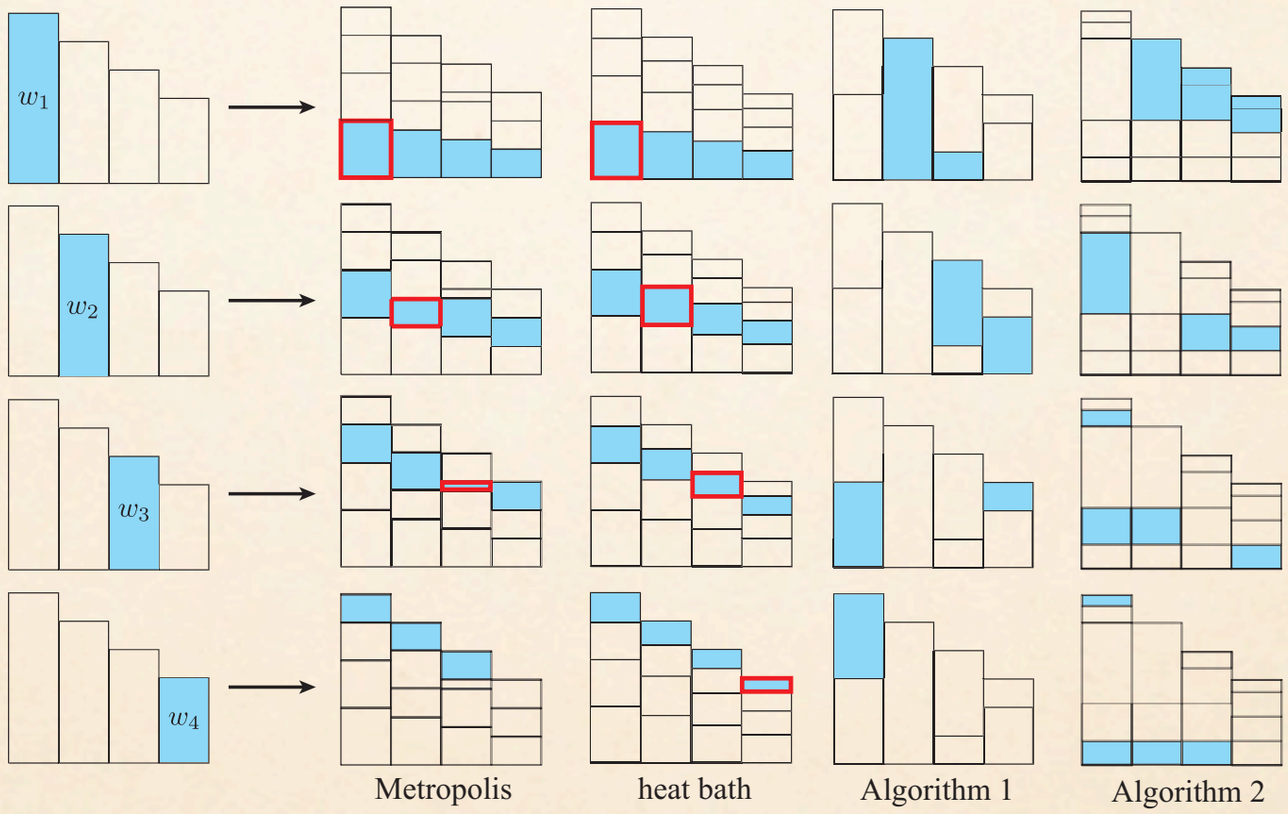
Sep. 1, 2011 Eighth IMACS Seminar on Monte Carlo Methods@Borovets, Bulgaria

Geometric Allocation Approach for Transition Kernel of Markov Chain



Outline

1. Review of Markov chain Monte Carlo
2. New geometric algorithms
3. Extension to continuous variables

Department of Applied Physics, The Univ. of Tokyo

Hidemaro Suwa

Collaborator : Synge Todo

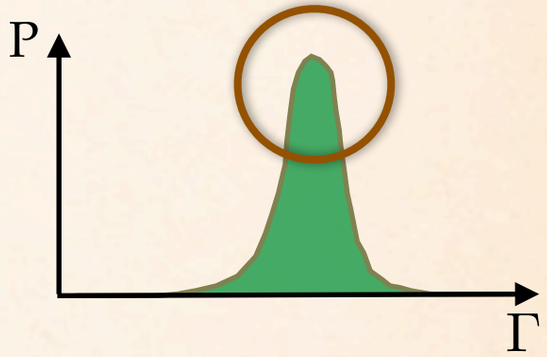
Markov Chain Monte Carlo (MCMC)

"Multiple degrees of freedom"

is essential in many models, especially in physics.

e.g. Ising model, Potts model, Heisenberg model, etc

How should samples be selected?



Monte Carlo integration → "Curse of dimensionality"

Solution : Importance sampling, Markov chain Monte Carlo

Samples can be generated from an arbitrary target distribution.
Possible to calculate expectation values asymptotically without approximation.

Broad Applications :

Physics, Chemistry, Biology, (Bayesian) Statistics, Machine learning, Finance, etc.

Autocorrelation Time

In MCMC, a configuration is generated with transition probabilities, that is, a transition kernel.

Markov chain : $\sigma_0 \to \sigma_1 \to \cdots \to \sigma_k \to \sigma_{k+1} \to \cdots$

Instead of Curse of dimensionality,

"Sample correlation" is the big problem.

Autocorrelation time (Asymptotic variance)

= needed number of steps for configuration to forget the history.

Variance of Monte Carlo average

s^2 : Naive variance

$$\left\langle \left(\frac{1}{M} \sum_i A_i - \langle A \rangle \right)^2 \right\rangle = \left\langle \frac{1}{M^2} \sum_i \left(A_i^2 - \langle A \rangle^2 \right) \right\rangle + \left\langle \frac{2}{M^2} \sum_{i=1}^{M} \sum_{t=1}^{M-i} \left(A_{i+t} A_i - \langle A \rangle^2 \right) \right\rangle \simeq \frac{s^2}{M} (1 + 2\tau_{\mathrm{int}})$$

Autocorrelation function :

$$C(t) = \frac{\langle A_{i+t} A_i \rangle - \langle A \rangle^2}{\langle A^2 \rangle - \langle A \rangle^2}$$

Autocorrelation time :

$$\tau_{\mathrm{int}} = \sum_{t=1}^{\infty} C(t)$$

number of effective samples

$$M \to \frac{M}{1 + 2\tau_{\mathrm{int}}}$$

needed computational cost $\propto$ correlation time

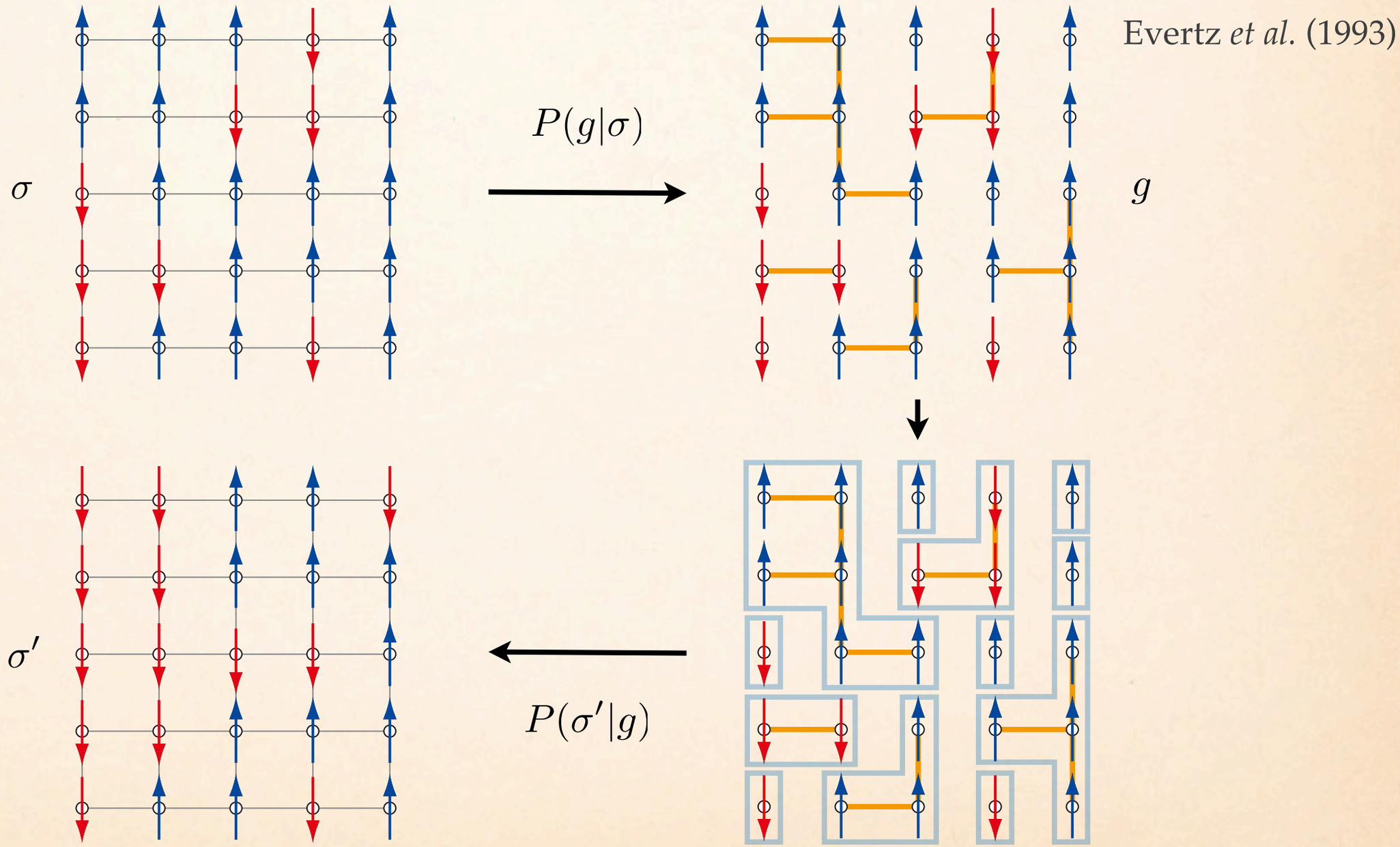
It is important to develop an update method with short correlation time!

Cluster Algorithm

・The Swendsen-Wang algorithm and the loop algorithm (quantum version), can overcome "critical slowing down" in some models.

Swendsen & Wang (1987)

Evertz *et al.* (1993)



Extended Ensemble Methods

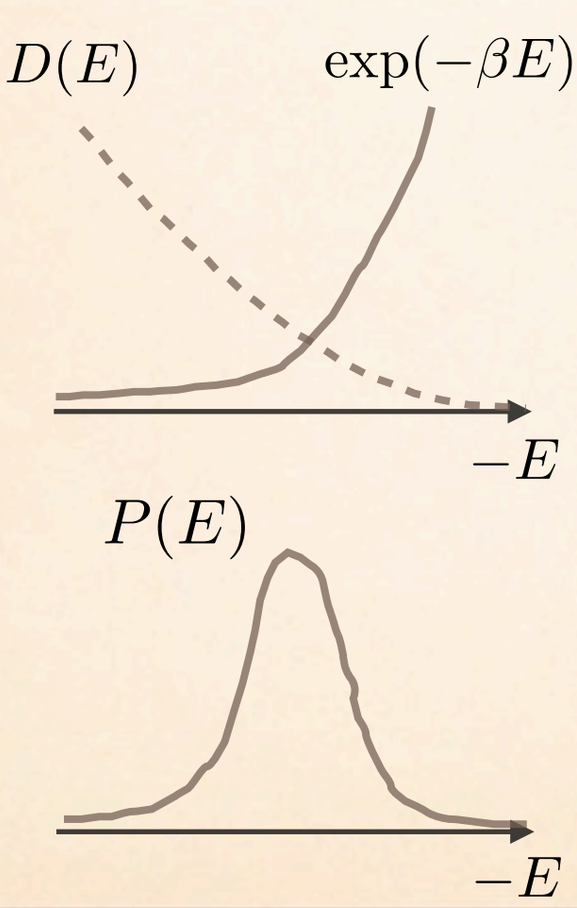
・The multicanonical method, and the exchange Monte Carlo method have been applied to various hardly-relaxing problems e.g., protein folding problems and spin glass systems.

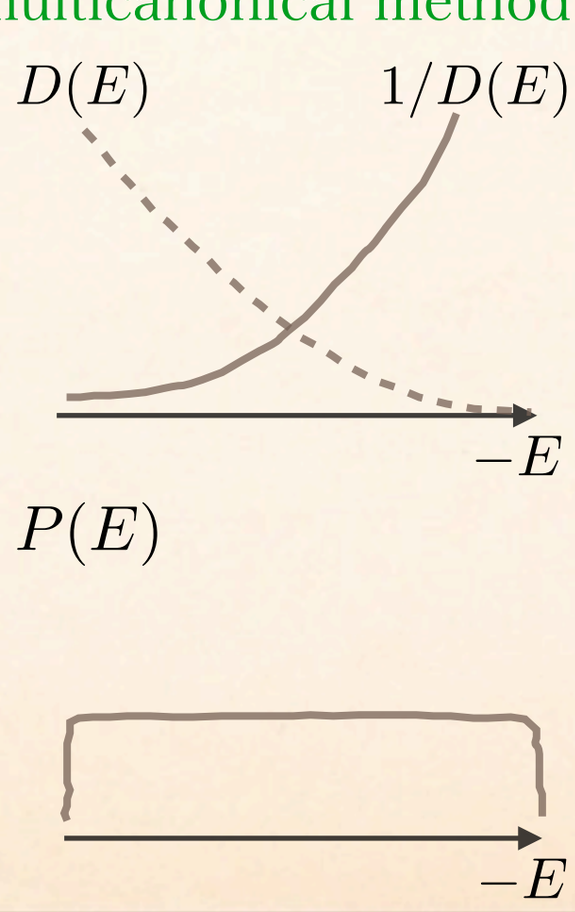
Keywords : Breaking constraints, Rare event sampling

Learning Wang & Landau (2001)
Berg at el. (1992)

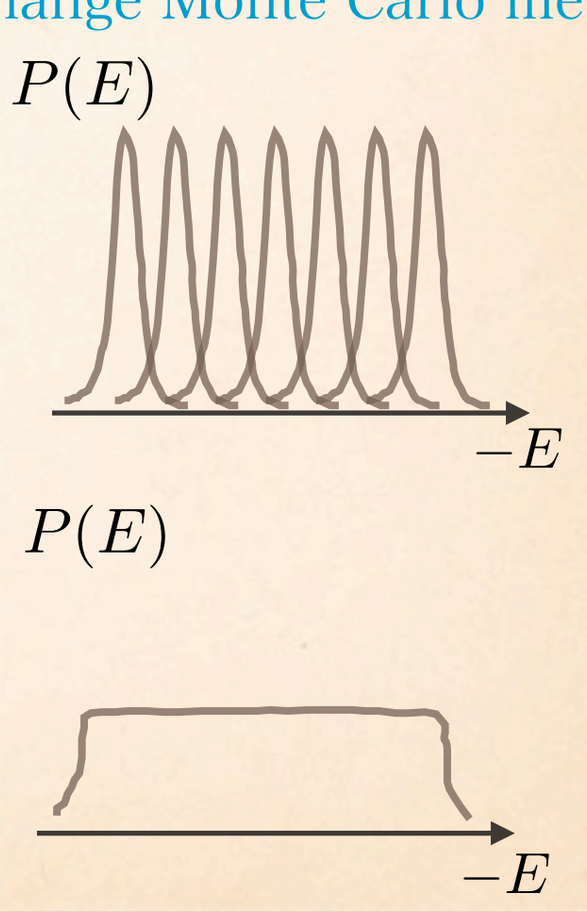
Hukushima & Nemoto (1996)

The multicanonical method





The exchange Monte Carlo method



Optimization of Probabilities

In MCMC, we must consider 3 problems.

1. Ensemble (reweighting) : Multicanonical, Exchange Monte Carlo
2. Candidate configurations : Cluster algorithms
3. Transition kernel (probability) : Metropolis, Heat bath (Gibbs sampler)

Let us consider a discrete variable now.

Heat bath algorithm (Gibbs sampler)

Baker (1965), Geman and Geman (1984)

$$P(c_i \to c_j) = \frac{W(c_j)}{\sum_k W(c_k)}$$

Metropolis
(Metropolis-Hastings) algorithm

Metropolis *et al.* (1953), Hastings (1970)

$$P(c_i \to c_j) = \min\left(1, \frac{W(c_j)}{W(c_i)}\right)$$

There are only few algorithms that construct the transition kernel.

These canonical algorithms are not optimal !

What is "Optimal" ?

We have one guideline : Minimize the rejection rate ! Peskun (1973)

Rejection is defined as the state does not move to others.

Metropolized sampling Hastings (1970) Liu (1996)

$$T_{ij}^{MG} = \begin{bmatrix} 0 & \frac{\pi_2}{1-\pi_1} & \frac{\pi_3}{1-\pi_1} & \cdots & \frac{\pi_n}{1-\pi_1} \\ \frac{\pi_1}{1-\pi_1} & 1-\cdots & \frac{\pi_3}{1-\pi_2} & \cdots & \frac{\pi_n}{1-\pi_2} \\ \frac{\pi_1}{1-\pi_1} & \frac{\pi_2}{1-\pi_2} & 1-\cdots & \cdots & \frac{\pi_n}{1-\pi_3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\pi_1}{1-\pi_1} & \frac{\pi_2}{1-\pi_2} & \frac{\pi_3}{1-\pi_3} & \cdots & 1-\cdots \end{bmatrix}$$

Locally optimal update (LOU) Pollet *et al.* (2004)

$$T_{ij}^{Opt} = \begin{bmatrix} 0 & \frac{W_2}{W_1}y_1 & \frac{W_3}{W_1}y_1 & \cdots & \frac{W_n}{W_1}y_1 \\ y_1 & 0 & \frac{W_3}{W_2}y_2 & \cdots & \frac{W_n}{W_2}y_2 \\ y_1 & y_2 & 0 & \cdots & \frac{W_n}{W_3}y_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_1 & y_2 & y_3 & \cdots & 1-y_1-y_2-\cdots \end{bmatrix}$$

$$T_{ij} = P(c_i \to c_j)$$

$$T_{ij}^{MG} = \min\left(\frac{\pi_j}{1-\pi_i}, \frac{\pi_j}{1-\pi_j}\right)$$

$$\pi_i = \frac{Wi}{\sum_j W_j}$$

$$y_1 = \frac{\pi_1}{1-\pi_1}$$

$$y_2 = \frac{(1-y_1)\pi_2}{1-\pi_1-\pi_2}$$

$$y_3 = \frac{(1-y_1-y_2)\pi_3}{1-\pi_1-\pi_2-\pi_3}$$

Also these methods fail to minimize the rejection rate !

Reversibility of Markov Chain

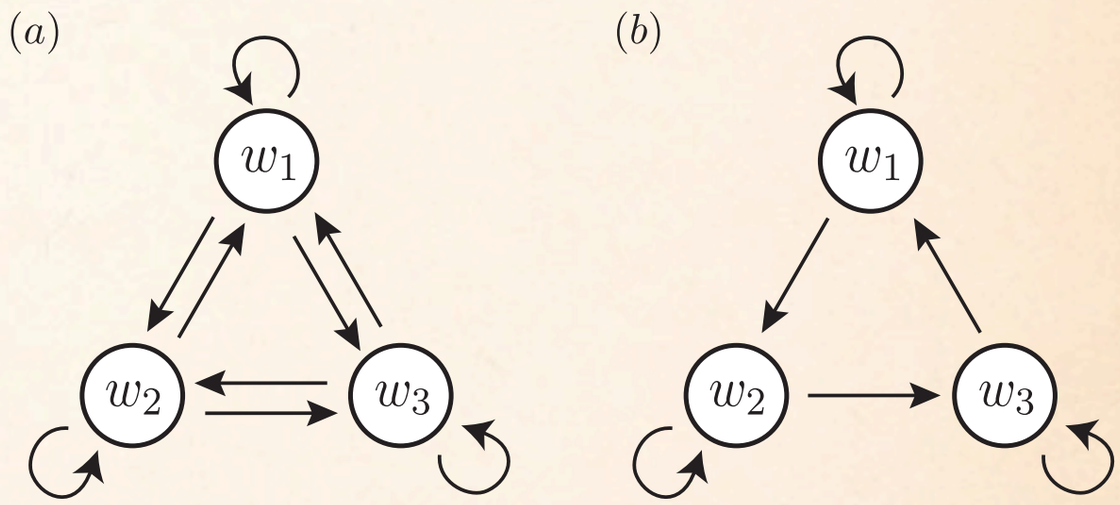
Total balance is imposed for sampling from a target distribution.

$$w_i = \sum_{j=1}^{n} w_i \, p_{i\to j} = \sum_{j=1}^{n} w_j \, p_{j\to i}$$

Sufficient condition

Detailed balance = Reversibility

$$w_i \, p_{i\to j} = w_j \, p_{j\to i} \qquad \forall \, i, \, j$$



Reversible Irreversible

1953

MCMC has evolved within the detailed balance since the invention,
e.g., Metropolis and heat bath algorithms

However, the reversibility is NOT a necessary condition.

Stochastic Flow

Problem : Determine transition probabilities from weights within the total balance

Guideline : Minimize the rejection rate (the state stays still) $\{w_i\} \to \{p_{i\to j}\}$

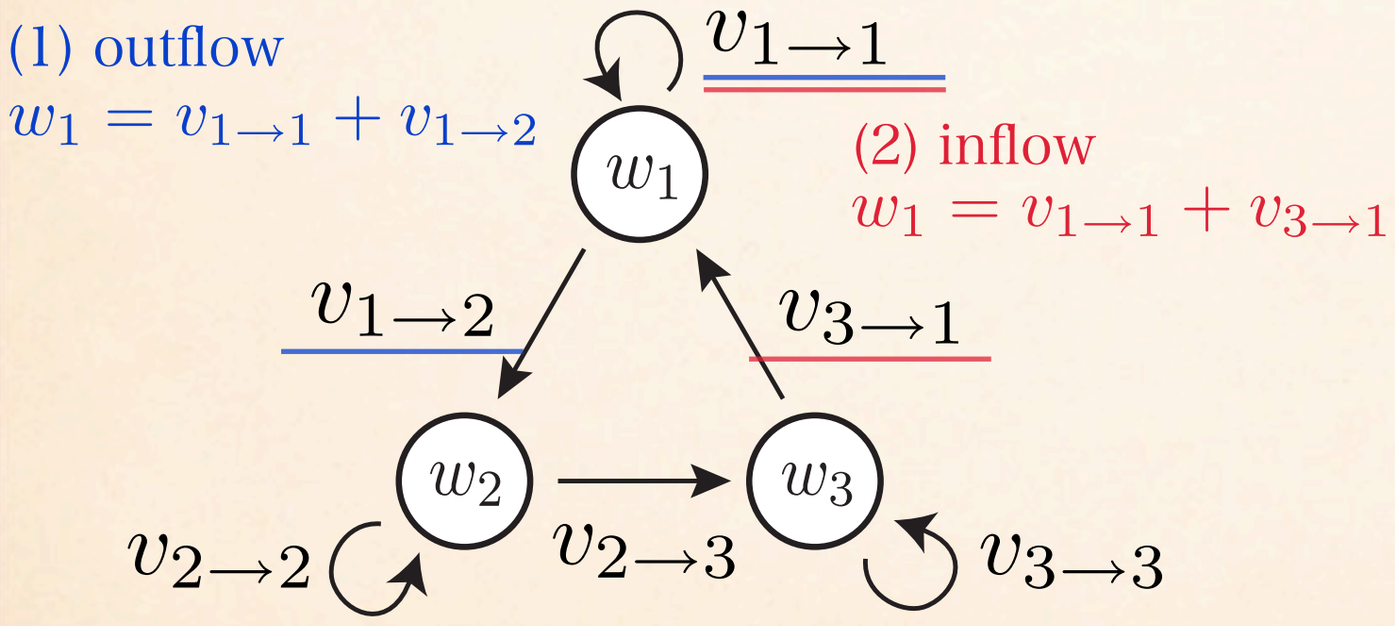
(Raw) Stochastic Flow ⇒ Visualize the condition

$$v_{i\to j} := w_i\, p_{i\to j}$$

n : Number of candidate states

Usual notation

$$w_i = \sum_{j=1}^{n} w_i\, p_{i\to j} = \sum_{j=1}^{n} w_j\, p_{j\to i}$$



Probability Conservation Law

$$w_i = \sum_{j=1}^{n} v_{i\to j} \quad \forall i \qquad (1)$$

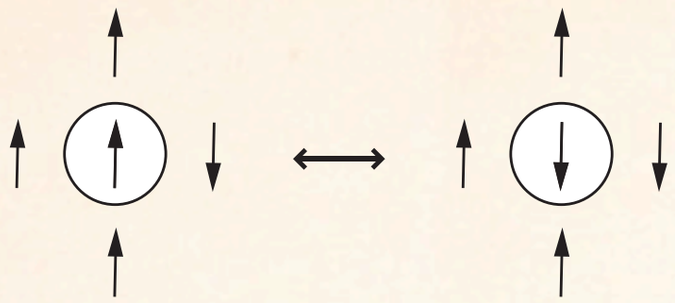
Total Balance with Stochastic Flow

$$w_j = \sum_{i=1}^{n} v_{i\to j} \quad \forall j \qquad (2)$$

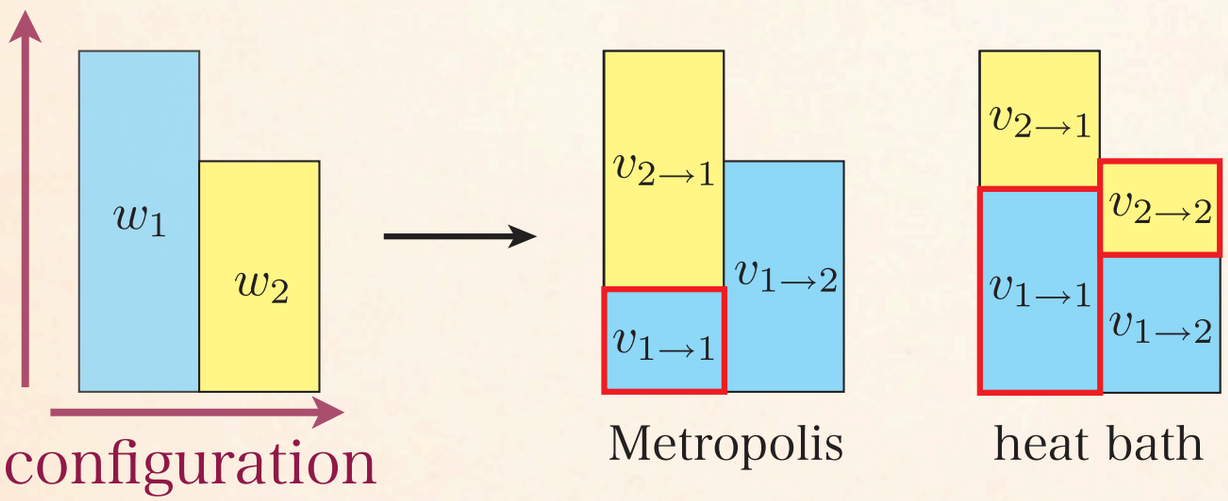
Find a set $\{v_{i\to j}\}$ that minimizes $\sum_i v_{i\to i} / \sum_i w_i$ under (1) and (2).

Geometric Allocation : n=2

For example : Single spin update for Ising (binary) spins



$w_1 = e^{2\beta}$ $w_2 = e^{-2\beta}$



Rejection rate $\propto$ Region with red frame

Probability Conservation Law

$$w_i = \sum_{j=1}^{n} v_{i\to j} \quad \forall i \qquad (1)$$

Total Balance with Stochastic Flow

$$w_j = \sum_{i=1}^{n} v_{i\to j} \quad \forall j \qquad (2)$$

Metropolis

$$v_{i\to j} = \frac{1}{n-1}\min(w_i, w_j) \qquad i \neq j$$

$$p_{i\to j} = v_{i\to j}/w_i$$

Heat bath

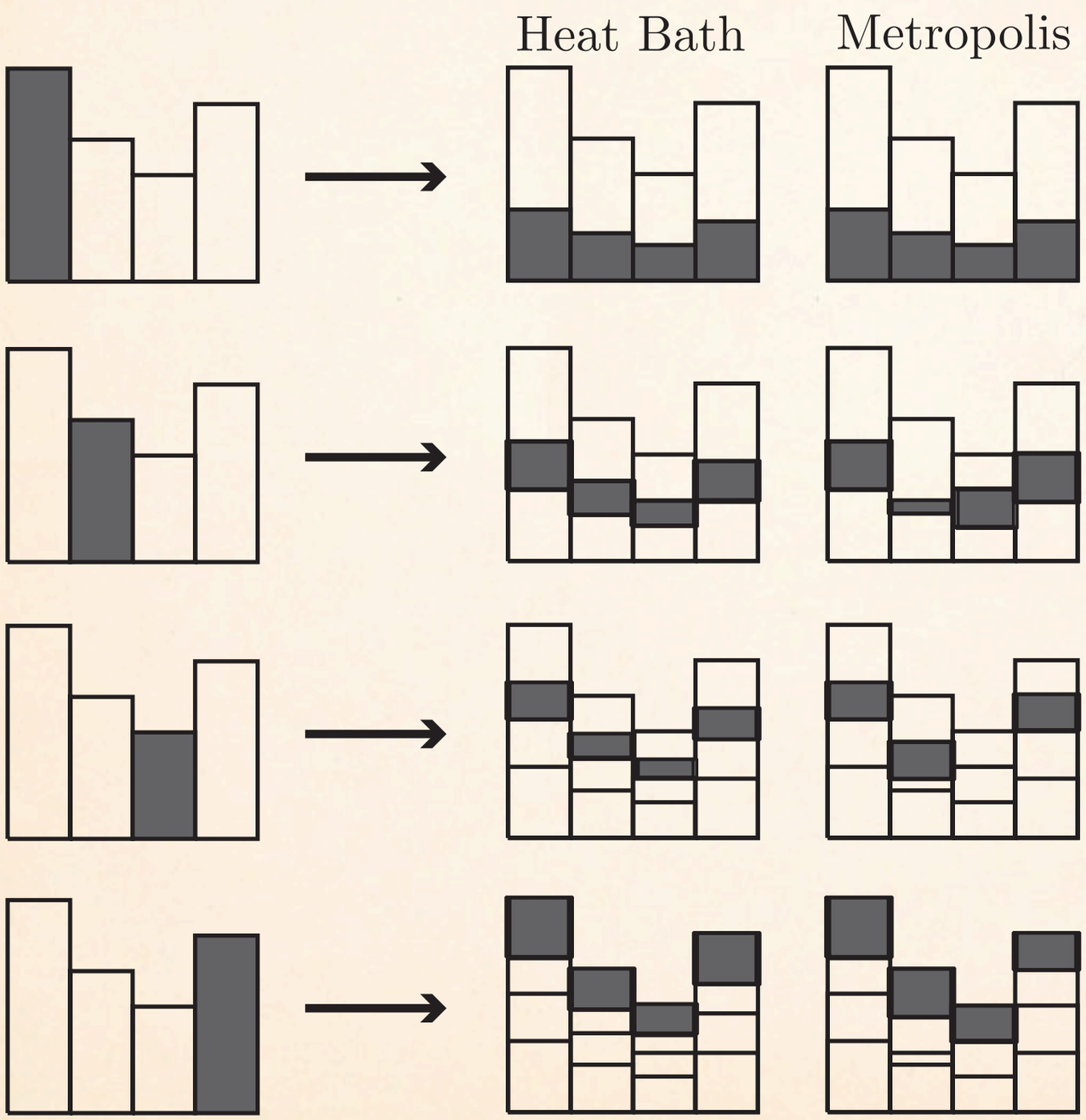
$$v_{i\to j} = \frac{w_i w_j}{\sum_{k=1}^{n} w_k} \qquad \forall i, j$$

Geometric Approach :
Allocate weights so that the entire shape of weight boxes is kept intact

Reversibility : $v_{i\to j} = v_{j\to i}$

Inevitable Rejection? : n=4

・Let us allocate weights in a 4-candidate case, like 4-state Potts model.



q-state Potts model : $\sigma_i = 1, 2, ..., q$

$$H = -\sum_{<i,j>} \delta_{\sigma_i \sigma_j}$$

These algorithms always leave the nonzero rejection rate.

Is it possible to reduce the rejection rate to zero ?

Yes!

Metropolis

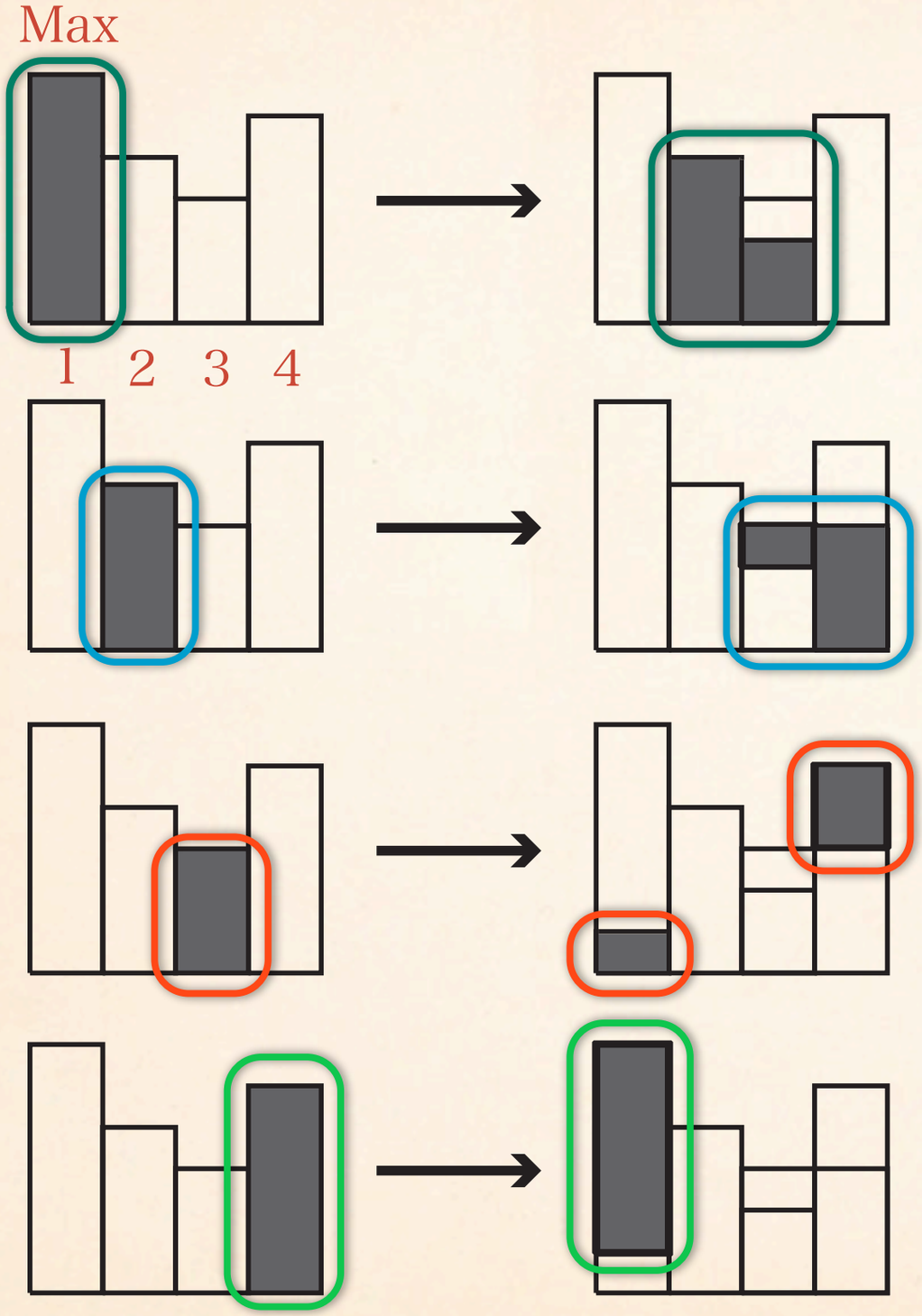
$$v_{i \to j} = \frac{1}{n-1} \min(w_i, w_j) \qquad i \neq j$$

Heat bath

$$v_{i \to j} = \frac{w_i w_j}{\sum_{k=1}^{n} w_k} \qquad \forall\, i, j$$

H.S. and Todo, (2010)
Phys. Rev. Lett., 105, 120603

New Algorithm



1. Choose a configuration that has the max weight and number it 1. Sort others in an arbitrary order.

2. Allocate the max weight w_1 to the next box (i=2). Reallocate the remain to the subsequent box (i=3).

3. Allocate the weight of the first filled box (i=2) to the subsequent position of Step 2. Continue to reallocate the remain until it vanishes.

4. Continue to allocate all weights likewise, keeping the starting order to the landfilled order.

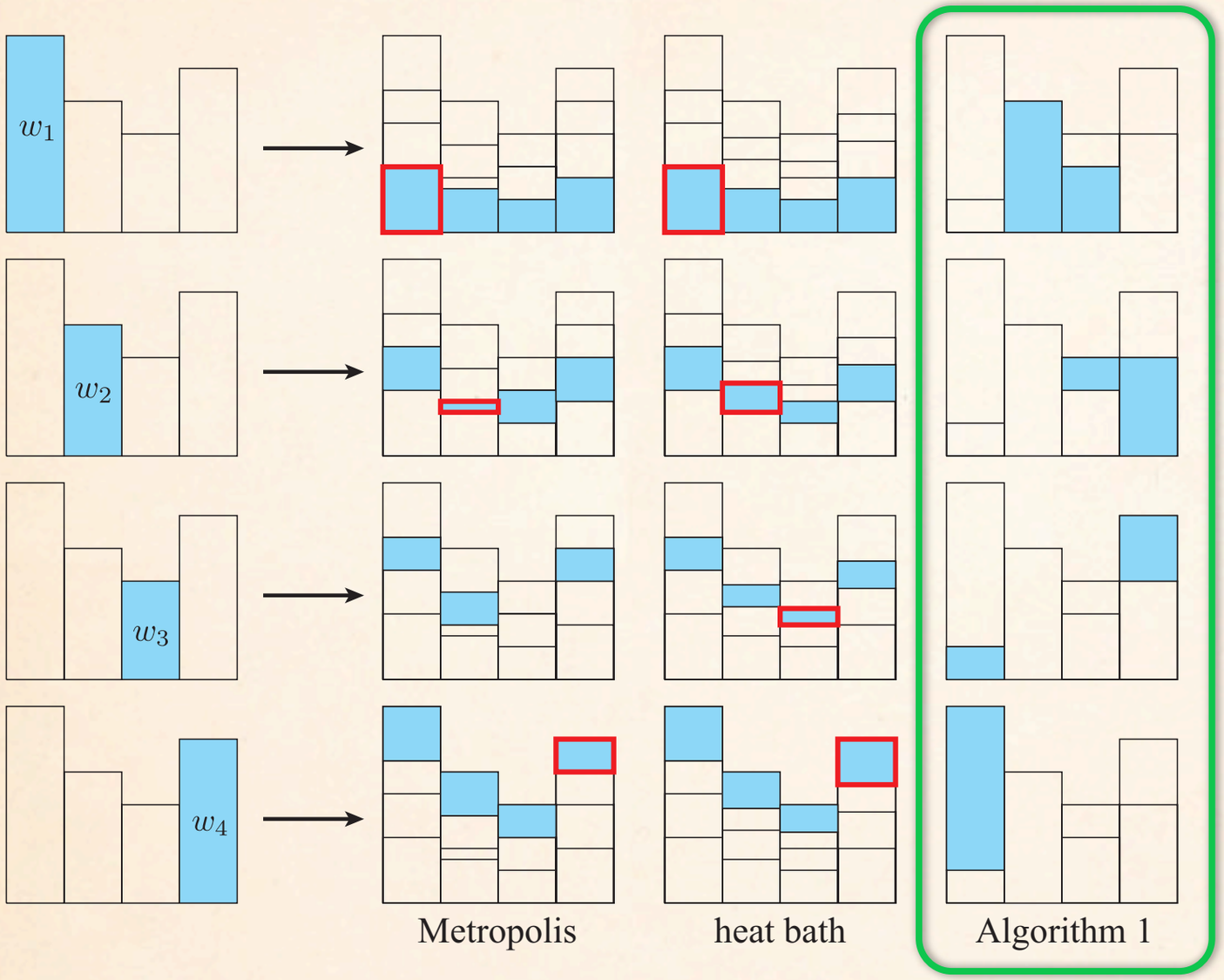
5. Define transition probabilities as,

$$p_{i\to j} = v_{i\to j}/w_i$$

- There is NO self-allocated weight !

Rejection-free Algorithm!

Comparison with Conventional Methods



$$\Delta_{ij} := S_i - S_{j-1} + w_1 \qquad 1 \le i, j \le n$$

$$S_i := \sum_{k=1}^{i} w_k \qquad 1 \le i \le n$$

$$S_0 := S_n.$$

$$v_{i\to j} = \max(0,\, \min(\Delta_{ij},\, w_i + w_j - \Delta_{ij},\, w_i,\, w_j)),$$

Average rejection rate

$$\frac{\sum_i v_{i\to i}}{\sum_i w_i} = \frac{\max(0,\, w_1 - \sum_{i=2}^{n} w_i)}{\sum_i w_i}$$

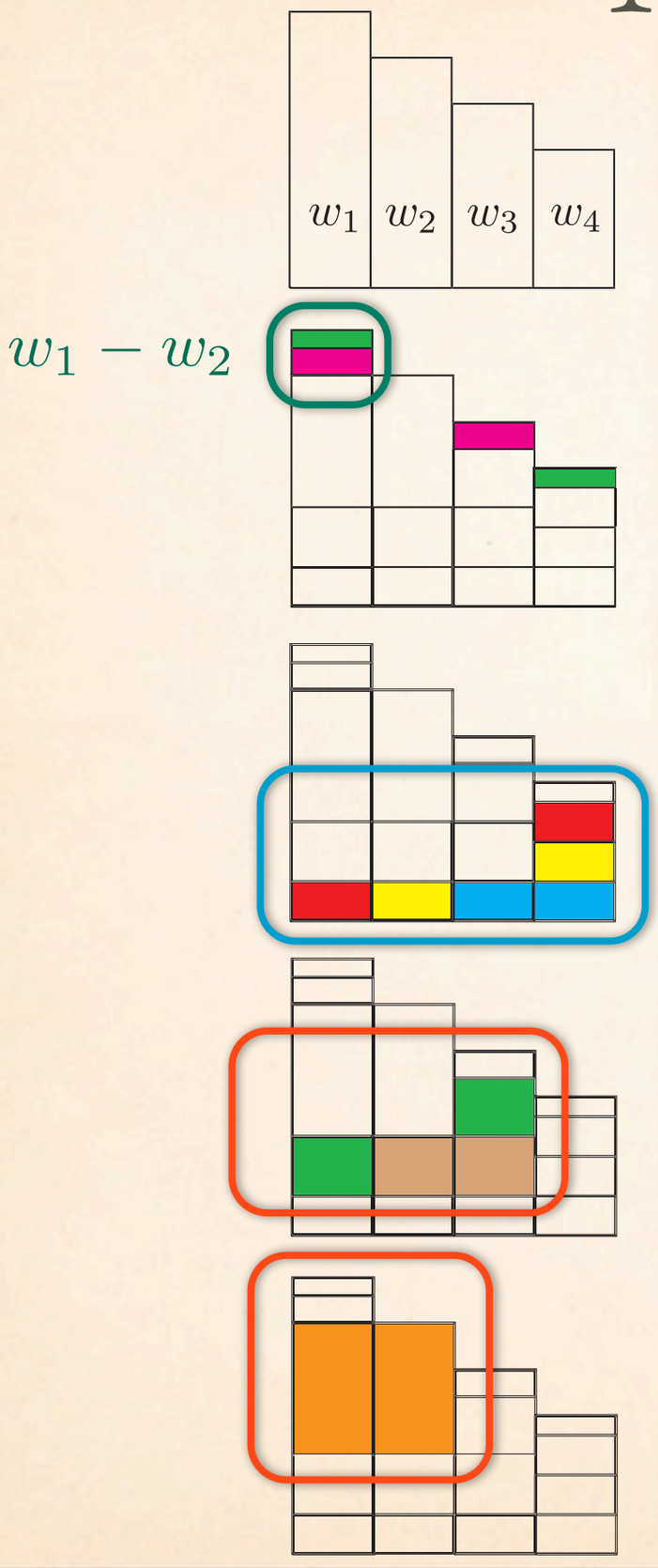
Rejection-free condition

$$w_1 \le \sum_{i=2}^{n} w_i$$

- No additional CPU time cost (making a table of probabilities before sampling)
- Breaks the detailed balance ---First versatile method without reversibility---

- Always minimizes the average rejection rate, given candidates !

New Algorithm 2



1. Sort n candidate configurations as

$$w_1 \geq w_2 \geq w_3 \geq \ldots \geq w_n (n \geq 3)$$

2. Allocate the weight $w_1 - w_2$ to the boxes (i > 2) in proportion to w_i, and swap the same weight.
Residual weights are

$$w'_1 = w'_2 \geq w'_3 \geq \ldots \geq w'_n$$

3. Swap the residual weight $w'_n/(n-1)$ equally with $1 \leq i \leq n-1$
Residuals are

$$w''_1 = w''_2 \geq w''_3 \geq \ldots \geq w''_n = 0$$

4. Continue to swap for n-1, n-2, ..., 2 likewise.

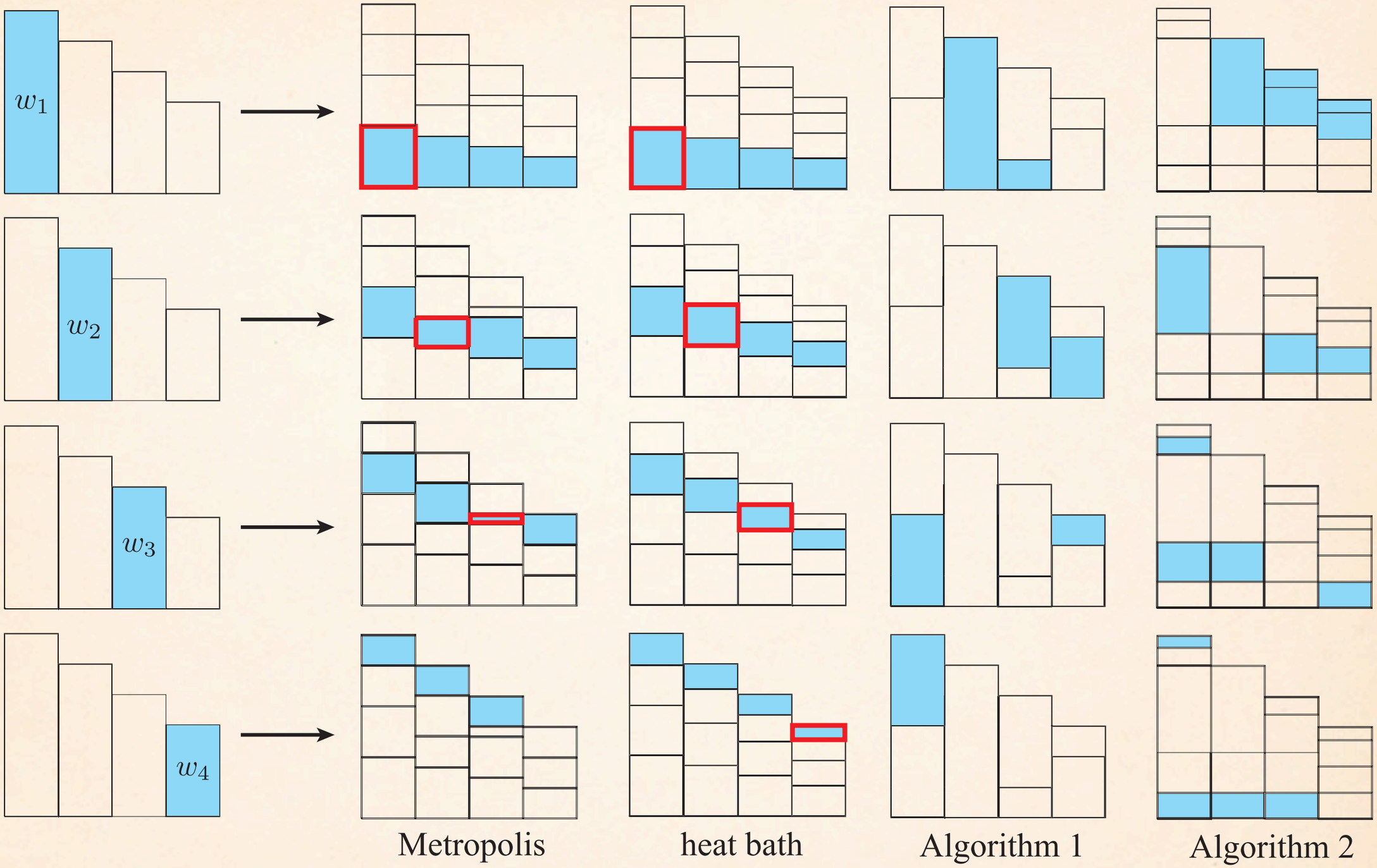
No rejection !

Reversibility :

$$v_{i \to j} = v_{j \to i}$$

This algorithm consists of only the swaps.
Therefore it satisfies the detailed balance (reversibility).

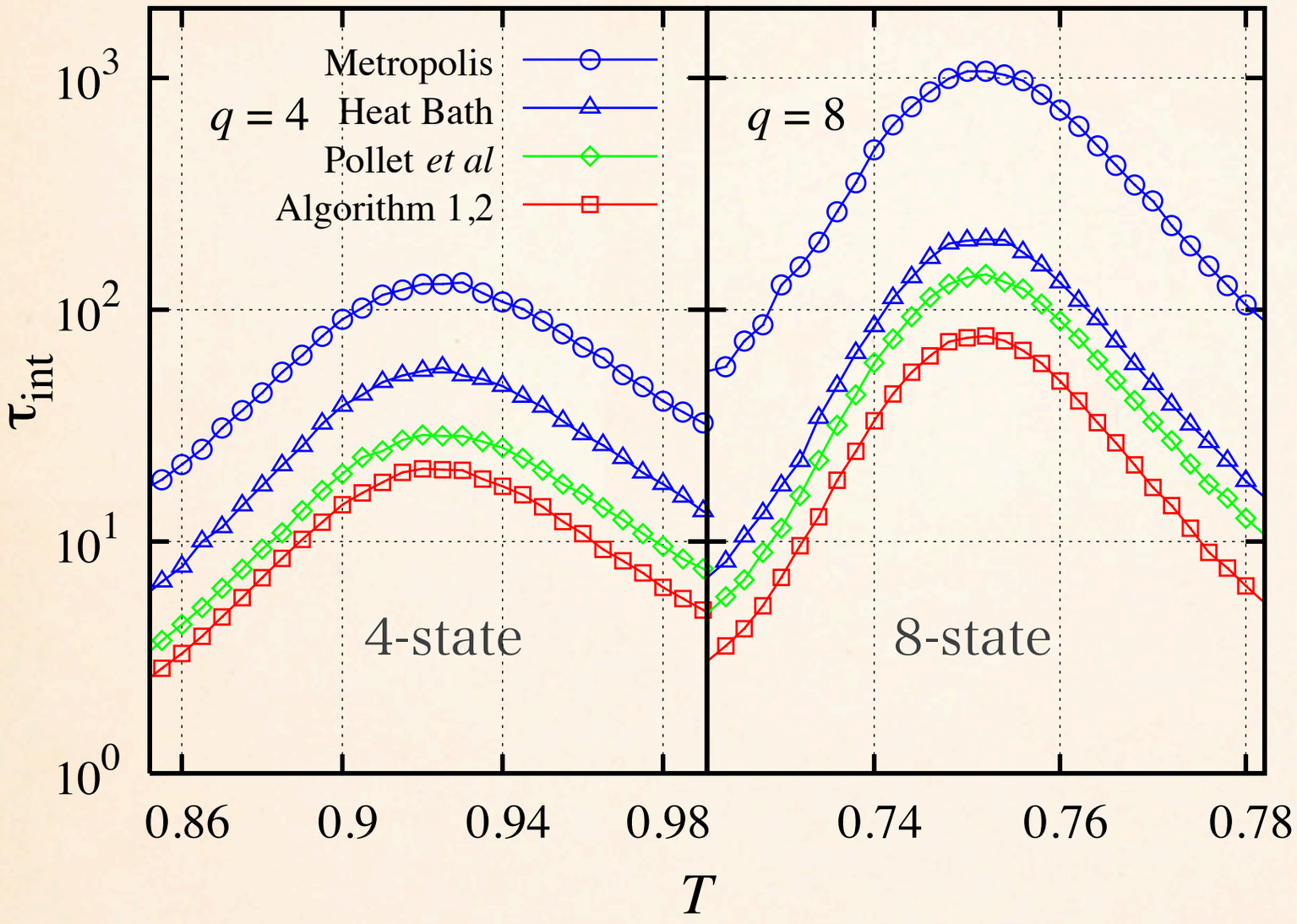
Geometric Allocation Picture



- New algorithms always minimizes the average rejection rate !

Acceleration in Potts Model

Structure Factor of Order Parameter



Square lattice $L = 16$

Ferromagnetic q-state Potts model $\sigma_i = 1, 2, ..., q$

Wu (1982)

$$H = -\sum_{<i,j>} \delta_{\sigma_i \sigma_j}$$

Phase transition temperature

$$T = 1/\ln(1+\sqrt{q})$$

Continuous $q \leq 4$

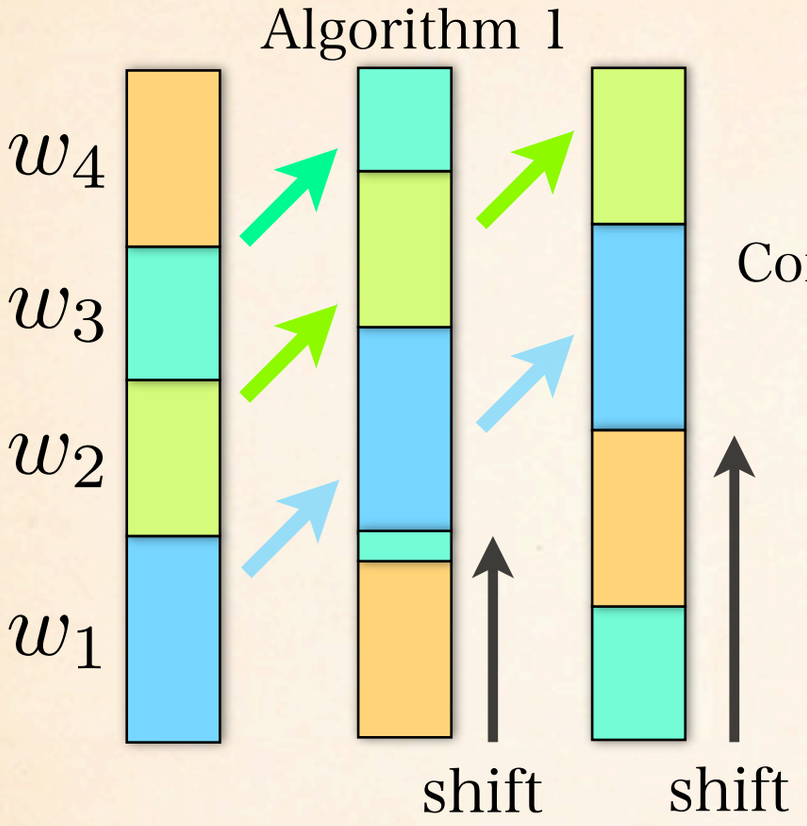
1st order $q > 4$

$T_c \simeq 0.910239$ $T_c \simeq 0.744904$

Metropolis	× 6.4	× 14
Heat bath	× 2.7	× 2.6

Significantly speed up !

Extension to Continuous Variable

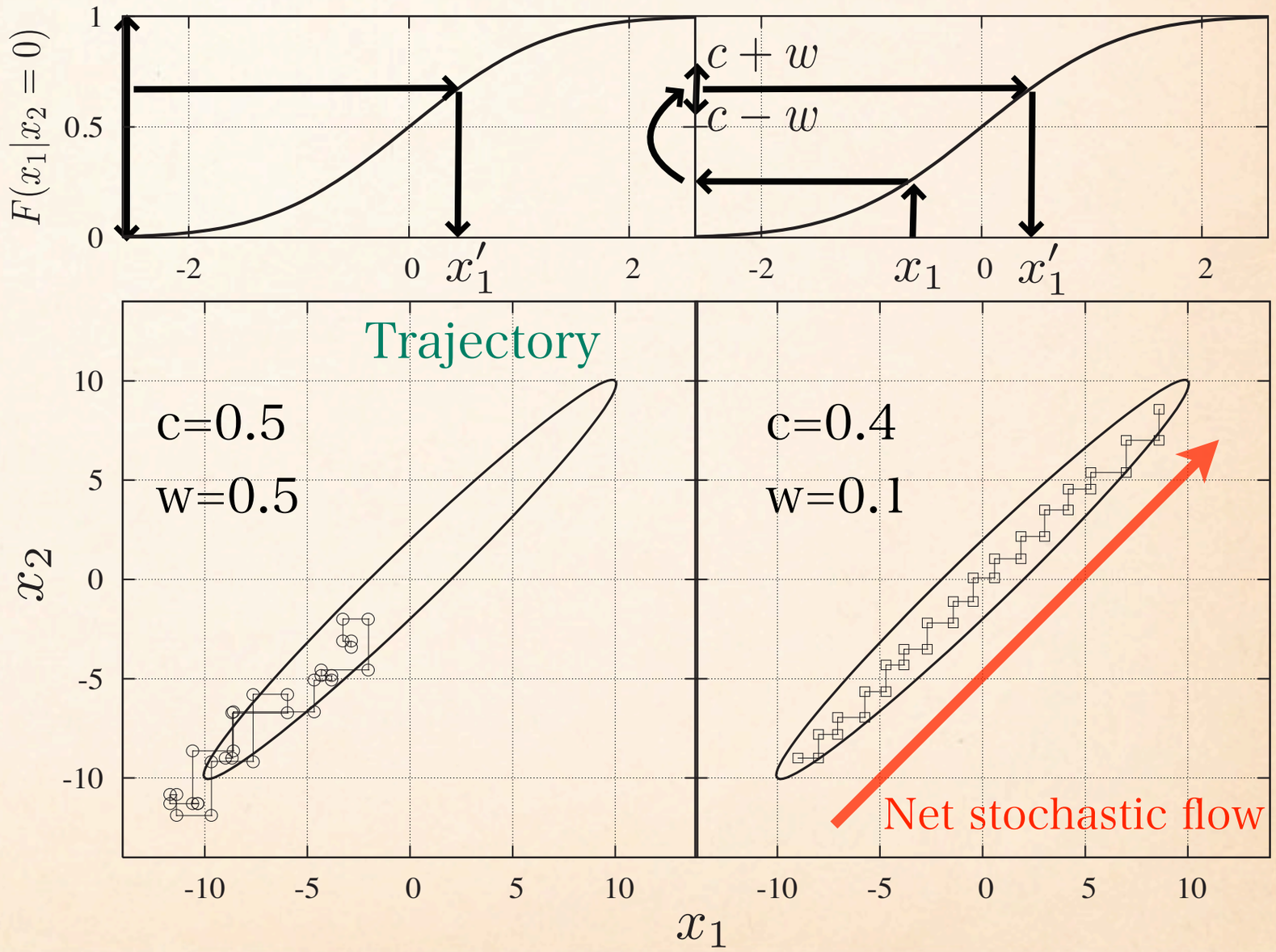


Bivariate Gaussian distribution

$$P(x_1, x_2) \propto e^{-\frac{(x_1-x_2)^2}{2\sigma_1^2} - \frac{(x_1+x_2)^2}{2\sigma_2^2}}$$

Conditional cumulative distribution $F(x_1|x_2) = \int_{-\infty}^{x_1} P(x|x_2)\,dx$

Heat bath (Gibbs sampler) present (irreversible)

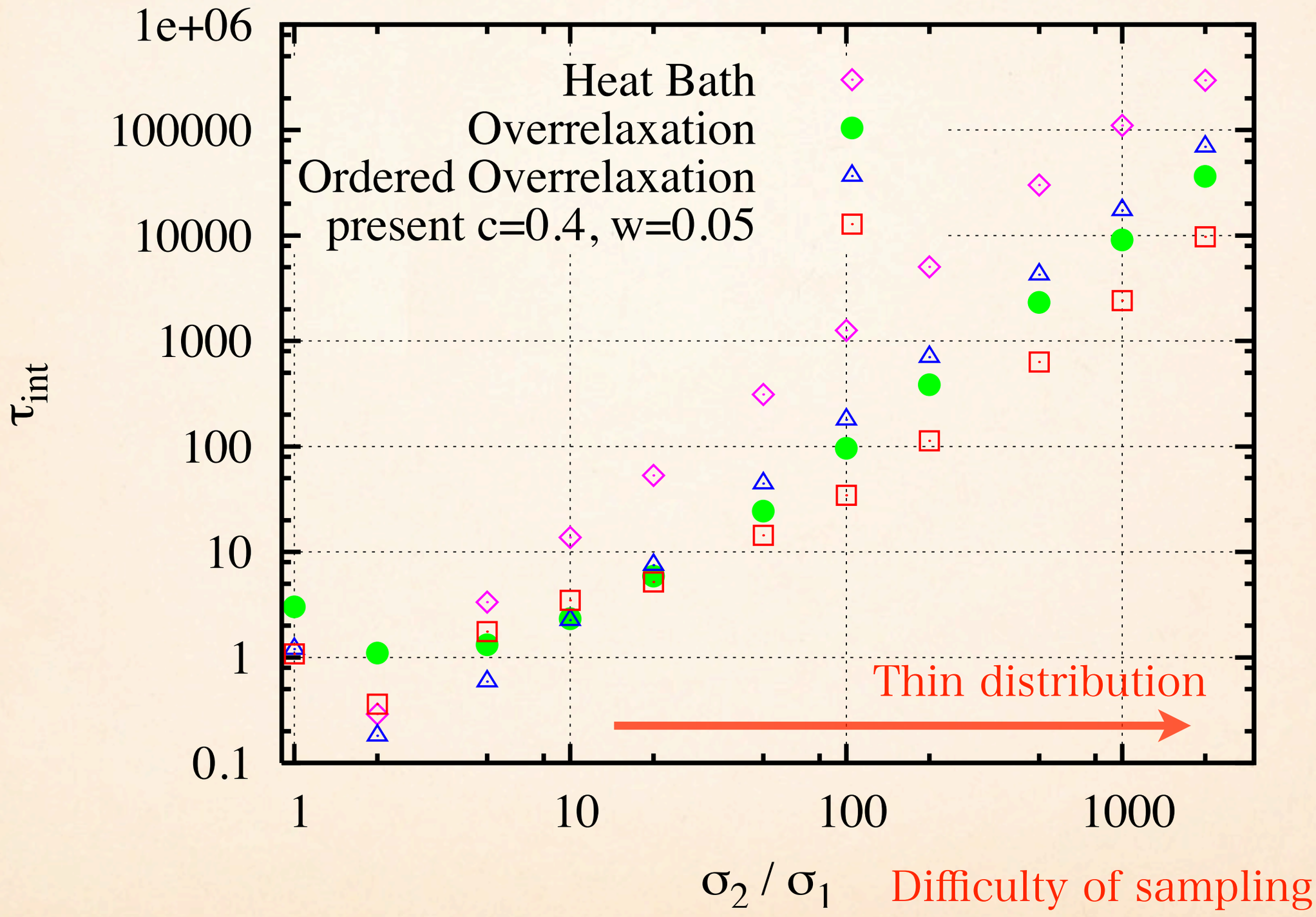


Periodic shift in
(cumulative) distribution.

Net stochastic flow boosts
up the convergence.

Comparison with Conventional Methods

Correlation times of $(x_1 + x_2)^2$ in the bivariate Gaussian dist.



Suppressing Random Walk Behavior

・Overrelaxation Adler (1981) Valid for Gaussian or simple distributions

Next configuration is chosen for "opposite" side in the Gaussian distribution.

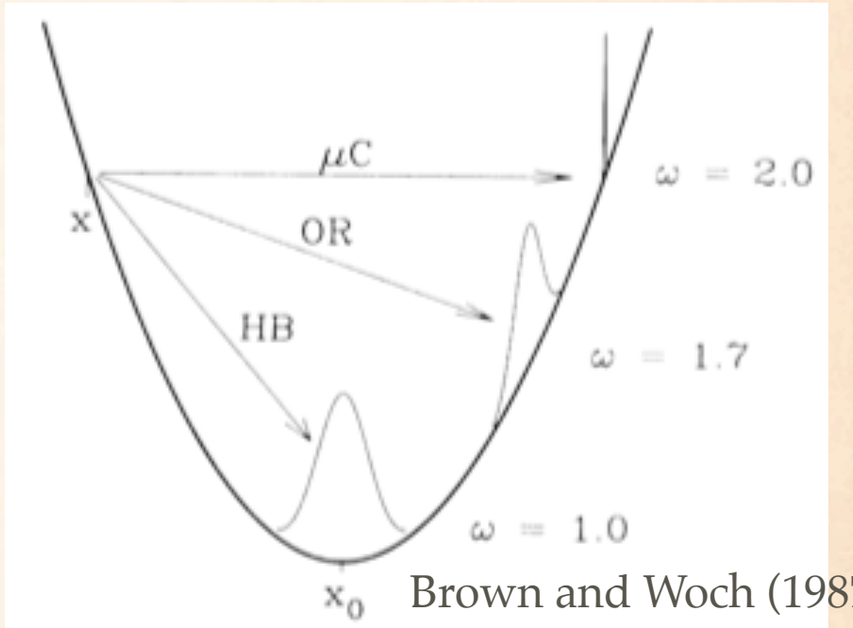
$$P(z_i|\cdot) \sim N(\mu_i, \sigma_i^2)$$

$$z_i' = \mu_i + \alpha(z_i - \mu_i) + \sigma_i\sqrt{1-\alpha_i^2}\nu$$

$\nu \sim N(0,1)$

$-1 < \alpha < 1$

$\alpha = 1$ no update
$\alpha = 0$ heat bath
$\alpha = -1$ microcanonical



Brown and Woch (1987)

・Hybrid Monte Carlo Duane et al. (1987)

1. Candidate configuration is chosen by the Newtonian dynamics and accepted by the Metropolis algorithm.

e.g. the leap-frog algorithm

$$x(t+\Delta t) = x(t) + \Delta t \left.\frac{\partial H}{\partial p}\right|_{t+\frac{\Delta t}{2}}$$

2. Momentums are updated by the heat bath algorithm.

These updates(1 and 2) are alternately repeated.

$$p\left(t+\frac{\Delta t}{2}\right) = p\left(t-\frac{\Delta t}{2}\right) - \Delta t \left.\frac{\partial H}{\partial x}\right|_{t}$$

Both of the overrelaxation and hybrid Monte Carlo indeed satisfy the DBC !

Beyond Metropolis

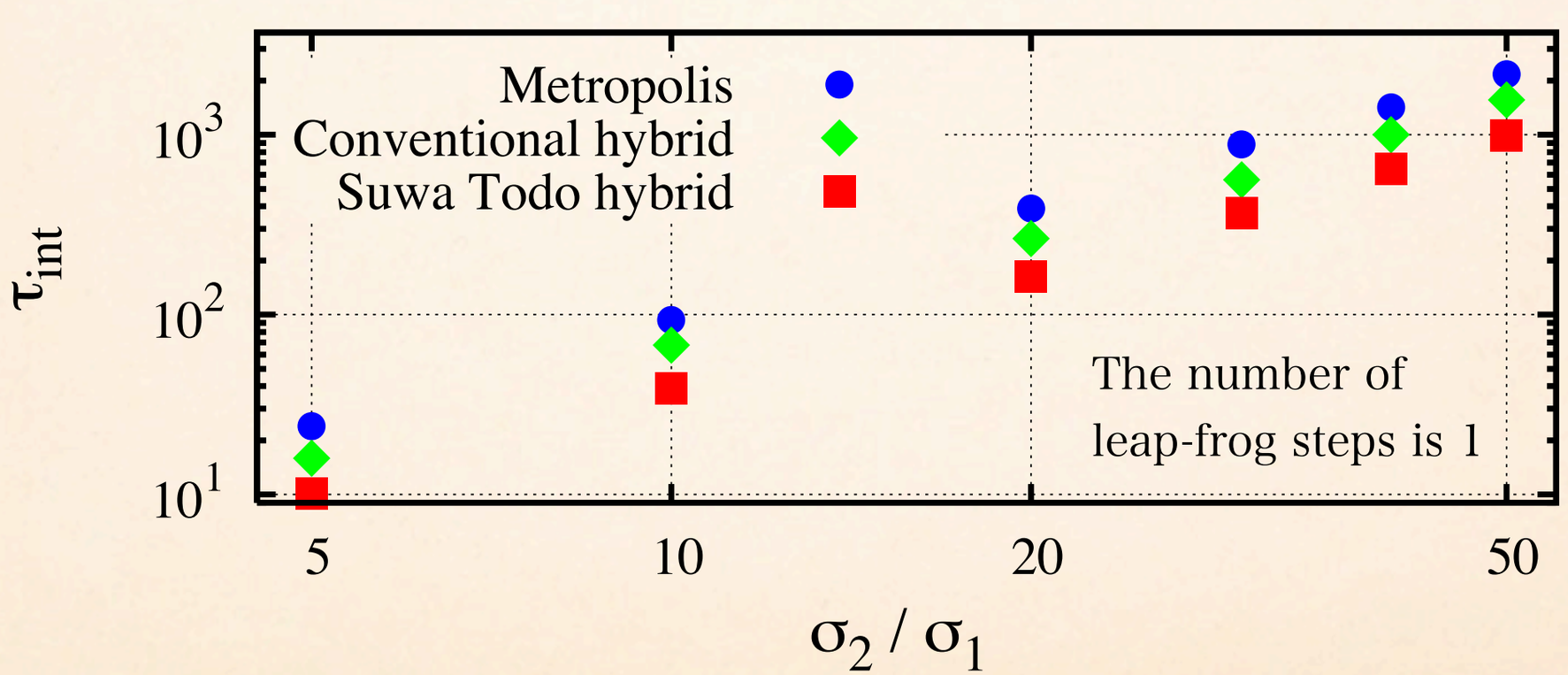
Rejection-free condition in our algorithms

$$w_1 \leq \sum_{i=2}^{n} w_i$$

→ The more candidate configurations, the smaller rejection rate !

We used our algorithm 1 (irreversible kernel) after making 3 candidates in Newtonian dynamics of the hybrid Monte Carlo, instead of conventional 2 candidates.

Correlation times of $(x_1+x_2)^2$ in the bivariate Gaussian distribution



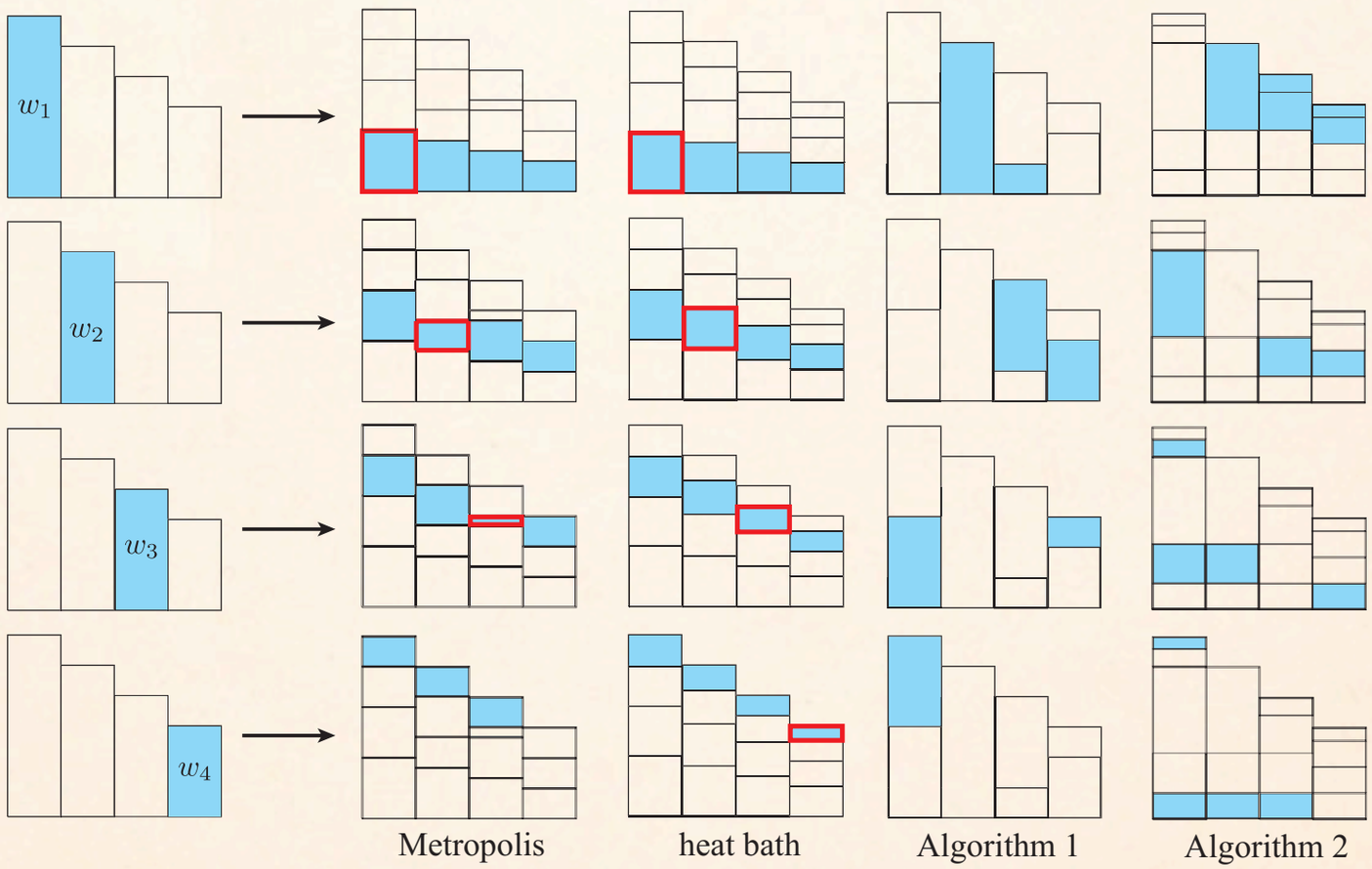
CPU time increases ? → Parallelization for making candidates !

cf. H.S. and Todo,
Phys. Rev. Lett., 105, 120603 (2010)

Summary

- We have developed new algorithms that construct the transition kernel, based on the geometric weight allocation approach.

- Our algorithms, which are new types of optimization method, always minimize the average rejection rate.



The present methods are applicable to almost all Markov chain Monte Carlo methods (also combinable with cluster and extended ensemble methods) and replace the conventional methods, e.g., the Metropolis algorithm and Gibbs sampler.