DENSO
Crafting the Core

量子アニーリングと
機械学習の交差点

2021/11/12

門脇　正史

(株)デンソー AI研究部 データサイエンス研究室
量子コンピューティング研究課

量子アニーリング (Quantum Annealing) ?

二値変数のコスト関数

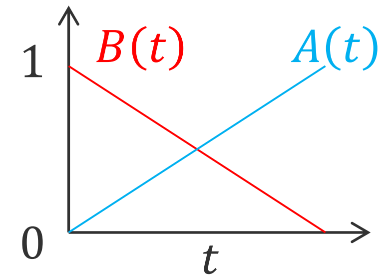
$J_{ij} \in \mathbb{R}, x_i \in \{-1,1\}$

σ_i^z, σ_i^x : パウリ行列

$$\mathcal{H}_C = -\sum_{i<j} J_{ij} x_i x_j$$

を、量子力学に拡張 ($x_i \to \sigma_i^z$)し、横磁場(~量子ゆらぎ)項 $\sum_i \sigma_i^x$ を加える。
(横磁場からコスト関数への変形をパラメータ化したい)

$$\mathcal{H}_Q(t) = -A(t) \sum_{i<j} J_{ij} \sigma_i^z \sigma_j^z - B(t) \sum_i \sigma_i^x$$



初期状態$|\psi(0)\rangle$を

$$\mathcal{H}_Q(0) = \sum_i \sigma_i^x$$

の基底状態*にえらび、シュレディンガー方程式によって系を時間発展する。

$$i \frac{d}{dt} |\psi(t)\rangle = \mathcal{H} |\psi(t)\rangle$$

断熱的に(~とってもゆっくりと)時間発展すると、断熱定理によりコスト関数の最適解が求まる。

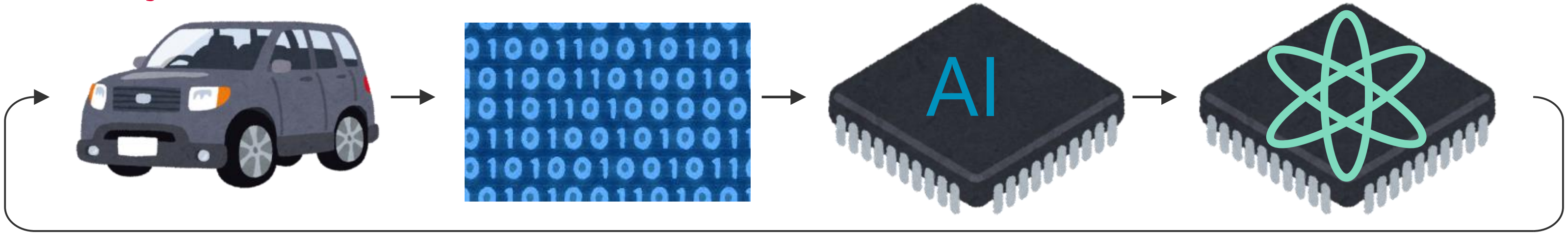
つまり、自明なハミルトニアンの解(すべての状態の重ね合わせ)から、非自明なハミルトニアンの解(知りたいコスト関数の最適解)が求まる。

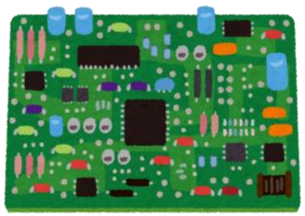
ちなみにシミュレーテッドアニーリングは、自明な分布(均一分布)から、非自明な分布(最適解でのデルタ関数)が求まる。

*基底状態(最小固有値の固有状態)はすべての計算基底の重ね合わせになっている(~全解空間の探索)

量子アニーリングで何を目指す？

DENSO Crafting the Core 製品 → データサイエンス → 機械学習(予測) → 最適化(制御)



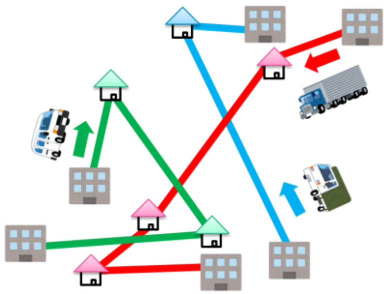


モノづくり

設計/開発/生産
・素材・部品
・生産・物流

サービス

モビリティ
・配車
・配送



【量子アニーリング業界】

今日の話

1. まだまだ基礎研究
   - もっと最適化を頑張る: 新しい相互作用、新しい制御、ダイナミクス
   - 最適化以外の出口: 量子多体系シミュレーション
2. 離散最適化ソルバーとして使う (イジングソルバーも含めて)

Future of product design

Robot Scientist
Cyber Physical System

Data generation

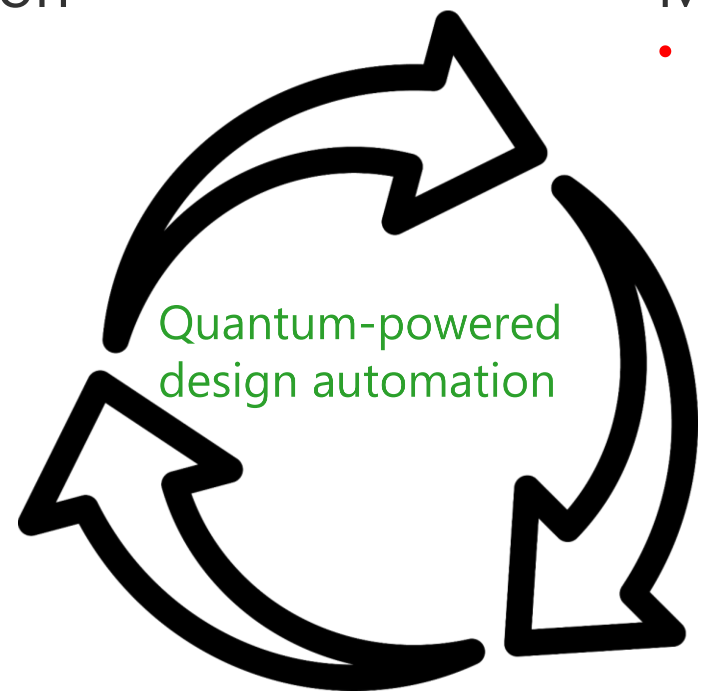
- experiments
- simulations
- robotics
- database

VQAs
Variational
Quantum
Algorithm

Model generation

- machine learning

QML
Quantum
Machine
Learning



Hypothesis generation

- optimization of cost functions

QA
Quantum
Annealing

1. ブラックボックス最適化　　機械学習 > 量子 な話
    - 非可逆行列圧縮
    - 基板設計最適化
2. 量子アニーリング　　量子 > 機械学習 な話
    - 次世代量子アニーラ向けアルゴリズム

Black-box optimization (BBO)

- Black-box optimization (BBO) of is a method to optimize unknown cost functions.
- Bayesian optimization (BO) is a well-known BBO algorithm, that acquire data points sequentially based on a model constructed from pre-acquired data set.
- This is a Bayesian approach of experimental design.
- BO iterates the 3 processes;
  - input-output data from difficult-to-evaluate function, such as experiments
  - construct a regression model (e.g. Gaussian process) & an acquisition function by using the regression model
  - find a new data point that maximizes the acquisition function
- BO is designed for “continuous” explanatory variables.
- BBO for “integer” variables (combinatorial optimization) is not studied so far.

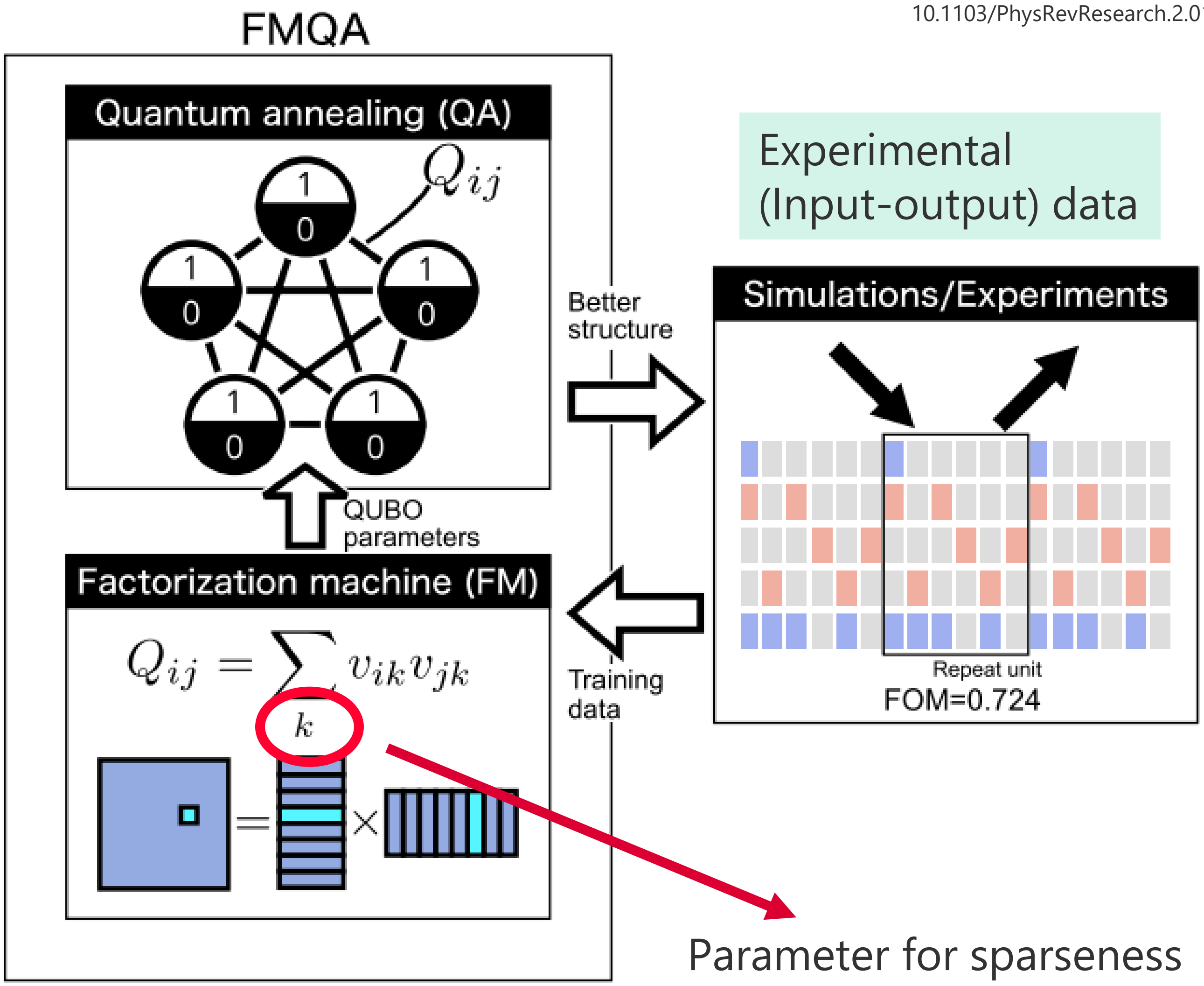
FMQA algorithm

Kitai et al (Phys Rev Res 2020)
10.1103/PhysRevResearch.2.013319

Maximize acquisition function

Modeling acquisition function

Experimental (Input-output) data



Parameter for sparseness

Nanostructure optimization of metamaterial

Kitai et al (Phys Rev Res 2020)
10.1103/PhysRevResearch.2.013319

The goal is to design complex structures of wavelength selective radiators with the thermal atmospheric transparency window.

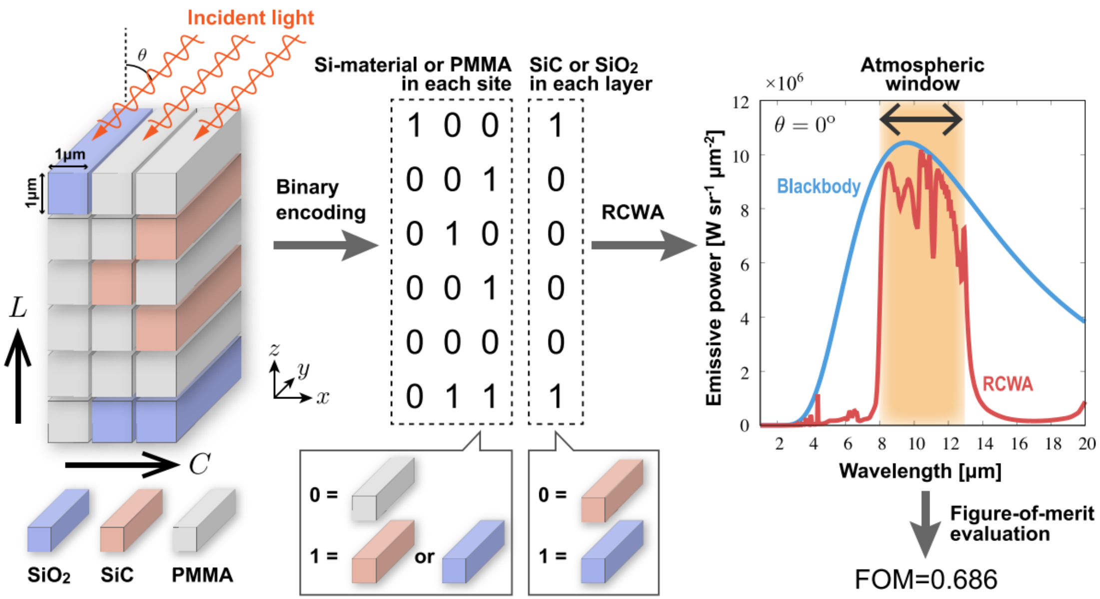
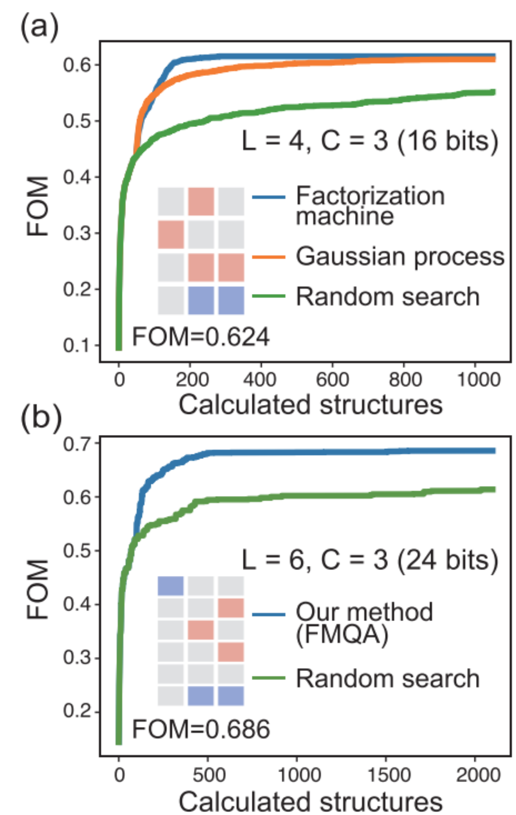


FIG. 2. Example of the target metamaterial structure for $L = 6$ and $C = 3$, the binary variables expressing it, and the emissive powers of target metamaterial (red curve) and blackbody (blue curve) calculated by RCWA.



Bayesian Optimization of Combinatorial Structures (BOCS)

Baptista and Poloczek (ICML2018)
arXiv:1806.08838

Algorithm 1 Bayesian Optimization of Combinatorial Structures

1: **Input:** Objective function $f(x) - \lambda \mathcal{P}(x)$; Sample budget N_{max}; Size of initial dataset N_0.
2: Sample initial dataset D_0.
3: Compute the posterior on $\boldsymbol{\alpha}$ given the prior and D_0.
4: **for** $t = 1$ **to** $N_{max} - N_0$ **do**
5: Sample coefficients $\boldsymbol{\alpha}_t \sim P(\boldsymbol{\alpha} \mid \mathbf{X}, \mathbf{y})$.
6: Find approximate solution $x^{(t)}$ for $\max_{x \in \mathcal{D}} f_{\boldsymbol{\alpha}_t}(x) - \lambda \mathcal{P}(x)$.
7: Evaluate $f(x^{(t)})$ and append the observation $y^{(t)}$ to $\mathbf{y}$.
8: Update the posterior $P(\boldsymbol{\alpha} \mid \mathbf{X}, \mathbf{y})$.
9: **end for**
10: **return** $\operatorname{argmax}_{x \in \mathcal{D}} f_{\boldsymbol{\alpha}_t}(x) - \lambda \mathcal{P}(x)$.

5. Acquisition function

6. Maximization

7. Input-output data

sparsity-inducing prior
(horseshoe prior)

BOCS on random Binary Quadratic Programming instances

Baptista and Poloczek (ICML2018)
arXiv:1806.08838

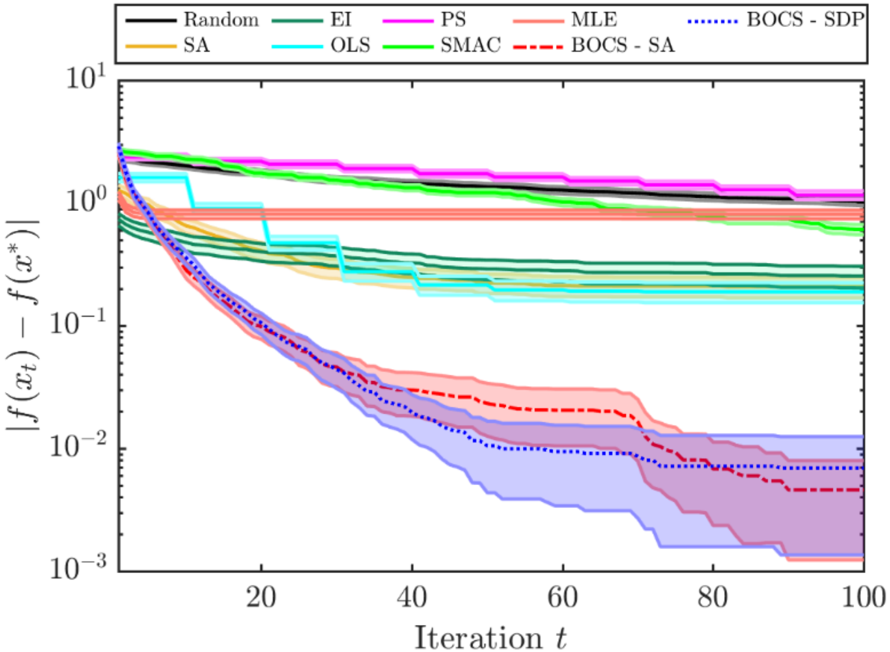


Figure 1. Random BQP instances with $L_c = 10$ and $\lambda = 0$: Both variants of BOCS outperform the competitors substantially.

Other algorithms

- EI: Expected Improvement
- PS: Sequential Monte Carlo Particle Search
- MLE: Maximum Likelihood Estimate
- OLS: Oblivious Local Search
- SA: Simulated Annealing
- SMAC: Sequential Model-based Algorithm Configuration

BOCS algorithms

- BOCS-SDP
- BOCS-SA

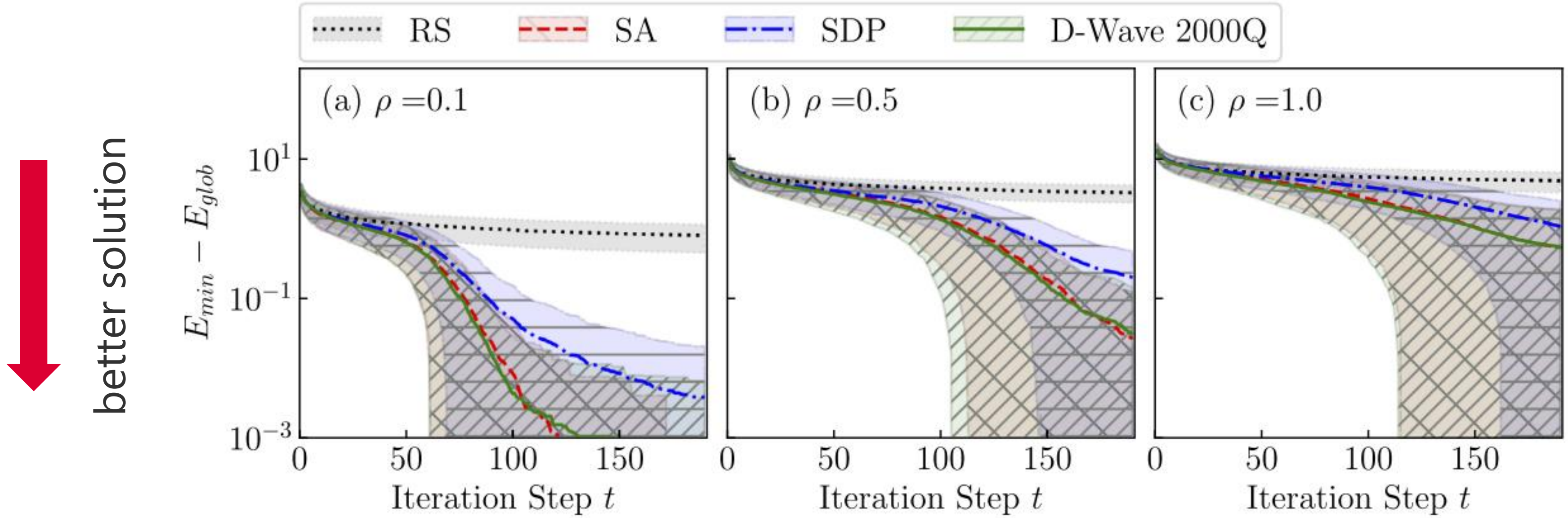
BOCS-QA

Koshikawa et al (J Phys Soc Jpn 2021)
10.7566/JPSJ.90.064001

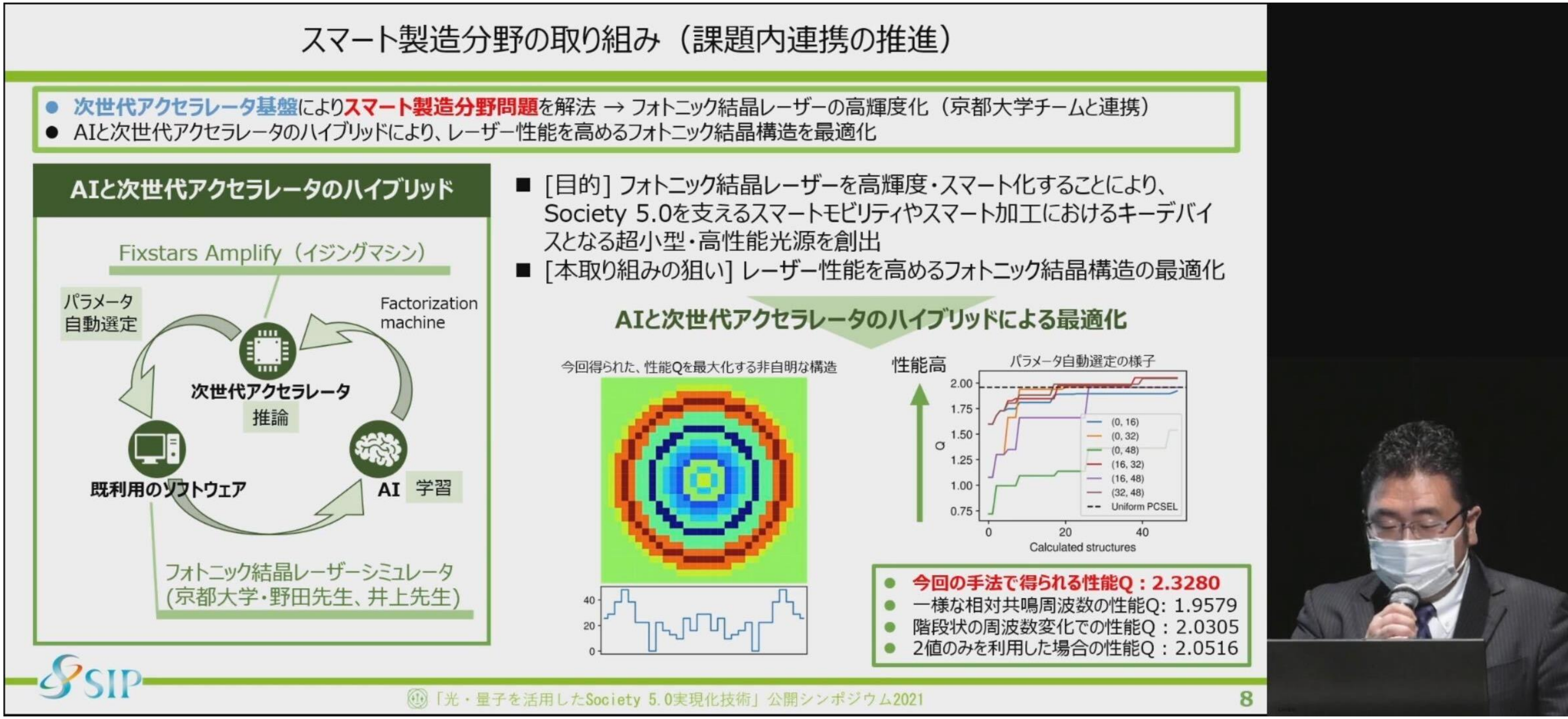
Finding ground state of diluted Sherrington-Kirkpatrick model from sequentially obtained state-energy data set without knowing the Hamiltonian.

$$\mathcal{H} = -\frac{1}{N}\sum_{i<j} J_{ij}\sigma_i\sigma_j \qquad J_{ij} \sim (1-\rho)\mathcal{N}(0,+0) + \rho\mathcal{N}(0,1)$$

N=20, 50 instances x 10 runs



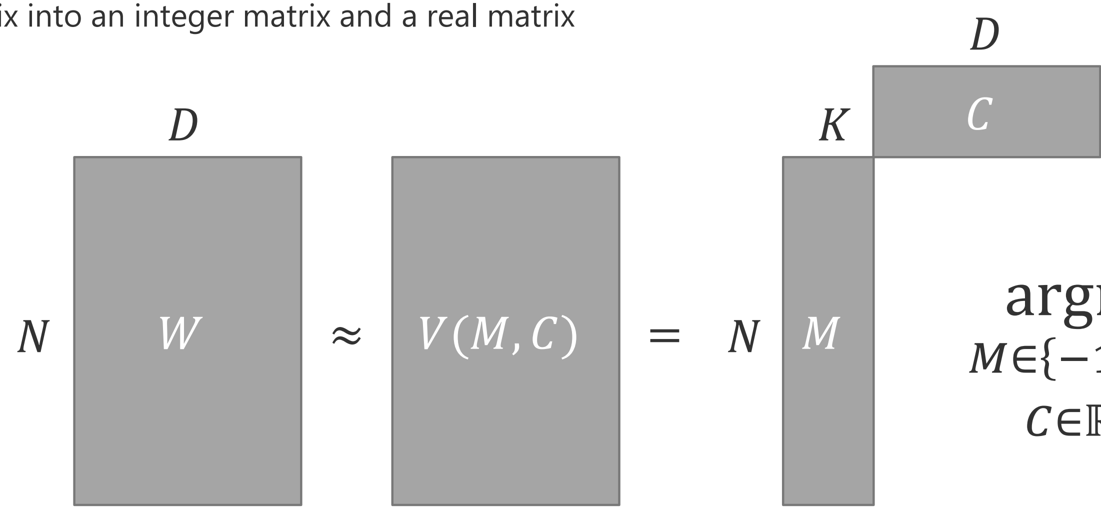
FMQA x photonic crystal



Integer decomposition (ID) 整数基底分解法 (非可逆圧縮)

Decompose a matrix into an integer matrix and a real matrix

Ambai and Sato (ECCV2014)



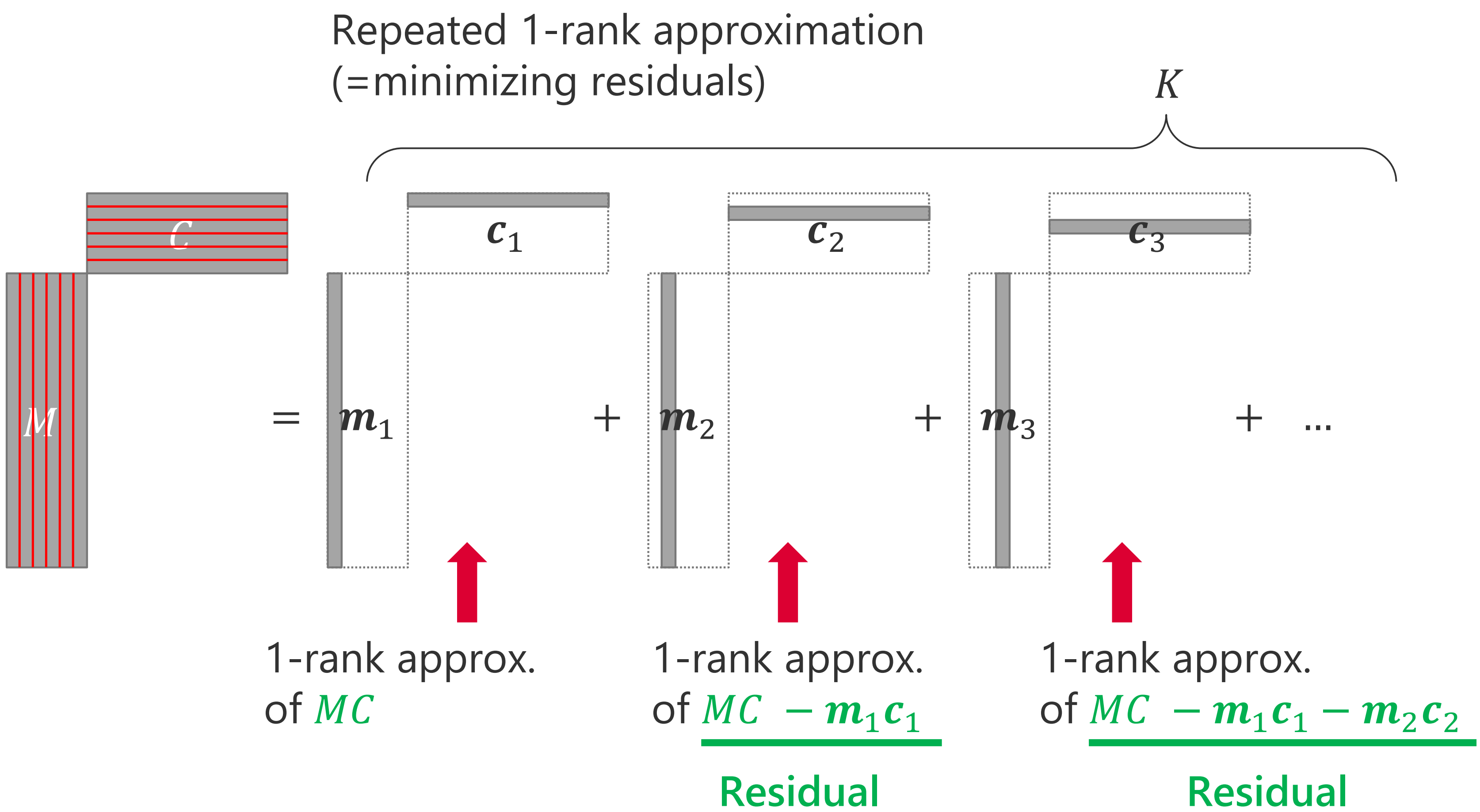
$$\underset{\substack{M \in \{-1,1\}^{NK} \\ C \in \mathbb{R}^{KD}}}{\operatorname{argmin}} |W - V|^2$$

W: target weight matrix
V: approximated matrix (function of M and C)
M: integer matrix $M \in \{-1,1\}^{NK}$
C: real matrix $c \in \mathbb{R}^{KD}$

The real matrix C is a dominant part of the memory footprint.
Memory size reduction $N \times D \rightarrow K \times D$ & speed up in matrix operations.
However, finding M and C is a mixed integer non-linear programming (MINLP) and NP hard problem.

Implementation of ID

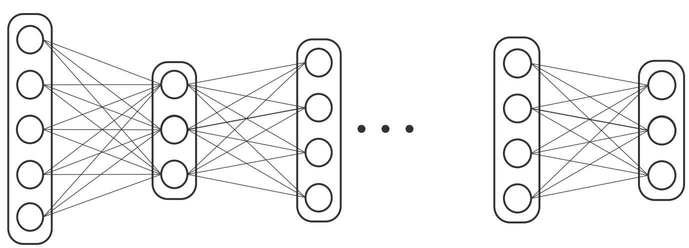
Ambai and Sato (ECCV2014)



Applications of ID

1. Image recognition
   - Linear SVM kernels
     Mitsuru Ambai and Ikuro Sato, SPADE: Scalar Product Accelerator by Integer Decomposition for Object Detection (ECCV2014)
2. Voice recognition
   - Deep neural networks
     太刀岡勇気, 安倍満, 整数基底分解法と量子化による音響モデルの圧縮 (日本音響学会2018年春季研究発表会)

Sound input



Prediction output

$= f(W^T x)$

f : activation function

ID demonstrated 1/20 memory footprint,
x15 acceleration with 1.5% increase of the error.

BBO for ID

We can convert the MINLP to
a nonlinear integer programming (NLIP).

In ID, for given M,

$$C = (M^T M)^{-1} M^T W$$

$$V(M, C) = MC = M(M^T M)^{-1} M^T W = V(M)$$

The problem is written as

$$\underset{M \in \{-1,1\}^{NK}}{\operatorname{argmin}} |W - V|^2$$

$$= \underset{M \in \{-1,1\}^{NK}}{\operatorname{argmin}} |W - M(M^T M)^{-1} M^T W|^2$$

MINLPの一般的な解法としてはBenders分解法などがある。

Non-linear & non-QUBO combinatorial optimization problem

How to solve it?

$$\underset{M}{\operatorname{argmin}} |W - M(M^T M)^{-1} M^T W|^2$$

The explicit expression will be very complicated!
We employ Black-box optimization of the function,

$$f(M) = |W - M(M^T M)^{-1} M^T W|^2$$

(Numerical) Experiment

data acquisition

$$z_i = f(M_i)$$

data modeling

$$z = g(M)$$

Machine Learning

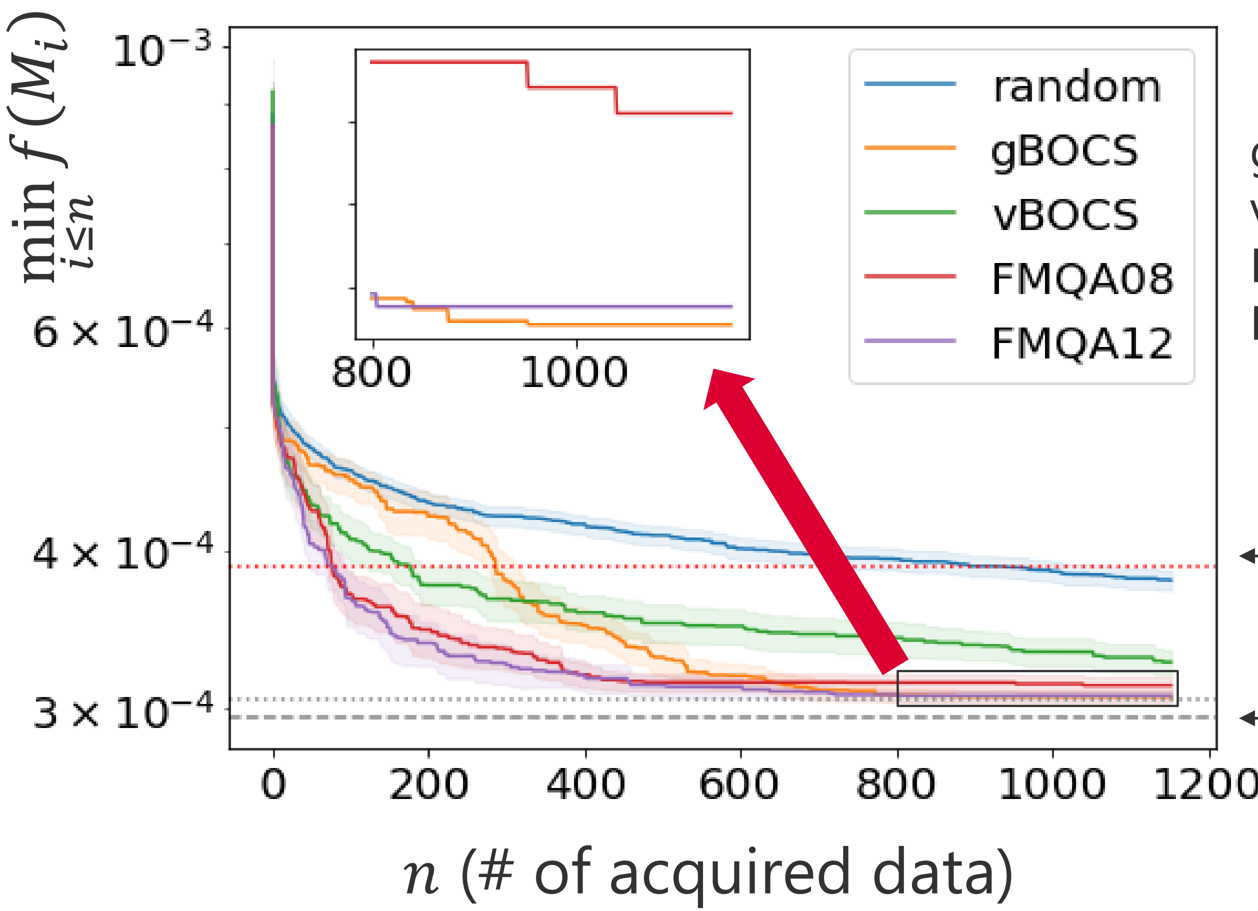
QUBO form

candidate generation

$$M_{i+1} = \operatorname{argmin}_M g(M)$$

Optimization

Results (8 x 100 => 8 x 3 & 3 x 100 [N=24])



gBOCS: tuned BOCS-SA
vBOCS: vanilla BOCS-SA
FMQA08: FMQA (k=8)
FMQA12: FMQA (k=12)

← Original repeated 1-rank approximation

← Ground state (Exact solution)

Quality of solutions and execution time

10 randomly generated instances
25 independent runs

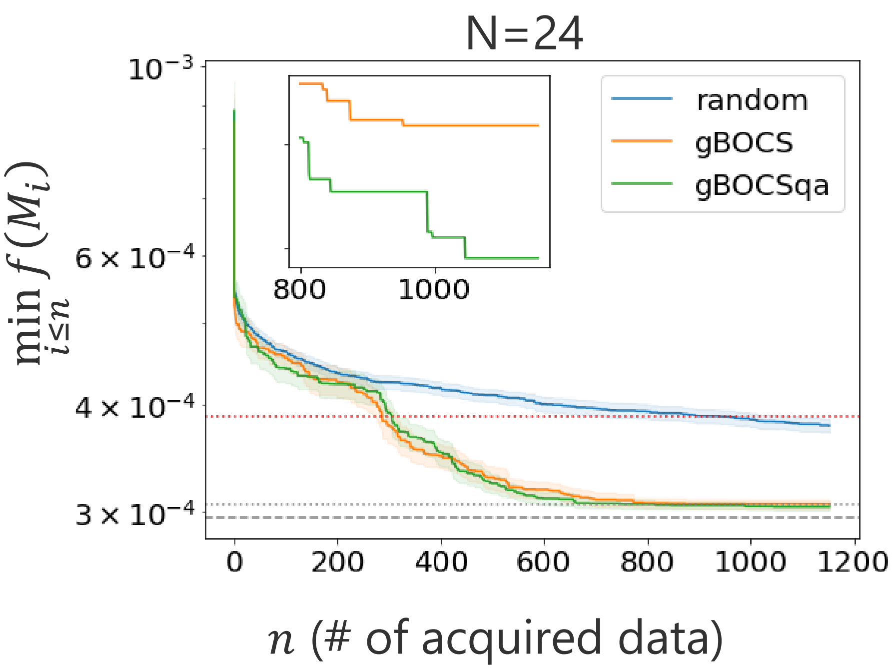
	random	gBOCS	vBOCS	FMQA08	FMQA12
GS count	3	62	6	58	53
GS win count	0	5	0	4	2
Energy win count	0	7	0	2	1
Time	0.35	51	24394	3719	3714

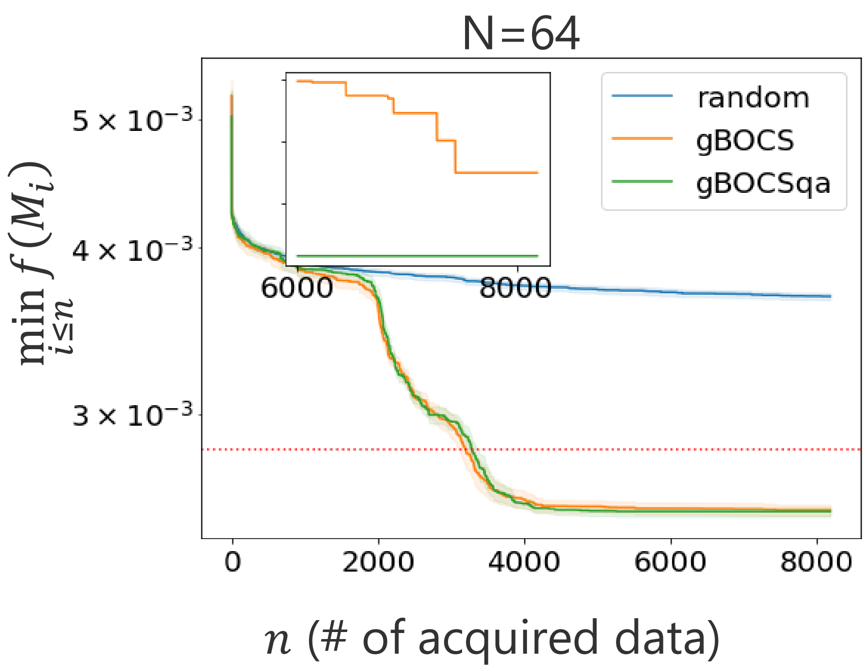
1/480

1/70

Comparable/better solutions with faster execution time

BOCS-SA and BOCS-QA



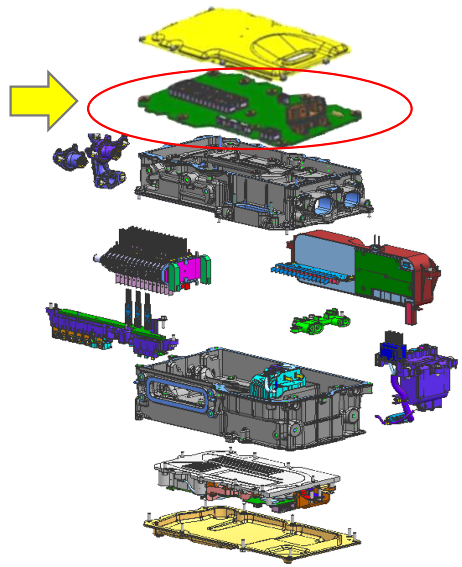


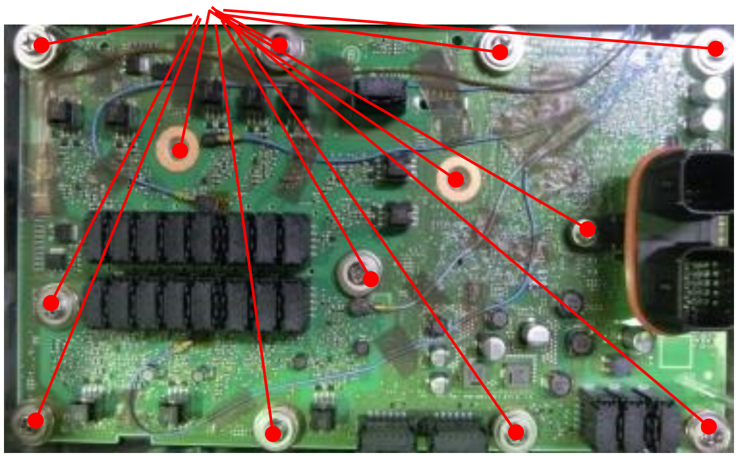
We see some differences between SA and QA, but not significant

Vibration-resistant design (VRD)

Power control unit

https://www.denso.com/jp/ja/news/newsroom/2019/20190522-01/-/media/8c44d0012d0148f893903eb515d52b67.ashx



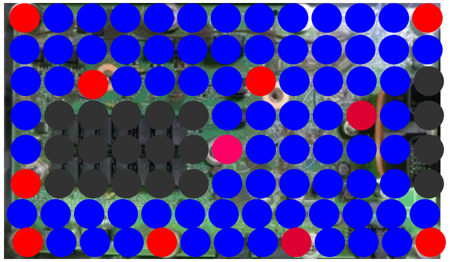


Printed circuit board (TOYOTA Yaris)

Vibrations in the power control unit will damage the printed circuit board. To reduce the damage, we design a PCB with a high resonance frequency. That generally causes many mounting holes. The goal is to design a PCB with a higher resonance frequency and fewer mounting holes.

Collaboration with T. Matsumori and M. Taki

Formulation of the problem



- Candidate points ($x_i = 0$)
- Selected points ($x_i = 1$)
- Keep out area

$$\underset{\mathbf{x}\in\{0,1\}^n}{\text{Minimize}} \quad F(\mathbf{x}) = \sum_{i=1}^{n} x_i,$$

Minimize # of mounting holes

$$\text{subject to} \quad G_1(\mathbf{x}) = \mathcal{G}(\mathbf{x}) - \overline{f} \geq 0$$

Resonance frequency $\mathcal{G}(x)$ constraint

by finite element method (FEM)

Modified cost function:

$$\underset{\mathbf{x}\in\{0,1\}^n}{\text{Minimize}} \quad \mathcal{H}(\mathbf{x}) = w_F(F(\mathbf{x}) - \overline{P})^2 - w_G\mathcal{G}.$$

Higher resonance frequency

Target # of mounting holes

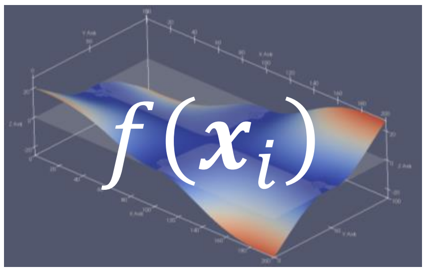
Non-linear & non-QUBO combinatorial optimization problem

How to solve it?

The problem to be solved:

$$\underset{x\in\{0,1\}^n}{\operatorname{argmin}}\, w_F(F(\boldsymbol{x}) - \bar{P})^2 - w_G\mathcal{G}(\boldsymbol{x})$$

$$f\,(\boldsymbol{x}) = w_F(F(\boldsymbol{x}) - \bar{P})^2 - w_G\mathcal{G}(\boldsymbol{x})$$



(Numerical) Experiment

data acquisition

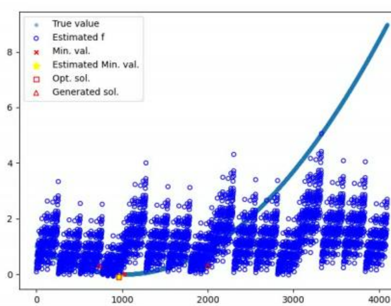
$z_i = f(\boldsymbol{x}_i)$

data modeling

$z = g(\boldsymbol{x})$

Machine Learning

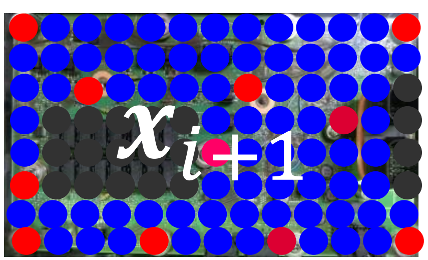
QUBO form



candidate generation

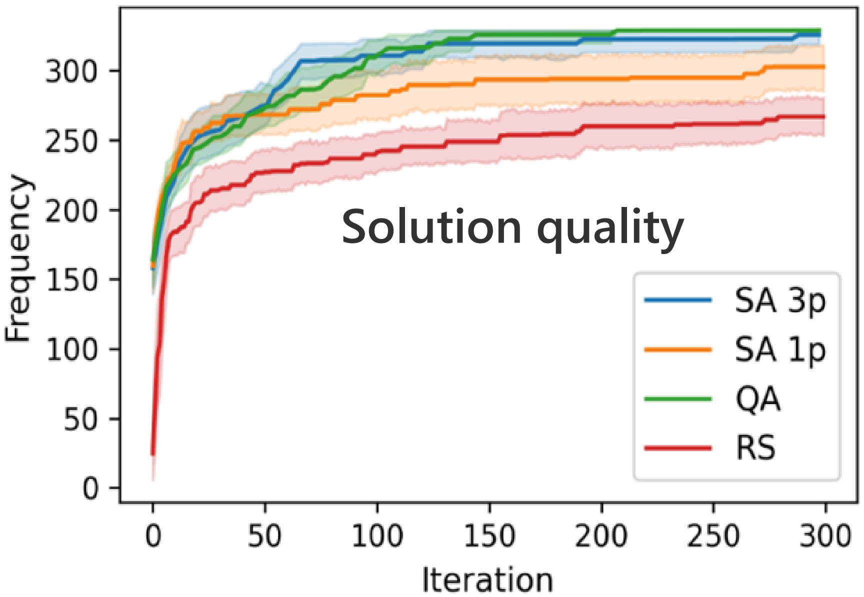
$\boldsymbol{x}_{i+1} = \operatorname{argmin}_x g(\boldsymbol{x})$

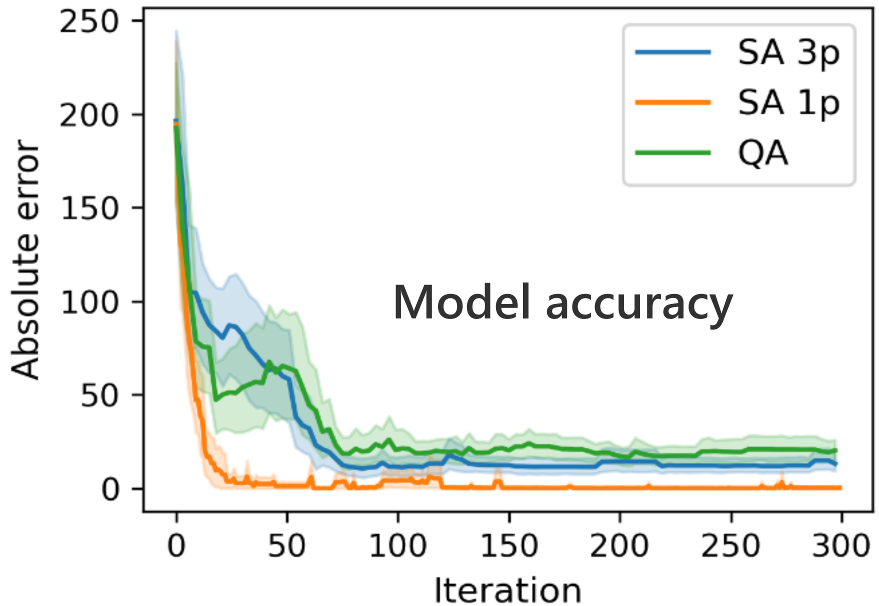
Optimization



Results (N=12) [$2^{12} = 4096$]

SA 3p: optimal + 2 neighbors
SA 1p: optimal





SA and QA outperform random search (RS).

1. ブラックボックス最適化
    - 非可逆行列圧縮
    - 基板設計最適化
2. 量子アニーリング
    - 次世代量子アニーラ向けアルゴリズム

Solving combinatorial optimization problems by qubits

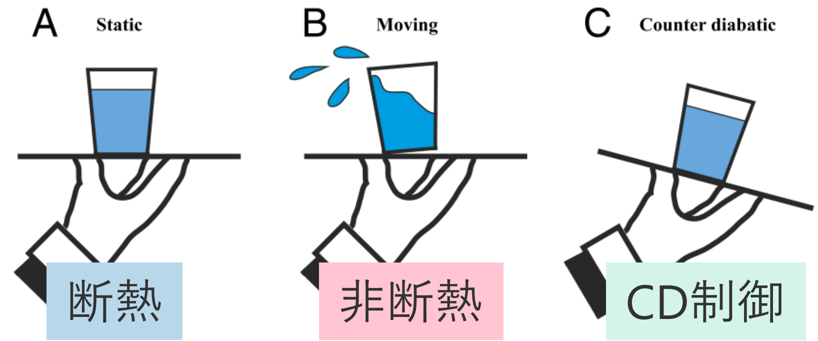
- Quantum Annealing (QA)
- Adiabatic Quantum Computation (AQC)
- Quantum Approximate Optimization Algorithm (QAOA)
- Variational Quantum Eigensolver (VQE)

Diabatic vs. Adiabatic, Global vs. Local search,

Unitary transformations (types & control parameters), etc…

Shorter processing times are desired in NISQ devices. We focus on dynamical aspects of QA with non-stoquastic Hamiltonian. A counter-diabatic driving and variational optimization are studied. We propose a new algorithm for combinatorial optimization. Comparison with conventional QA and simulated annealing (SA) will be presented, as well as a proposed oracle is discussed.

Model Hamiltonian with counter-diabatic driving



$$\mathcal{H} = A(t)\mathcal{H}_z + B(t)\,\mathcal{H}_x + \mathcal{H}_y$$

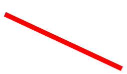
$\mathcal{H}_z = -\frac{1}{N-1}\sum_{i<j} \sigma_i^z \sigma_j^z$: cost function

$\mathcal{H}_x = -\sum_i \sigma_i^x$: uniform transverse field

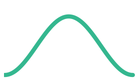
$\mathcal{H}_y = -\sum_i C_i(t)\sigma_i^y$: non-uniform y-field (non-stoquastic, counter-diabatic)

where the coefficients are

$A(t) = at/\tau$



$B(t) = b(1 - t/\tau)$



$C_i(t) = c_i \sin^2(\pi t/\tau)$

$a, b, \{c_i\}$: parameters

τ : annealing time

($\tau = a = 1$ for simplicity)

Sels 2017, Takahashi 2017, Prielinger 2021

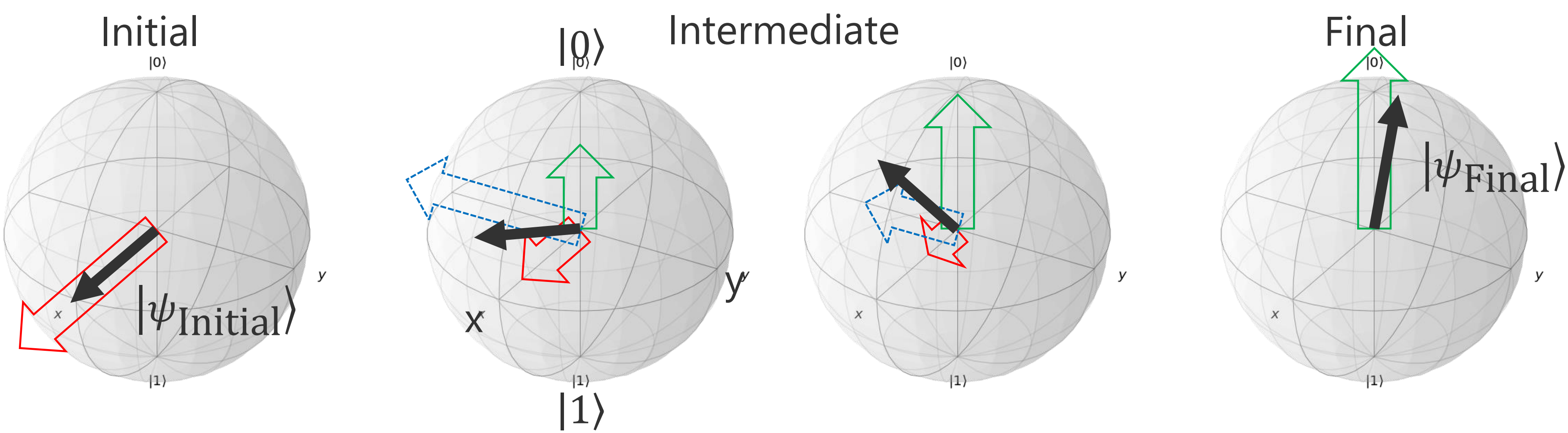
Time evolution of the system

$$\frac{d}{dt}|\psi\rangle = -i\mathcal{H}|\psi\rangle = -i\left(A(t)\mathcal{H}_z + B(t)\,\mathcal{H}_x + \mathcal{H}_y\right)|\psi\rangle$$

Initial : transverse field (makes superposition of all possible states)

Intermediate : (1) transverse field is decreasing,
(2) effective local field* in z is growing,
(3) non-uniform y-field is applied (we assume it's dominant.)

Final : effective local field



* This caused by spin-spin interactions in the cost function

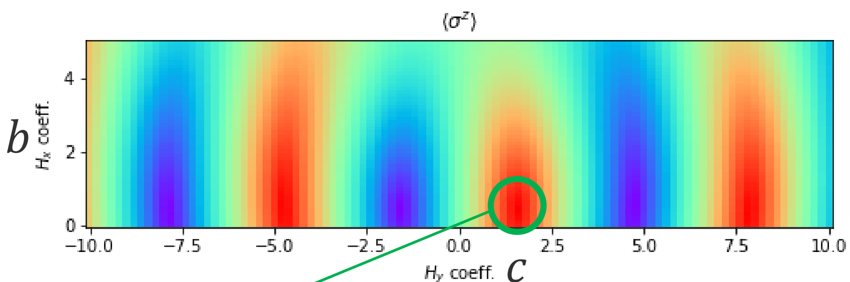
Mean-field model & variational optimization

Mean field approximation:

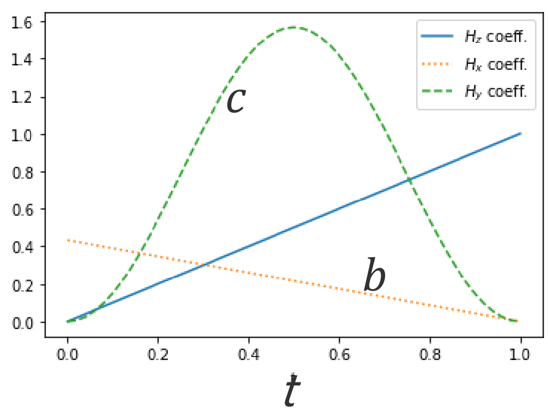
$$\mathcal{H} = -t\langle\sigma^z\rangle\sigma^z - b(1-t)\sigma^x - c\sin^2(\pi t)\,\sigma^y$$

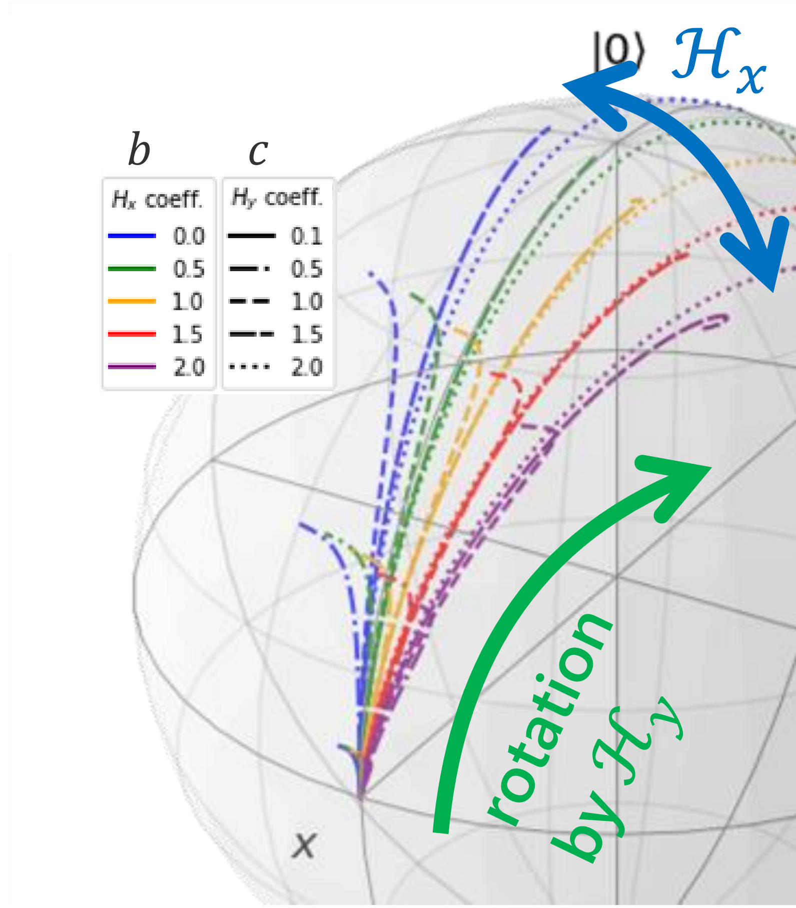
$$\langle\sigma^z\rangle = \langle\psi|\sigma^z|\psi\rangle$$

optimize magnetization $\langle\sigma^z\rangle$



$b_{\mathrm{opt}} = 0.539$
$c_{\mathrm{opt}} = 1.565$
$(\pi/2 = 1.571)$





Measures of optimization in many-body systems

We use two measures:

- Fidelity(P_{GS}) : probability to find the ground state from the final state

  $P_{\mathrm{GS}} = \langle \psi_{\mathrm{Final}} | \psi_{\mathrm{GS}} \rangle^2$, which can't be observed by experiment.

  A oracle to measure this probability is required.

  (This requirement will be relaxed in later analysis.)

  Error = $1 - P_{\mathrm{GS}}$

- Energy : energy of the final state $\langle \psi_{\mathrm{Final}} | \mathcal{H}_z | \psi_{\mathrm{Final}} \rangle$, which can

  be observed.

  Ground states and excited states are not distinguishable.

Variational optimization (Ferromagnetic, fully connected, $N = 20$)

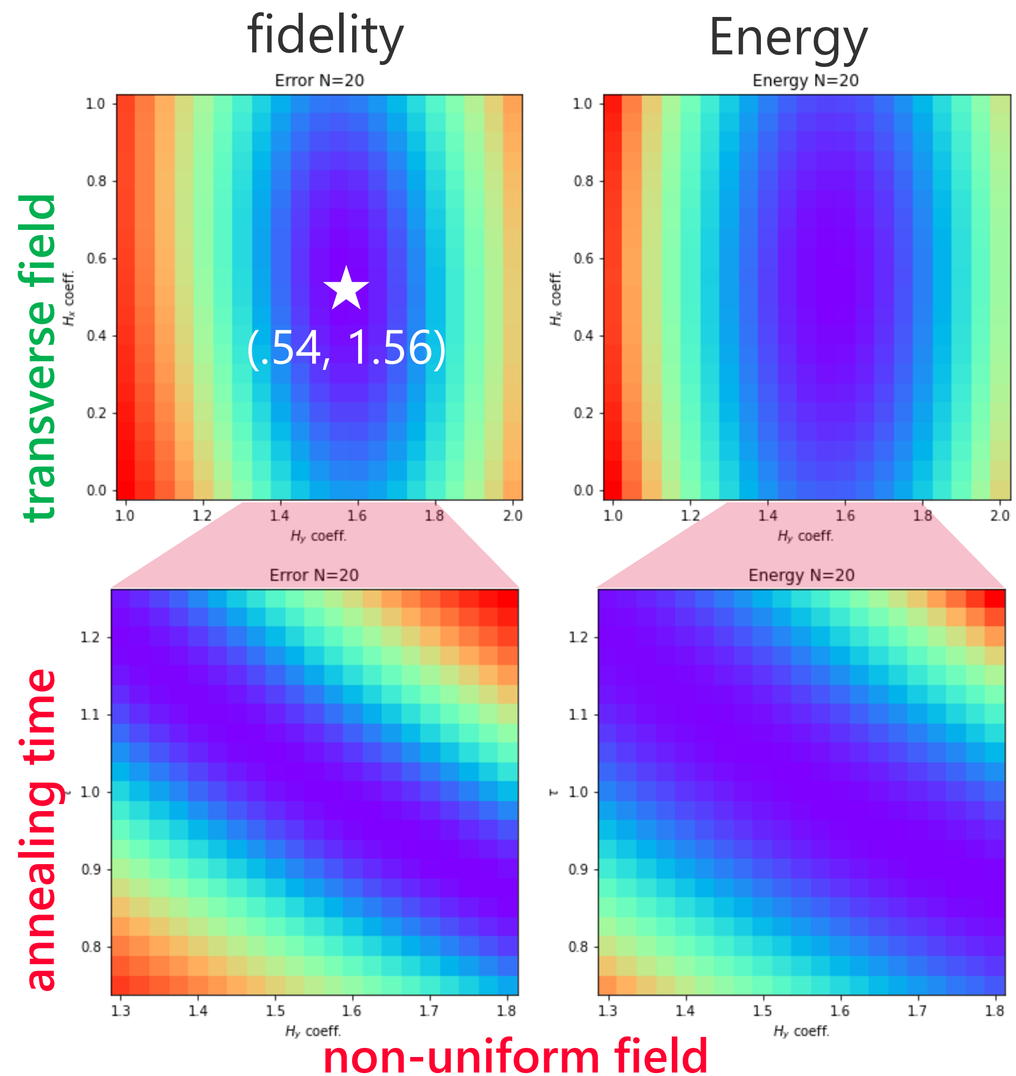
Parameters can be optimized by fidelity and energy.

<u>Robust:</u>

- transverse field

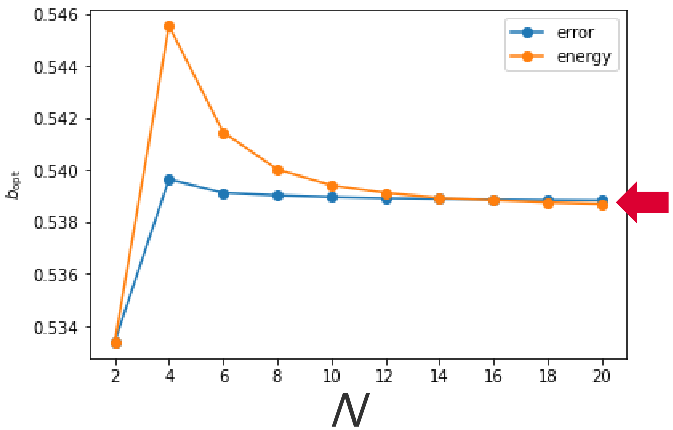
<u>Sensitive:</u> (reciprocal)

- non-uniform field
- annealing time

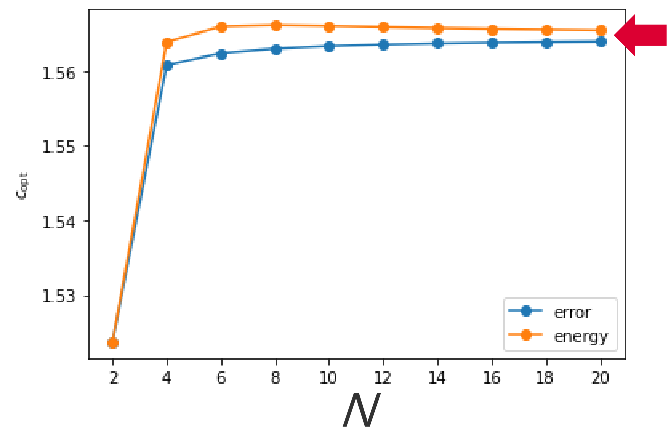


Variational optimization (size dependence)
Ferromagnetic, $N = 2 \sim 20$

b_{opt}^N

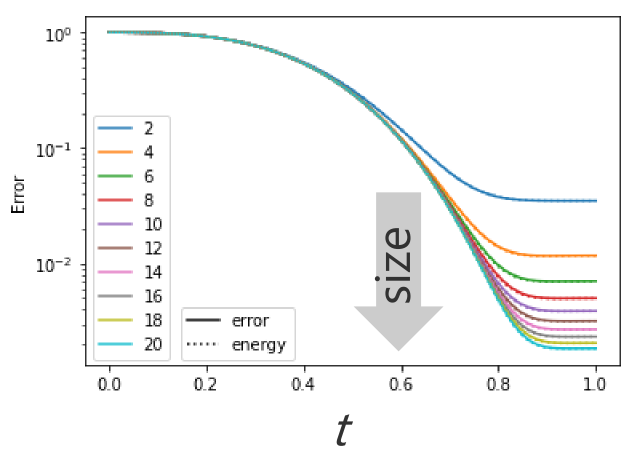


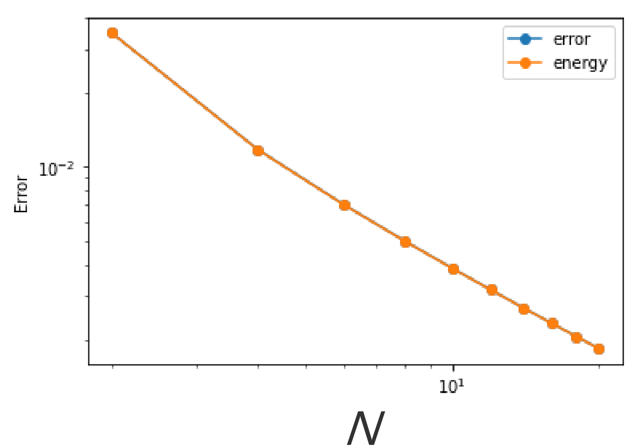
c_{opt}^N



use b_{opt}^N & c_{opt}^N for further analyses

Error ($= 1 - P_{\mathrm{GS}}$)





Scaling relation between size and optimization error is observed.

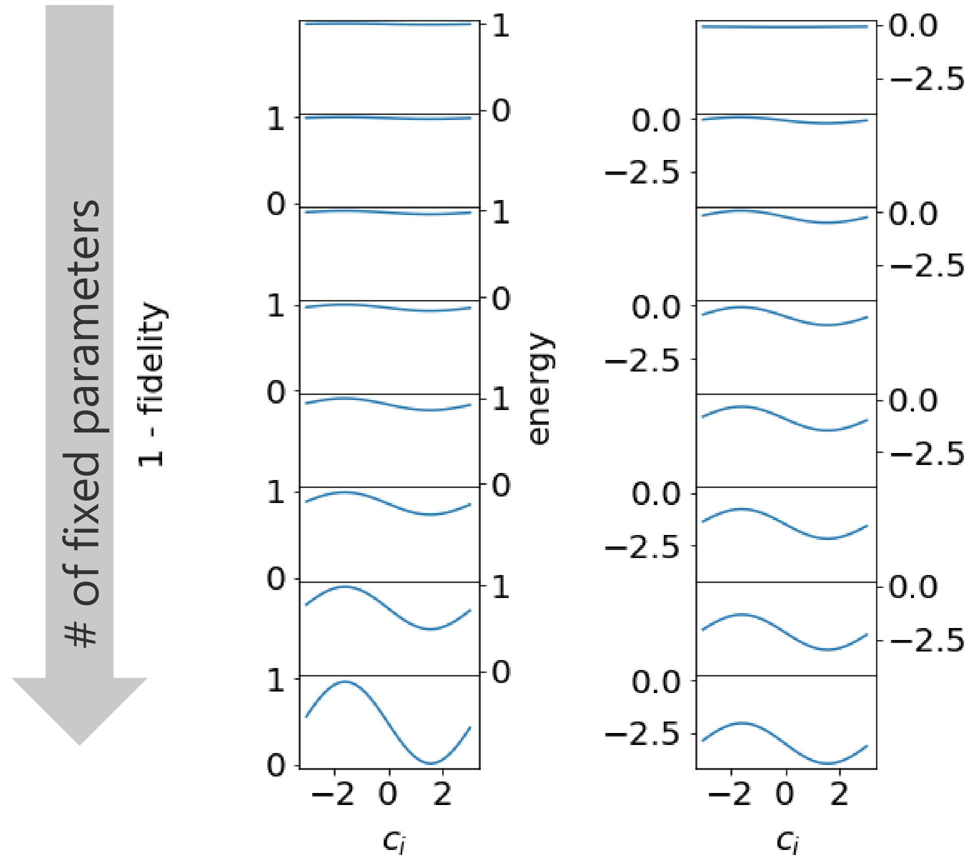
Sequential optimization (Ferromagnetic)

In general systems (e.g., random systems), we need to identify each parameter's sign (+ or −). For this purpose, one point estimation of the gradient at $c_i = 0$ is enough to identify the sign.

signs of gradients
= sign of parameters
= spin orientations in GS

Now, we need a little relaxed oracle to answer gradients(+/-) at $c_i = 0$.

Sequential estimation for N=8



Sequential QGO

Algorithm 1 Sequential QGO

Input: QA measure $f(b, \boldsymbol{c})$, b_{opt}^N, c_{opt}^N → parameters
- transverse field
- non-uniform field in Y

Output: a solution of the cost function

1: $b \leftarrow b_{\mathrm{opt}}^N$

2: $\boldsymbol{c} \leftarrow (0, \cdots, 0)$

3: **repeat**

4: $\quad \boldsymbol{g} \leftarrow \nabla f(b, \boldsymbol{c}) = \left(\frac{f(b, c_1+\Delta, \cdots, c_N) - f(b, \boldsymbol{c})}{\Delta}, \cdots, \frac{f(b, c_1, \cdots, c_N+\Delta) - f(b, \boldsymbol{c})}{\Delta}\right)$

5: $\quad i \leftarrow \arg\max_{j \in \{j | c_j = 0\}} |g_j|$ → sensitivity analysis at $c_i = 0$

6: $\quad c_i \leftarrow -c_{\mathrm{opt}}^N \operatorname{sign} g_i$

7: **until** $c_i \neq 0$ for all i

8: **return** $-\operatorname{sign} \boldsymbol{c}$

Sequential QGO of random systems

$$\mathcal{H} = A(t)\mathcal{H}_z + B(t)\,\mathcal{H}_x + \mathcal{H}_y$$

$\mathcal{H}_z = -\sum_{i<j} J_{ij}\sigma_i^z\sigma_j^z$: cost function

$\mathcal{H}_x = -\sum_i \sigma_i^x$: uniform transverse field

$\mathcal{H}_y = -\sum_i C_i(t)\sigma_i^y$: non-uniform y-field

This model is equivalent to Max-cut problem.

($J_{ij} = -w_{ij}$; w_{ij} is the weight matrix of the problem)

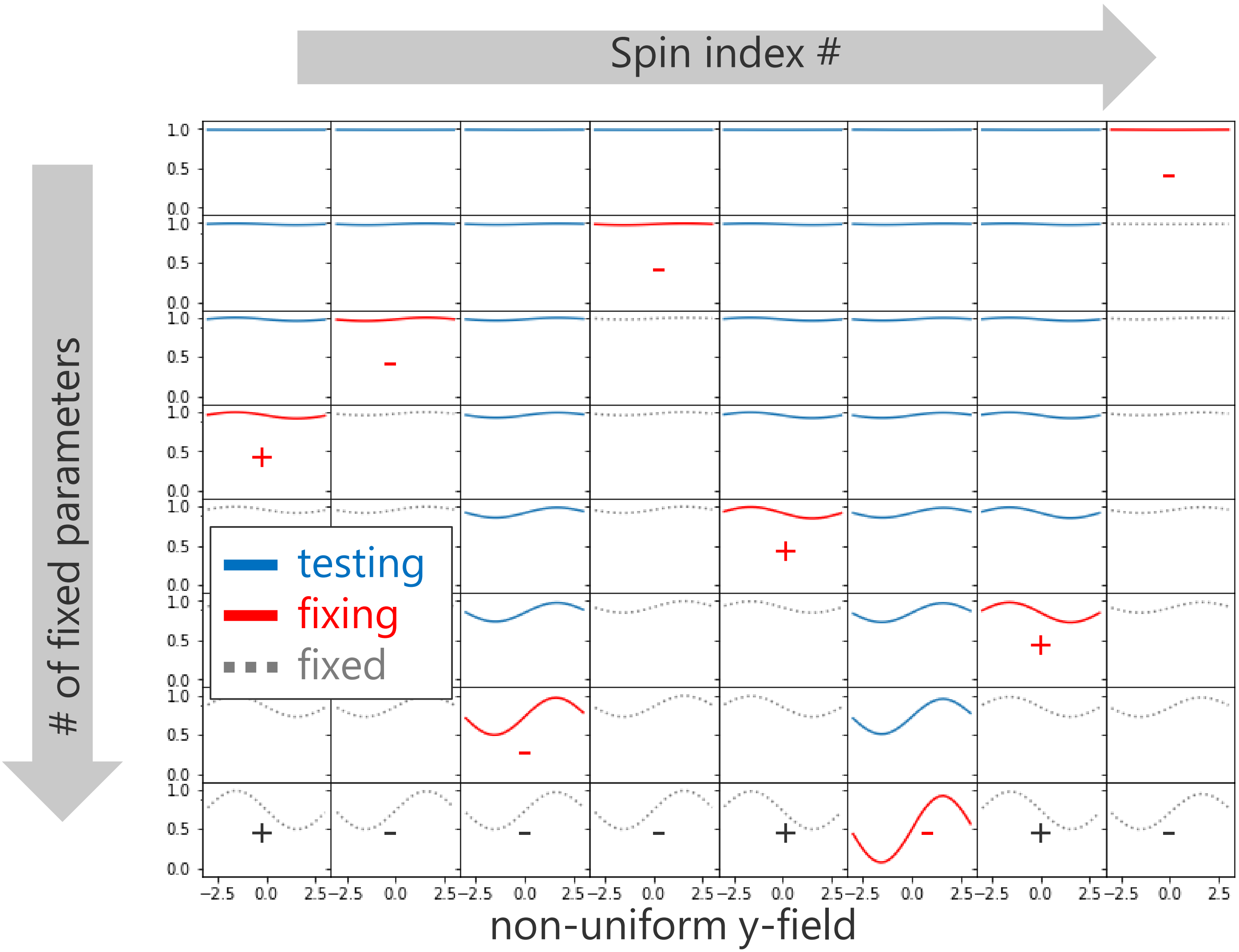
We employ SK model interaction:

$P(J_{ij}) \sim \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{J_{ij}^2}{2\sigma^2}\right)$, where $\sigma^2 = \frac{1}{N-1}$

Sequential QGO of random systems

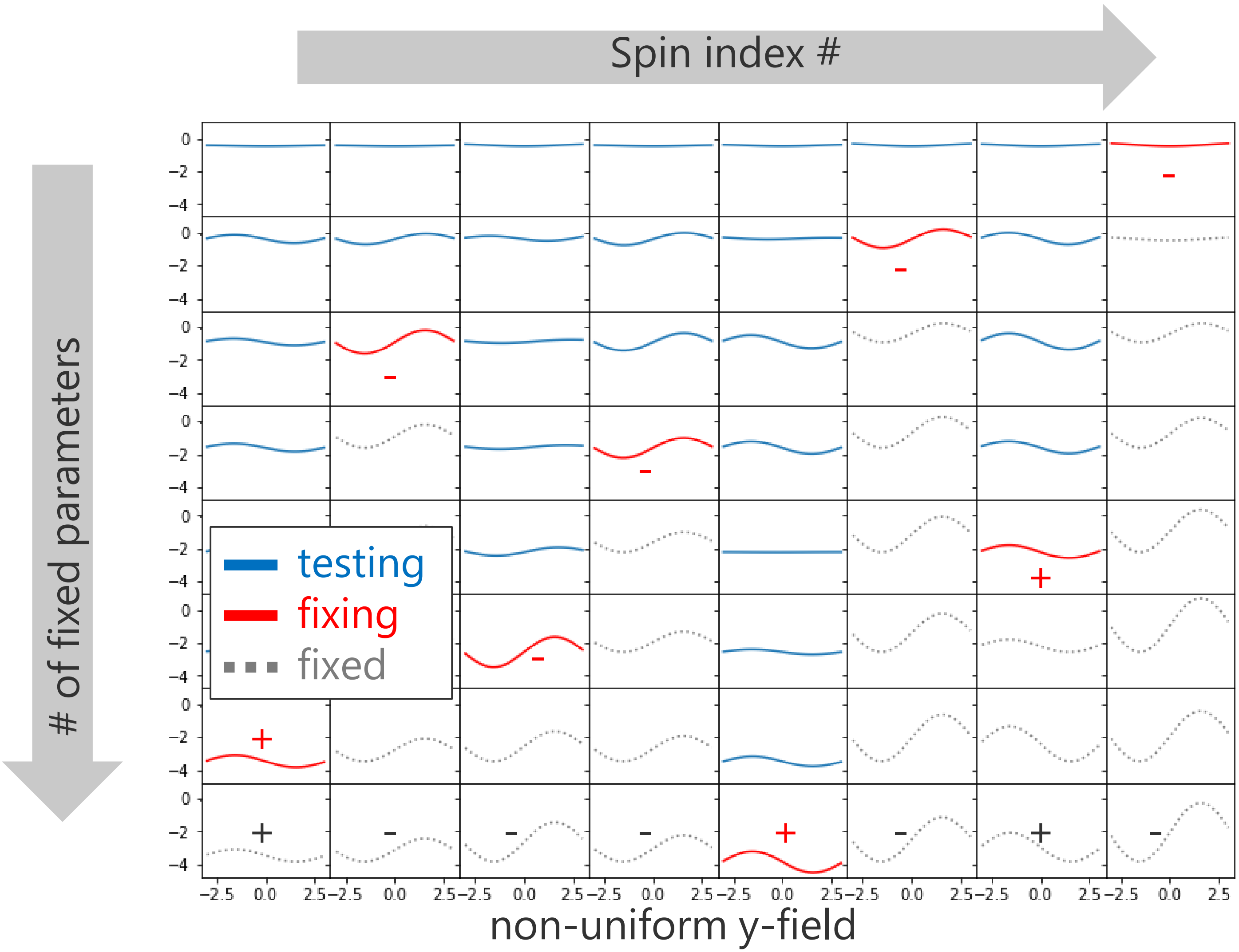
fidelity optimization

Success ratio 100%



Fidelity optimization always finds the ground state.

Sequential QGO of random systems



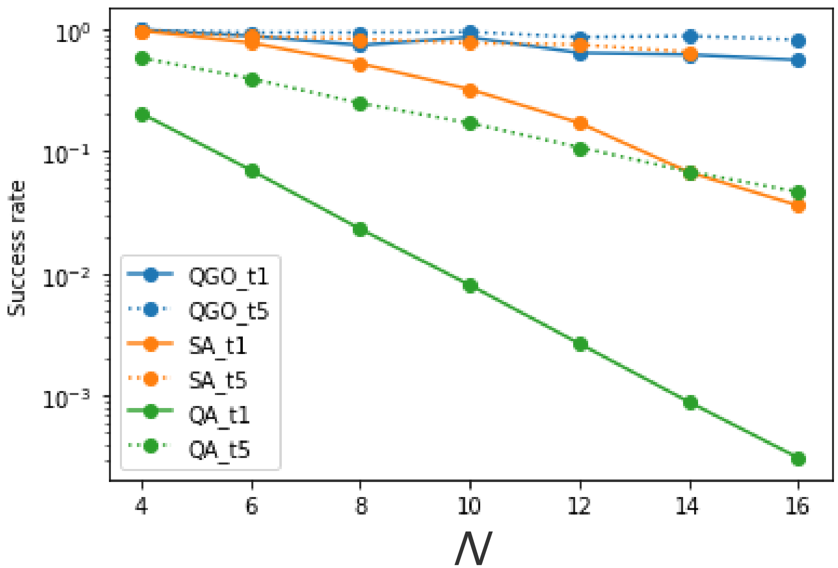
Energy optimization

Success ratio
??%

Energy optimization also finds the ground state.

Comparison with SA and QA

Short time (anneal time $\tau = 1, 5$) QA does not find the exact solution, while SA and QGO have a certain volume of basin.



100 random instances are analyzed and their performances is averaged.
While sequential QGO runs N(N+3)/2 QA tasks, no adjustment is applied.

QGO employs non-uniform y-field to probe the ground state or low energy states enriched during annealing process.

0. 量子アニーリング？

1. ブラックボックス最適化　　　　機械学習 > 量子 な話
   - 非可逆行列圧縮
   - 基板設計最適化

2. 量子アニーリング　　　　量子 > 機械学習 な話
   - 次世代量子アニーラ向けアルゴリズム

皆さん一緒に、

離散変数のブラックボックス最適化やりませんか？

DENSO
Crafting the Core