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 $V \setminus S$ such that the total weight of edges connecting the set $S$ and its complementary $V \setminus S$ 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)

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

\[
H(\sigma) = -\sum_{i,j \in E} J_{i,j} \sigma_i \sigma_j - \sum_{i \in V} h_i \sigma_i
\]

for each \( \sigma \in \Omega = \{-1, +1\}^V \). Let us consider the set of ground states of \( H \):

\[
\text{GS} = \{ \sigma : H(\sigma) = \min_\eta H(\eta) \}
\]
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

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

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

(1)
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

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

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

Let us consider the set of ground states of $H$:

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

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

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

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

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

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

(3) for $\sigma \in \{-1, 1\}^V$.

The weighted Max-Cut problem is equivalent to the minimization of $H$. 
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 $\tau = (\tau_t, i)_{N_t, i=1}^N \in \{0, 1\}^{N \times N}$, we have $\tau_t, i = 1$ if the city $i$ is occupied at time $t$, $\tau_t, i = 0$ if the city $i$ is NOT occupied at time $t$.

In that way, $\tau$ represents a legitimate trajectory iff $\sum_i \tau_t, i = 1$ for each $t$, and $\sum_t \tau_t, i = 1$ for each $i$. 

<table>
<thead>
<tr>
<th>City</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 0 0 0</td>
</tr>
<tr>
<td>2</td>
<td>0 0 0 0</td>
</tr>
<tr>
<td>3</td>
<td>0 0 0 0</td>
</tr>
<tr>
<td>4</td>
<td>0 0 0 0</td>
</tr>
<tr>
<td>5</td>
<td>0 0 0 0</td>
</tr>
</tbody>
</table>

City
Time
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 $\tau = (\tau_{t,i})_{t,i=1}^{N} \in \{0, 1\}^{N \times N}$, we have

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

(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
$
\tau = (\tau_{t,i})_{t,i=1}^N \in \{0, 1\}^{N \times N},$
we have

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

In that way, $\tau$ represents a legitimate trajectory iff

$$
\begin{align*}
\sum_i \tau_{t,i} &= 1 \quad \text{for each } t, \quad \text{and} \\
\sum_t \tau_{t,i} &= 1 \quad \text{for each } i.
\end{align*}
$$

In the example, the city 5 is not occupied at any time, while city 1 is occupied at times 1 and 3.
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 $\tau = (\tau_{t,i})_{t,i=1}^{N} \in \{0, 1\}^{N \times N}$, we have

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

(5)

In that way, $\tau$ represents a legitimate trajectory iff \[
\sum_{i} \tau_{t,i} = 1 \text{ for each } t, \text{ and } \sum_{t} \tau_{t,i} = 1 \text{ for each } i.
\]

Let us consider the Hamiltonian

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

(6)

where $\tau \in \{0, 1\}^{N \times N}$ 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

Quadratic Unconstrained Binary Optimization (QUBO) problems

NP-complete and NP-hard problems


Finance

Quadratic Unconstrained Binary Optimization (QUBO) problems

NP-complete and NP-hard problems


Finance


Chemistry

Quadratic Unconstrained Binary Optimization (QUBO) problems

NP-complete and NP-hard problems


Finance


Chemistry


Biology

Logistics and scheduling


Machine learning


VLSI design


Logistics and scheduling

Logistics and scheduling


Machine learning


Logistics and scheduling


Machine learning


VLSI design

Combinatorial Optimization (CO) is Vital to Our Society

- Finance: Portfolio Optimization
- Logistics: Traveling Salesperson
- Drug Discovery: Graph Mining

But an Exhaustive Search is Impractical for Large CO

Goal: Find an optimal route

Route 1  Route 2  Route 3  Opt. Route

$O(N!)$
Annealing Computation for CO using Ising Model

Find Ground (Minimum Energy) States of Ising Model

**Ising Model**
- \( \sigma_i \in \{+1, -1\} \): Spin
- \( J_{ij} \): Coupling Weight
- Energy Function
  \[
  H(\sigma) = -\frac{1}{2} \sum_{i \neq j} J_{ij} \sigma_i \sigma_j
  \]

**Annealing Process**
- Initial State
- Ground State
- Temperature Control

**CO Problem**
- Input: \( J \)

**CO Solution**
- Output: \( \sigma \)
Comparison of Annealing Policies

**SA** (Simulated Annealing)
- Random Select
- Flip?
- Single trial
- Single update

**DA** (Digital Annealing)
- Flip?
- Random Select
- Parallel trial
- Single update

**SCA** (Stochastic Cellular Automata Annealing)
- Flip?
- Do Nothing
- Parallel trial
- Parallel update

**RPA** (Ratio-controlled Parallel Annealing)
- Flip?
- Apply Prob. $\varepsilon$
- Parallel trial
- Managed parallel update
Simulated Annealing (SA)

Let us consider the Metropolis dynamics at inverse temperature $\beta$:

$$P_\beta(\sigma, \tau)/\sum_{i \in V} P_\beta(\sigma, \sigma_i)$$

if $\tau = \sigma_i$ for some $i \in V$,

$1 - \sum_{i \in V} P_\beta(\sigma, \sigma_i)$ if $\tau = \sigma$, and

0 otherwise; \hspace{1cm} (7)

where $\sigma_i$ is the configuration given by

$$\sigma_{i,j} = -\sigma_j \text{ if } j = i \sigma_j \text{ otherwise}, \hspace{1cm} (8)$$

and

$$E_i(\sigma) = H(\sigma_i) - H(\sigma). \hspace{1cm} (9)$$

Theorem (B. Hajek)

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

$$P(X_t = \sigma_t | X_{t-1} = \sigma_{t-1}, \ldots, X_0 = \sigma_0) = P_\beta_t(\sigma_{t-1}, \sigma_t)$$

for every $t \geq 1$ and $\sigma_0, \ldots, \sigma_t$ in $\Omega$. There is $\gamma_c > 0$ such that if we choose $\beta_n = 1/\gamma \log n$, then

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

holds if and only if $\gamma \geq \gamma_c$.\hspace{1cm}
Simulated Annealing (SA)

Let us consider the Metropolis dynamics at inverse temperature $\beta$:

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

where $\sigma^i$ is the configuration given by

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

and

$$E_i(\sigma) = H(\sigma^i) - H(\sigma).$$
Simulated Annealing (SA)

Let us consider the Metropolis dynamics at inverse temperature $\beta$:

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

(7)

where $\sigma^i$ is the configuration given by

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

(8)

and

$$E_i(\sigma) = H(\sigma^i) - H(\sigma).$$

(9)

Theorem (B. Hajek)

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

$$P(X_t = \sigma_t | X_{t-1} = \sigma_{t-1}, \ldots, X_0 = \sigma_0) = P(X_t = \sigma_t | X_{t-1} = \sigma_{t-1}) = P_\beta(\sigma_{t-1}, \sigma_t)$$

(10)

for every $t \geq 1$ and $\sigma_0, \ldots, \sigma_t$ in $\Omega$. There is $\gamma_c > 0$ such that if we choose $\beta_n = \frac{1}{\gamma} \log n$, then

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

(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

(b) The DA’s Algorithm

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
Comparison of Annealing Policies

**SA**
(Simulated Annealing)
- Random Select
- Flip?
- Single trial
- Single update

**DA**
(Digital Annealing)
- Flip?
- Random Select
- Parallel trial
- Single update

**SCA**
(Stochastic Cellular Automata Annealing)
- Flip?
- Do Nothing
- Parallel trial
- Parallel update

**RPA**
(Ratio-controlled Parallel Annealing)
- Flip?
- Apply Prob. $\varepsilon$
- Parallel trial
- Managed parallel update
In our framework, the Digital Annealer’s Algorithm transition matrix $P^{DA}_\beta$ at inverse temperature $\beta$ is defined by

$$P^{DA}_\beta(\sigma, \tau) = \begin{cases} \frac{1}{\left| S \right|} \prod_{j \in S} e^{-\beta E_j(\sigma)^+} \prod_{j \in V \setminus S} \left(1 - e^{-\beta E_j(\sigma)^+}\right) & \text{if } \tau = \sigma^i, \\ \prod_{j \in V} (1 - e^{-\beta E_j(\sigma)^+}) & \text{if } \tau = \sigma, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$
In our framework, the Digital Annealer's Algorithm transition matrix $P_{\beta}^{DA}$ at inverse temperature $\beta$ is defined by

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

The algorithm works as follows.
In our framework, the Digital Annealer's Algorithm transition matrix $P^{\text{DA}}_\beta$ at inverse temperature $\beta$ is defined by

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

The algorithm works as follows.

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

$$16 / 44$$
In our framework, the Digital Annealer’s Algorithm transition matrix $P^\text{DA}_\beta$ at inverse temperature $\beta$ is defined by

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

The algorithm works as follows.

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

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 $X_{t+1} = \sigma^i$; otherwise, nothing is done, and we consider $X_{t+1} = \sigma$. 
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

\[
P(X_t = \sigma_t | X_{t-1} = \sigma_{t-1}, \ldots, X_0 = \sigma_0) = P(X_t = \sigma_t | X_{t-1} = \sigma_{t-1}) = P^{DA}_{\beta t}(\sigma_{t-1}, \sigma_t) \tag{12}
\]

for every \(t \geq 1\) and \(\sigma_0, \ldots, \sigma_t\) in \(\Omega\).

There exists \(\gamma_c > 0\) such that the limit

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

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

\[
P(X_t = \sigma_t | X_{t-1} = \sigma_{t-1}, \ldots, X_0 = \sigma_0) = P(X_t = \sigma_t | X_{t-1} = \sigma_{t-1}) = P_{DA}^{\beta_t}(\sigma_{t-1}, \sigma_t)
\]

for every \(t \geq 1\) and \(\sigma_0, \ldots, \sigma_t\) in \(\Omega\).

There exists \(\gamma_c > 0\) such that the limit

\[
\lim_{t \to \infty} P(X_t \in GS) = 1
\]

holds if and only if

\[
\sum_{t=1}^{\infty} e^{-\beta_t \gamma_c} = +\infty.
\]

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

\[
\beta_t = \frac{1}{\gamma} \log t
\]

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

We say $\tau$ is reachable from $\sigma$ at height $E$ if there exists a path $\sigma = \sigma_0, \sigma_1, \ldots, \sigma_n = \tau$ such that $\max_{0 \leq k \leq n} H(\sigma_k) \leq E$.

Definition

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

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

The constant $\gamma_c$ coincides with the depth of the second deepest local minimum, i.e.,

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

Let us consider the following Hamiltonian:

\[ H(\sigma) = - \sum_{\{i,j\} \in E} J_{i,j} \sigma_i \sigma_j. \] (17)
Simulations

Let us consider the following Hamiltonian:

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

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

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

1. \textit{Max-cut problem}. The Hamiltonian is defined in an Erdős-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. \textit{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

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

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Success rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max-cut</td>
<td>7.52%</td>
</tr>
<tr>
<td>Spin-glass</td>
<td>5.08%</td>
</tr>
<tr>
<td></td>
<td>58.01%</td>
</tr>
<tr>
<td></td>
<td>40.72%</td>
</tr>
</tbody>
</table>
Comparison of Annealing Policies

**SA** (Simulated Annealing)
- Random Select
- Flip?
- Single trial
- Single update

**DA** (Digital Annealing)
- Flip?
- Random Select
- Parallel trial
- Single update

**SCA** (Stochastic Cellular Automata Annealing)
- Flip?
- Do Nothing
- Parallel trial
- Parallel update

**RPA** (Ratio-controlled Parallel Annealing)
- Flip?
- Apply Prob. $\varepsilon$
- Parallel trial
- Managed parallel update
The extended Hamiltonian $\tilde{H}$ is defined by

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

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

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

$$P_{SCA}^{\beta, q}(\sigma, \tau) = e^{-\beta \tilde{H}(\sigma, \tau)} \frac{1}{\sum_{\tau'} e^{-\beta \tilde{H}(\sigma, \tau')}}.$$ 

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

$$\pi_{SCA}^{\beta, q}(\sigma) = \frac{1}{\sum_{\tau} e^{-\beta \tilde{H}(\sigma, \tau)}} \frac{1}{\sum_{\sigma', \tau'} e^{-\beta \tilde{H}(\sigma', \tau')}}$$

is the stationary distribution for $P_{SCA}^{\beta, q}$. 
The extended Hamiltonian $\tilde{H}$ is defined by

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

(19)

for each pair $\sigma, \tau$ of configurations in $\{-1, +1\}^V$. 
The extended Hamiltonian $\hat{H}$ is defined by

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

(19)

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

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

$$P_{\beta,q}^{\text{SCA}}(\sigma, \tau) = \frac{e^{-\beta \hat{H}(\sigma, \tau)}}{\sum_{\tau'} e^{-\beta \hat{H}(\sigma, \tau')}}.$$  

(20)
The extended Hamiltonian $\tilde{H}$ is defined by

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

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

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

$$P_{\beta, q}^{SCA}(\sigma, \tau) = \frac{e^{-\beta \tilde{H}(\sigma, \tau)}}{\sum_{\tau'} e^{-\beta \tilde{H}(\sigma, \tau')}}.$$  \hspace{1cm} (20)$$

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

$$\pi_{\beta, q}^{SCA}(\sigma) = \frac{\sum_{\tau} e^{-\beta \tilde{H}(\sigma, \tau)}}{\sum_{\sigma', \tau'} e^{-\beta \tilde{H}(\sigma', \tau')}}.$$  \hspace{1cm} (21)$$

is the stationary distribution for $P_{\beta, q}^{SCA}$. 
Moreover, we can rewrite $P^{SCA}_{\beta,q}$ as

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

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

$$\tilde{h}_i(\sigma) = \sum_{j \in V} J_{i,j} \sigma_j + h_i. \quad (23)$$
Moreover, we can rewrite $P_{\beta,q}^{SCA}$ as

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

(22)

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

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

(23)

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

For any non-negative $q$, if $\beta$ is sufficiently small such that

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

(24)

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

$$t_{mix}(\varepsilon) \leq \left[ \frac{\log |V| - \log \varepsilon}{\log(1/r)} \right].$$

(25)
Moreover, we can rewrite $P_{\beta,q}^{SCA}$ as

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

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

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

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

*For any non-negative $q$, if $\beta$ is sufficiently small such that*

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

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

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

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 \( \rho_{TM}(P_{\beta,q}^{SCA}(\sigma,\cdot), P_{\beta,q}^{SCA}(\tau,\cdot)) \leq r \) for all \( \sigma, \tau \in \Omega \) with \( |D_{\sigma,\tau}| = 1 \). If \( |D_{\sigma,\tau}| \geq 2 \), then, by the triangle inequality along any sequence \( (\eta_0, \eta_1, \ldots, \eta_{|D_{\sigma,\tau}|}) \) of spin configurations that satisfy \( \eta_0 = \sigma, \eta_{|D_{\sigma,\tau}|} = \tau \) and \( |D_{\eta_{j-1},\eta_j}| = 1 \) for all \( j = 1, \ldots, |D_{\sigma,\tau}| \), we have

\[
\rho_{TM}(P_{\beta,q}^{SCA}(\sigma,\cdot), P_{\beta,q}^{SCA}(\tau,\cdot)) \leq \sum_{j=1}^{|D_{\sigma,\tau}|} \rho_{TM}(P_{\beta,q}^{SCA}(\eta_{j-1},\cdot), P_{\beta,q}^{SCA}(\eta_j,\cdot)) \leq r |D_{\sigma,\tau}|. \tag{26}
\]

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

\[
p(\sigma,y) = \frac{e^{\beta \left( \tilde{h}_y(\sigma) + q_y \sigma_y \right)}}{2 \cosh(\beta \left( \tilde{h}_y(\sigma) + q_y \sigma_y \right))} = \frac{1 + \tanh(\beta \left( \tilde{h}_y(\sigma) + q_y \sigma_y \right))}{2}. \tag{27}
\]

Notice that \( p(\sigma,y) \neq p(\sigma^x,y) \) only when \( y = x \) or \( y \in N_x \equiv \{ v \in V : J_{x,v} \neq 0 \} \). 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 \( P_{\beta,q}^{SCA}(\sigma,\cdot) \) and \( P_{\beta,q}^{SCA}(\sigma^x,\cdot) \) as

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

Denote the measure of this coupling by $P_{\sigma, \sigma^x}$ and its expectation by $E_{\sigma, \sigma^x}$. Then we obtain

$$
E_{\sigma, \sigma^x}[|D_X, Y|] = E_{\sigma, \sigma^x} \left[ \sum_{y \in V} \mathbb{I}\{X_y \neq Y_y\} \right] = \sum_{y \in V} P_{\sigma, \sigma^x}(X_y \neq Y_y) = \sum_{y \in V} |p(\sigma, y) - p(\sigma^x, y)|
$$

$$
= |p(\sigma, x) - p(\sigma^x, x)| + \sum_{y \in N_x} |p(\sigma, y) - p(\sigma^x, y)|,
$$

(29)

where, by using the rightmost expression above satisfies

$$
|p(\sigma, x) - p(\sigma^x, x)| \leq \frac{1}{2} \left| \tanh \left( \frac{\beta h_x(\sigma)}{2} + \frac{\beta q_x}{2} \right) - \tanh \left( \frac{\beta h_x(\sigma) - \beta q_x}{2} \right) \right|,
$$

(30)

and for $y \in N_x$,

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

(31)

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

$$
\rho_{TM}(P_{\beta, q}^{SCA}(\sigma, \cdot), P_{\beta, q}^{SCA}(\sigma^x, \cdot)) \leq E_{\sigma, \sigma^x}[|D_X, Y|] \leq \tanh \frac{\beta q_x}{2} + \sum_{y \in N_x} \tanh \frac{\beta |J_{x,y}|}{2} \leq r.
$$

(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 \( \lambda \) 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 \( \lim_{t \to \infty} \beta_t = \infty \), we have

\[
\sum_{t=1}^{\infty} \| \pi^{SCA}_{\beta_{t+1},q} - \pi^{SCA}_{\beta_t,q} \|_{TV} < \infty, \quad \lim_{t \to \infty} \| \pi^{SCA}_{\beta_t,q} - \pi^G_{\infty} \|_{TV} = 0. \tag{33}
\]

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

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

then we obtain

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

As a result, for any initial \( j \geq 1 \),

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

Step 1.
Let us show
$$\lim_{t \to \infty} \parallel. \parallel_\pi \mathcal{SCA} \beta t, q \parallel_\pi \mathcal{SCA} \beta t, q \parallel_\parallel_\parallel < 0 \quad (37)$$

We first define
$$\mu_\beta (\sigma, \tau) = e^{-\beta \tilde{H}(\sigma, \tau)} \sum_\xi, \eta e^{-\beta \tilde{H}(\xi, \eta)} \equiv e^{-\beta (\tilde{H}(\sigma, \tau) - \delta) \sum_\xi, \eta e^{-\beta (\tilde{H}(\xi, \eta) - \delta), (38)}$$

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

We conclude that
$$\mu_\beta (\sigma, \tau) = e^{-\beta (\tilde{H}(\sigma, \tau) - \delta)} \sum_\xi, \eta \parallel. \parallel \quad (39)$$

Summing this over \( \tau \) yields the second relation in (33).

Step 2.
Let us show
$$\sum_{t=1}^\infty \parallel. \parallel_\pi \mathcal{SCA} \beta t, q \parallel_\parallel_\parallel_\parallel < \infty \quad (40)$$

To show the first relation in (33), we note that
$$\frac{\partial \mu_\beta (\sigma, \tau)}{\partial \beta} = \left. \frac{\partial H}{\partial \beta} \right|_{\mu_\beta (\sigma, \tau)}$$

and that
$$\left. \frac{\partial H}{\partial \beta} \right|_{\mu_\beta (\sigma, \tau)} \equiv \sum_\sigma, \tau \tilde{H}(\sigma, \tau) \mu_\beta (\sigma, \tau)$$
tends to \( m \) as \( \beta \to \infty \), due to (39).
Proof

Step 1. Let us show

\[ \lim_{t \uparrow \infty} \| \pi_{SCA}^{\beta_t, q} - \pi_{\infty}^{G} \|_{TV} = 0 \quad (37) \]

We first define

\[ \mu_{\beta}(\sigma, \tau) = \frac{e^{-\beta \tilde{H}(\sigma, \tau)}}{\sum_{\xi, \eta} e^{-\beta \tilde{H}(\xi, \eta)}} \equiv \frac{e^{-\beta(\tilde{H}(\sigma, \tau)-m)}}{\sum_{\xi, \eta} e^{-\beta(\tilde{H}(\xi, \eta)-m)}}, \quad (38) \]

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

\[ \mu_{\beta}(\sigma, \tau) = \frac{e^{-\beta(\tilde{H}(\sigma, \tau)-m)}}{|\text{GS}| + \sum_{\xi, \eta: \tilde{H}(\xi, \eta)>m} e^{-\beta(\tilde{H}(\xi, \eta)-m)} \rightarrow \frac{1}{|\text{GS}|} \sum_{\sigma \in \text{GS}} \delta_{\sigma, \tau}. \quad (39) \]

Summing this over \( \tau \) yields the second relation in (33).
Proof

Step 1. Let us show
\[ \lim_{t \uparrow \infty} \| \pi^{SCA}_{t+1,q} - \pi^G_t \|_{TV} = 0 \]  
(37)

We first define
\[ \mu(\sigma, \tau) = \frac{e^{-\beta \tilde{H}(\sigma, \tau)}}{\sum_{\xi, \eta} e^{-\beta \tilde{H}(\xi, \eta)}} \equiv \frac{e^{-\beta (\tilde{H}(\sigma, \tau)-m)}}{\sum_{\xi, \eta} e^{-\beta (\tilde{H}(\xi, \eta)-m)}}, \]  
(38)

where \( m = \min_{\sigma, \eta} \tilde{H}(\sigma, \eta) \). We conclude that
\[ \mu(\sigma, \tau) = \frac{e^{-\beta (\tilde{H}(\sigma, \tau)-m)}}{|GS| + \sum_{\xi, \eta} e^{-\beta (\tilde{H}(\xi, \eta)-m)}} \rightarrow \frac{1}{|GS|} \sum_{\sigma \in GS} \delta_{\sigma, \tau}. \]  
(39)

Summing this over \( \tau \) yields the second relation in (33).

Step 2. Let us show
\[ \sum_{t=1}^{\infty} \| \pi^{SCA}_{t+1,q} - \pi^{SCA}_{t,q} \|_{TV} < \infty \]  
(40)

To show the first relation in (33), we note that
\[ \frac{\partial \mu(\sigma, \tau)}{\partial \beta} = \left( \mathbb{E}_{\mu} [\tilde{H}] - \tilde{H}(\sigma, \tau) \right) \mu(\sigma, \tau), \]  
(41)

and that \( \mathbb{E}_{\mu} [\tilde{H}] \equiv \sum_{\sigma, \tau} \tilde{H}(\sigma, \tau) \mu(\sigma, \tau) \) tends to \( m \) as \( \beta \uparrow \infty \), due to (39).
Therefore, $\frac{\partial}{\partial \beta} \mu_\beta(\sigma, \tau) > 0$ for all $\beta$ if $\tilde{H}(\sigma, \tau) = m$, while it is negative for sufficiently large $\beta$ if $\tilde{H}(\sigma, \tau) > m$. Let $n \in \mathbb{N}$ be such that, as long as $\beta \geq \beta_n$, (41) is negative for all pairs $(\sigma, \tau)$ satisfying $\tilde{H}(\sigma, \tau) > m$. As a result,

$$\sum_{t=n}^{N} \| \pi_{\beta_{t+1}, q}^{SCA} - \pi_{\beta_t, q}^{SCA} \|_{TV}$$

$$= \frac{1}{2} \sum_{\sigma \in GS} \sum_{t=n}^{N} |\pi_{\beta_{t+1}, q}^{SCA}(\sigma) - \pi_{\beta_t, q}^{SCA}(\sigma)| + \frac{1}{2} \sum_{\sigma \notin GS} \sum_{t=n}^{N} |\pi_{\beta_{t+1}, q}^{SCA}(\sigma) - \pi_{\beta_t, q}^{SCA}(\sigma)|$$

$$\leq \frac{1}{2} \sum_{\sigma \in GS} \sum_{t=n}^{N} (\mu_{\beta_{t+1}}(\sigma, \sigma) - \mu_{\beta_t}(\sigma, \sigma)) + \frac{1}{2} \sum_{\sigma \in GS} \sum_{\tau \neq \sigma} \sum_{t=n}^{N} (\mu_{\beta_t}(\sigma, \tau) - \mu_{\beta_{t+1}}(\sigma, \tau))$$

$$+ \frac{1}{2} \sum_{\sigma \notin GS} \sum_{t=n}^{N} (\pi_{\beta_{t+1}, q}^{SCA}(\sigma) - \pi_{\beta_t, q}^{SCA}(\sigma))$$

$$= \frac{1}{2} \sum_{\sigma \in GS} (\mu_{\beta_{N+1}}(\sigma, \sigma) - \mu_{\beta_n}(\sigma, \sigma)) + \frac{1}{2} \sum_{\sigma \in GS} \sum_{\tau \neq \sigma} (\mu_{\beta_n}(\sigma, \tau) - \mu_{\beta_{N+1}}(\sigma, \tau))$$

$$+ \frac{1}{2} \sum_{\sigma \notin GS} (\pi_{\beta_{N+1}, q}^{SCA}(\sigma) - \pi_{\beta_n, q}^{SCA}(\sigma))$$

$$\leq \frac{3}{2}$$

holds uniformly for $N \geq n$. 

(42)
Step 3. Let us show

\[ \sum_{t=1}^{\infty} \left( 1 - \delta(P_{\beta t, q}^{SCA}) \right) = \infty. \]  

(43)

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

\[
P_{\beta, q}^{SCA}(\sigma, \tau) = \prod_{x \in V} \frac{e^{\frac{\beta}{2}(\tilde{h}_x(\sigma) + q_x \sigma_x) \tau_x}}{2 \cosh(\frac{\beta}{2}(\tilde{h}_x(\sigma) + q_x \sigma_x))} \geq \prod_{x \in V} \frac{1}{1 + e^{\beta|\tilde{h}_x(\sigma) + q_x \sigma_x|}} \geq \prod_{x \in V} \frac{e^{-\beta \Gamma_x}}{2} = \frac{e^{-\beta \Gamma}}{2|V|}.
\]

(44)

Then, we obtain

\[ \sum_{t=1}^{\infty} \left( 1 - \delta(P_{\beta t, q}^{SCA}) \right) = \sum_{t=1}^{\infty} \min_{\sigma, \eta} \sum_{\tau} P_{\beta t, q}^{SCA}(\sigma, \tau) \land P_{\beta t, q}^{SCA}(\eta, \tau) \geq \sum_{t=1}^{\infty} e^{-\beta t \Gamma}, \]

(45)

which diverges, as required, under the cooling schedule (34). This completes the proof of the theorem.
Ratio-controlled Parallel Annealing (RPA or $\varepsilon$-SCA)

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

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

(46)

where we recall that $p_i(\sigma) = e^{-\beta 2 \tilde{h}_i(\sigma) \sigma_i 2 \cosh(\beta 2 \tilde{h}_i(\sigma))}$

(47)

is the probability of flipping the spin $\sigma_i$ from the configuration disregarding a pinning parameter at $i$. Theorem (Fukushima-Kimura, Kamijima, Kawamura, Sakai)

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

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

(48)

then $t_{\text{mix}}$ satisfies

$$t_{\text{mix}}(\delta) \leq \frac{\log V - \log \delta \log(1/\delta)}{r_{\text{mix}}},$$

(49)
**Ratio-controlled Parallel Annealing (RPA or $\varepsilon$-SCA)**

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

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

where we recall that

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

is the probability of flipping the spin $\sigma_i$ from the configuration $\sigma$ disregarding a pinning parameter at $i$. 

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

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

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

then $t_{\text{mix}}$ satisfies

$$t_{\text{mix}}(\delta) \leq \frac{\log V - \log \delta}{\log(1/\delta^r)}.$$
Ratio-controlled Parallel Annealing (RPA or $\varepsilon$-SCA)

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

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

where we recall that

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

is the probability of flipping the spin $\sigma_i$ from the configuration $\sigma$ disregarding a pinning parameter at $i$.

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

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

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

then $t_{\text{mix}}$ satisfies

$$t_{\text{mix}}(\delta) \leq \left[ \frac{\log |V| - \log \delta}{\log(1/r)} \right].$$
Comparison of Annealing Policies

SA (Simulated Annealing)
- Random Select
- Flip?
- Single trial
- Single update

DA (Digital Annealing)
- Flip?
- Random Select
- Parallel trial
- Single update

SCA (Stochastic Cellular Automata Annealing)
- Flip?
- Do Nothing
- Parallel trial
- Parallel update

RPA (Ratio-controlled Parallel Annealing)
- Flip?
- Apply Prob. $\varepsilon$
- Parallel trial
- Managed parallel update
Simulations

Histograms of minimal energy

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

(b) Spin-glass

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

The effect of $\varepsilon$ on the success rate

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

(b) Spin-glass

Figure: Success rate dependence on $\varepsilon$. 


Let us consider the following Hamiltonian:

$$H(\sigma) = -\sum_{\{i,j\} \in E} J_{ij} \sigma_i \sigma_j,$$  \hspace{1cm} (50)$$

where $P(J_{ij} = 1) = p_+$, $P(J_{ij} = -1) = p_-$, and $P(J_{ij} = 0) = 1 - (p_+ + p_-)$.
Table: Summary of the simulations

<table>
<thead>
<tr>
<th>Model</th>
<th>Success rate</th>
<th>$\varepsilon$-SCA</th>
<th>SCA</th>
<th>SA</th>
<th>DA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max-cut</td>
<td>85.9%</td>
<td>0%</td>
<td>7.52%</td>
<td>58.01%</td>
<td></td>
</tr>
<tr>
<td>Spin-glass</td>
<td>59.28%</td>
<td>40.82%</td>
<td>5.08%</td>
<td>40.72%</td>
<td></td>
</tr>
</tbody>
</table>

Let us consider the following Hamiltonian:

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

where

\[
P(J_{i,j} = 1) = p_+,
\]

\[
P(J_{i,j} = -1) = p_-,
\]

and

\[
P(J_{i,j} = 0) = 1 - (p_+ + p_-).
\]
Simulations

Parameter space for the Max-Cut problem

(a) eSCA vs SCA

(b) eSCA vs SA
Simulations

Parameter space for the Max–Cut problem

(a) $\varepsilon$-SCA vs DA
(b) SCA vs SA
Simulations

Parameter space for the Max-Cut problem

(a) SCA vs DA

(b) SA vs DA
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 $\epsilon$-SCA.
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 $\varepsilon$-SCA.
Colaborations


![STATICA chip (2021)](a)

![Amorphica chip (2023)](b)

**Specification Table**

<table>
<thead>
<tr>
<th></th>
<th>Technology</th>
<th>TSMC 40nm CMOS (LP)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Package</strong></td>
<td></td>
<td>QFN80</td>
</tr>
<tr>
<td><strong>Chip Size</strong></td>
<td></td>
<td>3mm x 3mm</td>
</tr>
<tr>
<td><strong>Core Area</strong></td>
<td>SRAM: 3.55mm² Logic: 1.48mm²</td>
<td></td>
</tr>
<tr>
<td><strong>Core V_{DD}</strong></td>
<td>0.8-1.1V</td>
<td></td>
</tr>
<tr>
<td><strong>I/O V_{DD}</strong></td>
<td>3.3V</td>
<td></td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td><a href="mailto:336MHz@1.1V">336MHz@1.1V</a> <a href="mailto:134MHz@0.8V">134MHz@0.8V</a></td>
<td></td>
</tr>
<tr>
<td><strong>Gate Count</strong></td>
<td>1.2M Gates</td>
<td></td>
</tr>
<tr>
<td><strong>SRAM</strong></td>
<td>WMEM: 8Mb IMEM: 64Kb DMEM: 64Kb Total: 8.125Mb</td>
<td></td>
</tr>
</tbody>
</table>
## Performance Comparison

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

## Evaluation

Best policy in Amorphica varies depending on the problem

**Power Consumption**

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


Thanks for your attention!