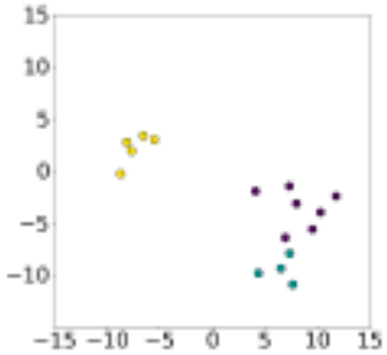
2000Q



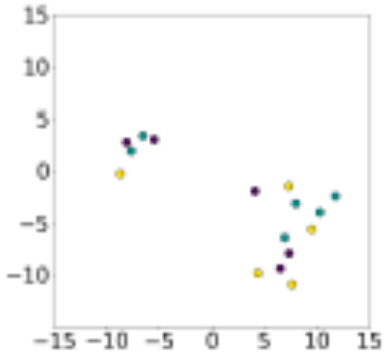
Advantage



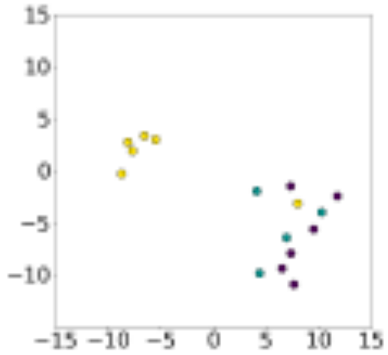
Advantage2 prototype



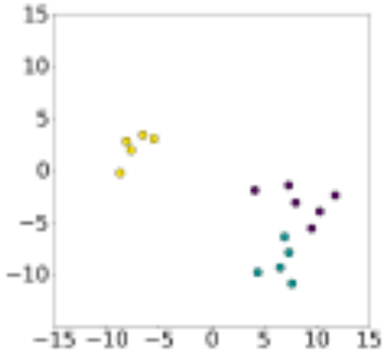
Hybrid



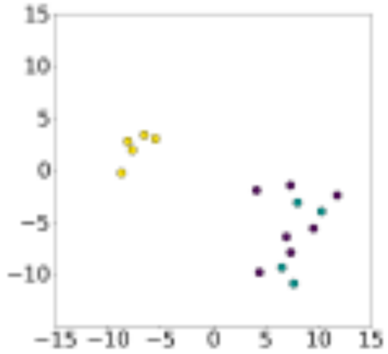
CQM



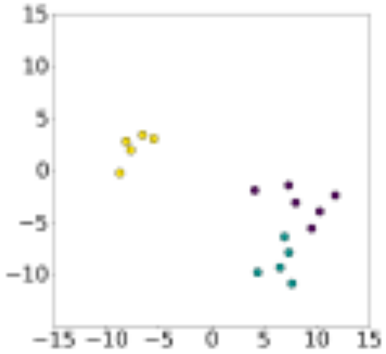
Neal



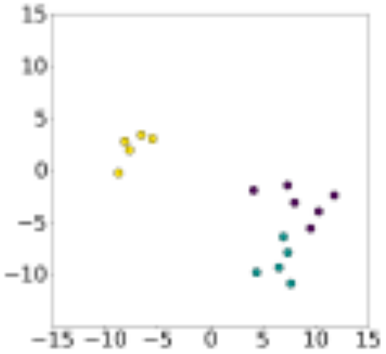
Vector Annealer (External)



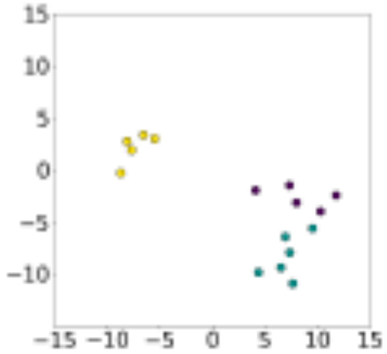
Vector Annealer (Internal)



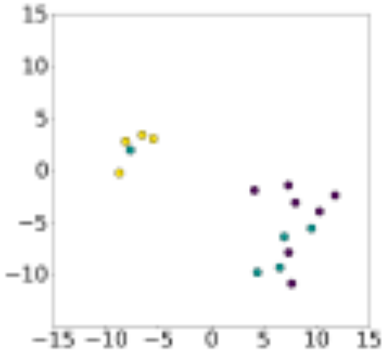
Vector Annealer 2 (External)



Amplify Engine

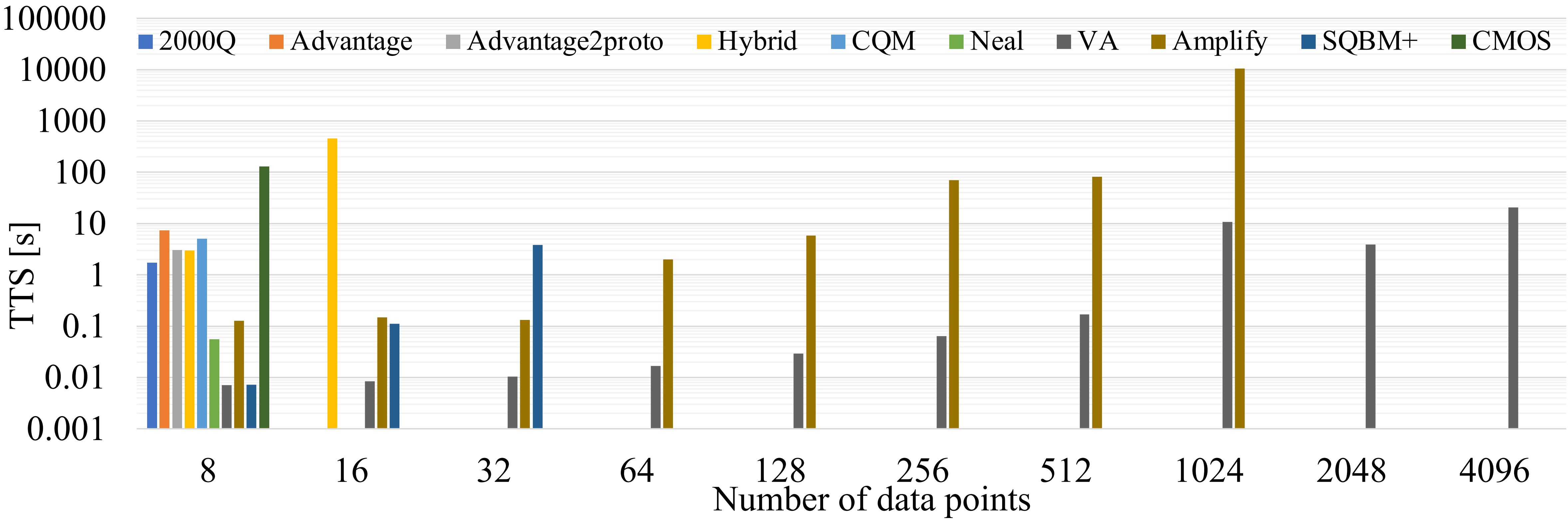


SQBM+



CMOS Annealer

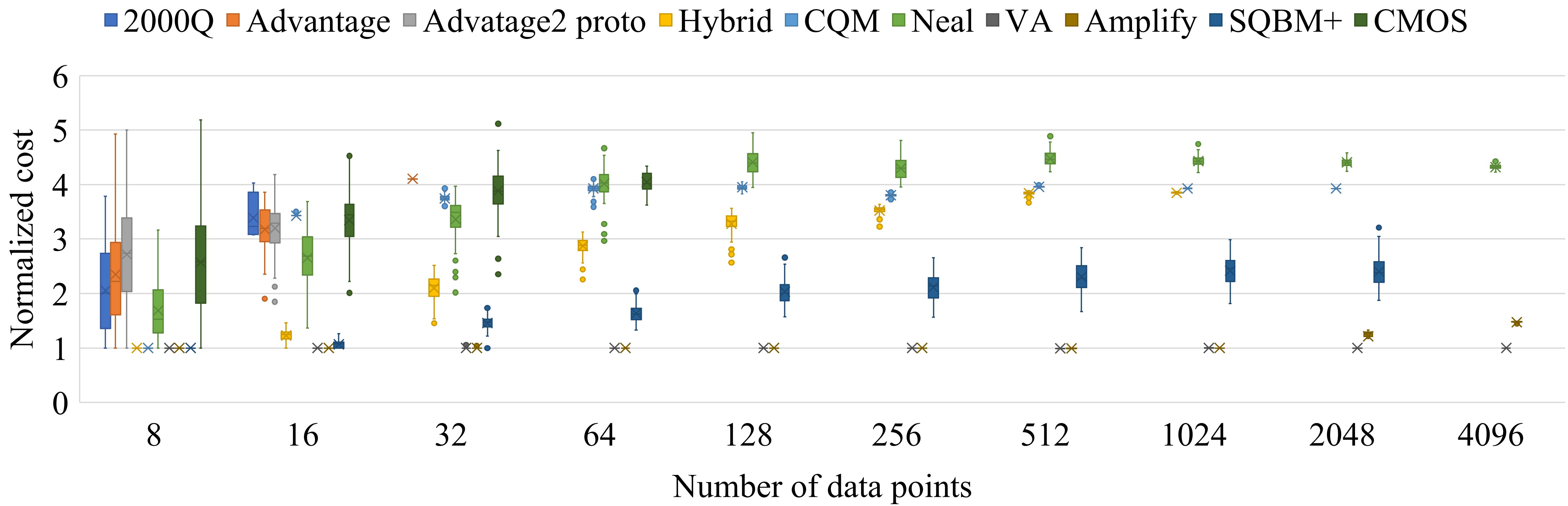
TTS



- VA-Ex < VA-In < AE < Neal < 2000Q < Leap < Advantage < CMOS
  - High accuracy clustering and fast execution time

- TTS cannot be calculated when a large amount of data
  - Insufficient number of qubits (2000Q, Advantage, CMOS)
  - Insufficient bit precision (Leap, Neal, VA-In)

Cost normalized the answer label



- VA-Ex, AE: Cost equivalent to the answer solution
- Others: large variation
- 16 or more data points
  - Do not reach the answer solution
  - No plots due to inability to run by the lack of bits

Summary

Evaluation of Ising machines

- Performance Comparison of Domestic and International Quantum Annealing Machines, Pseudo-Quantum Annealing Machines, and Pseudo-Branching Machines
  - D-wave, NEC, Fixstars, Toshiba, Hitachi
- Evaluation Benchmarks
  - Utilization of Ising Machines for Clustering
    - Problems that become increasingly difficult with larger datasets
- Evaluation Metrics
  - Time to Solution (TTS), Accuracy, Constraint Violation Rate, Execution Time
- Insights
  - Number of Bits, Bit Precision, Connection Methods, Mechanisms and Capabilities for Escaping Local Minima

Agenda

Introduction of feasibility study of quantum computing

- Quantum annealing group by NEC and Tohoku Univ.

Evaluation of annealing machines [QCE23, Komatsu]

- Performance investigations of Quantum and Quantum-inspired annealing machines

A Constraint Partition Method for Combinatorial Optimization Problems [MCSoC23, Onoda]

- Constraint Partion toward large constraint optimization problems

Constraints Problems

Constraints in combinatorial optimization problems

- Constraints integrated into the Hamiltonian
- Increasing the Hamiltonian when constraints are violated

The constraint function have an excessive influence

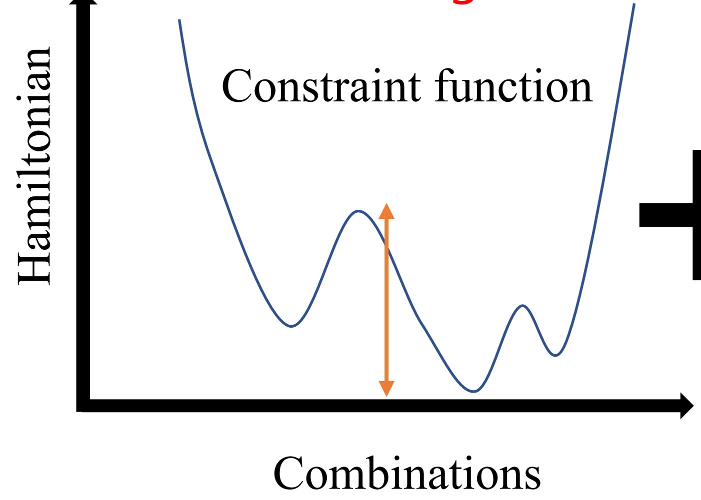
- Difficulty in reducing H of the objective function

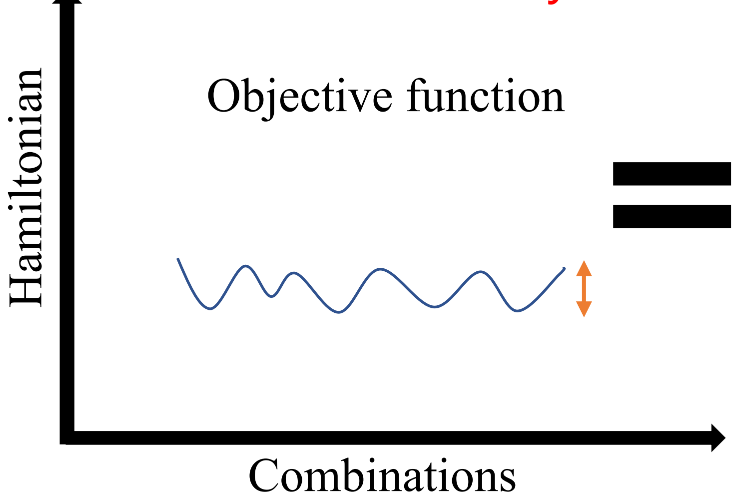
⇨ Resulting in a decrease in the solution accuracy

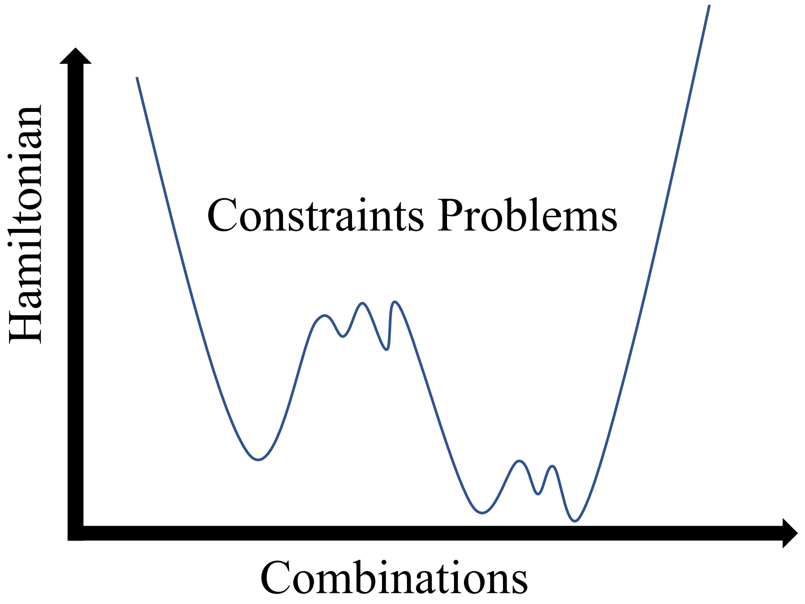
constraint function

$$H = \lambda \sum_{k} C_k + \sum_{i,j} Q_{ij} x_i x_j$$

Setting penalty coefficients λ enough **large**







Objective and Approach

Objective

- To improve the solution accuracy for constraint problems using Ising machines

Approach

- Partitioning a constraint function into terms
  - Assigning small penalty coefficients

A Constraint Partition Method

Partitioning constraint functions to reduce the Hamiltonian

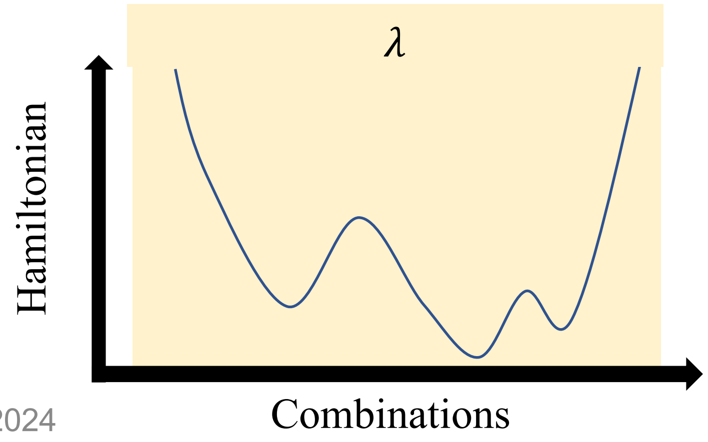
- Setting different coefficients λ_k for each term of constraints

The influence of the constraint function is reduced

→ **Improvement of the solution quality**

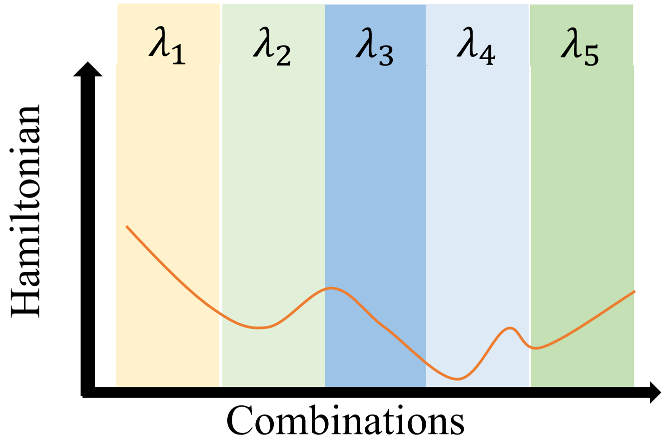
$$H = \lambda \sum_k C_k + \sum_{i,j} Q_{ij} x_i x_j$$

Conventional Constraint function



$$H = \sum_k \lambda_k C_k + \sum_{i,j} Q_{ij} x_i x_j$$

Proposal Constraint function



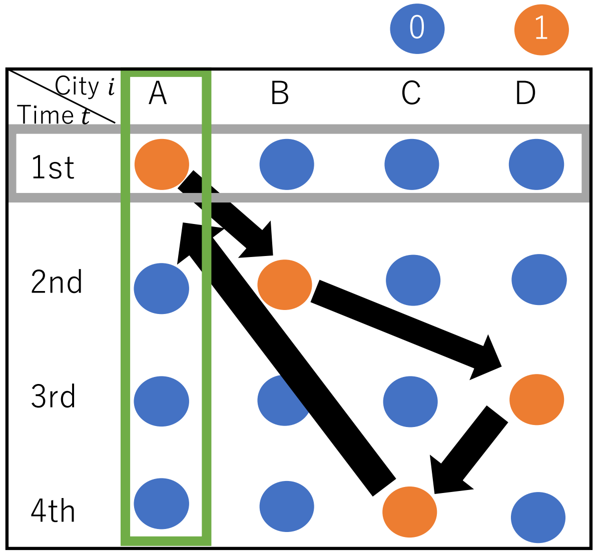
An example using TSP

Partitioning the constraint function to apply the different penalty coefficients for each city

- Setting λ_i^c for the i-th city

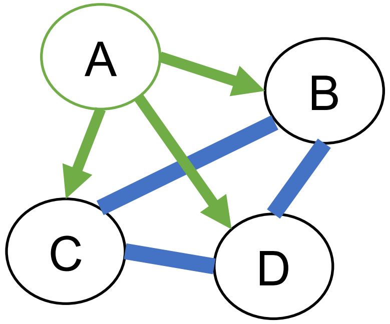
A city constraint for City A

- No need to consider for distances between cities B–C or B-D, C-D
- the distance between city A and the other cities
- A city constraint function has a different penalty coefficient from those of the other cities
  → The city constraints should be partitioned



A city constraint   A time constraint

$$H = \sum_{i=1}^{N} \lambda_i^c \left(1 - \sum_{t=1}^{N} x_{i,t}\right)^2 + \lambda^t \sum_{t=1}^{N} \left(1 - \sum_{i=1}^{N} x_{i,t}\right)^2 + \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{N} d_{i,j} x_{i,t} x_{j,t+1}$$



$$\lambda_{i=i_0}^c = \max(d_{i=i_0,j}), \lambda^t = \max(d_{i,j})$$

Experimental environments

Ising machines

- Fixstars Amplify AE
  - Nvidia V100

Datasets

- TSPLIB, burma14, bays29, eil51, eil76

Number of trials

- 100

The metric

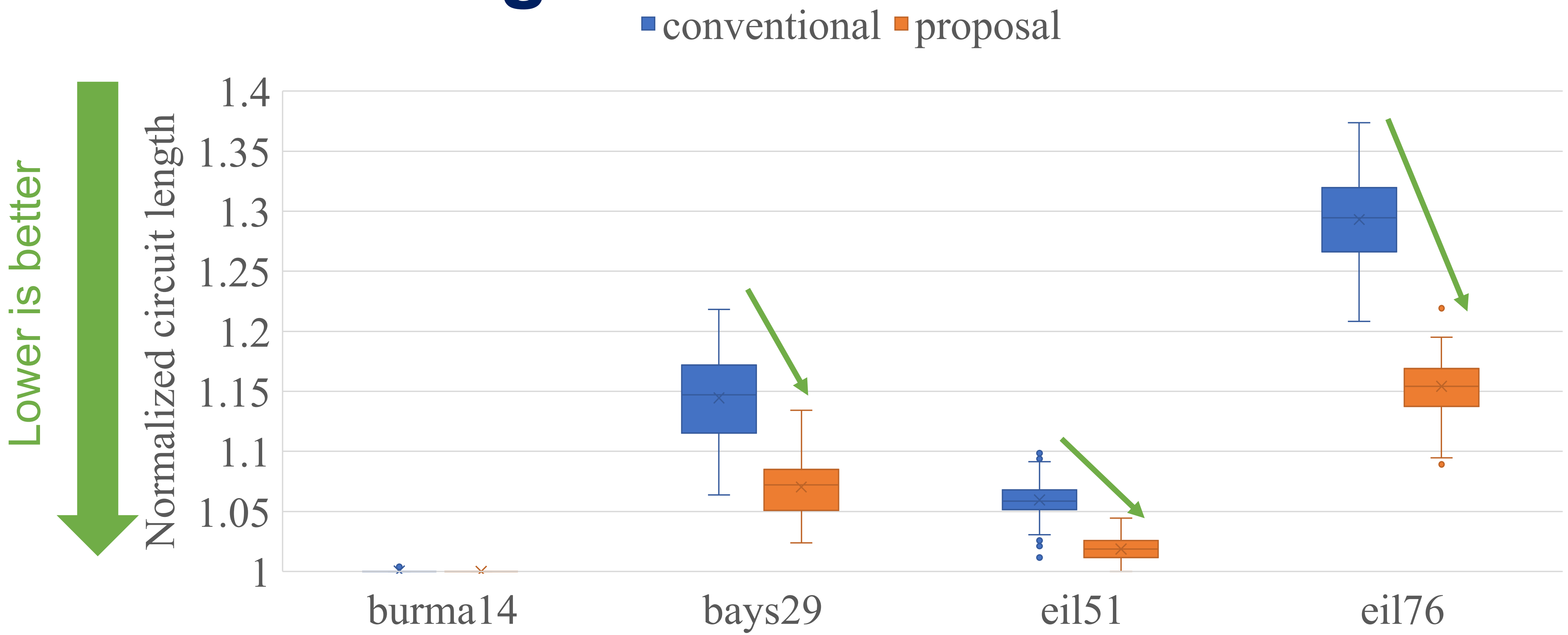
- Total distance of the route

Parameters

- Timeout
- Constraint relaxation rates

	burma14	bays29	eil51	eil76
timeout	10	100	1000	1000

Circuit length of TSP



- The proposed method achieves a shorter circuit length

Summary

Constraint partition method

- To improve the solution accuracy of constraint problems using Ising machines, partitioning a constraint function into terms
  - Assigning small penalty coefficients
- Experimental results shows that the proposed method significantly improves the quality of solutions

Conclusions

Feasibility study of quantum computing

- Quantum annealing group by NEC and Tohoku Univ.

Performance Evaluation of Ising machines

- Performance investigations of Quantum and Quantum-inspired annealing machines

A Constraint Partition Method for Combinatorial Optimization Problems

- Constraint Partion toward large constraint optimization problems