

The Method of Multicanonical Simulations  
in  
Different Models

Christoph Junghans<sup>1</sup>

Institute of Theoretical Physics  
Department of Physics and Geoscience  
University Leipzig

Supervisor:

Prof. Dr. Wolfhard Janke

Dr. Elmar Bittner  
Dr. Michael Bachmann

November 1, 2005

<sup>1</sup>E-Mail: christoph.junghans@itp.uni-leipzig.de



## **Abstract**

This theoretical report is about the method of multicanonical simulations, which became very common in the last ten years. I tested it for two different models, the Ising-Model and the AB-Model from Stilinger-Headgordon. For the Ising Model it works excellent, while for the AB-Model some difficulties, like getting trapped and problems with binning, appear which can be solved. The results look similar to previous done researches. So with a less more on input the multicanonical method provies much better results than a canonical simulation.



## Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>                    | <b>1</b>  |
| <b>2</b> | <b>Theory</b>                          | <b>1</b>  |
| 2.1      | Statistical Physics . . . . .          | 1         |
| 2.2      | Monte Carlo Simulations . . . . .      | 2         |
| 2.2.1    | Markov Process . . . . .               | 2         |
| 2.2.2    | Metropolis Update . . . . .            | 2         |
| 2.3      | Multicanonical Simulations . . . . .   | 2         |
| 2.3.1    | Determining the weights . . . . .      | 3         |
| 2.3.2    | Final Simulation . . . . .             | 5         |
| 2.3.3    | Reweighting back . . . . .             | 5         |
| 2.4      | Energy-Landscape-Paving(ELP) . . . . . | 5         |
| 2.5      | Models . . . . .                       | 6         |
| 2.5.1    | Ising-Model . . . . .                  | 6         |
| 2.5.2    | AB-Model . . . . .                     | 6         |
| <b>3</b> | <b>Results</b>                         | <b>7</b>  |
| 3.1      | Ising-Model . . . . .                  | 7         |
| 3.2      | AB-Model . . . . .                     | 8         |
| <b>4</b> | <b>Conclusions</b>                     | <b>10</b> |
|          | <b>References</b>                      | <b>10</b> |



## 1 Introduction

The physics of interesting properties are mostly described by states whose probability is low, e.g. the ground state, whose realizations are very rare, that is why the system will never pass through this state during simulation. These interesting states also represent transition states, e.g. the states with highest surface tension in the transition between droplet and strip in the Ising-Model. To overcome these barriers of improbability the sampled distribution of the simulation has to change from the Boltzmann distribution to a new one the multicanonical distribution. Another point is that a normal (canonical) simulation is performed at a fixed temperature and has a finite length, that is why the number of measured energies is also finite, so the energy distribution is only satisfactorily known in a certain interval. Below that interval the distribution is cut off because of the monotonic behavior of  $\Omega(E)$  and above this interval the factor  $e^{-\beta E}$  leads to a cut-off. To sample the whole energy space to get more information about the properties of the systems and calculate more accurate meanvalues. Maybe one can perform more canonical simulations at different temperatures or use reweighting techniques, but this is very time-consuming so I want to discuss the method of multicanonical simulations in this report.

## 2 Theory

### 2.1 Statistical Physics

It is known that the distribution of the states of a system in equilibration with a reservoir at temperature  $T$  is given by

$$p_\mu(\beta) \propto e^{-\beta E_\mu} \quad (1)$$

where  $\beta = 1/k_B T$  is called the inverse temperature. This distribution is known as the *Boltzmann Distribution*. It seems useful to rewrite it to a distribution of the energies

$$P(E, \beta) \propto \Omega(E) e^{-\beta E} \quad (2)$$

where  $\Omega(E)$  is the *density of states* which counts the number of states at a certain energy. The exact form of the density depends on the researched system and is unknown. For all normal systems one can assume that the density is continuous excepting some energy the system could not reach for structural reasons and the smallest energies, the ground states, always have much less realizations.

The exception value is the value of a quantity  $Q$  which measured in infinite long time of measurement. It is given by

$$\langle Q \rangle(\beta) = \frac{\sum_\mu Q_\mu p_\mu(\beta)}{\sum_\nu p_\nu(\beta)}, \quad (3)$$

where the sum goes over all possible states. With the help of equation 2 this can be rewritten

$$\langle Q \rangle(\beta) = \frac{\sum_E Q(E) P(E, \beta)}{\sum_E P(E, \beta)}, \quad (4)$$

but this is only possible if the quantity  $Q$  can be expressed completely in terms of the energy. The sum in equation 3 and 4 goes over the whole configuration

space which is very large and so the sum is not possible to process analytically for all systems.

## 2.2 Monte Carlo Simulations

Monte Carlo Simulations got very popular in the past fifty years. They are used to estimate statistical properties of large systems by replacing the whole configuration space by well chosen random subsets, which only contain the important states. This is called *importance sampling*.

### 2.2.1 Markov Process

A Markov process is used to produce such a random important subset of important states. It produce one configuration out of an other.

$$\mu \xrightarrow{P(\mu \rightarrow \nu)} \nu \xrightarrow{P(\nu \rightarrow \lambda)} \lambda \quad (5)$$

The  $P(\mu \rightarrow \nu)$  are called transition probabilities, which have to fulfill three conditions:

1. Normalization:  $\sum_{\nu} P(\mu \rightarrow \nu) = 1$   
This is necessary to ensure that the transition properties are normalized and that at least one transition is possible.
2. Ergodicity:  
It must be possible that every state can be reached from every state in a finite number of steps.
3. Detailed Balance:  $s_{\mu}P(\mu \rightarrow \nu) = s_{\nu}P(\nu \rightarrow \mu)$   
If this condition is satisfied, the occupation probability of a state  $\mu$  in the chain is given by  $s_{\mu}$ .

### 2.2.2 Metropolis Update

This update [1] is a special choice of the transition probabilities:

$$P(\mu \rightarrow \nu) = \min \left( 1, \frac{s_{\nu}}{s_{\mu}} \right) \quad (6)$$

## 2.3 Multicanonical Simulations

The basic idea [2] of the multicanonical simulations is to sample in an other distribution:

$$P_{\text{mu}}(E) = \Omega(E)W^{\text{mu}}(E) \approx \text{const.} \quad (7)$$

This is the multicanonical distribution which is different from the Boltzmann distribution and independent from the temperature. The new defined multicanonical weights  $W^{\text{mu}}(E)$  are unknown as the density of states is also unknown. So the simulation can be divided in:

1. Determining the weight
2. Final simulation with fixed weights and high statistics
3. Reweighting back to the canonical ensemble and calculation of the mean-values



### 2.3.1 Determining the weights

For this task I used the *Multicanonical Recursion*. One can also use the Wang-Landau algorithm or maybe a generic method. The derivations of the multicanonical recursion is adapted from the book of Bernd Berg [3]. The basic idea follows the assumption that the weights can be determined exactly from the density of states, which is of course unknown.

$$W^{\text{mu}}(E) \propto 1/\Omega(E) = e^{\ln \Omega(E)} = e^{-S(E)} \quad (8)$$

By using the *dimensionless, microcanonical free energy*

$$f(E) = \frac{F(E)}{T(E)} = \frac{U(E) - TS(E)}{T(E)} = \beta(E)E - S(E). \quad (9)$$

in equation 8 the weights can be rewritten as

$$W^{\text{mu}}(E) \propto e^{-\beta(E)E + f(E)}. \quad (10)$$

But  $f(E)$  is not independent of  $\beta(E)$ . The first fundamental law of thermodynamics

$$dU = TdS - pdV \quad (11)$$

gives the definition of the *microcanonical temperature*

$$\beta(E) = \frac{1}{T(E)} = \frac{\partial S}{\partial E} = \frac{S(E + \epsilon) - S(E)}{\epsilon}, \quad (12)$$

where  $\epsilon$  is the smallest energy difference. This is a simplification, but as I want to implement the method in a computer I need to discretize. So the connection of the temperature  $\beta(E)$  and the free energy  $f(E)$  is given by combining equations 9 and 12

$$f(E) - f(E - \epsilon) = (\beta(E) - \beta(E - \epsilon)) E. \quad (13)$$

The scale of the free energy is free to choose, that is why I choose  $f(E_{\text{max}}) = 0$ . Now I set the initial value to  $\beta_0(E) = 0$  and  $f_0(E) = 0$ , meaning that at the beginning all energies have the same weight. After the initial (0th) run the value will be replaced by  $\beta_1(E)$  and  $f_1(E)$ . So the  $n$ th run uses  $\beta_n(E)$  and  $f_n(E)$  to gain the  $n$ th histogram  $H_n(E)$ , out of which I calculate  $\beta_{n+1}(E)$  and with the help of equation 13  $f_{n+1}(E)$  follows for free.

An estimate  $\beta_{n+1}(E)$  called  $\tilde{\beta}_n(E)$  follows from the histogram  $H_n(E)$ . The idea is that an estimate of the density of states is given by the histogram so equation 8 leads to

$$H(E) \approx \Omega(E) \propto 1/W^{\text{mu}}(E). \quad (14)$$

Therefore the estimate of the associated multicanonical weight  $\tilde{W}_n^{\text{mu}}$  is simply given by

$$\tilde{W}_n^{\text{mu}} \propto e^{-\tilde{S}_n(E)} \propto \frac{W_n^{\text{mu}}}{H_n(E)}, \quad (15)$$

which under the use of equation 12 also means

$$\tilde{\beta}_n(E) = \beta_n(E) + (\ln H_n(E + \epsilon) - \ln H_n(E)) / \epsilon. \quad (16)$$

There is a problem with empty histogram entries, which is of course solved later. Now the  $\beta_{n+1}(E)$  is a error weighted mixture of the  $\tilde{\beta}_n(E)$  and  $\beta_n(E)$

$$\beta_{n+1}(E) = \kappa(E)\tilde{\beta}_n(E) + (1 - \kappa(E))\beta_n(E), \quad (17)$$

where  $\kappa$  is inversely proportional to the variance of  $\tilde{\beta}_n(E)$ . Equation 16 gives

$$\sigma^2(\tilde{\beta}_n(E)) = \sigma^2(\beta_n(E)) + \sigma^2(\ln H_n(E + \epsilon))/\epsilon + \sigma^2(\ln H_n(E))/\epsilon. \quad (18)$$

As  $\beta_n(E)$  is fixed in every single simulation its variance vanishes. For the histogram it holds that

$$\sigma^2(\ln H_n(E + \epsilon)) = (\ln(H_n + \Delta H_n) - \ln(H_n))^2, \quad (19)$$

where  $\Delta H_n$  is the fluctuation in the  $n$ th histogram which is known to grow with square root of the number of entries, what is in fact a assumption for uncorrelated histogram entries. As  $\Delta H_n \propto \sqrt{H_n}$  is smaller in comparison to the number of entries an expansion of the  $\ln(H_n + \Delta H_n)$  gives

$$\ln(H_n + \Delta H_n) - \ln(H_n) \approx \ln H_n + \frac{\Delta H_n}{H_n} - \ln H_n = \frac{\Delta H_n}{H_n} \propto \frac{1}{\sqrt{H_n}}. \quad (20)$$

Equation 18, 20 and of course uncorrelated histogram entries leads to

$$\sigma^2(\tilde{\beta}_n(E)) \propto \frac{1}{H_n(E + \epsilon)} + \frac{1}{H_n(E)}. \quad (21)$$

This means that if the statistic weight of  $\tilde{\beta}_n(E)$  grows inversely proportional to its variance it also grows with the number of entries. This is understandable as double amount of entries should have double weight.  $1/\sigma^2(\tilde{\beta}_n(E))$  can be written as

$$1/\sigma^2(\tilde{\beta}_n(E)) \propto p(E) = \frac{H_n(E + \epsilon)H_n(E)}{H_n(E + \epsilon) + H_n(E)}. \quad (22)$$

This also solves the problem with empty entries in the histogram, because  $p(E) = 0$  for  $H_n(E) = 0$  or  $H_n(E + \epsilon) = 0$ . All simulations should be taken in account in the same way. So  $\kappa(E)$  in equation 17 is just  $p(E)$  normalized on all previous simulations.

$$\kappa(E) = \frac{p(E)}{p(E) + p_n(E)}, \quad (23)$$

where  $p_n(E)$  is just sum of all  $p(E)$  from previous runs. With the help of equation 16 the recursion in equation 17 simplifies to

$$\beta_{n+1}(E) = \beta_n(E) + \kappa(E) (\ln H_n(E + \epsilon) - \ln H_n(E)) / \epsilon. \quad (24)$$

This equation can be transformed to an equation of the ratios of the weights which are defined by

$$R(E) = \frac{W^{\text{mu}}(E)}{W^{\text{mu}}(E + \epsilon)} = \frac{e^{-\beta(E)E + f(E)}}{e^{-\beta(E+\epsilon)(E+\epsilon) + f(E+\epsilon)}} = e^{\epsilon\beta(E)}, \quad (25)$$

where the last step is performed with the help of equation 13. Now the recursion in equation 24 looks much simpler:

$$R_{n+1}(E) = R_n(E) \left( \frac{H_n(E + \epsilon)}{H_n(E)} \right)^{\kappa(E)} \quad (26)$$

This recursion has no problems with empty histogram entries, because  $\kappa(E)$  vanishes for those energies.

I want to summarize this section. The determination of the weights works like this:

1.  $W_0^{\text{mu}} = 1$  and  $p_0(E) = 0$  for all  $E$ ,  $n = 0$
2. Perform a simulation with a certain number of updates using a Metropolis update with  $s_\mu = W_n^{\text{mu}}(E_\mu)$
3. Update the weights as shown in equation 24 or 26,  $p_{n+1}(E) = p_n(E) + p(E)$
4.  $n = n + 1$
5. Go to 2 or make final simulation

### 2.3.2 Final Simulation

The final simulation is longer than the simulations in the recursion in order to have more statistics which means smaller errors. It uses, like the recursion simulations, a Metropolis update with  $s_\mu = W^{\text{mu}}(E_\mu)$ . For that reason the produced Markov chain is multicanonical distributed.

### 2.3.3 Reweighting back

As I was interested in the meanvalues in the canonical ensemble I have to reweight back. A estimator for a physical quantity  $Q$  is given by

$$Q(\beta) = \frac{\sum_i Q(E_i)(W^{\text{mu}}(E_i))^{-1}e^{-\beta E_i}}{\sum_j (W^{\text{mu}}(E_j))^{-1}e^{-\beta E_j}}, \quad (27)$$

where the summation goes over all measurements of the final simulation. In the case of discretized energies, as in the Ising model, or small bin sizes, as in the AB-Model, one can also do a reweighting out of the histogram  $H(E)$

$$Q(\beta) = \frac{\sum_E Q(E)H(E)(W^{\text{mu}}(E))^{-1}e^{-\beta E}}{\sum_E H(E)(W^{\text{mu}}(E))^{-1}e^{-\beta E}}. \quad (28)$$

By using this equation one can also get a estimate for the density of states except for a scale factor  $C$ :

$$\Omega(E) = C \cdot H(E)/W^{\text{mu}}(E) \quad (29)$$

As far as I know there are no possibilities to estimate errors after calculating the density as shown above.

## 2.4 Energy-Landscape-Paving(ELP)

This relatively new method was invented by Ulrich Hansmann and Luc Wille [4]. It is designed to find global energy minima. The idea is pretty simple, one just replaces the energy  $E$  by some effective energy  $E_{\text{eff}}$

$$E \rightarrow E_{\text{eff}} = E + H(E) \quad (30)$$

where  $H(E)$  is the histogram at energy  $E$  at this step of the simulation. The rest of the simulation is just an ordinary canonical simulation at low temperature,  $T \approx 0.1$ . What happens now is that the local minima are filled up with histogram entries, that means that after the system has been in such a minima for a certain number of steps it is supposed to leave it again, because the effective energy increases with every steps. After certain, maybe long, time the system has found the global minima. Because of the time depending form of the energy all statistic properties of the simulation vanish and only a global minima can be found.

## 2.5 Models

### 2.5.1 Ising-Model

The Ising-Model is a very popular model for testing algorithms. It describes a simple ferromagnet with spins  $s_i$ , pointing up or down, sitting on every lattice site. The energy is given by

$$E = -J \sum_{\langle i,j \rangle} s_i \cdot s_j, \quad (31)$$

where  $J$  is the coupling constant and the sum goes over the nearest neighbors. I take  $J = 1$  and choose a 2D square lattice with length 16, so that there are  $16 \times 16 = 256$  spins on the lattice. The energy can take values from  $-512$  to  $512$ , the smallest energy difference is four, which comes from the number of neighbors. As simple as this model is, there are still some exceptions, the energies 508 and  $-508$  are not possible to visit due to energy change by flipping one spins in the maximum or minimum configuration.

The update is very simple, it is just a flip of a randomly chosen spin. The energy changes from  $E_\mu$  to  $E_\nu$  and a Metropolis update with  $s_\mu = W^{\text{mu}}(E_\mu)$  is done.

### 2.5.2 AB-Model

The AB-Model is a course grained polymer model. It became very popular in the past few years. The system consists of a chain of monomers (see Figure 1) with distance one, which have to have different types, A or B. A stands for hydrophobic and B for hydrophilic monomers. The energy is given by

$$E = \frac{1}{4} \sum_{k=1}^{N-2} (1 - \cos \vartheta_k) + 4 \sum_{i=1}^{N-2} \sum_{j=i+2}^N \left( \frac{1}{r_{ij}^{12}} - \frac{C(\sigma_i, \sigma_j)}{r_{ij}^6} \right) \quad (32)$$

where:

$$C(\sigma_i, \sigma_j) = \begin{cases} +1 & : \sigma_i = \sigma_j = A \\ +1/2 & : \sigma_i = \sigma_j = B \\ -1/2 & : \sigma_i \neq \sigma_j \end{cases} \quad (33)$$

The first term describes the energy stored in the bending of the chain, the second term is the Lennard-Jones term which takes in account the radii of the monomers and also the interaction of the different types. I choose the sequence  $A_3B_2AB_2ABAB_2ABA$ , which is part of the sequence 20.6 from [5].



Figure 1: Researched sequence of monomers ( $A_3B_2AB_2ABAB_2ABA$ )

The update is a little more complicated for this model. The changes happen in the angles between two monomers. I choose one monomer (not the first) randomly and allowed a change of the coordinates of this monomer with two restrictions. First the distance stays one, secondly the opening angle of the cone with all possible points is at most 5 degree. I made this limitation to avoid big configuration changes in one step. The rest chain of monomers behind the changing monomer stay unchanged. Meaning that the tail is treated as hard and moves along with the update. The coordinates on the surface of the cone are chosen to be uniformly distributed. In formulas this is done in the following way:

- Take three random number  $x$ ,  $y$  and  $z$  out of  $[0, 1)$
- Calculate the random angles  $\varphi^r = 2\pi \times x$  and  $\vartheta^r = \arccos(1 - 0.01 \times y)$ , where the 0.01 comes from the maximum opening angle of 5 degree
- Calculate the angles changes  $\Delta\varphi = \arcsin(\sin \vartheta^r \times \sin \varphi^r)$  and  $\Delta\vartheta = \arcsin(\sin \vartheta^r \times \cos \varphi^r)$ , these angles changes are relative to the spheric coordinates of the monomer
- Choose a random monomer by  $n^r = z \times (N - 1) + 1$  where  $N$  is the total number of monomers
- Calculate the distance vector  $\vec{r} = \vec{x}_{n^r} - \vec{x}_{n^r-1}$  and its spheric coordinates  $\vartheta$  and  $\varphi$
- Replace  $\vartheta$  by  $\vartheta + \Delta\vartheta$  and  $\varphi$  by  $\varphi + \Delta\varphi$  and calculate the related change vector  $\vec{r}^r$
- $\vec{r}^r$  is now the displacement of all monomers behind the chosen one including the chosen one itself.

### 3 Results

#### 3.1 Ising-Model

I simulated the Ising-Model of the size  $16 \times 16$  and made 128 recursion simulations with  $257 \cdot 10^3$  single spin updates. The multicanonical recursion makes (first picture of Figure 2) the histograms flatter and flatter and the weight changed much less in a later simulation run because  $\kappa(E)$  got very small as  $p_n(E)$  got bigger with every run. The update procedure only changes the ratios of the weights (equation 26) and after that the weight are calculated iteratively from the right side, from higher to smaller energies. That is why unvisited energies always have the same weight as their low energy neighbors, this effect can be seen in the second picture of Figure 2.

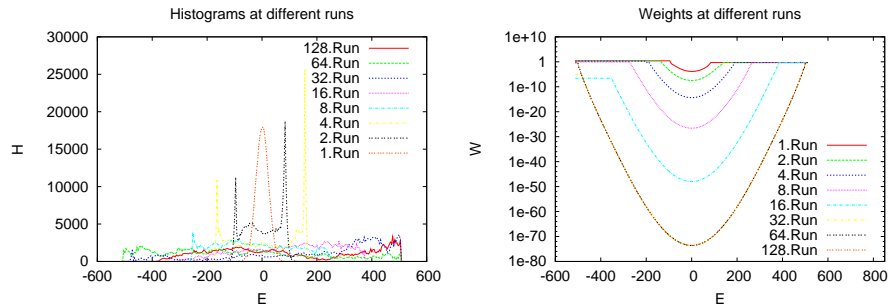


Figure 2: In recursion runs of the Ising-Modell the histograms got flatter with every run. The multicanonical weight changed very much in the beginning and less at the end, unvisited energy always stay at the same weight as their low energy neighbors

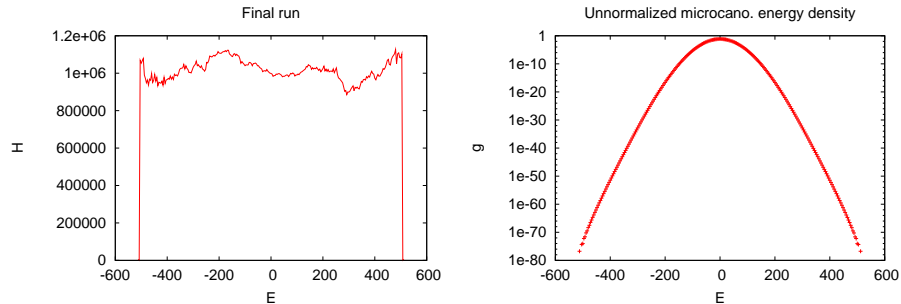


Figure 3: The histogram of the final simulation is sufficiently flat. The calculated density of states  $g$  is symmetric and goes over 80 orders of magnitude.

After the recursion I perform a final simulation with  $257 \cdot 10^6$  single spin updates. The final histogram got sufficiently flat (first picture of Figure 3), meaning that the multicanonical method works. Out of that I calculated the density of states, which may not be the common way, but makes no difference for discrete energies. The density goes over 80 orders of magnitude (see second picture of Figure 3), meaning that the states at the lowest and highest energy were exponentially suppressed and would never be visited in a canonical simulation. With the help of the density I calculate the meanvalue of the energy per spin and the specific heat per spin, the curves look very good in comparison to the exact values (see Figure 4).

### 3.2 AB-Model

I researched the sequence  $A_3B_2AB_2ABAB_2ABA$  which is a part of sequence 20.6 from [5]. I set the bin size to  $\Delta E = 0.01$  which is a very small difference and helped avoiding the problems that normally occurs with binning. The researched energy range was  $-25, \dots, 25$  which is a good choice, because the lowest energy found was around  $E_{\min} = -12.5$ . The lowest energy I found within a multicanonical simulation was  $E_{\min} = -12.58$ , as cross-check I did a

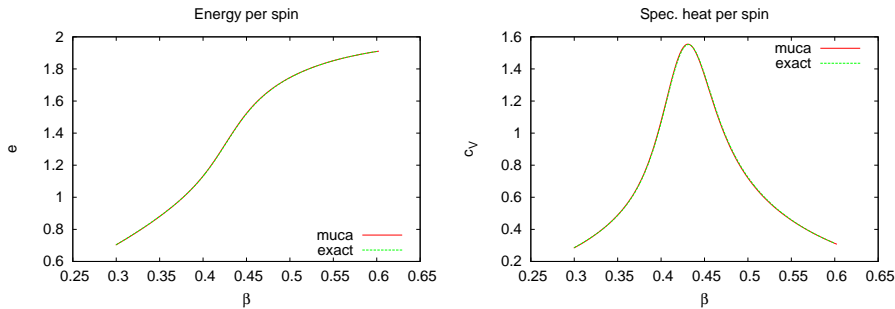


Figure 4: The meanvalue of the energy per spin and the specific heat per spin are very similar to the exact values.

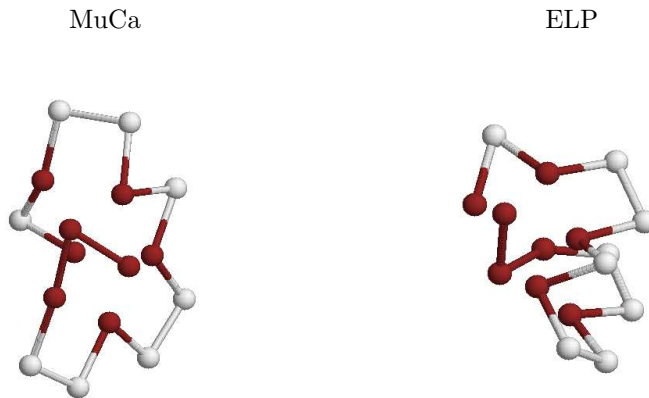


Figure 5: States with the lowest energy found in multicanonical simulations at  $E_{\min} = -12.578097$  and in ELP simulation at  $E_{\min} = -12.599549$ .

ELP simulation and found a ground state of  $E_{\min} = -12.60$  (for comparison see Figure 5). I think this energy difference comes from the too short length of the final run, but it also shows that the multicanonical simulation uses the right energy range. I did 22 recursion simulations with  $16 \cdot 10^5$  single angle updates to find the multicanonical weights, I tried more recursion runs, but the system always got stuck in some configuration around the ground state that is why the weights have not changed essentially.

The final run had a length of  $16 \cdot 10^7$  single angle updates. After that I calculated the density of states (first picture of Figure 6), which could be done without losing informations, because the bin size is so small. It also goes over 40 orders of magnitude. Out of the density the specific heat was evaluated (second picture of Figure 6), which was very similar to sequence 20.6.

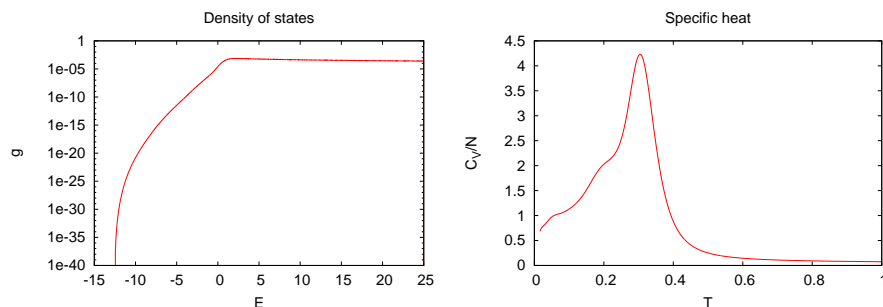


Figure 6: The density of states and the specific heat per monomer which is very similar to sequence 20.6.

## 4 Conclusions

Multicanonical simulations work for different models and help to overcome barriers caused by improbability of certain configurations. Also this method is very general and can be used for nearly every model. Small problems appear if some configurations, e.g. the ground state, have very high multicanonical weights, than the acceptance probability of the Metropolis update is near to zero and the system is getting trapped in that configuration for the rest of the simulation, so the final histogram won't be flat anymore. Longer final runs or other initial configurations can help in such occasions. An other problem is the binning, because for large energy ranges and small bin sizes the histogram has many shafts which have to be filled with entries what can take very long, but big bin sizes lead to information loses.

In summary it can be ascertained that the method of multicanonical simulation is much better than normal canonical simulation especially in low temperature regions. It is a technical method to help avoiding improbability and cut-off problems, but no *real physical* method like cluster algorithms [6] [7] and parallel tempering [8]. However it is a very good method because it uses the same input as a canonical simulation expect the estimates of the multicanonical weights. The output contains much more informations because results for all temperatures can be calculated out of one run. In comparison, a canonical simulation provides only informations about one temperature in one run. That is why the multicanonical method should be used if possible and advisable, because much better results follow nearly for free.

## References

- [1] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.
- [2] B.A. Bernd and T. Neuhaus. Multicanonical ensemble: A new approach to simulate first-order phase transistions. *Physical Review Letters*, 68(1):9–12, 1992.



- 
- [3] B.A. Berg. *Markov Chain Monte Carlo Simulations and Their Staistical Analysis*, chapter 5, pages 236–269. World Scientific, 2005.
  - [4] U.H.E. Hansmann and L.T. Wille. Global optimization by energy landscape paving. *Physical Review Letters*, 88(4), 2002.
  - [5] H. Arkun, M. Bachmann, and W. Janke. Multicanonical study of coarse-grained off-lattice models for folding heteropolymers. *Physical Review E*, 71(031906), 1998.
  - [6] R.H. Swendsen and J.-S. Wang. Nonuniversal critical dynamics in monte carlo simulations. *Physical Review Letters*, 58(2):86–88, 1987.
  - [7] U. Wolff. Collective monte carlo updating for spin systems. *Physical Review Letters*, 62(4):361–364, 1989.
  - [8] K. Hukushima and K. Nemeto. Exchange monte carlo method and application to spin glass simulations. *Journal of Physical Society of Japan*, 65(6):1604–1608, 1996.