Mathematical foundation of various MCMC methods
@ French Japanese Conference on Probability & Interactions

Akira Sakai

Hokkaido University

March 6, 2024

- Want to find a ground state of the energy landscape H (= Hamiltonian) for an Ising model on a finite graph $G = (V, E)$:

$$H(\boldsymbol{\sigma}) = -\sum_{\{x,y\}\in E} J_{x,y}\sigma_x\sigma_y - \sum_{x\in V} h_x\sigma_x, \qquad \text{GS} = \arg\min_{\boldsymbol{\sigma}\in\{\pm1\}^V} H(\boldsymbol{\sigma}),$$

where $[J_{x,y}]_{V\times V}$ is symmetric (with 0 on the diagonals), $\{h_x\}_{x\in V} \in \mathbb{R}^V$.

- Why does GS matter? $\Rightarrow$ Many combinatorial optimization problems can be mapped to Ising models:

1. Max-cut: Divide the vertex set V of a weighted graph into S and $V\setminus S$, while maximizing the sum of the weights of the cut edges.
   - $w_{x,y} \geq 0$: the weight on the edge $\{x,y\}\in E$ $(w_{x,x}=0,\ w_{x,y}=w_{y,x})$.
   - $C = \sum_{\{x,y\}\in E} w_{x,y}\big(\mathbb{1}_{\{x\in S\}}\mathbb{1}_{\{y\in V\setminus S\}} + \mathbb{1}_{\{x\in V\setminus S\}}\mathbb{1}_{\{y\in S\}}\big)$: the total weight to be maximized.
   - $\sigma_x = \mathbb{1}_{\{x\in S\}} - \mathbb{1}_{\{x\in V\setminus S\}} = \begin{cases} 1 & [x\in S] \\ -1 & [x\in V\setminus S] \end{cases} \quad \Rightarrow \quad C = \frac{1}{2}\sum_{\{x,y\}\in E} w_{x,y}(1-\sigma_x\sigma_y).$
   - $H(\boldsymbol{\sigma}) = \sum_{\{x,y\}\in E} w_{x,y}\sigma_x\sigma_y$ is to be minimized.
2. Traveling salesman: Find a route that minimizes the total cost among those which allow a salesman to return to the first city by going through all cities once.
3. Knapsack: Determine which item should be selected to maximize the total value of the items packed in an knapsack of fixed volume.

- Want to find a ground state of the energy landscape H (= Hamiltonian) for an Ising model on a finite graph $G = (V, E)$:

$$H(\boldsymbol{\sigma}) = -\sum_{\{x,y\}\in E} J_{x,y}\sigma_x\sigma_y - \sum_{x\in V} h_x\sigma_x, \qquad \text{GS} = \arg\min_{\boldsymbol{\sigma}\in\{\pm1\}^V} H(\boldsymbol{\sigma}),$$

where $[J_{x,y}]_{V\times V}$ is symmetric (with 0 on the diagonals), $\{h_x\}_{x\in V} \in \mathbb{R}^V$.

- Why does GS matter? $\Rightarrow$ Many combinatorial optimization problems can be mapped to Ising models:

1. Max-cut: Divide the vertex set V of a weighted graph into S and $V \setminus S$, while maximizing the sum of the weights of the cut edges.
   - $w_{x,y} \geq 0$: the weight on the edge $\{x, y\} \in E$ $(w_{x,x} = 0,\ w_{x,y} = w_{y,x})$.
   - $C = \sum_{\{x,y\}\in E} w_{x,y}\big(\mathbb{1}_{\{x\in S\}}\mathbb{1}_{\{y\in V\setminus S\}} + \mathbb{1}_{\{x\in V\setminus S\}}\mathbb{1}_{\{y\in S\}}\big)$: the total weight to be maximized.
   - $\sigma_x = \mathbb{1}_{\{x\in S\}} - \mathbb{1}_{\{x\in V\setminus S\}} = \begin{cases} 1 & [x \in S] \\ -1 & [x \in V \setminus S] \end{cases} \Rightarrow C = \frac{1}{2}\sum_{\{x,y\}\in E} w_{x,y}(1 - \sigma_x\sigma_y).$
   - $H(\boldsymbol{\sigma}) = \sum_{\{x,y\}\in E} w_{x,y}\sigma_x\sigma_y$ is to be minimized.
2. Traveling salesman: Find a route that minimizes the total cost among those which allow a salesman to return to the first city by going through all cities once.
3. Knapsack: Determine which item should be selected to maximize the total value of the items packed in an knapsack of fixed volume.

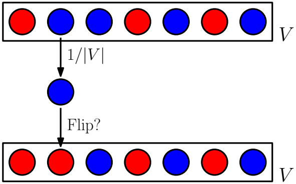
- Difficult to find GS, because
  - the configuration space $\{\pm 1\}^V$ may be humongous (depending on $|V|$),
  - the energy landscape H may be complicated (depending on $[J_{x,y}]$ and $\{h_x\}$).
- Use Markov Chain Monte Carlo (MCMC) methods to sample the Gibbs distribution $\pi_\beta(\boldsymbol{\sigma}) \propto e^{-\beta H(\boldsymbol{\sigma})}$ at inverse temperature $\beta \geq 0$, with
  - $\underset{\beta\downarrow 0}{\Longrightarrow}$ uniform on the entire $\{\pm 1\}^V$,
  - $\underset{\beta\uparrow\infty}{\Longrightarrow}$ uniform on GS.
- Compare the conventional MCMC methods (Glauber, Metropolis) with
  - Digital Annealer (DA) by Fujitsu Laboratories,
  - Stochastic Cellular Automata (SCA),
  - its variant (called ε-SCA).

- Conventional single-spin flip MCs:

$$(\sigma^x)_y = \begin{cases} -\sigma_y & [y = x], \\ \sigma_y & [y \neq x], \end{cases} \qquad \nabla_x^+ H(\sigma) = \Big(H(\sigma^x) - H(\sigma)\Big)^+.$$

1. Glauber dynamics: $$P_\beta^{\mathrm{g}}(\sigma, \tau) = \begin{cases} \frac{1}{|V|} \frac{e^{-\beta H(\sigma^x)}}{e^{-\beta H(\sigma)} + e^{-\beta H(\sigma^x)}} & [\tau = \sigma^x], \\ 1 - \sum_{x \in V} P_\beta^{\mathrm{g}}(\sigma, \sigma^x) & [\tau = \sigma]. \end{cases}$$

2. Metropolis algorithm: $$P_\beta^{\mathrm{m}}(\sigma, \tau) = \begin{cases} \frac{1}{|V|} e^{-\beta \nabla_x^+ H(\sigma)} & [\tau = \sigma^x], \\ 1 - \sum_{x \in V} P_\beta^{\mathrm{m}}(\sigma, \sigma^x) & [\tau = \sigma]. \end{cases}$$



- Both are aperiodic and irreducible and satisfy the detailed balance condition:

$$^\forall\boldsymbol{\sigma},\boldsymbol{\tau}\in\{\pm1\}^V,\qquad \pi_\beta(\boldsymbol{\sigma})\,P^\bullet_\beta(\boldsymbol{\sigma},\boldsymbol{\tau})=\pi_\beta(\boldsymbol{\tau})\,P^\bullet_\beta(\boldsymbol{\tau},\boldsymbol{\sigma}).$$

E.g., for Metropolis, $\boldsymbol{\tau}=\boldsymbol{\sigma}^x \quad\Rightarrow\quad e^{-\beta H(\boldsymbol{\sigma})}e^{-\beta\nabla_x^+H(\boldsymbol{\sigma})}=e^{-\beta\left(H(\boldsymbol{\sigma})\vee H(\boldsymbol{\sigma}^x)\right)}$. ■

$$\therefore\ \pi_\beta=\pi_\beta*P^\bullet_\beta,\qquad\qquad {}^\forall\mu*(P^\bullet_\beta)^{*n}\underset{n\uparrow\infty}{\Longrightarrow}\pi_\beta.$$

Theorem 1 (e.g., Catoni (1999))

Let $\{X_n\}_{n=0}^\infty$ be the $\{\pm1\}^V$-valued MC generated by $P^\bullet_\beta=P^g_\beta$ or P^m_β. Then,

$$^\forall\varepsilon>0,\ {}^\exists N\in\mathbb{N},\ {}^\exists\beta\ge0\ \Big[n\ge N\ \Rightarrow\ \min_{\boldsymbol{\sigma}}P^\bullet_\beta\big(X_n\in\text{GS}\,\big|\,X_0=\boldsymbol{\sigma}\big)\ge1-\varepsilon\Big].$$

- Other properties:
  - Mixing time for $\beta\ll1$ (e.g., Levin & Peres & Wilmer (2008)):

$$T^\bullet_{\text{mix}}\equiv\min\Big\{n:\max_{\boldsymbol{\sigma}}\big\|\delta_{\boldsymbol{\sigma}}*(P^\bullet_\beta)^{*n}-\pi_\beta\big\|_{\text{TV}}\le\tfrac12\Big\}\ \begin{cases}\le{}^\exists C|V|\log|V|,\\ \ge{}^\exists c|V|.\end{cases}$$

  - Simulated annealing (e.g., Catoni (1999)): $^\exists\beta_n=O(\log n)$ such that

$$\mu*P^\bullet_{\beta_1}*P^\bullet_{\beta_2}*\cdots*P^\bullet_{\beta_n}\underset{n\uparrow\infty}{\Longrightarrow}\frac{\mathbb{1}_{\text{GS}}}{|\text{GS}|}.$$

- Both are aperiodic and irreducible and satisfy the detailed balance condition:

$$^\forall\boldsymbol{\sigma},\boldsymbol{\tau}\in\{\pm1\}^V,\qquad \pi_\beta(\boldsymbol{\sigma})\,P^\bullet_\beta(\boldsymbol{\sigma},\boldsymbol{\tau})=\pi_\beta(\boldsymbol{\tau})\,P^\bullet_\beta(\boldsymbol{\tau},\boldsymbol{\sigma}).$$

☺ E.g., for Metropolis, $\boldsymbol{\tau}=\boldsymbol{\sigma}^x \quad\Rightarrow\quad e^{-\beta H(\boldsymbol{\sigma})}e^{-\beta\nabla_x^+H(\boldsymbol{\sigma})}=e^{-\beta\left(H(\boldsymbol{\sigma})\vee H(\boldsymbol{\sigma}^x)\right)}$. ■

$$\therefore\ \pi_\beta=\pi_\beta*P^\bullet_\beta,\qquad\qquad ^\forall\mu*(P^\bullet_\beta)^{*n}\underset{n\uparrow\infty}{\Longrightarrow}\pi_\beta.$$

Theorem 1 (e.g., Catoni (1999))

Let $\{X_n\}_{n=0}^\infty$ be the $\{\pm1\}^V$-valued MC generated by $P^\bullet_\beta=P^g_\beta$ or P^m_β. Then,

$$^\forall\varepsilon>0,\ ^\exists N\in\mathbb{N},\ ^\exists\beta\ge0\ \Big[n\ge N\ \Rightarrow\ \min_{\boldsymbol{\sigma}}P^\bullet_\beta\big(X_n\in\mathit{GS}\,\big|\,X_0=\boldsymbol{\sigma}\big)\ge1-\varepsilon\Big].$$

- Other properties:
  - Mixing time for $\beta\ll1$ (e.g., Levin & Peres & Wilmer (2008)):

$$T^\bullet_{\text{mix}}\equiv\min\Big\{n:\max_{\boldsymbol{\sigma}}\big\|\delta_{\boldsymbol{\sigma}}*(P^\bullet_\beta)^{*n}-\pi_\beta\big\|_{\text{TV}}\le\tfrac12\Big\}\begin{cases}\le{}^\exists C|V|\log|V|,\\ \ge{}^\exists c|V|.\end{cases}$$

  - Simulated annealing (e.g., Catoni (1999)): $^\exists\beta_n=O(\log n)$ such that

$$\mu*P^\bullet_{\beta_1}*P^\bullet_{\beta_2}*\cdots*P^\bullet_{\beta_n}\underset{n\uparrow\infty}{\Longrightarrow}\frac{\mathbb{1}_{\text{GS}}}{|\text{GS}|}.$$

- A $\{\pm1\}^V$-valued MC generated by a transition matrix P_β is said to have rare transition with rate function V if (with the convention $\log 0 = -\infty$)

$$^\forall\boldsymbol{\sigma},\boldsymbol{\tau}\in\{\pm1\}^V,\quad \lim_{\beta\uparrow\infty}\frac{-1}{\beta}\log P_\beta(\boldsymbol{\sigma},\boldsymbol{\tau}) = V(\boldsymbol{\sigma},\boldsymbol{\tau}). \qquad \therefore\ P_\beta(\boldsymbol{\sigma},\boldsymbol{\tau}) \underset{\beta\uparrow\infty}{=} e^{-\beta V(\boldsymbol{\sigma},\boldsymbol{\tau})+o(\beta)}.$$

E.g., for both Glauber and Metropolis, $V(\boldsymbol{\sigma},\boldsymbol{\tau}) = \begin{cases} \nabla_x^+ H(\boldsymbol{\sigma}) & [\boldsymbol{\tau}=\boldsymbol{\sigma}^x], \\ 0 & [\boldsymbol{\tau}=\boldsymbol{\sigma}], \\ \infty & [\text{o/w}]. \end{cases}$

- Define the virtual energy U as

$$U(\boldsymbol{\sigma}) = \min_{g\in G(\boldsymbol{\sigma})}\sum_{(\boldsymbol{\eta},\boldsymbol{\xi})\in g} V(\boldsymbol{\eta},\boldsymbol{\xi}) - \min_{\boldsymbol{\tau}\in\{\pm1\}^V}\min_{g\in G(\boldsymbol{\tau})}\sum_{(\boldsymbol{\eta},\boldsymbol{\xi})\in g} V(\boldsymbol{\eta},\boldsymbol{\xi}),$$

where $G(\boldsymbol{\sigma})$ is the set of oriented spanning trees on $\{\pm1\}^V$ rooted at $\boldsymbol{\sigma}$ such that every $\boldsymbol{\tau}\in\{\pm1\}^V\setminus\{\boldsymbol{\sigma}\}$ has outgoing degree 1.

Theorem 2 (e.g., Catoni (1999))

Let $\{X_n\}_{n=0}^\infty$ be the $\{\pm1\}^V$-valued MC generated by an aperiodic P_β having rare transitions with irreducible rate function V (i.e., the matrix $[e^{-V(\sigma,\tau)}]$ is irreducible). Then the unique stationary distribution obeys $\mu_\beta(\sigma) \underset{\beta\uparrow\infty}{=} e^{-\beta U(\sigma)+o(\beta)}$. In particular,

$$^\forall\varepsilon>0,\ ^\exists N\in\mathbb{N},\ ^\exists\beta\ge0\ \Big[n\ge N \Rightarrow \min_\sigma P_\beta\big(X_n\in\arg\min U \,\big|\, X_0=\sigma\big)\ge 1-\varepsilon\Big].$$

- $^\exists f(\sigma)+V(\sigma,\tau)=f(\tau)+V(\tau,\sigma) \Rightarrow U(\sigma)=f(\sigma)-\min f(\tau)$

- A $\{\pm1\}^V$-valued MC generated by a transition matrix P_β is said to have rare transition with rate function V if (with the convention $\log 0 = -\infty$)

$$^\forall\boldsymbol{\sigma},\boldsymbol{\tau}\in\{\pm1\}^V,\quad \lim_{\beta\uparrow\infty}\tfrac{-1}{\beta}\log P_\beta(\boldsymbol{\sigma},\boldsymbol{\tau})=V(\boldsymbol{\sigma},\boldsymbol{\tau}).\qquad \therefore\ P_\beta(\boldsymbol{\sigma},\boldsymbol{\tau})\underset{\beta\uparrow\infty}{=}e^{-\beta V(\boldsymbol{\sigma},\boldsymbol{\tau})+o(\beta)}.$$

E.g., for both Glauber and Metropolis, $V(\boldsymbol{\sigma},\boldsymbol{\tau})=\begin{cases}\nabla_x^+H(\boldsymbol{\sigma}) & [\boldsymbol{\tau}=\boldsymbol{\sigma}^x],\\ 0 & [\boldsymbol{\tau}=\boldsymbol{\sigma}],\\ \infty & [\text{o/w}].\end{cases}$

- Define the virtual energy U as

$$U(\boldsymbol{\sigma})=\min_{g\in G(\boldsymbol{\sigma})}\sum_{(\boldsymbol{\eta},\boldsymbol{\xi})\in g}V(\boldsymbol{\eta},\boldsymbol{\xi})-\min_{\boldsymbol{\tau}\in\{\pm1\}^V}\min_{g\in G(\boldsymbol{\tau})}\sum_{(\boldsymbol{\eta},\boldsymbol{\xi})\in g}V(\boldsymbol{\eta},\boldsymbol{\xi}),$$

where $G(\boldsymbol{\sigma})$ is the set of oriented spanning trees on $\{\pm1\}^V$ rooted at $\boldsymbol{\sigma}$ such that every $\boldsymbol{\tau}\in\{\pm1\}^V\setminus\{\boldsymbol{\sigma}\}$ has outgoing degree 1.

Theorem 2 (e.g., Catoni (1999))

Let $\{X_n\}_{n=0}^\infty$ be the $\{\pm1\}^V$-valued MC generated by an aperiodic P_β having rare transitions with irreducible rate function V (i.e., the matrix $[e^{-V(\boldsymbol{\sigma},\boldsymbol{\tau})}]$ is irreducible). Then the unique stationary distribution obeys $\mu_\beta(\boldsymbol{\sigma})\underset{\beta\uparrow\infty}{=}e^{-\beta U(\boldsymbol{\sigma})+o(\beta)}$. In particular,

$$^\forall\varepsilon>0,\ ^\exists N\in\mathbb{N},\ ^\exists\beta\ge0\ \Big[n\ge N\ \Rightarrow\ \min_{\boldsymbol{\sigma}}P_\beta\Big(X_n\in\arg\min U\ \Big|\ X_0=\boldsymbol{\sigma}\Big)\ge1-\varepsilon\Big].$$

- $^\exists f(\boldsymbol{\sigma})+V(\boldsymbol{\sigma},\boldsymbol{\tau})=f(\boldsymbol{\tau})+V(\boldsymbol{\tau},\boldsymbol{\sigma})\quad\Rightarrow\quad U(\boldsymbol{\sigma})=f(\boldsymbol{\sigma})-\min_{\boldsymbol{\tau}}f(\boldsymbol{\tau}).$

☺ [Proof of Theorem 1] Recall that both Glauber and Metropolis are aperiodic and irreducible and satisfy

$$\begin{aligned} H(\sigma) + V(\sigma, \sigma^x) = H(\sigma) + \nabla_x^+ H(\sigma) &= H(\sigma) \vee H(\sigma^x) \\ &= H(\sigma^x) + V(\sigma^x, \sigma). \end{aligned}$$

Therefore $U(\sigma) = H(\sigma) - \min_\tau H(\tau)$, hence $\arg\min U$ = GS. ■

- Theorem 2 is obtained from the following general result of MCs:

Proposition 3 (e.g., Catoni (1999))

The unique stationary distribution μ of an aperiodic and irreducible transition matrix P on $\{\pm 1\}^V$ equals

$$\mu(\sigma) = \sum_{g \in G(\sigma)} \prod_{(\eta,\xi) \in g} P(\eta, \xi) \Bigg/ \sum_{\tau \in \{\pm 1\}^V} \sum_{g \in G(\tau)} \prod_{(\eta,\xi) \in g} P(\eta, \xi)$$

☺ [Proof of Theorem 1] Recall that both Glauber and Metropolis are aperiodic and irreducible and satisfy

$$H(\sigma) + V(\sigma,\sigma^x) = H(\sigma) + \nabla_x^+ H(\sigma) = H(\sigma) \vee H(\sigma^x) = H(\sigma^x) + V(\sigma^x,\sigma).$$

Therefore $U(\sigma) = H(\sigma) - \min_\tau H(\tau)$, hence $\arg\min U$ = GS. ■

- Theorem 2 is obtained from the following general result of MCs:

Proposition 3 (e.g., Catoni (1999))

The unique stationary distribution μ of an aperiodic and irreducible transition matrix P on $\{\pm 1\}^V$ equals

$$\mu(\sigma) = \sum_{g\in G(\sigma)} \prod_{(\eta,\xi)\in g} P(\eta,\xi) \Bigg/ \sum_{\tau\in\{\pm 1\}^V} \sum_{g\in G(\tau)} \prod_{(\eta,\xi)\in g} P(\eta,\xi)$$

Key points so far:

1. Find the rate function V and the virtual energy U:

$$V(\sigma,\tau) = \lim_{\beta\uparrow\infty} \frac{-1}{\beta} \log P_\beta(\sigma,\tau),$$

$$U(\sigma) = \min_{g\in G(\sigma)} \sum_{(\eta,\xi)\in g} V(\eta,\xi) - \min_{\tau\in\{\pm1\}^V} \min_{g\in G(\tau)} \sum_{(\eta,\xi)\in g} V(\eta,\xi),$$

or

$$= \lim_{\beta\uparrow\infty} \frac{-1}{\beta} \log \mu_\beta(\sigma) \quad \text{if } [e^{-V(\sigma,\tau)}] \text{ is aperiodic and irreducible.}$$

$$^\exists f(\sigma) + V(\sigma,\tau) = f(\tau) + V(\tau,\sigma) \quad\Rightarrow\quad U(\sigma) = f(\sigma) - \min_\tau f(\tau).$$

2. $^\forall\varepsilon > 0,\ ^\exists N \in \mathbb{N},\ ^\exists\beta \ge 0 \left[n \ge N \;\Rightarrow\; \min_\sigma P_\beta\Big(X_n \in \arg\min U \,\Big|\, X_0 = \sigma\Big) \ge 1-\varepsilon \right].$

3. $\arg\min U$ = GS for both Glauber and Metropolis.

4. $\arg\min U^{\text{DA}}$ = GS.

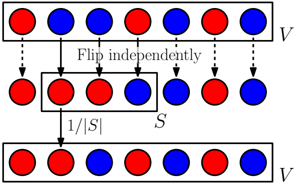
5. $\arg\min U_q^{\text{SCA}}$ = GS if $q > \frac{1}{2}\times$ the largest eigenvalue of $[-J_{x,y}]_{V\times V}$.

6. $\arg\min U^\varepsilon \supset$ GS.

7. DA and ε-SCA are the best in performance among today's MCs.

- Digital annealer (invented in 2017 by Fujitsu Laboratories):

$$P_\beta^{\mathrm{DA}}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \begin{cases} \displaystyle\sum_{x\in S\subset V} \frac{1}{|S|} \prod_{y\in S} e^{-\beta\nabla_y^+ H(\sigma)} \prod_{z\in V-S} \left(1-e^{-\beta\nabla_z^+ H(\sigma)}\right) & [\boldsymbol{\tau}=\boldsymbol{\sigma}^x], \\ \displaystyle\prod_{z\in V}\left(1-e^{-\beta\nabla_z^+ H(\sigma)}\right) & [\boldsymbol{\tau}=\boldsymbol{\sigma}]. \end{cases}$$



- Rate function: $V^{\mathrm{DA}}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \begin{cases} \nabla_x^+ H(\boldsymbol{\sigma}) & [\boldsymbol{\tau}=\boldsymbol{\sigma}^x], \\ 0 & [\boldsymbol{\tau}=\boldsymbol{\sigma}], \\ \infty & [\text{o/w}]. \end{cases}$ $\therefore$ arg min U^{DA} = GS.

$$P_\beta^{\mathrm{DA}}(\boldsymbol{\sigma},\boldsymbol{\sigma}^x) = e^{-\beta\nabla_x^+ H(\sigma)} \underbrace{\sum_{S'\subset V-x} \frac{1}{|S'|+1} \prod_{y\in S'} e^{-\beta\nabla_y^+ H(\sigma)} \prod_{z\in V-x-S'} \left(1-e^{-\beta\nabla_z^+ H(\sigma)}\right)}_{\equiv R_{\beta,x}(\boldsymbol{\sigma})\in[1/|V|,1]}.$$

Theorem 4 (with Fukushima-Kimura, Kawamoto and Noda (2023))

1. $^\forall\varepsilon > 0,\ ^\exists N \in \mathbb{N},\ ^\exists\beta \geq 0\ \Big[n \geq N \Rightarrow \min_{\sigma} P_\beta^{\text{DA}}\big(X_n \in \text{GS} \,\big|\, X_0 = \sigma\big) \geq 1 - \varepsilon\Big].$
2. $(\pi_\beta * P_\beta^{\text{DA}} - \pi_\beta)(\sigma) = \sum_{x \in V} e^{-\beta\nabla_x^+ H(\sigma)}\big(R_{\beta,x}(\sigma^x) - R_{\beta,x}(\sigma)\big).$
3. $J \equiv 0 \Rightarrow \pi_\beta * P_\beta^{\text{DA}} = \pi_\beta.$
4. *If* $J \geq 0,\ h \equiv 0$, *then* $J \equiv 0 \Leftrightarrow \pi_\beta * P_\beta^{\text{DA}} = \pi_\beta.$

[Proof of Theorem 4 4]

- By 3, it suffices to show $J \not\equiv 0 \Rightarrow \pi_\beta * P_\beta^{\text{DA}} \neq \pi_\beta$.
- Let $X_z = \begin{cases} 1 & [\text{w.p. } e^{-\beta\nabla_z^+ H(\sigma)}], \\ 0 & [\text{w.p. } 1 - e^{-\beta\nabla_z^+ H(\sigma)}], \end{cases}$ $Y_z = \begin{cases} 1 & [\text{w.p. } e^{-\beta\nabla_z^+ H(\sigma^x)}], \\ 0 & [\text{w.p. } 1 - e^{-\beta\nabla_z^+ H(\sigma^x)}]. \end{cases}$ Then,

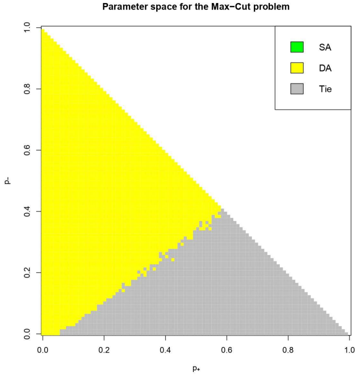
$$R_{\beta,x}(\sigma) = \mathbb{E}\left[\frac{1}{\sum_{z \in V - x} X_z + 1}\,\middle|\, X_x = 1\right], \quad R_{\beta,x}(\sigma^x) = \mathbb{E}\left[\frac{1}{\sum_{z \in V - x} Y_z + 1}\,\middle|\, Y_x = 1\right].$$

- If $\sigma \in \text{GS}$, $z \neq x$, then $\nabla_z^+ H(\sigma^x) = \nabla_z^+ H(\sigma) - 4J_{x,z}$, hence $R_{\beta,x}(\sigma^x) \leq R_{\beta,x}(\sigma)$.
- Since $J \not\equiv 0$, $^\exists x \in V$ such that $R_{\beta,x}(\sigma^x) < R_{\beta,x}(\sigma)$. ■

- E.g., max-cut on the complete graph with $|V| = 128$:

$$H(\boldsymbol{\sigma}) = -\frac{1}{2}\sum_{x,y\in V} J_{x,y}\sigma_x\sigma_y, \quad \text{with i.i.d. } J_{x,y} = \begin{cases} 1 & [\text{w.p. } p_+], \\ -1 & [\text{w.p. } p_-], \\ 0 & [\text{w.p. } 1 - p_+ - p_-]. \end{cases}$$

- $N = 20000$ steps in each of 128 trials.
- Exponential cooling schedule: $\beta_n = \beta_0(\beta_N/\beta_0)^{n/N}$ with $\beta_0 = 0.001, \beta_N = 20$.



- The actual statement for the exponential cooling schedule:

Theorem 5 (with Fukushima-Kimura, Kawamoto and Noda (2023))

$^\exists A$ (= the 1st critical depth), $^\exists B$ (= the difficulty), $^\forall a > A$, $^\forall b \in (0, B)$, $^\forall q \in \mathbb{N}$,

$$\beta_n = \frac{\log(N/q)}{a}\left(\frac{a}{b\epsilon}\right)^{\frac{1}{q}\left[\frac{n-1}{N/q}\right]} \qquad [n = 1, 2, \dots, N]$$

$$\Rightarrow \quad \limsup_{N\to\infty} \frac{\log P^{\mathit{DA}}_{\beta_1} * \cdots * P^{\mathit{DA}}_{\beta_N}\big(U^{\mathit{DA}}(X_N) \geq \epsilon \,\big|\, X_0 = \boldsymbol{\sigma}\big)}{\log N} \leq \frac{-1}{B}\left(\frac{b\epsilon}{a}\right)^{1/q} \equiv -\gamma.$$

Intuitively, for large N,

$$P^{\mathit{DA}}_{\beta_1} * \cdots * P^{\mathit{DA}}_{\beta_N}\big(U^{\mathit{DA}}(X_N) \geq \epsilon \,\big|\, X_0 = \boldsymbol{\sigma}\big) \lesssim N^{-\gamma}.$$

- Stochastic Cellular Automata (SCA): $P^{\text{SCA}}_{\beta,q}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \dfrac{e^{-\beta \tilde{H}_q(\boldsymbol{\sigma},\boldsymbol{\tau})}}{\sum_{\boldsymbol{\eta}} e^{-\beta \tilde{H}_q(\boldsymbol{\sigma},\boldsymbol{\eta})}}$, where

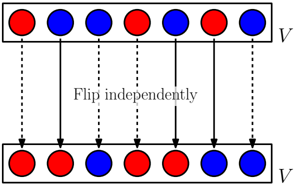
$$\tilde{H}_q(\boldsymbol{\sigma},\boldsymbol{\tau}) = -\frac{1}{2}\sum_{x,y} J_{x,y}\sigma_x\tau_y - \frac{1}{2}\sum_x h_x(\sigma_x+\tau_x) - \frac{q}{2}\sum_x \sigma_x\tau_x.$$

- Since $\tilde{H}_q(\boldsymbol{\sigma},\boldsymbol{\sigma}) = H(\boldsymbol{\sigma}) - \frac{q}{2}|V|$,

$$P^{\text{SCA}}_{\beta,q}(\boldsymbol{\sigma},\boldsymbol{\tau}) \underset{q\uparrow\infty}{\longrightarrow} \pi_\beta(\boldsymbol{\sigma})\,\delta_{\boldsymbol{\sigma},\boldsymbol{\tau}}.$$

- Since $\tilde{H}_q(\boldsymbol{\sigma},\boldsymbol{\tau}) = -\frac{1}{2}\sum_x \Big(\underbrace{\sum_y J_{x,y}\sigma_y + h_x}_{\tilde{h}_x(\boldsymbol{\sigma})} + q\sigma_x\Big)\tau_x - \frac{1}{2}\sum_x h_x\sigma_x$,

$$P^{\text{SCA}}_{\beta,q}(\boldsymbol{\sigma},\boldsymbol{\tau}) = \prod_{x\in V} \frac{e^{\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q\sigma_x)\tau_x}}{2\cosh\left(\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q\sigma_x)\right)}.$$



- $P^{\text{SCA}}_{\beta,q}$ is aperiodic and irreducible and $\pi^{\text{SCA}}_{\beta,q}(\sigma) \propto \sum_\eta e^{-\beta \tilde{H}_q(\sigma,\eta)}$ satisfies

$$^\forall\sigma,\eta \in \{\pm 1\}^V, \qquad \pi^{\text{SCA}}_{\beta,q}(\sigma) P^{\text{SCA}}_{\beta,q}(\sigma,\eta) = \pi^{\text{SCA}}_{\beta,q}(\eta) P^{\text{SCA}}_{\beta,q}(\eta,\sigma).$$

$$\therefore \;\; \pi^{\text{SCA}}_{\beta,q} = \pi^{\text{SCA}}_{\beta,q} * P^{\text{SCA}}_{\beta,q} \qquad\qquad {}^\forall\mu * (P^{\text{SCA}}_{\beta,q})^{*n} \underset{n\uparrow\infty}{\Longrightarrow} \pi^{\text{SCA}}_{\beta,q}.$$

- Rate function: $V^{\text{SCA}}_q(\sigma,\tau) = \tilde{H}_q(\sigma,\tau) - \min_\eta \tilde{H}_q(\sigma,\eta) \equiv \tilde{H}_q(\sigma,\tau) - m_q(\sigma)$.

$$☺ \quad \frac{-1}{\beta} \log \frac{1}{\sum_\tau e^{-\beta \tilde{H}_q(\sigma,\tau)}} = -m_q(\sigma) + \frac{1}{\beta} \log \sum_\tau e^{-\beta V^{\text{SCA}}_q(\sigma,\tau)} \underset{\beta\uparrow\infty}{\longrightarrow} -m_q(\sigma). \qquad \blacksquare$$

- Since $m_q(\sigma) + V^{\text{SCA}}_q(\sigma,\tau) = m_q(\tau) + V^{\text{SCA}}_q(\tau,\sigma)$,

$$U^{\text{SCA}}_q(\sigma) = m_q(\sigma) - \min_\eta m_q(\eta).$$

Theorem 6 (with Fukushima-Kimura *et al.* (2023); Okuyama *et al.* (2019))

1. $T^{\text{SCA}}_{\text{mix}} = O(\log|V|)$ *for* $\beta \ll 1$ *(depending on* $\{J_{x,y}\}, \{h_x\}, q$*)* *(cf.,* $T^{g}_{\text{mix}} \geq {}^\exists c|V|$*).*
2. $\arg\min U^{SCA}_q = GS$ *if* $q > \frac{1}{2} \times$ *the largest eigenvalue of* $[-J_{x,y}]_{V\times V}$*. As a result,*

$$^\forall\varepsilon > 0,\ {}^\exists\beta \geq 0,\ {}^\exists N \in \mathbb{N}\ \Big[n \geq N \ \Rightarrow\ \min_\sigma P^{SCA}_{\beta,q}\big(X_n \in GS \,\big|\, X_0 = \sigma\big) \geq 1 - \varepsilon \Big].$$

- $P^{\text{SCA}}_{\beta,q}$ is aperiodic and irreducible and $\pi^{\text{SCA}}_{\beta,q}(\boldsymbol{\sigma}) \propto \sum_{\boldsymbol{\eta}} e^{-\beta \tilde{H}_q(\boldsymbol{\sigma},\boldsymbol{\eta})}$ satisfies

$$^{\forall}\boldsymbol{\sigma},\boldsymbol{\eta} \in \{\pm 1\}^V, \quad \pi^{\text{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma}) P^{\text{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \pi^{\text{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\eta}) P^{\text{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\eta},\boldsymbol{\sigma}).$$

$$\therefore \; \pi^{\text{SCA}}_{\beta,q} = \pi^{\text{SCA}}_{\beta,q} * P^{\text{SCA}}_{\beta,q} \qquad {}^{\forall}\mu * (P^{\text{SCA}}_{\beta,q})^{*n} \underset{n\uparrow\infty}{\Longrightarrow} \pi^{\text{SCA}}_{\beta,q}.$$

- Rate function: $V^{\text{SCA}}_q(\boldsymbol{\sigma},\boldsymbol{\tau}) = \tilde{H}_q(\boldsymbol{\sigma},\boldsymbol{\tau}) - \min_{\boldsymbol{\eta}} \tilde{H}_q(\boldsymbol{\sigma},\boldsymbol{\eta}) \equiv \tilde{H}_q(\boldsymbol{\sigma},\boldsymbol{\tau}) - m_q(\boldsymbol{\sigma})$.

$$\frac{-1}{\beta}\log\frac{1}{\sum_{\boldsymbol{\tau}} e^{-\beta\tilde{H}_q(\boldsymbol{\sigma},\boldsymbol{\tau})}} = -m_q(\boldsymbol{\sigma}) + \frac{1}{\beta}\log\sum_{\boldsymbol{\tau}} e^{-\beta V^{\text{SCA}}_q(\boldsymbol{\sigma},\boldsymbol{\tau})} \underset{\beta\uparrow\infty}{\longrightarrow} -m_q(\boldsymbol{\sigma}). \quad \blacksquare$$

- Since $m_q(\boldsymbol{\sigma}) + V^{\text{SCA}}_q(\boldsymbol{\sigma},\boldsymbol{\tau}) = m_q(\boldsymbol{\tau}) + V^{\text{SCA}}_q(\boldsymbol{\tau},\boldsymbol{\sigma})$,

$$U^{\text{SCA}}_q(\boldsymbol{\sigma}) = m_q(\boldsymbol{\sigma}) - \min_{\boldsymbol{\eta}} m_q(\boldsymbol{\eta}).$$

Theorem 6 (with Fukushima-Kimura *et al.* (2023); Okuyama *et al.* (2019))

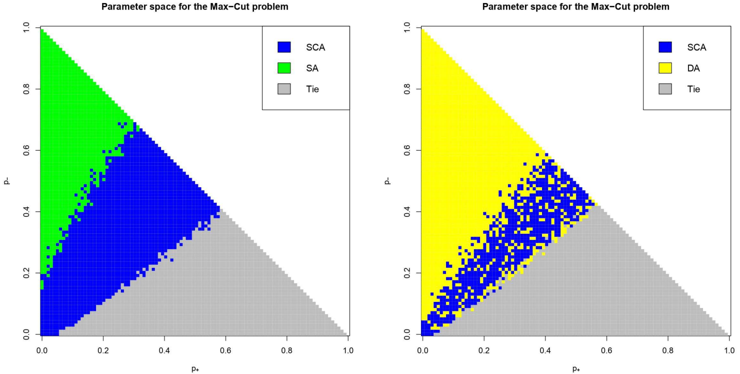
1. $T^{SCA}_{mix} = O(\log|V|)$ *for* $\beta \ll 1$ *(depending on* $\{J_{x,y}\}, \{h_x\}, q$*)* *(cf.,* $T^{g}_{mix} \geq {}^{\exists}c|V|$*)*.
2. $\arg\min U^{SCA}_q = GS$ *if* $q > \frac{1}{2}\times$ *the largest eigenvalue of* $[-J_{x,y}]_{V\times V}$*. As a result,*

$$^{\forall}\varepsilon > 0, \; ^{\exists}\beta \geq 0, \; ^{\exists}N \in \mathbb{N} \; \left[n \geq N \;\Rightarrow\; \min_{\boldsymbol{\sigma}} P^{SCA}_{\beta,q}\left(X_n \in GS \,\middle|\, X_0 = \boldsymbol{\sigma}\right) \geq 1 - \varepsilon \right].$$

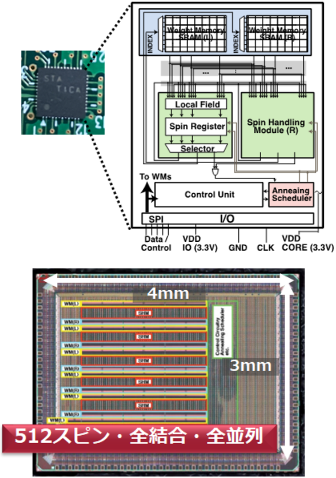
- E.g., max-cut on the complete graph with $|V| = 128$:

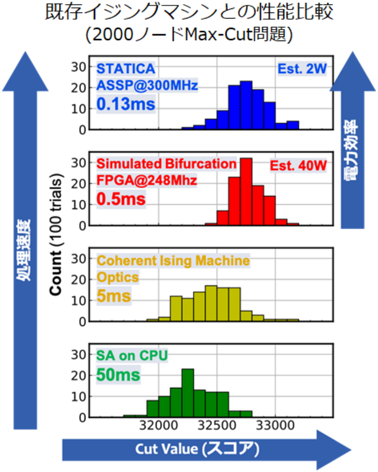
$$H(\boldsymbol{\sigma}) = -\frac{1}{2}\sum_{x,y\in V} J_{x,y}\sigma_x\sigma_y, \quad \text{with i.i.d. } J_{x,y} = \begin{cases} 1 & [\text{w.p. } p_+], \\ -1 & [\text{w.p. } p_-], \\ 0 & [\text{w.p. } 1 - p_+ - p_-]. \end{cases}$$

- $N = 20000$ steps in each of 128 trials.
- Exponential cooling schedule: $\beta_n = \beta_0(\beta_N/\beta_0)^{n/N}$ with $\beta_0 = 0.001, \beta_N = 20$.



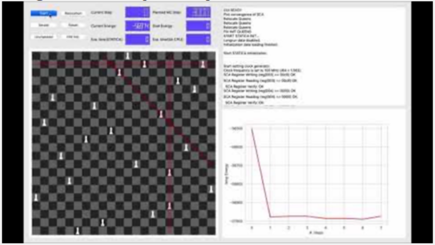
Ⓐ STATICAプロトタイプチップ

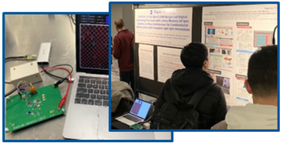




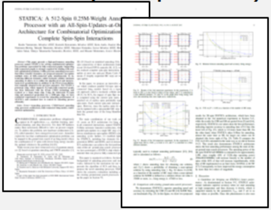
STATICA関連の成果・業績

ISSCC2020発表，およびチップ動作実演デモ：N-Queen問題 (N=22)





IEEE Journal of Solid State Circuits (IEEE Xplore掲載済)



日経新聞 朝刊

毎日新聞 朝刊

IEEE Spectrum

アニーリングの社会応用関係企業からのコンタクト多数

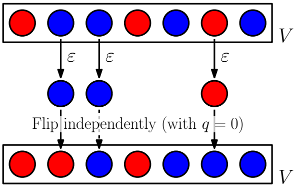
現在: スケールアップ・実応用展開に向けて，SCAの収束挙動解析とGPU実行環境を整備中

- Let $D_{\sigma,\tau}=\{x:\sigma_x\neq\tau_x\}$ and rewrite $P^{\text{SCA}}_{\beta,q}(\sigma,\tau)$ by isolating the effect of q as

$$\begin{aligned}P^{\text{SCA}}_{\beta,q}(\sigma,\tau)&=\prod_{x\in V}\frac{e^{\frac{\beta}{2}q\sigma_x\tau_x}\cosh\left(\frac{\beta}{2}\tilde{h}_x(\sigma)\right)}{\cosh\left(\frac{\beta}{2}(\tilde{h}_x(\sigma)+q\sigma_x)\right)}\frac{e^{\frac{\beta}{2}\tilde{h}_x(\sigma)\tau_x}}{2\cosh\left(\frac{\beta}{2}\tilde{h}_x(\sigma)\right)}\\&=\prod_{x\in D_{\sigma,\tau}}\varepsilon_x(\sigma)p_x(\sigma)\prod_{y\in V\setminus D_{\sigma,\tau}}\Big(1-\varepsilon_y(\sigma)p_y(\sigma)\Big),\qquad\varepsilon_x(\sigma)=O_\sigma(e^{-\beta q}).\end{aligned}$$

- ε-SCA:

$$\begin{aligned}P^{\varepsilon}_{\beta}(\sigma,\tau)&=\prod_{x\in D_{\sigma,\tau}}\varepsilon p_x(\sigma)\prod_{y\in V\setminus D_{\sigma,\tau}}\Big(1-\varepsilon p_x(\sigma)\Big)\\&=\sum_{S:D_{\sigma,\tau}\subset S\subset V}\varepsilon^{|S|}(1-\varepsilon)^{|V\setminus S|}\prod_{x\in D_{\sigma,\tau}}p_x(\sigma)\prod_{y\in S\setminus D_{\sigma,\tau}}\Big(1-p_x(\sigma)\Big).\end{aligned}$$



- Since P_β^ε is aperiodic and irreducible, there must be a unique stationary distribution π_β^ε, which is not Gibbs in general.
- Rate function: $V^\varepsilon(\boldsymbol{\sigma},\boldsymbol{\tau}) = \sum_{x\in D_{\boldsymbol{\sigma},\boldsymbol{\tau}}} \big(\tilde{h}_x(\boldsymbol{\sigma})\,\sigma_x\big)^+$.

☺
$$\tfrac{-1}{\beta}\log P_\beta^\varepsilon(\boldsymbol{\sigma},\boldsymbol{\tau}) = \sum_{x\in D_{\boldsymbol{\sigma},\boldsymbol{\tau}}} \tfrac{-1}{\beta}\log p_x(\boldsymbol{\sigma}) + o(1)$$
$$= \sum_{x\in D_{\boldsymbol{\sigma},\boldsymbol{\tau}}} \Big(\tfrac{1}{2}\tilde{h}_x(\boldsymbol{\sigma})\sigma_x + \tfrac{1}{\beta}\log\big(2\cosh\big(\tfrac{\beta}{2}\tilde{h}_x(\boldsymbol{\sigma})\big)\big)\Big) + o(1). \qquad \blacksquare$$

- Since V^ε is irreducible, there is a virtual energy $U^\varepsilon(\boldsymbol{\sigma}) = \lim_{\beta\uparrow\infty} \tfrac{-1}{\beta}\log \pi_\beta^\varepsilon(\boldsymbol{\sigma})$.

  In fact, ${}^\exists b_\varepsilon \in (0,1)$ (independent of β) such that
$$b_\varepsilon e^{-\beta U^\varepsilon(\boldsymbol{\sigma})} \le \pi_\beta^\varepsilon(\boldsymbol{\sigma}) \le b_\varepsilon^{-1} e^{-\beta U^\varepsilon(\boldsymbol{\sigma})}.$$

Theorem 7 (with Fukushima-Kimura *et al.* (202?))

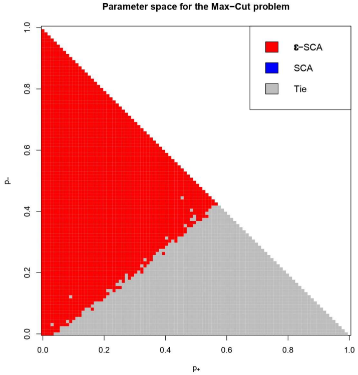
1. $\pi_\beta^\varepsilon \underset{\varepsilon\downarrow 0}{\Longrightarrow} \pi_\beta$ *uniformly in* β; $\lim_{\beta\uparrow\infty}\pi_\beta^\varepsilon$ *exists for every* $\varepsilon\in(0,1)$.
2. $\arg\min U^\varepsilon \supset GS$ *(i.e., GS* $\setminus \arg\min U^\varepsilon = \varnothing$*).*

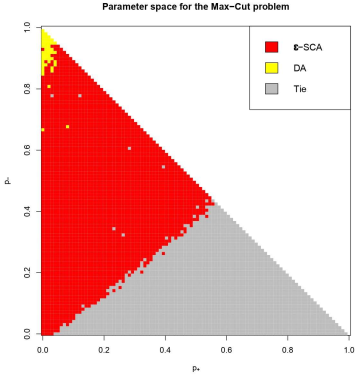
☺ [Proof of Theorem 7 2] $\boldsymbol{\sigma}\in \mathrm{GS}\setminus\arg\min U^\varepsilon \;\Rightarrow\; 0 = \lim_{\varepsilon\downarrow 0}\lim_{\beta\uparrow\infty}\pi_\beta^\varepsilon(\boldsymbol{\sigma})$
$= \lim_{\beta\uparrow\infty}\lim_{\varepsilon\downarrow 0}\pi_\beta^\varepsilon(\boldsymbol{\sigma}) = \lim_{\beta\uparrow\infty}\pi_\beta(\boldsymbol{\sigma}) = \frac{1}{|\mathrm{GS}|}$, which is a contradiction. $\blacksquare$

- E.g., max-cut on the complete graph with $|V| = 128$:

$$H(\boldsymbol{\sigma}) = -\frac{1}{2}\sum_{x,y\in V} J_{x,y}\sigma_x\sigma_y, \quad \text{with i.i.d. } J_{x,y} = \begin{cases} 1 & [\text{w.p. } p_+], \\ -1 & [\text{w.p. } p_-], \\ 0 & [\text{w.p. } 1 - p_+ - p_-]. \end{cases}$$

- $N = 20000$ steps in each of 128 trials.
- Exponential cooling schedule: $\beta_n = \beta_0(\beta_N/\beta_0)^{n/N}$ with $\beta_0 = 0.001, \beta_N = 20$.





Today's key points:

1. Find the rate function V and the virtual energy U:

$$V(\sigma,\tau)=\lim_{\beta\uparrow\infty}\frac{-1}{\beta}\log P_\beta(\sigma,\tau),$$

$$U(\sigma)=\min_{g\in G(\sigma)}\sum_{(\eta,\xi)\in g}V(\eta,\xi)-\min_{\tau\in\{\pm1\}^V}\min_{g\in G(\tau)}\sum_{(\eta,\xi)\in g}V(\eta,\xi),$$

or

$$=\lim_{\beta\uparrow\infty}\frac{-1}{\beta}\log\mu_\beta(\sigma)\quad\text{if }\ [e^{-V(\sigma,\tau)}]\ \text{ is aperiodic and irreducible.}$$

$$^{\exists}f(\sigma)+V(\sigma,\tau)=f(\tau)+V(\tau,\sigma)\quad\Rightarrow\quad U(\sigma)=f(\sigma)-\min_\tau f(\tau).$$

2. $^{\forall}\varepsilon>0,\ ^{\exists}N\in\mathbb{N},\ ^{\exists}\beta\geq 0\ \Big[n\geq N\ \Rightarrow\ \min_\sigma P_\beta\big(X_n\in\arg\min U\ \big|\ X_0=\sigma\big)\geq 1-\varepsilon\Big].$
3. $\arg\min U$ = GS for both Glauber and Metropolis.
4. $\arg\min U^{\text{DA}}$ = GS.
5. $\arg\min U_q^{\text{SCA}}$ = GS if $q>\frac{1}{2}\times$ the largest eigenvalue of $[-J_{x,y}]_{V\times V}$.
6. $\arg\min U^\varepsilon\supset$ GS.
7. DA and ε-SCA are the best in performance among today's MCs.