Lecture 6

Thermal Motion, Annealing and Monte Carlo Schemes

Bao-Jun Cai, 4/8/2026

Introduction to Algorithms for Data Science and Physics IMP@Fudan, 2026

Topics of this lecture:

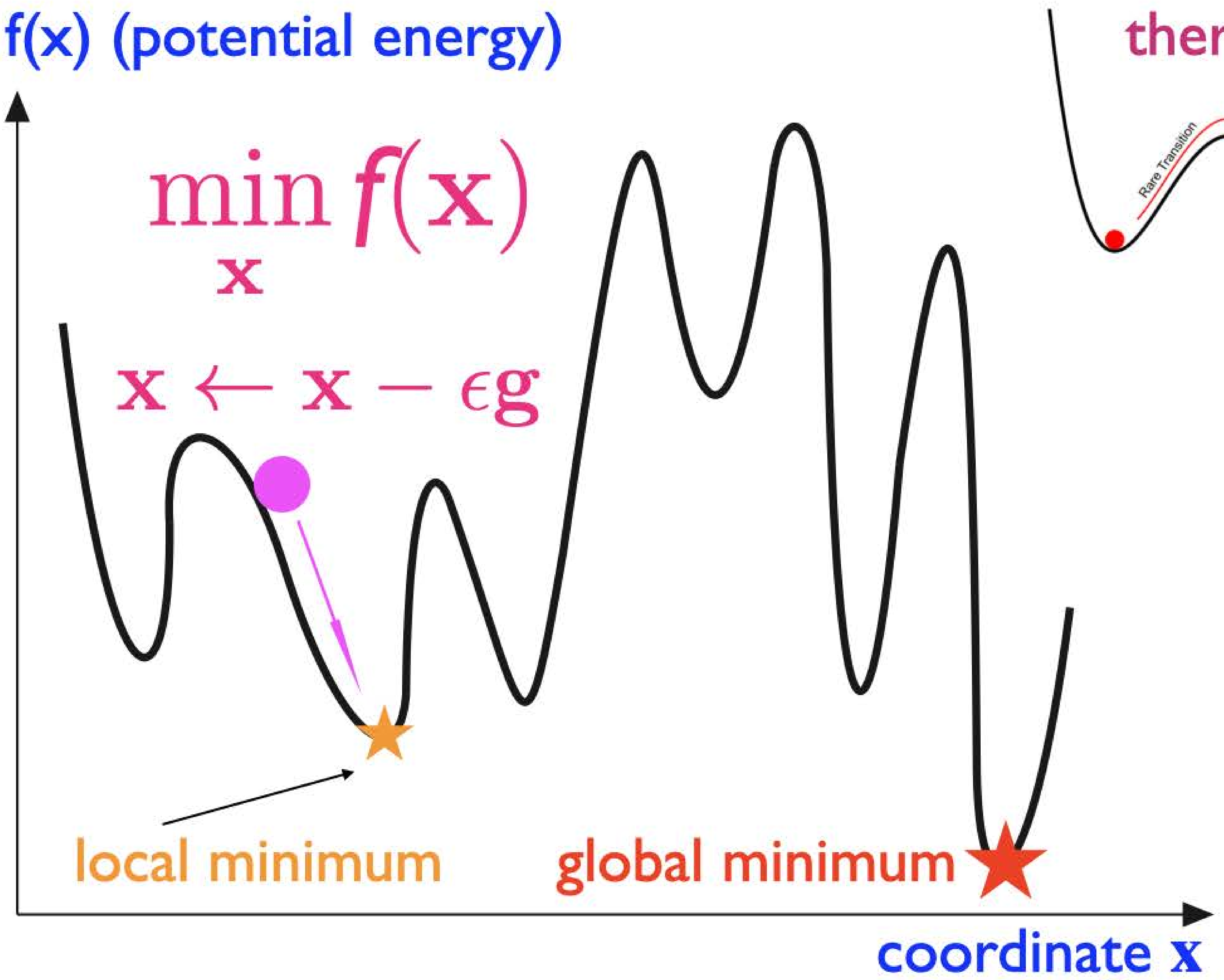
- thermal motion, kinetic energy $3k_BT \sim m\vec{u}^2/2$
- simulated annealing $T_0 \to T_i/T_0 < 1$
- Metropolis algorithm prob $\sim \min[1, e^{-\beta\Delta E}]$
- 2D Ising model by simulation $E \sim -J\sum_{\langle ij\rangle} s_is_j$
- inverse, rejection and importance
- variance reduction $\sigma_M \sim 1/\sqrt{M}$

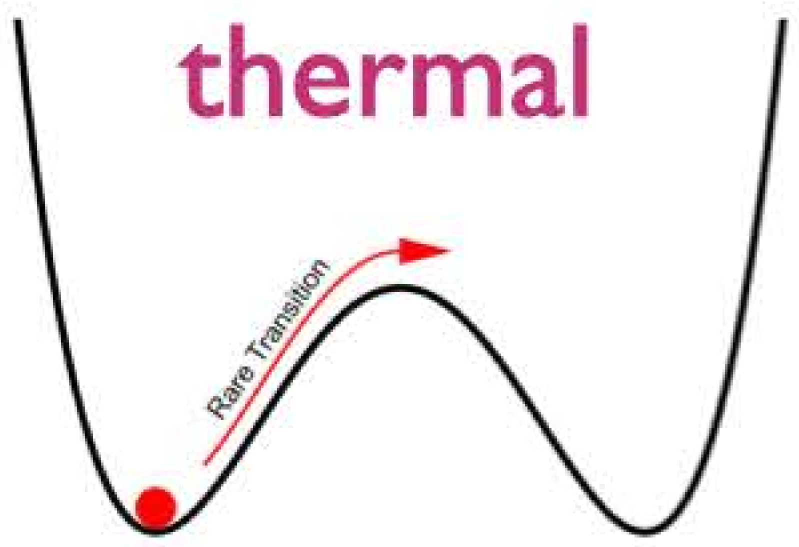
$$q(\mathbf{x}) \to p(\mathbf{x})$$

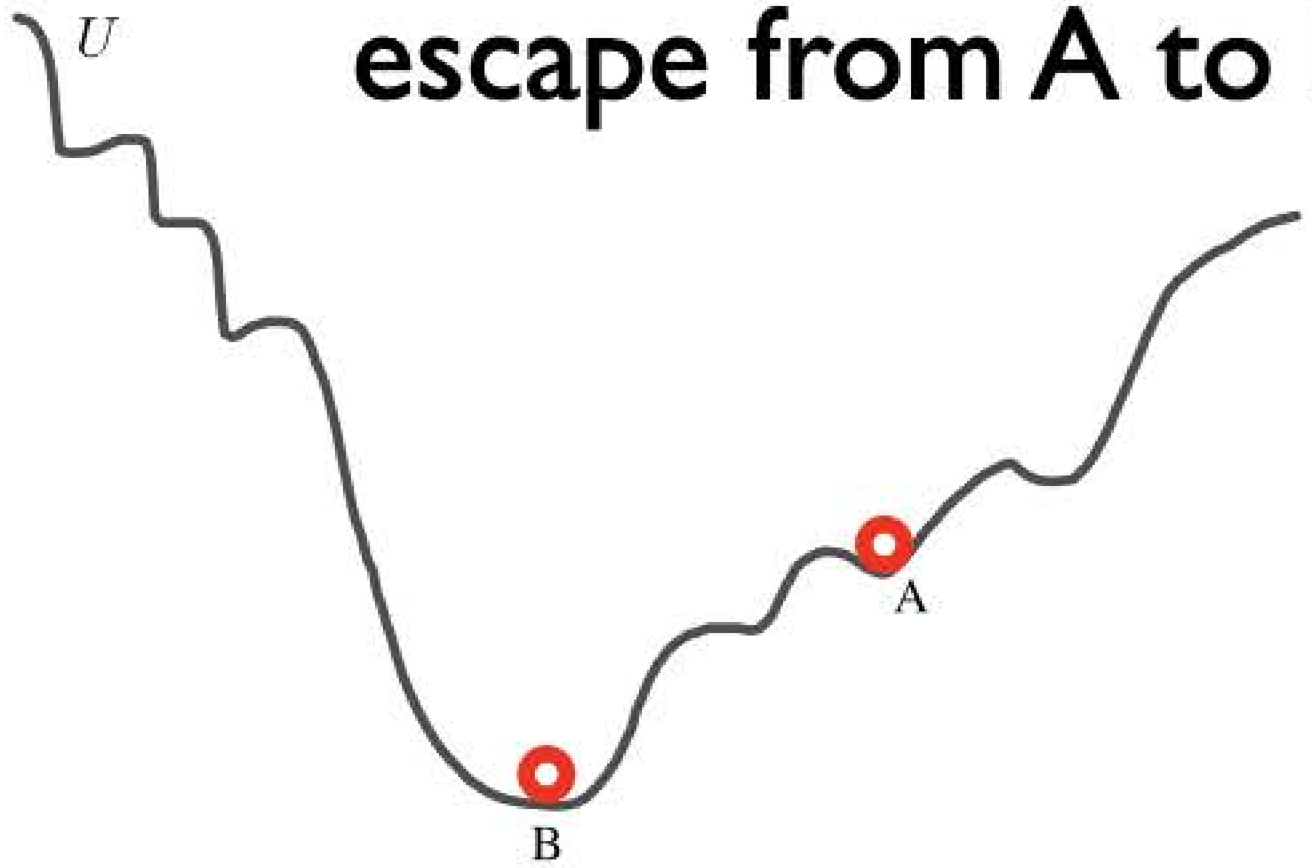
$$\mathrm{E}[f] = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

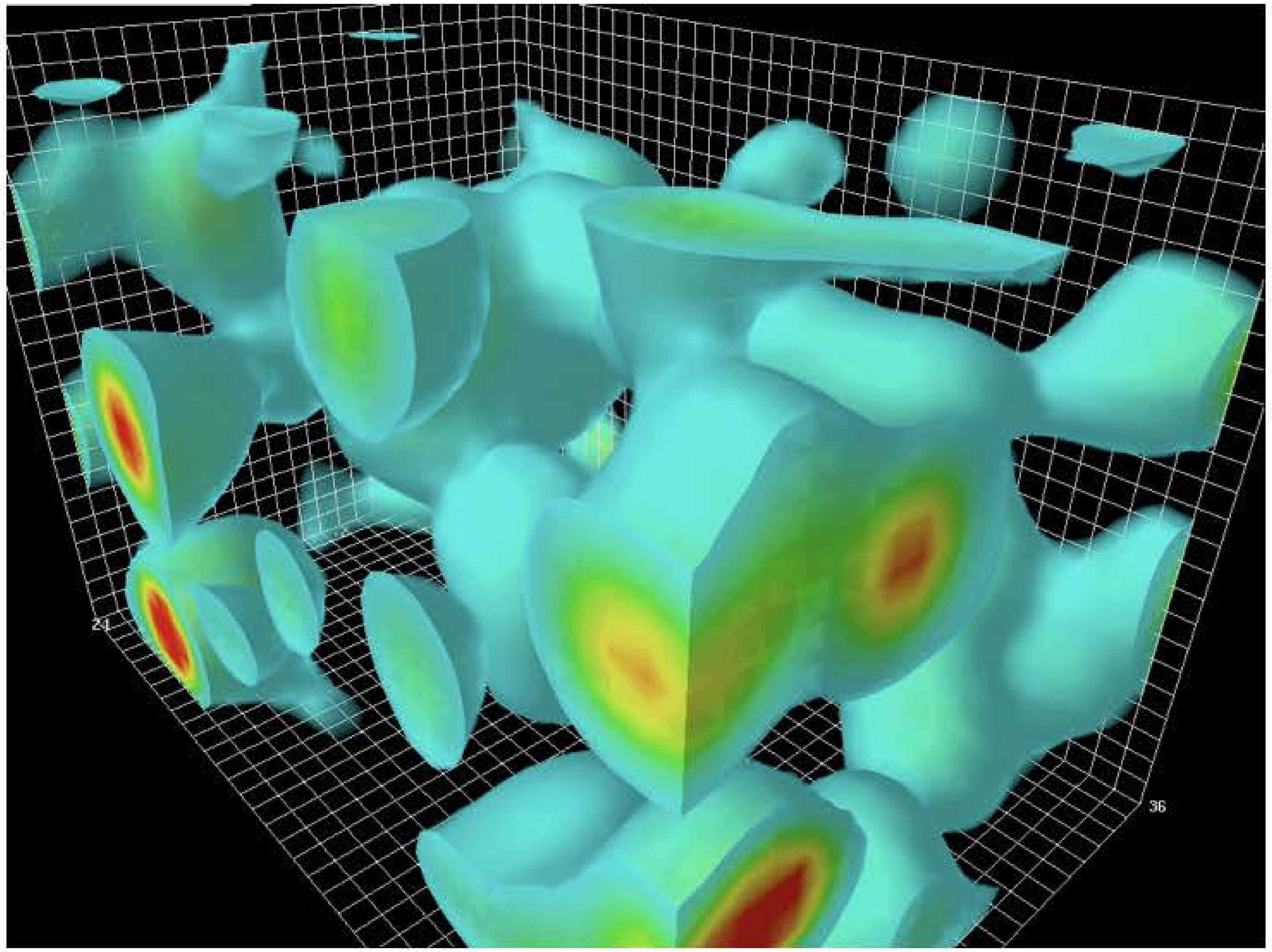
Trapped in a local minimum

How can the particle escape from A to B?



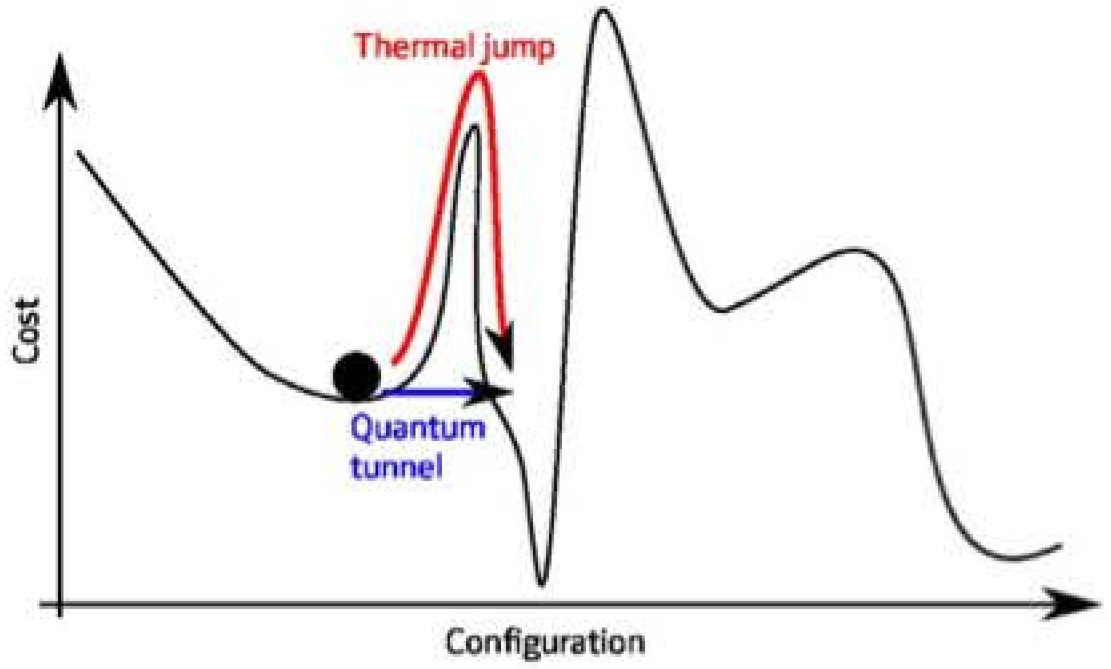






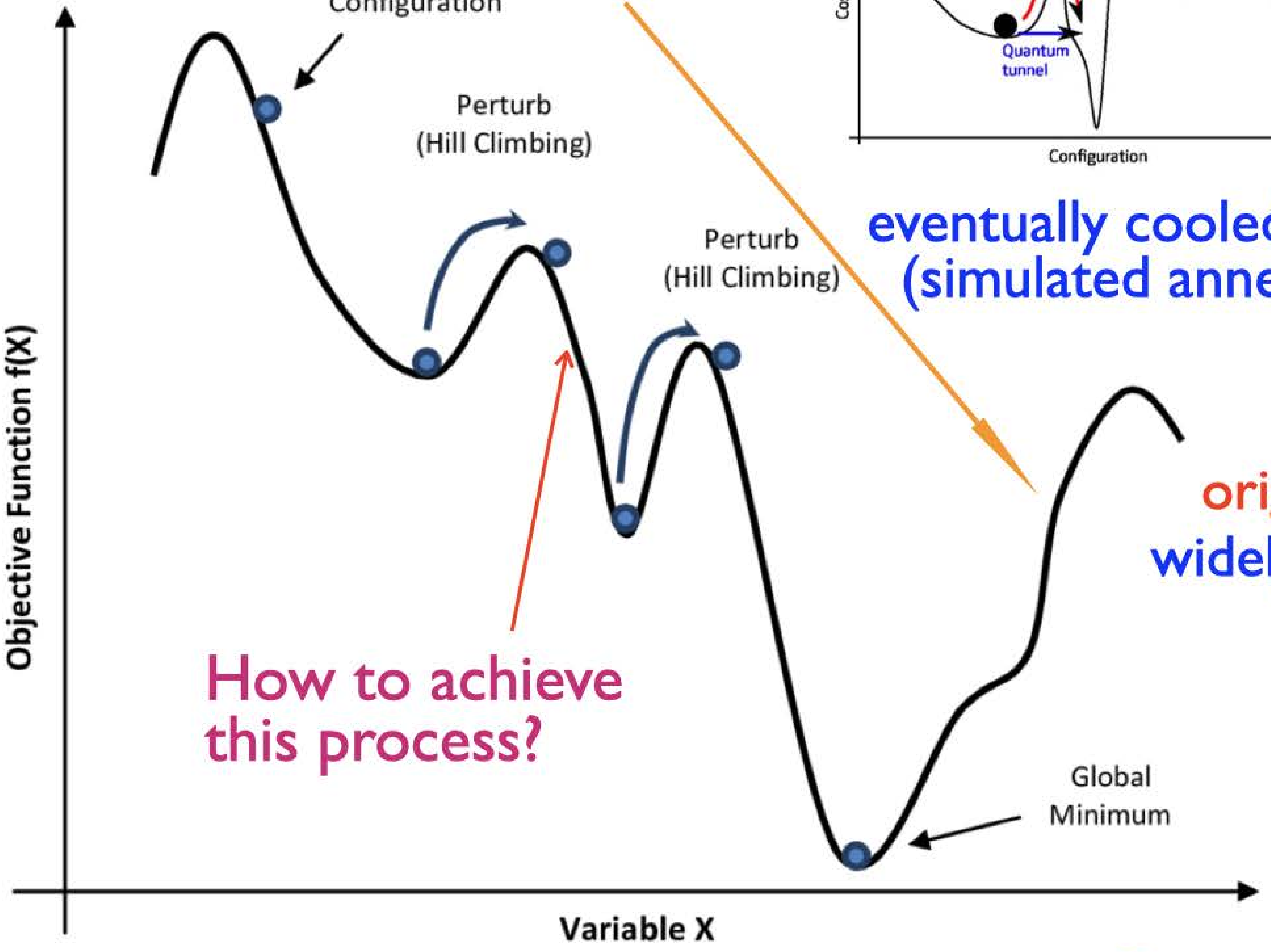
Simulated annealing

quantum coherence

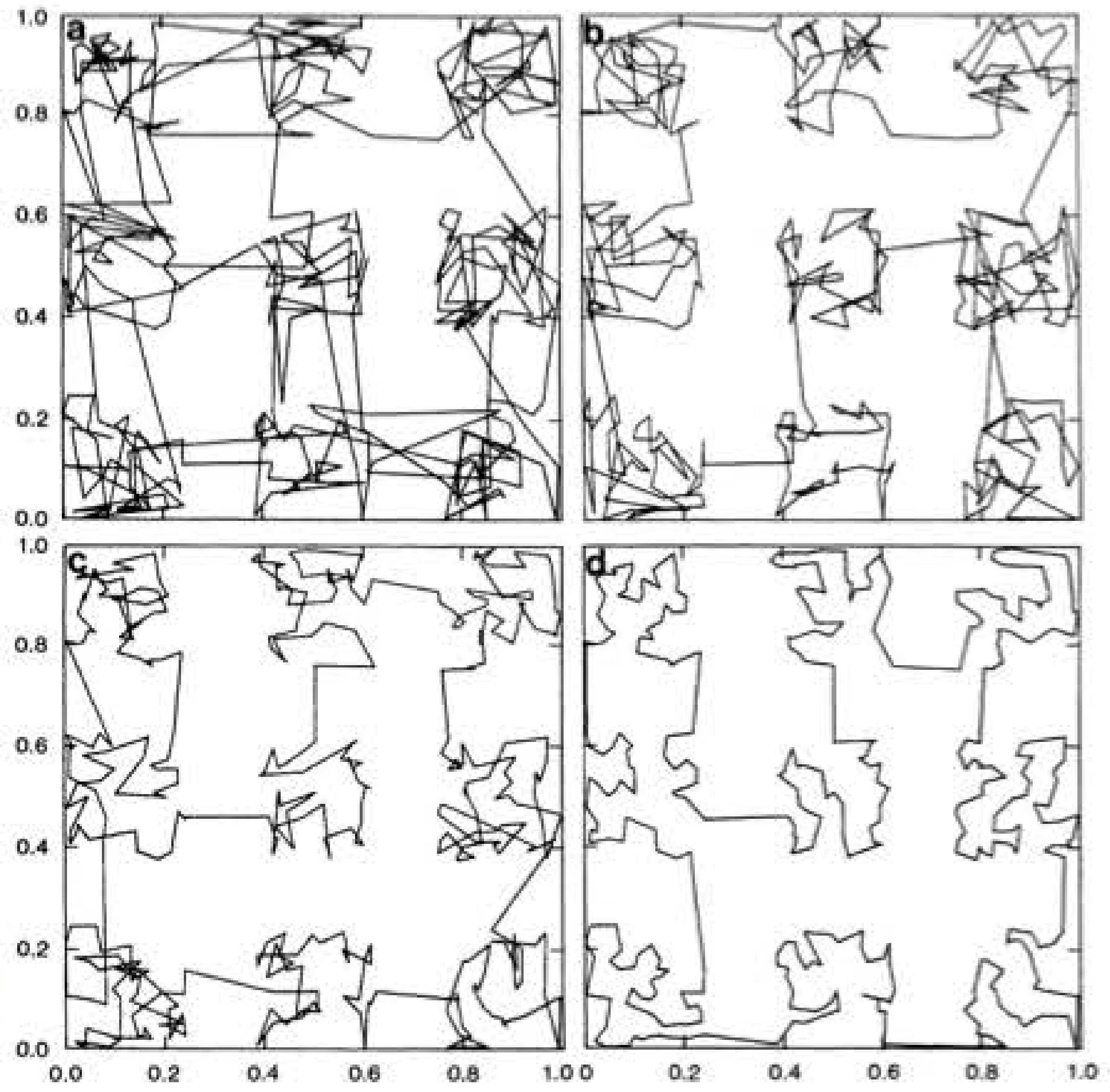


initial T_0

eventually cooled down (simulated annealing)



How to achieve this process?



Traveling Salesmen

Quantitative analysis of the simulated annealing algorithm or comparison between it and other heuristics requires problems simpler than physical design of computers. There is an extensive literature on algorithms for the traveling salesman problem (3, 4), so it provides a natural context for this discussion.

If the cost of travel between two cities is proportional to the distance between them, then each instance of a traveling salesman problem is simply a list of the positions of N cities. For example, an arrangement of N points positioned at random in a square generates one instance. The distance can be calculated in either the Euclidean metric or a "Manhattan" metric, in which the distance between two points is the sum of their separations along the two coordinate axes. The latter is appropriate for physical design applications, and easier to compute, so we will adopt it.

We let the side of the square have length $N^{1/2}$, so that the average distance between each city and its nearest neighbor is independent of N. It can be shown

13 May 1983, Volume 220, Number 4598 **SCIENCE**

originated from statistical physics

widely used in optimization problems

with N, so that in practice exact solutions can be attempted only on problems involving a few hundred cities or less. The traveling salesman belongs to the large class of NP-complete (nondeterministic polynomial time complete) problems, which has received extensive study in the past 10 years (3). No method for exact solution with a computing effort bounded by a power of N has been found for any of these problems, but if such a solution were found, it could be mapped into a procedure for solving all members of the class. It is not known what features of the individual problems in the NP-complete class are the cause of their difficulty.

Since the NP-complete class of problems contains many situations of practical interest, heuristic methods have been developed with computational require-

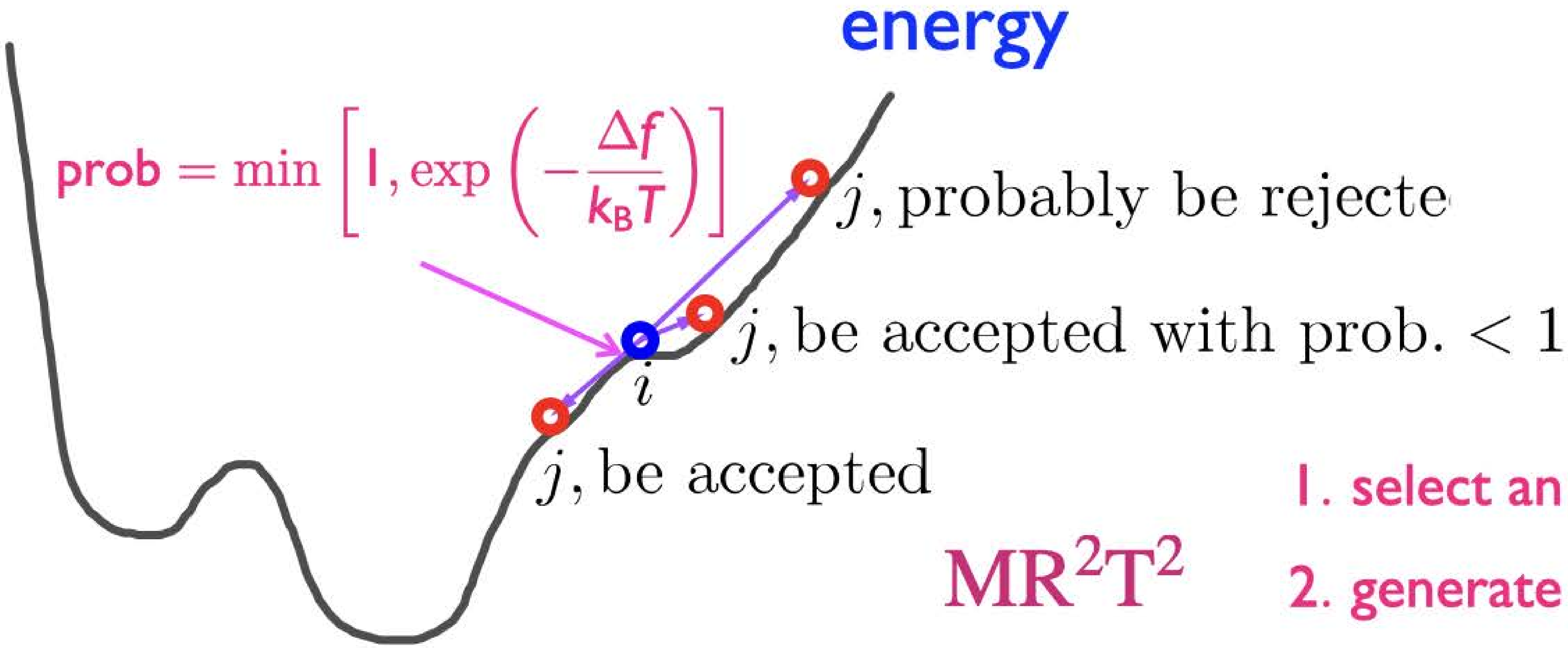
Optimization by Simulated Annealing

S. Kirkpatrick, C. D. Gelatt, Jr., M. P. Vecchi

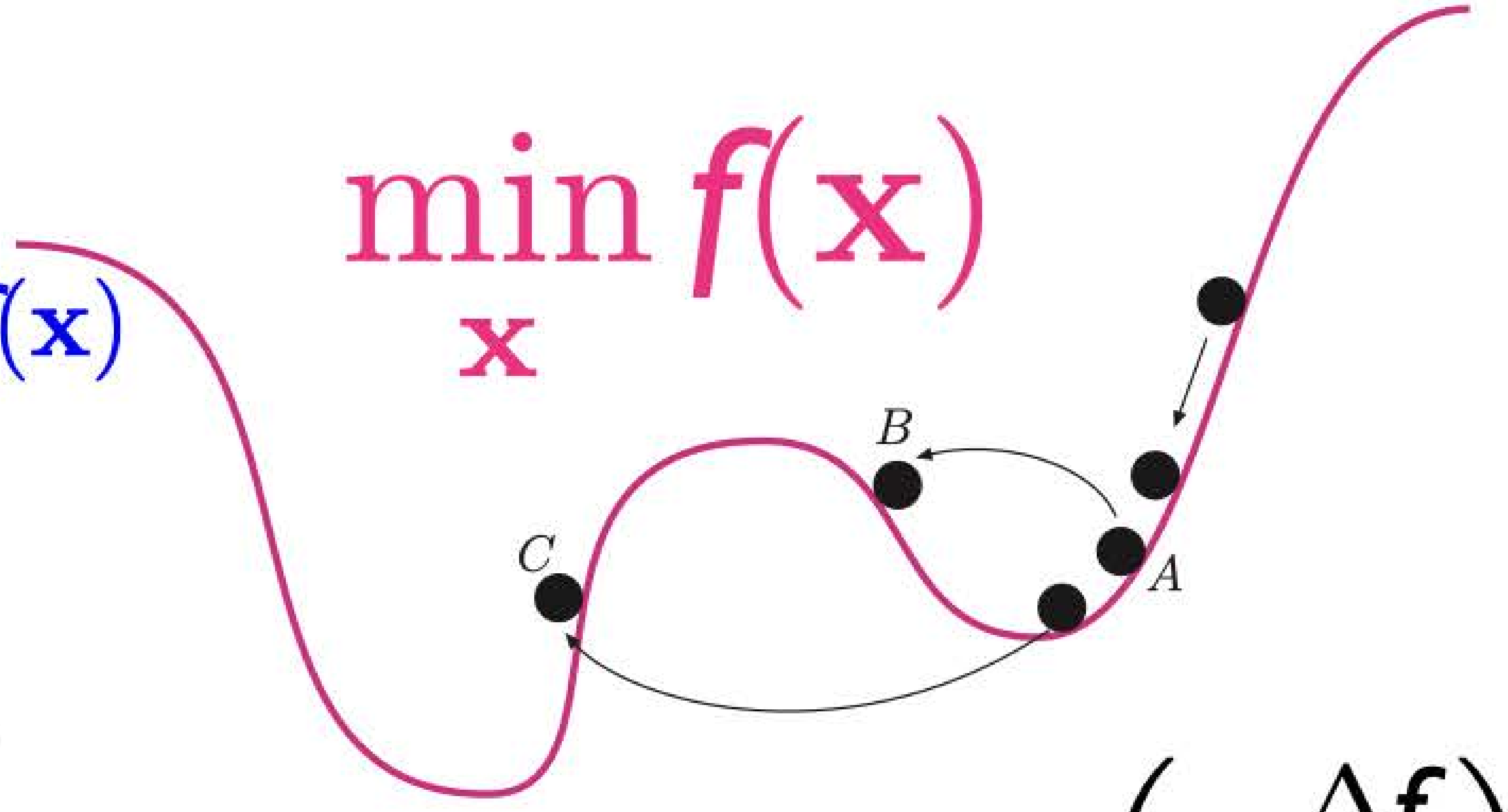
In this article we briefly review the central constructs in combinatorial optimization and in statistical mechanics and then develop the similarities between the two fields. We show how the Metropolis algorithm for approximate numerical simulation of the behavior of a many- sure of the "goodness" of some complex system. The cost function depends on the detailed configuration of the many parts of that system. We are most familiar with optimization problems occurring in the physical design of computers, so examples used below are drawn from

Metropolis algorithm

$$\mathbf{x} \to \mathbf{x} + \Delta\mathbf{x},\ f(\mathbf{x}) \to f(\mathbf{x} + \Delta\mathbf{x}),\ \Delta f = f(\mathbf{x} + \Delta\mathbf{x}) - f(\mathbf{x})$$



$$\min_{\mathbf{x}} f(\mathbf{x})$$



$$\text{prob} = \exp\left(-\frac{\Delta f}{k_B T}\right)$$

$\mathrm{MR^2T^2}$

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

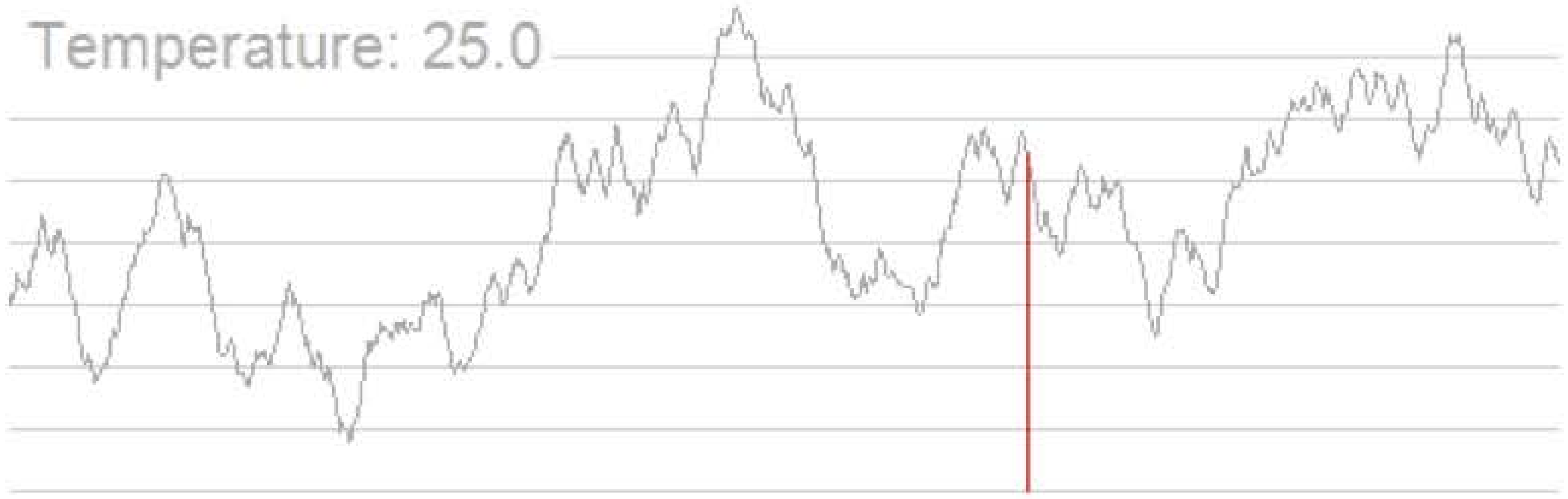
A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

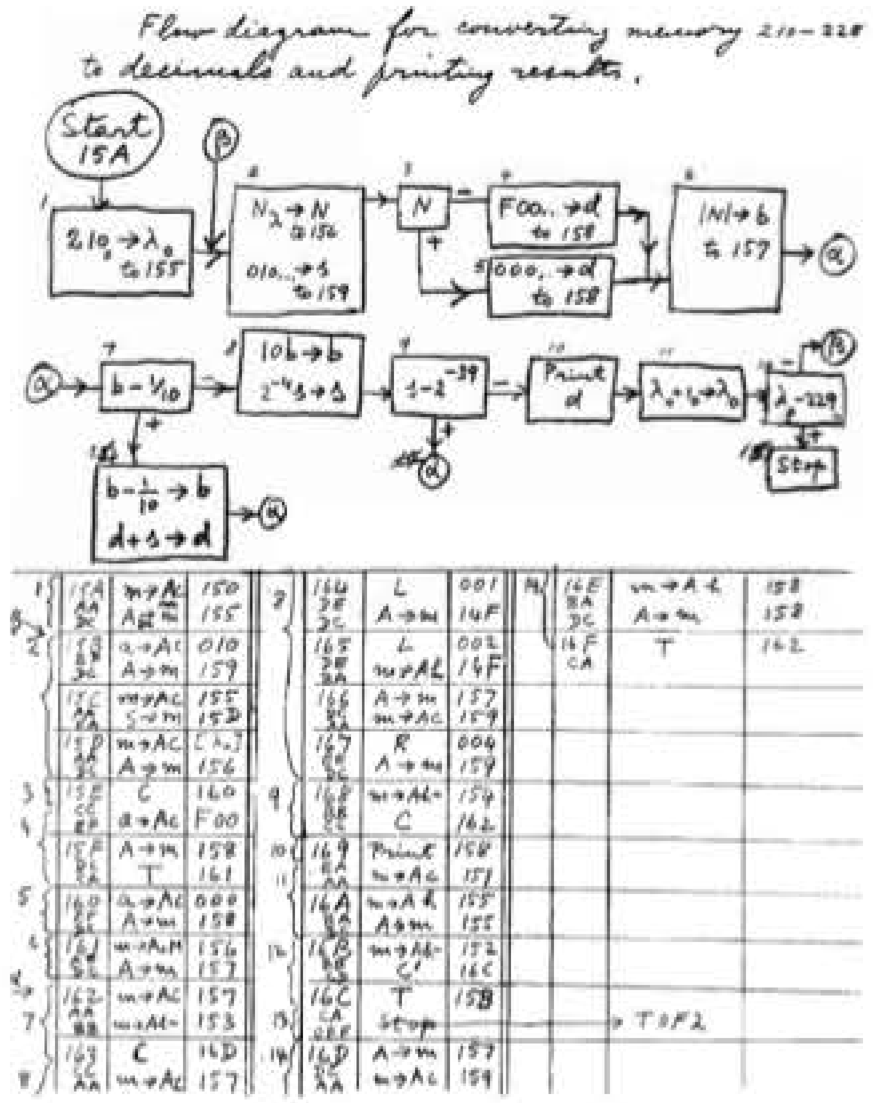
1. select an initial position $\mathbf{x}$
2. generate $\mathbf{x} \to \mathbf{x} + \Delta\mathbf{x}$
3. compute $\Delta f = f(\mathbf{x} + \Delta\mathbf{x}) - f(\mathbf{x})$
4. if $\Delta f < 0$, accept the new state without doubt

   if $\Delta f \geq 0$, accept with some probability < 1
5. cool down the temperature (annealing pattern)

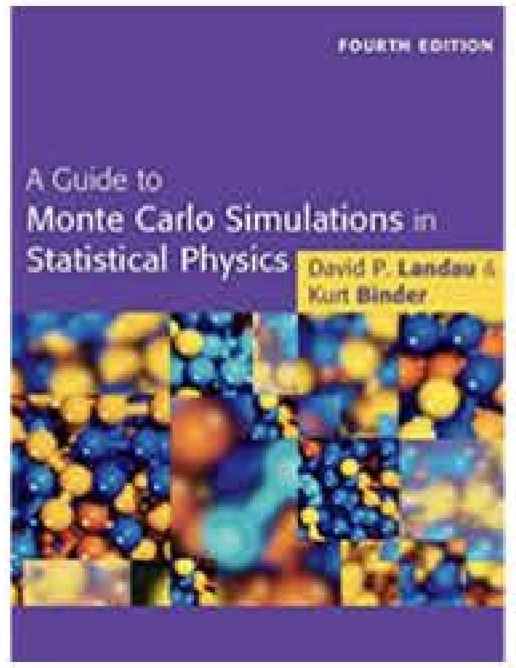
$u \sim \text{Unif}(0, 1)$

Monte Carlo, MANIAC, . . .

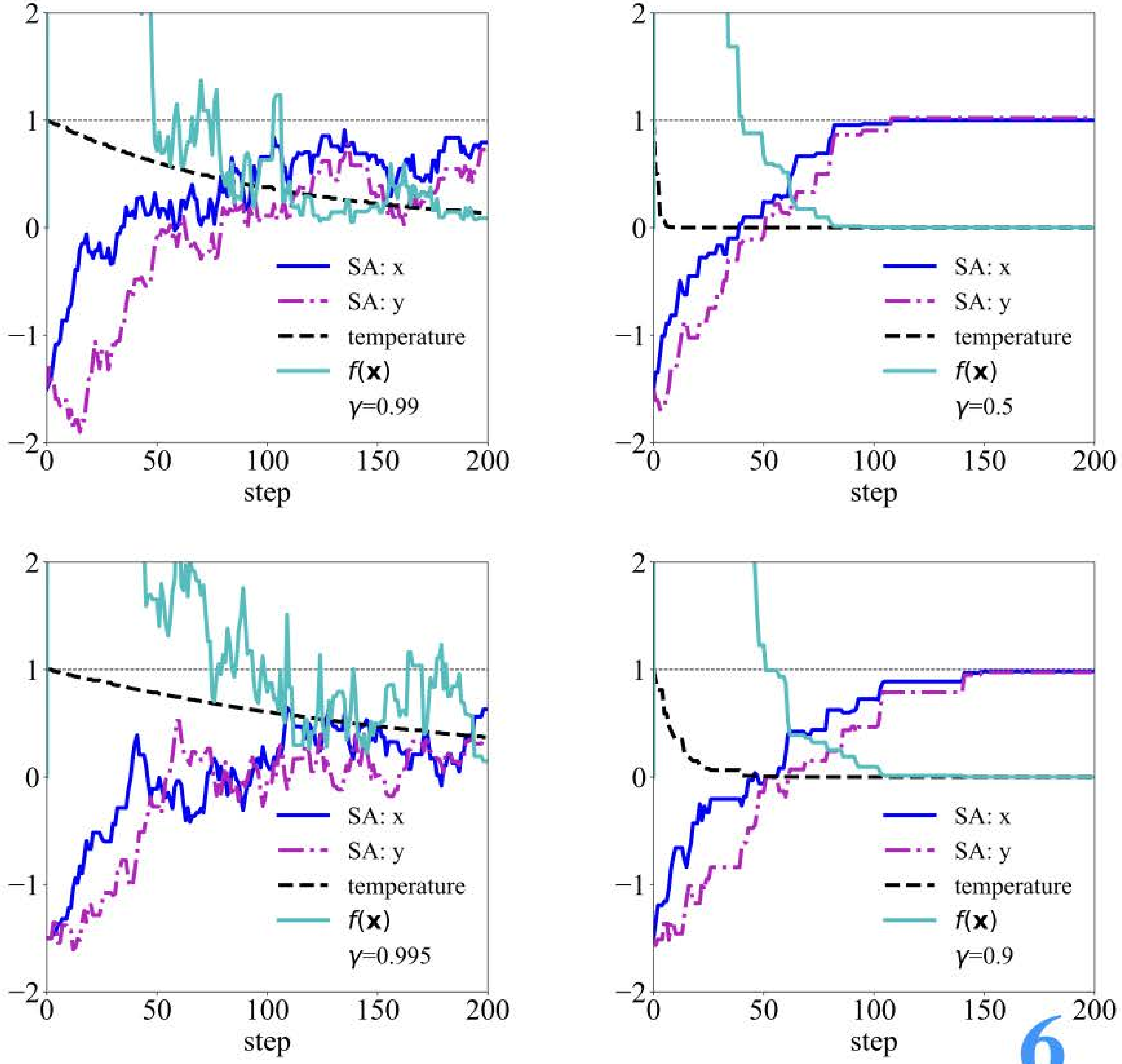


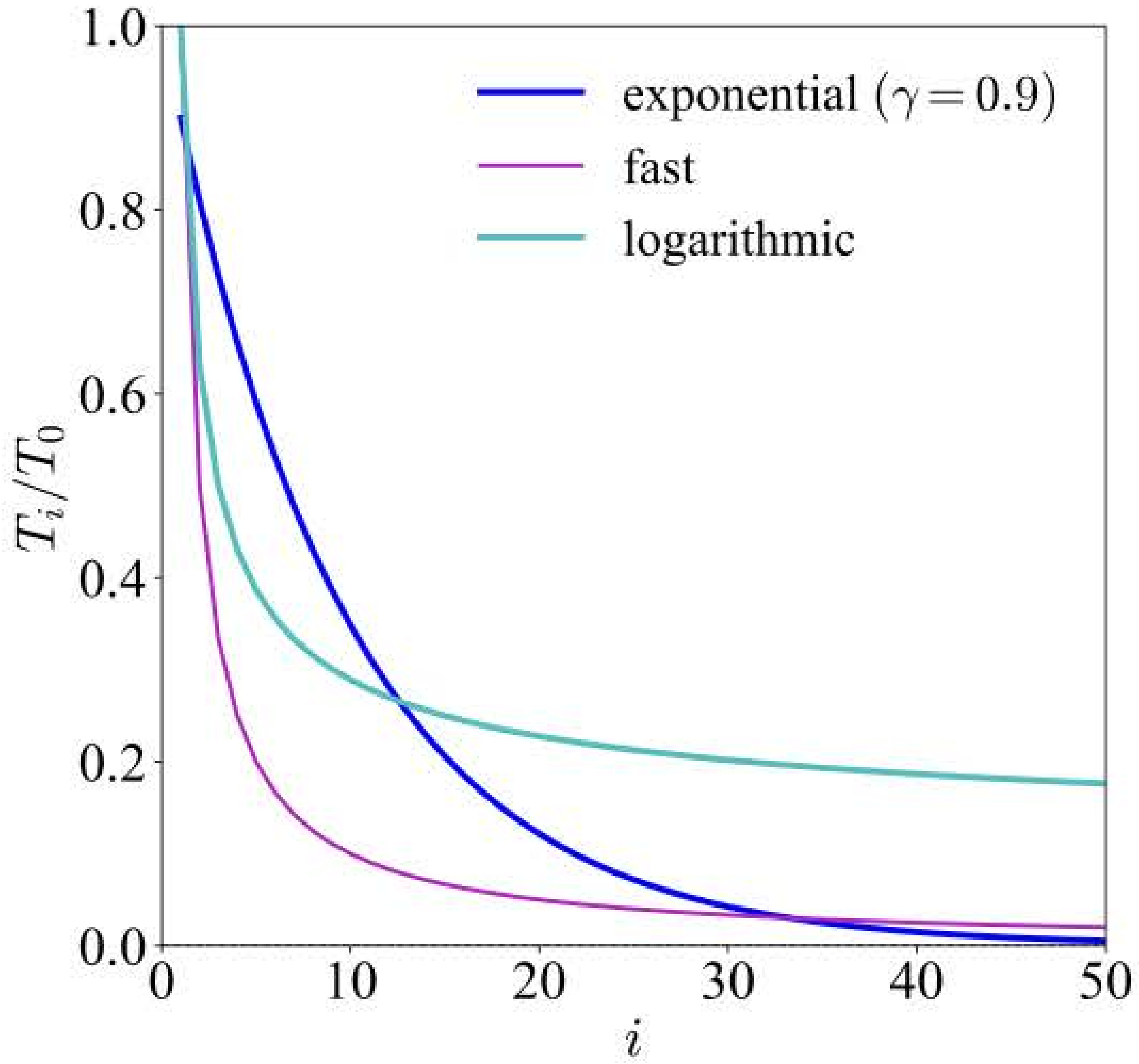



C. N. Yang and Edward Teller (1992) 13



Annealing pattern, Rosenbrock function





$$f(\mathbf{x}) = (x - a)^2 + b(x^2 - y)^2$$

fast: $T_i = T_0/i$

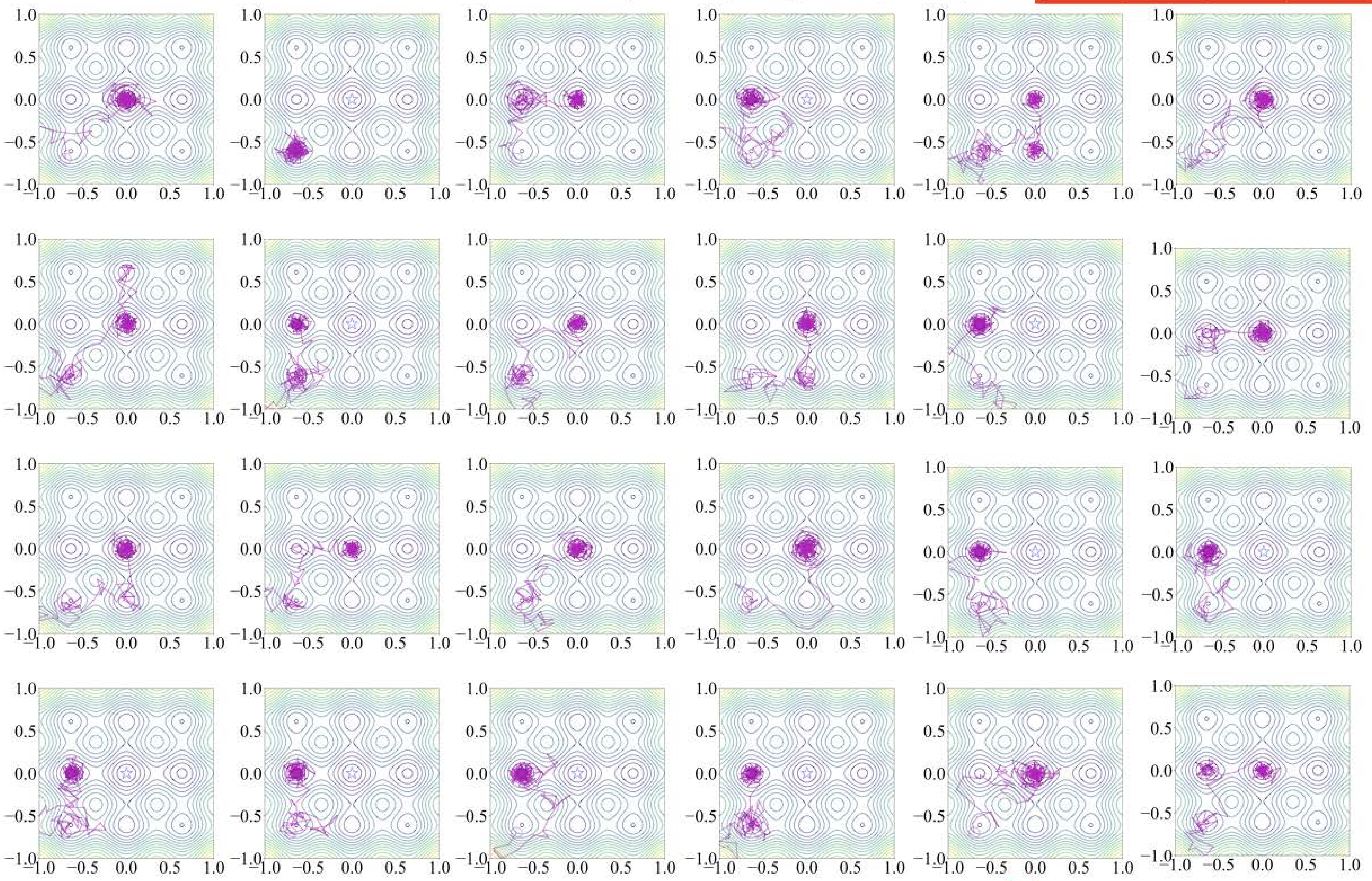
exponential: $T_i = T_i\gamma^i$

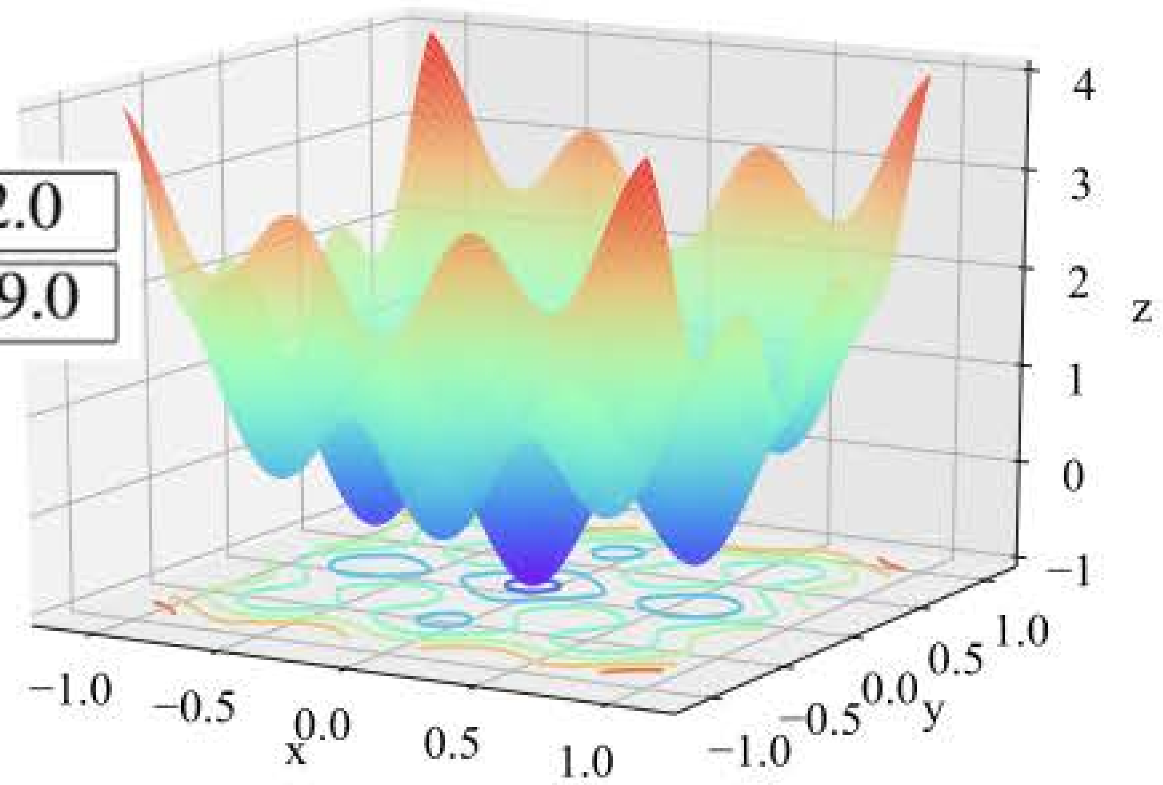
logarithmic: $T_i = T_0 \ln 2/ln(1 + i)$

Multiple local minima

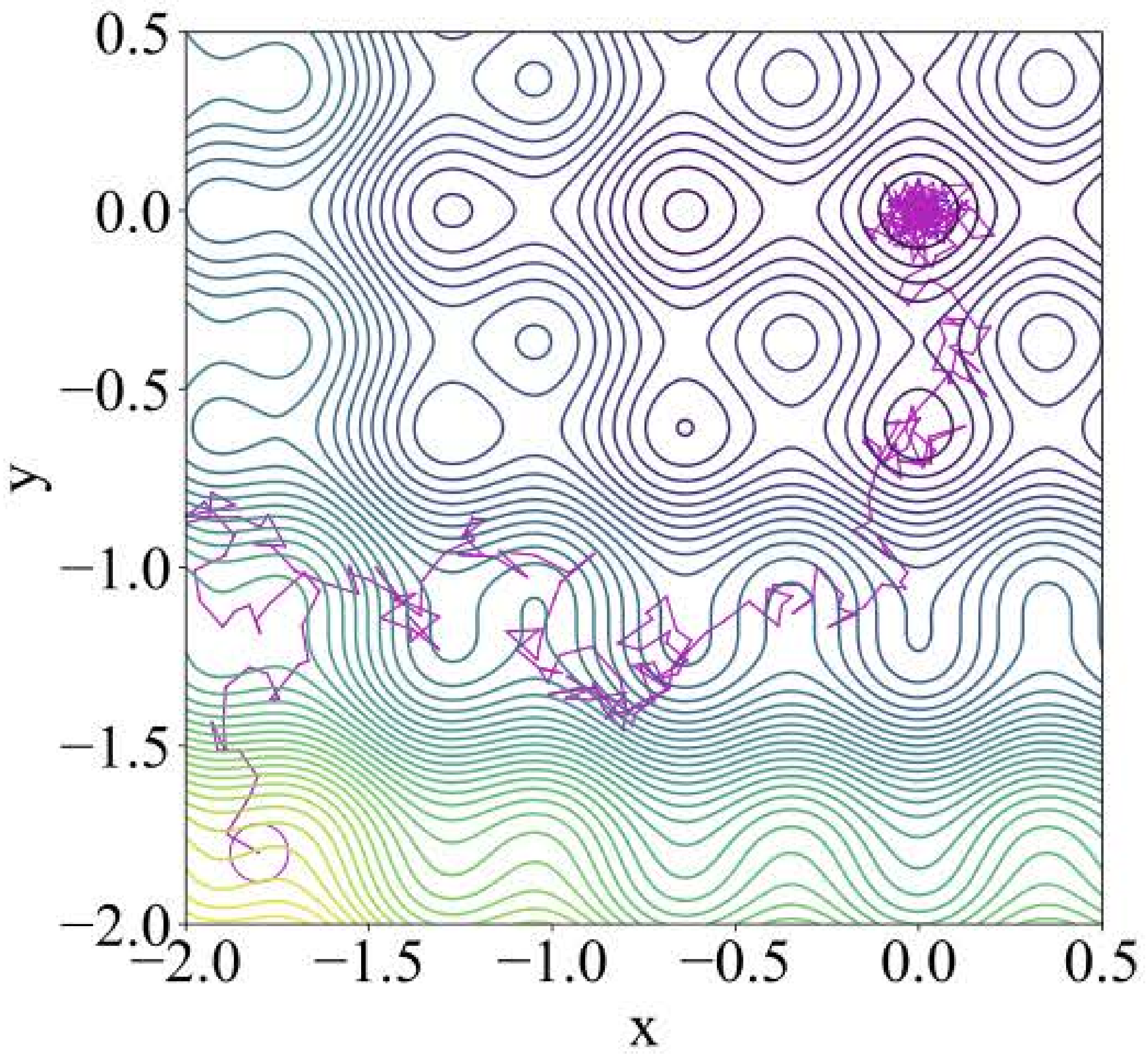
successful probability

d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	2.0
$P\,(\%)$	1.1	10.5	33.5	69.4	98.8	99.6	99.6	99.2	99.6	98.2	89.0





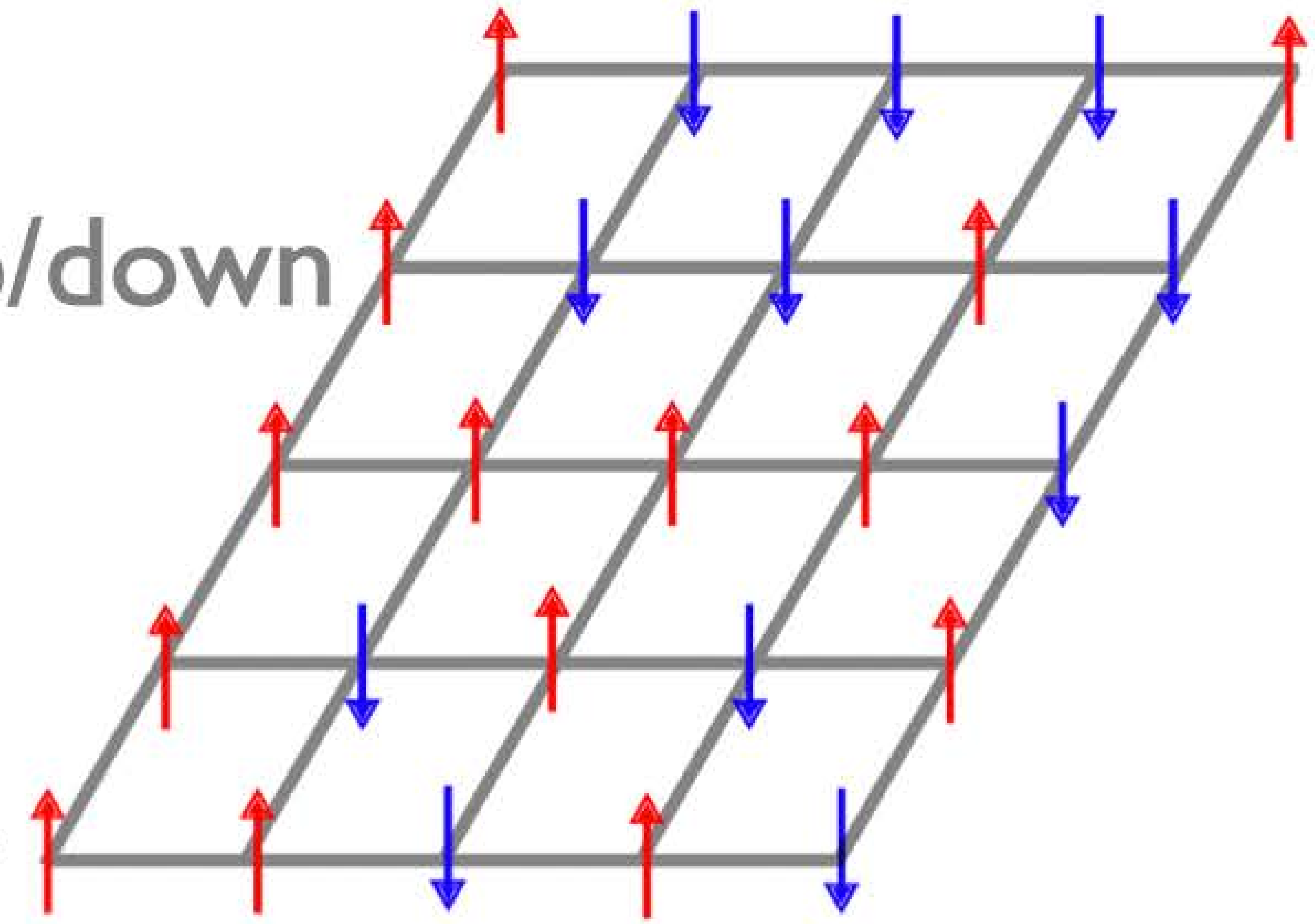
$$f(\mathbf{x}) = x^2 + 2y^2 + a[\cos(3\pi x) + \cos(3\pi y)]$$



fast annealing

2D Ising model: simulation

$s_i = \pm 1$, up/down



$$E = -J \sum_{\langle ij \rangle} s_i s_j, \; p(E_\mu) \sim e^{-\beta E_\mu}, \; \beta = \frac{1}{k_B T}$$

Boltzmann factor

size $= n = L^2$

internal energy

specific heat

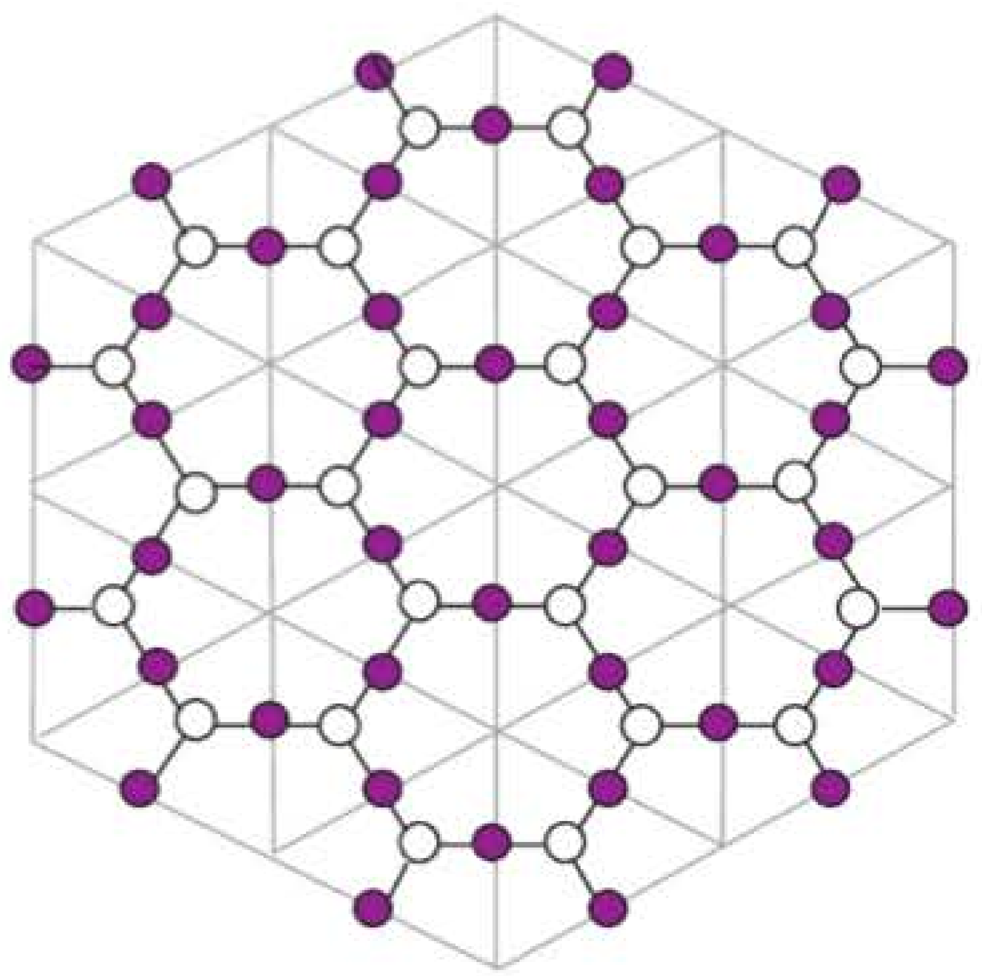
$$U = \langle E \rangle = \frac{\sum_\mu E_\mu e^{-\beta E_\mu}}{\sum_\mu e^{-\beta E_\mu}}, \; C = \frac{1}{n} \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2} = \frac{\sigma_E^2}{n k_B T^2}$$

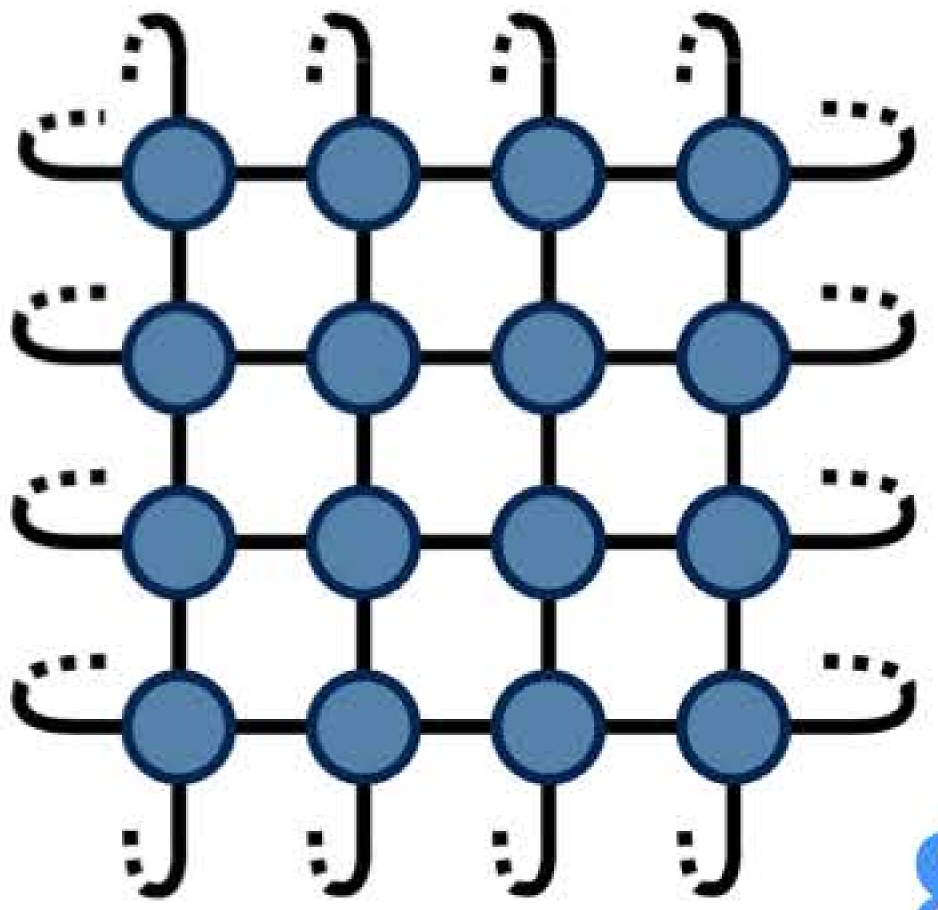
$$\text{prob} = e^{-\Delta E / k_B T}$$

Metropolis simulation:

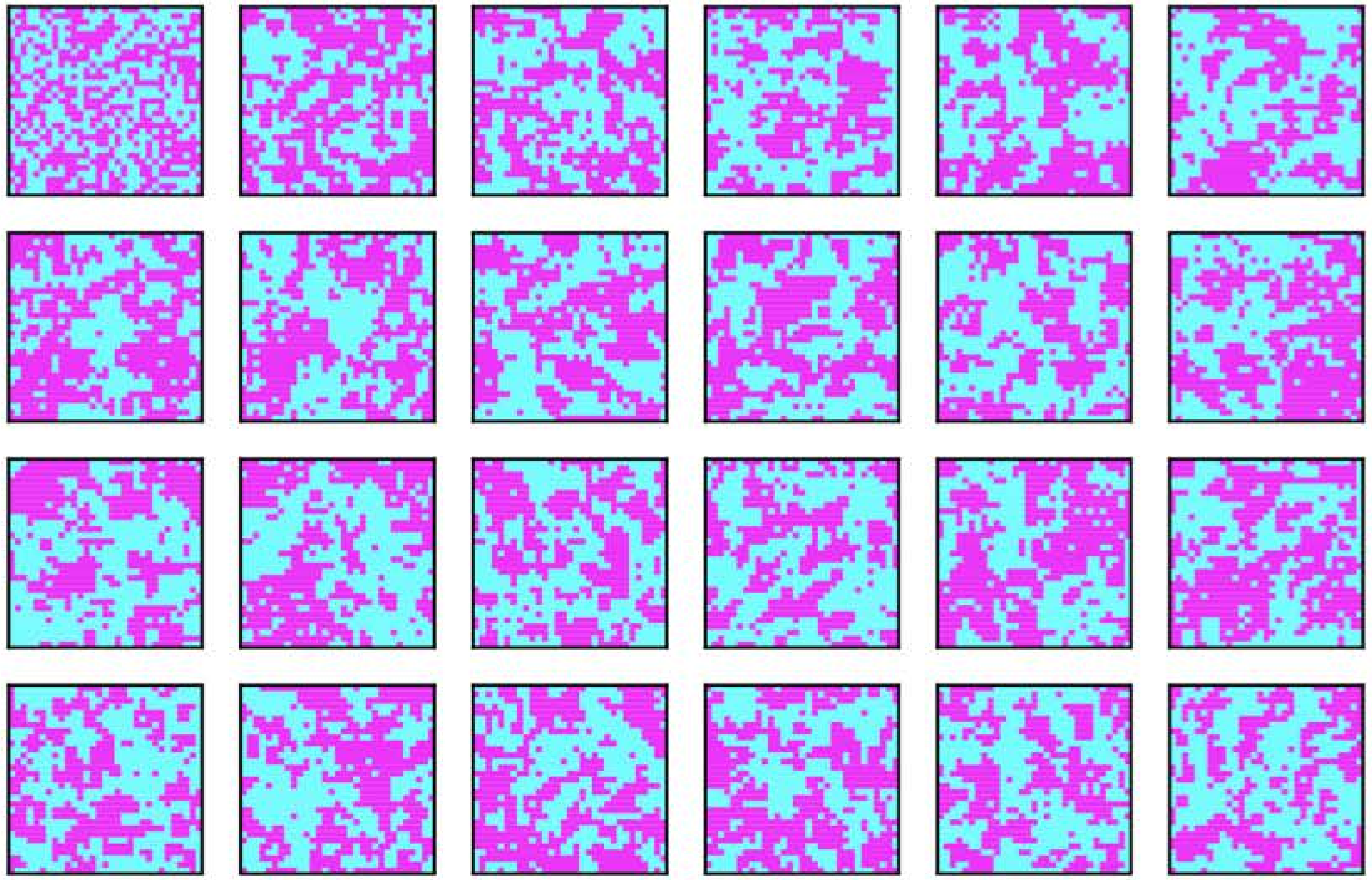
1. select an initial configuration μ, calculate E_μ
2. randomly flip a spin and obtain $E_{\mu'}$
3. calculate the energy difference $\Delta E = E_{\mu'} - E_\mu$
4. if $\Delta E < 0$, accept the new state without doubt

   if $\Delta E \geq 0$, accept with some probability < 1
5. continue to be stationary



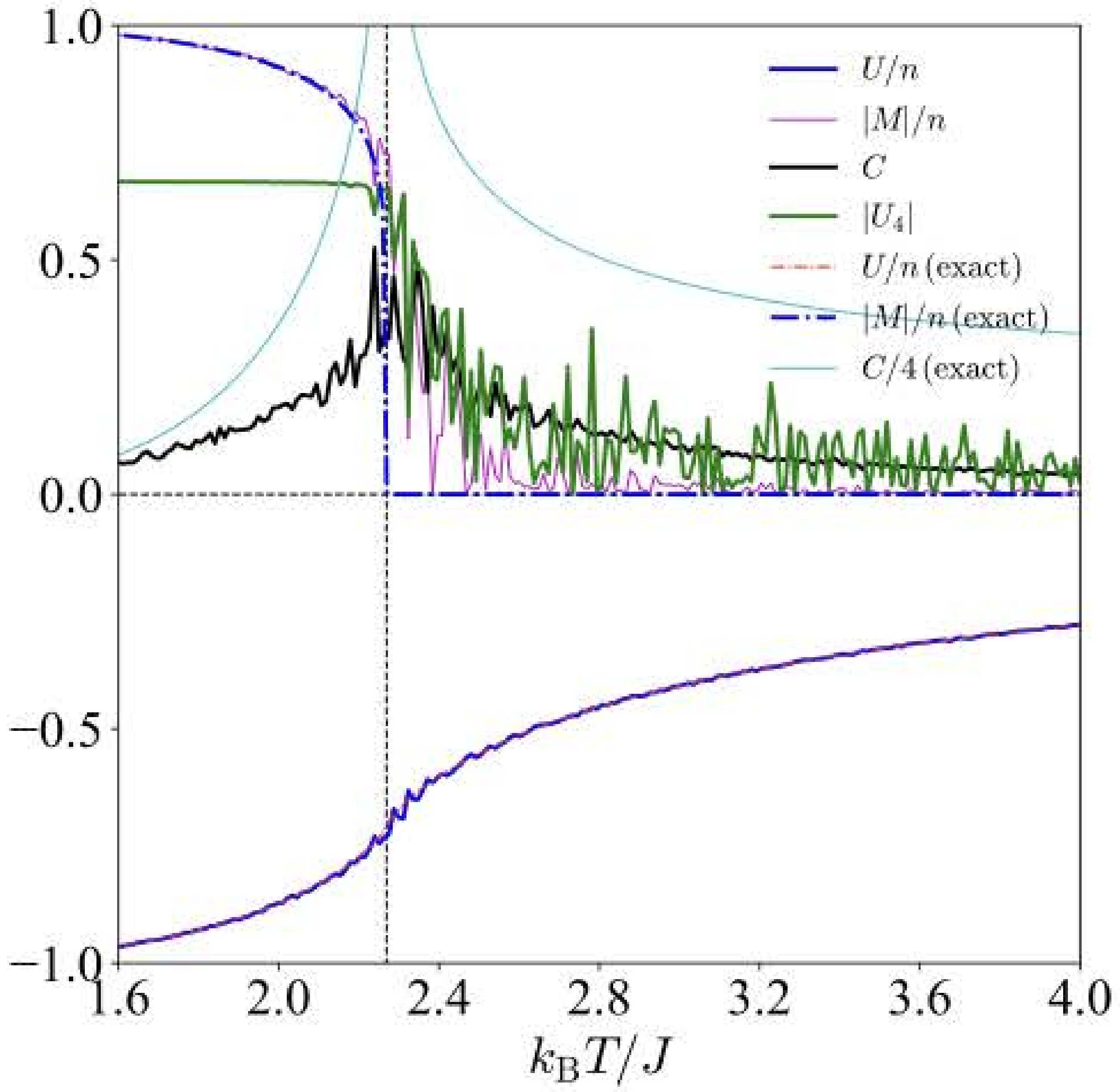


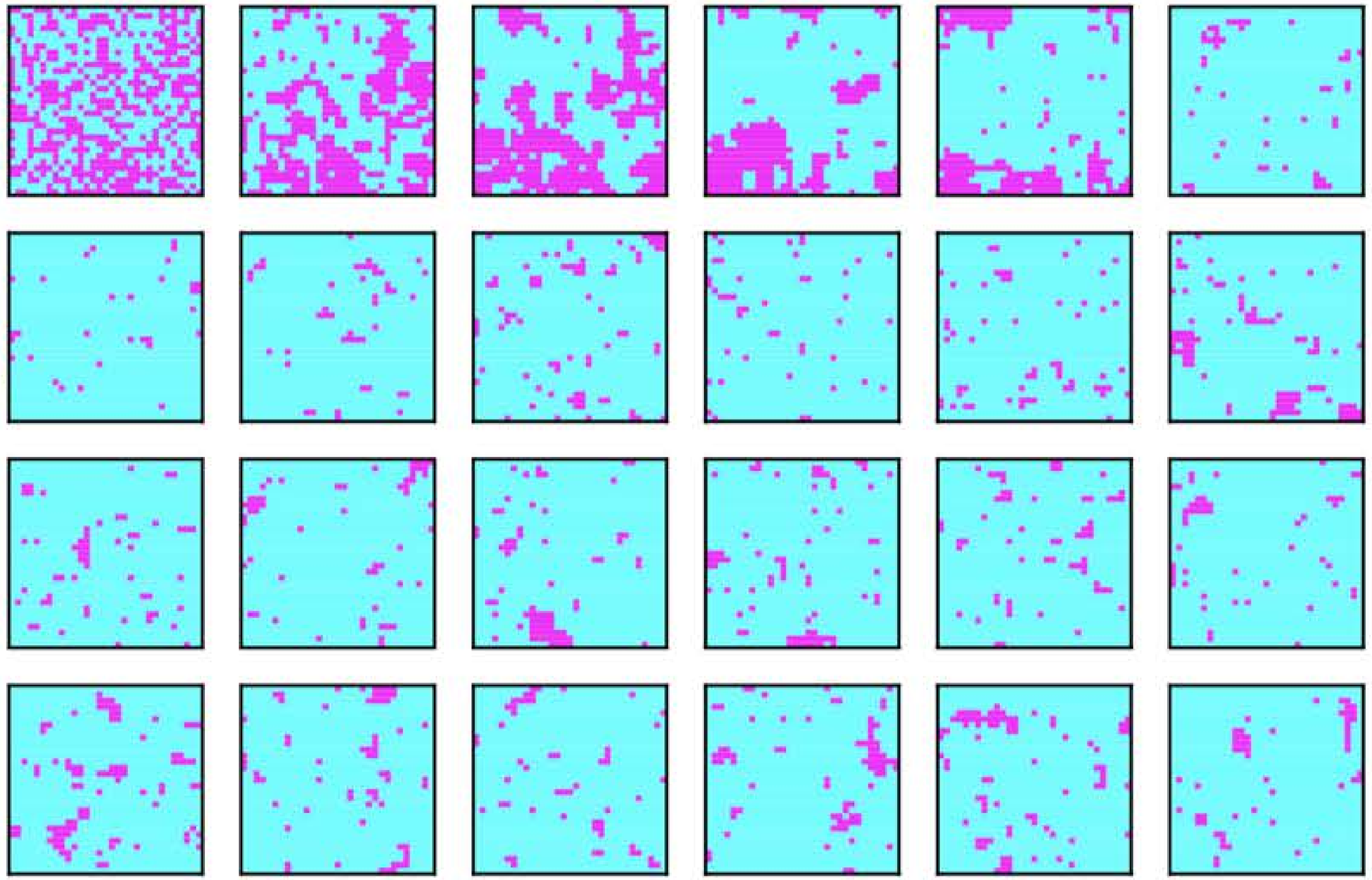
Configuration evolution with flipping



$$m = n^{-1} \sum_i s_i$$

$$U_4 = 1 - \frac{\langle m^4 \rangle}{3\langle m^2 \rangle^2}$$



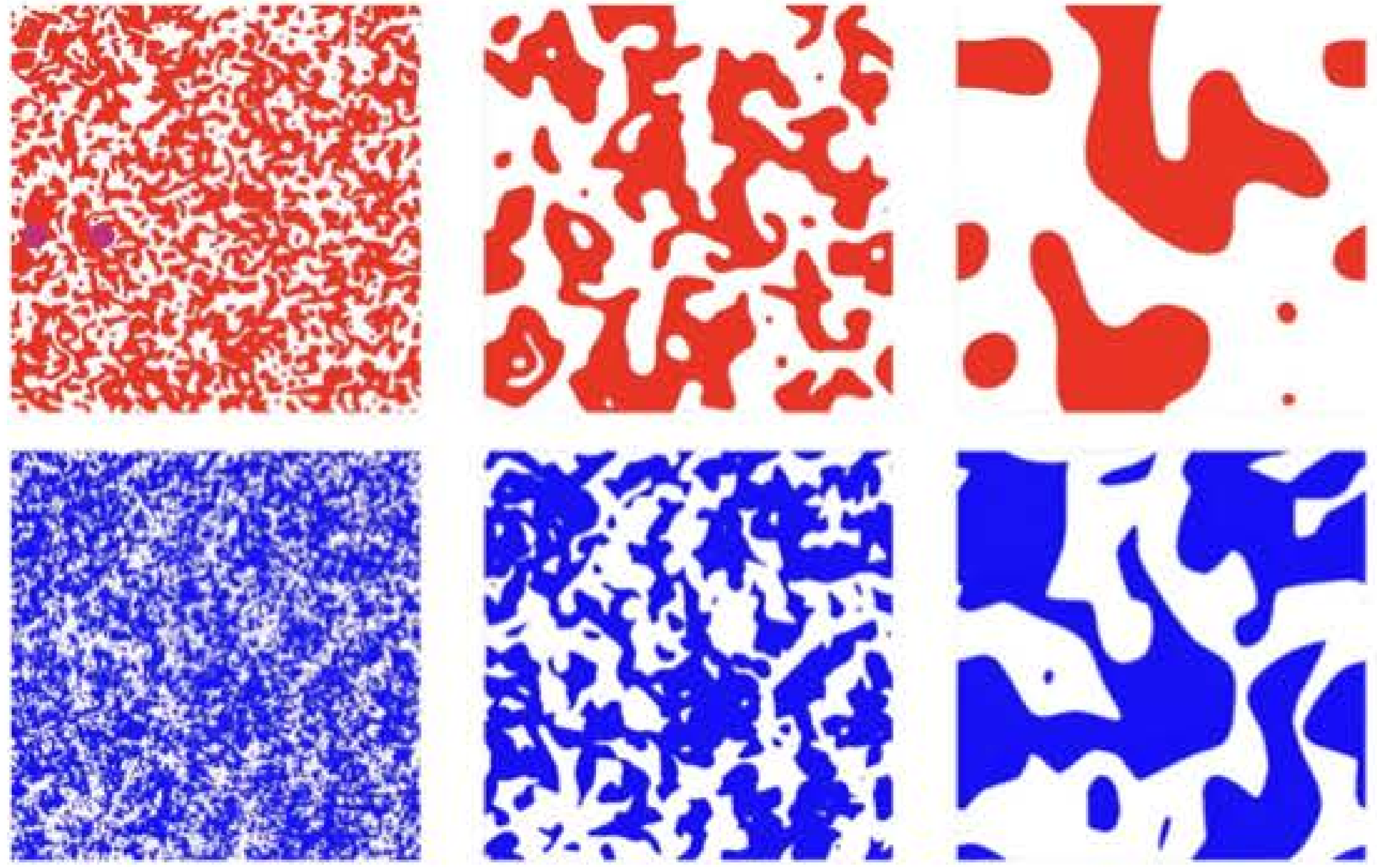


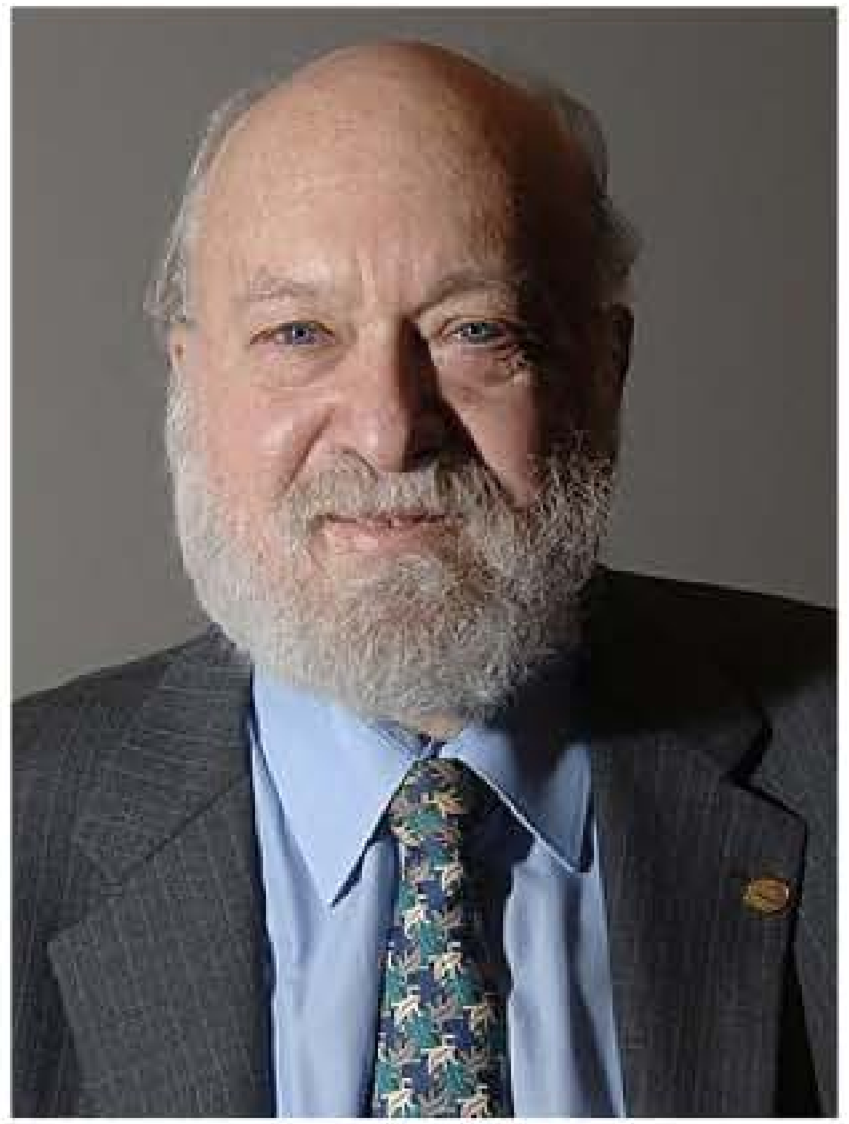
$k_{\rm B}T/J = 3.0$

$k_{\rm B}T/J = 2.1$

theory: $k_{\rm B}T/J_{\rm c} = \dfrac{2}{\ln(1+\sqrt{2})} \approx 2.269$

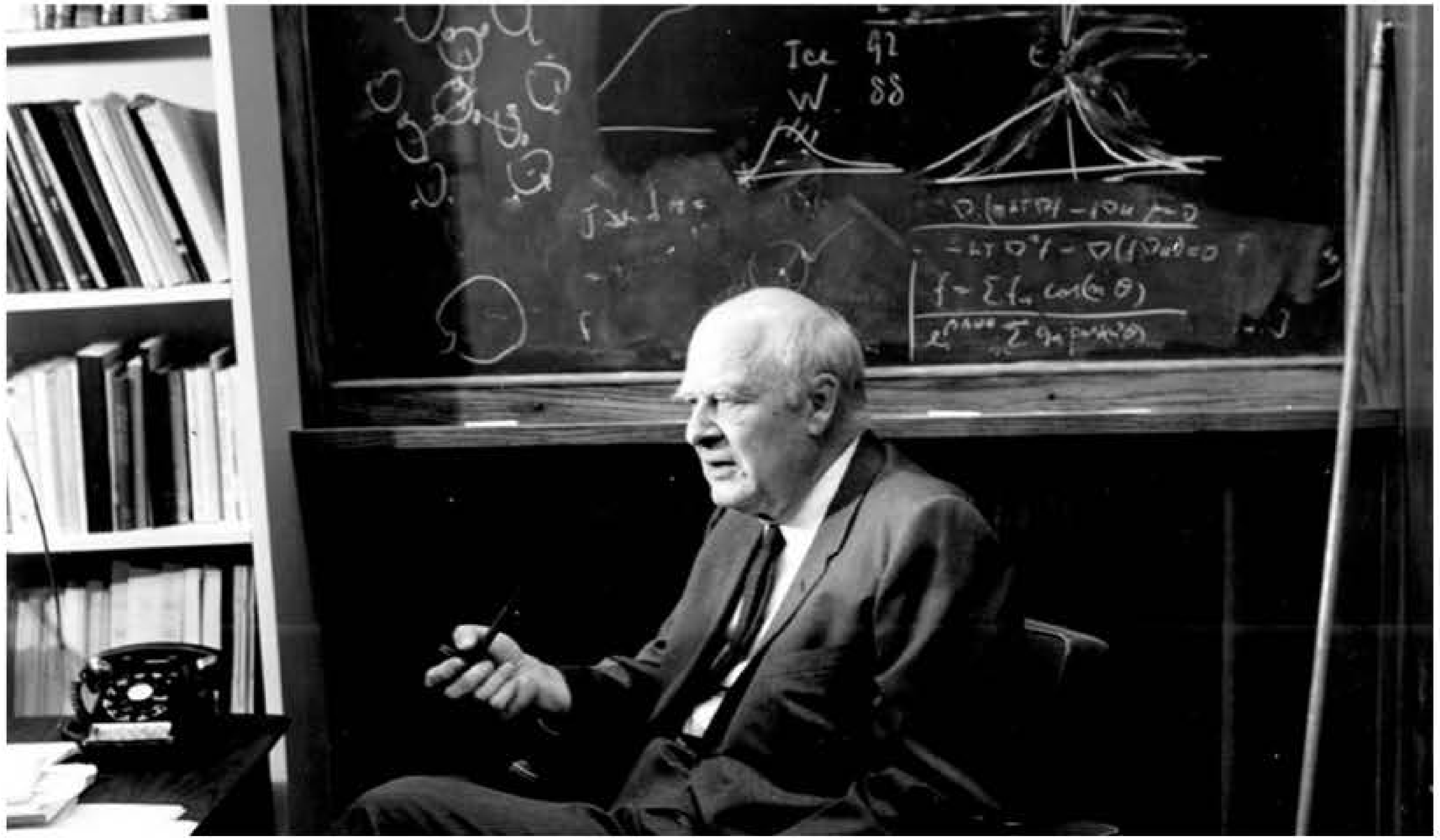
Onsager, C.N. Yang, Kadanoff, Wilson, ...





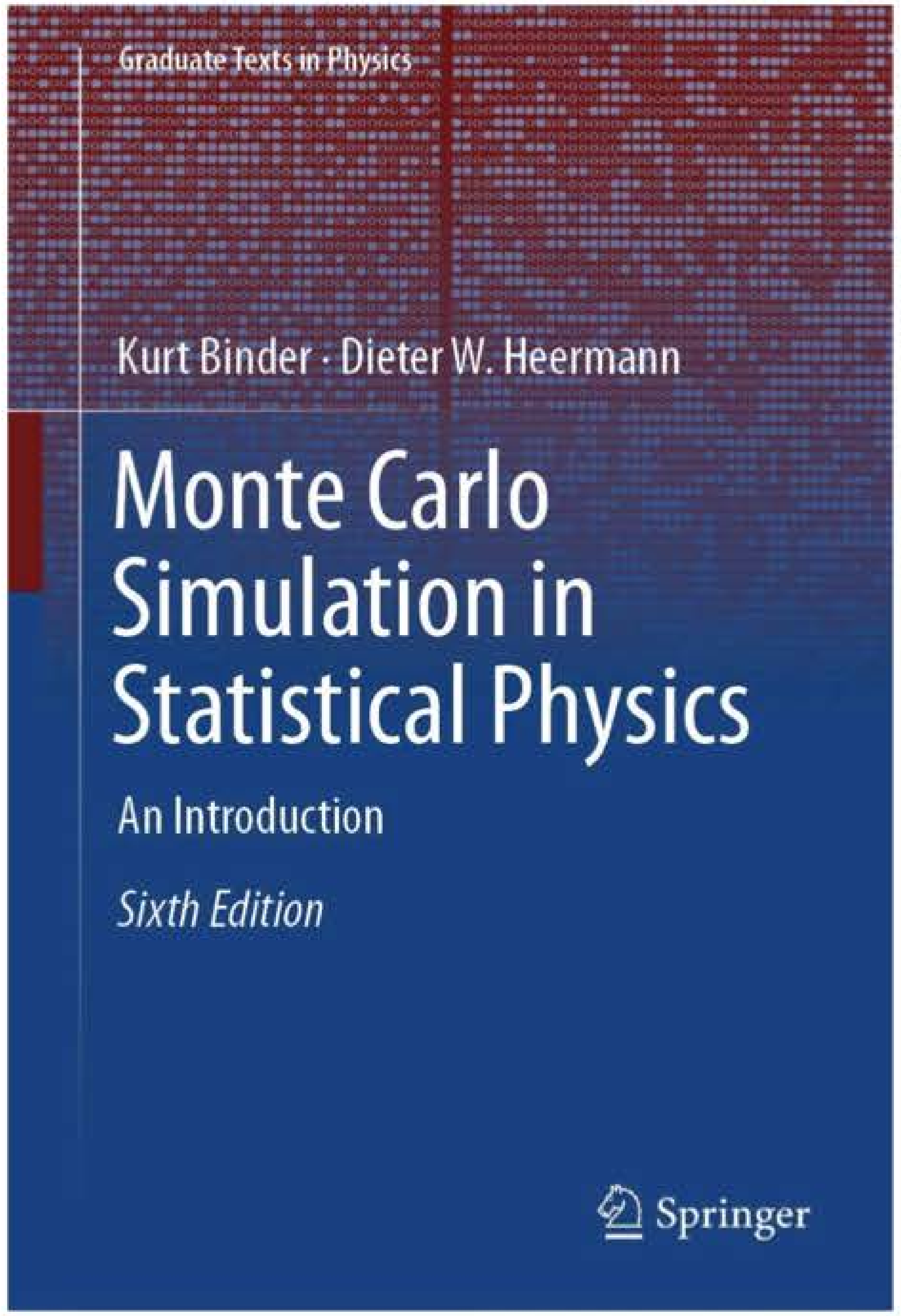
Wilson: $K^* = R(K^*)$

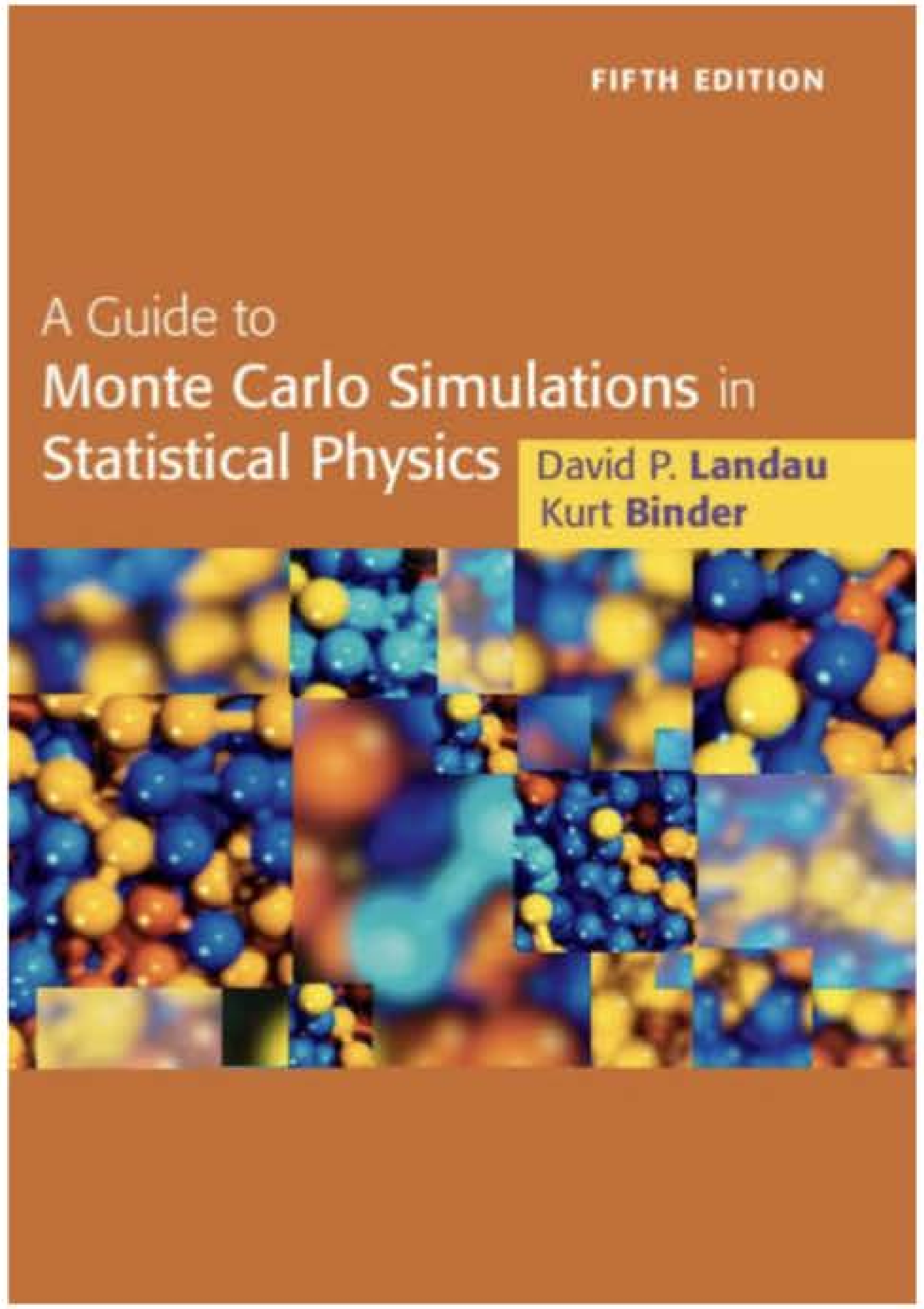
Onsager: $k_B T/J_c = \frac{2}{\ln(1+\sqrt{2})} \approx 2.269$

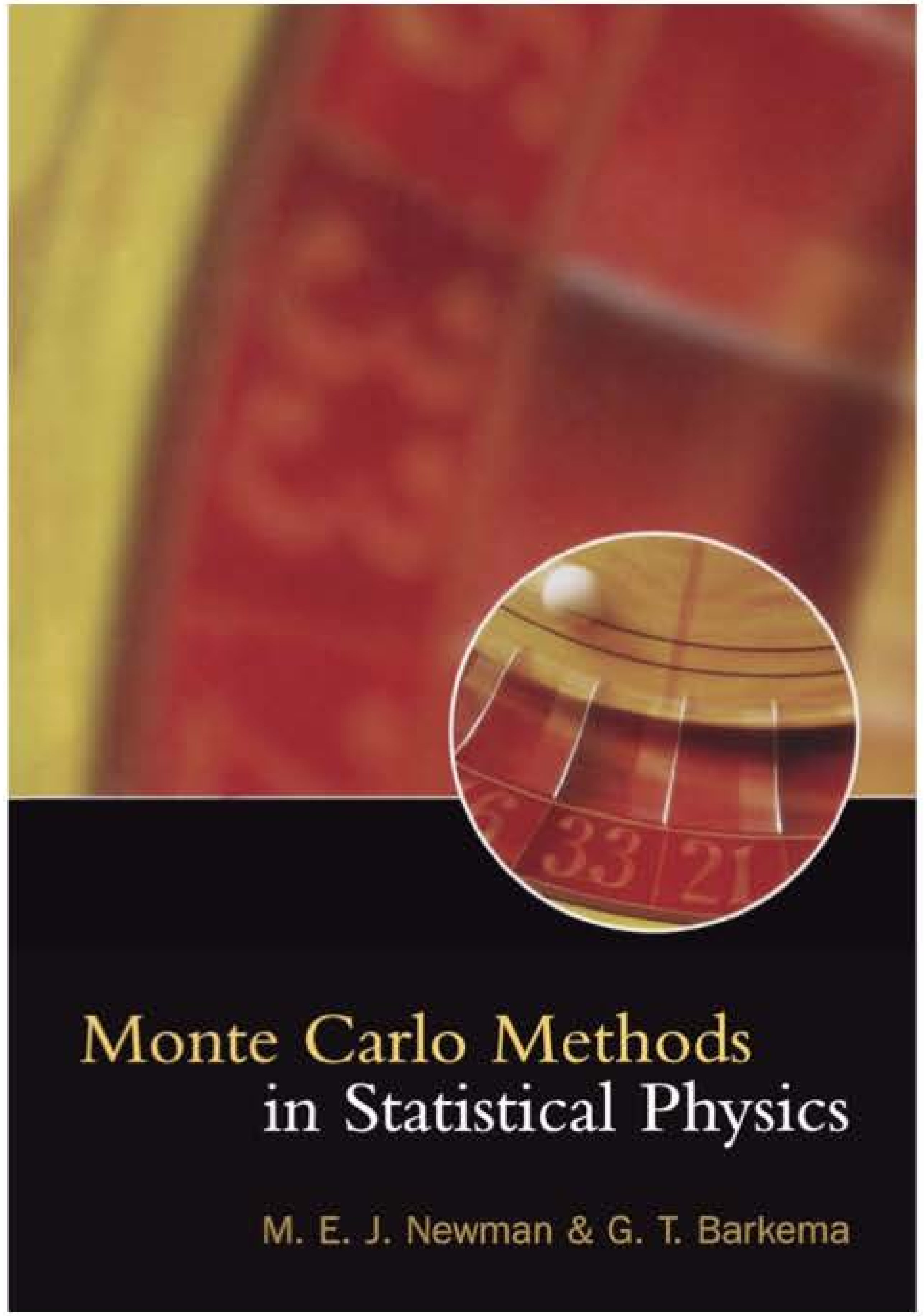


C.N. Yang: $m(T) = \left[1 - \sinh^{-4}\left(\frac{2J}{k_B T}\right)\right]^{1/8}$

A few textbooks on Monte Carlo in Statistical Physics



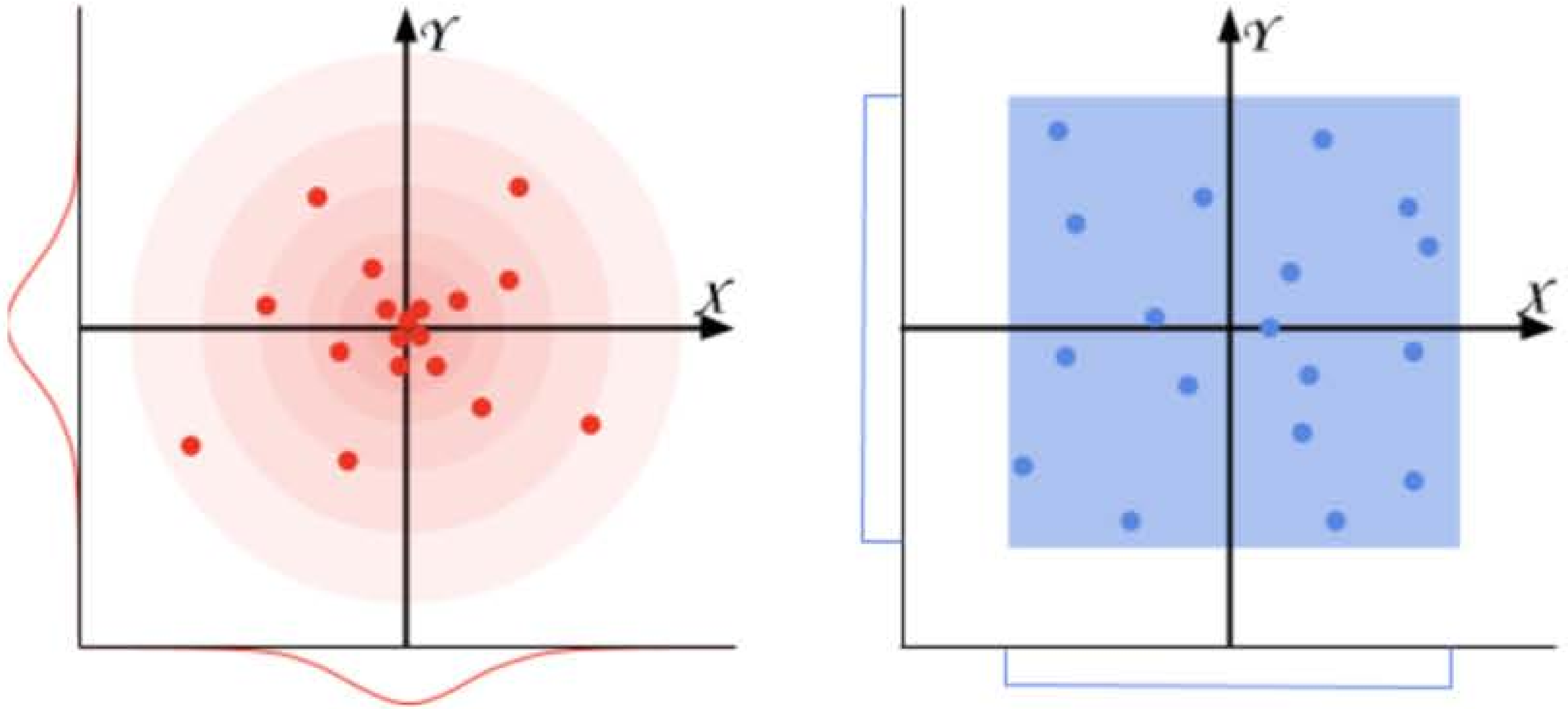




Wolff algorithms; correlation time; event-chain; Metropolis-Hastings, …

Inverse function method

Q: assume that we have only uniformly distributed random numbers, how to generate other distributed number?



Box-Muller algorithm for Gaussian

inverse function:

$$u = P(x) = \int_{-\infty}^{x} p(x')dx' \to x = P^{-1}(u), \; u \sim \mathrm{Unif}(0, 1)$$

Ex.: Prove it!

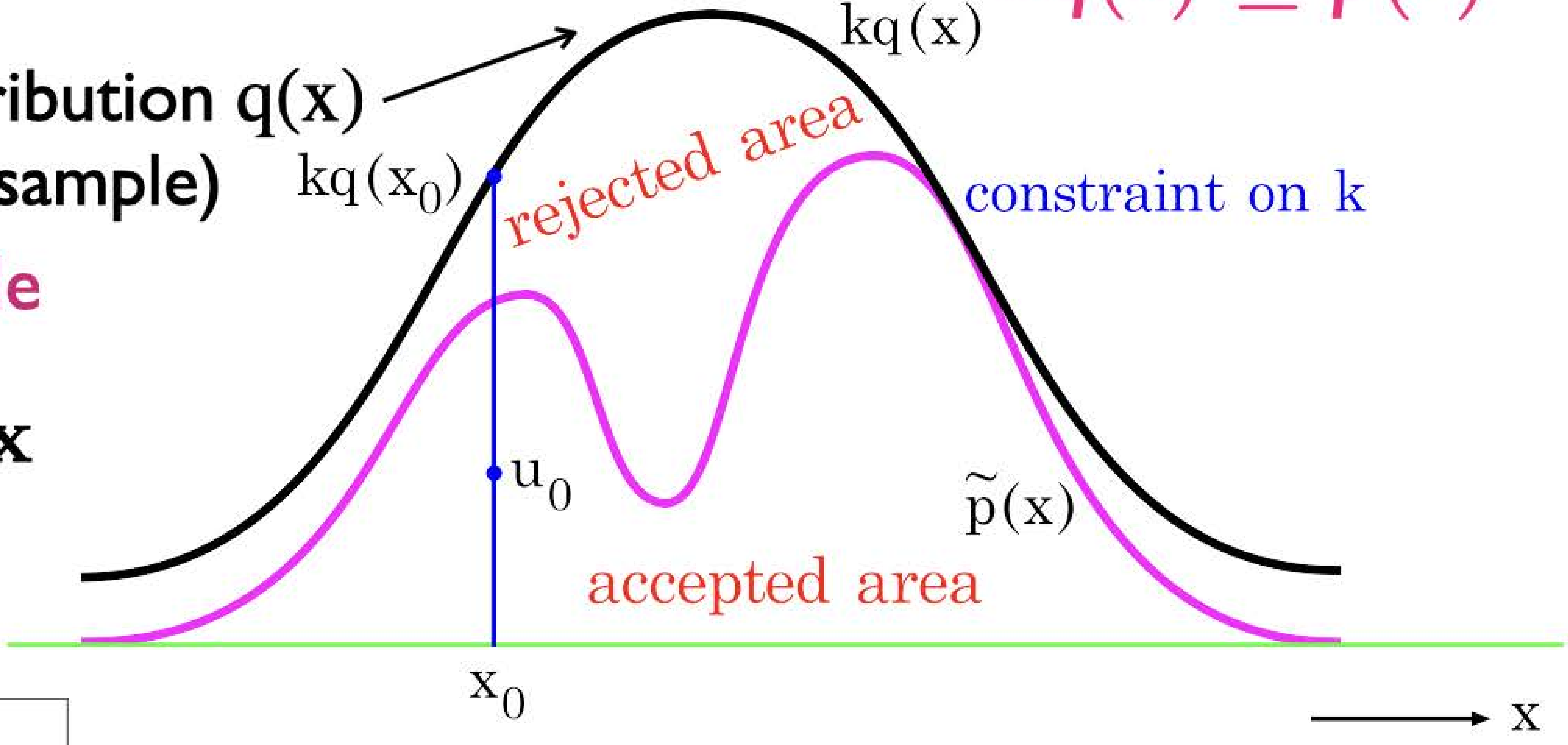
$$P_\lambda(x) = \lambda e^{-\lambda x} \to P_\lambda(x) = 1 - e^{-\lambda x} \to -\frac{\ln(1-u)}{\lambda} \sim p_\lambda(x) \to -\frac{\ln u}{\lambda} \sim p_\lambda(x)$$

Ex.: $p_n(x) = nx^{n-1}, 0 \leq x \leq 1; \; p(x) = \sin x, 0 \leq x \leq \pi/2$

Rejection sampling

$kq(x) \geq \widetilde{p}(x)$

proposal distribution q(x)
(easy to sample)

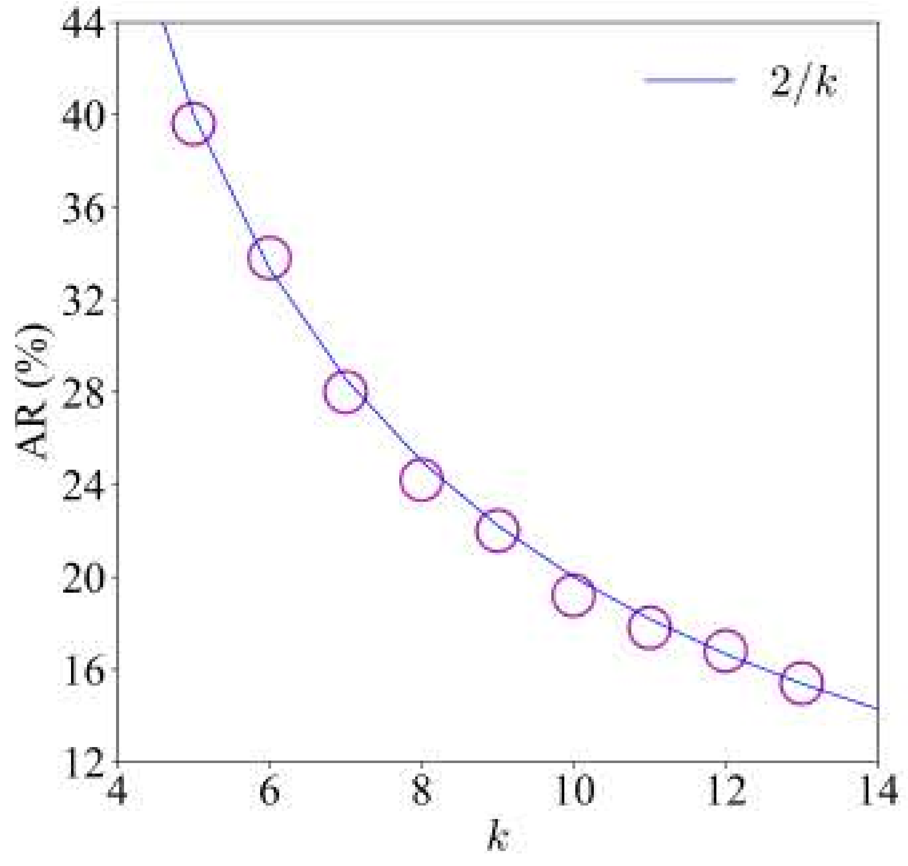


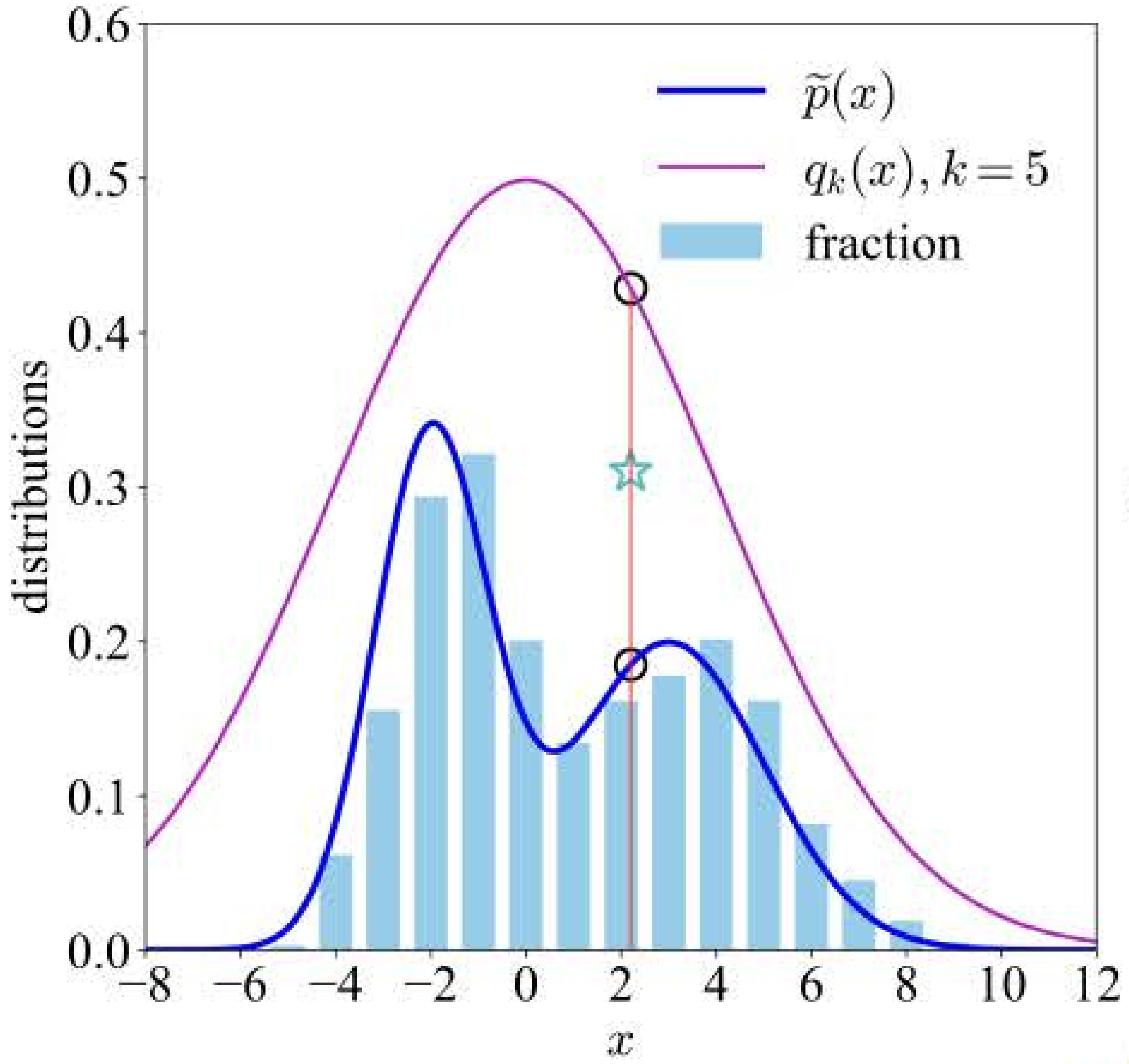
target distribution
p(x) : easy to evaluate/difficult to sample

$$p(\mathbf{x}) = \widetilde{p(\mathbf{x})}/Z_p, \;\; Z_p = \int p(\mathbf{x})dx$$

$$\widetilde{p}(x) = \sum_{i=1}^{2} \frac{\nu_i}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right)$$

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$





probability of accepting a suggested sample

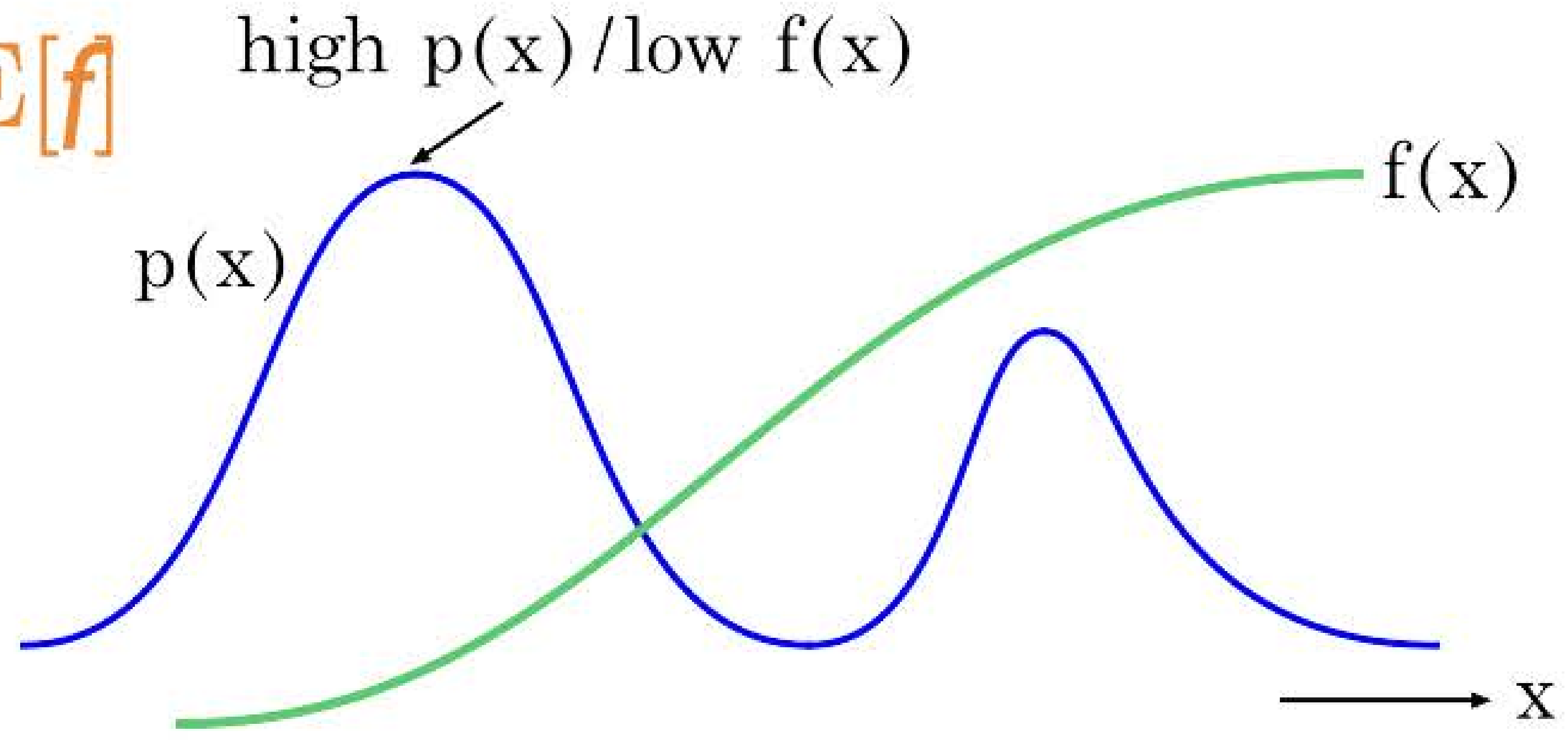
rejection sampling:

- $*$ sample $x \sim q(x)$
- $*$ generate $u \sim \mathrm{Unif}(0, kq(x))$
- $*$ accept if $u < \widetilde{p}(x)$ with prob$=\widetilde{p}(x)/kq(x)$

$$\text{prob} = \int \frac{\widetilde{p}(x)}{kq(x)} q(x)dx = \frac{1}{k}\int \widetilde{p}(x)dx = \frac{Z_p}{k} \sim k^{-1}$$

Importance sampling

Ex.: prove $\mathrm{E}[\hat{f}] = \mathrm{E}[f]$



$$\mathrm{E}[f] = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

$$\hat{f} = \frac{1}{M}\sum_{i=1}^{M} f(\mathbf{x}^{(i)}) \qquad \mathbf{x}^{(i)} \sim p(\mathbf{x})$$

$$\hat{f} = \frac{1}{M}\sum_{i=1}^{M} f(\mathbf{x}^{(i)})p(\mathbf{x}^{(i)}) \qquad \mathbf{x}^{(i)} \sim \text{Unif}$$

$$p(\mathbf{x}) = \tilde{p}(\mathbf{x})/Z_p, Z_p = \int p(\mathbf{x})d\mathbf{x}$$

$$q(\mathbf{x}) = \tilde{q}(\mathbf{x})/Z_q, Z_q = \int q(\mathbf{x})d\mathbf{x}$$

proposal distribution q(x)

$$\mathrm{E}[f] = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x} = \int f(\mathbf{x})\frac{p(\mathbf{x})}{q(\mathbf{x})}q(\mathbf{x})d\mathbf{x} \approx \frac{1}{M}\sum_{i=1}^{M}\frac{p(\mathbf{x}^{(i)})}{q(\mathbf{x}^{(i)})}f(\mathbf{x}^{(i)}), \quad \mathbf{x}^{(i)} \sim q(\mathbf{x})$$

$$\frac{Z_p}{Z_q} = \frac{1}{Z_q}\int \tilde{p}(\mathbf{x})d\mathbf{x} = \int \frac{\tilde{p}(\mathbf{x})}{\tilde{q}(\mathbf{x})}q(\mathbf{x})d\mathbf{x} \approx \frac{1}{M}\sum_{i=1}^{M}\tilde{r}_i, \; Z_q = \frac{\tilde{q}(\mathbf{x})}{q(\mathbf{x})}$$

$$\tilde{r}_i = \tilde{p}(\mathbf{x}^{(i)})/\tilde{q}(\mathbf{x}^{(i)})$$

$$\mathrm{E}[f] = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x} = \frac{Z_q}{Z_p}\int f(\mathbf{x})\frac{\tilde{p}(\mathbf{x})}{\tilde{q}(\mathbf{x})}q(\mathbf{x})d\mathbf{x} \approx \frac{Z_q}{Z_p}\frac{1}{M}\sum_{i=1}^{M}\tilde{r}_i f(\mathbf{x}^{(i)}) \approx \sum_{i=1}^{M} w_i f(\mathbf{x}^{(i)})$$

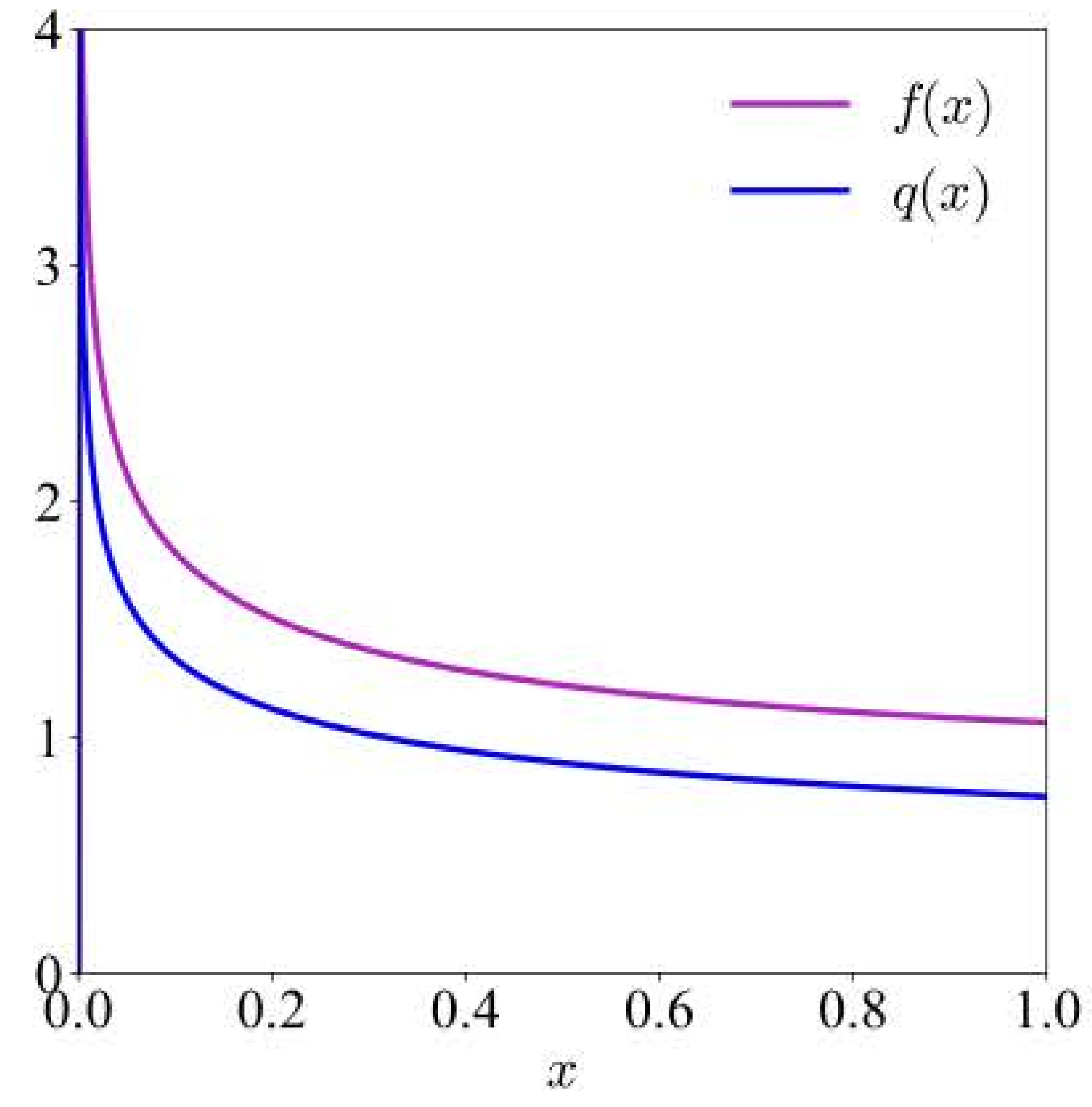
importance ratio

$$w_i = \frac{\tilde{r}_i}{\sum_{i'}\tilde{r}_{i'}} = \frac{\tilde{p}(\mathbf{x}^{(i)})/q(\mathbf{x}^{(i)})}{\sum_{i'}\tilde{p}(\mathbf{x}^{(i')})/q(\mathbf{x}^{(i')})}$$

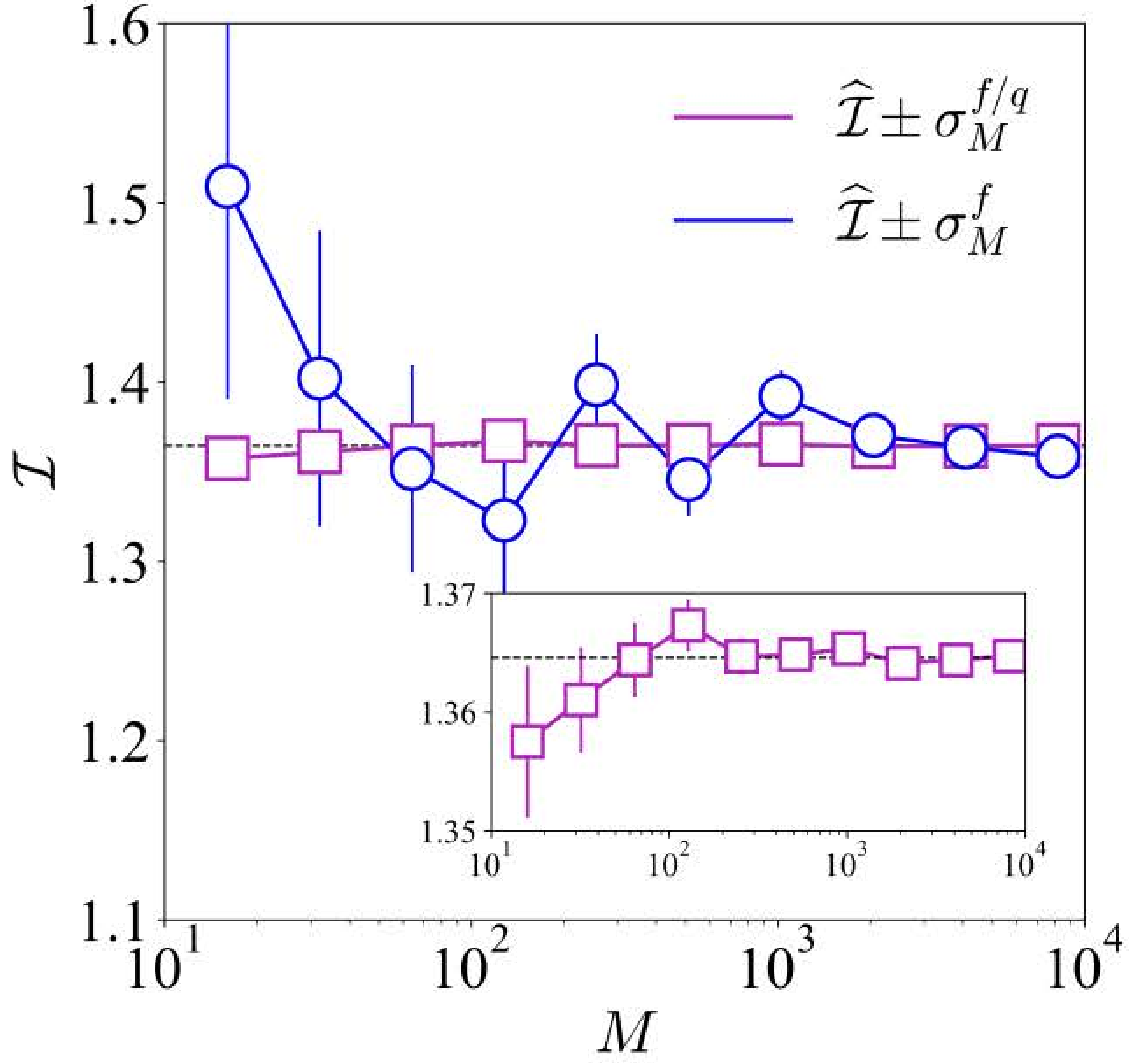
Variance reduction: example

$f(x) = x^{-1/4} + x/16, 0 \leq x \leq 1$

$$\sigma_M = \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{M-1}} \sim \frac{1}{\sqrt{M}}, \widehat{f} \approx \frac{1}{M}\sum \cdots$$



$q(x) = 3x^{3/4}/4$



$\sigma_M^f \approx 0.459/\sqrt{M}$

$\sigma_M^{f/q} \approx 0.025/\sqrt{M}$

$$\int_0^1 f(x)dx = \int_0^1 \frac{f(x)}{q(x)} q(x)dx = \int_0^1 \frac{f(x(y))}{q(x(y))} dy$$

$$\uparrow$$

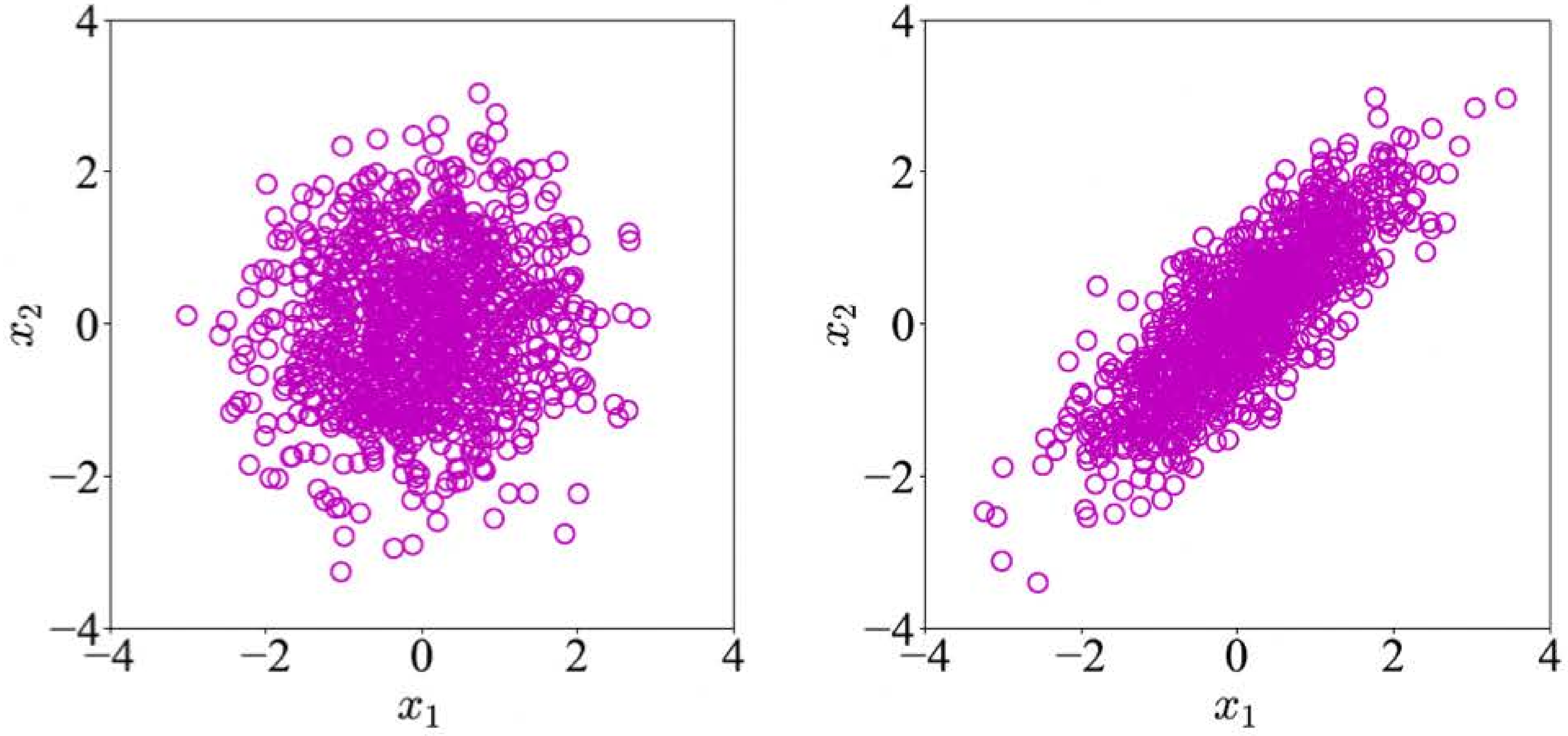
$$y = \int_0^x q(x')dx' = x^{3/4} \rightarrow x = y^{4/3}, y \sim \text{Unif}(0, 1)$$

Ex.: Under which dimension Simpson's rule wins over Monte Carlo?

*Gibbs sampling: high dimensional problems

$$p(\mathbf{x})p_{\mathbf{xy}} = p(\mathbf{y})p_{\mathbf{yx}}$$

applications:
noise images (German&German, 1984); biological processes



based on random walks

aim: $\vec{x} \sim p(\vec{x})$

update x_i using the marginal probability

$x_1 \neq y_1$, other components fixed

$p_{\vec{x}\to\vec{y}} \equiv p_{\vec{x}\vec{y}} = d^{-1}p(y_1|x_2,\cdots,x_d)$

$p_{\vec{y}\vec{x}} = d^{-1}p(x_1|y_2,\cdots,y_d) = d^{-1}p(x_1|x_2,\cdots,x_d)$

$$p_{\mathbf{yx}} = \frac{1}{d}\frac{p(\mathbf{x})}{p(x_2,x_3,\cdots,x_d)}$$

$$p_{\mathbf{xy}} = \frac{1}{d}\frac{p(y_1|x_2,x_3,\cdots,x_d)p(x_2,x_3,\cdots,x_d)}{p(x_2,x_3,\cdots,x_d)} = \frac{1}{d}\frac{p(y_1,x_2,x_3,\cdots,x_d)}{p(x_2,x_3,\cdots,x_d)} = \frac{1}{d}\frac{p(\mathbf{y})}{p(x_2,x_3,\cdots,x_d)}$$