Computer Simulations of Generalized Ensembles

Bernd A. Berg

Florida State University

GBA Theoretical Chemistry Lecture Series, Boston, 11/29/2006

Overview

1. Reweighting
2. Umbrella Sampling
3. Binder's Method for Estimating Interface Tensions
4. Multicanonical Simulations
   - First order phase transitions
   - Groundstates and rugged free energy landscapes
   - Simulations of peptides
   - How to Get the Weights (Wang-Landau Recursion)?
   - MUCA Performance
   - Second order phase transitions and cluster algorithms
5. Replica Exchange Method (Parallel Tempering)
   - Molecular dynamics
   - Protein folding
   - Hamiltonian replica exchange and λ scaling

The talk will make some effort to present each topic in its chronological order (talk and references will be posted on the web, but I may not be aware of all relevant contributions).

Reweighting – Single Histogram Method

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

1959: A first attempt to calculate the partition function by MCMC is made in a paper by Salsburg, Jacobson, Fickett, and Wood, using a method called in the modern language reweighting. As noticed by the authors their method is restricted to very small lattices.

1967: McDonald and Singer used reweighting to evaluate physical quantities over a small range of temperatures. Thereafter the approach appeared to be dormant.

1982: Reweighting was rediscovered by Falcioni, Marinari, Paciello, Parisi and Taglienti who focused on calculating complex zeros of the partition function.

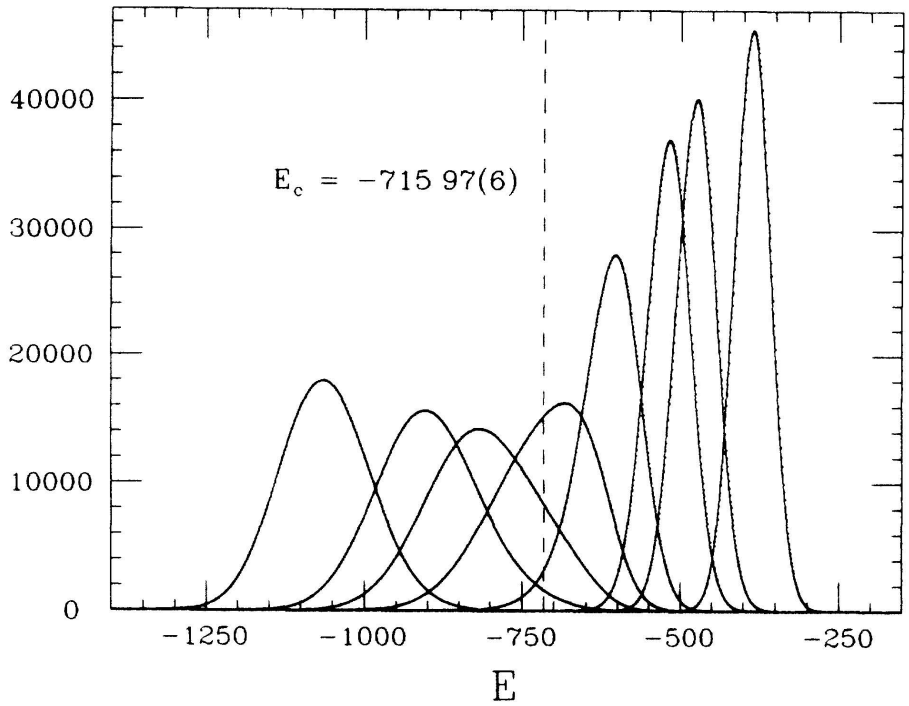
1988: Ferrenberg and Swendsen formulated clearly for what the method is good, and for what it is not: Reweighting allows to focus on maxima or minima of appropriate observables, but not to cover a finite temperature range in the infinite volume limit.

Reweighting – Multi Histogram Method

1972: Valleau and Card introduced the use of overlapping bridging distributions and called their method multistage sampling.

1989: Ferrenberg and Swendsen proposed a recursion for creating a joint distribution.

1990: Alves, Berg and Villanova patched histograms by minimizing χ^2 for the reweighted overlap: $\nu = 0.6303\,(14)$ for 3D Ising model ($\nu = 0.6301\,(4)$ nowadays best, Pelissetto & Vicari 2002). 14^3 lattice:



Umbrella Sampling

Patching histograms of canonical simulations faces a number of limitations:

1. Bottlenecks in the canonical ensemble can create quasi-ergodicity.
2. There is no way to focus on canonically rare events.

1976: Torrie and Valleau introduced umbrella sampling, which copes with these difficulties by allowing arbitrary sampling distributions.

Patching of of weighting factors from umbrella or window potentials is then done (e.g., Chandler 1987). But applications remained rather limited, although MCMC simulations flourished in the 1980s.

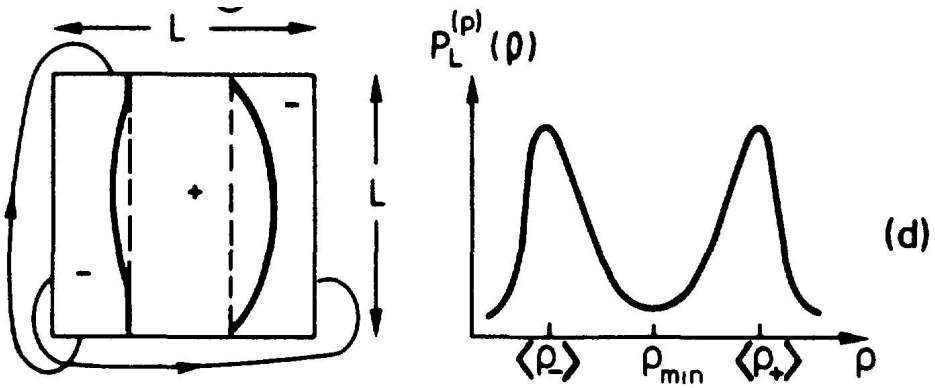
Li and Scheraga, 1988: *The difficulty of finding such weighting factors has prevented wide applications of umbrella sampling.*

Furthermore, there appeared to be no focus on identifying challenging problems, which could be overcome by using umbrella sampling.

Binder's Method for Estimating Interface Tensions

Simulations with periodic boundary conditions (Binder 1982).

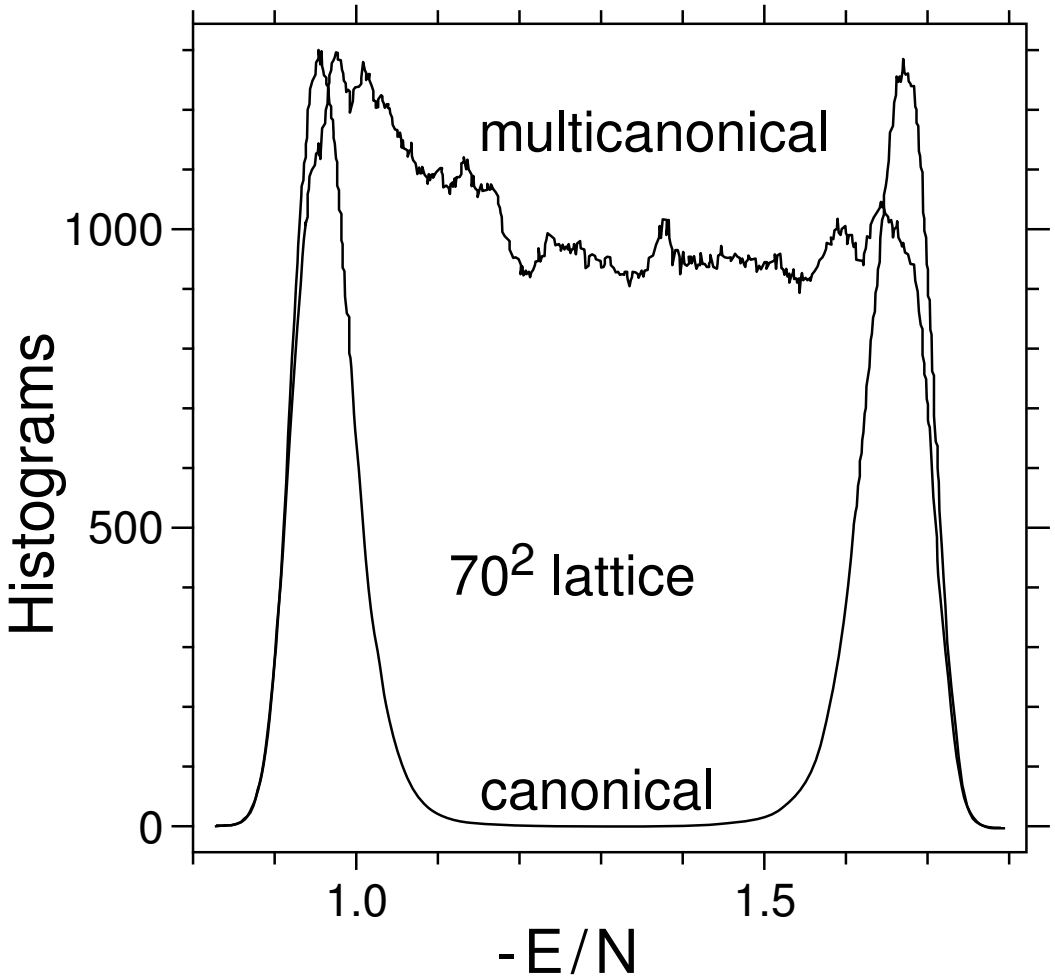
From equal heights double peaks : $$F_L^s = -\frac{1}{L^{(D-1)}} \ln\left(\frac{P_L^{\min}}{P_L^{\max}}\right) .$$



But in practice results remained pitiful: $L^{\max} = 12$ for the 3D Ising model with far off estimates like $F_\infty^s = 0.0050\,(25)$ at $\beta = 0.27227$. Reason: $P_L^{\min}$ is exponentially suppressed in the canonical ensemble.

Multicanonical Simulations (Berg and Neuhaus 1991/1992)

2D 10-state Potts model:



... achieved by sampling in an appropriate energy range $E^{\min} \leq E \leq E^{\max}$ with an approximation of the weights

$$W_{\mathrm{MUCA}}(E) = \frac{1}{n(E)} = e^{-b(E)\,E+\alpha(E)}$$

and reweighting to the canonical ensemble. Thus, the multicanonical approach requires two steps (to be detailed later):

1. Obtain a working estimate of the weights $W_{\mathrm{MUCA}}(E)$.
2. Perform a MCMC simulation with fixed weights.

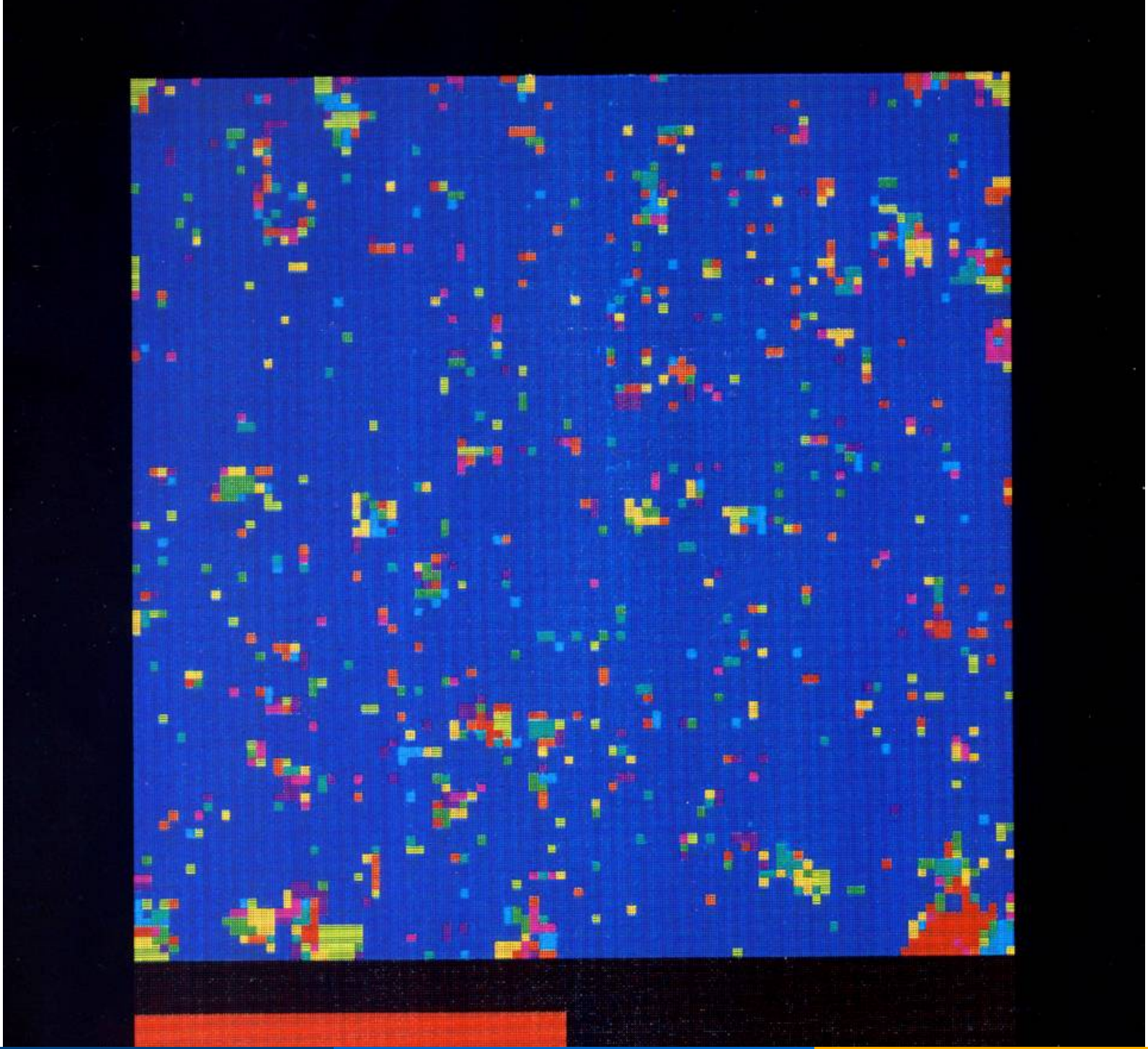
Working estimate means that the Markov chain cycles (tunnels) in the energy range:

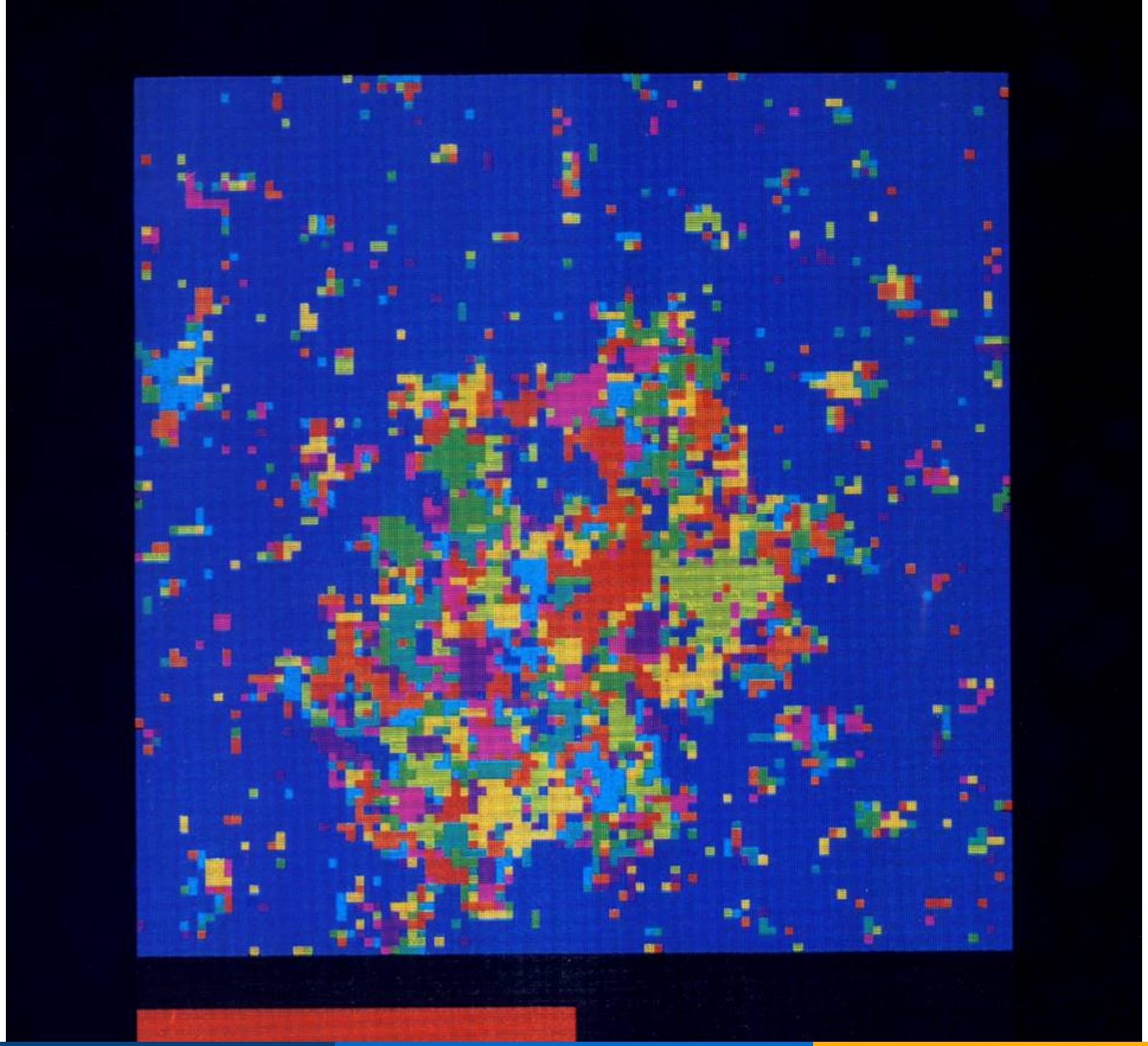
$$E^{\min} \leq E \leq E^{\max}$$

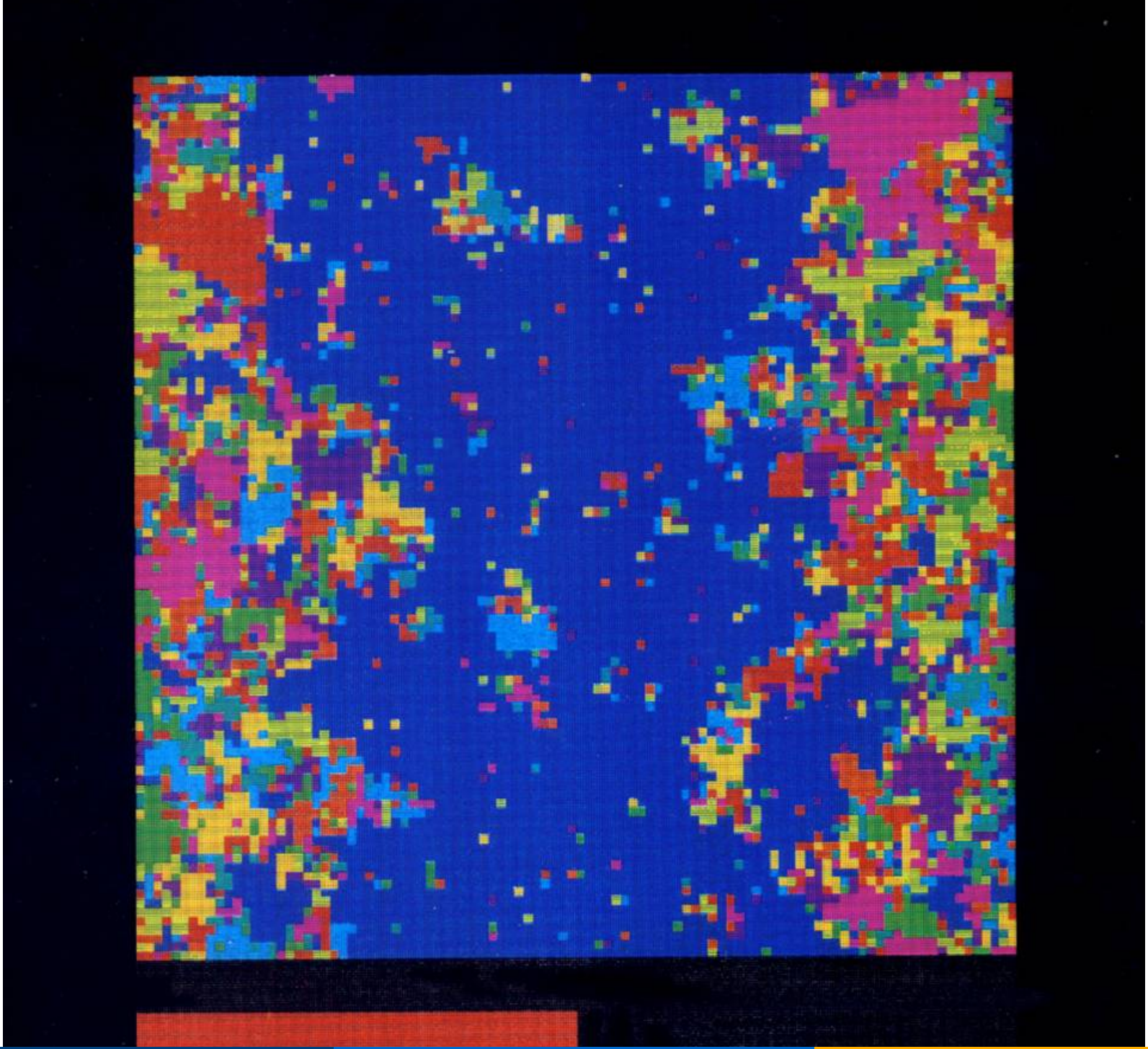
which corresponds to a canonical temperature range (multicanonical)

$$\beta^{\min} = b(E^{\max}) \leq \beta \leq b(E^{\min}) = \beta^{\max}\,.$$

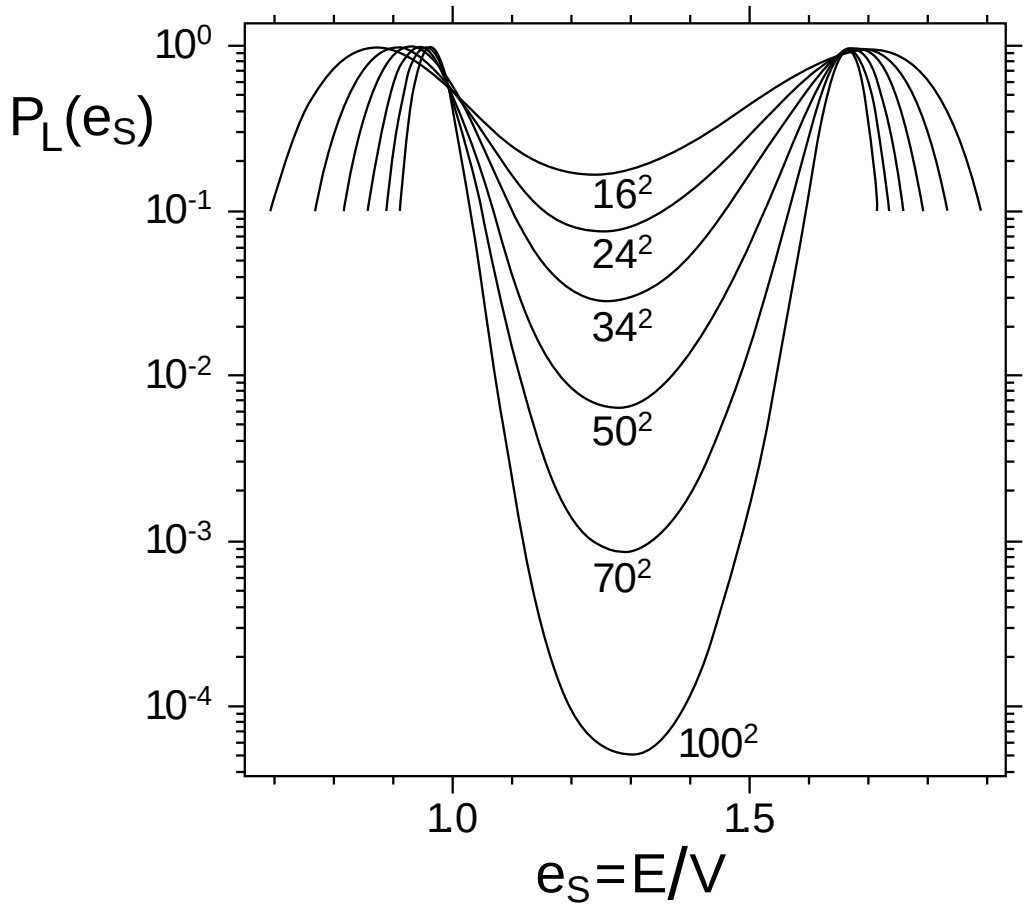
Rare configurations are now sampled (2D 10-state Potts model):

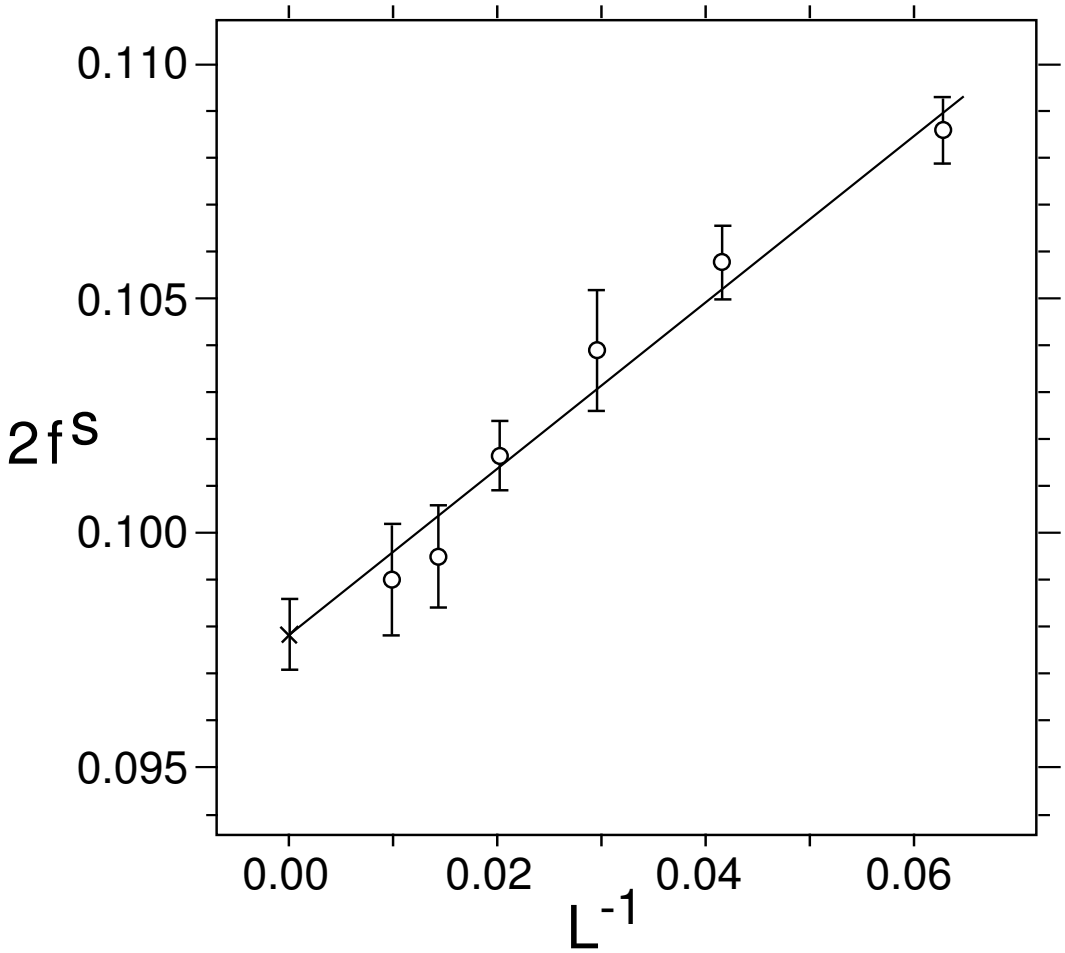






Energy per spin histogram




0.110
0.105
0.100
0.095
2f^s
0.00
0.02
0.04
0.06
L^-1

Potts models: infinite volume interface tensions estimates.

MUCA:

$$q = 10: \quad F^s_\infty = 0.09781\ (75)$$

(Berg and Neuhaus 1992, submitted July 1991)

$$q = 7: \quad F^s_\infty = 0.0241\,(10)$$

(Janke, Berg, and Katoot 1992, submitted March 1992)

Before:

$$q = 7: \quad F^s_\infty = 0.1886\,(12) \quad \text{(Boston 1989)}$$
$$q = 7: \quad F^s_\infty \approx 0.20 \quad \text{(Helsinki 1989)}$$

Potts models: infinite volume interface tensions estimates.

MUCA:

$$q = 10: \quad F^s_\infty = 0.09781\ (75)$$

(Berg and Neuhaus 1992, submitted July 1991)

$$q = 7: \quad F^s_\infty = 0.0241\,(10)$$

(Janke, Berg, and Katoot 1992, submitted March 1992)

Before:

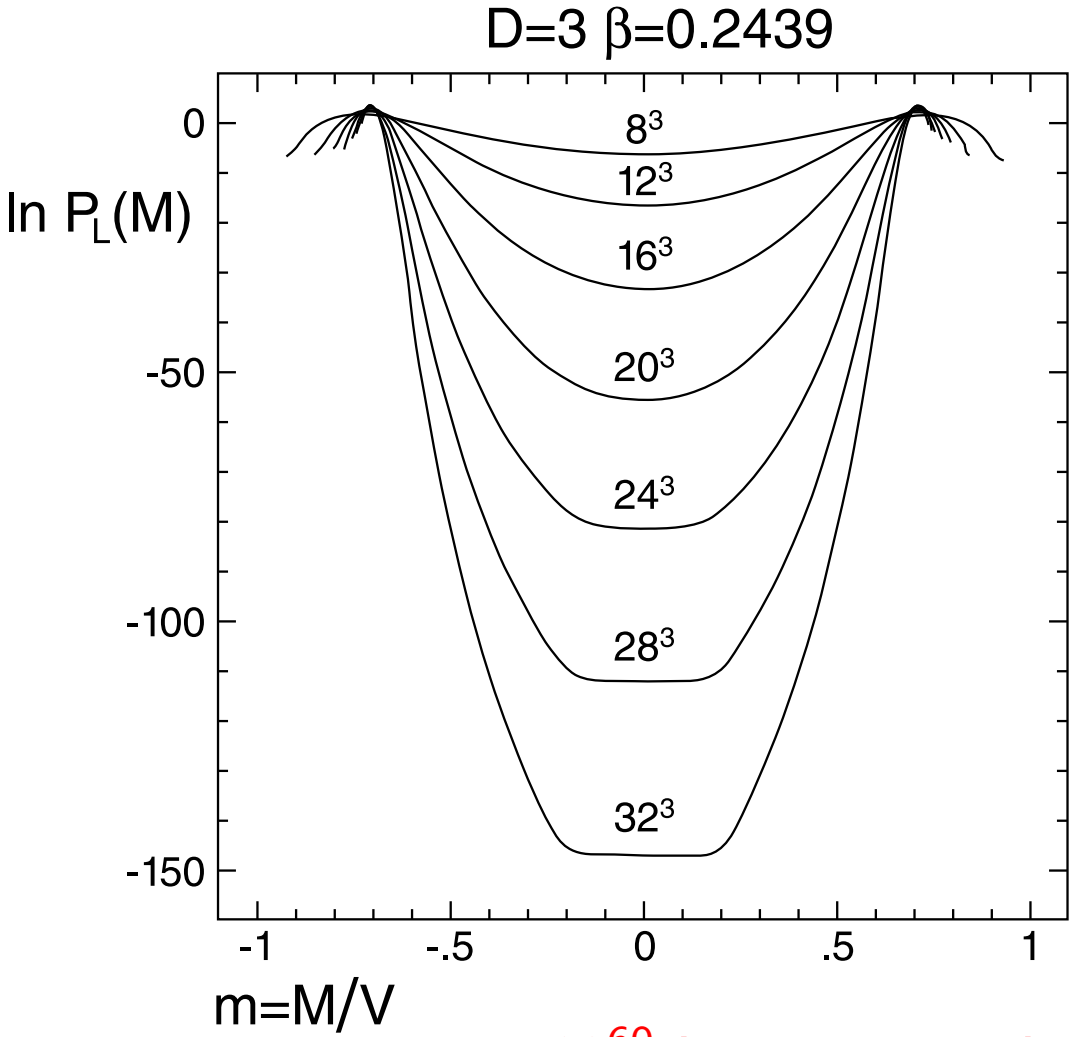
$$q = 7: \quad F^s_\infty = 0.1886\,(12) \quad \text{(Boston 1989)}$$
$$q = 7: \quad F^s_\infty \approx 0.20 \quad \text{(Helsinki 1989)}$$

Exact: (Borgs and Janke 1992, submitted September 1992)

$$q = 10: \quad F^s_\infty = 0.094701\ldots$$
$$q = 7: \quad F^s_\infty = 0.020792\ldots$$

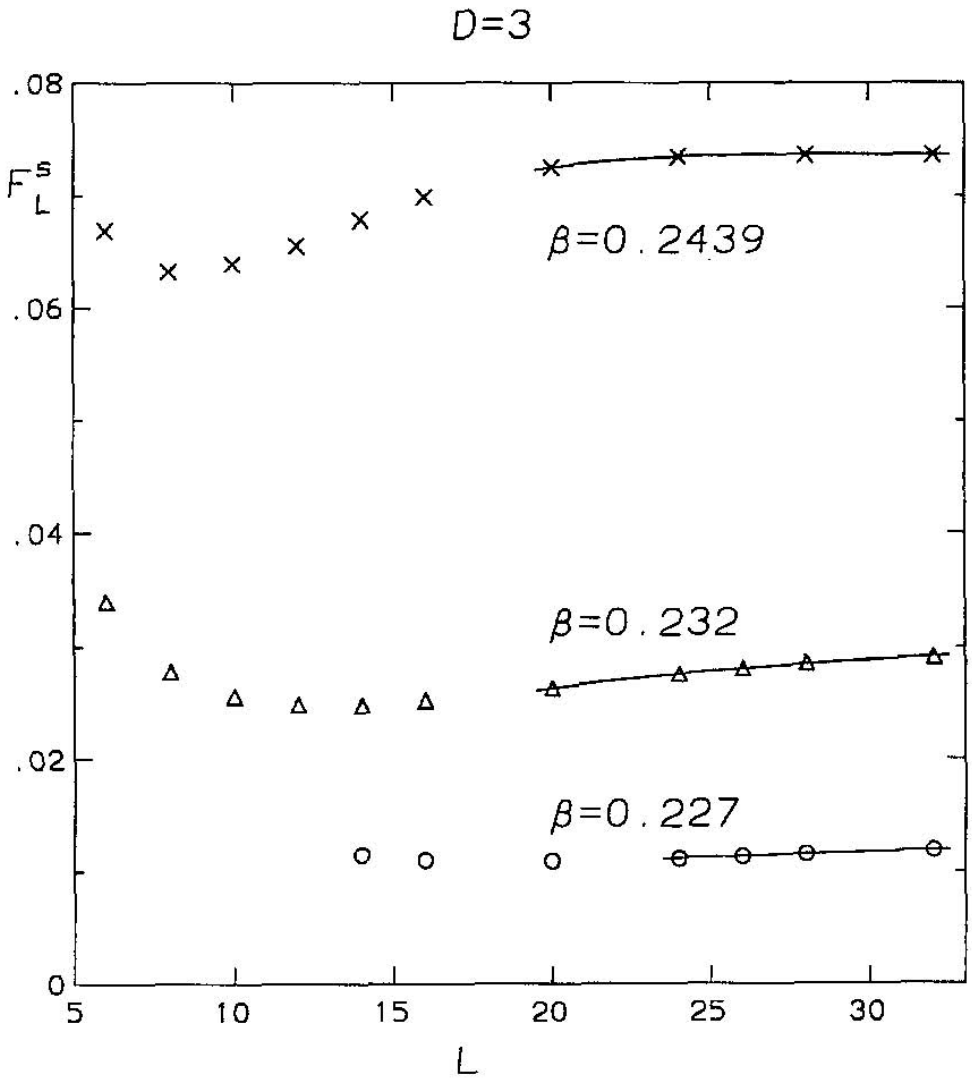
Multimagnetical Simulations – Back to the 3D Ising Model

(Berg, Hansmann, and Neuhaus, 1993)



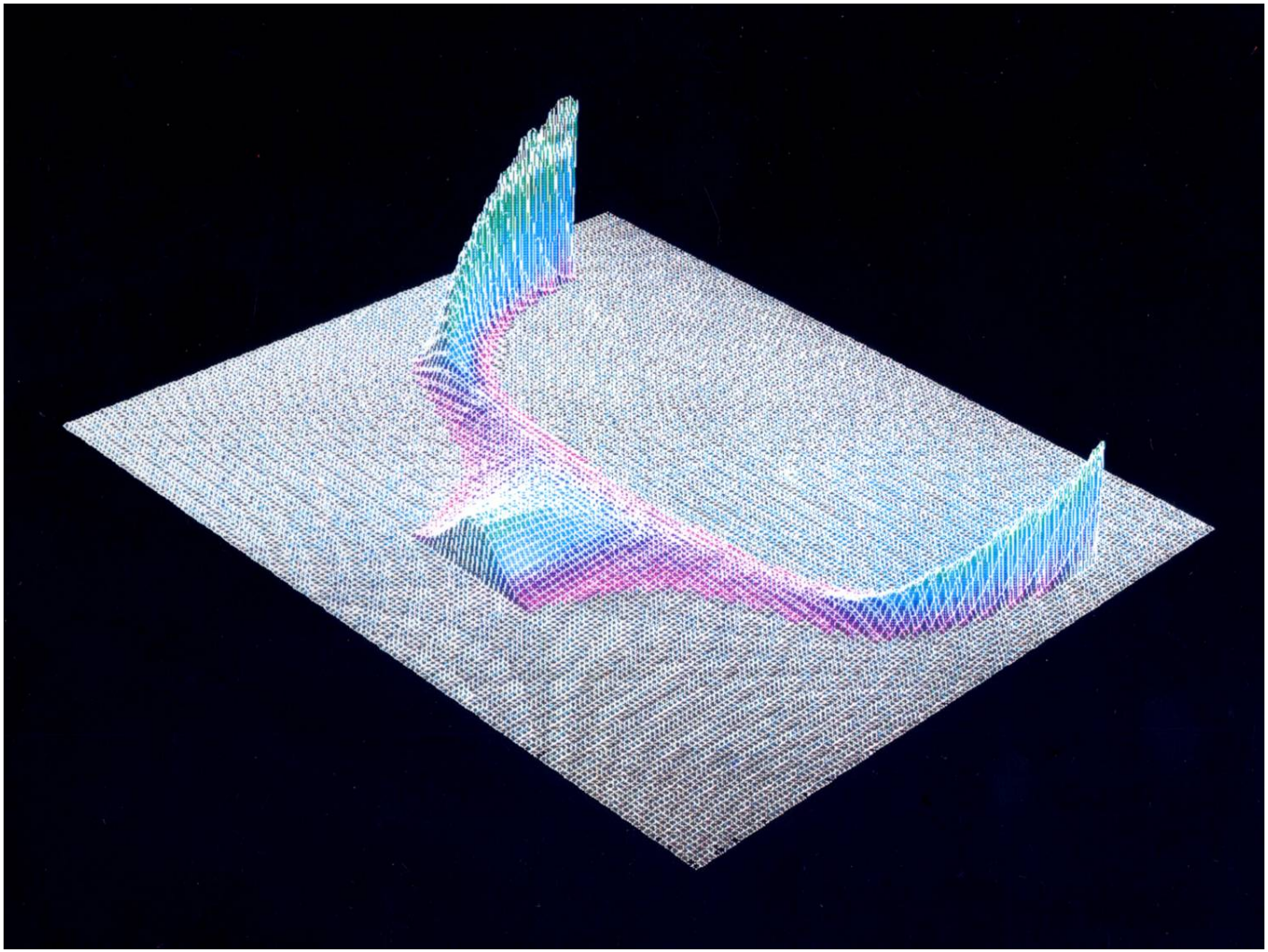
10^{60} improvement in statistics!

$$F^s_\infty = 0.01293\ (17) \quad \text{at} \quad \beta = 0.272 \quad (150\%\ \text{up}).$$



Rugged Free Energy Landscapes and Groundstate Entropy

(Berg and Celik 1992)



Distribution of the 2D Ising Model Magnetization versus Temperature.

Edward-Anderson Ising Spin Glass:

$$E = -\sum_{\langle ij \rangle} J_{ij}\, s_i\, s_j\,, \quad s_i = \pm 1\,.$$

The exchange coupling constants J_{ij} are quenched random variables. MUCA simulations allow to equilibrate in the low temperature region. One averages over many J_{ij} realizations (each requires to simulate an Ising model with random couplings). Groundstates can be sampled on small systems (infinite volume extrapolations: Berg, Celik, and Hansmann 1994).

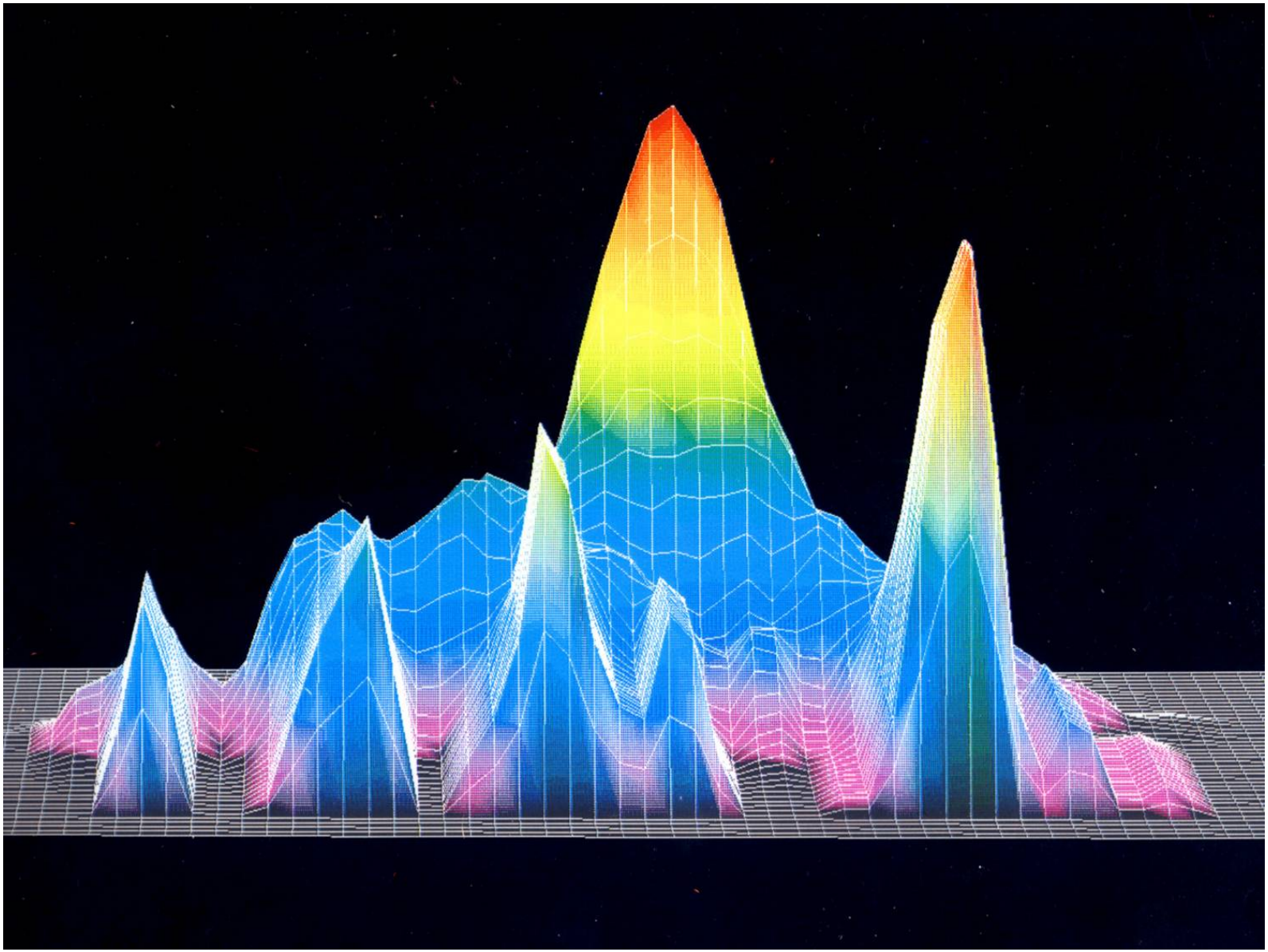
The magnetization is no longer an order parameter. The substitute is the Parisi overlap parameter:

$$q = \sum_i s_i^1\, s_i^2$$

The superscripts 1 and 2 label replica of the same realization.

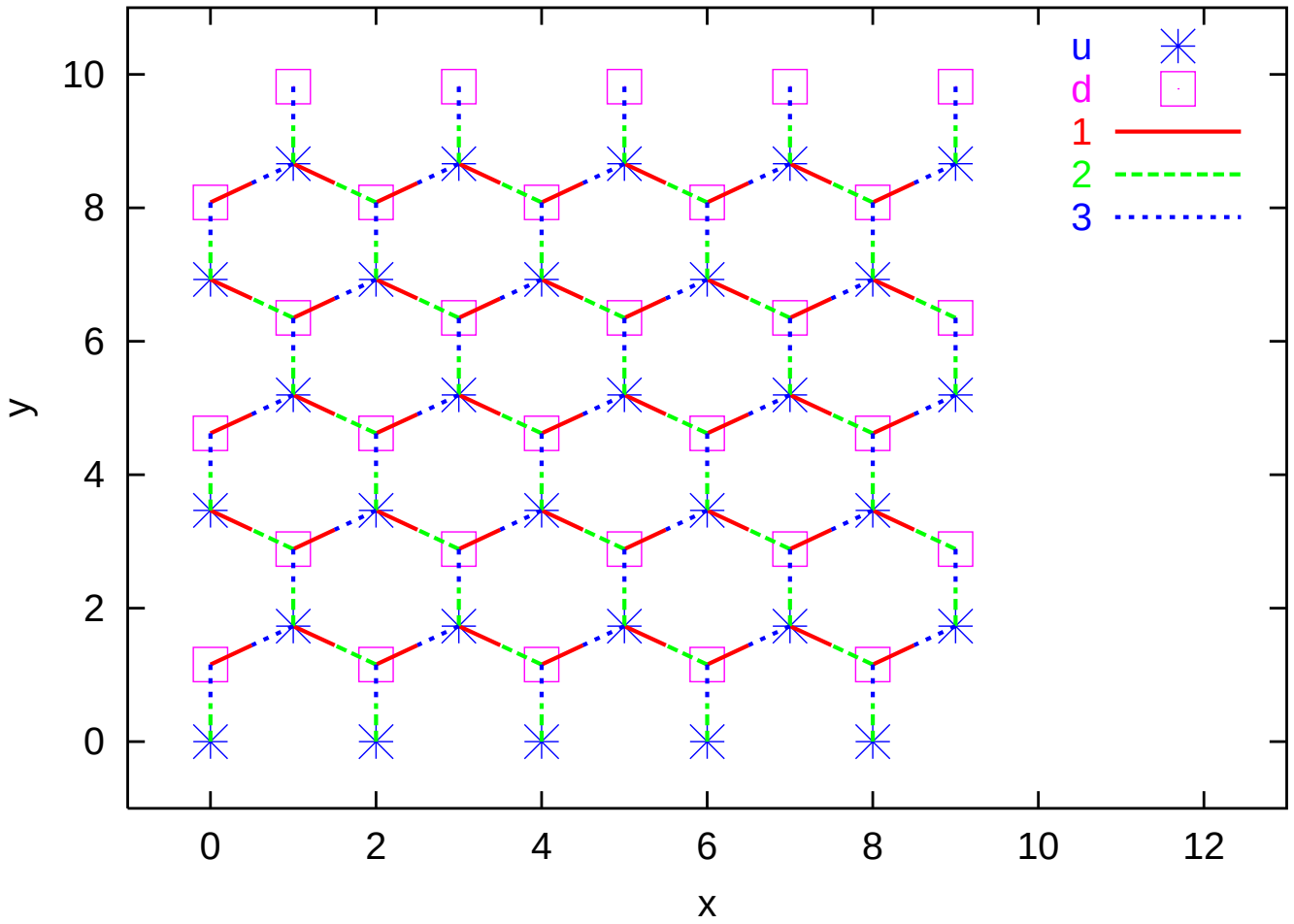
For one realization:

Parisi order parameter distribution versus temperature.



Residual Entropy of Ordinary Ice:

(Berg, Muguruma, and Okamoto 2006)



Ice rule: All links between H_2O molecules are hydrogen bonds.

Residual entropy per molecule:

$S^0 = \ln W_1$ where $W = (W_1)^N$ is the total number of configurations, which fulfill the ice rule.

Ising-like model (Linus Pauling): Hydrogen can take two positions on the bond $W_0 = 2^{2N}$. Without correlations: $6/16$ probability of correct hydrogen positions at a molecule and (1935)

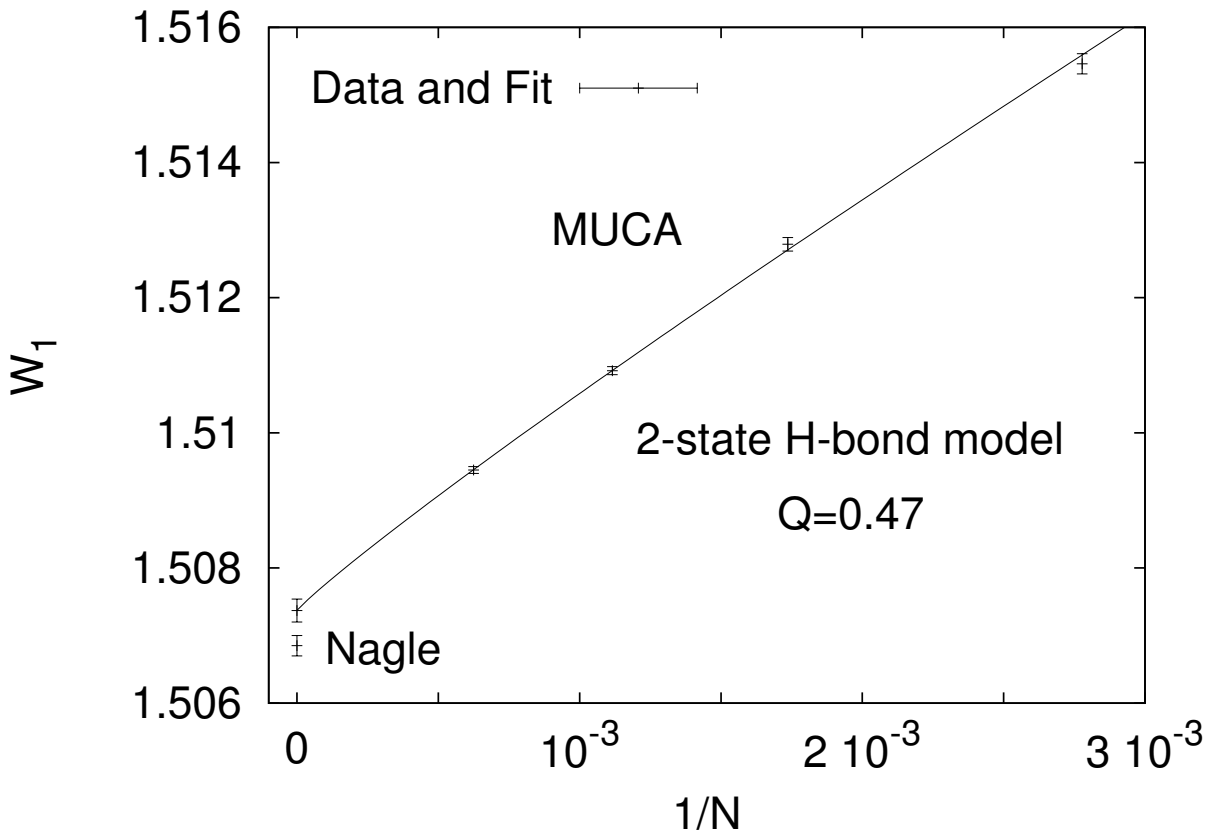
$$W^{\mathrm{Pauling}} = \left(\frac{6}{16}\right)^N 4^N \;\Rightarrow\; W_1^{\mathrm{Pauling}} = \frac{3}{2}\,.$$

Onsager and Dupuis showed that this is a lower bound. Onsager's student Nagle devoted his Ph.D. thesis to improve Pauling's estimate by a series expansion method and (1966): $W_1^{\mathrm{Nagle}} = 1.50684\,(15)$.

Experimentalists (1936): $W_1^{\mathrm{meas}} = 1.507\,(20)$ (reduced error bar).

Notes: The residual entropy of ice was experimentally discovered by Giauque and Ashley before the hydrogen bond picture was fully developed (1933). H-bond clusters are important in water at room temperature.

MUCA simulation: $W_1^{\mathrm{MUCA}} = 1.50738\,(16)$.



Theoretical predictions with an accuracy better than one per mille!

Experimental efforts to confirm the prediction!?

Multicanonical Simulations of Peptides

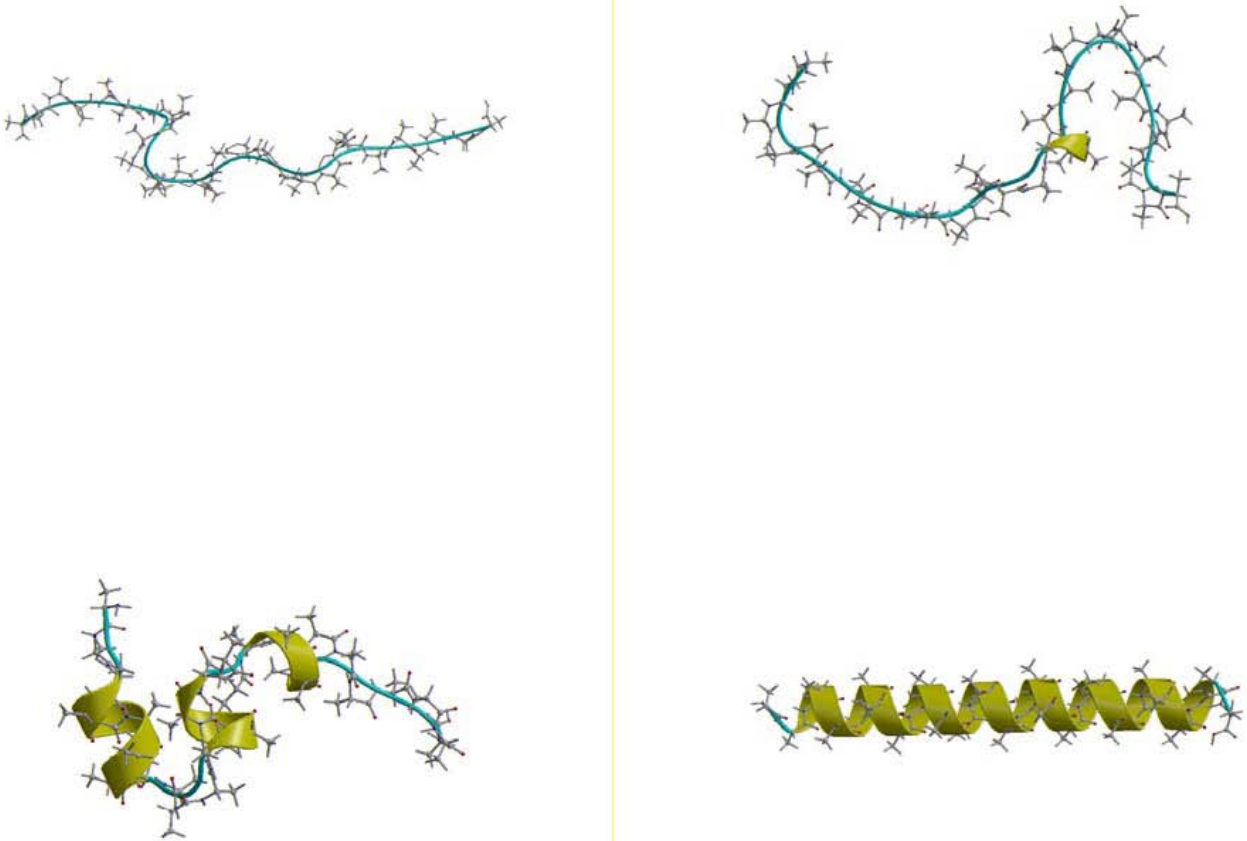
1993 Hansmann and Okamoto: Predictions of peptide conformation by multicanonical algorithm (Met-Enkephalin).

1994 Hao and Scheraga: Monte Carlo simulations of a first-order transition for protein folding using entropic sampling (lattice protein).

Entropic sampling is simply another name for the multicanonical method (Berg, Hansmann, and Okamoto 1995). Broad or flat histogram, density of states methods are other names in use. Still others names are presumably in the making ...

Most general: Transition Matrix Monte Carlo Method (Swendsen and Wang 2002).

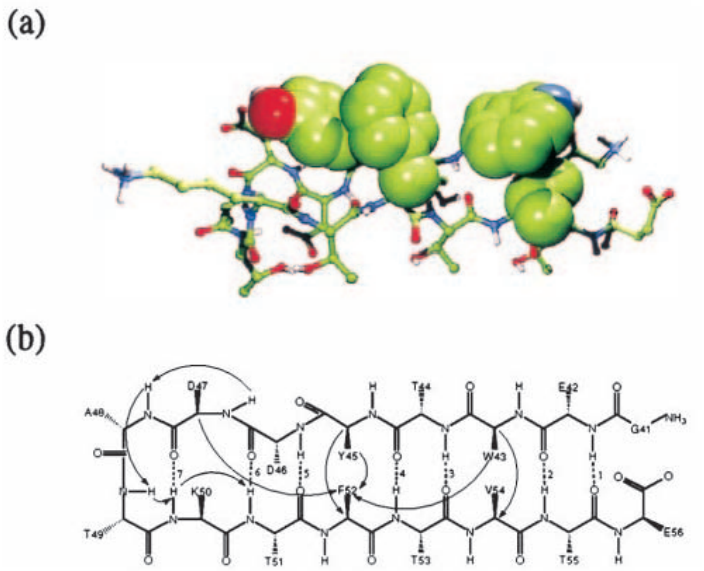
Poly-alanine: (also animation)

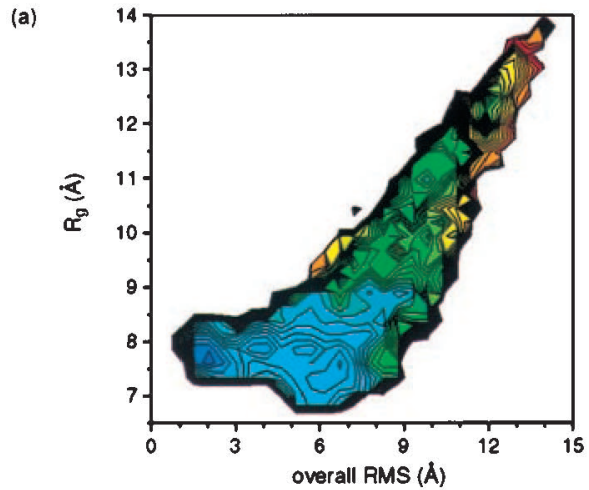


(Hansmann and Okamoto 1999)

Understanding of β-hairpin formation

GEWTYDDATKFTFTVTE 16-residue peptide corresponding to residues 41-56 of the streptococcal protein G. Native structure and free energy landscape (color coded):





(Dinner, Lazaridis and Karplus 1999)

How to get the Weights?

1. Overlapping constrained (microcanonical) MC simulations. Problems: Tedious and ergodicity of constrained simulations likely to break down in a rugged free energy landscapes.
2. Finite Size Scaling (FSS) Estimates. Best when it works! Problem: There may be no FSS theory.
3. Recursions. In general most convenient. Problem: May deteriorate quickly with increasing system size.

Example: Multicanonical Recursion (Berg 1996). Iterates microcanonical temperature $b(E)$. Animation for the 2D 10-state Potts model on an 80×80 lattice (100 iterations of 100 sweeps each).

Wang-Landau Recursion (Wang and Landau 2001)

Updates are performed with estimators $g(E)$ of the density of states:

$$p(E_1 \to E_2) = \min\left[\frac{g(E_1)}{g(E_2)}, 1\right] \ .$$

Each time an energy level is visited, the estimator is updated multiplicatively:

$$g(E) \to g(E)\, f \ .$$

Initially $g(E) = 1$ and $f = f_0 = e^1$. Once the desired energy range is adequately covered, the factor f is refined

$$f_1 = \sqrt{f}, \ f_{n+1} = \sqrt{f_{n+1}}$$

until a value sufficiently close to 1. is reached. Means: The system keeps on cycling with frozen weights.

Switch to the usual MUCA production run as soon as possible!

(To the contrary, Wang-Landau advertise to keep on iterating forever.)

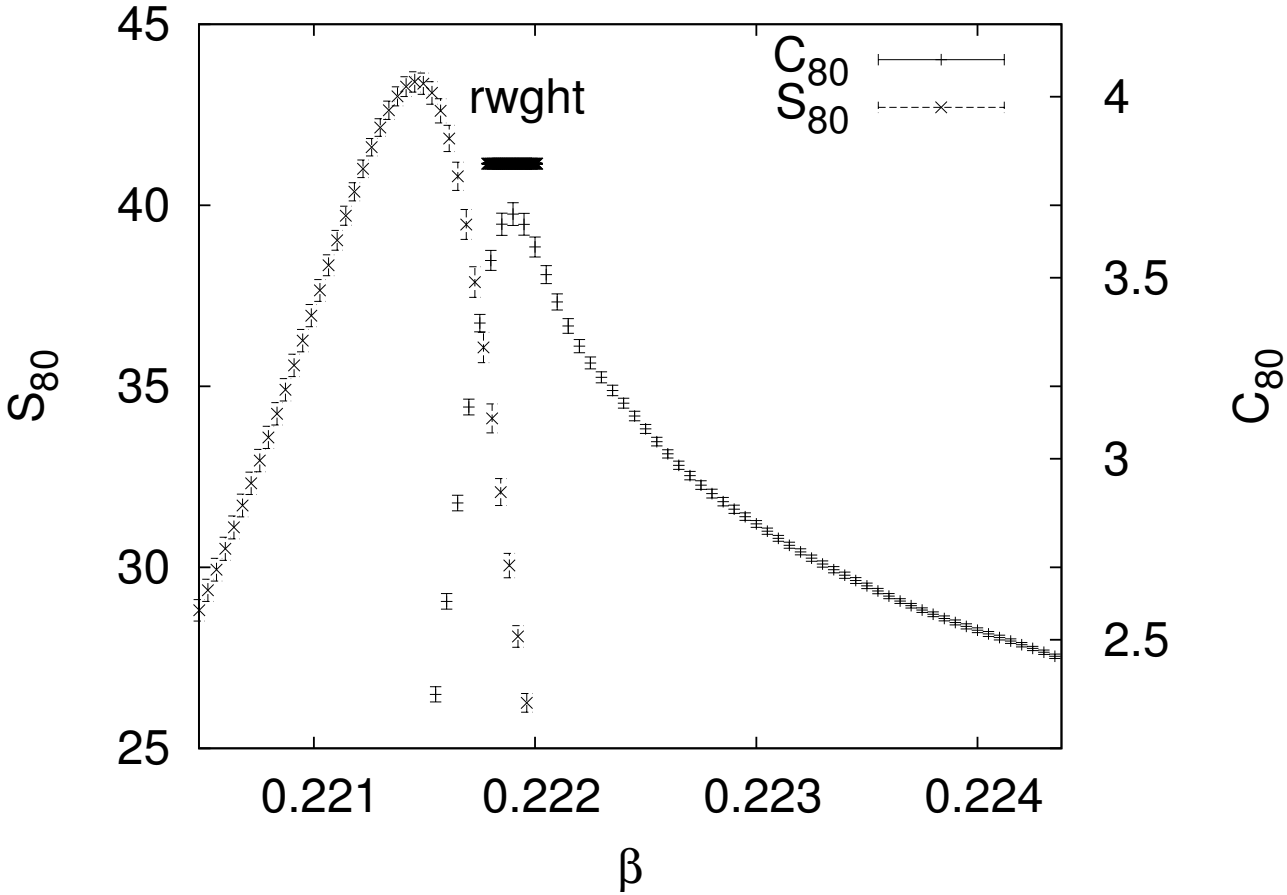
MUCA Performance

Slowing down in units of updates:

1. Optimum: $\propto N^2$ perfect random walk in the energy. Compare to $\propto \exp(+f^s\, L^{D-1}),\ N = L^D$ canonical slowing down for first order phase transitions.
2. 2D 10-state Potts model (transition range): Effective $\propto N^{2.325\,(10)}$ observed (Berg and Neuhaus 1992). Expectation for large L: Subleading exponential (Neuhaus and Hager 2003).
3. 2D EAI spin glass: $\propto N^{3.2\,(2)}$ observed (Berg and Celik 1992). Still exponential slowing down expected for large L.
4. 2D Ising model (entire energy range): $\propto N^{2.4}$ improved to $N^2\ \ln(N)$ by optimizing cycling times (Trebst, Huse, and Troyer 2004).

Second order phase transitions and cluster algorithms

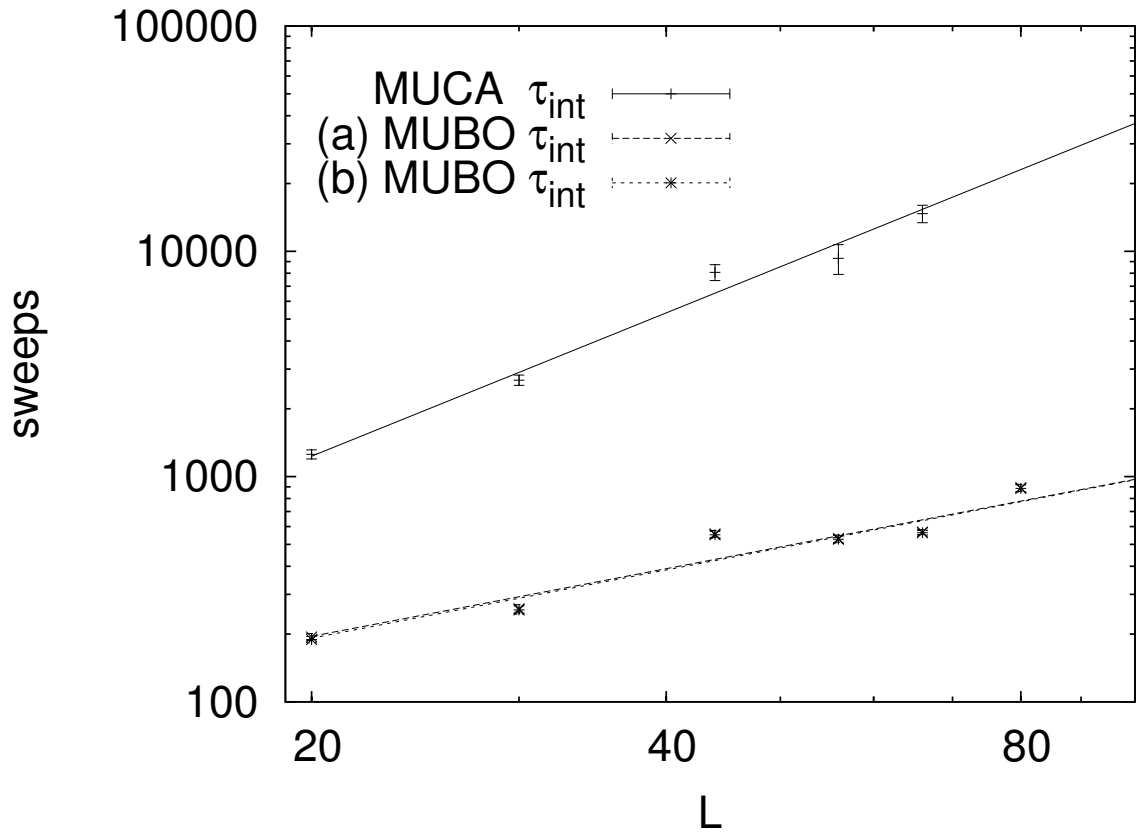
(Berg and Janke 2006)



$L^3,\ L = 80^3$ Ising model at β_c : Desired reweighting range (entire axis) versus actual canonical reweighting range (rwght).

Conventional Wang-Landau/MUCA simulations lack the efficiency of cluster algorithms (Swendsen and Wang 1987, Wolff 1989).

Combination is possible using bonds instead of energy (MUBO).



For 3D Ising: $\tau_{\rm int}^{\rm MUCA} \propto L^{2.22\,(11)}$ and $\tau_{\rm int}^{\rm MUBO} = L^{1.05\,(5)}$ sweeps.

Replica Exchange Method (Parallel Tempering)

1988 Swendsen and Wang: (Too) General replica exchange method.

1990 Frantz, Freemann, and Doll: Jump-walking feeds replica from a high into a low temperature simulation. Does not fulfill balance.

1991 Geyer: Multiple Markov chains (later called replica exchange).

1992 Lyubartsev, Martsinovski, Shevkanov, and Vorontsov-Velyaminov: Expanded ensembles (no exchange of replica).

1992 Marinari and Parisi: Simulated tempering, which is a special case of the Russian method (coined the tempering notation).

1996 Hukusima and Nemoto: Replica Exchange, often called parallel tempering (when temperatures are exchanged).

1997 Hansmann: First application to biomolecules (Met-Enkephalin).

Parallel Tempering Algorithm

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 with uniform probability and accepted with the Metropolis probability

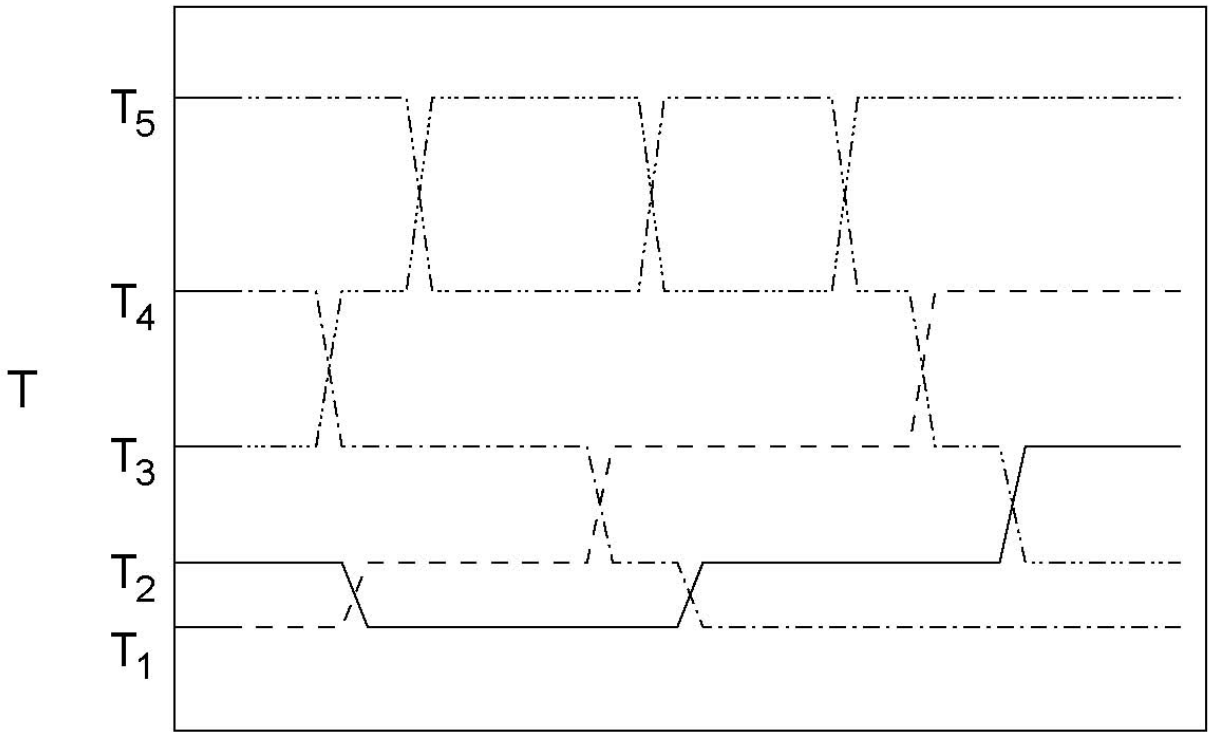
$$P_{\rm acpt} = \min\left\{ 1,\, \exp\left[+(\beta_k - \beta_l) \left(U(\vec{r}_1^{\,k}, \ldots, \vec{r}_N^{\,k}) - U(\vec{r}_1^{\,l}, \ldots, \vec{r}_N^{\,l}) \right) \right] \right\}$$

so that balance holds. Refreshes low temperature simulations through high temperature excursions. Refinements:

Other variables than temperatures can be exchanged as well.

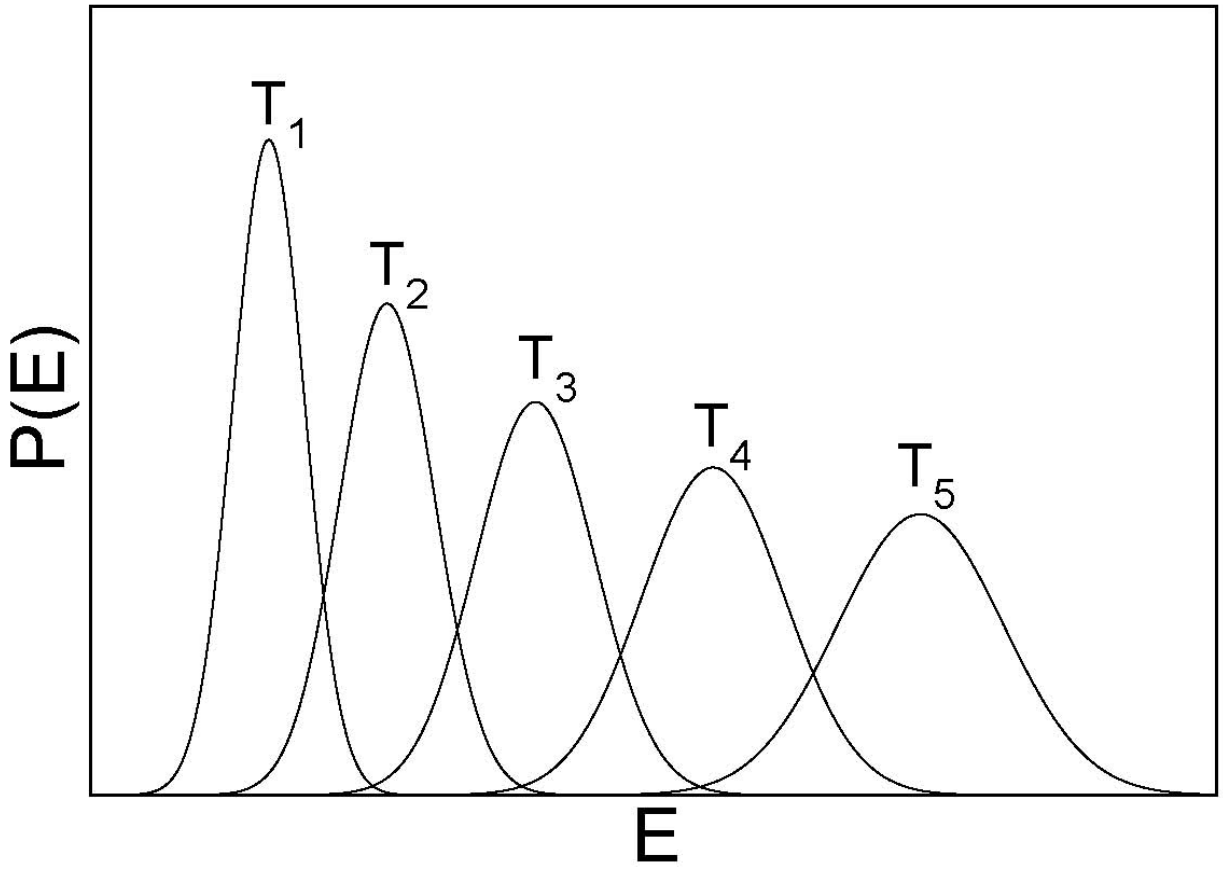
Generalized ensembles can be exchanged, ...

Temperature Exchange



Typical time series.

Energy Histograms



Overlap of energy histograms between adjacent replicas allows for acceptance of temperature swaps.

Molecular Dynamics (MD)

(Sugita and Okamoto 1999)

MD relies on Newton's equations of motion. After an exchange is accepted the new momenta are rescaled

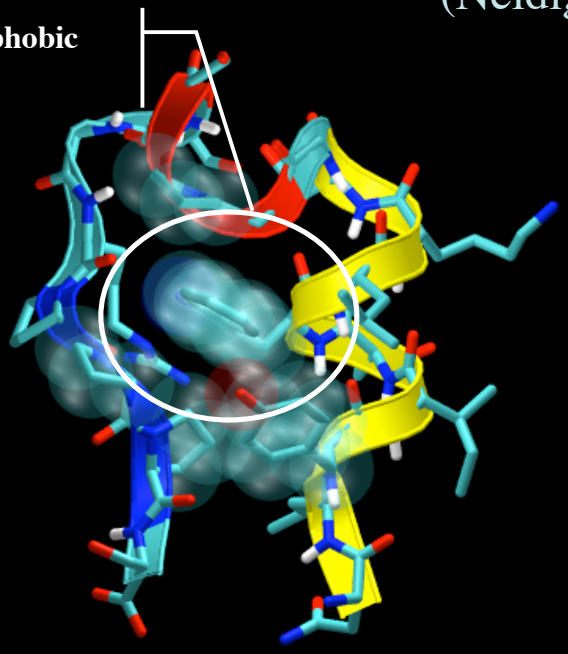
$$\vec{p}_i^{\,\mathrm{new}} = \sqrt{\frac{T^{\mathrm{new}}}{T^{\mathrm{old}}}}\,\vec{p}_i^{\,\mathrm{old}}$$

so that the average kinetic energy remains $3Nk_BT/2$.

An explosion of applications followed.

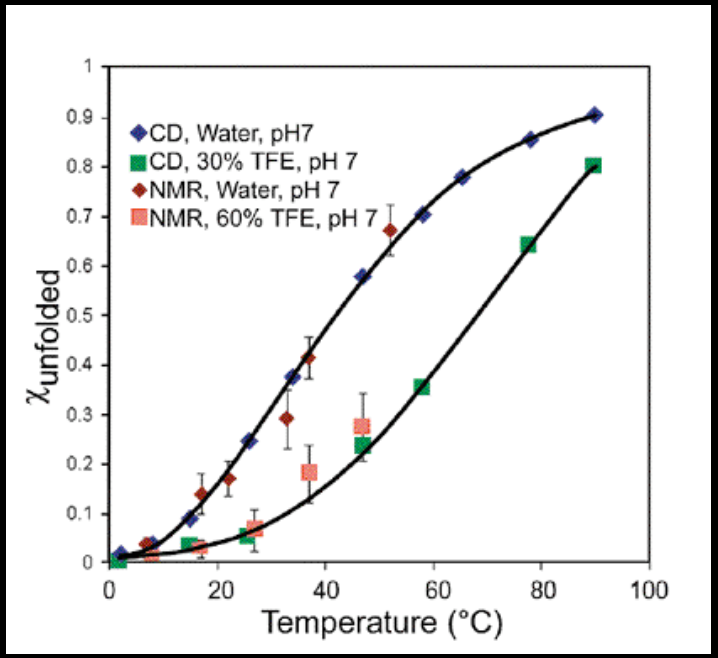
Trp-cage Protein

(Neidigh, Fesinmeyer & Andersen, 2002)



N LYIQWLKD G GPS SG RPPP S

(Qiu, et al. 2002)

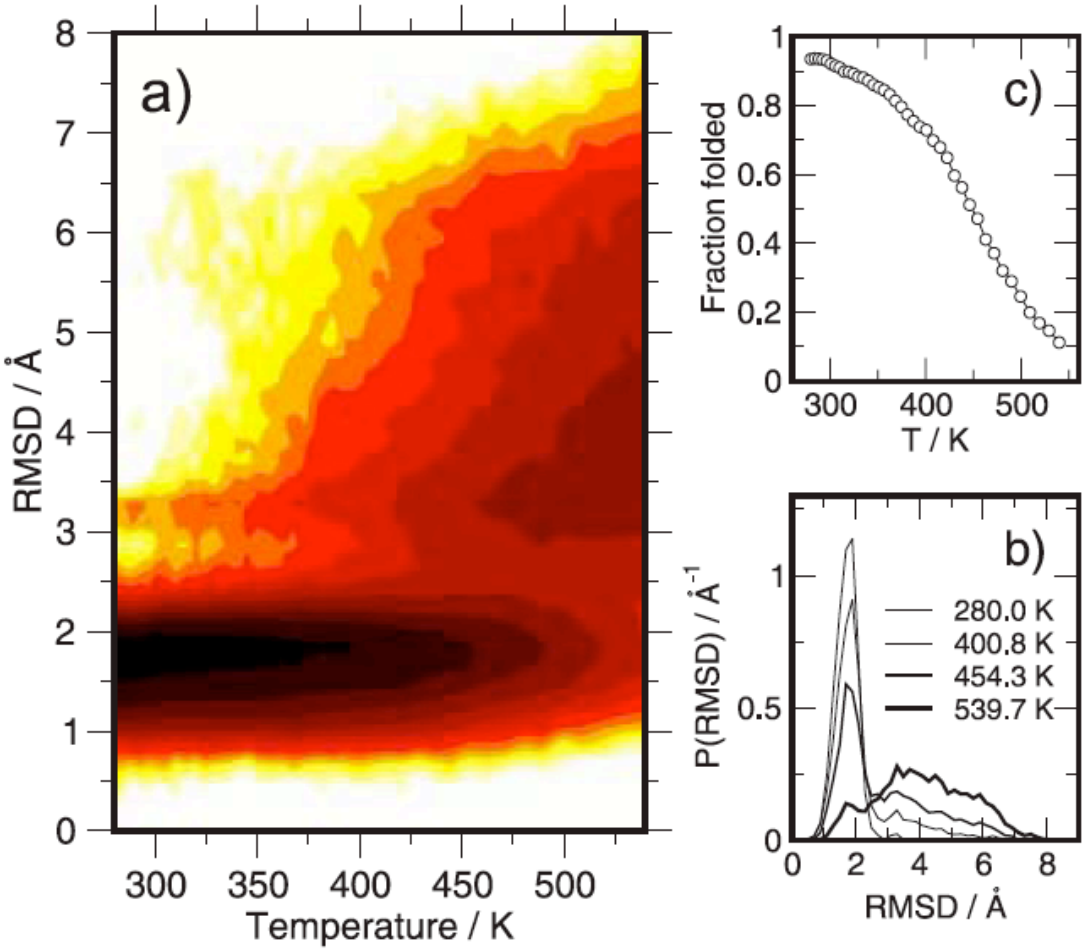


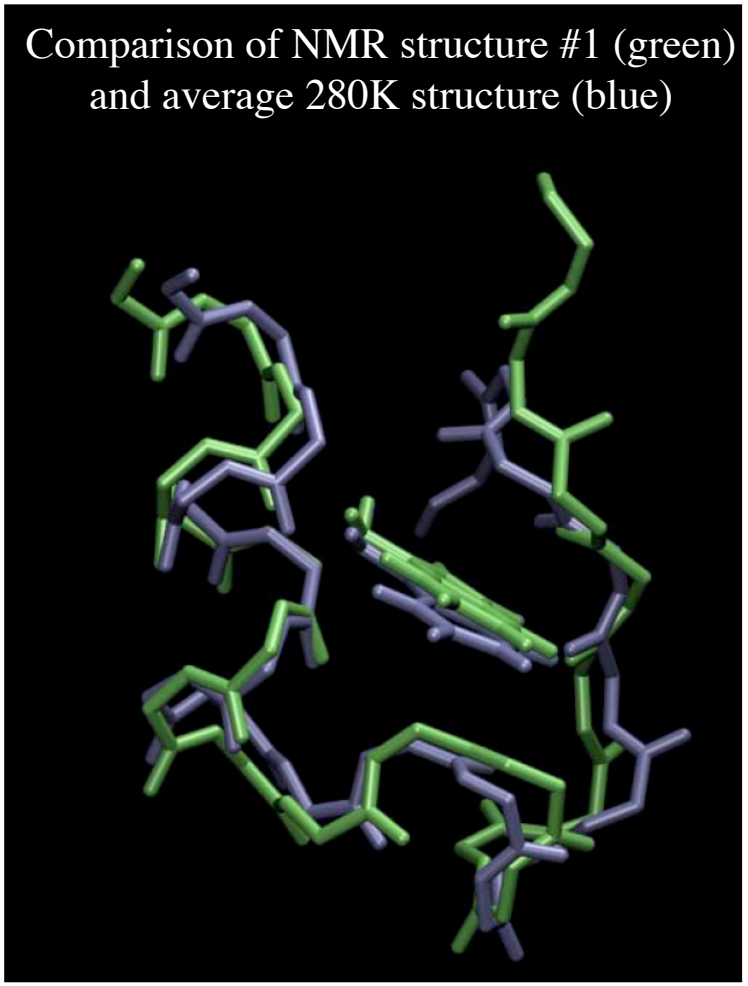
$\tau_{fold} \approx 4\ \mu s$ $\Delta H \approx -11.6$ kcal/mol

$\Delta S \approx -37.0$ cal/mol/K

(Paschek, Nymeyer & Garcia 2006)

REMD w/ AMBER94
40 replicas from 280K to 540K
100ns per replica (40ns equil + 60 prod)





Hamiltonian Replica Exchange

In essence one can exchange all kind of stuff, but the right pick is not trivial. Selected examples;

1999 Yan and de Pablo: Hyperparallel tempering Monte Carlo.

2000 Sugita et al.: Multidimensional replica-exchange.

2002 Fukunishi et al.: Hamiltonian replica exchange.

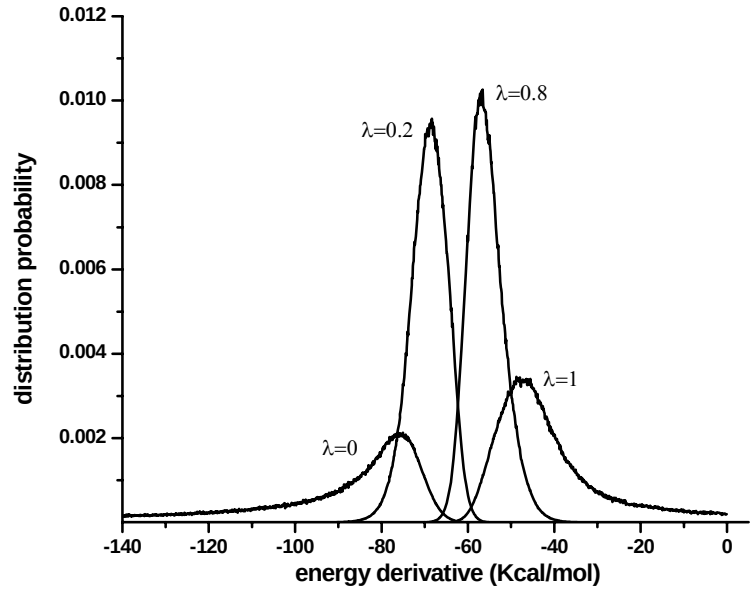
2003 Jang et al.: Exchange of generalized effective potential.

2005 Liu et al. (Bruce Berne group): (Explicit) Solute tempering.

Free Energy Differences by λ Scaling (Wei Yang Group)

$$U = (1-\lambda)U_S^A(\vec{X}) + \lambda U_S^B(\vec{X}') + U_E$$

1. HREM Variant: (Min et al. 2006)



λ_1=1 MM1

λ_2

λ_3

λ_4=0 MM0

From histogram overlap analysis

$$\Delta A_{AB} = \sum_{i=1}^{N-1} \Delta A_{i,i+1} = \sum_{i=1}^{N-1} \Delta\lambda_{i,i+1} \frac{\partial U}{\partial \lambda}_{i,i+1}$$

2. Simulated Scaling Approach (Li et al. 2006)

(Multicanonical Variant, Wang-Landau Recursion)

Flattening the λ distribution:

$$\exp\left(-\frac{\lambda_m U_s + U_e}{RT_o} + a(\lambda_m)\right)$$

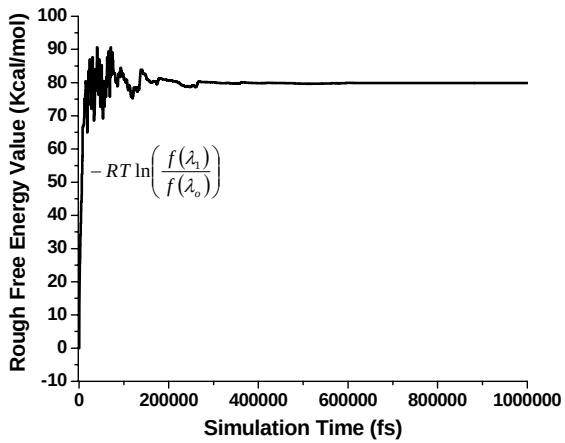
$$p_{acpt}^{biased}\left[(\lambda_o \rightarrow \lambda_1)\right] = \min\left\{1, \exp\left(-\beta\Delta\lambda\frac{\partial U}{\partial\lambda}\right)\frac{\exp\left[a(\lambda_1)\right]}{\exp\left[a(\lambda_o)\right]}\right\}$$

$$= \min\left\{1, \exp\left(-\beta\Delta\lambda\frac{\partial U}{\partial\lambda}\right)\frac{f(\lambda_1)}{f(\lambda_o)}\right\}$$

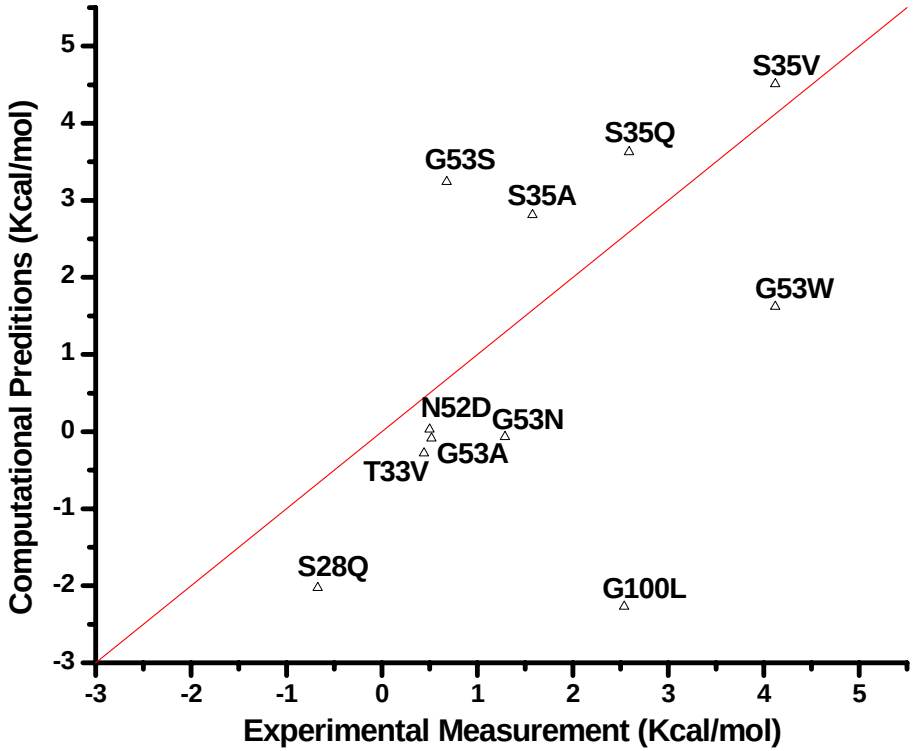
Free Energy difference is obtained via:

$$\Delta A(\lambda_o \rightarrow \lambda_1) = -RT\left[a(\lambda_1) - a(\lambda_o)\right]$$

$$= -RT\ln\left(\frac{f(\lambda_1)}{f(\lambda_o)}\right)$$



Applications in the Protein Design



AQC2 Antibody Design

Conclusions:

From a referee report on a NSF grant proposals: There are already so many numerical methods, why research on more?

There is a point, but research on computational methods will go on by the same reasons why Chemistry, Physics, or Mathematics departments do not close down!

We have seen that there can be large factors of efficiency in MCMC or MD simulations due to choosing the right methods: Larger than Avogadro's number, which determines Nature's parallel nodes in the canonical ensemble. So, try to be thoughtful when choosing your method and developing your code! No algorithm can do this part!

Many papers published smart (even more papers trivial) variations and improvements of the methods discussed here. It is difficult to consolidate them all (no clear benchmarks exist). In the course of time major breakthroughs should become apparent.