Challenging Collaborations with T-QARD

Graduate School of Information Sciences,
Tohoku University

Masayuki Ohzeki

TOHOKU
UNIVERSITY

Our update

- T-QARD: Tohoku university Quantum Annealing Research and Development
  - Starts from October 2017, make education and active research networks of QA
  - From this April 2019, start more prominent collaborations with many companies

- Jij Inc. is established as an achievement of JST-START project
  - CEO: Yu Yamashiro and CTO: Koji Nishimura
  - Advisors: Masayuki Ohzeki and Masamichi J. Miyama

- D-Wave 2000Q and Pegasus will be (we wish) installed
  - From this April 2019, we sign a contract with D-Wave Systems
  - We establish a consortium (head: Prof. H. Nishimori) for setting the D-Wave machine in Japan
    - Various companies joined this consortium already and gather more colleagues

Agenda

- T-QARD: Tohoku university Quantum Annealing Research and Development
  - Quantum annealing for deep neural network (QML-BQT, Spain)
  - Tsunami evacuation (Qubits Europe in Munich)
  - Quantum Clustering and many others (AQC2018 in Mountain View)
  - Automated Guided Vehicles (Qubits North America in Noxville)

- Our collaboration"s” with T-QARD
  - Forecasting stock attractiveness by Masaya Abe (Nomura Asset Management)
  - Graph partition in quantum simulation by Takako Mashiko (Kyocera)
  - Bus Scheduling Problem by Naoki Maruyama (Tohoku・Hachinohe high school)
  - Black-box optimization by Ami Koshikawa (Tohoku・DENSO)
  - Calibration for Auto-Transmission Shift Control System by Miyama (Tohoku・Aisin)

forecasting stock relative attractiveness

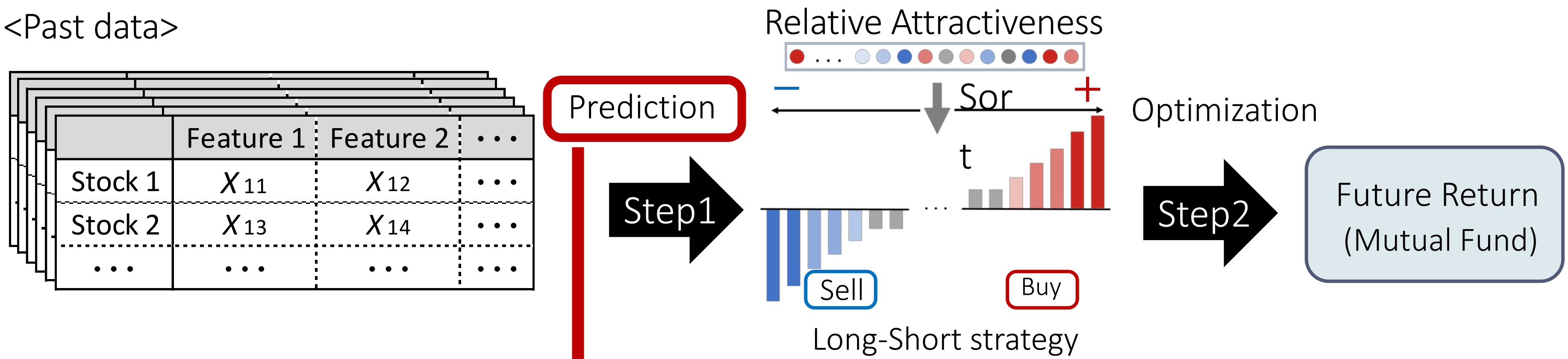
Masaya Abe
Nomura Asset Management Co. Ltd.
T-QARD collaboration with Nomura Asset Management Co. Ltd.

A sampling technique of the D-Wave to implement Restricted Boltzmann Machine for forecasting stock relative attractiveness

Masaya Abe[1] , Masayuki Ohzeki[2,3], Masamichi Miyama[2]

[1] Innovation Lab, Nomura Asset Management Co., Ltd.
[2] Graduate School of Information Sciences, Tohoku University
[3] Institute of Innovative Research, Tokyo Institute of Technology

Motivation



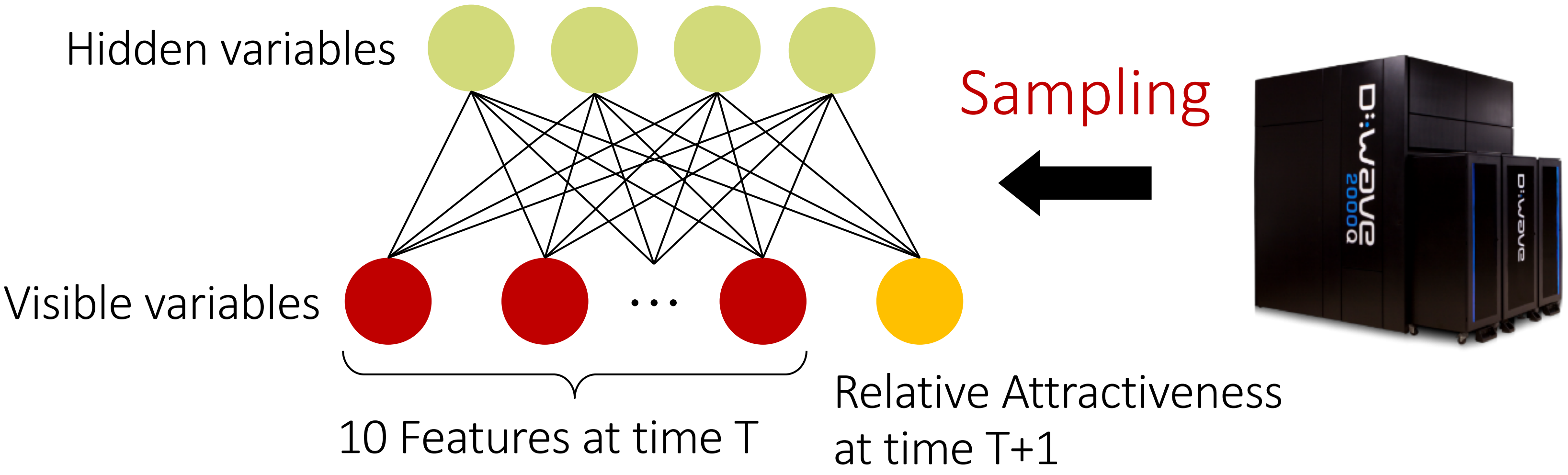
Cross-sectional approach via Restricted Boltzmann Machine

Our Model: Restricted Boltzmann Machine

NOMURA

TOHOKU UNIVERSITY

- Restricted Boltzmann Machine



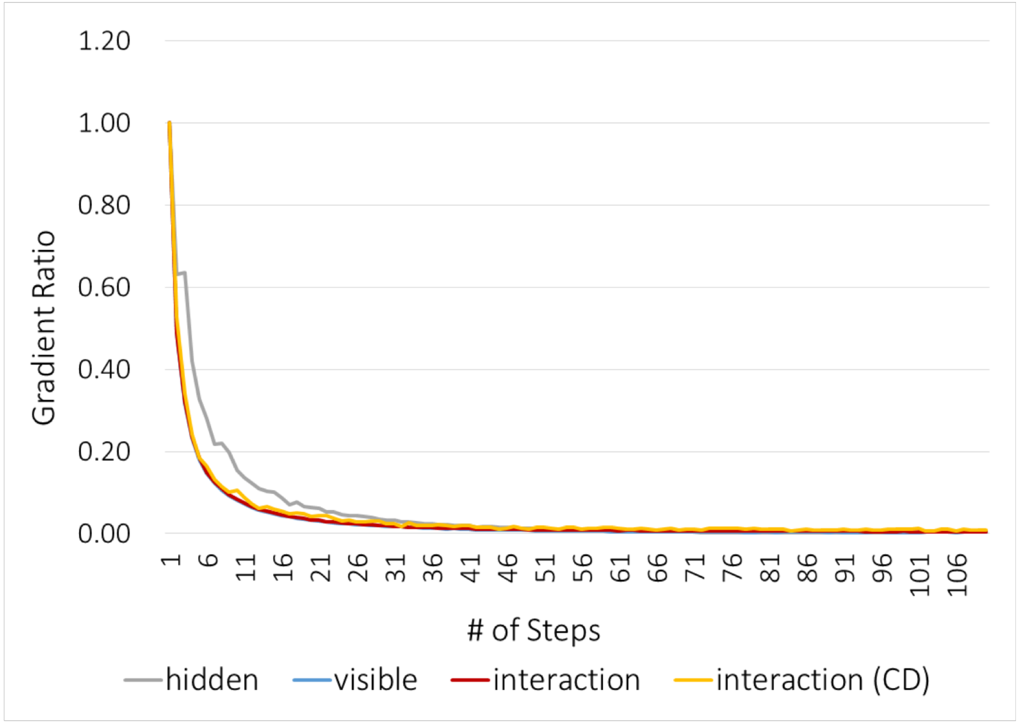
Our Model: Input data sets

- Input : Past data sets
  - ✓ {Features of each stock at time T , Relative attractiveness of each stock at time T+1※}
    ※Only used for training
    - Preprocessing for cross-sectional data:
      Converting each feature and stock returns to {0, 1}
      by cross-sectional median within stock universe
      at each time point
  - ✓ # of features : 10
  - ✓ Stock universe : TOPIX500 (approximately 500 stocks)

No.	Feature (Factor)
1	Book-to-market ratio
2	Earnings-to-price ratio
3	Dividend yield
4	Return on equity
5	Return on asset
6	Current ratio
7	EPS Revision(1 month)
8	EPS Revision(3 months)
9	Past stock return(1 month)
10	Past stock return(12-1 months)

Tentative Result: Convergence in the learning step

- Gradient Ratio
  - Past data sets: 120 months (Dec 2008 – Nov 2018)

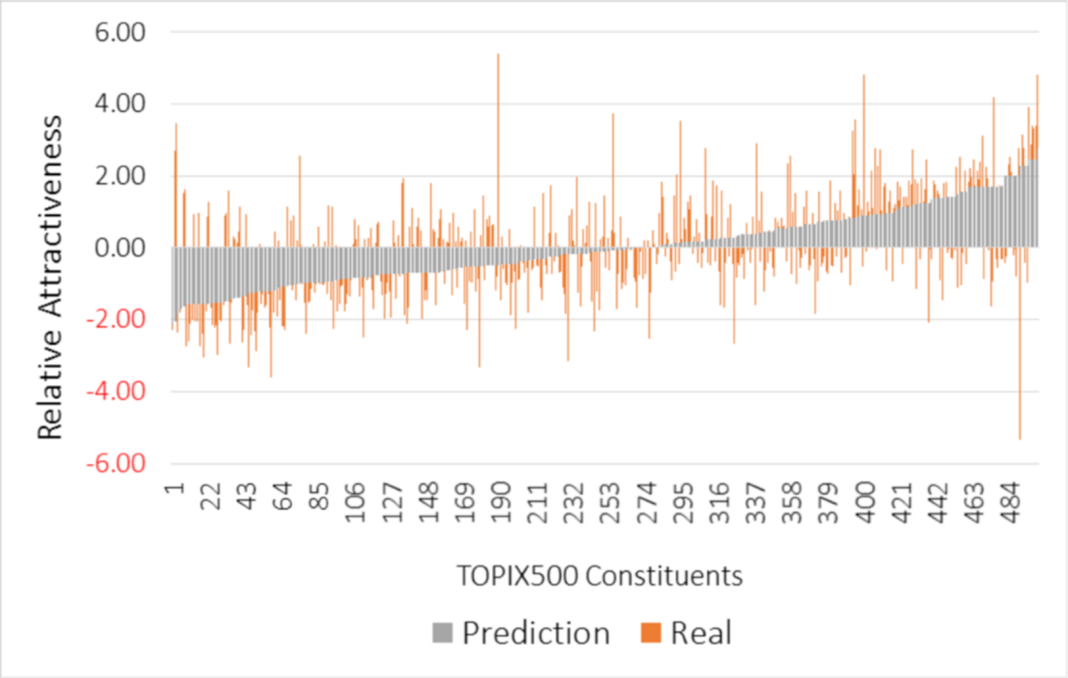


※CD: Contrastive Divergence

Tentative Result: Back test for Single Period

- Performance from Jan 2019 to Feb 2019 (Fixed model)

  - ✓ Rank Correlation: +0.14 (Spearman)
  - ✓ Performance(avg): +1.35% (Long – Short: Quintile)



Ascending Order by Prediction →

Discussion

- Done

  Single period prediction →

- Next

  Performance evaluation by sequential updates

- Future

  Multiple period predictions:

  Expand output variable from one time point (T+1) to multiple time points (e.g. T+1, T+2, T+3)

Graph partitioning

Takako Mashiko

Kyocera

T-QARD collaboration with Kyocera

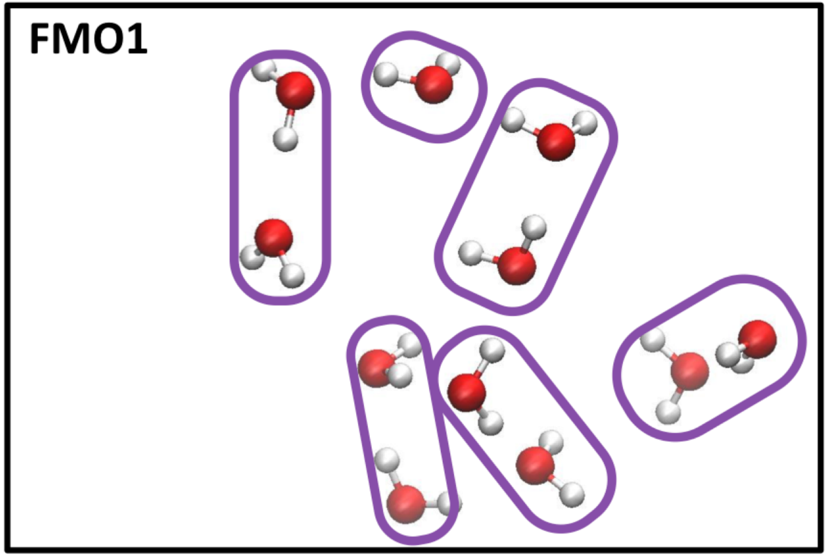
Graph partitioning based on pair interaction energy using quantum annealing on the D-wave system

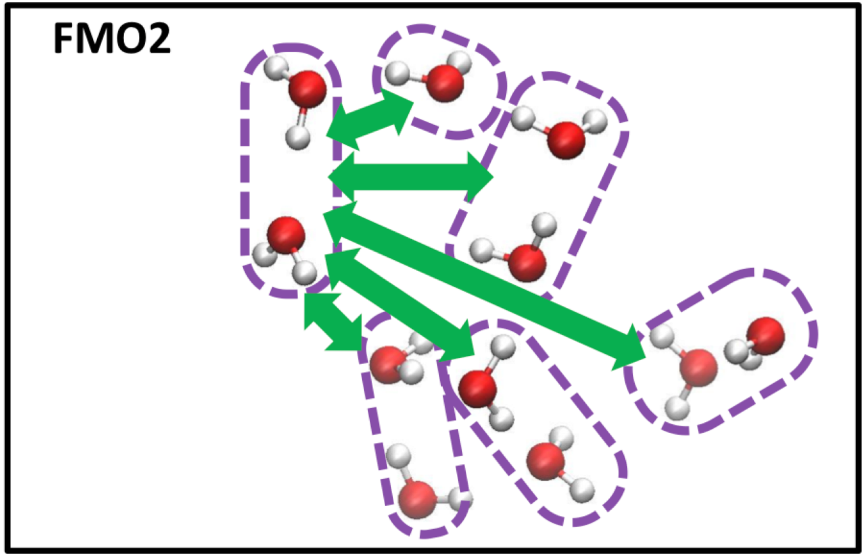
KYOCERA Corporation

MAR. 27th, 2019

1. Intro. | What Fragment Molecular Orbital method ?

Fragment Molecular Orbital (FMO) method[1]





$$E^{\mathrm{FMO1}} = \sum_{I} E_I$$

fragment

$$E^{\mathrm{FMO2}} = \sum_{I} E_I + \sum_{I>J} \Delta E_{IJ}$$

fragment   fragment pair

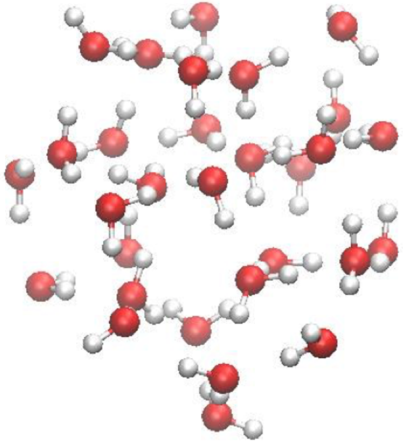
fragment : I, J
fragment pair : IJ
$\Delta E_{IJ} = E_{IJ} - E_I - E_J$

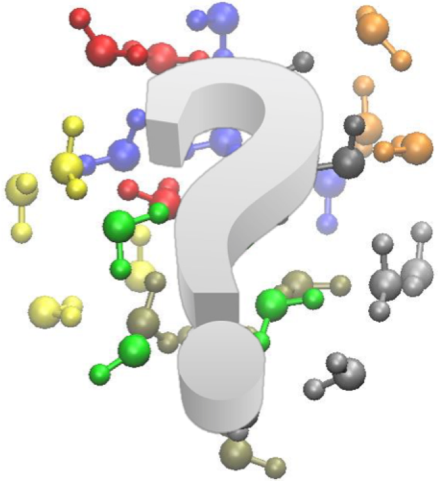
How to divide the fragments is important, but it relies on experience.

[1] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayashi, *Chem. Phys. Lett.*, 1999, 313, 701

2. Purpose

Replacing the fragment partitioning of FMO method with a graph partitioning problem, realizing automated fragmentation.



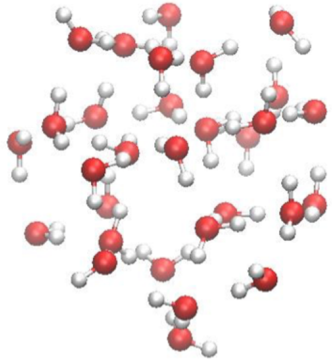


The fragmentation of water clusters was calculated by D-wave, and compared with the results of METIS.

3. Computational detail | flow chart of analysis

Flow chart

1. Prepare for the model of the molecular structure/assembles
2. Perform FMO2 as one molecule one fragment
3. Get the pair interaction energy (PIE)
4. Perform the graph partitioning considering the PIE as edge weight
5. Evaluate (energy difference between FMO2 and QM without fragment)



Step 1, 2, and 3

Software	:GAMESS 2018
Method	:HF
Basis set	:6-31G*
Solvation model	:PCM (water)

Step 4. Perform the graph partitioning

Considering the minimization of the Ising Hamiltonian.

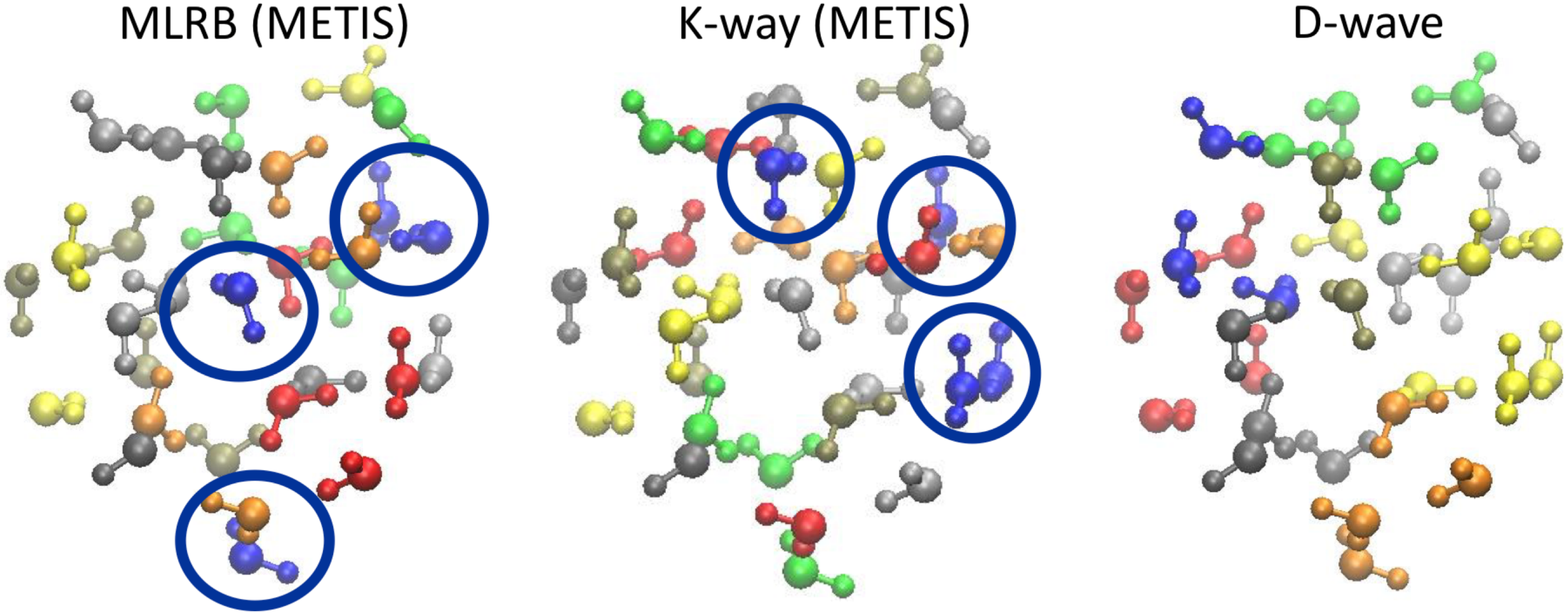
$$H^{1st} = \sum_{i<j} J_{ij}\sigma_i\sigma_j \quad \text{where } \sigma_i = \{-1,1\}$$

Step 5. Evaluate (FMO2 and QM without fragment)

Software	:GAMESS 2018
Method	:HF-D3, MP2, B3LYP-D3, wB97XD
Basis set	:6-31G*
Solvation model	:gas phase, PCM (water)

4. Results | method dependency of 8 cut problem

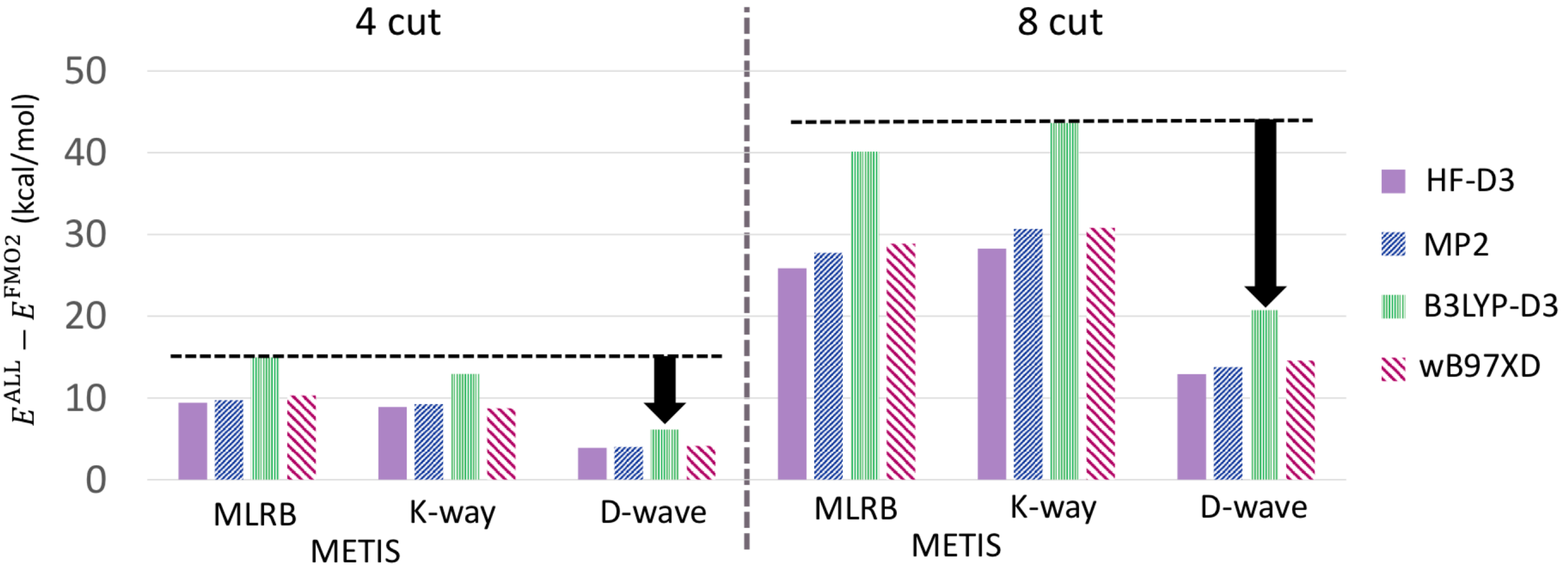
Water molecules are colored by group (green, blue, red, yellow, orange, black, gray, khaki).



For the MLRB and K-way methods, the blue group is spread wide apart.
On the other hand, the results for D-wave shows that molecules in each group are kept close together.

The partitioning using D-wave seems most reasonable.

4. Results | Energy difference between FMO, E^{FMO2}, and QM without fragmentation, E^{All}



For both 4 cut and 8 cut problems,

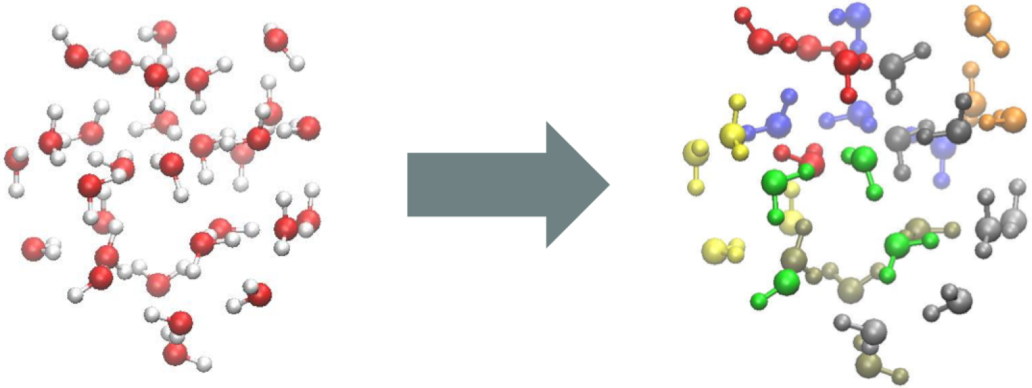
- ✓ The best graph cutting method was D-wave.
- ✓ D-wave energy difference results are half the value of the other methods.

D-wave results proved to be the best choice for automated fragmentation.

5. Conclusion

Replacing the fragment partitioning of FMO method with a graph partitioning problem, realizing automated fragmentation.

The fragmentation of water clusters is calculated by D-wave, and compared with the results of METIS.



When we performed graph partitioning considering the pair interaction energy as edge weight, D-wave results proved to be the best choice for the fragmentation.

Bus Scheduling Problem

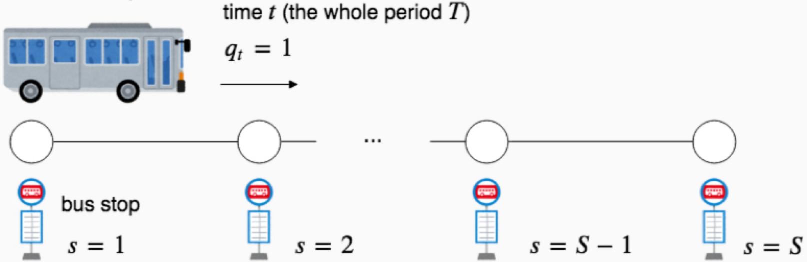
Naoki Maruyama
Tohoku University
T-QARD collaboration with Hachinohe high school

Bus Scheduling Problem

- We addressed optimization of bus schedule using D-Wave machine with Hachinohe high school students.
- In Hachinohe, the buses they often take are crowded every morning. So we have wanted to reduce the congestion of buses.
- The purpose of this problem is to efficiently carry passengers and reduce the total number of buses.
- We evaluate the bus transportation by average number of passengers and EWT(Excess Waiting Time).

Problem setting

- For simplicity, all buses proceed without delay after leaving the first bus stop.
- $q_t = \begin{cases} 1 & \text{if the bus departs at time } t \\ 0 & \text{if it does not} \end{cases}$

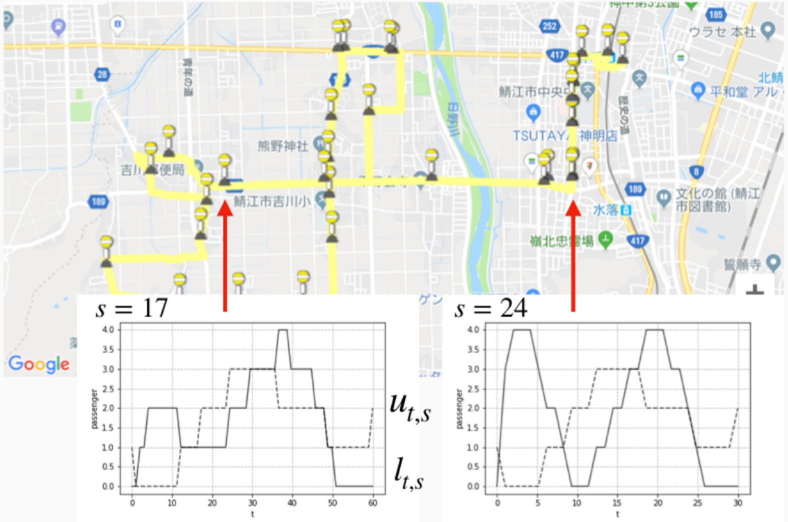


(Source: http://8-bus.com/ohashi-loop.html)

Bus Scheduling Problem

Variables

- C_t : the target number of passengers (uniformly $C_t = C = 21$)
- $l_{t,s}$: the number of passengers getting on the bus departing at each stop s at t
- $u_{t,s}$: the number of passengers getting off at t each stop s
- We used the real data on the passengers in Sabae city, Fukui prefecture. (Source: https://fukuno.jig.jp/app/bus/busgraph.html)



$T = 30$

$S = 32$

optimize →

$(s = 1)$

Answer	Departure time
$q_1 = 1$	7:00
$q_2 = 0$	~~7:10~~
$q_3 = 1$	7:20
	...
$q_{T-1} = 0$	~~11:50~~
$q_T = 1$	12:00

QUBO formulations

Constrained formulation

- $\sum_{t=1}^{T} q_t$: to reduce the number of operating buses

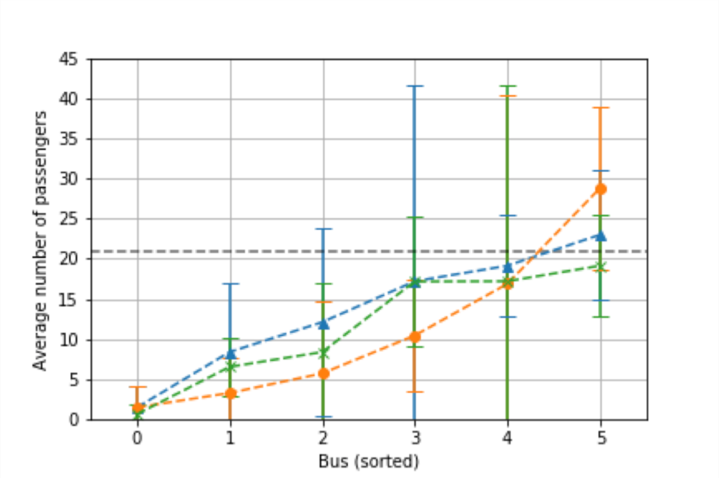
- $\frac{1}{S}\sum_{s=1}^{S}\sum_{s'=1}^{s}\left(l_{t,s'} - u_{t,s'}\right) q_t = C$: to maintain average number of passengers at approximately C for each bus departing at t
- $\sum_{u=1}^{N}\left(1 - q_{t-u}\right) < N$: not to permit the absence of bus transportation from $t-1$ to $t-N$ in order to reduce EWT (N : the capable waiting time)

Unconstrained (QUBO) formulation

$$E(\mathbf{q}) = \sum_{t=1}^{T} q_t + \lambda_1 \sum_{t=1}^{T}\left(\frac{1}{S}\sum_{s=1}^{S}\sum_{s'=1}^{s}\left(l_{t,s'} - u_{t,s'}\right) q_t - C\right)^2 + \lambda_2 \sum_{t=1}^{T} q_t \sum_{n=1}^{N}\left(1 - q_{t-n}\right)$$

Result 1: vs equal intervals and real schedule

- The optimal solution is more efficient of each bus operation than the real schedule and the case considering equal intervals.
- The variance of the optimal solution is slightly larger than that of the others.



▲: optimal solution

●: equal intervals

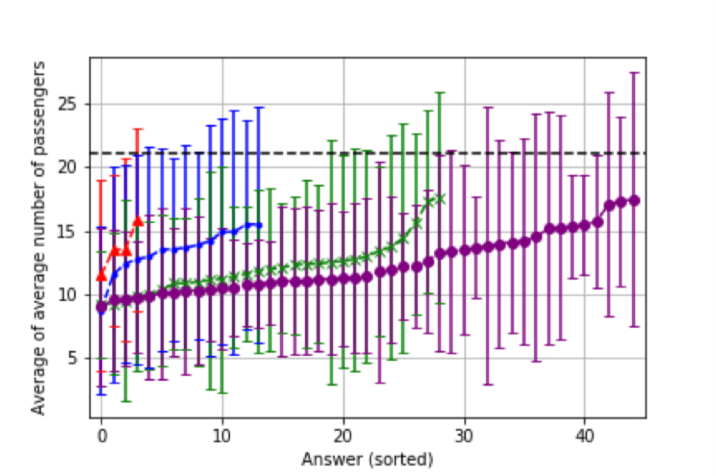
×: real schedule

Notes

- We set $\lambda_1 = 0.5$, $\lambda_2 = 1.2$.
- We used D-Wave 2000Q and set annealing time: 20 micro sec and num_reads: 1000.
- We selected the solutions which satisfy the hard constraint and include the same number of buses as real schedule.

Result 2: optimization for various EWT

- The larger the capable waiting time N is, The more the number of solutions which satisfy the hard constraint is.
- However, those average numbers of passengers gradually decrease.



▲: $N = 5$

●: $N = 6$

×: $N = 7$

●: $N = 8$

Notes

- We set $\lambda_1 = 0.5$,$\lambda_2 = 1.2$.
- We used D-Wave 2000Q and set annealing time: 20 micro sec and num_reads: 1000.
- We selected the solutions which satisfy the hard constraint.

Black-box optimization

Ami Koshikawa

Tohoku University

T-QARD collaboration with DENSO

BLACK-BOX OPTIMIZATION [arXiv:1806.08838]

- Bayesian approach to combinatorial global optimization

$$\vec{x}_i \longrightarrow \boxed{f(\vec{x})} \longrightarrow E_i$$

$\vec{x} \in \{0,1\}^N$

└A large evaluation cost (e.g. experiment)

Find a global optimizer of $f(\vec{x})$

<u>Bayesian Optimization</u>

$\{\vec{x}_i, E_i\}_{i=0,1,\dots}$: Data

$\underline{f_\alpha(\vec{x}) = \vec{x}^{\mathsf{T}} \alpha \vec{x}}$: Acquisition function w/ a horseshoe prior over α

Our work: D-Wave 2000Q, Fujitsu Digital Annealer

→Potential to solve any combinatorial optimization problems

RESULTS (Ferromagnetic Interaction)

- N: 10 spins, # of initial data: 10, $H = -\sum_{i,j,k} J_{ijk}\sigma_i\sigma_j\sigma_k - \sum_i h_i\sigma_i, \quad \sigma_i = \{-1, 1\}$
- The convergence is observed even if the blackbox has 3 body interaction!

Calibration for Auto-Transmission Shift Control System

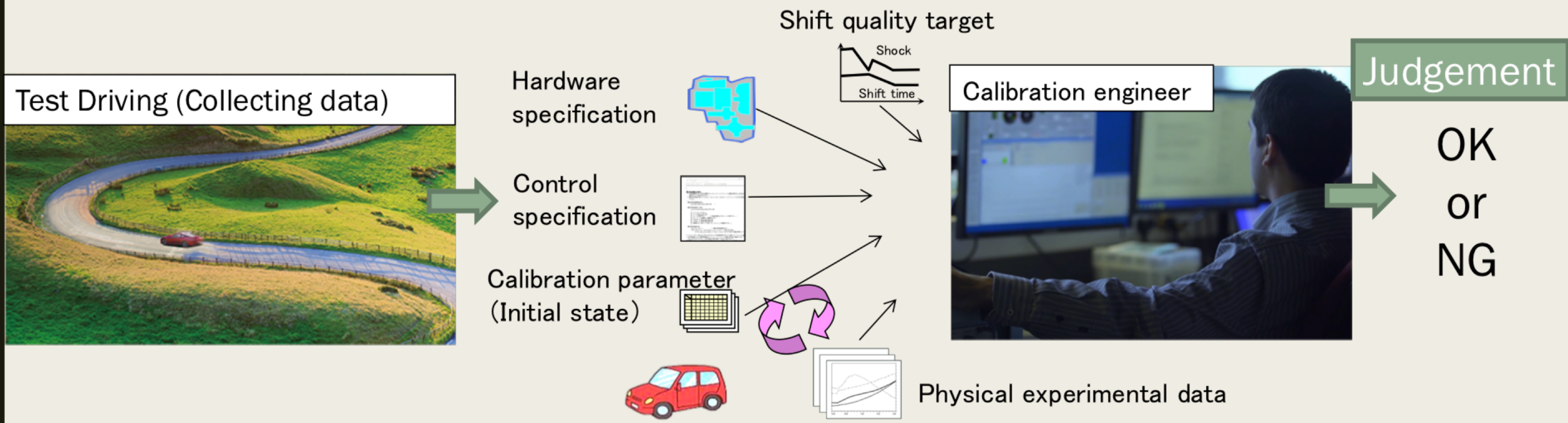
Masamichi J. Miyama
Tohoku University
T-QARD collaboration with AISIN AW

Q-BOOST BASED CALIBRATION TEST FOR AT SHIFT CONTROL SYSTEM

AISIN AW Co., LTD., Tohoku University*

Kiyohisa Tomita, Yasuhiko Kobayashi, Masamichi J. Miyama*

Our Motivation: Calibration test for AT shift control system

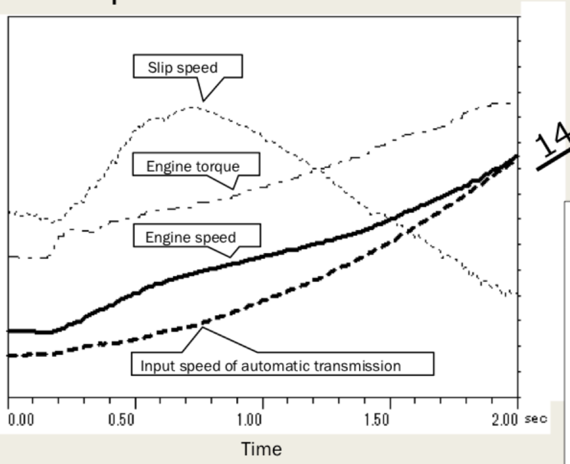


Only well-trained calibration engineer can judge whether the settings is OK or NG.
It takes **several years** to train such an expert engineer!

➡ Binary classifier that can decide OK or NG from a given data.

Previous study by AISIN-AW (DL)

Time series measurement data of Lock-up clutch



14 signals

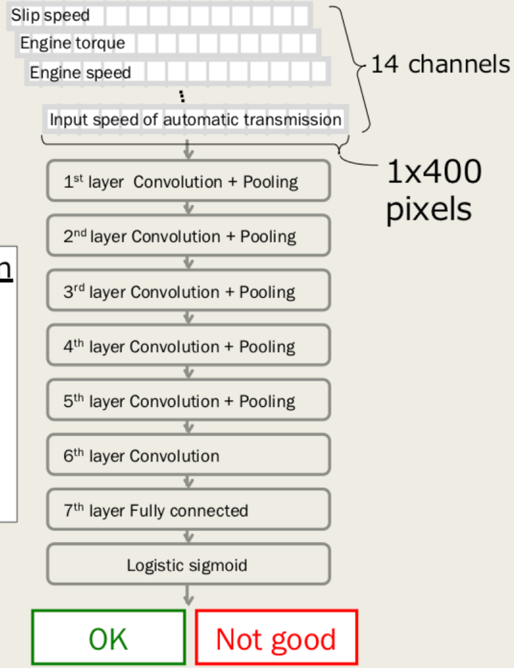
Normalization of data

$$x' = \frac{x-\mu}{\sigma}$$

where

μ : Mean

σ : Standard deviation



Result of 5-fold cross validation

	Accuracy
Test 1	89.4%
Test 2	88.7%
Test 3	90.4%
Test 4	89.9%
Test 5	89.2%
Average	89.5%

Dataset

Prepared 5,536 data

→ 4,429 training data → Augmented 48,719 data

→ 1,107 test data

Q-Boost: QUBO formulation

	Teacher	Learner1	Learner2	Learner3	...
Data1	OK=1	OK	NG	OK	
Data2	NG=0	OK	OK	NG	
Data3	NG=0	OK	NG	NG	
Data4	OK=1	NG	OK	OK	
:	:	:	:	:	

Binary variables: $q_i \in \{0$ (NOT USE), 1 (USE)$\}$

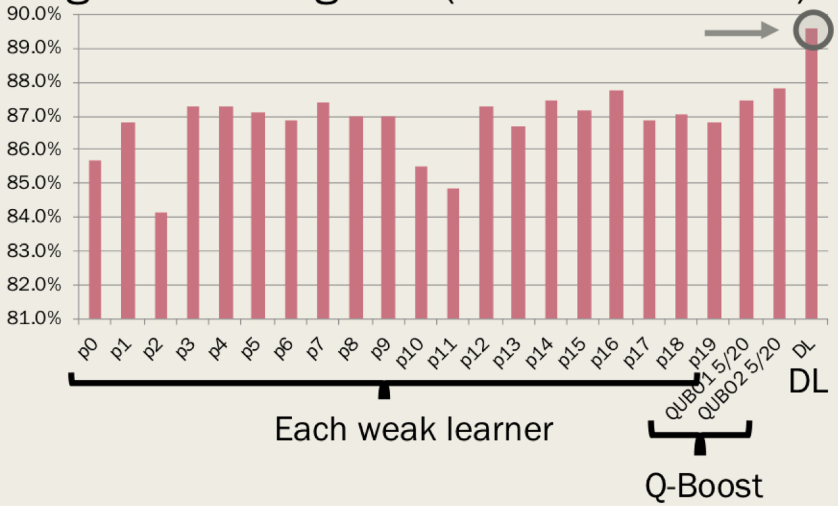
for the *i*-th weak leaner.

QUBO Formulation: Selecting K from M learners.

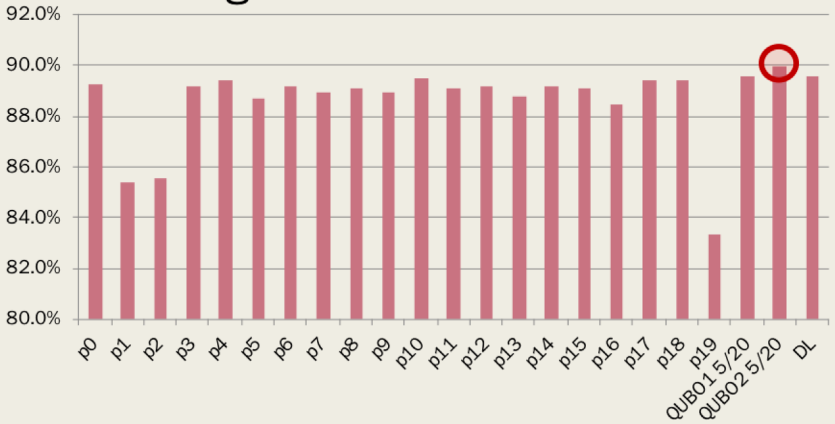
$$\mathcal{H}(\boldsymbol{q}) = \underbrace{\sum_{i=1}^{N}\sum_{j=1}^{M}(T_j - D_{ij})^2 q_i}_{\text{Decreasing the error}} + \lambda \underbrace{\left(\sum_{i=1}^{N} q_i - K\right)^2}_{\text{Selecting K learners}}$$

Main Results: Comparison between Q-Boost (0.1 million steps x 20 learners) and previous study, DL (1million steps)

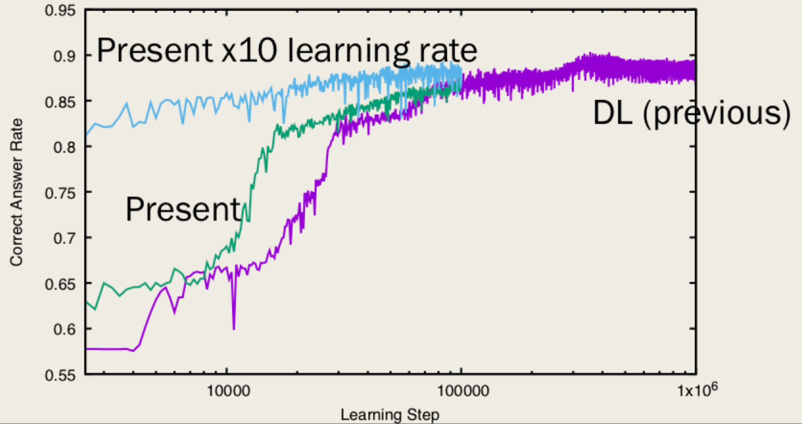
Original learning rate (fine-tuned for DL)



x10 learning rate



Each weak learner is 6-layer DL with less learning step and 10x learning rate. The learning time is 80% shortened than previous study by 6-layer DL with 1million steps.



Conclusion

	Expert Engineer	6-Layers Deep Learning	Q-Boost (20 weak learners)
Computer Resource	Human	GPU GeForce GTX 1080	D-Wave 2000Q
Accuracy Rate	100% (Teacher)	89.5%	**90.0%**
Training Time	20 years	4.5 hours (1)	**49 mins. (x0.18)**
#Training Data		48,179	48,179
Time to judge	7.5 hours	1.8 secs.	1.8 secs.

To be honest…, the optimization of selection from 20 weak learners can be easily done by standard digital computers. DW2000Q can deal with 64 learners, and how to deal with more…??? → Next Ohzeki-san’s talk!

New method for solving constraints

Masayuki Ohzeki
Tohoku University

Why do we use penalty method?

- Quadratic unconstrained binary optimization problem
  - Ising spin-glass model with two-body interaction and biases is embedded on the Ising machines

$$H(\mathbf{q}) = \mathbf{q}^{\mathrm{T}} Q \mathbf{q}$$

  - How to deal with constrained problems?
    - Penalty method

$$\boxed{F_i(\mathbf{q}) = K_i \quad \forall i} \Rightarrow H(\mathbf{q}) = H_0(\mathbf{q}) + \boxed{\frac{1}{2}\sum_i \lambda_i \left(F_i(\mathbf{q}) - K_i\right)^2}$$

  - However, it generates fully-connected Ising model

This leads to an obstacle in quantum annealing

How to solve constraints?

- Standard way to deal with the constraints in optimization problem
  - Penalty method

$$F_i(\mathbf{q}) = K_i \quad \forall i \Rightarrow H(\mathbf{q}) = H_0(\mathbf{q}) + \frac{1}{2}\sum_i \lambda_i \left(F_i(\mathbf{q}) - K_i\right)^2$$

    - Very simple to implement it
    - A large coefficient is necessary

Suffering from architecture of the D-Wave 2000Q

  - Lagrange multiplier method

$$F_i(\mathbf{q}) = K_i \quad \forall i \Rightarrow H(\mathbf{q}) = H_0(\mathbf{q}) + \sum_i \lambda_i \left(F_i(\mathbf{q}) - K_i\right)$$

    - It does not yield more two-body interactions
    - Instead, Lagrange multipliers are adaptively changed while solving the optimization problem

Lagrange multiplier method is more suitable for the D-Wave 2000Q

Protocol of our method

- A classical approach to reduce the squared term: Huburd-Stratnovich transformation
  - Partition function with Penalty method

$$Z = \sum_{\mathbf{q}} \exp\left(-\beta H_0(\mathbf{q}) - \frac{\beta}{2} \sum_{i=1}^{N} \lambda_i \left(F_i(\mathbf{q}) - K_i \right)^2 \right)$$

  - Partition function with Lagrange multiplier method can be obtained

$$Z \propto \sum_{\mathbf{q}} \prod_i \int dz_i \exp\left(-\sum_i \frac{1}{2} z_i^2 + i \sum_i \sqrt{\beta \lambda_i} z_i \left(F_i(\mathbf{q}) - C_i \right) - \beta H_0(\mathbf{q}) \right).$$

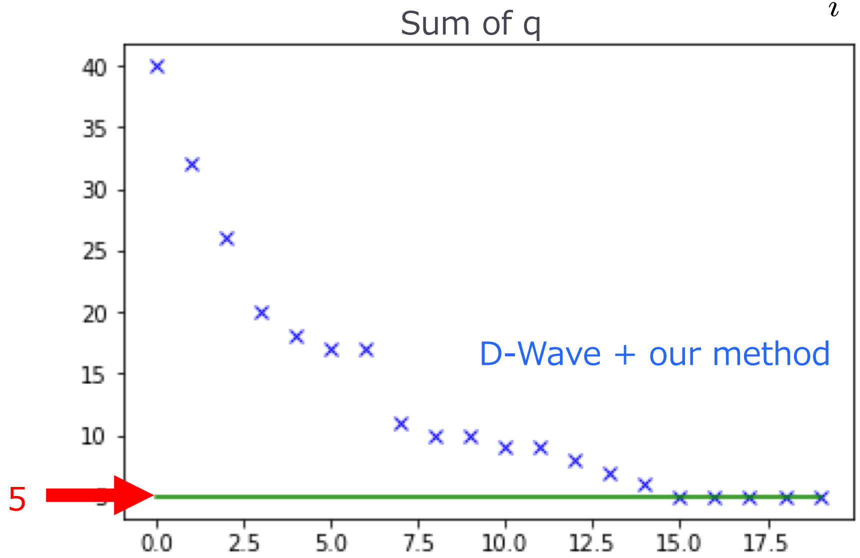
    - Solving the saddle-point equation in the large-beta limit in adequate axis and scale
    - Adaptively changing the Lagrange multipliers, following statistical mechanics

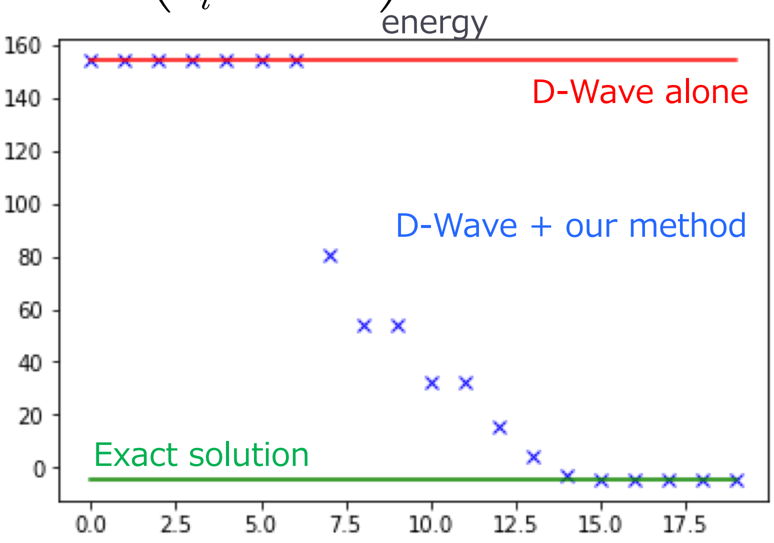
Lagrange multiplier method is more suitable for the D-Wave 2000Q

Results

- Simplest problem: selection variable in order of descent
  - From N=40 variables, select K=5

$$H(\mathbf{q}) = -\sum_i h_i q_i + \frac{\lambda}{2}\left(\sum_i q_i - K\right)^2$$

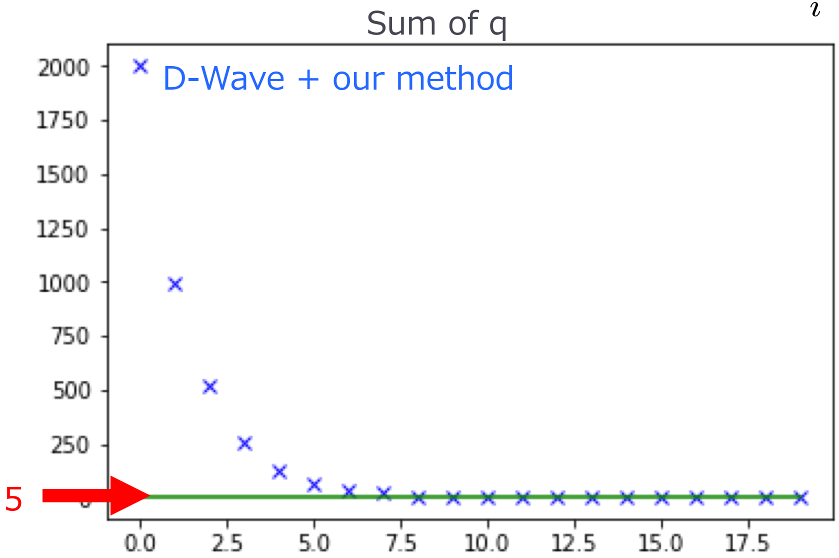


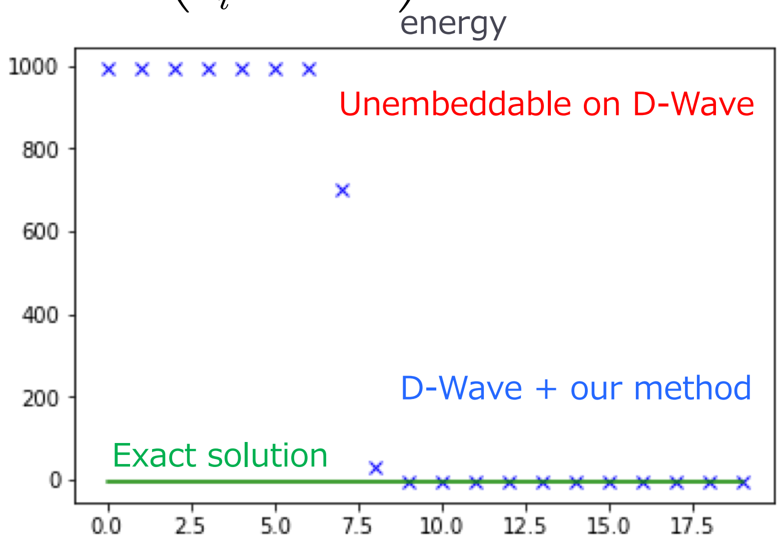


Results

- Simplest problem: selection variable in order of descent (unembeddable in penalty method!)
  - From N = 2000 variables, select K = 5

$$H(\mathbf{q}) = -\sum_i h_i q_i + \frac{\lambda}{2}\left(\sum_i q_i - K\right)^2$$



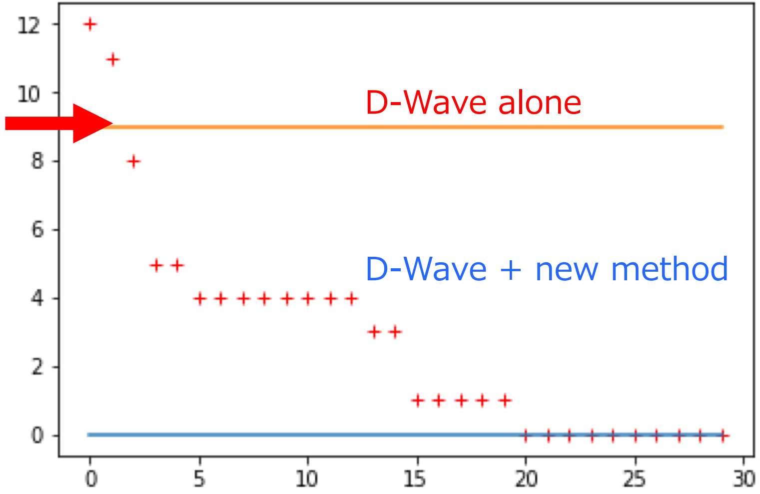


Results

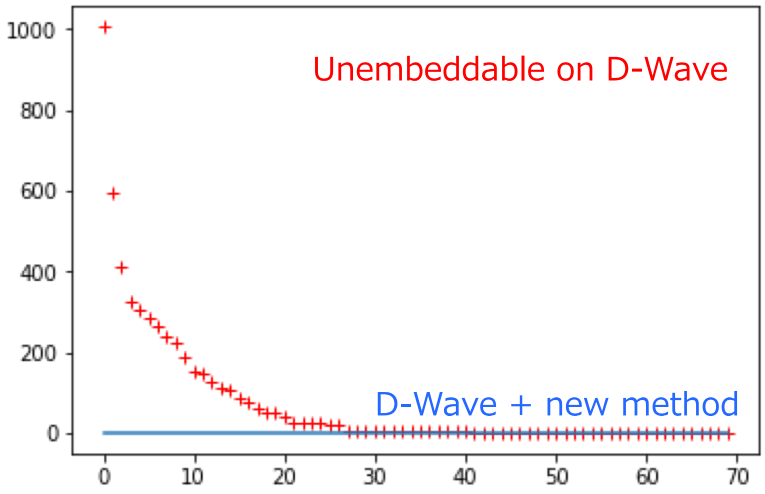
- CDMA in digital communication (Essentially the same as NBMF problem as in VW work)
  - Reconstruction of N input from M insufficient linear measurements $\mathbf{y} = A\mathbf{q}_0$

$$H(\mathbf{q}) = \frac{1}{2}\|\mathbf{y} - A\mathbf{q}\|_2^2$$

Residual (N=30・M=20)



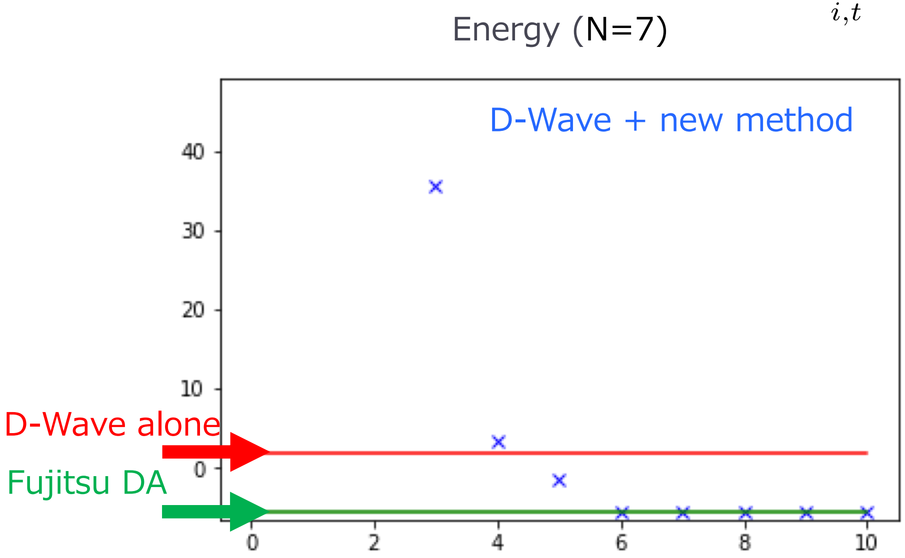
Residual (N=2000・M=1600)

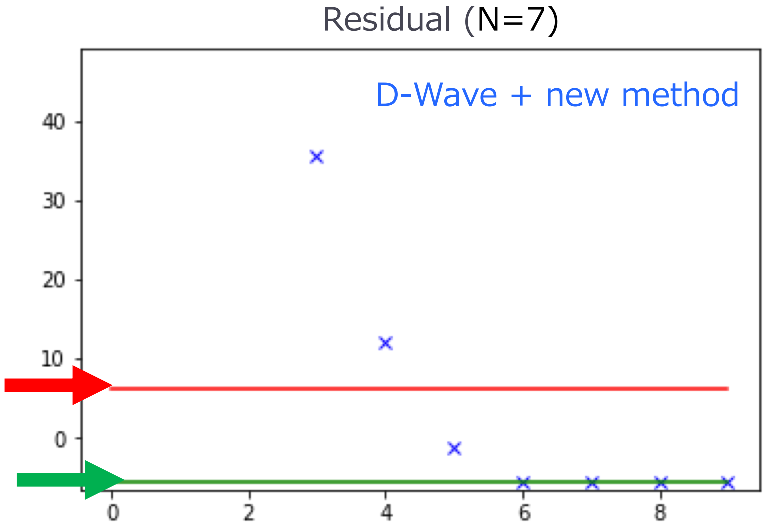


Results

- Double constraints as in traveling salesman problem
  - N-item selection (logical spins are N × N)

$$H(\mathbf{q}) = -\sum_{i,t} h_{i,t} q_{i,t} + \frac{\lambda_1}{2} \sum_i \left(\sum_t q_{i,t} - K_1 \right)^2 + \frac{\lambda_2}{2} \sum_t \left(\sum_i q_{i,t} - K_2 \right)^2$$

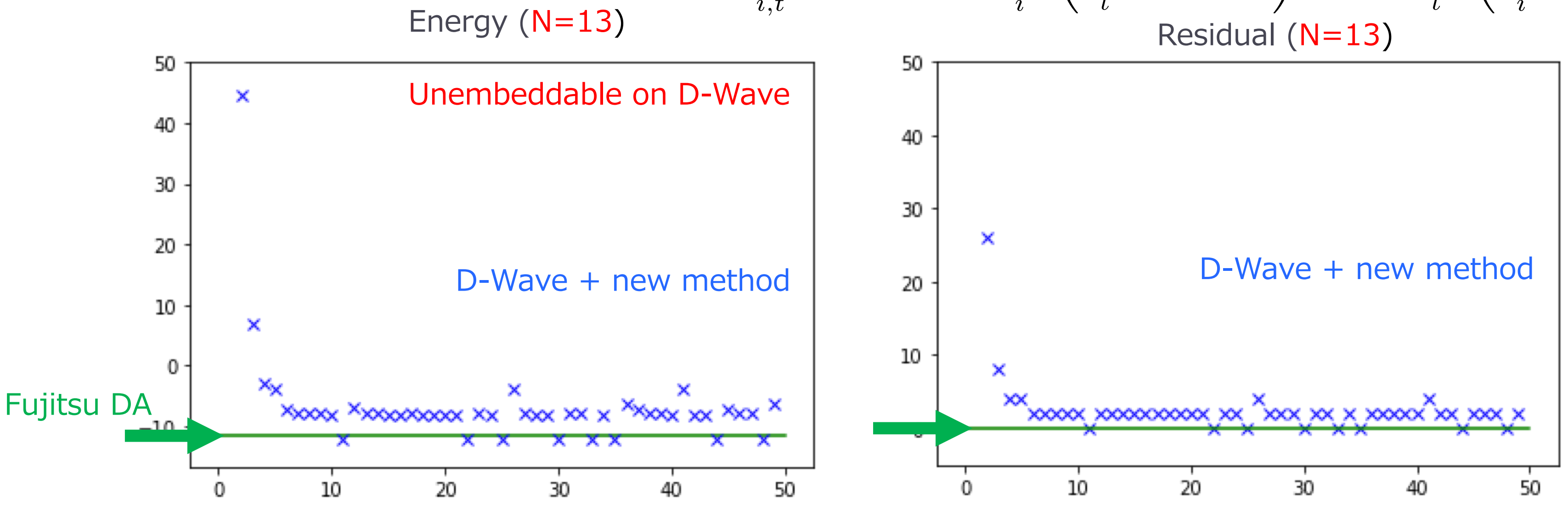




Results

- Double constraints as in traveling salesman problem
  - N-item selection problem (logical spins are N × N)

$$H(\mathbf{q}) = -\sum_{i,t} h_{i,t} q_{i,t} + \frac{\lambda_1}{2}\sum_i \left(\sum_t q_{i,t} - K_1\right)^2 + \frac{\lambda_2}{2}\sum_t \left(\sum_i q_{i,t} - K_2\right)^2$$

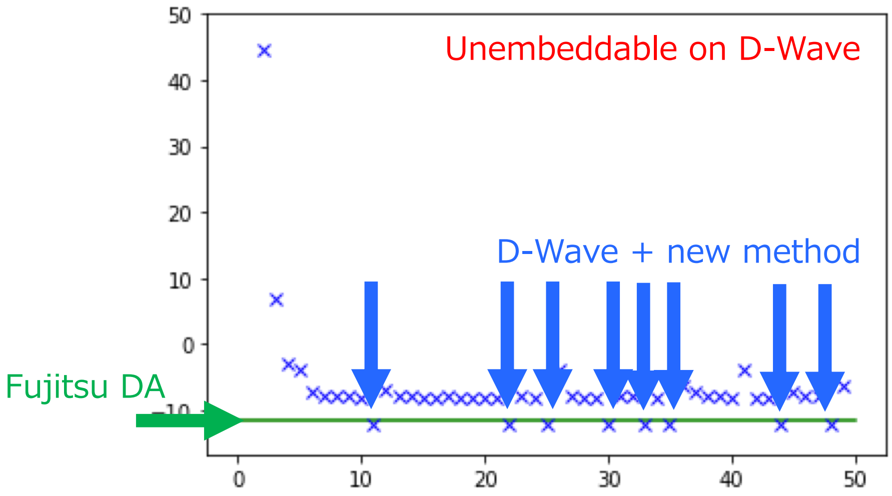


Results

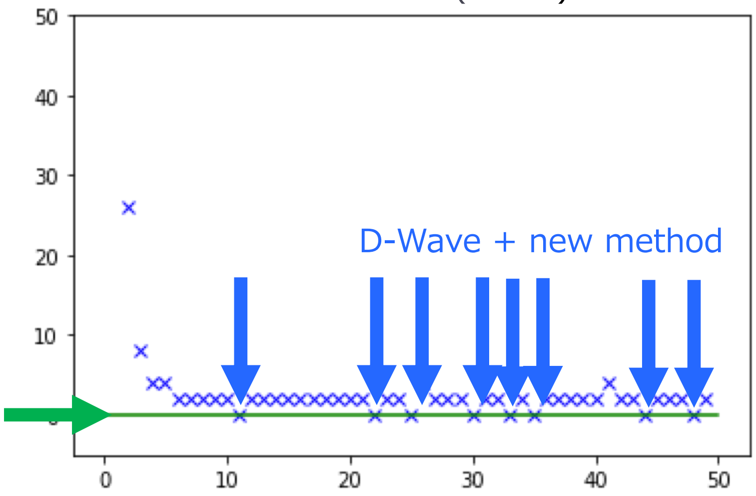
- Double constraints as in traveling salesman problem
  - N-item selection problem (logical spins are N × N)

$$H(\mathbf{q}) = -\sum_{i,t} h_{i,t} q_{i,t} + \frac{\lambda_1}{2}\sum_i \left(\sum_t q_{i,t} - K_1\right)^2 + \frac{\lambda_2}{2}\sum_t \left(\sum_i q_{i,t} - K_2\right)^2$$

Energy (N=13)



Residual (N=13)

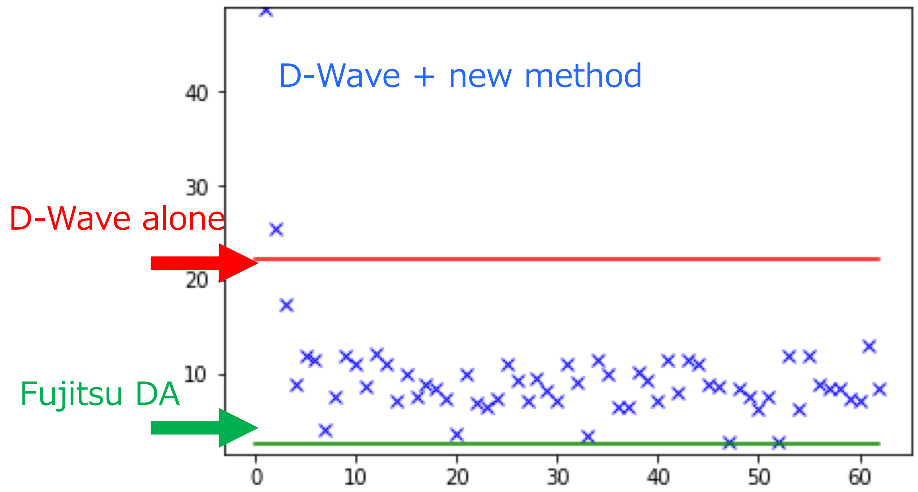


Results

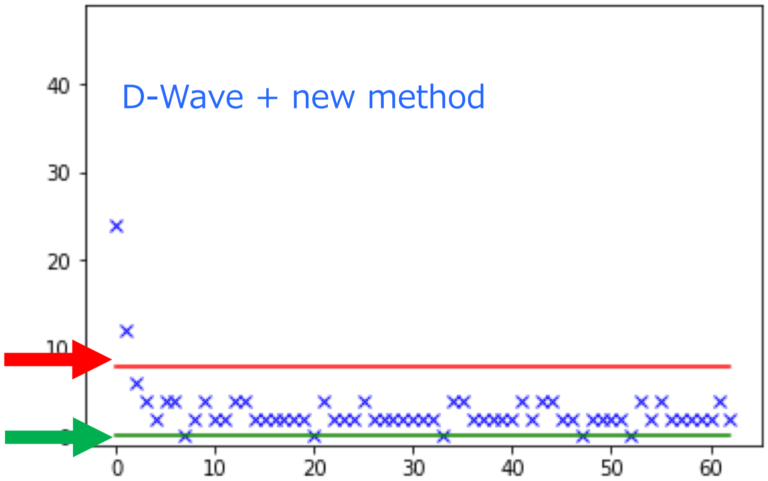
- Test in travelling salesman problem
  - N-city problem (logical spins are N × N)

$$H(\mathbf{q}) = H_0(\mathbf{q}) + \frac{\lambda_1}{2}\sum_i\left(\sum_t q_{i,t} - K_1\right)^2 + \frac{\lambda_2}{2}\sum_t\left(\sum_i q_{i,t} - K_2\right)^2$$

Energy (N=7)



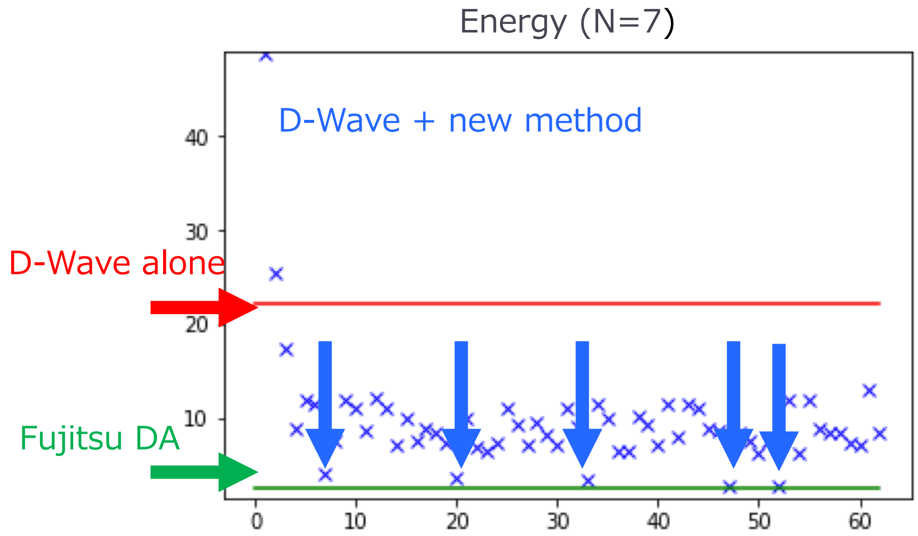
Residual (N=7)

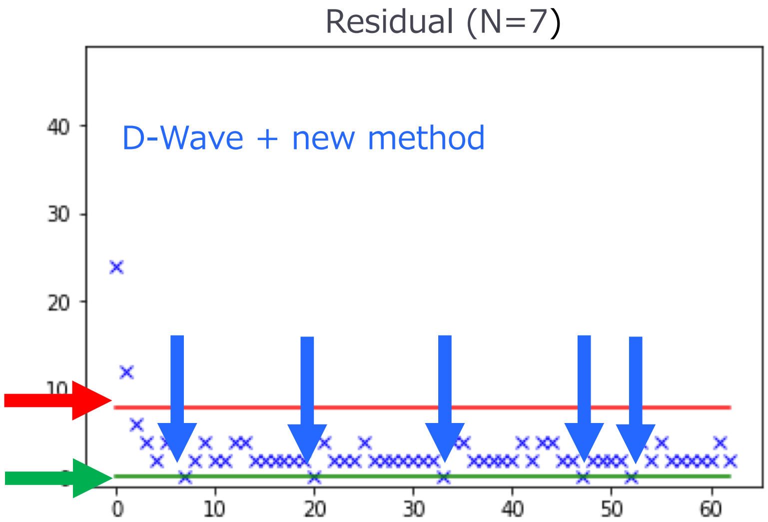


Results

- Test in travelling salesman problem
  - N-city problem (logical spins are N × N)

$$H(\mathbf{q}) = H_0(\mathbf{q}) + \frac{\lambda_1}{2}\sum_i\left(\sum_t q_{i,t} - K_1\right)^2 + \frac{\lambda_2}{2}\sum_t\left(\sum_i q_{i,t} - K_2\right)^2$$





Discussion

- Lagrange multiplier method with sampling works well
  - Unembeddable cases on the D-Wave 2000Q can be dealt with
  - Precision of solutions can increase
  - Constraints can ben satisfied

- Not yet well established at the moment
  - Theoretical/Experimental assessments are not provided yet
  - Probably a number of iterations for large-N system will remain (harmful for many users)
  - Probably the case on double constraints as in traveling salesman problem is difficult
  - However our method would be helpful for applications of sparse modeling as CDMA and NBMF

Summary

- We show several collaborations with T-QARD
  - Forecasting stock attractiveness by Masaya Abe (Nomura Asset Management)
  - Graph partition in quantum simulation by Takako Mashiko (Kyocera)
  - Bus Scheduling Problem by Naoki Maruyama (Tohoku・Hachinohe high school)
  - Black-box optimization by Ami Koshikawa (Tohoku・DENSO)
  - Calibration for Auto-Transmission Shift Control System by Miyama (Tohoku・Aisin AW)

- New method for solving constraints in quantum annealer
  - Lagrange multiplier method for dealing with constraints
  - Theoretical/Experimental assessment will remain in the future work