A Theoretical Approach to the Stochastic Cellular Automata Annealer and the Digital Annealer's Algorithm

Bruno Hideki Fukushima Kimura

Hokkaido University, Japan

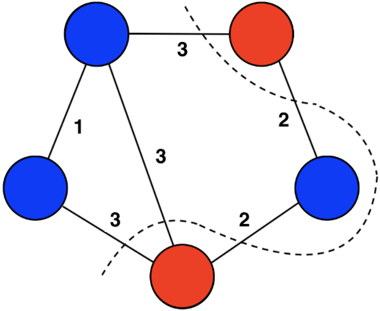
Discrete Mathematics Seminars
University of Amsterdam

February 23, 2023

Motivation

Weighted Max-Cut problem

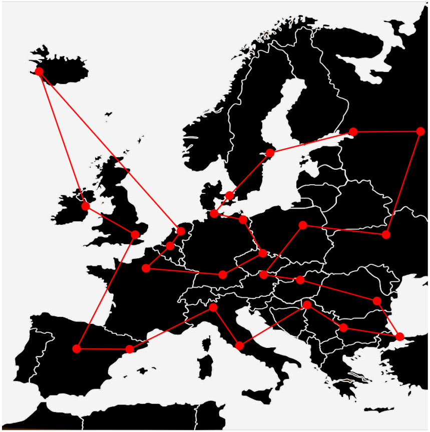
Given a graph $G = (V, E)$, find a partition of the vertex set into two sets S and $V \backslash S$ such that the total weight of edges connecting the set S and its complementary $V \backslash S$ is as large as possible.



Motivation

Traveling salesman problem

Given N cities and the distances $(d_{i,j})_{i,j=1}^{N}$ between each pair of cities, what is the shortest possible route that visits each city exactly once and returns to the origin city?”



Quadratic Unconstrained Binary Optimization (QUBO)

Quadratic Unconstrained Binary Optimization (QUBO)

Ising model:

Given a finite simple graph $G = (V, E)$ and spin-spin coupling constants $(J_{i,j})_{i,j \in V}$ and external fields $(h_i)_{i \in V}$, where $J_{i,j} = J_{j,i}$, let us define the Hamiltonian H by

$$H(\boldsymbol{\sigma}) = -\sum_{\{i,j\} \in E} J_{i,j}\sigma_i\sigma_j - \sum_{i \in V} h_i\sigma_i \tag{1}$$

for each $\boldsymbol{\sigma} \in \Omega = \{-1, +1\}^V$.

Quadratic Unconstrained Binary Optimization (QUBO)

Ising model:

Given a finite simple graph $G = (V, E)$ and spin-spin coupling constants $(J_{i,j})_{i,j \in V}$ and external fields $(h_i)_{i \in V}$, where $J_{i,j} = J_{j,i}$, let us define the Hamiltonian H by

$$H(\boldsymbol{\sigma}) = -\sum_{\{i,j\} \in E} J_{i,j}\sigma_i\sigma_j - \sum_{i \in V} h_i\sigma_i \tag{1}$$

for each $\boldsymbol{\sigma} \in \Omega = \{-1, +1\}^V$.

Let us consider the set of ground states of H:

$$\text{GS} = \{\boldsymbol{\sigma} : H(\boldsymbol{\sigma}) = \min_{\boldsymbol{\eta}} H(\boldsymbol{\eta})\}. \tag{2}$$

Example

Weighted Max-Cut problem

Example

Weighted Max-Cut problem

Given a graph $G = (V, E)$ and a family of weights $(w_{i,j})_{i,j \in V}$ such that $w_{i,j} = w_{j,i}$ and $w_{i,j} = 0$ if $\{i, j\} \notin E$. Then, let us consider the Hamiltonian

$$H(\boldsymbol{\sigma}) = \sum_{\{i,j\} \in E} w_{i,j}(1 - \sigma_i \sigma_j)/2 \tag{3}$$

for $\boldsymbol{\sigma} \in \{-1, 1\}^V$.

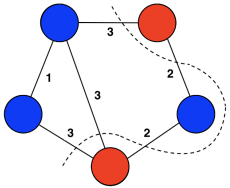
Example

Weighted Max-Cut problem

Given a graph $G = (V, E)$ and a family of weights $(w_{i,j})_{i,j \in V}$ such that $w_{i,j} = w_{j,i}$ and $w_{i,j} = 0$ if $\{i,j\} \notin E$. Then, let us consider the Hamiltonian

$$H(\boldsymbol{\sigma}) = \sum_{\{i,j\} \in E} w_{i,j}(1 - \sigma_i \sigma_j)/2 \tag{3}$$

for $\boldsymbol{\sigma} \in \{-1, 1\}^V$.



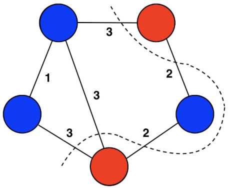
Example

Weighted Max-Cut problem

Given a graph $G = (V, E)$ and a family of weights $(w_{i,j})_{i,j \in V}$ such that $w_{i,j} = w_{j,i}$ and $w_{i,j} = 0$ if $\{i,j\} \notin E$. Then, let us consider the Hamiltonian

$$H(\boldsymbol{\sigma}) = \sum_{\{i,j\} \in E} w_{i,j}(1 - \sigma_i \sigma_j)/2 \quad (3)$$

for $\boldsymbol{\sigma} \in \{-1, 1\}^V$.



The weighted Max-Cut problem is equivalent to the minimization of H.

Example

Traveling salesman problem

Example

Traveling salesman problem

Given N cities and their distances $(d_{i,j})_{i,j=1}^{N}$, where $d_{i,j} = d_{j,i}$. For a spin configuration $\boldsymbol{\tau} = (\tau_{t,i})_{t,i=1}^{N} \in \{0,1\}^{N \times N}$, we have

$$\begin{cases} \tau_{t,i} = 1 & \text{if the city i is occupied at time t,} \\ \tau_{t,i} = 0 & \text{if the city i is NOT occupied at time t.} \end{cases} \tag{4}$$

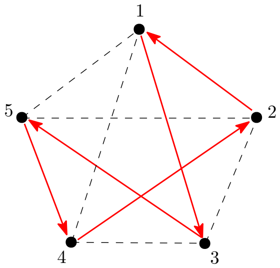
Example

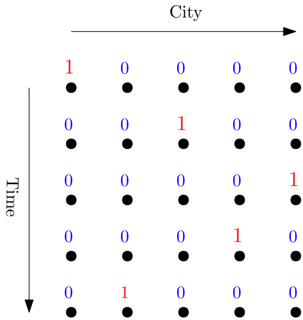
Traveling salesman problem

Given N cities and their distances $(d_{i,j})_{i,j=1}^{N}$, where $d_{i,j} = d_{j,i}$. For a spin configuration $\boldsymbol{\tau} = (\tau_{t,i})_{t,i=1}^{N} \in \{0,1\}^{N \times N}$, we have

$$\begin{cases} \tau_{t,i} = 1 & \text{if the city i is occupied at time t,} \\ \tau_{t,i} = 0 & \text{if the city i is NOT occupied at time t.} \end{cases} \tag{4}$$

In that way, $\boldsymbol{\tau}$ represents a legitimate trajectory iff $\begin{cases} \sum_i \tau_{t,i} = 1 & \text{for each } t \text{, and} \\ \sum_t \tau_{t,i} = 1 & \text{for each } i. \end{cases}$





Example

Traveling salesman problem

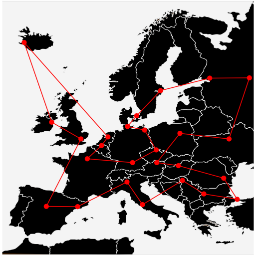
Given N cities and their distances $(d_{i,j})_{i,j=1}^{N}$, where $d_{i,j} = d_{j,i}$. For a spin configuration $\boldsymbol{\tau} = (\tau_{t,i})_{t,i=1}^{N} \in \{0,1\}^{N\times N}$, we have

$$\begin{cases} \tau_{t,i} = 1 & \text{if the city i is occupied at time t,} \\ \tau_{t,i} = 0 & \text{if the city i is NOT occupied at time t.} \end{cases} \quad (5)$$

In that way, $\boldsymbol{\tau}$ represents a legitimate trajectory iff $\begin{cases} \sum_i \tau_{t,i} = 1 & \text{for each } t \text{, and} \\ \sum_t \tau_{t,i} = 1 & \text{for each } i. \end{cases}$

Let us consider the Hamiltonian

$$H(\boldsymbol{\tau}) = A \sum_{t=1}^{N} \left(1 - \sum_{i=1}^{N} \tau_{t,i}\right)^2 + A \sum_{i=1}^{N} \left(1 - \sum_{t=1}^{N} \tau_{t,i}\right)^2 + B \sum_{t=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} d_{i,j} \tau_{t,i} \tau_{t+1,j} \quad (6)$$



where $\boldsymbol{\tau} \in \{0,1\}^{N\times N}$ is defined by $\tau_{t,i} = (1 + \sigma_{t,i})/2$.

If $0 < B \max\{d_{i,j}\} < A$, then the TSP is equivalent to the minimization of H.

Quadratic Unconstrained Binary Optimization (QUBO) problems

Quadratic Unconstrained Binary Optimization (QUBO) problems

NP-complete and NP-hard problems

- Černý V. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *J Optim Theory Appl* . 45 (1985) 41–51.
- Lucas A. Ising formulations of many NP problems. *Front Phys.* (2014) 12:5.

Quadratic Unconstrained Binary Optimization (QUBO) problems

NP-complete and NP-hard problems

- Černý V. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *J Optim Theory Appl* . 45 (1985) 41–51.
- Lucas A. Ising formulations of many NP problems. *Front Phys.* (2014) 12:5.

Finance

- Rosenberg G, Haghnegahdar P, Goddard P, Carr P, Wu K, de Prado ML. Solving the optimal trading trajectory problem using a quantum annealer. *IEEE J Select Top Signal Process.* (2016) 10:1053.

Quadratic Unconstrained Binary Optimization (QUBO) problems

NP-complete and NP-hard problems

- Černý V. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *J Optim Theory Appl* . 45 (1985) 41–51.
- Lucas A. Ising formulations of many NP problems. *Front Phys.* (2014) 12:5.

Finance

- Rosenberg G, Haghnegahdar P, Goddard P, Carr P, Wu K, de Prado ML. Solving the optimal trading trajectory problem using a quantum annealer. *IEEE J Select Top Signal Process.* (2016) 10:1053.

Chemistry

- Hernandez M, Zaribafiyan A, Aramon M, Naghibi M. A novel graph-based approach for determining molecular similarity. *arXiv:1601.06693.* (2016).
- Hernandez M, Aramon M. Enhancing quantum annealing performance for the molecular similarity problem. *Quantum Inform Process.* (2017) 16:133.

Quadratic Unconstrained Binary Optimization (QUBO) problems

NP-complete and NP-hard problems

- Černý V. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *J Optim Theory Appl* . 45 (1985) 41–51.
- Lucas A. Ising formulations of many NP problems. *Front Phys.* (2014) 12:5.

Finance

- Rosenberg G, Haghnegahdar P, Goddard P, Carr P, Wu K, de Prado ML. Solving the optimal trading trajectory problem using a quantum annealer. *IEEE J Select Top Signal Process.* (2016) 10:1053.

Chemistry

- Hernandez M, Zaribafiyan A, Aramon M, Naghibi M. A novel graph-based approach for determining molecular similarity. *arXiv:1601.06693.* (2016).
- Hernandez M, Aramon M. Enhancing quantum annealing performance for the molecular similarity problem. *Quantum Inform Process.* (2017) 16:133.

Biology

- Perdomo-Ortiz A, Dickson N, Drew-Brook M, Rose G, Aspuru-Guzik A. Finding low-energy conformations of lattice protein models by quantum annealing. *Sci Rep.* (2012) 2:571.
- Li RY, Di Felice R, Rohs R, Lidar DA. Quantum annealing versus classical machine learning applied to a simplified computational biology problem. *NPJ Quantum Inf.* (2018) 4:14.

Logistics and scheduling

- Venturelli D, Marchand DJJ, Rojo G. Quantum annealing implementation of job-shop scheduling. *arXiv:1506.08479v2.* (2015).
- Neukart F, Von Dollen D, Compostella G, Seidel C, Yarkoni S, Parney B. Traffic flow optimization using a quantum annealer. *Front ICT*. (2017) 4:29.

Logistics and scheduling

- Venturelli D, Marchand DJJ, Rojo G. Quantum annealing implementation of job-shop scheduling. *arXiv:1506.08479v2.* (2015).
- Neukart F, Von Dollen D, Compostella G, Seidel C, Yarkoni S, Parney B. Traffic flow optimization using a quantum annealer. *Front ICT*. (2017) 4:29.

Machine learning

- Crawford D, Levit A, Ghadermarzy N, Oberoi JS, Ronagh P. Reinforcement learning using quantum Boltzmann machines. *arXiv:1612.05695v2.* (2016).
- Khoshaman A, Vinci W, Denis B, Andriyash E, Amin MH. Quantum variational autoencoder. *Quantum Sci Technol.* (2019) 4:014001.
- Henderson M, Novak J, Cook T. Leveraging adiabatic quantum computation for election forecasting. *arXiv:1802.00069.* (2018).
- Levit A, Crawford D, Ghadermarzy N, Oberoi JS, Zahedinejad E, Ronagh P. Free energy-based reinforcement learning using a quantum processor. *arXiv:1706.00074.* (2017).

Logistics and scheduling

- Venturelli D, Marchand DJJ, Rojo G. Quantum annealing implementation of job-shop scheduling. *arXiv:1506.08479v2.* (2015).
- Neukart F, Von Dollen D, Compostella G, Seidel C, Yarkoni S, Parney B. Traffic flow optimization using a quantum annealer. *Front ICT*. (2017) 4:29.

Machine learning

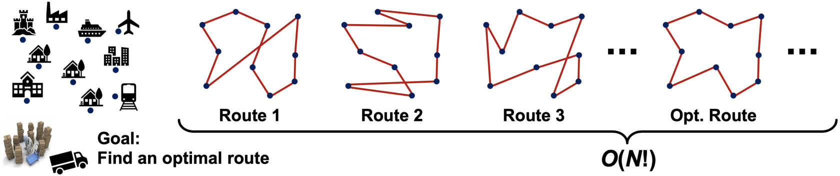
- Crawford D, Levit A, Ghadermarzy N, Oberoi JS, Ronagh P. Reinforcement learning using quantum Boltzmann machines. *arXiv:1612.05695v2.* (2016).
- Khoshaman A, Vinci W, Denis B, Andriyash E, Amin MH. Quantum variational autoencoder. *Quantum Sci Technol.* (2019) 4:014001.
- Henderson M, Novak J, Cook T. Leveraging adiabatic quantum computation for election forecasting. *arXiv:1802.00069.* (2018).
- Levit A, Crawford D, Ghadermarzy N, Oberoi JS, Zahedinejad E, Ronagh P. Free energy-based reinforcement learning using a quantum processor. *arXiv:1706.00074.* (2017).

VLSI design

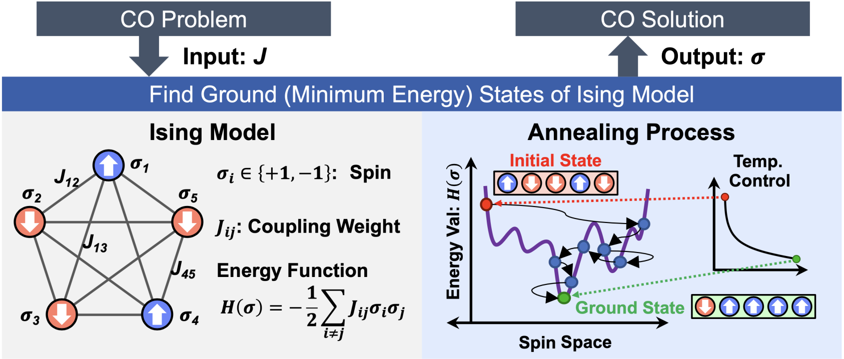
- S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220 (1983) 671–680.
- D.F. Wong, H.W. Leong, and H.W. Liu. Simulated Annealing for VLSI Design. *The Springer International Series in Engineering and Computer Science*. Springer US, 1988.

Combinatorial Optimization (CO) is Vital to Our Society

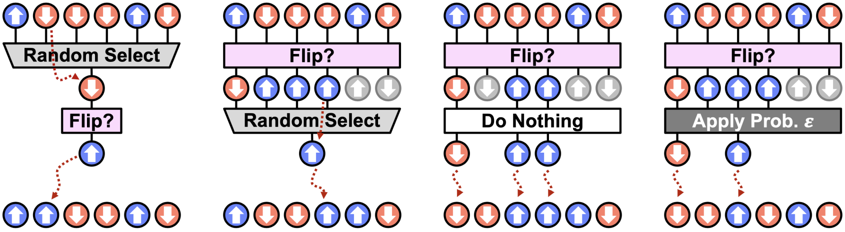
But an Exhaustive Search is Impractical for Large CO



Annealing Computation for CO using Ising Model



Comparison of Annealing Policies



SA
(Simulated Annealing)

- Single trial
- Single update

DA
(Digital Annealing)

- Parallel trial
- Single update

SCA
(Stochastic Cellular Automata Annealing)

- Parallel trial
- Parallel update

RPA
(Ratio-controlled Parallel Annealing)

- Parallel trial
- Managed parallel update

Simulated Annealing (SA)

Simulated Annealing (SA)

Let us consider the Metropolis dynamics at inverse temperature β:

$$P_\beta(\boldsymbol{\sigma},\boldsymbol{\tau}) \begin{cases} \frac{1}{|V|} \cdot e^{-\beta E_i(\boldsymbol{\sigma})^+} & \text{if } \boldsymbol{\tau} = \boldsymbol{\sigma}^i \text{ for some } i \in V, \\ 1 - \sum_{i \in V} P_\beta(\boldsymbol{\sigma},\boldsymbol{\sigma}^i) & \text{if } \boldsymbol{\tau} = \boldsymbol{\sigma}, \text{ and} \\ 0 & \text{otherwise;} \end{cases} \tag{7}$$

where $\boldsymbol{\sigma}^i$ is the configuration given by

$$(\boldsymbol{\sigma}^i)_j = \begin{cases} -\sigma_j & \text{if } j = i \\ \sigma_j & \text{otherwise,} \end{cases} \tag{8}$$

and

$$E_i(\boldsymbol{\sigma}) = H(\boldsymbol{\sigma}^i) - H(\boldsymbol{\sigma}). \tag{9}$$

Simulated Annealing (SA)

Let us consider the Metropolis dynamics at inverse temperature β:

$$P_\beta(\boldsymbol{\sigma},\boldsymbol{\tau}) \begin{cases} \frac{1}{|V|} \cdot e^{-\beta E_i(\boldsymbol{\sigma})^+} & \text{if } \boldsymbol{\tau} = \boldsymbol{\sigma}^i \text{ for some } i \in V, \\ 1 - \sum_{i \in V} P_\beta(\boldsymbol{\sigma},\boldsymbol{\sigma}^i) & \text{if } \boldsymbol{\tau} = \boldsymbol{\sigma}, \text{ and} \\ 0 & \text{otherwise;} \end{cases} \tag{7}$$

where $\boldsymbol{\sigma}^i$ is the configuration given by

$$(\boldsymbol{\sigma}^i)_j = \begin{cases} -\sigma_j & \text{if } j = i \\ \sigma_j & \text{otherwise,} \end{cases} \tag{8}$$

and

$$E_i(\boldsymbol{\sigma}) = H(\boldsymbol{\sigma}^i) - H(\boldsymbol{\sigma}). \tag{9}$$

Theorem (B. Hajek)

Let $(X_t)_{t \geq 0}$ be the discrete-time inhomogeneous Markov chain satisfying

$$\mathbb{P}(X_t = \boldsymbol{\sigma}_t | X_{t-1} = \boldsymbol{\sigma}_{t-1}, \dots, X_0 = \boldsymbol{\sigma}_0) = \mathbb{P}(X_t = \boldsymbol{\sigma}_t | X_{t-1} = \boldsymbol{\sigma}_{t-1}) = P_{\beta_t}(\boldsymbol{\sigma}_{t-1}, \boldsymbol{\sigma}_t) \tag{10}$$

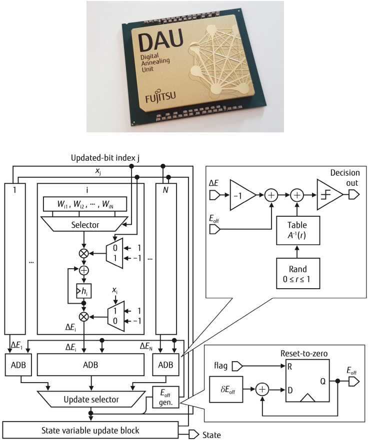
for every $t \geq 1$ and $\boldsymbol{\sigma}_0, \dots, \boldsymbol{\sigma}_t$ in Ω. There is $\gamma_c > 0$ such that if we choose $\beta_n = \frac{1}{\gamma} \log n$, then

$$\lim_{n \to \infty} \mathbb{P}(X_n \in GS) = 1 \tag{11}$$

holds if and only if $\gamma \geq \gamma_c$.

Digital Annealer's Algorithm

Fujitsu Laboratories has recently developed a CMOS hardware designed to solve fully connected quadratic unconstrained binary optimization (QUBO) problems, known as the Digital Annealer (DA).

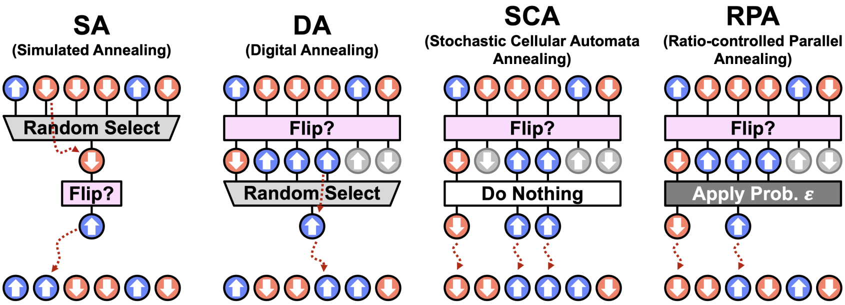


(a) The DA architecture

initial state ← random state;
for *each run* **do**
  initialize to initial state;
  for *each MC step* **do**
    update the temperature;
    for *each variable j, in parallel* **do**
      propose a flip using $e^{-\beta E_j^+}$;
      if accepted, record;
    end
    if *at least one flip accepted* **then**
      choose one flip uniformly at random among them;
      update the state and cavity fields in parallel;
    end
  end
end

(b) The DA's Algorithm

Comparison of Annealing Policies



SA
- Single trial
- Single update

DA
- Parallel trial
- Single update

SCA
- Parallel trial
- Parallel update

RPA
- Parallel trial
- Managed parallel update

In our framework, the Digital Annealer's Algorithm transition matrix P_β^{DA} at inverse temperature β is defined by

$$P_\beta^{\text{DA}}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \begin{cases} \displaystyle\sum_{\substack{S\subseteq V\\ S\ni i}} \frac{1}{|S|} \prod_{j\in S} e^{-\beta E_j(\boldsymbol{\sigma})^+} \prod_{j\in V\setminus S} (1 - e^{-\beta E_j(\boldsymbol{\sigma})^+}) & \text{if } \boldsymbol{\tau} = \boldsymbol{\sigma}^i, \\ \displaystyle\prod_{j\in V} (1 - e^{-\beta E_j(\boldsymbol{\sigma})^+}) & \text{if } \boldsymbol{\tau} = \boldsymbol{\sigma}, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

In our framework, the Digital Annealer's Algorithm transition matrix P_β^{DA} at inverse temperature β is defined by

$$P_\beta^{\mathrm{DA}}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \begin{cases} \sum\limits_{\substack{S\subseteq V\\ S\ni i}} \dfrac{1}{|S|} \prod\limits_{j\in S} e^{-\beta E_j(\boldsymbol{\sigma})^+} \prod\limits_{j\in V\setminus S} (1 - e^{-\beta E_j(\boldsymbol{\sigma})^+}) & \text{if } \boldsymbol{\tau} = \boldsymbol{\sigma}^i, \\ \prod\limits_{j\in V} (1 - e^{-\beta E_j(\boldsymbol{\sigma})^+}) & \text{if } \boldsymbol{\tau} = \boldsymbol{\sigma}, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

The algorithm works as follows.

In our framework, the Digital Annealer's Algorithm transition matrix P_β^{DA} at inverse temperature β is defined by

$$P_\beta^{\mathrm{DA}}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \begin{cases} \displaystyle\sum_{\substack{S\subseteq V\\ S\ni i}} \frac{1}{|S|} \prod_{j\in S} e^{-\beta E_j(\boldsymbol{\sigma})^+} \prod_{j\in V\setminus S} (1 - e^{-\beta E_j(\boldsymbol{\sigma})^+}) & \text{if } \boldsymbol{\tau} = \boldsymbol{\sigma}^i, \\ \displaystyle\prod_{j\in V} (1 - e^{-\beta E_j(\boldsymbol{\sigma})^+}) & \text{if } \boldsymbol{\tau} = \boldsymbol{\sigma}, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

The algorithm works as follows.

1. Given a state $X_t = \boldsymbol{\sigma}$ at time t, we propose a parallel-trial where each spin variable σ_j is assigned as eligible to be flipped with probability $\exp(-\beta_t E_j(\boldsymbol{\sigma})^+)$.

In our framework, the Digital Annealer's Algorithm transition matrix P_β^{DA} at inverse temperature β is defined by

$$P_\beta^{\mathrm{DA}}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \begin{cases} \displaystyle\sum_{\substack{S\subseteq V\\ S\ni i}} \frac{1}{|S|} \prod_{j\in S} e^{-\beta E_j(\boldsymbol{\sigma})^+} \prod_{j\in V\setminus S} (1-e^{-\beta E_j(\boldsymbol{\sigma})^+}) & \text{if } \boldsymbol{\tau}=\boldsymbol{\sigma}^i, \\ \displaystyle\prod_{j\in V}(1-e^{-\beta E_j(\boldsymbol{\sigma})^+}) & \text{if } \boldsymbol{\tau}=\boldsymbol{\sigma}, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

The algorithm works as follows.

1. Given a state $X_t = \boldsymbol{\sigma}$ at time t, we propose a parallel-trial where each spin variable σ_j is assigned as eligible to be flipped with probability $\exp(-\beta_t E_j(\boldsymbol{\sigma})^+)$.
2. If the set S of all vertices associated with eligible spin variables contains at least one element, then a vertex i is chosen uniformly at random from S, and we place $X_{t+1} = \boldsymbol{\sigma}^i$; otherwise, nothing is done, and we consider $X_{t+1} = \boldsymbol{\sigma}$.

Theorem (Fukushima-Kimura, Kawamoto, Noda, Sakai)

Let $(\beta_t)_{t\geq 1}$ *be a nondecreasing sequence of positive numbers such that* $\lim_{t\to\infty}\beta_t = +\infty$, *and let* $(X_t)_{t\geq 0}$ *be the discrete-time inhomogeneous Markov chain satisfying*

$$\mathbb{P}(X_t = \boldsymbol{\sigma}_t | X_{t-1} = \boldsymbol{\sigma}_{t-1}, \ldots, X_0 = \boldsymbol{\sigma}_0) = \mathbb{P}(X_t = \boldsymbol{\sigma}_t | X_{t-1} = \boldsymbol{\sigma}_{t-1}) = P^{DA}_{\beta_t}(\boldsymbol{\sigma}_{t-1}, \boldsymbol{\sigma}_t) \tag{12}$$

for every $t \geq 1$ *and* $\boldsymbol{\sigma}_0, \ldots, \boldsymbol{\sigma}_t$ *in* Ω.

There exists $\gamma_c > 0$ *such that the limit*

$$\lim_{t\to\infty} \mathbb{P}(X_t \in GS) = 1 \tag{13}$$

holds if and only if

$$\sum_{t=1}^{\infty} e^{-\beta_t \gamma_c} = +\infty. \tag{14}$$

Theorem (Fukushima-Kimura, Kawamoto, Noda, Sakai)

Let $(\beta_t)_{t\geq 1}$ be a nondecreasing sequence of positive numbers such that $\lim_{t\to\infty}\beta_t = +\infty$, and let $(X_t)_{t\geq 0}$ be the discrete-time inhomogeneous Markov chain satisfying

$$\mathbb{P}(X_t = \boldsymbol{\sigma}_t | X_{t-1} = \boldsymbol{\sigma}_{t-1}, \ldots, X_0 = \boldsymbol{\sigma}_0) = \mathbb{P}(X_t = \boldsymbol{\sigma}_t | X_{t-1} = \boldsymbol{\sigma}_{t-1}) = P^{DA}_{\beta_t}(\boldsymbol{\sigma}_{t-1}, \boldsymbol{\sigma}_t) \tag{12}$$

for every $t \geq 1$ and $\boldsymbol{\sigma}_0, \ldots, \boldsymbol{\sigma}_t$ in Ω.

There exists $\gamma_c > 0$ such that the limit

$$\lim_{t\to\infty} \mathbb{P}(X_t \in GS) = 1 \tag{13}$$

holds if and only if

$$\sum_{t=1}^{\infty} e^{-\beta_t \gamma_c} = +\infty. \tag{14}$$

In particular, if $(\beta_t)_{t\geq 1}$ assumes the form

$$\beta_t = \frac{1}{\gamma} \log t \tag{15}$$

then, equation (13) holds if and only if $\gamma \geq \gamma_c$.

Definition

We say $\boldsymbol{\tau}$ is reachable from $\boldsymbol{\sigma}$ at height E if there exists a path $\boldsymbol{\sigma} = \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_n = \boldsymbol{\tau}$ such that $\max_{0 \le k \le n} H(\boldsymbol{\sigma}_k) \le E$.

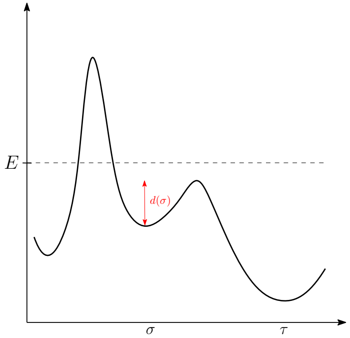
Definition

We say $\boldsymbol{\sigma}$ is a local minimum if there is no $\boldsymbol{\tau}$ satisfying $H(\boldsymbol{\tau}) < H(\boldsymbol{\sigma})$ which is reachable from $\boldsymbol{\sigma}$ at height $H(\boldsymbol{\sigma})$. So, the depth of a local minimum $\boldsymbol{\sigma}$ which is not a ground state is defined as

$$d(\boldsymbol{\sigma}) = \min\{E > 0 : \exists \boldsymbol{\tau} \text{ with } H(\boldsymbol{\tau}) < H(\boldsymbol{\sigma}) \text{ that is reachable from } \boldsymbol{\sigma} \text{ at height } H(\boldsymbol{\sigma}) + E\}. \quad (16)$$

The constant γ_c coincides with the depth of the second deepest local minimum, i.e.,

$$\gamma_c = \max\{d(\boldsymbol{\sigma}) : \boldsymbol{\sigma} \text{ is a local minimum not in GS}\}.$$



Simulations

Let us consider the following Hamiltonian:

$$H(\boldsymbol{\sigma}) = -\sum_{\{i,j\}\in E} J_{i,j}\sigma_i\sigma_j. \tag{17}$$

Simulations

Let us consider the following Hamiltonian:

$$H(\boldsymbol{\sigma}) = -\sum_{\{i,j\}\in E} J_{i,j}\sigma_i\sigma_j. \tag{17}$$

1. *Max-cut problem.* The Hamiltonian is defined in an Erdös-Rényi random graph $G(N,p)$, with spin-spin coupling satisfying $J_{i,j} = -1$ if $\{i,j\}$ is an edge of the graph and $J_{i,j} = 0$ otherwise.

Simulations

Let us consider the following Hamiltonian:

$$H(\boldsymbol{\sigma}) = -\sum_{\{i,j\}\in E} J_{i,j}\sigma_i\sigma_j. \tag{17}$$

1. *Max-cut problem.* The Hamiltonian is defined in an Erdös-Rényi random graph $G(N,p)$, with spin-spin coupling satisfying $J_{i,j} = -1$ if $\{i,j\}$ is an edge of the graph and $J_{i,j} = 0$ otherwise.
2. *Spin glasses.* Let us consider a spin glass Hamiltonian in a complete graph with N vertices, where the values for the spin-spin couplings $J_{i,j} = J_{j,i}$ are realizations of i.i.d. normal random variables.

Simulations

Let us consider the following Hamiltonian:

$$H(\boldsymbol{\sigma}) = -\sum_{\{i,j\}\in E} J_{i,j}\sigma_i\sigma_j. \qquad (17)$$

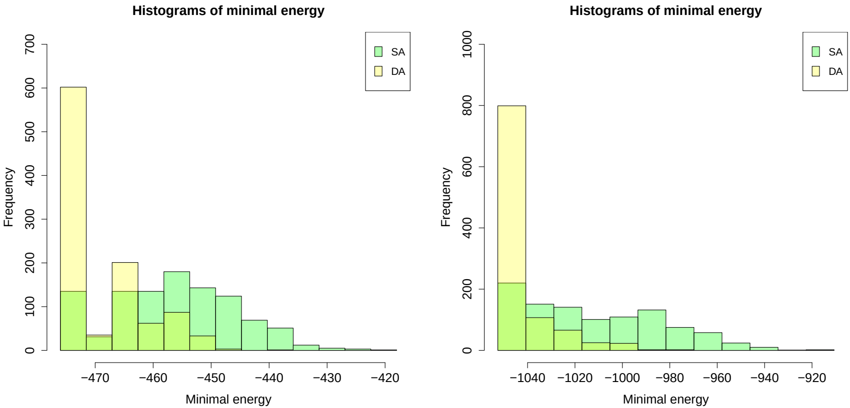
1. *Max-cut problem.* The Hamiltonian is defined in an Erdös-Rényi random graph $G(N,p)$, with spin-spin coupling satisfying $J_{i,j} = -1$ if $\{i,j\}$ is an edge of the graph and $J_{i,j} = 0$ otherwise.
2. *Spin glasses.* Let us consider a spin glass Hamiltonian in a complete graph with N vertices, where the values for the spin-spin couplings $J_{i,j} = J_{j,i}$ are realizations of i.i.d. normal random variables.

Each plot in the following figure illustrates the histogram of minimal energy achieved by the DA and Metropolis dynamics Considering a graph with $N = 128$ vertices and $M = 1024$ annealing trials, where on each trial we applied $L = 20000$ Markov chain steps and considered the exponential cooling schedule with initial temperature $T_{\text{init}} = 1000$ and final temperature $T_{\text{fin}} = 0.05$, explicitly, we considered

$$\frac{1}{\beta_t} = T_{\text{init}}\left(\frac{T_{\text{fin}}}{T_{\text{init}}}\right)^{\frac{t-1}{L-1}} \qquad (18)$$

for $t = 1, 2, \ldots, L$.

Simulations



(a) Max-cut problem, $p = 0.25$ (b) Spin-glass

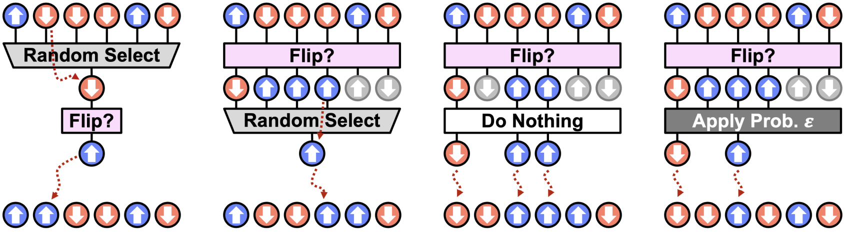
Figure: Histograms obtained by using the SA and DA, where $N = 128$.

Table: Summary of the simulations

Model	Success rate	
	SA	DA
Max-cut	7.52%	58.01%
Spin-glass	5.08%	40.72%

Comparison of Annealing Policies

SA (Simulated Annealing) | **DA** (Digital Annealing) | **SCA** (Stochastic Cellular Automata Annealing) | **RPA** (Ratio-controlled Parallel Annealing)



SA
- Single trial
- Single update

DA
- Parallel trial
- Single update

SCA
- Parallel trial
- Parallel update

RPA
- Parallel trial
- Managed parallel update

SCA

SCA

The extended Hamiltonian $\tilde{H}$ is defined by

$$\tilde{H}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = -\frac{1}{2} \sum_{i,j \in V} J_{i,j} \sigma_i \tau_j - \frac{1}{2} \sum_{i \in V} h_i (\sigma_i + \tau_i) - \frac{1}{2} \sum_{i \in V} q_i \sigma_i \tau_i \tag{19}$$

for each pair $\boldsymbol{\sigma}, \boldsymbol{\tau}$ of configurations in $\{-1, +1\}^V$.

SCA

The extended Hamiltonian $\tilde{H}$ is defined by

$$\tilde{H}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = -\frac{1}{2}\sum_{i,j \in V} J_{i,j}\sigma_i\tau_j - \frac{1}{2}\sum_{i \in V} h_i(\sigma_i + \tau_i) - \frac{1}{2}\sum_{i \in V} q_i\sigma_i\tau_i \qquad (19)$$

for each pair $\boldsymbol{\sigma}, \boldsymbol{\tau}$ of configurations in $\{-1, +1\}^V$.

Let us define the SCA transition probability $P^{SCA}_{\beta,q}$ by

$$P^{SCA}_{\beta,q}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \frac{e^{-\beta \tilde{H}(\boldsymbol{\sigma}, \boldsymbol{\tau})}}{\sum_{\boldsymbol{\tau}'} e^{-\beta \tilde{H}(\boldsymbol{\sigma}, \boldsymbol{\tau}')}}. \qquad (20)$$

SCA

The extended Hamiltonian $\tilde{H}$ is defined by

$$\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau}) = -\frac{1}{2}\sum_{i,j\in V} J_{i,j}\sigma_i\tau_j - \frac{1}{2}\sum_{i\in V} h_i(\sigma_i+\tau_i) - \frac{1}{2}\sum_{i\in V} q_i\sigma_i\tau_i \tag{19}$$

for each pair $\boldsymbol{\sigma},\boldsymbol{\tau}$ of configurations in $\{-1,+1\}^V$.

Let us define the SCA transition probability $P^{SCA}_{\beta,q}$ by

$$P^{SCA}_{\beta,q}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \frac{e^{-\beta\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau})}}{\sum_{\boldsymbol{\tau}'} e^{-\beta\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau}')}}. \tag{20}$$

It is straightforward to verify that the distribution $\pi^{SCA}_{\beta,q}$ defined by

$$\pi^{SCA}_{\beta,q}(\boldsymbol{\sigma}) = \frac{\sum_{\tau} e^{-\beta\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau})}}{\sum_{\boldsymbol{\sigma}',\boldsymbol{\tau}'} e^{-\beta\tilde{H}(\boldsymbol{\sigma}',\boldsymbol{\tau}')}} \tag{21}$$

is the stationary distribution for $P^{SCA}_{\beta,q}$.

Moreover, we can rewrite $P^{SCA}_{\beta,q}$ as

$$P^{SCA}_{\beta,q}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \prod_{i\in V} \frac{e^{\frac{\beta}{2}(\tilde{h}_i(\boldsymbol{\sigma})+q_i\sigma_i)\tau_i}}{2\cosh(\frac{\beta}{2}(\tilde{h}_i(\boldsymbol{\sigma})+q_i\sigma_i))}, \tag{22}$$

where the cavity fields $\tilde{h}_i(\boldsymbol{\sigma})$ are given by

$$\tilde{h}_i(\boldsymbol{\sigma}) = \sum_{j\in V} J_{i,j}\sigma_j + h_i. \tag{23}$$

Moreover, we can rewrite $P_{\beta,q}^{SCA}$ as

$$P_{\beta,q}^{SCA}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \prod_{i\in V} \frac{e^{\frac{\beta}{2}(\tilde{h}_i(\boldsymbol{\sigma})+q_i\sigma_i)\tau_i}}{2\cosh(\frac{\beta}{2}(\tilde{h}_i(\boldsymbol{\sigma})+q_i\sigma_i))}, \tag{22}$$

where the cavity fields $\tilde{h}_i(\boldsymbol{\sigma})$ are given by

$$\tilde{h}_i(\boldsymbol{\sigma}) = \sum_{j\in V} J_{i,j}\sigma_j + h_i. \tag{23}$$

Theorem (Fukushima-Kimura, Handa, Kamijima, Kamakura, Kawamura, Sakai)

For any non-negative $\boldsymbol{q}$*, if* β *is sufficiently small such that*

$$r \equiv \max_{x\in V}\left(\tanh\frac{\beta q_x}{2} + \sum_{y\in V}\tanh\frac{\beta|J_{x,y}|}{2}\right) < 1, \tag{24}$$

then $t_{\mathrm{mix}}^{\mathrm{SCA}}(\varepsilon)$ *obeys*

$$t_{\mathrm{mix}}(\varepsilon) \leq \left\lceil \frac{\log|V| - \log\varepsilon}{\log(1/r)} \right\rceil. \tag{25}$$

Moreover, we can rewrite $P^{SCA}_{\beta,q}$ as

$$P^{SCA}_{\beta,q}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \prod_{i\in V} \frac{e^{\frac{\beta}{2}(\tilde{h}_i(\boldsymbol{\sigma})+q_i\sigma_i)\tau_i}}{2\cosh(\frac{\beta}{2}(\tilde{h}_i(\boldsymbol{\sigma})+q_i\sigma_i))}, \tag{22}$$

where the cavity fields $\tilde{h}_i(\boldsymbol{\sigma})$ are given by

$$\tilde{h}_i(\boldsymbol{\sigma}) = \sum_{j\in V} J_{i,j}\sigma_j + h_i. \tag{23}$$

Theorem (Fukushima-Kimura, Handa, Kamijima, Kamakura, Kawamura, Sakai)

For any non-negative $\boldsymbol{q}$, *if* β *is sufficiently small such that*

$$r \equiv \max_{x\in V}\left(\tanh\frac{\beta q_x}{2} + \sum_{y\in V}\tanh\frac{\beta|J_{x,y}|}{2}\right) < 1, \tag{24}$$

then $t^{\mathrm{SCA}}_{\mathrm{mix}}(\varepsilon)$ *obeys*

$$t_{\mathrm{mix}}(\varepsilon) \le \left\lceil \frac{\log|V| - \log\varepsilon}{\log(1/r)} \right\rceil. \tag{25}$$

T.P. Hayes and A. Sinclair (2007) proved that the mixing time for the Glauber dynamics is $\Omega(|V|\log|V|)$.

Proof.

It suffices to show $\rho_{\mathrm{TM}}(P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\cdot),P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\tau},\cdot))\le r$ for all $\boldsymbol{\sigma},\boldsymbol{\tau}\in\Omega$ with $|D_{\boldsymbol{\sigma},\boldsymbol{\tau}}|=1$. If $|D_{\boldsymbol{\sigma},\boldsymbol{\tau}}|\ge 2$, then, by the triangle inequality along any sequence $(\boldsymbol{\eta}_0,\boldsymbol{\eta}_1,\dots,\boldsymbol{\eta}_{|D_{\boldsymbol{\sigma},\boldsymbol{\tau}}|})$ of spin configurarions that satisfy $\boldsymbol{\eta}_0=\boldsymbol{\sigma}$, $\boldsymbol{\eta}_{|D_{\boldsymbol{\sigma},\boldsymbol{\tau}}|}=\boldsymbol{\tau}$ and $|D_{\boldsymbol{\eta}_{j-1},\boldsymbol{\eta}_j}|=1$ for all $j=1,\dots,|D_{\boldsymbol{\sigma},\boldsymbol{\tau}}|$, we have

$$\rho_{\mathrm{TM}}\Big(P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\cdot),P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\tau},\cdot)\Big)\le\sum_{j=1}^{|D_{\boldsymbol{\sigma},\boldsymbol{\tau}}|}\rho_{\mathrm{TM}}\Big(P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\eta}_{j-1},\cdot),P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\eta}_j,\cdot)\Big)\le r|D_{\boldsymbol{\sigma},\boldsymbol{\tau}}|. \tag{26}$$

Suppose that $D_{\boldsymbol{\sigma},\boldsymbol{\tau}}=\{x\}$, i.e., $\boldsymbol{\tau}=\boldsymbol{\sigma}^x$. For any $\boldsymbol{\sigma}\in\Omega$ and $y\in V$, we let $p(\boldsymbol{\sigma},y)$ be the conditional SCA probability of $\sigma_y\to 1$ given that the others are fixed:

$$p(\boldsymbol{\sigma},y)=\frac{e^{\frac{\beta}{2}(\tilde h_y(\boldsymbol{\sigma})+q_y\sigma_y)}}{2\cosh(\frac{\beta}{2}(\tilde h_y(\boldsymbol{\sigma})+q_y\sigma_y))}=\frac{1+\tanh(\frac{\beta}{2}(\tilde h_y(\boldsymbol{\sigma})+q_y\sigma_y))}{2}. \tag{27}$$

Notice that $p(\boldsymbol{\sigma},y)\ne p(\boldsymbol{\sigma}^x,y)$ only when $y=x$ or $y\in N_x\equiv\{v\in V:J_{x,v}\ne 0\}$. Using this as a threshold function for i.i.d. uniform random variables $\{U_y\}_{y\in V}$ on $[0,1]$, we define the coupling (X,Y) of $P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\cdot)$ and $P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma}^x,\cdot)$ as

$$X_y=\begin{cases}+1 & [U_y\le p(\boldsymbol{\sigma},y)],\\ -1 & [U_y> p(\boldsymbol{\sigma},y)],\end{cases}\qquad Y_y=\begin{cases}+1 & [U_y\le p(\boldsymbol{\sigma}^x,y)],\\ -1 & [U_y> p(\boldsymbol{\sigma}^x,y)].\end{cases} \tag{28}$$

□

Proof.

Denote the measure of this coupling by $\mathbf{P}_{\boldsymbol{\sigma},\boldsymbol{\sigma}^x}$ and its expectation by $\mathbf{E}_{\boldsymbol{\sigma},\boldsymbol{\sigma}^x}$. Then we obtain

$$\mathbf{E}_{\boldsymbol{\sigma},\boldsymbol{\sigma}^x}[|D_{X,Y}|] = \mathbf{E}_{\boldsymbol{\sigma},\boldsymbol{\sigma}^x}\left[\sum_{y\in V} \mathbb{1}_{\{X_y \neq Y_y\}}\right] = \sum_{y\in V} \mathbf{P}_{\boldsymbol{\sigma},\boldsymbol{\sigma}^x}(X_y \neq Y_y) = \sum_{y\in V} |p(\boldsymbol{\sigma}, y) - p(\boldsymbol{\sigma}^x, y)|$$
$$= |p(\boldsymbol{\sigma}, x) - p(\boldsymbol{\sigma}^x, x)| + \sum_{y\in N_x} |p(\boldsymbol{\sigma}, y) - p(\boldsymbol{\sigma}^x, y)|, \quad (29)$$

where, by using the rightmost expression above satisfies

$$|p(\boldsymbol{\sigma}, x) - p(\boldsymbol{\sigma}^x, x)| \le \frac{1}{2}\left|\tanh\left(\frac{\beta\tilde{h}_x(\boldsymbol{\sigma})}{2} + \frac{\beta q_x}{2}\right) - \tanh\left(\frac{\beta\tilde{h}_x(\boldsymbol{\sigma})}{2} - \frac{\beta q_x}{2}\right)\right|, \quad (30)$$

and for $y \in N_x$,

$$|p(\boldsymbol{\sigma}, y) - p(\boldsymbol{\sigma}^x, y)| \le \frac{1}{2}\left|\tanh\left(\frac{\beta(\sum_{v\neq x} J_{v,y}\sigma_v + h_y + q_y\sigma_y)}{2} + \frac{\beta J_{x,y}}{2}\right)\right.$$
$$\left. - \tanh\left(\frac{\beta(\sum_{v\neq x} J_{v,y}\sigma_v + h_y + q_y\sigma_y)}{2} - \frac{\beta J_{x,y}}{2}\right)\right|. \quad (31)$$

Since $|\tanh(a+b) - \tanh(a-b)| \le 2\tanh|b|$ for any a, b, we can conclude

$$\rho_{\mathrm{TM}}\left(P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\cdot), P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma}^x,\cdot)\right) \le \mathbf{E}_{\boldsymbol{\sigma},\boldsymbol{\sigma}^x}[|D_{X,Y}|] \le \tanh\frac{\beta q_x}{2} + \sum_{y\in N_x} \tanh\frac{\beta|J_{x,y}|}{2} \le r. \quad (32)$$

Theorem (Fukushima-Kimura, Handa, Kamijima, Kamakura, Kawamura, Sakai)

Suppose that the pinning parameters $q = (q_i)_{i \in V}$ *satisfy* $q_i \geq \lambda/2$*, where* λ *is the largest eigenvalue of the matrix* $(-J_{i,j})_{i,j \in V}$*. For any non-decreasing sequence* $(\beta_t)_{t \geq 1}$ *satisfying* $\lim_{t\uparrow\infty} \beta_t = \infty$*, we have*

$$\sum_{t=1}^{\infty} \|\pi^{SCA}_{\beta_{t+1},q} - \pi^{SCA}_{\beta_t,q}\|_{TV} < \infty, \qquad \lim_{t\uparrow\infty} \|\pi^{SCA}_{\beta_t,q} - \pi^{G}_{\infty}\|_{TV} = 0. \tag{33}$$

In particular, if we choose $(\beta_t)_{t \geq 1}$ *as*

$$\beta_t = \frac{\log t}{\Gamma}, \qquad \Gamma = \sum_{i \in V} \Gamma_i, \qquad \Gamma_i = \sum_{j \in V} |J_{i,j}| + |h_i| + q_i, \tag{34}$$

then we obtain

$$\sum_{t=1}^{\infty} \left(1 - \delta(P^{SCA}_{\beta_t,q})\right) = \infty. \tag{35}$$

As a result, for any initial $j \geq 1$*,*

$$\lim_{t\to\infty} \sup_{\mu} \left\|\mu P^{SCA}_{\beta_j,q} P^{SCA}_{\beta_{j+1},q} \cdots P^{SCA}_{\beta_t,q} - \pi^{G}_{\infty}\right\|_{TV} = 0. \tag{36}$$

Proof

Proof

Step 1. Let us show

$$\lim_{t\uparrow\infty} \|\pi^{SCA}_{\beta_t,q} - \pi^{G}_{\infty}\|_{TV} = 0 \tag{37}$$

We first define

$$\mu_\beta(\boldsymbol{\sigma},\boldsymbol{\tau}) = \frac{e^{-\beta\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau})}}{\sum_{\boldsymbol{\xi},\boldsymbol{\eta}} e^{-\beta\tilde{H}(\boldsymbol{\xi},\boldsymbol{\eta})}} \equiv \frac{e^{-\beta(\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau})-m)}}{\sum_{\boldsymbol{\xi},\boldsymbol{\eta}} e^{-\beta(\tilde{H}(\boldsymbol{\xi},\boldsymbol{\eta})-m)}}, \tag{38}$$

where $m = \min_{\boldsymbol{\sigma},\boldsymbol{\eta}} \tilde{H}(\boldsymbol{\sigma},\boldsymbol{\eta})$. We conclude that

$$\mu_\beta(\boldsymbol{\sigma},\boldsymbol{\tau}) = \frac{e^{-\beta(\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau})-m)}}{|\mathrm{GS}| + \sum_{\boldsymbol{\xi},\boldsymbol{\eta}:\tilde{H}(\boldsymbol{\xi},\boldsymbol{\eta})>m} e^{-\beta(\tilde{H}(\boldsymbol{\xi},\boldsymbol{\eta})-m)}} \xrightarrow[\beta\uparrow\infty]{} \underbrace{\frac{\mathbf{1}_{\boldsymbol{\sigma}\in\mathrm{GS}}}{|\mathrm{GS}|}}_{\pi^{\mathrm{G}}_{\infty}(\boldsymbol{\sigma})} \delta_{\boldsymbol{\sigma},\boldsymbol{\tau}}. \tag{39}$$

Summing this over $\boldsymbol{\tau}$ yields the second relation in (33).

Proof

Step 1. Let us show

$$\lim_{t\uparrow\infty} \|\pi^{SCA}_{\beta_t,q} - \pi^{G}_{\infty}\|_{TV} = 0 \tag{37}$$

We first define

$$\mu_\beta(\boldsymbol{\sigma},\boldsymbol{\tau}) = \frac{e^{-\beta\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau})}}{\sum_{\boldsymbol{\xi},\boldsymbol{\eta}} e^{-\beta\tilde{H}(\boldsymbol{\xi},\boldsymbol{\eta})}} \equiv \frac{e^{-\beta(\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau})-m)}}{\sum_{\boldsymbol{\xi},\boldsymbol{\eta}} e^{-\beta(\tilde{H}(\boldsymbol{\xi},\boldsymbol{\eta})-m)}}, \tag{38}$$

where $m = \min_{\boldsymbol{\sigma},\boldsymbol{\eta}} \tilde{H}(\boldsymbol{\sigma},\boldsymbol{\eta})$. We conclude that

$$\mu_\beta(\boldsymbol{\sigma},\boldsymbol{\tau}) = \frac{e^{-\beta(\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau})-m)}}{|\mathrm{GS}| + \sum_{\boldsymbol{\xi},\boldsymbol{\eta}:\tilde{H}(\boldsymbol{\xi},\boldsymbol{\eta})>m} e^{-\beta(\tilde{H}(\boldsymbol{\xi},\boldsymbol{\eta})-m)}} \xrightarrow[\beta\uparrow\infty]{} \underbrace{\frac{\mathbf{I}_{\boldsymbol{\sigma}\in\mathrm{GS}}}{|\mathrm{GS}|}}_{\pi^{\mathrm{G}}_{\infty}(\boldsymbol{\sigma})} \delta_{\boldsymbol{\sigma},\boldsymbol{\tau}}. \tag{39}$$

Summing this over $\boldsymbol{\tau}$ yields the second relation in (33).

Step 2. Let us show

$$\sum_{t=1}^{\infty} \|\pi^{SCA}_{\beta_{t+1},q} - \pi^{SCA}_{\beta_t,q}\|_{TV} < \infty \tag{40}$$

To show the first relation in (33), we note that

$$\frac{\partial\mu_\beta(\boldsymbol{\sigma},\boldsymbol{\tau})}{\partial\beta} = \Big(\mathbb{E}_{\mu_\beta}[\tilde{H}] - \tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau})\Big)\mu_\beta(\boldsymbol{\sigma},\boldsymbol{\tau}), \tag{41}$$

and that $\mathbb{E}_{\mu_\beta}[\tilde{H}] \equiv \sum_{\boldsymbol{\sigma},\boldsymbol{\tau}} \tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau})\,\mu_\beta(\boldsymbol{\sigma},\boldsymbol{\tau})$ tends to m as $\beta\uparrow\infty$, due to (39).

Therefore, $\frac{\partial}{\partial \beta}\mu_\beta(\boldsymbol{\sigma},\boldsymbol{\tau}) > 0$ for all β if $\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau}) = m$, while it is negative for sufficiently large β if $\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau}) > m$. Let $n \in \mathbb{N}$ be such that, as long as $\beta \geq \beta_n$, (41) is negative for all pairs $(\boldsymbol{\sigma},\boldsymbol{\tau})$ satisfying $\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\tau}) > m$. As a result,

$$\begin{aligned}
&\sum_{t=n}^{N} \|\pi^{\mathrm{SCA}}_{\beta_{t+1},\boldsymbol{q}} - \pi^{\mathrm{SCA}}_{\beta_t,\boldsymbol{q}}\|_{\mathrm{TV}} \\
&= \frac{1}{2}\sum_{\boldsymbol{\sigma}\in\mathrm{GS}}\sum_{t=n}^{N}|\pi^{\mathrm{SCA}}_{\beta_{t+1},\boldsymbol{q}}(\boldsymbol{\sigma}) - \pi^{\mathrm{SCA}}_{\beta_t,\boldsymbol{q}}(\boldsymbol{\sigma})| + \frac{1}{2}\sum_{\boldsymbol{\sigma}\notin\mathrm{GS}}\sum_{t=n}^{N}|\pi^{\mathrm{SCA}}_{\beta_{t+1},\boldsymbol{q}}(\boldsymbol{\sigma}) - \pi^{\mathrm{SCA}}_{\beta_t,\boldsymbol{q}}(\boldsymbol{\sigma})| \\
&\leq \frac{1}{2}\sum_{\boldsymbol{\sigma}\in\mathrm{GS}}\sum_{t=n}^{N}\big(\mu_{\beta_{t+1}}(\boldsymbol{\sigma},\boldsymbol{\sigma}) - \mu_{\beta_t}(\boldsymbol{\sigma},\boldsymbol{\sigma})\big) + \frac{1}{2}\sum_{\boldsymbol{\sigma}\in\mathrm{GS}}\sum_{\boldsymbol{\tau}\neq\boldsymbol{\sigma}}\sum_{t=n}^{N}\big(\mu_{\beta_t}(\boldsymbol{\sigma},\boldsymbol{\tau}) - \mu_{\beta_{t+1}}(\boldsymbol{\sigma},\boldsymbol{\tau})\big) \\
&\quad + \frac{1}{2}\sum_{\boldsymbol{\sigma}\notin\mathrm{GS}}\sum_{t=n}^{N}\big(\pi^{\mathrm{SCA}}_{\beta_t,\boldsymbol{q}}(\boldsymbol{\sigma}) - \pi^{\mathrm{SCA}}_{\beta_{t+1},\boldsymbol{q}}(\boldsymbol{\sigma})\big) \\
&= \frac{1}{2}\sum_{\boldsymbol{\sigma}\in\mathrm{GS}}\big(\mu_{\beta_{N+1}}(\boldsymbol{\sigma},\boldsymbol{\sigma}) - \mu_{\beta_n}(\boldsymbol{\sigma},\boldsymbol{\sigma})\big) + \frac{1}{2}\sum_{\boldsymbol{\sigma}\in\mathrm{GS}}\sum_{\boldsymbol{\tau}\neq\boldsymbol{\sigma}}\big(\mu_{\beta_n}(\boldsymbol{\sigma},\boldsymbol{\tau}) - \mu_{\beta_{N+1}}(\boldsymbol{\sigma},\boldsymbol{\tau})\big) \\
&\quad + \frac{1}{2}\sum_{\boldsymbol{\sigma}\notin\mathrm{GS}}\big(\pi^{\mathrm{SCA}}_{\beta_n,\boldsymbol{q}}(\boldsymbol{\sigma}) - \pi^{\mathrm{SCA}}_{\beta_{N+1},\boldsymbol{q}}(\boldsymbol{\sigma})\big) \\
&\leq \frac{3}{2} \qquad (42)
\end{aligned}$$

holds uniformly for $N \geq n$.

Step 3. Let us show

$$\sum_{t=1}^{\infty}\left(1-\delta(P^{SCA}_{\beta_t,q})\right)=\infty. \tag{43}$$

To show the equation above, we use the following bound on $P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}$, which holds uniformly in $(\boldsymbol{\sigma},\boldsymbol{\tau})$:

$$\begin{aligned} P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\boldsymbol{\tau})=\prod_{x\in V}\frac{e^{\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x)\tau_x}}{2\cosh(\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x))} \quad &\geq \quad \prod_{x\in V}\frac{1}{1+e^{\beta|\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x|}} \\ &\geq \quad \prod_{x\in V}\frac{e^{-\beta\Gamma_x}}{2}=\frac{e^{-\beta\Gamma}}{2^{|V|}}. \end{aligned} \tag{44}$$

Then, we obtain

$$\sum_{t=1}^{\infty}\left(1-\delta(P^{\mathrm{SCA}}_{\beta_t,\boldsymbol{q}})\right)=\sum_{t=1}^{\infty}\min_{\boldsymbol{\sigma},\boldsymbol{\eta}}\sum_{\boldsymbol{\tau}}P^{\mathrm{SCA}}_{\beta_t,\boldsymbol{q}}(\boldsymbol{\sigma},\boldsymbol{\tau})\wedge P^{\mathrm{SCA}}_{\beta_t,\boldsymbol{q}}(\boldsymbol{\eta},\boldsymbol{\tau})\geq\sum_{t=1}^{\infty}e^{-\beta_t\Gamma}, \tag{45}$$

which diverges, as required, under the cooling schedule (34). This completes the proof of the theorem.

Ratio-controlled Parallel Annealing (RPA or ε-SCA)

Ratio-controlled Parallel Annealing (RPA or ε-SCA)

Given the inverse temperature $\beta \geq 0$ and a number $\varepsilon \in [0,1]$, let the transition kernel of the ε-SCA be defined by

$$P_{\beta,\varepsilon}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \prod_{i:\sigma_i=-\tau_i} \big(\varepsilon p_i(\boldsymbol{\sigma})\big) \prod_{j:\sigma_j=\tau_j} \big(1-\varepsilon p_j(\boldsymbol{\sigma})\big), \tag{46}$$

where we recall that

$$p_i(\boldsymbol{\sigma}) = \frac{e^{-\frac{\beta}{2}\tilde{h}_i(\boldsymbol{\sigma})\sigma_i}}{2\cosh(\frac{\beta}{2}\tilde{h}_i(\boldsymbol{\sigma}))} \tag{47}$$

is the probability of flipping the spin σ_i from the configuration $\boldsymbol{\sigma}$ disregarding a pinning parameter at i.

Ratio-controlled Parallel Annealing (RPA or ε-SCA)

Given the inverse temperature $\beta \geq 0$ and a number $\varepsilon \in [0,1]$, let the transition kernel of the ε-SCA be defined by

$$P_{\beta,\varepsilon}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \prod_{i:\sigma_i=-\tau_i} \big(\varepsilon p_i(\boldsymbol{\sigma})\big) \prod_{j:\sigma_j=\tau_j} \big(1-\varepsilon p_j(\boldsymbol{\sigma})\big), \tag{46}$$

where we recall that

$$p_i(\boldsymbol{\sigma}) = \frac{e^{-\frac{\beta}{2}\tilde{h}_i(\boldsymbol{\sigma})\sigma_i}}{2\cosh\left(\frac{\beta}{2}\tilde{h}_i(\boldsymbol{\sigma})\right)} \tag{47}$$

is the probability of flipping the spin σ_i from the configuration $\boldsymbol{\sigma}$ disregarding a pinning parameter at i.

Theorem (Fukushima-Kimura, Kamijima, Kawamura, Sakai)

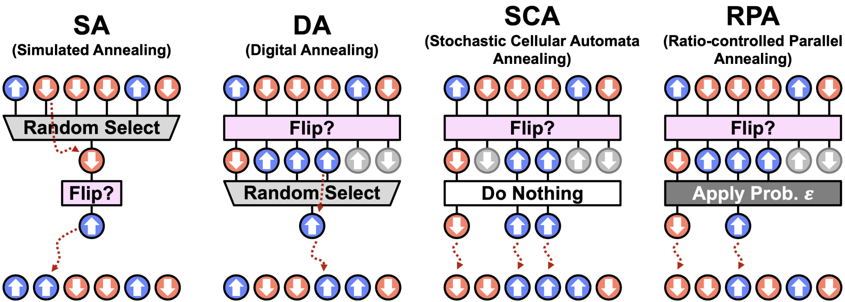
For any parameter $\varepsilon \in (0,1]$, if β is sufficiently small such that

$$r \equiv (1-\varepsilon) + \varepsilon \max_{i\in V}\left(\sum_{j\in V} \tanh\frac{\beta|J_{i,j}|}{2}\right) < 1, \tag{48}$$

then t_{mix} satisfies

$$t_{\mathrm{mix}}(\delta) \leq \left\lceil \frac{\log|V| - \log\delta}{\log(1/r)} \right\rceil. \tag{49}$$

Comparison of Annealing Policies



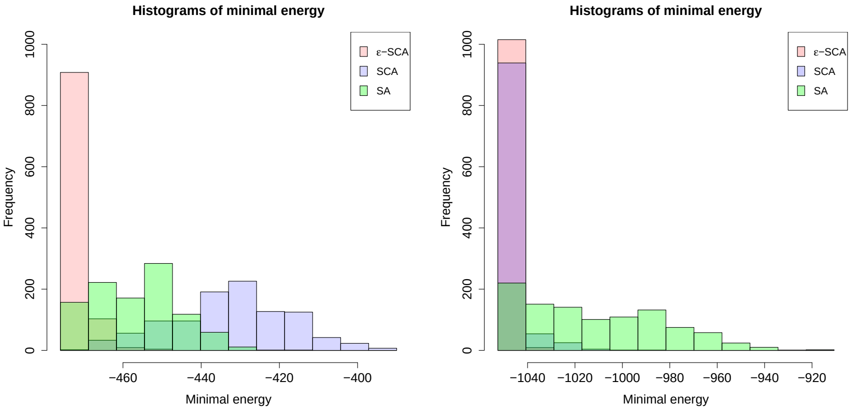
SA
- Single trial
- Single update

DA
- Parallel trial
- Single update

SCA
- Parallel trial
- Parallel update

RPA
- Parallel trial
- Managed parallel update

Simulations

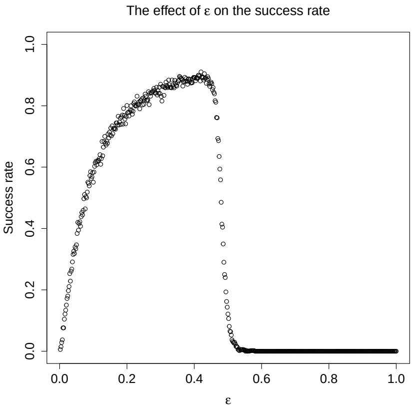


(a) Max-cut problem, $p = 0.25$

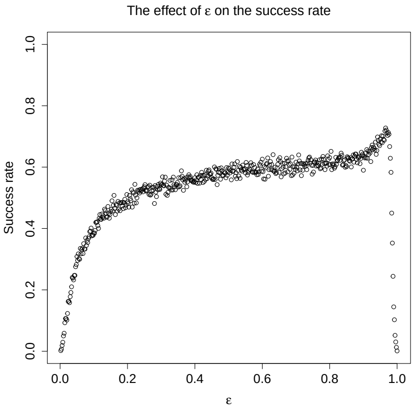
(b) Spin-glass

Figure: Histograms obtained by using the ε-SCA, SCA and Glauber dynamics, where $N = 128$.

Simulations



(a) Max-cut problem, $p = 0.25$



(b) Spin-glass

Figure: Success rate dependence on ε.

Table: Summary of the simulations

Model	Success rate			
	ε-SCA	SCA	SA	DA
Max-cut	85.9%	0%	7.52%	58.01%
Spin-glass	59.28%	40.82%	5.08%	40.72%

Table: Summary of the simulations

Model	Success rate			
	ε-SCA	SCA	SA	DA
Max-cut	85.9%	0%	7.52%	58.01%
Spin-glass	59.28%	40.82%	5.08%	40.72%

Let us consider the following Hamiltonian:

$$H(\boldsymbol{\sigma}) = -\sum_{\{i,j\}\in E} J_{i,j}\sigma_i\sigma_j, \tag{50}$$

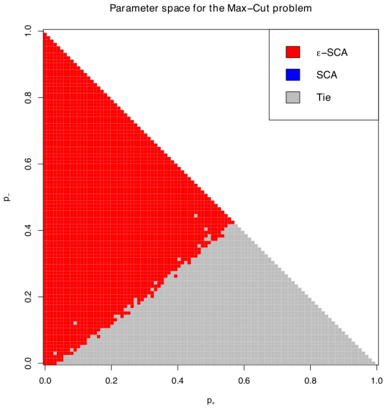
where

$$\mathbb{P}(J_{i,j} = 1) = p_+,$$
$$\mathbb{P}(J_{i,j} = -1) = p_-,$$

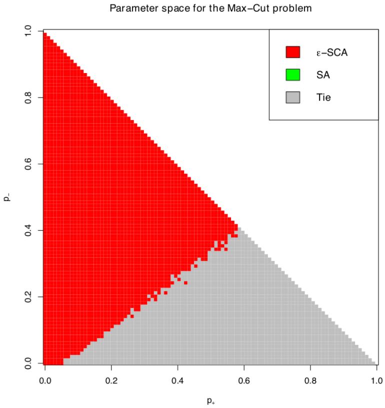
and

$$\mathbb{P}(J_{i,j} = 0) = 1 - (p_+ + p_-).$$

Simulations

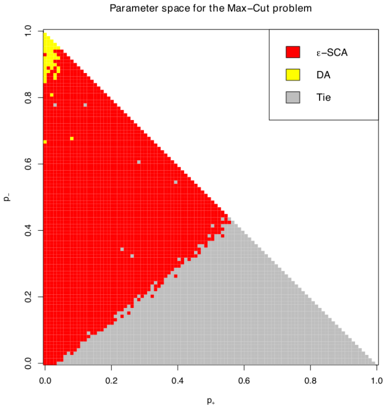


(a) eSCA vs SCA

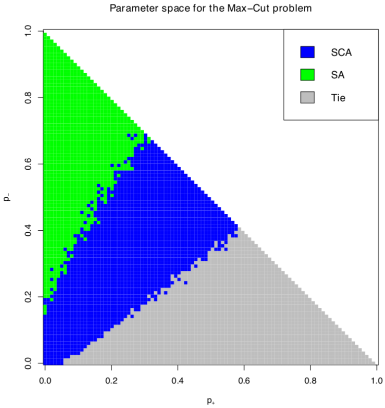


(b) eSCA vs SA

Simulations

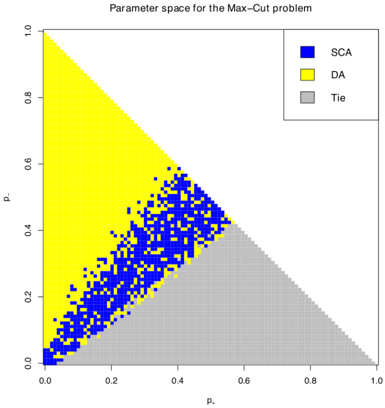


(a) eSCA vs DA

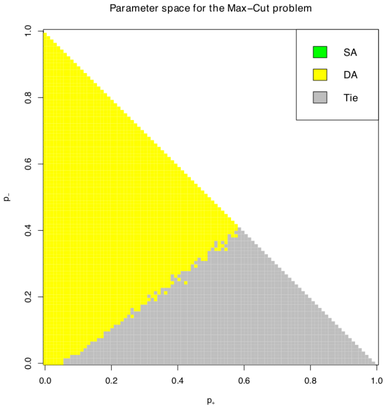


(b) SCA vs SA

Simulations



(a) SCA vs DA



(b) SA vs DA

Next goals

Next goals

- Prove rigorous results for exponential cooling schedules.

Next goals

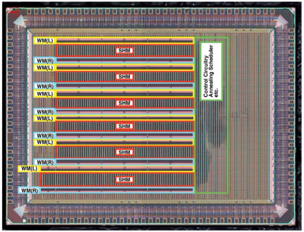
- Prove rigorous results for exponential cooling schedules.
- Derive results that are not asymptotic, that is, consider finite time simulation.

Next goals

- Prove rigorous results for exponential cooling schedules.
- Derive results that are not asymptotic, that is, consider finite time simulation.
- Provide rigorous results for the ε-SCA.

Colaborations

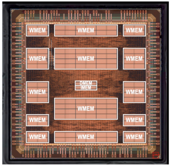
- K. Yamamoto, et al. STATICA: A 512-Spin 0.25M-Weight Annealing Processor With an All-Spin-Updates-at-Once Architecture for Combinatorial Optimization With Complete Spin–Spin Interactions. *IEEE Journal of Solid-State Circuits*, 56 (2021): 165–178.
- K. Kawamura, et al. 2.3 Amorphica: 4-replica 512 fully connected spin 336MHz metamorphic annealer with programmable optimization strategy and compressed-spin-transfer multi-chip extension. *In 2023 IEEE International Solid-State Circuits Conference, 2023.*



Technology	TSMC 65nm CMOS
Chip Size	12mm²
Core V_{DD}	1.1V
I/O V_{DD}	3.3V
Frequency	320MHz
Power	649mW
Gate Count	336K
SRAM	256word,140bit×18 256word,40bit×2

(a) STATICA chip (2021)

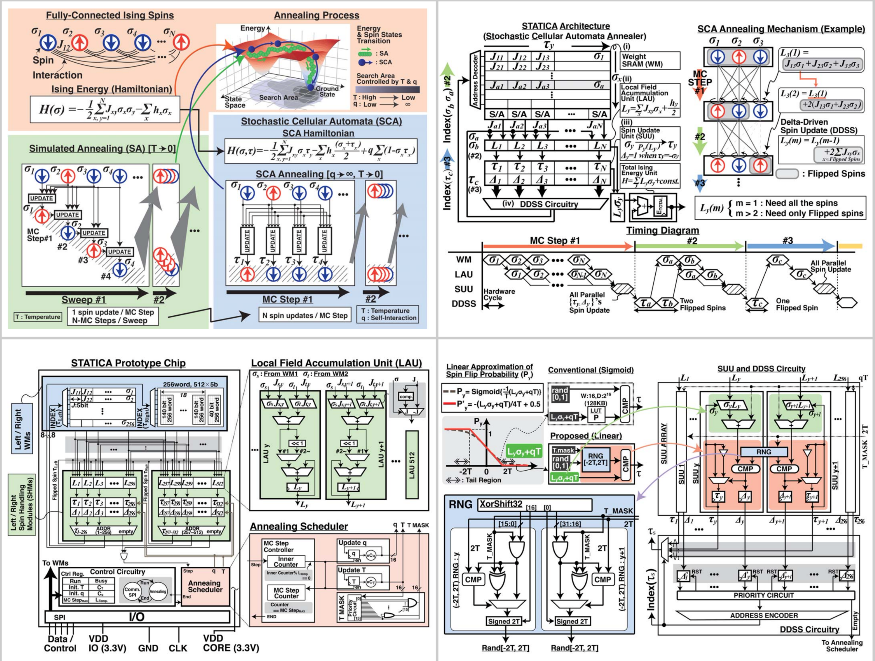
Micrograph



Specification Table

Technology	TSMC 40nm CMOS (LP)
Package	QFN80
Chip Size	3mm x 3mm
Core Area	SRAM: 3.55mm² Logic: 1.48mm²
Core V_{DD}	0.8-1.1V
I/O V_{DD}	3.3V
Max Freqency	336MHz@1.1V 134MHz@0.8V
Gate Count	1.2M Gates
SRAM	WMEM: 8Mb DMEM: 64Kb IMEM: 64Kb Total: 8.125Mb

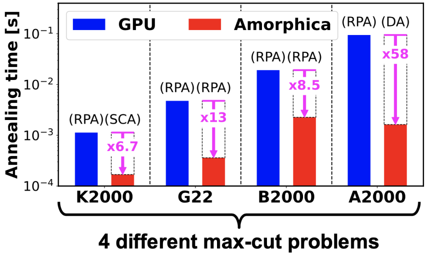
(b) Amorphica chip (2023)


Fully-Connected Ising Spins
Annealing Process
Ising Energy (Hamiltonian)
Stochastic Cellular Automata (SCA)
SCA Hamiltonian
Simulated Annealing (SA) [T→0]
SCA Annealing [q→∞, T→0]
Sweep #1
MC Step #1
1 spin update / MC Step
N-MC Steps / Sweep
N spin updates / MC Step
T : Temperature
q : Self-Interaction
STATICA Architecture
(Stochastic Cellular Automata Annealer)
Weight SRAM (WM)
Local Field Acumulation Unit (LAU)
Spin Update Unit (SUU)
Total Ising Energy Unit
DDSS Circuitry
SCA Annealing Mechanism (Example)
Delta-Driven Spin Update (DDSS)
Flipped Spins
Timing Diagram
STATICA Prototype Chip
Local Field Accumulation Unit (LAU)
Annealing Scheduler
Linear Approximation of Spin Flip Probability (Py)
Conventional (Sigmoid)
Proposed (Linear)
SUU and DDSS Circuitry
RNG
XorShift32
PRIORITY CIRCUIT
ADDRESS ENCODER
To Annealing Scheduler

Performance Comparison

	STATICA	ISSCC2021 4.6	VLSI2021 JFS2-6	ISSCC2022 16.5	Amorphica
Technology	65nm CMOS	40nm CMOS	65nm CMOS	65nm CMOS	40nm CMOS
Inter-Spin Couplings	Full/Complete	Local/Sparse	Local/Sparse	Local/Sparse	Full/Complete
#Spins / Replica	512	16K	560	256 or 1K	2K
#Replicas	1	1	1	1	4
#Couplings / Spin	512	8	8	28 or 7	2K
Weight Width	5bit	5bit	3bit	8bit	8bit
Multi-Chip Extension	No	Up to 9	No	Up to 2	Up to 4
Annealing Algorithm	SCA	SA	SA	SA	Metamorphic Annealing
Operating Power	649mW	N/A	9.9mW@0.8V	1.167mW	151.6–474.9mW @1.1V, 320MHz

Evaluation



Best policy in Amorphica varies depending on the problem

Power Consumption

- ✓ [GPU, Nvidia RTX2080] ≈ 250W
- ✓ [Amorphica] < 500mW

References

- B.H. Fukushima-Kimura, N. Kawamoto, E. Noda, A. Sakai. Mathematical aspects and simulated annealing for the Digital Annealer's Algorithm. *Manuscript in preparation.*
- K. Kawamura, et al. 2.3 Amorphica: 4-replica 512 fully connected spin 336MHz metamorphic annealer with programmable optimization strategy and compressed-spin-transfer multi-chip extension. *In 2023 IEEE International Solid-State Circuits Conference, 2023.*
- B.H. Fukushima-Kimura, S. Handa, K. Kamakura, Y. Kamijima, K. Kawamura, A. Sakai. Mixing time and simulated annealing for the stochastic cellular automata. *arXiv preprint arXiv:2007.11287* (2023).
- B.H. Fukushima-Kimura, Y. Kamijima, K. Kawamura and A. Sakai. Stochastic optimization: stochastic cellular automata versus Glauber dynamics. *Transactions of the Institute of Systems, Control and Information Engineers* 36(1):9–16, 2023.
- B.H. Fukushima-Kimura, Y. Kamijima, K. Kawamura and A. Sakai. Stochastic optimization via parallel dynamics: rigorous results and simulations. *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications*, 2022, 65-71, 2022.
- B.H. Fukushima-Kimura, A. Sakai, H. Toyokawa and Y. Ueda. Stability of energy landscape for Ising models. *Physica A: Statistical Mechanics and its Applications*, 583:126208, 2021.
- K. Yamamoto, et al. STATICA: A 512-Spin 0.25M-Weight Annealing Processor With an All-Spin-Updates-at-Once Architecture for Combinatorial Optimization With Complete Spin–Spin Interactions. *IEEE Journal of Solid-State Circuits*, 56 (2021): 165–178.

Thanks for your attention!