Short History of Generalized Ensembles in Markov Chain Monte Carlo Simulations.

Bernd A. Berg

Florida State University

Telluride Workshop, July 14–18, 2008

Overview

1. Metropolis Algorithm and Hastings' Extension
2. Umbrella Sampling
3. Multicanonical and Multimagnetical Simulations
4. New Horizons
5. Replica Exchange Method and Expanded Ensembles
6. Summary and Conclusions

The talk will try to clear up some confusion about generalized ensemble methods (Get to the bottom of Charlie Brooks statement: "It is all umbrella sampling.") by presenting developments close to their chronological order.

Metropolis Algorithm and Hastings' Extension

Markov Chain Monte Carlo (MCMC) simulations started in earnest with the following famous paper:

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.

we move each of the particles in succession according to the following prescription:

$$X \rightarrow X + \alpha\xi_1$$
$$Y \rightarrow Y + \alpha\xi_2, \qquad (3)$$

where α is the maximum allowed displacement, which for the sake of this argument is arbitrary, and ξ_1 and ξ_2 are random numbers§ between (-1) and 1. Then, after

We then calculate the change in energy of the system ΔE, which is caused by the move. If $\Delta E<0$, i.e., if the move would bring the system to a state of lower energy, we allow the move and put the particle in its new position. If $\Delta E>0$, we allow the move with probability $\exp(-\Delta E/kT)$; i.e., we take a random number ξ_3 between 0 and 1, and if $\xi_3<\exp(-\Delta E/kT)$, we move the particle to its new position. If $\xi_3 > \exp(-\Delta E/kT)$, we return it to its old position. Then, whether the move has been allowed or not, i.e., whether we are in a different configuration or in the original configuration, we consider that we are in a new configuration for the purpose of taking our averages. So

Famous accept/reject criterion.

When a move is rejected: Count present configuration again!

Notable early developments:

1959 Salsburg, Jacobson, Fickett, and Wood, J. Chem. Phys. 30, 65–72: An attempt to calculate the partition function by MCMC using a method called in the modern language reweighting:

$$\exp(-\beta\, E) \;\rightarrow\; \exp(-\beta\, E - \triangle\beta\, E) = \exp(-\beta'\, E)\ .$$

As noticed by the authors their method is restricted to very small lattices.

1963 Glauber, J. Math. Phys. 4, 294-307: discussion of heatbath and other dynamics for the Ising model.

1967 McDonald and Singer, Discussions Faraday Soc. 43, 40–49: Used reweighting to evaluate physical quantities over a small range of temperatures.

1972 Valleau and Card, J. Chem. Phys. 37, 5457–5462: Multistage sampling using overlapping bridging distributions.

Extension:

Biometrika (1970), **57**, 1, *p.* 97
Printed in Great Britain

Monte Carlo sampling methods using Markov chains and their applications

BY W. K. HASTINGS
University of Toronto

SUMMARY

A generalization of the sampling method introduced by Metropolis *et al.* (1953) is presented along with an exposition of the relevant theory, techniques of application and methods and difficulties of assessing the error in Monte Carlo estimates. Examples of the

Arbitrary distributions π:

2·2. *Construction of the transition matrix*

In order to use this method for a given distribution $\boldsymbol{\pi}$, we must construct a Markov chain $\mathbf{P}$ with $\boldsymbol{\pi}$ as its stationary distribution. We now describe a general procedure for doing this which contains as special cases the methods which have been used for problems in statistical mechanics, in those cases where the matrix $\mathbf{P}$ was made to satisfy the reversibility condition that for all i and j

$$\pi_i p_{ij} = \pi_j p_{ji}. \tag{4}$$

The property ensures that $\Sigma \pi_i p_{ij} = \pi_j$, for all j, and hence that $\boldsymbol{\pi}$ is a stationary distribution of $\mathbf{P}$. The irreducibility of $\mathbf{P}$ must be checked in each specific application. It is only necessary to check that there is a positive probability of going from state i to state j in some finite number of transitions, for all pairs of states i and j.

We assume that p_{ij} has the form

$$p_{ij} = q_{ij}\alpha_{ij} \quad (i \neq j), \tag{5}$$

with

$$p_{ii} = 1 - \sum_{j \neq i} p_{ij},$$

where $\mathbf{Q} = \{q_{ij}\}$ is the transition matrix of an arbitrary Markov chain on the states $0, 1, \ldots, S$ and α_{ij} is given by

$$\alpha_{ij} = \frac{s_{ij}}{1 + \dfrac{\pi_i q_{ij}}{\pi_j q_{ji}}}, \tag{6}$$

where s_{ij} is a symmetric function of i and j chosen so that $0 \leqslant \alpha_{ij} \leqslant 1$ for all i and j. With this form for p_{ij} it is readily verified that $\pi_i p_{ij} = \pi_j p_{ji}$, as required. In order to simulate this process we carry out the following steps for each time t:

(i) assume that $X(t) = i$ and select a state j using the distribution given by the ith row of $\mathbf{Q}$;

(ii) take $X(t+1) = j$ with probability α_{ij} and $X(t+1) = i$ with probability $1 - \alpha_{ij}$.

For the choices of s_{ij} we will consider, only the quantity $(\pi_j q_{ji})/(\pi_i q_{ij})$ enters into the simulation and we will henceforth refer to it as the test ratio.

Reversibility conditions: Detailed balance.

Irreducibility: Ergodicity.

Normalization of the distribution π_i drops out.

Metropolis updating is a special case:

$$s_{ij}^M = \begin{cases} 1 + \dfrac{\pi_i q_{ij}}{\pi_j q_{ji}} & \text{for} \quad \dfrac{\pi_j q_{ji}}{\pi_i p_{ij}} \geq 1 \quad \Rightarrow \quad \alpha_{ij}^M = 1\,, \\ 1 + \dfrac{\pi_j q_{ji}}{\pi_i q_{ij}} & \text{for} \quad \dfrac{\pi_j q_{ji}}{\pi_i p_{ij}} < 1 \quad \Rightarrow \quad \alpha_{ij}^M = \dfrac{\pi_j q_{ji}}{\pi_i q_{ij}}\,, \end{cases}$$

which goes for $q_{ij} \neq q_{ji}$ sometimes under the name biased Metropolis updating.

Symmetry:

$$q_{ij} = q_{ji} \quad \Rightarrow \quad \alpha_{ij}^{M} = \begin{cases} 1 \quad \text{for} \quad \frac{\pi_j}{\pi_i} \geq 1\,, \\ \frac{\pi_j}{\pi_i} \quad \text{for} \quad \frac{\pi_j}{\pi_i} < 1\,. \end{cases}$$

Boltzmann distribution:

$$\pi_i = \exp(-\beta E_i) \quad \Rightarrow \quad \text{Metropolis paper.}$$

Umbrella Sampling

Torrie and Valleau, Chem. Phys. Lett. 28 (1974) 578: estimation of free energies by sampling on distributions designed for this purpose (no reference to Hastings). Further Elaboration:

A Monte Carlo method for obtaining the interionic potential of mean force in ionic solution

G. N. Patey and J. P. Valleau

Lash Miller Chemical Laboratories, University of Toronto, Toronto, Ontario, Canada M5S 1A1
(Received 28 April 1975)

Since we want to explore a wider region of ionic separations, we used a modified importance sampling scheme. The configuration were chosen with a frequency proportional to $w(r_{12})\exp(-U/kT)$, where the weighting function $w(r_{12})$ was chosen to spread the sampling over the desired range of r_{12}. ... Accepted or rejected depending on the ratio

$$\frac{\pi'}{\pi} = \frac{w(r'_{12})}{w(r_{12})} \exp\left(-\frac{U'-U}{kT}\right) .$$

J. Comp. Phys 23, 187–199 (1977):

Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling

G. M. Torrie and J. P. Valleau

Lash Miller Chemical Laboratories, University of Toronto, Toronto, Ontario, Canada

Received May 7, 1976; revised June 16, 1976

The free energy difference between a model system and some reference system can easily be written as an ensemble average, but the conventional Monte Carlo methods of obtaining such averages are inadequate for the free-energy case. That is because the Boltzmann-weighted sampling distribution ordinarily used is extremely inefficient for the purpose. This paper describes the use of arbitrary sampling distributions chosen to facilitate such estimates. The methods have been tested successfully on the Lennard–Jones system

... statistical properties such as the entropy and free energy, because they cannot be expressed as ensemble averages, have not been so easily accessible. The conventional technique has been numerical integration, ... This somewhat cumbersome method is least efficient or altogether unworkable when the system undergoes a phase transition, because of the difficulty of defining a path of integration on which the necessary ensemble averages can be reliable measured, ...

In the calculations described in the following sections weighting functions were used which brought about sampling of a range of ΔU^* up to three times that of a conventional Monte Carlo experiment, allowing accurate determination of values of $f_0(\Delta U^*)$

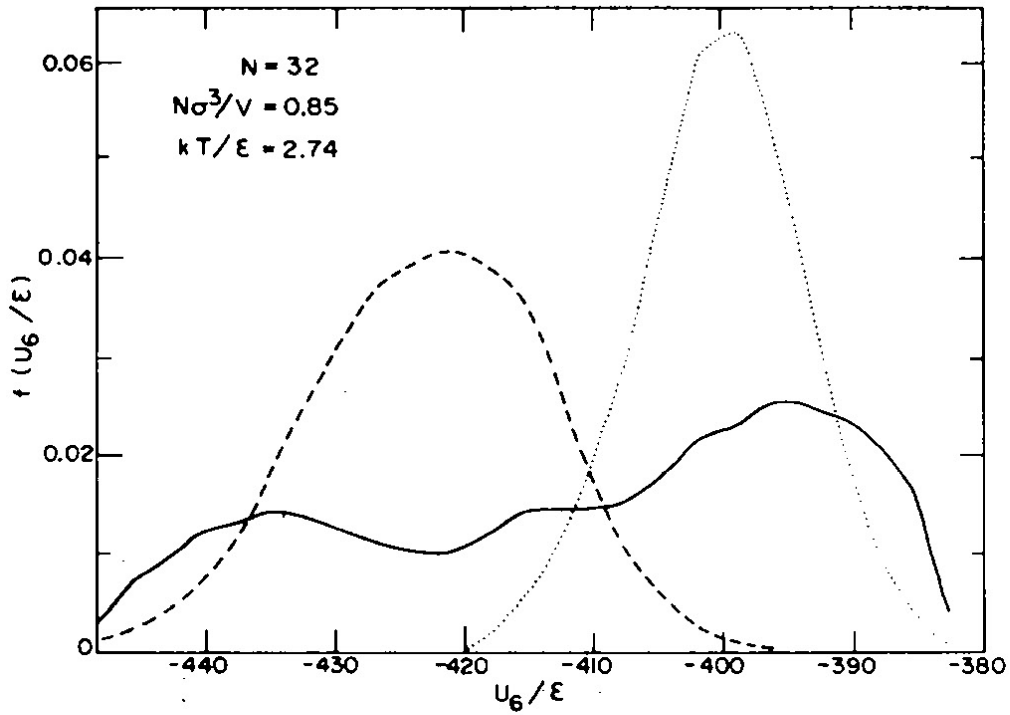


FIG. 1. Probability density functions for U_6 in a 32-particle soft-sphere fluid at $kT/\epsilon = 2.74$, $N\sigma^3/V = 0.85$. Solid line, f_w, the biased probability density. Dotted line, f_0, the unbiased probability density obtained by reweighting f_w. Broken line, relative values of $f_6(U_6/\epsilon)\exp(-U_6/kT)$ normalized to unity.

Superficially, the most serious limitation of the sampling techniques described here may appear to be the lack of a direct and straightforward way of determining the weighting function to use for a given problem. Instead, $w(\mathbf{q}^N)$ must be determined by a trial-and-error procedure for each case, often beginning with the information available from the distribution in a very short Boltzmann-weighted experiment which is then broadened in stages through subsequent short test runs with successively greater bias of the sampling. What this rather inelegant procedure lacks aesthetically is more than compensated by the efficiency of the ultimate umbrella-sampling experiment.

Summary:

1. Weighting functions $w(q_1, \ldots, q_N)$ depend on (generalized) coordinates.
2. Determination via trial and error.

Conclusion eleven years later (Li and Scheraga, J. Mol. Struct. (Theochem) 179, 333-352 (1988)):

The difficulty of finding such weighting factors has prevented wide applications of umbrella sampling.

Umbrella Sampling in Practice

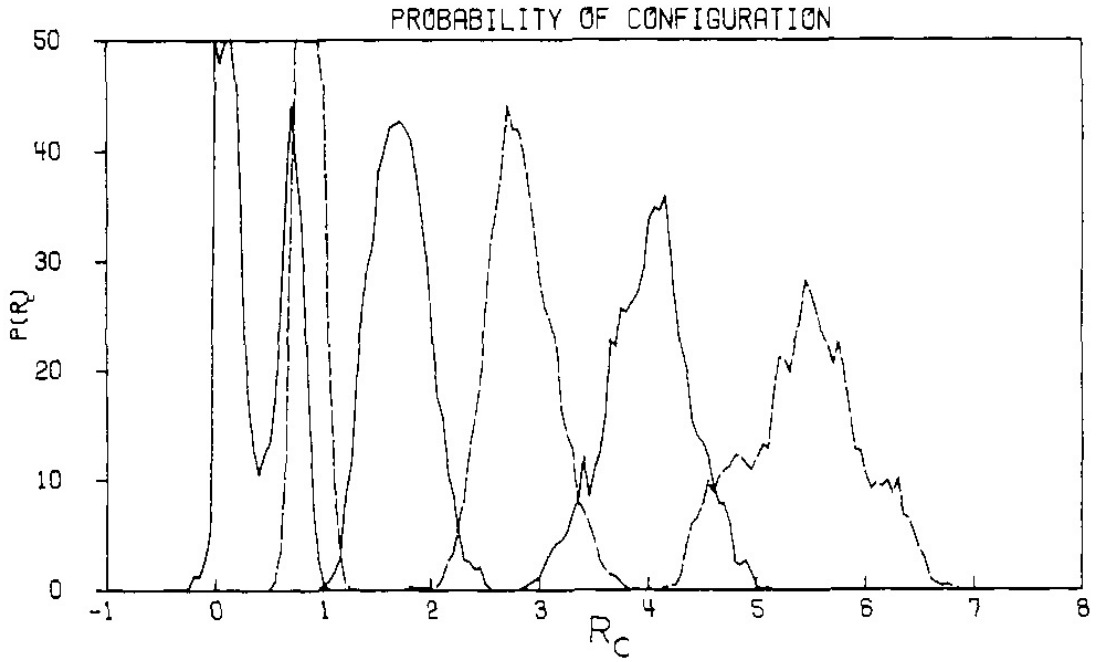
J. Am. Chem. Soc. 107, 154–163 (1985):

Theoretical Examination of the S_N2 Reaction Involving Chloride Ion and Methyl Chloride in the Gas Phase and Aqueous Solution

Jayaraman Chandrasekhar,* Scott F. Smith, and William L. Jorgensen*

Contribution from the Department of Chemistry, Purdue University, West Lafayette, Indiana 47907. *Received May 29, 1984*

Simulations with importance sampling were carried out over six windows, corresponding to different choices of the umbrella potential, u'. As in previous applications, harmonic forces centered at different values of r_c were used in the present study.[14] The force constants were chosen to be progressively smaller when going from the transition-state region to the products region, reflecting the expected steepness of the energy profile. In addition to the harmonic force, it was found necessary to add an exponential biasing potential to ensure uniform sampling, especially near the transition state. The final choice of the umbrella potential is given



Gas-Phase vs. Solution Energy Profiles. The *unnormalized* solute distributions, $P_i(r_c)$, for the six windows are shown in Figure 5. Smooth Gaussian distributions with widths inversely proportional to the harmonic force constants, k_i, would be obtained if the true relative energy of the solution is exactly counterbalanced by the exponential term in the biasing potential (eq 9). This is roughly the case for windows 2–6, whose $P_i(r_c)$ are smooth distributions spanning increasing ranges of r_c. The distribution for the first window has two peaks, indicating that the variation in the pmf in this region is not perfectly compensated by the exponential biasing term. Nevertheless, the region from $r_c = 0$ to

Summary:

Tedious patching of umbrella potentials.

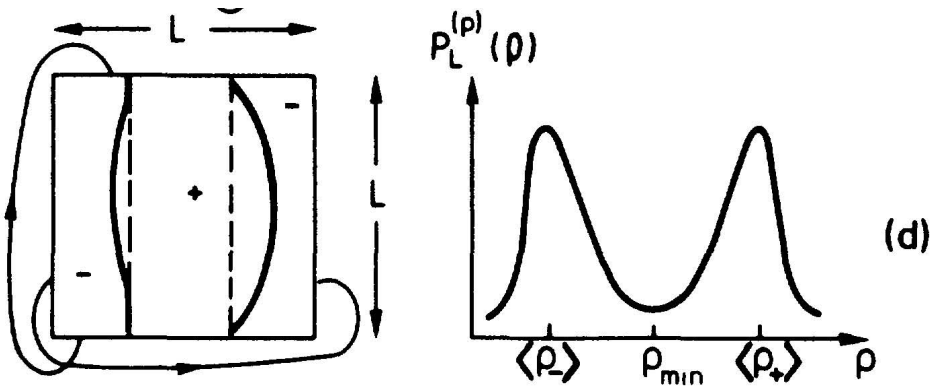
Let's now backup a bit in time to follow another line of developments:

Binder's Method for Estimating Interface Tensions

Binder, Phys. Rev. A 25, 1699–1709 (1982).

Simulations with periodic boundary conditions.

From equal heights double peaks : $$F_L^s = -\frac{1}{L^{(D-1)}} \ln\left(\frac{P_L^{\min}}{P_L^{\max}}\right) .$$



First results with canonical simulations remained pitiful. Reason: $P_L^{\min}$ is exponentially suppressed in the canonical ensemble.

Chandler: Introduction to Modern Statistical Mechanics, Oxford University Press 1987. Chapter 6.3: Non-Boltzmann Sampling:

Non-Boltzmann sampling is a powerful tool when the reference or unperturbed energy, $E_\nu^{(0)}$, creates a trajectory that is close to that of E_ν. Comment: There is no need for the trajectories to stay close.

Non-Boltzmann sampling can also be useful in removing the bottlenecks that create quasi-ergodic problems and in focusing attention on rare events. Comment: This is on the right track.

The practice: To illustrate this methodology, we consider ... the computation of the free energy function, $\tilde{A}(M)$. This function is defined for the Ising magnet by

$$\exp\left[-\beta\tilde{A}(M)\right] = \sum_v \Delta\left(M - \mu\sum_{i=1}^{N} s_i\right) \exp\left(-\beta E_\nu\right) ,$$

where $\Delta(x)$ is the Kronecker delta ... .

Clearly,

$$\exp\left[-\beta\tilde{A}(M)\right] \propto P(M) = \left\langle \Delta\left(M - \mu\sum_{i=1}^{N} s_i\right)\right\rangle ,$$

where $P(M)$ is the probability of observing the Ising magnet with magnetization M. ... if we consider the situation of broken symmetry (i.e., when $T < T_c$), and plan to compute $\tilde{A}(M)$ for a wide range of M values, we immediately encounter a serious problem ... $\tilde{A}(M)$ is a bistable function of M, ... visitation of states with $M = 0$ is an infrequent event ...

The method of umbrella sampling, however, avoids this difficulty. We chose a set of umbrella or window potentials

$$\begin{aligned} W_\nu &= 0, \quad \text{for } M_i - w/2 \le \mu\sum_{j=1}^{N} s_j \le M_i \le M - i + w/2 \\ &= \infty, \quad \text{otherwise.} \end{aligned}$$

... $\tilde{A}(M)$ is determined in each window to within an additive constant ... , which must be adjusted from one window to the next, ...

Let τ denote the required computer time to acquire such (accurate) statistics in each window. ... Note that

$$\tau \propto w^2 \,.$$

Therefore, the total computation time required to determine $\tilde{A}(M)$ by the method of umbrella sampling is

$$t_{\mathrm{CPU}} \propto nw^2 \,.$$

Now, how much time would it have taken if we did not use this method? As a lower bound, let us assume that $\tilde{A}(M)$ does not vary more than a few $k_B T$ over the entire range of M. The size of this range is nw. Therefore, the time to sample this range is proportional to $(nw)^2 = n\, t_{\mathrm{CPU}}$. Hence, without the windows, the computation time would be n times longer than that with the n windows.

Exercise BB1: What is wrong with Chandler's argument?

(Replacing Exercise 6.8 of the book.)

The advantage (i.e., lower computation time) of umbrella sampling is, of course, even greater than this when regions of M have relatively high values of $\tilde{A}(M)$ and thus relatively low probabilities.

Comment: This is the (only) major advantage!

400 spin Ising magnet, ... 10 windows were used between $M = 0$ and $M = 400\mu$. From the figure (next transparency):

$$\frac{P^{\min}}{P^{\max}} \approx e^{-14} = 8.3 \times 10^{-7} .$$

Comment: Configurations sampled, which are in Boltzmann simulations suppressed by 6 to 7 orders of magnitude.

No mention of Binder or any attempt to calculate the interface tension (known exactly since Onsager's 1944 solution of the 2D Ising model). Umbrella potentials by Tedious patching of windows remains (in Chandler's example chosen to be constant).

Multicanonical and Multimagnetical Simulation

Also in 1987: Umbrella sampling in Chandler's version was discovered independently by Bhanot and collaborators, and used in a series of papers: Phys. Rev. Lett. 59 (1987) 803 and references given therein.

Due to the tedious patching of windows their simulations remained confined to rather small lattices.

This changed with

VOLUME 68, NUMBER 1 PHYSICAL REVIEW LETTERS 6 JANUARY 1992

Multicanonical Ensemble: A New Approach to Simulate First-Order Phase Transitions

Bernd A. Berg [1),(2),(a)] and Thomas Neuhaus [1)]

[1)] *Fakultät für Physik, Universität Bielefeld, D-4800 Bielefeld, Federal Republic of Germany*

[2)] *Supercomputer Computations Research Institute, Tallahassee, Florida 32306*

(Received 19 July 1991)

The multicanonical MC algorithm samples configurations with the weight

$$\mathcal{P}_L^{\mathrm{MC}}(S) \sim e^{(\alpha_L^k + \beta_L^k S)} \quad \text{for} \quad S_L^k < S \leq S_L^{k+1} \tag{7}$$

instead of sampling with the usual Boltzmann factor $P_L^B(S) \sim \exp(\beta_L^f S)$ corresponding to the canonical ensemble. Here we partitioned the total action interval $0 \leq S \leq 2V$ into $k = 0, \ldots, N$ ($N+1$ odd) intervals $I_k = (S_L^k, S_L^{k+1}]$. The idea of the multicanonical MC algorithm is to choose intervals I_k and values of β_L^k and α_L^k at the pseudocritical point β_L^f in such a way that the resulting multicanonical action density $\mathcal{P}_L(S)$ has an approximately flat behavior for values of the action in the interval $[S_L^{1,\max}, S_L^{2,\max}]$; that is to say, configurations dominated by the interface are no longer exponentially suppressed

... In practice the appropriate choice of the parameters in Eq. (7) is obtained by making from the given systems an FSS prediction of the density distribution $P_L(S)$ for the next larger system. A second run may then be performed with optimized parameters [Citation End].

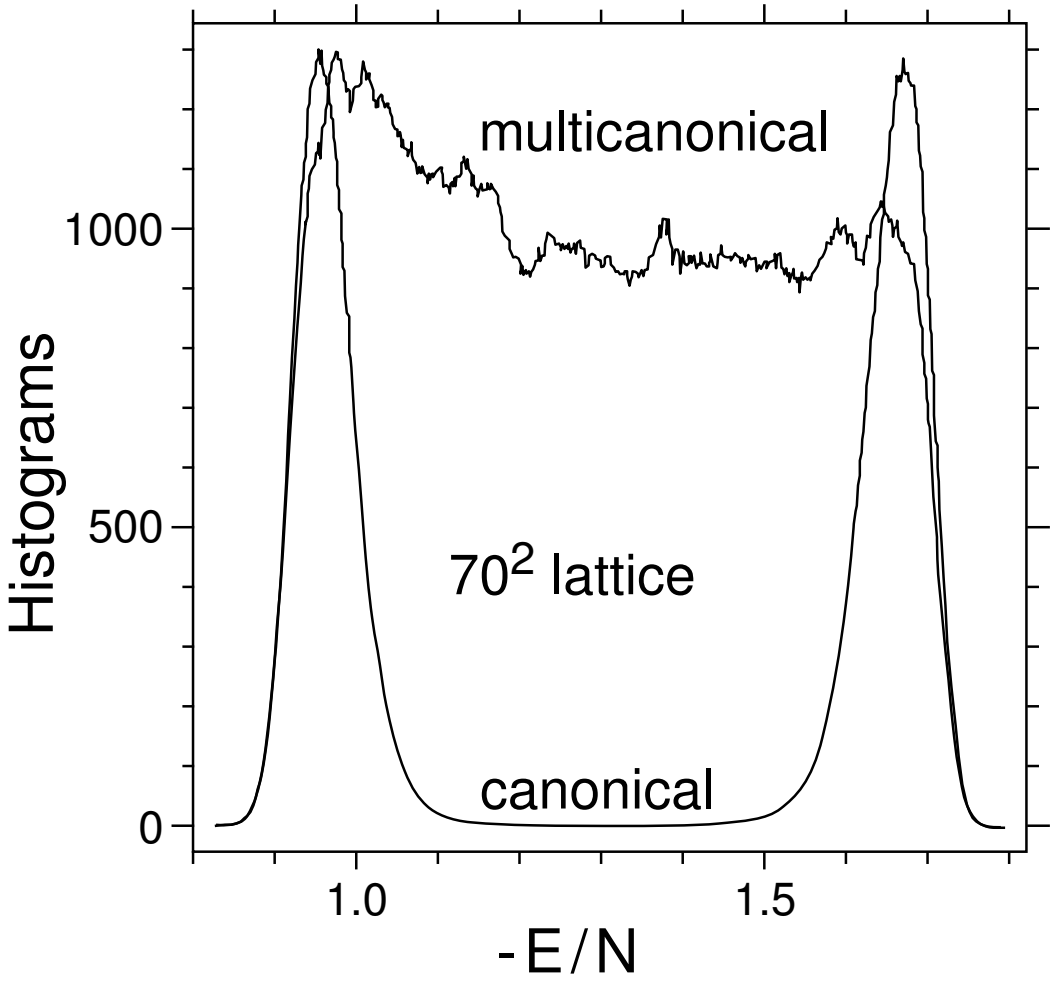
Changes compared to umbrella sampling:

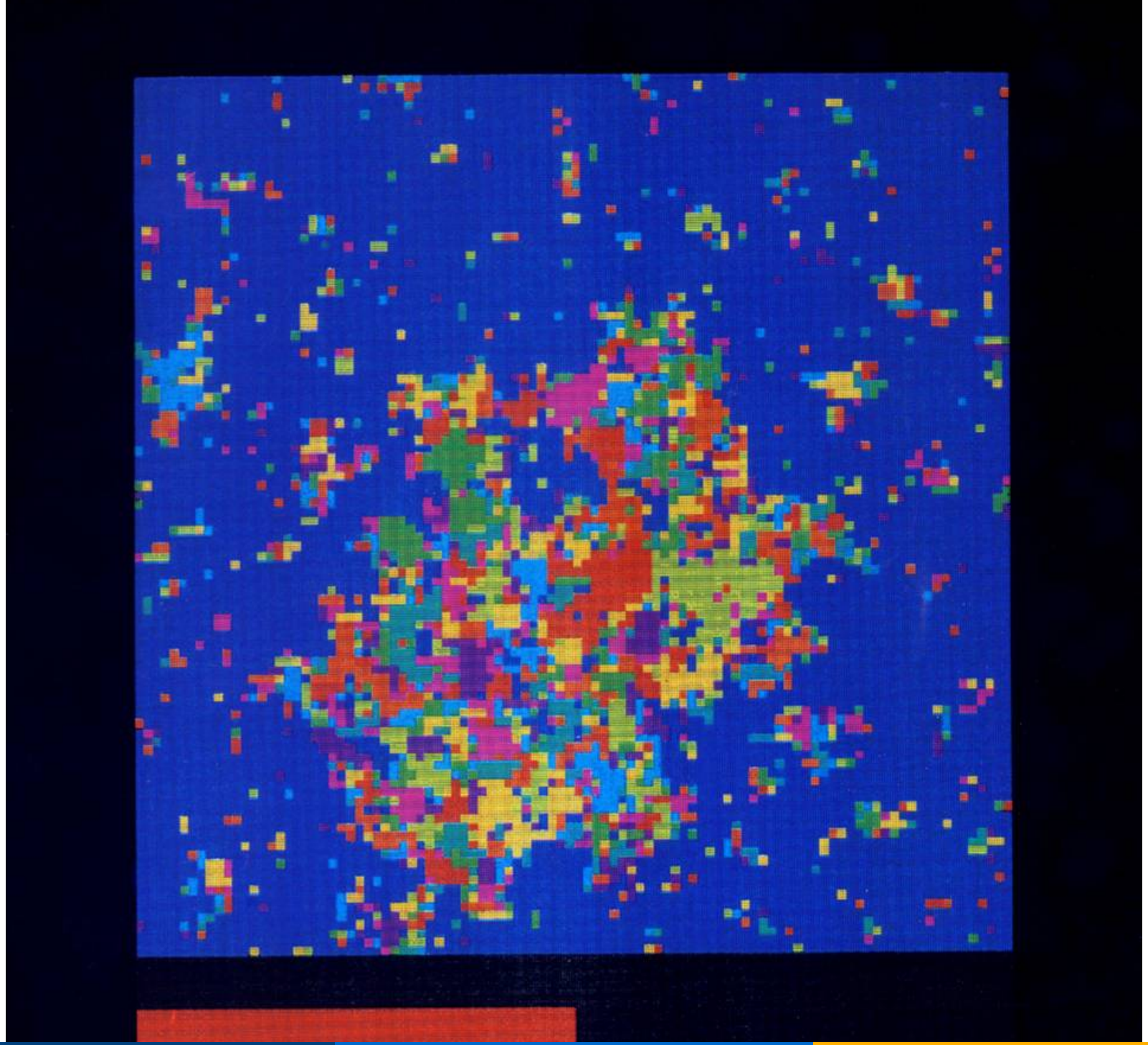
No windows anymore, all in one ensemble.

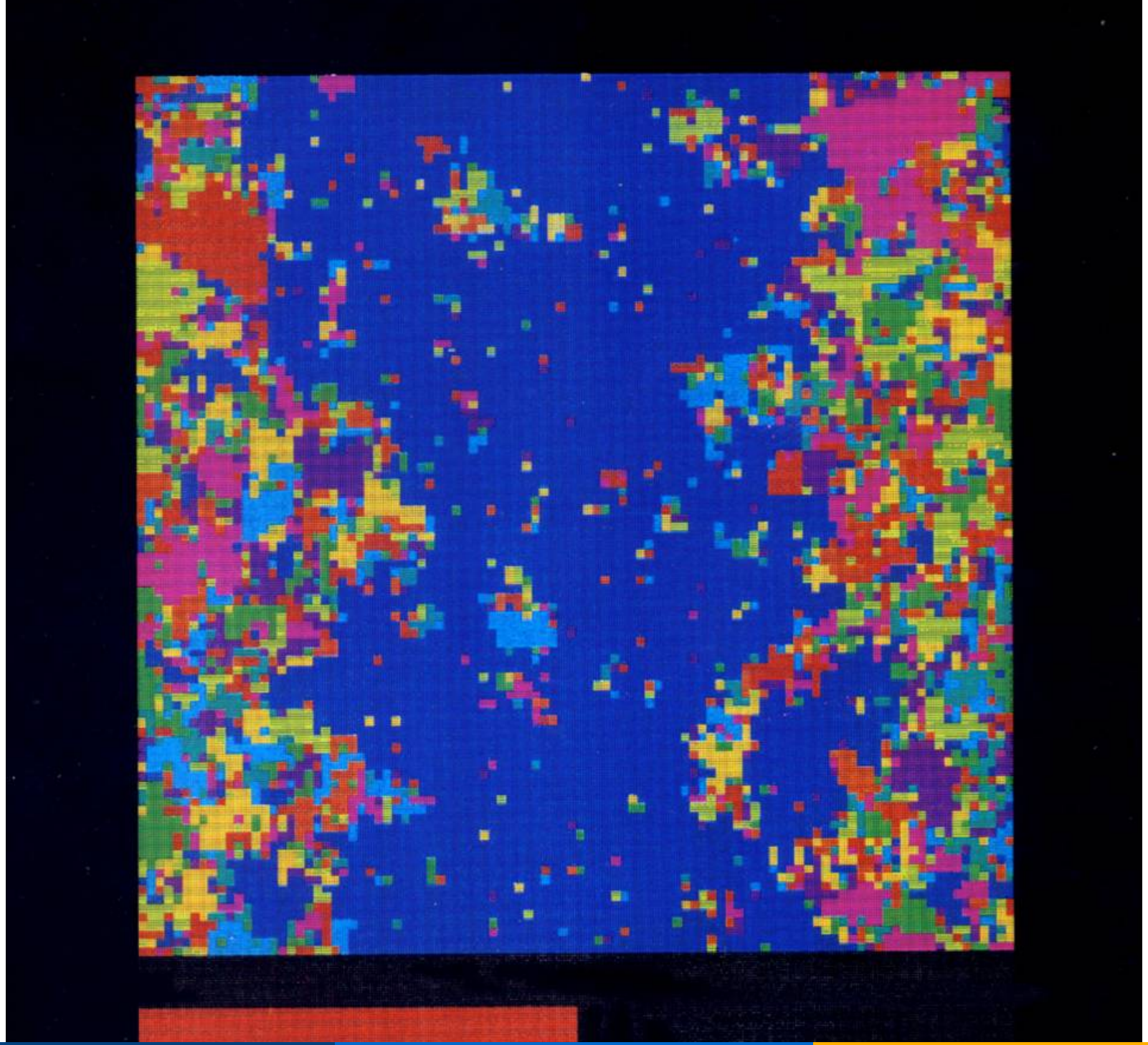
Weights are not umbrella potentials $w(q_1, \ldots, q_N)$, but functions of the energy: $w(S)$, here S energy.

Weights are obtained in one (plus one) step(s).

2D 10-state Potts model:

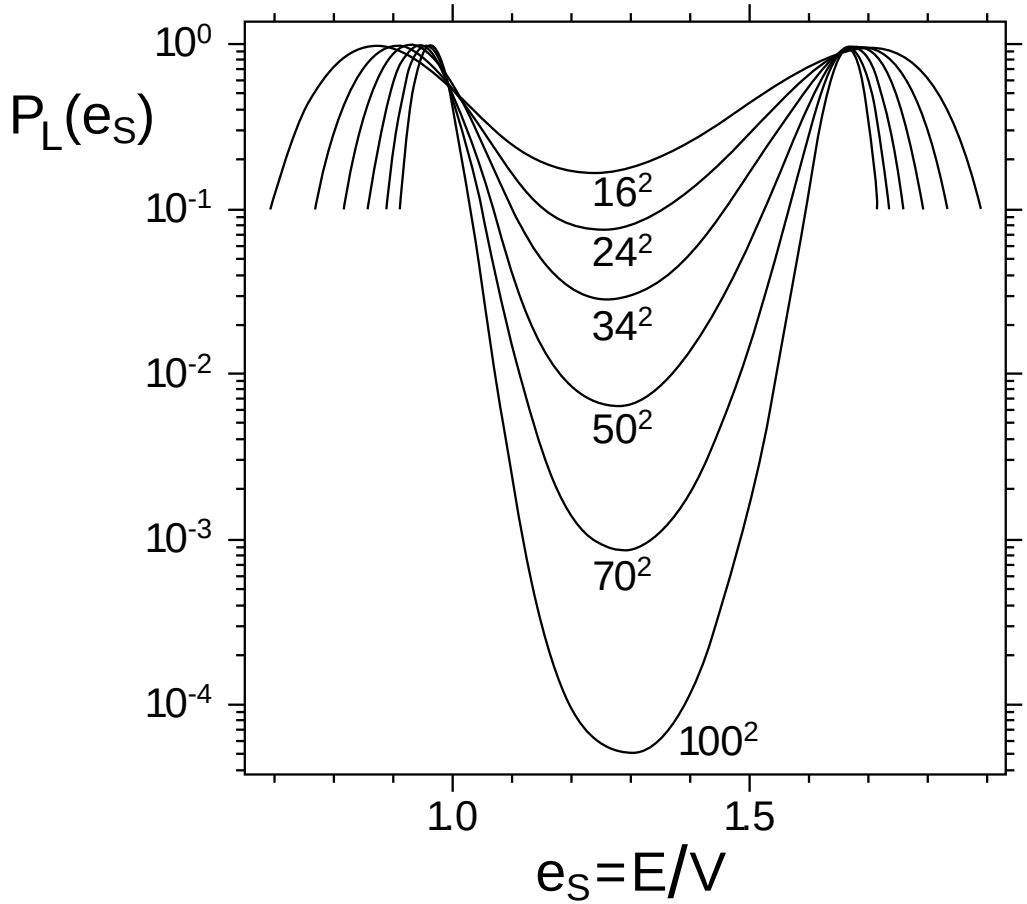






Energy per spin histogram

allows then for accurate estimates of interface tension



Back to Chandler's Case

PHYSICAL REVIEW B VOLUME 47, NUMBER 1 1 JANUARY 1993-I

Simulation of an ensemble with varying magnetic field: A numerical determination of the order-order interface tension in the $D=2$ Ising model

B. A. Berg and U. Hansmann
Department of Physics, The Florida State University, Tallahassee, Florida 32306
and Supercomputer Computations Research Institute Tallahassee, The Florida State University, Tallahassee, Florida 32306

T. Neuhaus
Fakultät für Physik, Universität Bielefeld, D-4800 Bielefeld, Federal Republic of Germany
(Received 13 September 1991; revised manuscript received 30 June 1992)

In analogy with a recently proposed multicanonical ensemble we introduce an ensemble where the partition function is simulated with a term in the action containing a varying magnetic field. Using this ensemble we demonstrate on lattices with periodic boundary conditions that it is possible to enhance the appearance of order-order interfaces by many orders of magnitude. To perform a stringent test of the method we consider the $D=2$ Ising model at $\beta=0.5$ and simulate square lattices up to size 100×100. By a finite-size scaling analysis, the order-order interface tension per unit area is obtained. Our best infinite-volume extrapolation is in excellent agreement with Onsager's exact result.

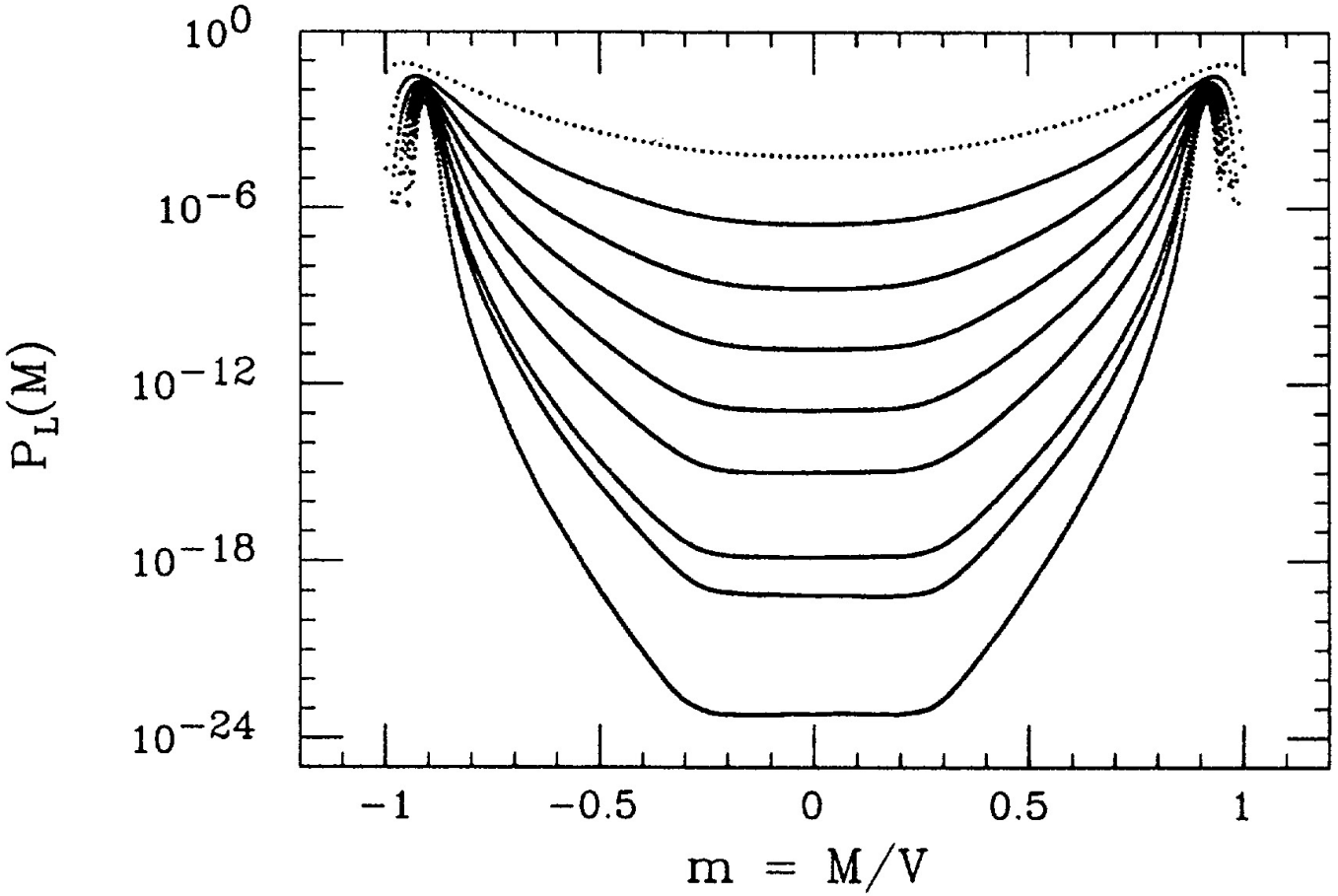


FIG. 1. Boltzmann probability distributions $P_L(M)$ for the magnetization.

$L = 10,\ 20,\ 30,\ 40,\ 50,\ 60,\ 74,\ 80,\ 100.$

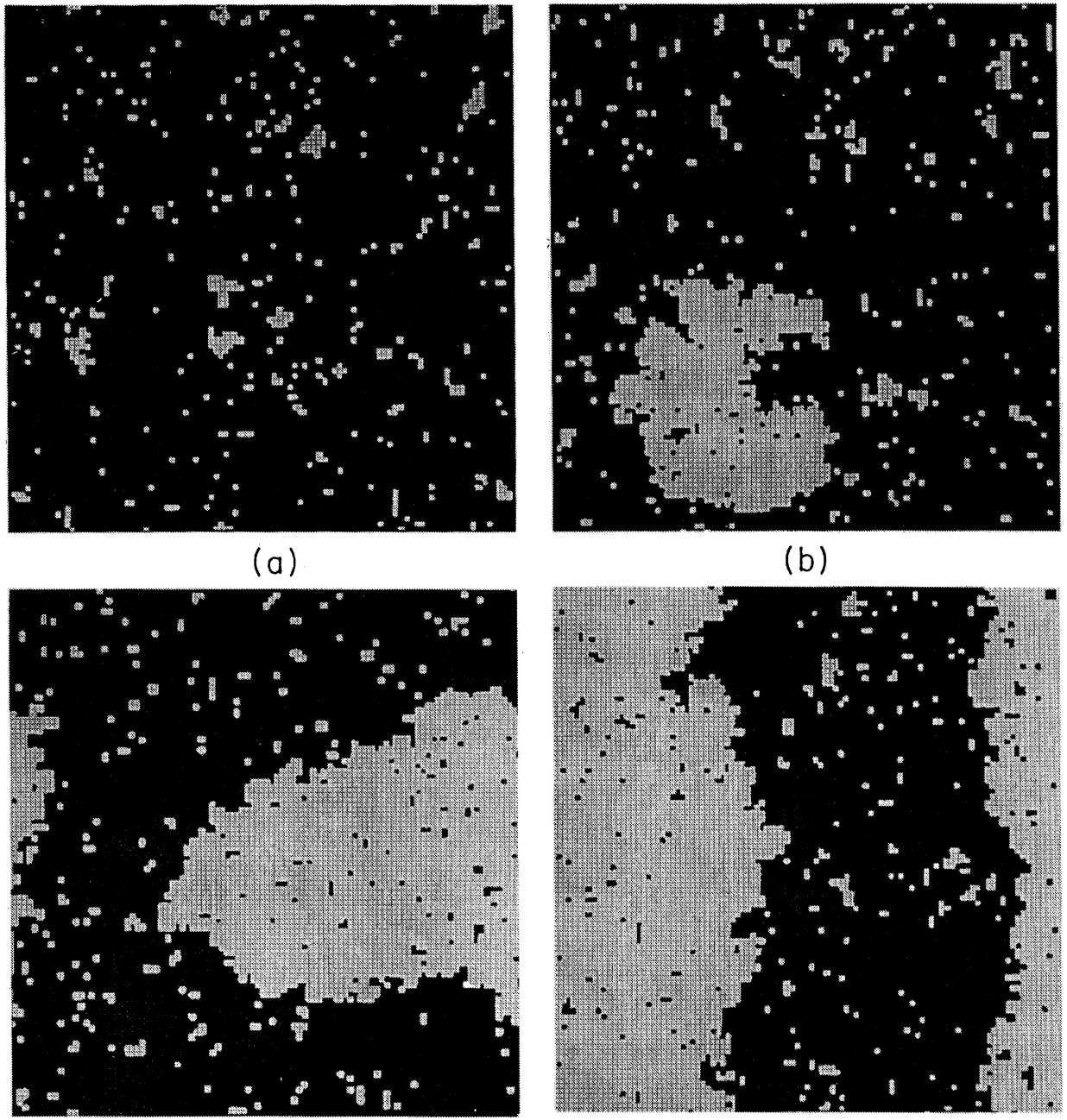
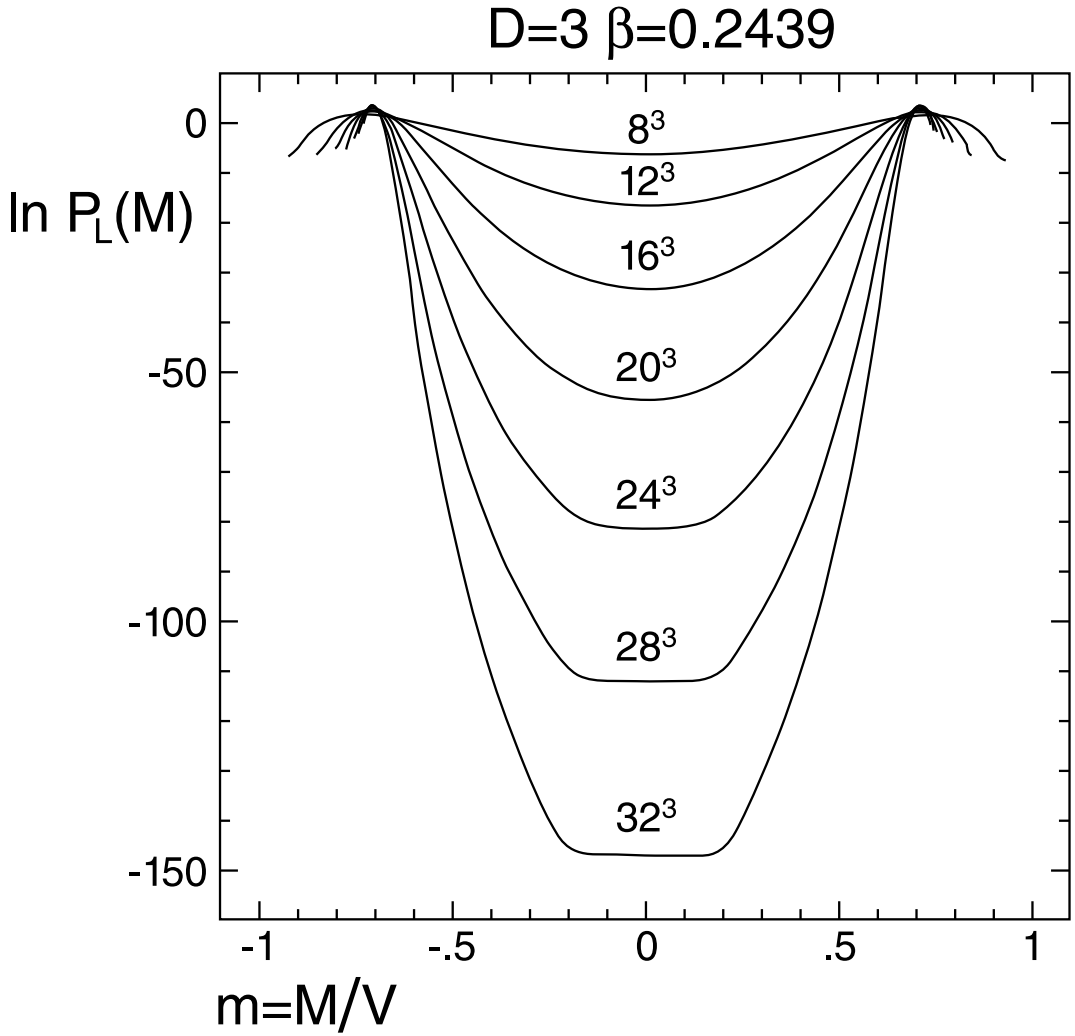


FIG. 4. (a) Configuration at $m \approx -0.9$ on the 100×100 lattice. (b) Configuration at $m \approx -0.6$ on the 100×100 lattice. (c) Configuration at $m \approx -0.4$ on the 100×100 lattice. (d) Configuration at $m \approx 0$ on the 100×100 lattice.

3D Ising model

Berg, Hansmann, and Neuhaus, Z. Phys. B 90 (1993) 229-239



10^{60} improvement over simulation with Boltzmann weights!

New Horizons

VOLUME 69, NUMBER 15 PHYSICAL REVIEW LETTERS 12 OCTOBER 1992

New Approach to Spin-Glass Simulations

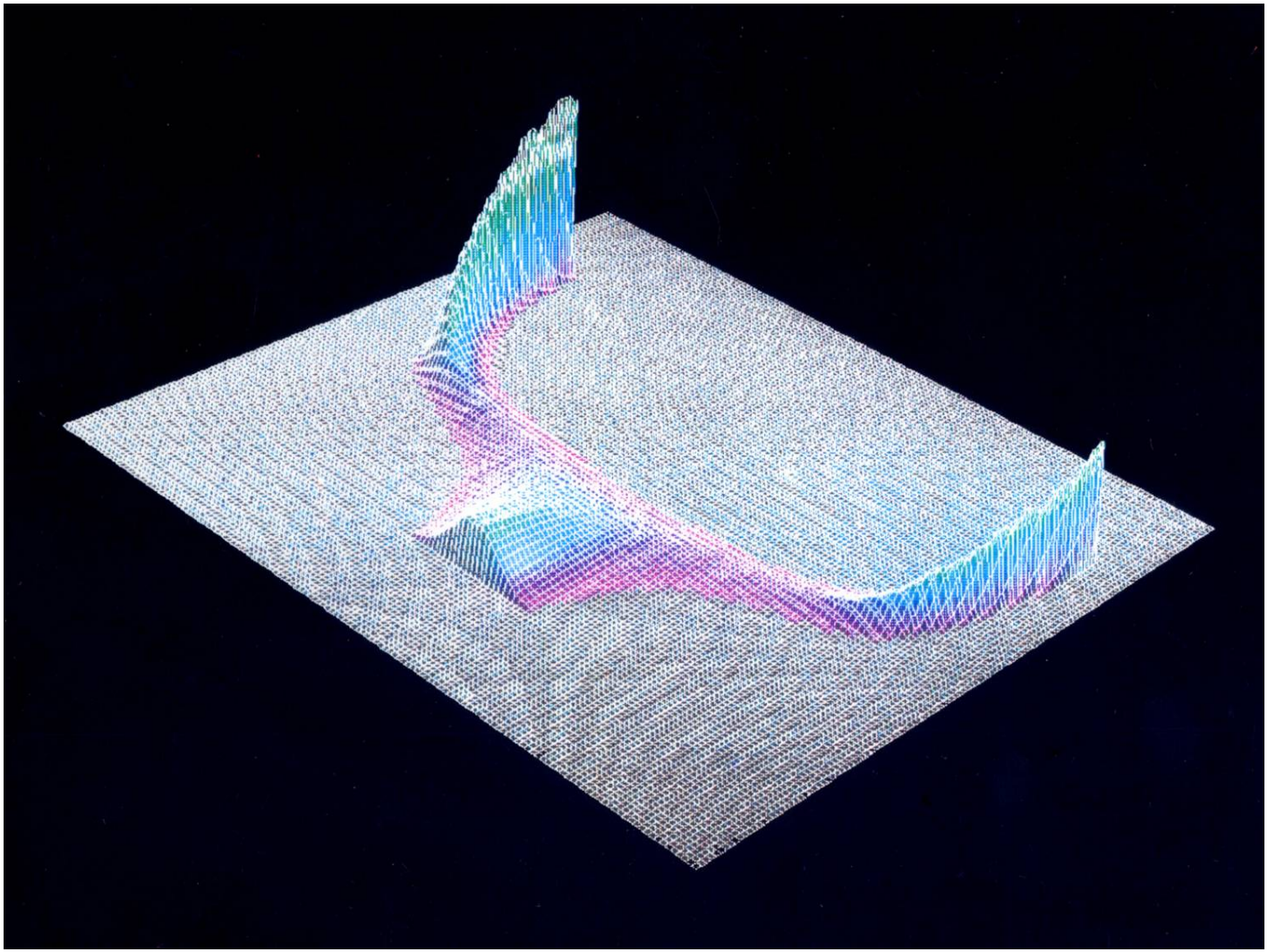
Bernd A. Berg [1),(2)] and Tarik Celik [1),(a)]
[1)] *Supercomputer Computations Research Institute, Tallahassee, Florida 32306*
[2)] *Department of Physics, The Florida State University, Tallahassee, Florida 32306*
(Received 13 April 1992)

We present a recursive procedure to calculate the parameters of the recently introduced multicanonical ensemble and explore the approach for spin glasses.

We suggest that in a large class of situations, in particular those where canonical simulations face severe ergodicity problems, it is more efficient to reconstruct the Gibbs ensemble form a simulation of a multicanonical ensemble [7] than to simulate is directly.

[Citation End].

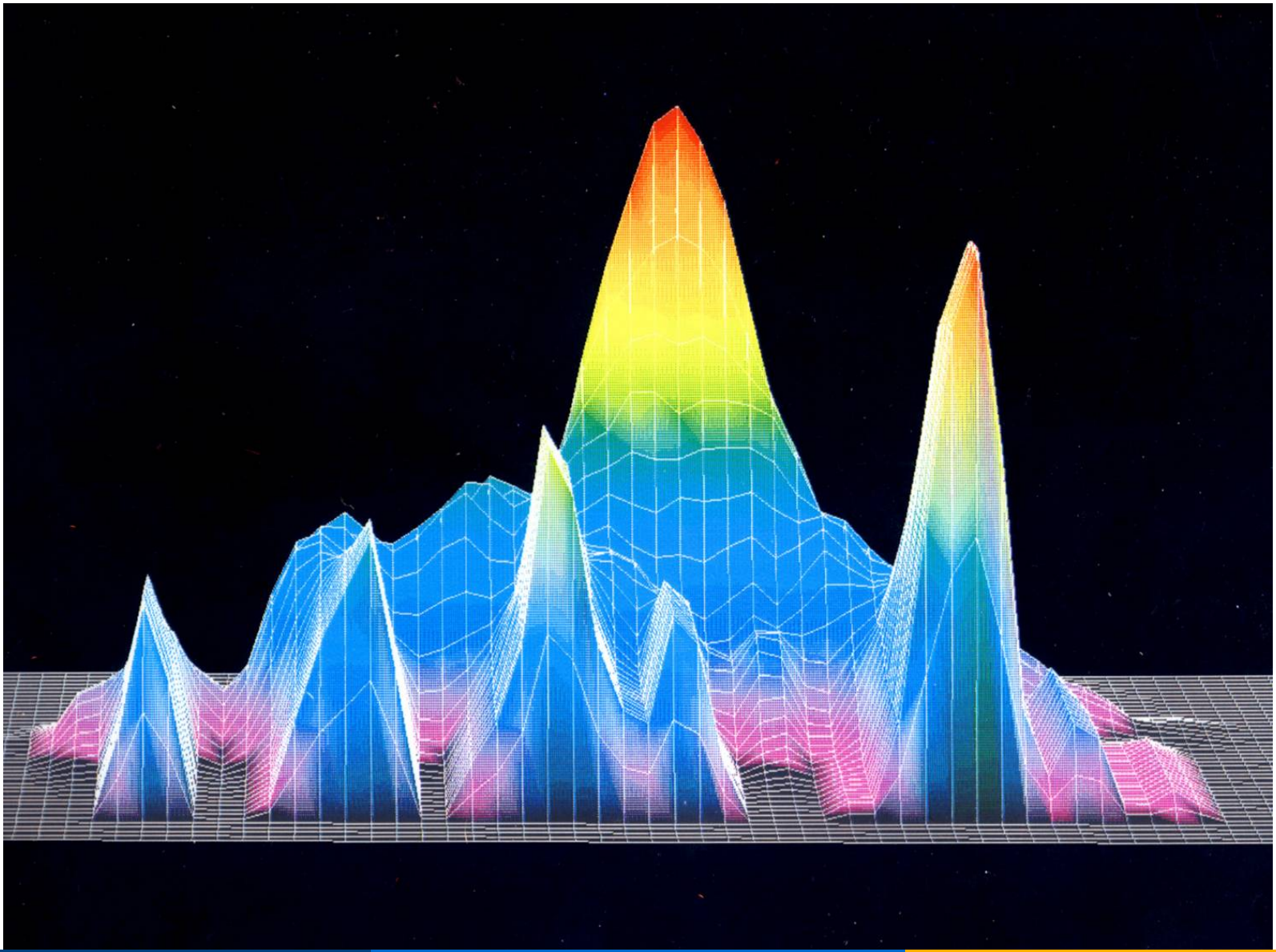
Illustration of Rugged Free Energy Landscapes



Distribution of the 2D Ising Model Magnetization versus Temperature.

One Edwards-Anderson Ising (EAI) Spin Glass Realization:

Parisi order parameter distribution versus temperature.



Exercise BB2: Explain why patching windows cannot work well for exploring these free energy landscapes.

Berg-Celik concluding remarks:

> Our results make clear that the multicanonical approach is certainly a relevant enrichment of the options one has with respect to spin-glass simulations. The similarities of spin glasses to other problems with conflicting constraints [13] suggest that multicanonical simulations may be of value for a wide range of investigations: optimization problems like the traveling salesman, neural networks, protein folding, and others.

Jülich Conference, Berg, Int. J. Mod. Phys. C 3 (1992) 1083-1098: Review up to then.

Prediction of Peptide Conformation by Multicanonical Algorithm: New Approach to the Multiple-Minima Problem

Ulrich H.E. Hansmann[1]† and Yuko Okamoto[2]*†

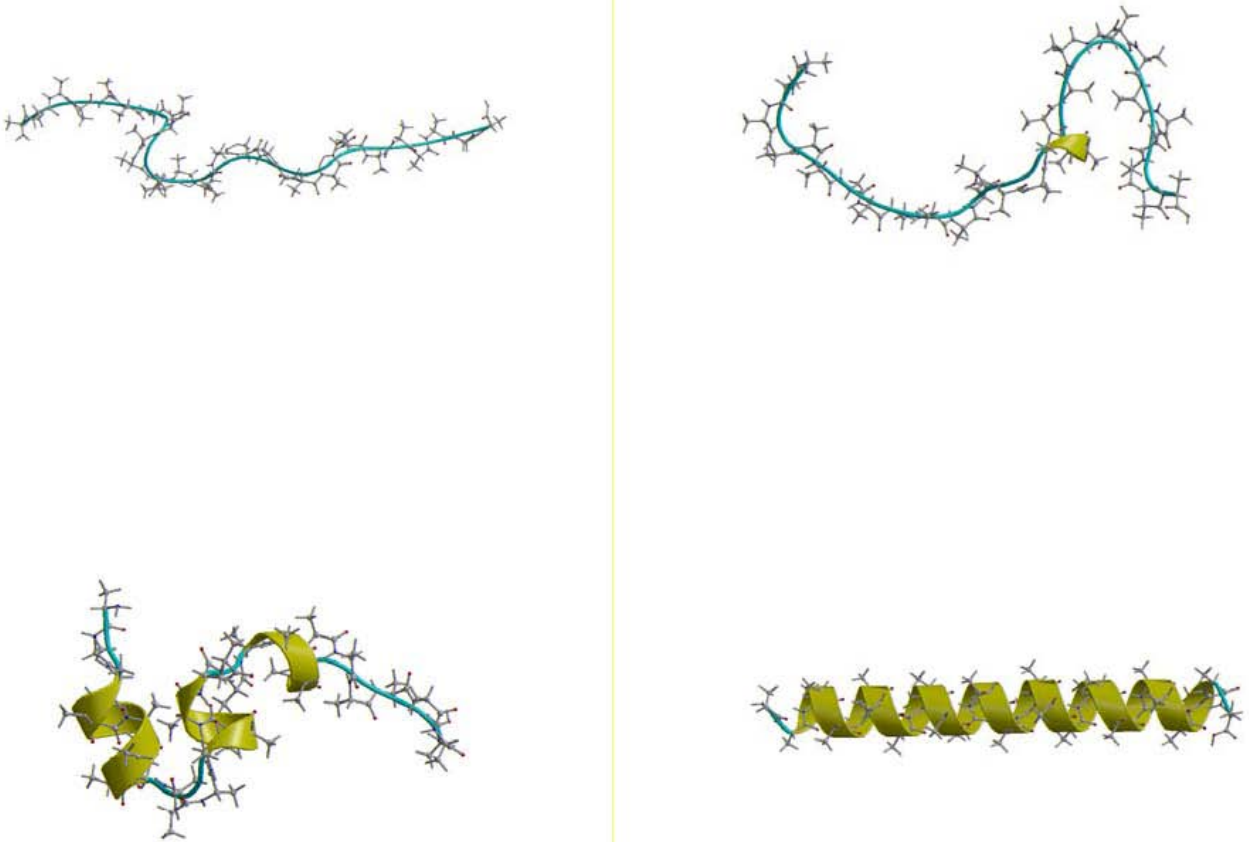
[1]*Department of Physics and Supercomputer Computations Research Institute, The Florida State University, Tallahassee, Florida 32306, and* [2]*Stanford Linear Accelerator Center, Stanford University, Stanford, California 94309*

Received 27 January 1993; accepted 27 May 1993

We apply a recently developed method, the multicanonical algorithm, to the problem of tertiary structure prediction of peptides and proteins. As a simple example to test the effectiveness of the algorithm, met-enkephalin is studied and the ergodicity problem, or multiple-minima problem, is shown to be overcome by this algorithm. The lowest-energy conformation obtained agrees with that determined by other efficient methods such as Monte Carlo simulated annealing. The superiority of the present method to simulated annealing lies in the fact that the relationship to the canonical ensemble remains exactly controlled. Once

Their review The Generalized-Ensemble Approach for Protein Folding Simulations, Ann. Rev. Comp. Phys. 6, 129-157 (1999) coined the name Generalized Ensembles.

Poly-alanine:



Hansmann and Okamoto, J. Chem. Phys. 110, 1267-1276 (1999); erratum 111, 1339.

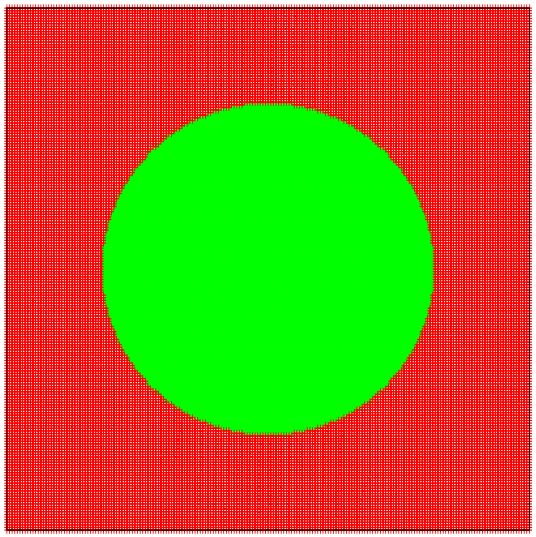
MUCA Performance

Slowing down in units of updates:

1. Optimum: $\propto N^2$ perfect random walk in the energy.
2. 2D 10-state Potts model (transition range): Effective $\propto N^{2.325\,(10)}$ observed (Berg and Neuhaus 1992).
3. 2D EAI spin glass: $\propto N^{3.2\,(2)}$ observed (Berg and Celik 1992).

Still exponential slowing down expected due to hidden energy barriers.

Even for simple first order phase transitions: Hager and Neuhaus, J. Stat. Phys. 113 (2003) 47–83: $\Leftrightarrow$

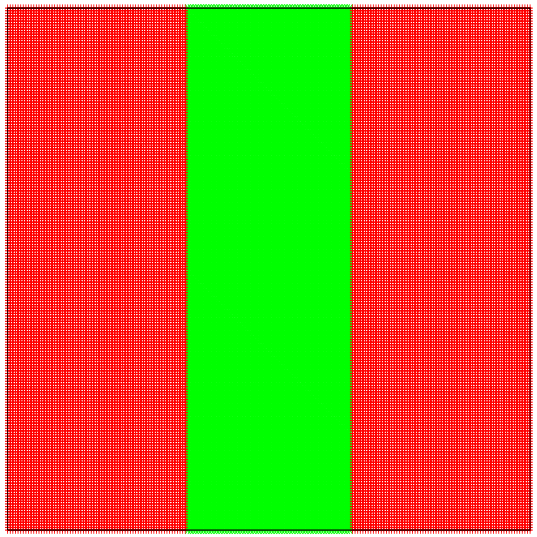


Transitions

only through

rare or suppressed

configurations.



Replica Exchange Method and Expanded Ensembles

The late 1980s and early 1990s saw a flurry of innovative activities on MCMC methods.

Swendsen and Wang, Phys. Rev. 58 (1987) 86-88; Wolff Phys. Ref. Lett. 62 (1989) 361-363: Cluster algorithms.

Ferrenberg and Swendsen, Phys. Rev. Lett. 61 (1988) 2635-2638, 63 (1989) 1658: Popularized histogram reweighting of canonical MCMC simulations and focused them on determinations of finite-Volume maxima of divergent quantities in studies of phase transitions.

Ferrenberg and Swendsen, Phys. Rev. Lett. 63 (1989) 1195-1198; Alves, Berg and Villanova, Phys. Rev. B 41 (1990) 383: Multi-histogram reweighting.

Question: Which of the earlier references relate to reweighting?

Replica Exchange:

Swendsen and Wang, Phys. Rev. 57 (1986) 2607-2609: Introduced (too) general replica exchange method.

Frantz, Freemann, and Doll, J. Chem. Phys. 93 (1990) 2769-2784: Jump-walking feeds replica from a high- into a low-temperature simulation. Does not fulfill balance (not replica exchange).

Geyer, in Proceedings of the 23rd Symposium on the Interface, Keramidas (editor), Interface Foundation, Fairfax, Virginia (1991) 156-163: Multiple Markov chains (the same as replica exchange).

Hukusima and Nemoto, J. Phys. Soc. Japan 65 (196) 1604–1608: Replica exchange. Often called parallel tempering when temperatures are exchanged.

Hansmann, Chem. Phys. Lett. 281 (1997) 1267-1276: First application to a biomolecule (Met-Enkephalin).

Parallel Tempering:

For $\beta_1 < \beta_2 < ... < \beta_n$ consider the joint partition function

$$Z = \prod_{k=1}^{n} \left\{ \prod_{i=1}^{N} \frac{(2\pi\, m_i\, k_B\, T_k)^{3/2}}{N!} \int d^3 x_i^k \, \exp\left[-\beta_k\, U(\vec{r}_1^{\,k}, \ldots, \vec{r}_N^{\,k}) \right] \right\}$$

of a molecular system, where the momenta are integrated out. Swaps between (normally neighboring) temperatures are proposed and accepted with the Metropolis probability

$$P_{\rm acpt} = \min\left\{ 1,\, \exp\left[+(\beta_k - \beta_l) \left(U(\vec{r}_1^{\,k}, \ldots, \vec{r}_N^{\,k}) - U(\vec{r}_1^{\,l}, \ldots, \vec{r}_N^{\,l}) \right) \right] \right\}$$

so that balance holds. The differences $\beta_k - \beta_l$ have to be chosen small enough to ensure reasonable acceptance rates.

Well-suited for parallel processing.

Other advantages: To be discussed.

Generalized ensembles can be exchanged as well, ...

Do not confuse with

Expanded Ensembles:

Lyubartsev, Martsinovski, Shevkanov, and Vorontsov- Velyaminov, J. Chem. Phys 96 (1992) 1776-1783.

Marinari and Parisi, Europhys. Lett. 19 (1992) 451-458: Simulated tempering.

Weights

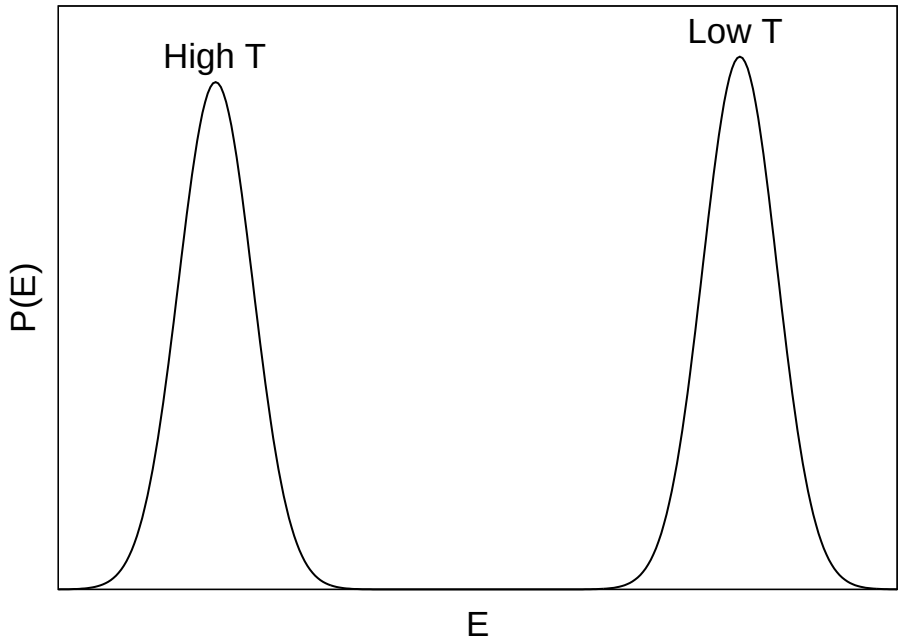
$$\exp(-\beta_m E_k + g_m), \quad m = 1, \ldots, M, \quad k \text{ configuration},$$

and m is considered a new dynamical variable. Moves

$$m \to m \pm 1$$

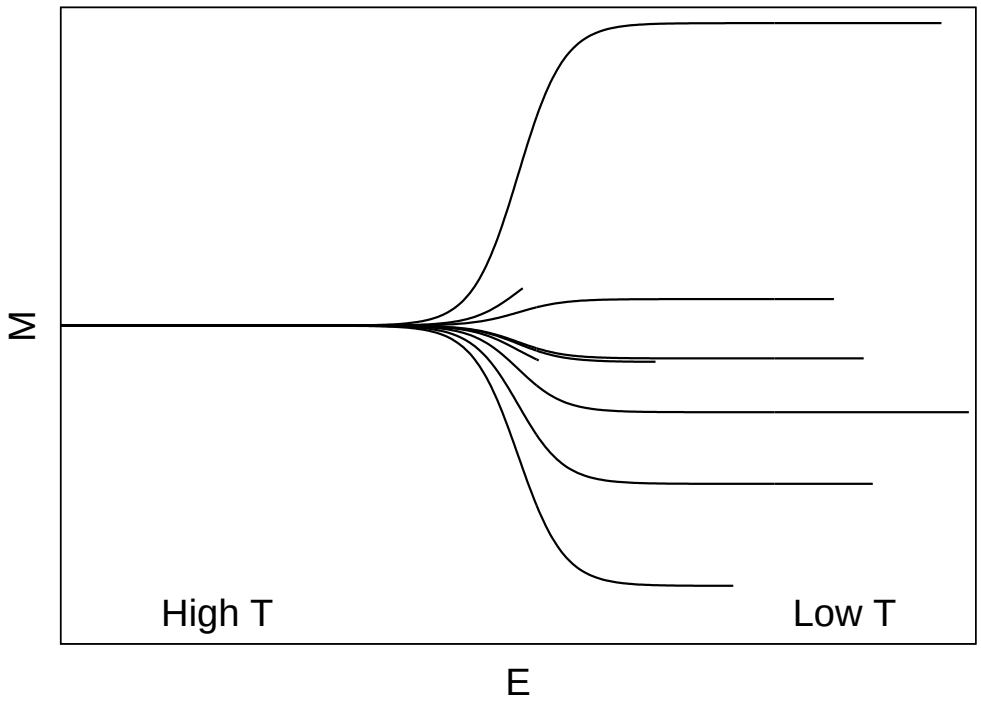
are proposed and rejected/accepted according to Metropolis. The constants g_m need to be adjusted to get reasonable acceptance rates.

How to overcome a free energy barrier?



1. Flatten the barrier. This requires that the barrier is explicit in a known reaction coordinate. Probabilities of transition states are then greatly enhanced.
2. Jump the barrier via parallel or simulated tempering. Initial low-temperature configuration(s) problematic! Probabilities of transition states stay unchanged (tiny). Parallel tempering will not estimate entropy and free energy across the barrier.

Exploration of low temperature configurations in a rugged free energy landscape:



If there are no major barriers in the energy variable, all of the introduced methods explore all low temperature branches and yield their relative probabilities as well as the global energy minimum.

Due to barriers in M: Large performance enhancement versus low temperature Boltzmann simulations.

Example (left over from my last MCMC course at FSU):

Parallel tempering with 8 processes for the 2D 10-state Potts model on a 20×20 lattice. Some parameters:

$$h_q = 0,\ \beta_{\min} = 0.65,\ \beta_{\max} = 0.75,\ (\text{note } \beta_t = 0.71303\ldots).$$

	Parallel Tempering		Canonical Simulation	
β	$\tau_{\rm int}(E)$	$\tau_{\rm int}(M)$	$\tau_{\rm int}(E)$	$\tau_{\rm int}(M)$
0.7133...	$761\,(21)$	$434\,(13) \times 10$	$1199\,(51)$	$181\,(10) \times 10^2$
0.7261...	$54.0\,(4.9)$	$611\,(19) \times 10$	$13.63\,(22)$	$979\,(48) \times 10^4$

Units are sweeps. Assembled statistics: 10×2^{20} per replica for PT, 10×2^{20} cano at $\beta = 0.7133...$, $10^3 \times 2^{22}$ cano at $\beta = 0.7261...$.

Exercise BB3: Explain why the integrated autocorrelation times $\tau_{\rm int}(E)$ and $\tau_{\rm int}(M)$ behave so differently.

Summary and Conclusions:

This talk presented use of generalized ensembles in MCMC simulations close to the chronological order of their development. Obviously, my mini-review cannot be complete, I had to make omissions and may have overlooked some relevant papers.

A message is that mastering simulation methods is not the trivial part of a biophysical, chemical or physical MCMC study. Astronomically large efficiency factors can float around between making it right or wrong.

There have been hundreds of papers refining and improving the methods outlined here. It is practically impossible to follow up on all of them, all the more to verify them. Let me just pick four points:

1. Wang-Landau, PRL 86 (2001) 361-363: Introduced a now widely used algorithm, which can be employed as a recursion to get working estimates of the MUCA weights.
2. Trebst, Huse and Troyer, PRE 70 (2004) 046710: Optimized the MUCA weight with respect to cycling times. Improvement factors appear to be around 3 to 5 compared to relying on flat histograms.
3. Hamiltonian replica exchange, alchemical transitions and related ideas have been studies by a number of authors, including Uli Hansmann and Wei Yang, who are here and may tell you more.
4. Data analysis problems should not be neglected. For instance, reweighting often requires a logarithmic coding, because factors just get too large to be allowed in Fortran or C. That is a good occasion to advertise my book for your library:

This book teaches modern Markov chain Monte Carlo (MC) simulation techniques step by step. The material should be accessible to advanced undergraduate students and is suitable for a course. It ranges from elementary statistics concepts (the theory behind MC simulations), through conventional Metropolis and heat bath algorithms, autocorrelations and the analysis of the performance of MC algorithms, to advanced topics including the multicanonical approach, cluster algorithms and parallel computing. Therefore, it is also of interest to researchers in the field. The book relates the theory directly to Web-based computer code. This allows readers to get quickly started with their own simulations and to verify many numerical examples easily. The present code is in Fortran 77, for which compilers are freely available. The principles taught are important for users of other programming languages, like C or C++.

Markov Chain Monte Carlo Simulations and their Statistical Analysis

With Web-Based Fortran Code

Key Features

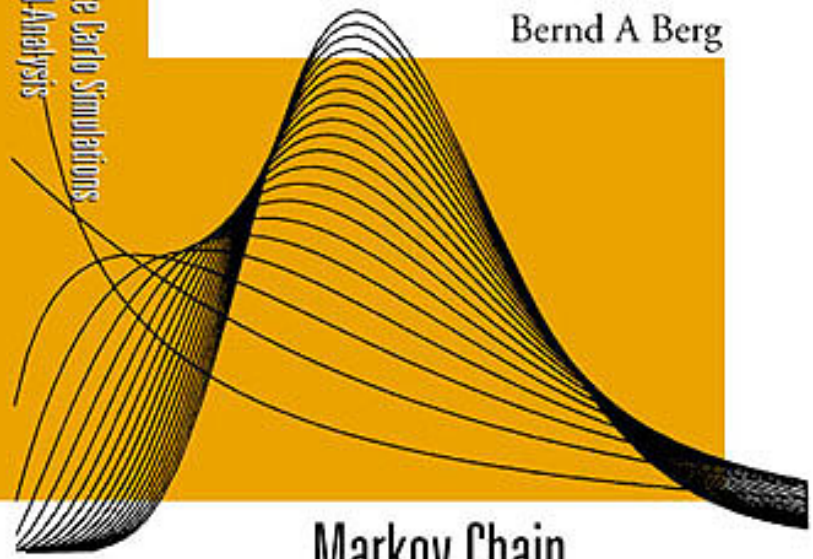
- Teaches Markov chain Monte Carlo simulations step by step
- Ranges from elementary statistics concepts to advanced Markov chain Monte Carlo simulations
- The only book on Monte Carlo simulations for which Web-based computer code allows the reader to verify many numerical examples easily

World Scientific
www.worldscientific.com
5602 hc

ISBN 981-238-935-0
9 789812 389350

Markov Chain Monte Carlo Simulations and their Statistical Analysis

Berg



Bernd A Berg

Markov Chain Monte Carlo Simulations and Their Statistical Analysis

With Web-Based Fortran Code

Really huge improvement will get attention, because they would be enabling methods for simulations which otherwise cannot be done. However, attention is not enough. Required are independent verifications, which too often never happen.

Where are we stuck?

Apparently at many fronts, but we focus here on the simulation of well-defined models (all-atom in biophysics, classical spin glasses in statistical physics, and so on).

1. Obviously there are often hidden barriers and we are unable to find reaction coordinates in which the barriers become explicit.
2. The optimal performance of the discussed methods is limited by a diffusive process. For large systems that can still be far too slow. In some situations that is greatly improved by the collective updating of cluster algorithms, but their application range has remained fairly limited.

Challenges: Many. Combinations with MD simulations as for instance pioneered by Wei Yang appear to me one of the promising roads.