

Introduction to Multicanonical Monte Carlo Simulations

Bernd A. Berg
Department of Physics
Florida State University
Tallahassee, Florida 32306, USA

E-mail: berg@hep.fsu.edu; Web: <http://www.hep.fsu.edu/~berg>.

Abstract. Monte Carlo simulation with *a-priori* unknown weights have attracted recent attention and progress has been made in understanding (i) the technical feasibility of such simulations and (ii) classes of systems for which such simulations lead to major improvements over conventional Monte Carlo simulations. After briefly sketching the history of multicanonical calculations and their range of application, a general introduction in the context of the statistical physics of the d -dimensional generalized Potts models is given. Multicanonical simulations yield canonical expectation values for a range of temperatures or any other parameter(s) for which appropriate weights can be constructed. We shall address in some details the question how the multicanonical weights are actually obtained. Subsequently miscellaneous topics related to the considered algorithms are reviewed. Then multicanonical studies of first order phase transitions are discussed and finally applications to complex systems such as spin glasses and proteins.

1 Introduction and Summary

One of the questions which ought to be addressed before performing a large scale computer simulation is “What are suitable weight factors for the problem at hand?” Since the 1970s it has been expert wisdom that Monte Carlo (MC) simulations with *a-priori* unknown weight factors are feasible and deserve to be considered [78]. With focus on narrow classes of applications, this idea was occasionally re-discovered, for instance [4]. With the work of ref.[7, 8] it became a more widely accepted idea, which is employed for an increasing number of applications in Physics, Chemistry, Structural Biology and other areas [21].

Lecture given at the Fields Institute *Workshop on Monte Carlo Methods*, October 1998. The author was supported in part by DOE Grants DE-FG05-87ER40319 and DE-FG05-85ER2500.

In 1988 the Italian APE¹ collaboration [3] raised the question whether the $SU(3)$ deconfining phase transition in lattice gauge theory is really of first order, as everyone believed, or possible of second order. The first order nature of the transition was ultimately confirmed and in our context it is only of interest that it was this question which triggered the theoretical high energy physics community to perform a number of numerical precision studies of first order phase transitions. In particular difficulties to calculate the interfacial tension were noted, as is reflected in the papers by Potvin and Rebbi [73], as well as Kajantie et al. [59], where new methods were introduced and tested for the² $2d$ 7-state Potts model (and later applied to $SU(3)$ lattice gauge theory). The author of this review was involved in a number of studies of the $SU(3)$ deconfining transitions, which cumulated in ref.[1] with systematic applications and developments of re-weighting techniques [37] for $SU(3)$ gauge theory. In this context the papers with Neuhaus [7, 8] originated where we introduced the *multicanonical* method and succeeded for the $2d$ 10-state Potts model to get accurate interface tension estimates out of Binder's [26] histogram technique. Looking back, it is to some extent astonishing that we did not realize immediately the connection with Torrie and Valeau's [78] *umbrella sampling* techniques. However, Binder's histogram method (known to us by personal contacts) was by then ten years old and more or less dormant due to the problem of supercritical slowing down discussed below. It was just not a natural idea to search the literature under the assumption that everyone may have overlooked an existing numerical method that gets it working. The umbrella method had stayed confined to small expert circles. The major reason for this was, most likely, that non-experts could never figure out how to get the *a-priori* unknown weight factors in the first place. With respect to this point the multicanonical papers initiated progress beyond simply making the method popular.

We are here concerned with MC simulations of the Gibbs canonical ensemble, see the next section for preliminaries. Multicanonical simulations calculate canonical expectation values in a temperature range, whereas conventional canonical simulations [66] calculate at a fixed temperature T and can, by re-weighting techniques, only be extrapolated to a vicinity of this temperature. Interest into re-weighting techniques, an idea which can be traced back to the late 1950's [75], exploded with the paper by Ferrenberg and Swendsen [37]. Apparently, the time was then right for the radical step of transgressing entirely beyond conventional, canonical MC simulations. Still, a lucky accident helped the rapid acceptance of the multicanonical method. Namely, for the $2d$ 7-state Potts model ref. [73, 59] had produced interface tension estimates which were by an entire order of magnitude larger than the multicanonical estimates [56] for the same model. Normally, such numerical discrepancies are difficult to resolve. However, for the $2d$ q -state Potts model a miracle happened: Building on results of [62, 63, 32] the *exact* $2d$ Potts model interface tensions were derived [29] shortly after the simulations were completed. The exact results and the multicanonical estimates were found to be in excellent agreement and the previous controversy converted to a significant boost for the new method.

In the application to calculations of interface tensions the emphasis is on enhancing rare configurations. Canonical MC simulations [66] sample configurations

¹Array processor with emulator, the collaboration named itself after their first dedicated computer.

²Here and in the following d denotes the dimension of the considered system.

k with the Boltzmann³ weights

$$\widehat{w}_B(k) = w_B(E^{(k)}) = e^{-\beta E^{(k)}} \quad (1.1)$$

where $E^{(k)}$ is the energy of configuration k , $\beta = 1/T$ and units are chosen such that $k_B = 1$ holds for the Boltzmann constant. The resulting probability distribution for the energy is

$$P(E) = c_\beta w_B(E) = c_\beta n(E) e^{-\beta E} \quad (1.2)$$

where $n(E)$ is the spectral density, *i.e.* the number of configurations of energy E . The normalization constant c_β is needed to ensure $\sum_E P(E) = 1$. Let L characterize the lattice size (for instance $N = L^d$ spins). For first order phase transitions pseudocritical points $\beta^c(L)$ exist such that the energy distributions $P(E) = P(E; L)$ become double peaked and $\beta^c(L)$ can be chosen such that the maxima at $E_{\max}^1 < E_{\max}^2$ are of equal height $P_{\max} = P(E_{\max}^1) = P(E_{\max}^2)$. In-between these values a minimum is located at some energy E_{\min} . Configurations at E_{\min} are exponentially suppressed like

$$P_{\min} = P(E_{\min}) = c_f L^p \exp(-f^s A) \quad (1.3)$$

where f^s is the interface tension and A is the minimal area between the phases, $A = 2L^{d-1}$ for an L^d lattice, c_f and p are constants, $p = d - 1$ in the capillary-wave approximation [31, 39, 70]. To determine the interface tension one has to calculate the quantities

$$f^s(L) = -\frac{1}{A(L)} \ln R(L) \quad \text{with} \quad R(L) = \frac{P_{\min}(L)}{P_{\max}(L)} \quad (1.4)$$

and make a finite size scaling (FSS) extrapolation of $f^s(L)$ for $L \rightarrow \infty$. However, for large systems a canonical MC simulation (1.2) will practically never visit configurations at energy $E = E_{\min}$ and estimates of the ratio $R(L)$ will be very inaccurate. The terminology *supercritical slowing down* was coined to characterize such an exponential deterioration of simulation results with lattice size. Ref. [8] overcame this problem by sampling, in an appropriate energy range, with an *approximation*

$$\widehat{w}_{mu}(k) = w_{mu}(E^{(k)}) = e^{-b(E^{(k)}) E^{(k)} + a(E^{(k)})} \quad (1.5)$$

to the weights

$$\widehat{w}_{1/n}(k) = w_{1/n}(E^{(k)}) = \frac{1}{n(E^{(k)})}. \quad (1.6)$$

Here the function $b(E)$ is the *microcanonical temperature* at energy E and $a(E)$ is some kind of fugacity. Whereas approximations to the weight factors⁴ $1/n(E)$ were already used in umbrella sampling [78], the parameterization (1.5) in terms of the microcanonical temperature $b(E)$ was introduced in [7] and is typical for the multicanonical approach. The function $b(E)$ has a relatively smooth dependence on its arguments which makes it a very useful quantity when dealing with the weight factors.

With an approximation $w_{mu}(E^{(k)})$ to the weights (1.6) one samples instead of the canonical energy distribution $P(E)$ a new multicanonical distribution

$$P_{mu}(E) = c_{mu} n(E) w_{mu}(E) \approx c_{mu}. \quad (1.7)$$

³Recently Tsallis [79] weights received also attention [46].

⁴The first detailed discussion of the connection of the multicanonical weights (1.5) with the inverse spectral density was given in [9].

The desired canonical distribution (1.2) is obtained by re-weighting

$$P(E) = \frac{c_\beta}{c_{mu}} \frac{P_{mu}(E)}{w_{mu}(E)} e^{-\beta E}. \quad (1.8)$$

This relation is rigorous, because the weights $w_{mu}(E)$ used in the actual simulation are exactly known. With the approximate relation (1.7) the average number of configurations sampled does not longer depend strongly on the energy and accurate estimates of the ratio (1.4) $R(L) = P_{\min}/P_{\max}$ become possible. Namely for $i = 1, 2$ the equation

$$R(L) = R_{mu}(L) \frac{w_{mu}(E_{\max}^i) \exp(-\beta^c E_{\min})}{w_{mu}(E_{\min}) \exp(-\beta^c E_{\max})} \text{ with } R_{mu}(L) = \frac{P_{mu}(E_{\min})}{P_{mu}(E_{\max}^i)} \quad (1.9)$$

holds and the statistical errors are those of the ratio R_{mu} times the exactly known factor. The errors of $R_{mu}(L)$ do not suffer from supercritical slowing down, because in the multicanonical simulation E_{\min} is about as frequently visited as E_{\max}^1 or E_{\max}^2 .

The attentive non-expert will raise the objection that the argument appears to be circular. The weights (1.6) are *a-priori* unknown and would they be known we could choose $w_{mu}(E^{(k)}) = w_{1/n}(E^{(k)})$ and have $R_{mu} = P_{mu}(E_{\min})/P_{mu}(E_{\max}) = 1$, *i.e.* there is no need anymore to perform a simulation as $R(L)$ of equation (1.9) is then exactly known. The answer to this is that the multicanonical method consists of two steps

1. Obtain a *working estimate* $\hat{w}_{mu}(k)$ of the weights $\hat{w}_{1/n}(k)$. Working estimate means that the approximation to (1.6) has to be good enough to ensure movement in the desired energy range, but deviations of $P_{mu}(E)$ from the constant behavior (1.7) by a factor of, say, ten are tolerable.
2. Perform a Markov chain MC simulation with the weights $\hat{w}_{mu}(k)$. Canonical expectation values are found by re-weighting to the Gibbs ensemble and standard jackknife methods [36, 10] allow reliable error estimates.

For the $2d$ 10-state Potts model figure 1 reproduces thus obtained multicanonical and re-weighted canonical energy histograms of ref.[8]. The interface tension estimate of ref.[8] is depicted in figure 2.

Assume a researcher, familiar with the canonical MC method of Metropolis et al. [66], likes to get started with multicanonical simulations. For canonical simulation the weights (1.1) are exactly known. The first stumbling block is now to get the weight factors $\hat{w}_{mu}(k)$. For small to medium sized discrete systems my recommendation is to implement the general purpose recursion [17] which is discussed in section 3. For first order phase transitions an alternative is to rely on the FSS scaling behavior of the energy distribution, see section 5. For large enough systems this approach has advantages, but its applicability is limited to systems for which a FSS theory of $P(E)$ exist. Some people recommend constrained microcanonical MC simulations, which are also shortly sketched at the end of section 5.

Once the weights $\hat{w}_{mu}(k)$ are fixed the simulation is done by Markov chain MC in the usual way. The second complication, which is minor compared to the first, is that canonical expectation values of physical observables are no longer simple arithmetic averages of the generated (“measured”) values, but more involved equations have to be employed. This expands the programming efforts, but causes no real difficulties. There is a major award for this additional work: Multicanonical simulations allow for immediate estimates of the *free energy* and the *entropy*,

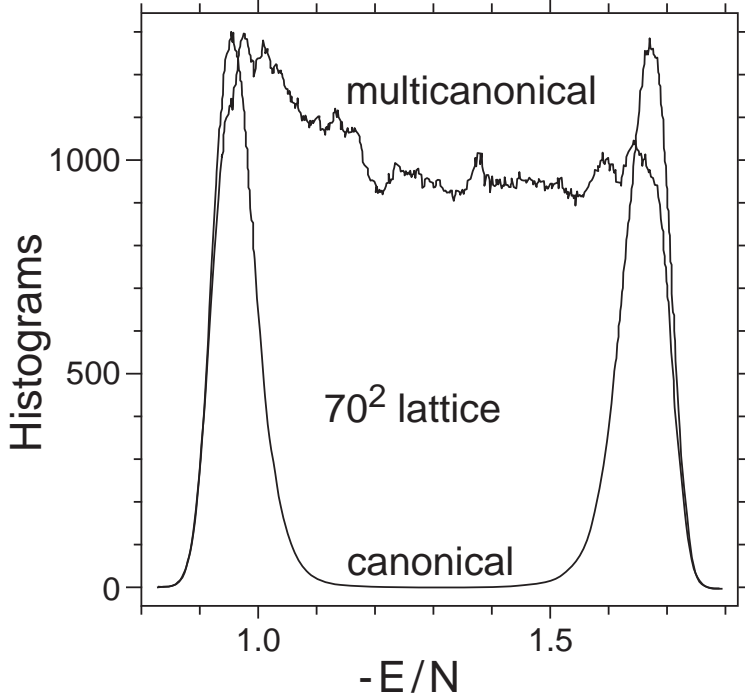


Figure 1 Multicanonical $P_{mu}(E)$ versus canonical $P(E)$ energy distribution as obtained in ref.[8] for the $2d$ 10-state Potts model on a 70×70 lattice.

whereas their reconstruction from results of canonical simulations requires tedious integration techniques. These issues are reviewed in the next section.

General properties of the method are discussed in section 4. This includes a discussion of its slowing down, *static* and *dynamic* aspects of the algorithm, and a summary of papers which have reported variants and new progress. In addition two related methods are sketched: *parallel tempering* and *random walk algorithms*. We then turn to applications in the next sections.

For first order phase transitions the multicanonical method is now well-established and a number of applications are summarized in section 5. Next, it has been realized early [11] that multicanonical simulations do not only allow to enhance the weights of relevant rare configurations, but can also be employed to improve the dynamics of the Markov process, *i.e.* its movement through configuration space in the presence of free energy barriers. Equation (1.3) is a particularly simple example of an explicitly controlled free energy barrier. Of major interest in nowadays research are complex systems, characterized by a rough free energy landscape for which an explicit parameterization is not known. The most important complex systems are presumably proteins for which multicanonical investigations have been pioneered by Hansmann and Okamoto [43]. Applications of the multicanonical method to these systems are summarized in section 6. The lack of explicit parameterizations of the

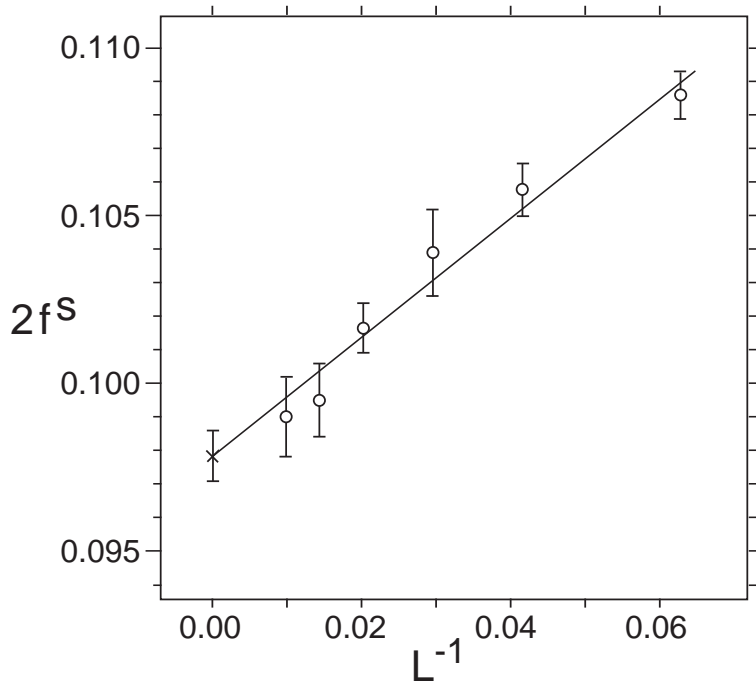


Figure 2 Interface tensions (1.4) and their $L \rightarrow \infty$ extrapolation of ref.[8].

free energy barriers makes complex systems difficult to study. For simulations one has in principle two strategies:

1. Enhance the suppressed configurations and move over the barriers.
2. Move around the barriers.

Only the first strategy allows for an explicit calculation of barrier heights, because in the second configurations with barriers are still not sampled. It is with respect to the second strategy, but not the first, that the multicanonical approach is in competition with the method of multiple Markov chains [40] (see *parallel tempering* in section 4). The canonical weights are not changed in the latter method, configurations which are rare in the Gibbs ensemble (for instance configurations with interfaces) stay rare and the strength of the two methods can be quite distinct, although a considerable overlap in the range of potential applications exists.

Finally, conclusions and outlook are given in section 7.

2 Preliminaries

MC simulations of systems described by the Gibbs canonical ensemble aim at calculating estimators of physical observables \mathcal{O} . For discrete systems (on a computer all systems are discrete) the expectations values at temperature $\beta = 1/T$

are defined by

$$\mathcal{O} = \mathcal{O}(\beta) = \langle \mathcal{O}^{(k)} \rangle = Z^{-1} \sum_{k=1}^K \mathcal{O}^{(k)} e^{-\beta E^{(k)}} \quad (2.1)$$

where

$$Z = Z(\beta) = \sum_{k=1}^K e^{-\beta E^{(k)}} \quad (2.2)$$

is the *partition function*. The sum $k = 1, \dots, K$ goes over all configurations (or microstates) of the system and $E^{(k)}$ is the energy of microstate k .

Here we focus our discussion on generalized Potts models in an external magnetic field on d -dimensional hypercubic lattices. Without being overly complicated, these models are general enough to illustrate the essential features we are interested in. In addition, various subcases of these models are by themselves of considerable physical interest. Conceptual generalizations to other models are straightforward, but technical complications arise in some cases.

For generalized Potts models the energy of the system is given by

$$E^{(k)} = E_0^{(k)} + M^{(k)} \quad (2.3)$$

where

$$E_0^{(k)} = - \sum_{\langle ij \rangle} J_{ij}(q_i^{(k)}, q_j^{(k)}) \delta(q_i^{(k)}, q_j^{(k)}) \quad \text{with} \quad \delta(q_i, q_j) = \begin{cases} 1 & \text{for } q_i = q_j \\ 0 & \text{for } q_i \neq q_j \end{cases} \quad (2.4)$$

and

$$M^{(k)} = H \sum_{i=1}^N \delta(1, q_i^{(k)}) . \quad (2.5)$$

In equation (2.4) the sum $\langle ij \rangle$ is over the nearest neighbour lattice sites and $q_i^{(k)}$ is the state of configuration k at site i . For the q -state Potts model $q_i^{(k)}$ takes on the values $1, \dots, q$. The $J_{ij}(q_i, q_j)$, ($q_i = 1, \dots, q$; $q_j = 1, \dots, q$) functions denote exchange coupling constants between the states at site i and site j . For

$$J_{ij}(q_i, q_j) \equiv J > 0 \quad (\text{conventionally } J = 1) \quad (2.6)$$

the original Potts model is recovered and $q = 2$ becomes equivalent to the Ising ferromagnet. Potts glasses [27] and Ising spin glasses [35] are obtained when the exchange constants are quenched random variables. Other choices of the J_{ij} include anti-ferromagnets and the fully frustrated Ising model [82].

The sum in equation (2.5) goes over all N sites i of the lattice. The external magnetic field is chosen to interact with the state $q_i = 1$ at each site i , but not with the other states. Each configuration (microstate of the system) k defines a particular arrangement of all states at the sites and, vice versa, each arrangement of the states at the sites determines uniquely a configuration:

$$k = \{q_1^{(k)}, \dots, q_N^{(k)}\} . \quad (2.7)$$

As there are q possible states at each site, the total number of microstates is

$$Z(0) = K = q^N , \quad (2.8)$$

where we have used the definition (2.2) of Z . Already for moderately sized systems q^N is an enormously large number. If L_n , ($n = 1, \dots, d$) are the edge lengths of the lattice, the number of sites is given by

$$N = \prod_{n=1}^d L_n, \quad (2.9)$$

i.e. $N = L^d$ for a symmetric lattice with $L_1 = \dots = L_N = L$.

Metropolis et al. [66] introduced *importance sampling* for the Gibbs canonical ensemble⁵. Their original approach imitates the thermal fluctuations of nature and weights configurations with the Boltzmann factors (1.1). The implementation relies on a Markov process and generalizes immediately to arbitrary weights $\widehat{w}(k)$. Given a configuration k , new configurations k' are proposed with probabilities of an explicitly known transition matrix

$$p^0(k', k) = p^0(k, k'), \quad \sum_{k'} p^0(k', k) = 1. \quad (2.10)$$

The symmetry condition is known as *detailed balance* and can be replaced by the somewhat weaker condition of balance, see for instance [25]. The newly proposed configuration k' is accepted with the probability

$$p^a(k', k) = \min[1, \widehat{w}(k', k)] \quad \text{with} \quad \widehat{w}(k', k) = \frac{\widehat{w}(k')}{\widehat{w}(k)} \quad (2.11)$$

and otherwise rejected. Here $\widehat{w}(k)$ can be the Boltzmann factors (1.1), the multi-canonical weights (1.5) or any other function of k . Putting equations (2.10) and (2.11) together, the transition probability for $k \rightarrow k'$ becomes

$$p(k', k) = p^0(k', k) p^a(k', k) \quad \text{for} \quad k' \neq k \quad \text{and} \quad p(k, k) = 1 - \sum_{k' \neq k} p(k', k). \quad (2.12)$$

Note that rejected proposals $k \rightarrow k'$ have to be counted as $k \rightarrow k$ transitions. It is a well-known beginner's mistake to undercount the at hand configuration k .

It was first proven in [66] that the procedure (2.12) generates configurations k with the desired weights $w(k)$ provided the property of *ergodicity* holds: Every configurations has to be reachable in a finite number of steps.

For generalized Potts models implementation of the Metropolis algorithm is straightforward. The transition matrix $p^0(k', k)$ can be defined by the following procedure

1. Pick one site i at random, *i.e.* with probability $1/N$.
2. Assign one of the states $1, \dots, q$ to q'_i , each with probability $1/q$.

These rules account for

$$p^0(k', k) = \begin{cases} 1/q & \text{for } k' = k, \\ (q-1)/(qN) & \text{for each configuration } k' \text{ with } q'_i \neq q_i \text{ at one site } i. \end{cases} \quad (2.13)$$

The new configuration k' is then accepted or rejected according to the rule (2.11). Ergodicity is fulfilled, because with (2.13) every configuration can be reached in a minimum number of N update steps. The random choice of a site insures detailed balance. A similar procedure where states are updated in the systematic order of

⁵For an introduction to the canonical ensemble and statistical physics in general see for instance [52].

a permutation π_1, \dots, π_N leads still to the desired distribution. Although detailed balance is then violated the weaker condition of balance still holds.

The Metropolis process needs to run some while to establish *equilibrium* with respect to the *ensemble* defined by the weights with which it samples, see for instance [25] for an introduction and *coupling from the past* [74] is a recent rigorous approach. In the following we assume that equilibrium has been reached. Subsequently the Metropolis process generates a number of equilibrium configurations k_i , $i = 1, \dots, n$ on which observables \mathcal{O} are *measured* similarly as in real experiments. Let us focus on the internal energy E . When Boltzmann weights $\hat{w}_B(k)$ (1.1) are used the Metropolis process generates $E^i = E^{(k_i)}$ values in the canonical ensemble and the estimator \bar{E} of $\mathcal{O} = E$ in (2.1) is simply the arithmetic average

$$\bar{E} = \frac{1}{n} \sum_{i=1}^n E^i . \quad (2.14)$$

When other weights are used this is no longer true. After equilibrium is reached, a Metropolis process with the weights $\hat{w}_{mu}(k)$ (1.5) will generate configurations in the *multicanonical ensemble*. To obtain estimators of canonical expectation values (2.1) one has to re-weight to the canonical ensemble and equation (2.14) becomes

$$\bar{E}(\beta) = \sum_{i=1}^n E^i w_{mu}^{-1}(E^i) \exp(-\beta E^i) \bigg/ \sum_{i=1}^n w_{mu}^{-1}(E^i) \exp(-\beta E^i) . \quad (2.15)$$

It is instructive to compare the histograms $H_i(E) = H_{\beta_i}(E)$, ($i = 1, 2$) of two canonical simulations ($\beta_1 < \beta_2$) with the histogram $H_{mu}(E)$ generated by a multicanonical Metropolis process. These histograms are estimators of the corresponding probability densities $P(E)$ (1.2) and $P_{mu}(E)$ (1.7). A typical situation is depicted in figure 3. Histograms from canonical simulations at sufficiently distinct temperatures will not overlap, whereas a histogram from a multicanonical simulation can bridge the entire range and allows to calculate canonical expectation values for the range $\beta_1 \leq \beta \leq \beta_2$ (and, in the case depicted, also for some $\beta > \beta_2$).

Obviously the re-weighting (2.15) can only work in a temperature range where the multicanonical histogram includes the canonical histogram from a simulation at that temperature. For insufficient statistics (or a false choice of multicanonical weights) the entries from outside the $H_{\beta}(E)$ range may lead to spurious estimates of $\bar{E}(\beta)$ (2.15). This is an example of a *bias* problem. For dealing with such problems the jackknife method, see for instance [36, 10], is recommended. Already for the statistical analysis of results from canonical simulations this should really be the method of first choice and as such be taught in introductory books on MC simulations. For multicanonical simulations bias problems tend to become more subtle and the use of jackknife estimators becomes often a must.

It is well-known [25] that a canonical MC simulation at temperature $\beta = 1/T$ provides estimates for many but not all thermodynamical quantities of interest. Obtained are estimators for the internal energy E , quantities related to its derivatives, like the specific heat

$$C_V = -\frac{\partial E}{\partial T} = \beta^2 \frac{\partial E}{\partial \beta} = \beta^2 (\langle E^{(k)} E^{(k)} \rangle - \langle E^{(k)} \rangle \langle E^{(k)} \rangle) , \quad (2.16)$$

correlation functions, the magnetization and many others. However, canonical simulations fails with respect to a number of important quantities, most prominently

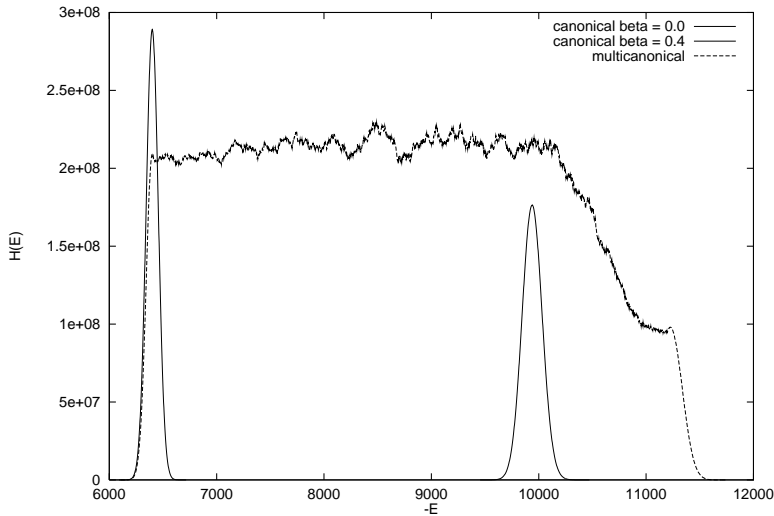


Figure 3 Simulation of a $2d$ Ising model on a 80^2 lattice: Canonical histograms $H_1(E)$ and $H_2(E)$ from simulations at $\beta_1 = 0$ and $\beta_2 = 0.4$ versus a multicanonical histogram $H_{mu}(E)$ which includes their ranges.

the partition function $Z(\beta)$ itself and, related to it, for the Helmholtz free energy

$$F(\beta) = -\beta^{-1} \ln Z(\beta) \quad (2.17)$$

and the entropy⁶

$$S = \frac{E - F}{T} . \quad (2.18)$$

The reason is that the histograms $H(E)$ of canonical sampling yield the shape of the probability density $P(E)$ but do not estimate the normalization constant c_β (1.2) which, due to $c_\beta \sum_E n(E) \exp(-\beta E) = 1$, is

$$c_\beta = 1/Z(\beta) . \quad (2.19)$$

Instead the normalization constant for the generated histogram is simply the inverse (arbitrary) number of generated configurations. In contrast to this multicanonical simulations allow to calculate $Z(\beta)$ by exploiting that $Z(0) = K = q^N$ is known (2.8). After the multicanonical simulation the estimator for the canonical probability density (1.8) is

$$P(E) = \frac{c_\beta}{c_{mu}^n} \frac{H_{mu}(E)}{w_{mu}(E)} e^{-\beta E} \quad (2.20)$$

where the superscript n on c_{mu}^n indicates that this is the constant needed when using the histogram which relies on n configurations generated in the multicanonical ensemble. Assume that the β -range for which re-weighting is valid includes $\beta = 0$. The constant c_{mu}^n follows then from the known result $c_0 = 1/Z(0) = 1/K$. Once c_{mu}^n is known c_β (and hence the partition function) follows from the normalization $\sum_E P(E) = 1$, where β has of course to stay in the admissible re-weighting range.

⁶This is the canonical entropy which has to be distinguished from the microcanonical entropy used in the next section.

3 Multicanonical Recursion

In this section a recursion [17] is presented which allows to obtain working estimates $w_{mu}(k)$ of the weight factors (1.6). Essentially, it is a simplified and improved version of a recursion which was first published in [16] after being extensively tested in spin glass simulations [15].

It is recommended to start the recursion in the disordered phase of a system, where the system moves freely under MC updates. The *zeroth* simulation is done with the weights

$$w^0(k) = 1 \text{ for all } k. \quad (3.1)$$

The most obvious recursion towards the weights (1.6) goes as follows: Simulation n , ($n = 0, 1, 2, \dots$) is carried out with the estimate $w^n(k)$ and yields the histogram $H^n(E)$. Estimate $n + 1$ for the weight factors is then given by

$$w^{n+1}(k) = \frac{w^n(k)}{H^n(E^{(k)})}. \quad (3.2)$$

This simple-minded approach fails due to a number of difficulties

1. What to do with histogram entries $H^n(E) = 0$ or small?
2. Each $H^n(E)$ calculation starts with zero statistics. Assume, someone has given us the exact weights (1.6) and we use $w^0(k) = w_{1/n}(k)$: The next estimate $w^1(k)$ will be worse. A noisy left-over of $w_{1/n}(k)$.
3. The initial weights $w^0(k) = 1$ correspond to temperature infinity and are bad in the limit $E^{(k)} \rightarrow E_g$, where E_g is the groundstate of the systems which is approached in the zero temperature limit.

The recursion derived in the following overcomes these difficulties. Let us first discuss the relationship [7, 16] of the weights (1.5) with the microcanonical temperature $b(E)$ and the fugacity $a(E)$, because it turns out to be advantageous to formulate the recursion in terms of these quantities. We have

$$w(k) = e^{-S(E^{(k)})} = e^{-b(E^{(k)}) E^{(k)} + a(E^{(k)})} \quad (3.3)$$

where $S(E)$ is the microcanonical entropy and, by definition,

$$b(E) = \frac{\partial S(E)}{\partial E}. \quad (3.4)$$

This determines the fugacity function $a(E)$ up to an (irrelevant) additive constant: We consider the case of a discrete minimal energy ϵ and choose

$$b(E) = [S(E + \epsilon) - S(E)] / \epsilon. \quad (3.5)$$

The identity $S(E) = b(E) E - a(E)$ implies

$$S(E) - S(E - \epsilon) = b(E) E - b(E - \epsilon)(E - \epsilon) - a(E) + a(E - \epsilon).$$

Inserting $\epsilon b(E - \epsilon) = S(E) - S(E - \epsilon)$ yields

$$a(E - \epsilon) = a(E) + [b(E - \epsilon) - b(E)] E \quad (3.6)$$

and $a(E)$ is fixed by defining $a(E_{\max}) = 0$. In summary, once $b(E)$ is given, $a(E)$ follows for free. The starting condition (3.1) becomes (other $b^0(E)$ choices are of course possible)

$$b^0(E) = a^0(E) = 0. \quad (3.7)$$

To avoid $H(E) = 0$ we replace for the moment

$$H(E) \rightarrow \hat{H}(E) = \max[h_0, H(E)], \quad (3.8)$$

where $0 < h_0 < 1$. Our final equations will allow for the limit $h_0 \rightarrow 0$. With this replacement we translate equation (3.2) into an equation for $b(E)$. Subscripts $_0$ are used to indicate that those quantities are not yet our final estimators from the n^{th} simulation. Let

$$w_0^{n+1}(E) = e^{-S_0^{n+1}(E)} = c \frac{w^n(E)}{\hat{H}^n(E)},$$

where the (otherwise irrelevant) constant c is introduced to ensure that $S_0^{n+1}(E)$ can be an estimator of the microcanonical entropy. It follows

$$S_0^{n+1}(E) = -\ln c + S^n(E) + \ln \hat{H}^n(E). \quad (3.9)$$

Inserting this relation into (3.5) gives

$$b_0^{n+1}(E) = b^n(E) + [\ln \hat{H}^n(E + \epsilon) - \ln \hat{H}^n(E)]/\epsilon \quad (3.10)$$

The estimator of the variance of $b_0^{n+1}(E)$ is obtained from

$$\sigma^2[b_0^{n+1}(E)] = \sigma^2[b^n(E)] + \sigma^2[\ln \hat{H}^n(E + \epsilon)]/\epsilon + \sigma^2[\ln \hat{H}^n(E)]/\epsilon.$$

Now $\sigma^2[b^n(E)] = 0$ as $b^n(E)$ is the fixed function used in the n^{th} simulation and the fluctuations are governed by the sampled histogram $H^n = H^n(E)$

$$\sigma^2[\ln \hat{H}^n] = \sigma^2[\ln(H^n)] = [\ln(H^n + \Delta H^n) - \ln(H^n)]^2$$

where ΔH^n is the fluctuation of the histogram, which is known to grow with the square root of the number of entries $\Delta H^n \sim \sqrt{H^n}$. Hence,

$$\sigma^2[b_0^{n+1}(E)] = \frac{c'}{H^n(E + \epsilon)} + \frac{c'}{H^n(E)}, \quad (3.11)$$

holds where c' is an unknown constant and this equation emphasizes that the variance is infinite when there is zero statistics, *i.e.* $H^n(E) = 0$ or $H^n(E + \epsilon) = 0$. The statistical weight for $b_0^{n+1}(E)$ is inversely proportional to its variance and the over-all constant is irrelevant. Choosing a convenient over-all constant we get

$$g_0^n(E) = \frac{H^n(E + \epsilon) H^n(E)}{H^n(E + \epsilon) + H^n(E)}. \quad (3.12)$$

Note that $g_0^n(E) = 0$ for $H^n(E + \epsilon) = 0$ or $H^n(E) = 0$. The n^{th} simulation was carried out using $b^n(E)$. It is now straightforward to combine $b_0^{n+1}(E)$ and $b^n(E)$ according to their respective statistical weights into the desired estimator:

$$b^{n+1}(E) = \hat{g}^n(E) b^n(E) + \hat{g}_0^n(E) b_0^{n+1}(E), \quad (3.13)$$

where the normalized weights

$$\hat{g}_0^n(E) = \frac{g_0^n(E)}{g^n(E) + \hat{g}_0^n(E)} \text{ and } \hat{g}^n(E) = 1 - \hat{g}_0^n(E)$$

are determined by the recursion

$$g^{n+1}(E) = g^n(E) + g_0^n(E), \quad g^0(E) = 0. \quad (3.14)$$

We can eliminate $b_0^{n+1}(E)$ from equation (3.13) by inserting its definition (3.10) and get

$$b^{n+1}(E) = b^n(E) + \hat{g}_0^n(E) \times [\ln \hat{H}^n(E + \epsilon) - \ln \hat{H}^n(E)]/\epsilon \quad (3.15)$$

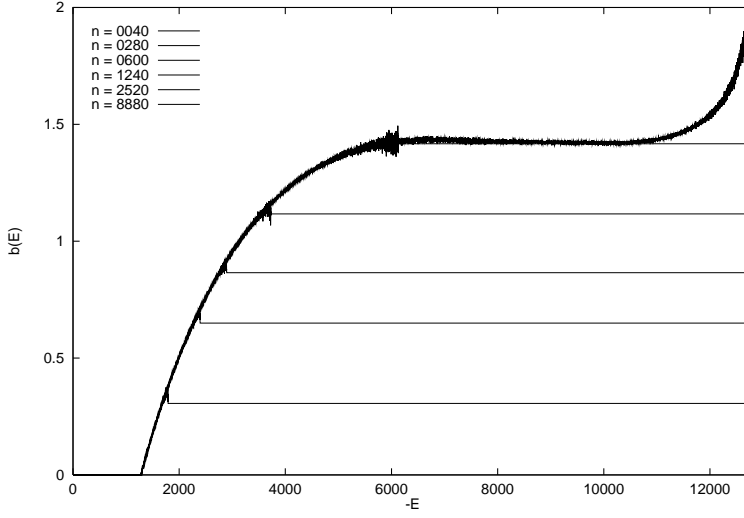


Figure 4 Recursion $b^n(E)$ for the 10-state Potts model on an 80^2 lattice.

Notice that it is now save to perform the limit

$$h_0 \rightarrow 0$$

in the definition of \hat{H} (3.8). Namely, $\hat{g}_0^n(E) = 0$ for $H^n(E) = 0$ or $H^n(E + \epsilon) = 0$ and

$$\hat{g}_0^n \times \lim_{h_0 \rightarrow 0} [\ln \hat{H}^n(E + \epsilon) - \ln \hat{H}^n(E)]$$

is well-defined for all $H^n(E)$.

In contrast to the naive recursion (3.2) the recursion (3.15) includes the entire information assembled. Frequent iterations are allowed, implying increased stability and decreased CPU time consumption. In addition it is recommended to implement a suitable $b(E)$ guess for the not yet covered $E \rightarrow E_{\min}$ energy values. Figure 4 depicts how the recursion works for the $2d$ 10-state Potts model (magnetic field $H = 0$) on an 80^2 lattice. The straight horizontal lines indicate the $b(E)$, $E \rightarrow E_g$, guess used after the corresponding recursion step and the noise at their off-spring points comes from the last recursions done before the snapshot was taken. Subsequent recursion steps were separated by one thousand MC sweeps.

Finally, equation (3.15) can be converted into a direct recursion for ratios of the weight factor neighbors. We define

$$R^n(E) = e^{b^n(E)} = \frac{w^n(E)}{w^n(E + \epsilon)} \quad (3.16)$$

and get [18]

$$R^{n+1}(E) = R^n(E) \left[\frac{\hat{H}^n(E + \epsilon)}{\hat{H}^n(E)} \right]^{\hat{g}_0^n(E)}. \quad (3.17)$$

This version is of interest when RAM limitations are an issue.

4 Properties of the Algorithm and Miscellaneous Topics

Before we come to the two major fields of applications, first order phase transitions and complex systems, it is helpful to summarize a number of algorithmic issues. A reader who is not interested in more technical points may want to skip this section.

4.1 Slowing down. Our typical situation is

$$E_{\max} - E_{\min} \sim V$$

The optimum for a flat energy distribution is given by a random walk in the energy [7]. This implies a CPU time increase

$$\sim V^2$$

to keep the number of $E_{\max} \rightarrow E_{\min} \rightarrow E_{\max}$ transitions constant. The recursion (3.15) needs an additional $\sim V^{0.5}$ (optimum) attempts to cover the entire range. It follows:

$$\text{slowing down} \sim V^{2.5} \text{ or worse.}$$

For first order phase transitions (see the next section) a *recursion alternative* is patching of overlapping constraint MC simulations:

$$\text{number of (fixed size) patches} \sim V.$$

When results can be obtained by keeping the number of updates per spin (sweeps) in each patch constant, another CPU factor $\sim V$ follows. In this case we can get:

$$\text{optimal performance} \sim V^2.$$

However, in practice this approach may face ergodicity problems.

4.2 Static and dynamic aspects of the algorithm. The actual performance of the algorithm tends to be close to the optimal performance, when one focuses on static applications. Here *static* means that some canonically rare configurations have to be enhanced. In *dynamical* applications one uses the method to by-pass rare configurations without actually enhancing them. Whether the slowing down is still close to the optimum or not depends then on whether a suitable parameterization of the problem can be found and for complex systems this is an open issue.

Static Examples:

- Magnetic field driven first-order phase transitions: Configurations with zero (or small) magnetic fields are exponentially suppressed at low temperatures and they exhibit domain walls.
- Temperature driven first-order phase transitions: Configurations with domain walls are exponentially suppressed.
- These rare configurations can be enhanced [8, 12].

Dynamic Examples:

- The following transitions may be induced by multicanonical simulations which include the high temperature region [11]:
- Low temperature transitions between magnetic states (for instance the up-down states of the Ising model below the Curie temperature, ...).
- Transition between low temperature states in systems with conflicting constraints: Spin glasses, proteins, the traveling salesman problem, ...

4.3 Variants of the multicanonical methods. *Multicanonical* refers to calculations of canonical expectation values for a temperature range and re-weighting has to be done in the internal energy. Similarly, other physical quantities can be considered, *e.g.* *multimagnetical* [12] refers to simulations which give results for a certain range of the magnetic field. A variant for cluster updates due to Janke and Kappler [58] is called *multibondic*. Recently the *multi-overlap* algorithm was introduced [19] which focuses on the Parisi order parameter for spin glasses. All these algorithms fall into the general class of *umbrella sampling* methods, which knows similar distinctions, like temperature-scaling MC [78] and the later introduced density-scaling MC [80].

4.4 Technical progress. A number of papers have reported further progress along the lines of multicanonical algorithms. For asymmetric distributions Borgs and Kappler [30] suggest to use an equal weight instead of the equal height criterium employed in equation (1.4). Hesselbo and Stinchcombe [50] made an attempt to optimize the weights and propose

$$w_{HS}(E) = 1 / \sum_{E' \leq E} n(E) \quad (4.1)$$

instead of (1.6). Their arguments cover the static situation (pertinence), but as one has no *a-priori* control over the dynamics of the systems, the question of optimal weights remains one of trial and error. Combining multicanonical with multigrid methods has been explored by Janke and Sauer [57]. For molecular dynamics, Langevin and hybrid MC variants see Hansmann et al. [44] and, with emphasize on lattice gauge theory, Arnold et al. [2]. Connections with adaption and linear response theory were explored in [71] and optimization of estimators is discussed in [38]. Occasionally attempts have been made to use bivariate multicanonical weighting [51], with most spectacular results claimed recently [48]. Concerning the latter reference, it appears to be too early for a final judgement. In general numerical methods tend to become unstable when the parameter space becomes too big.

4.5 Parallel tempering. The developments sketched above have to be distinguished from the *parallel tempering* approach which is for some applications the major competing method and should by no means be confused with multicanonical sampling.

The first paper on parallel tempering is due to Geyer [40] who introduces it as the method of multiple Markov chains. Independently it was developed in other papers. The *method of expanded ensembles* was introduced by Lyubartsev et al. [65] and proposes to enlarge the configuration space by introducing new dynamical variables. The better known *simulated tempering* [67] method of Marinari and Parisi can be considered as the special case where the temperature becomes the new dynamical variable. Building on simulated tempering the method of *parallel tempering* was introduced and tested by Hukusima and Nemoto [53]. It is identical with Geyer's [40] proposal and well suited for clusters of workstations and massively parallel computer architectures (although a moderately fast network between the nodes is sufficient).

Parallel tempering performs n canonical MC simulations at different β -values with Boltzmann weight factors

$$w_{B,i}(E^{(k)}) = e^{-\beta_i E^{(k)}}, \quad i = 1, \dots, n, \quad \beta_1 < \beta_2 < \dots < \beta_{n-1} < \beta_n$$

and allows exchange of neighbouring β -values

$$\beta_{i-1} \longleftrightarrow \beta_i \text{ for } i = 2, \dots, n. \quad (4.2)$$

These transitions lead to the energy change

$$\begin{aligned} \Delta E &= \left(-\beta_{i-1} E_i^{(k)} - \beta_i E_{i-1}^{(k')} \right) - \left(-\beta_i E_i^{(k)} - \beta_{i-1} E_{i-1}^{(k')} \right) \\ &= (\beta_i - \beta_{i-1}) \left(E_i^{(k)} - E_{i-1}^{(k')} \right) \end{aligned} \quad (4.3)$$

which is accepted or rejected according to the Metropolis algorithm. The β_i -values have to be determined such that a reasonably large acceptance rate is obtained for the β exchange (4.2) and ref.[53] employs a recursive method due to Kerler and Rehberg [61]. This is similar to the recursion needed at the beginning of a multicanonical simulation.

Remark: The method works for dynamical but not for statical supercritical slowing down. Each member of the discrete set of weight factors samples still a Boltzmann distribution (*e.g.* it is not a valid method for calculating interfacial tensions).

4.6 Random Walk Algorithms. In ref.[14] a class of algorithms was designed to perform a random walk in the energy (or any other function of the microstates). Assume the Metropolis proposal probabilities $p^0(k', k)$ (2.10) are defined which allow for $N^{(k)}$ moves out of configuration k (for the generalized Potts model probabilities (2.13) we have $N^{(k)} = (q-1)N$ independently of k). We may divide the $N^{(k)}$ moves into three classes:

$$N^{(k)} = N^+ + N^0 + N^- \quad (4.4)$$

with

1. N^+ moves with $\Delta E_i > 0$, $i = 1, \dots, N^+$.
2. N^0 moves with $\Delta E_i = 0$, $i = N^+ + 1, \dots, N^+ + N^0$.
3. N^- moves with $\Delta E_i < 0$, $i = N^+ + N^0 + 1, \dots, N^{(k)}$.

The observation of [14] is that it is easy to define transition probabilities p_i^+ , p_i^0 and p_i^- ,

$$\sum_{i=1}^{N^+} p_i^+ + \sum_{i=N^++1}^{N^++N^0} p_i^0 + \sum_{i=N^++N^0+1}^{N^{(k)}} p_i^- = 1,$$

such that

$$\sum_{i=1}^{N^+} p_i^+ \Delta E_i^+ = - \sum_{i=N^++N^0+1}^{N^{(k)}} p_i^- \Delta E_i^- \quad (4.5)$$

holds (but at the extrema). For any such choice the algorithm performs (away from the extrema) a random walk in the energy and, hence, samples a broad energy distribution.

In [14] configuration dependent transition probabilities

$$p_i^+(k), p_i^0(k) \text{ and } p_i^-(k) \quad (4.6)$$

were used, where the dependence on the configurations k exceeds a mere energy dependence. The algorithm sacrifices then an exact relationship with the canonical ensemble in favor of having a-priori well-defined defined transition probabilities and was conjectured to be of interest for optimization problems, where one focuses on

minima and not in the canonical ensemble. With this in mind the name *random cost algorithm* was chosen in [14].

Wang's [83] recent random walk algorithm is a special case of equations (4.5), but uses microcanonical averages

$$p_i^+(E), p_i^0(E) \text{ and } p_i^-(E). \quad (4.7)$$

The configuration dependence (4.6) is then reduced to energy dependence and canonical expectation values can be recovered by using the equation

$$n(E) N(E, \Delta E) = n(E + \Delta E) N(E + \Delta E, -\Delta E) \quad (4.8)$$

where $N(E, \Delta E)$ is the microcanonical average for the number of transitions from configurations with energy E to configurations with energy $E' = E + \Delta E$. For proofs of (4.8) see the appendix of [20] or ref.[72].

It seems [83] that in the random walk MC the use of estimators for $N(E, \Delta E)$ instead of their unknown exact values faces more serious problems than in multicanonical simulations. Therefore, the best application of equation (4.8) might be to employ it as input in the iteration towards the multicanonical weights (1.5). Such a use of transition probabilities was first proposed by Smith and Bruce [76, 77] and apparently deserves renewed attention.

5 First Order Phase Transitions

Multicanonical simulations are best established for investigations of first-order phase transitions. The range of applications goes from studies of mathematically ambitious topics to chemistry oriented ones. For instance an investigation [23] of the mathematically rigorous Borgs-Kotecký [28] FSS theory reveals that very strong phase transitions or very large lattices are needed to observe the asymptotic behavior predicted by their theory. To give two examples from the chemistry side, an investigation of the coexistence curve of the Lennard-Jones fluid was performed in ref.[84, 54] and liquid-vapor transitions in fluids were studied in ref.[85, 86]. The Lennard-Jones fluid has also been a major field of applications for the original umbrella sampling method and variants thereof [78, 34, 81].

A lot of work has focused on calculations of interfacial tensions and an overview is given in the forthcoming.

5.1 2d Potts Models. The pioneering study was performed for the 2d 10-state Potts model [8] and $2f^s = 0.0978(8)$ was found through FSS study of the equation (1.4). Only afterwards the exact value was discovered to be $2f^s = 0.094701\dots$ [29]. Once, the exact result was known, the remaining, small discrepancy could be eliminated by improving the finite volume estimators [24]. For these simulations the slowing down is around $\sim V^{2.3}$, *i.e.* reasonably close to the optimal performance. For a related investigation of the 7-state 2d Potts model see [56].

5.2 2d and 3d Ising Model. Many real physical systems fall into the universality class of the 3d Ising model. Despite its simplicity, it is therefore a very rewarding object to study. Although many of its universal parameters have already been determined with high precision, others are still in the making. In particular, there has been interest in the universal surface tension and the critical-isotherm amplitude ratios [87]. To obtain them, one needs accurate interfacial tension results below the Curie temperature and multimagnetical simulations [12, 13] have become the enabling technique for Binder's [26] histogram method, which was originally proposed in this context.

For the $2d$ Ising model Onsager's exact result was reproduced with good accuracy [12]. However, for the $3d$ model the temperature dependence of the interfacial tension [13] has, unfortunately, come out erratic. Therefore, the results of ref.[47] appear to be the up-to-date best estimates. Considerable technical improvements of multimagnetical calculations are nowadays feasible and it seems worthwhile to start off a new multimagnetical $3d$ Ising model simulation.

5.3 $SU(3)$ Gauge Theory. One is interested in the interfacial tension for the confinement/deconfinement phase transition. The use of multicanonical techniques has been explored by Grossmann, Laursen et al. [41, 42]. In particular, they noticed that it is suitable to use an asymmetric lattice, $V = L_z L^2 L_t$ with $L_z \geq 3L_t$. This forces the interfaces into the L^2 plane and ensures a flat region for the minimum of equation (1.3), thus greatly facilitating the extraction of finite-lattice values for the interfacial tension and, consequently, the FSS analysis.

For $SU(3)$ gauge theory the interfacial tension is usually denoted by the symbol σ and estimates are conveniently given as multiples of $T_c^3 = (T_c)^3$, where T_c is the deconfinement temperature. Using the conventions of [55], the estimates of [42] are

$$\sigma = 0.052(4) T_c^3, (L_t = 2) \text{ and } \sigma = 0.020(2) T_c^3, (L_t = 4)$$

This may be compared with the later estimate by Iwasaki et al. [55]

$$\sigma = 0.02925(22) T_c^3, (L_t = 4) \text{ and } \sigma = 0.0218(33) T_c^3, (L_t = 6)$$

The discrepancy (only $L_t = 4$ can be compared) is presumably due to too small lattice sizes in [42]. Physically, one is interested in the $L_t \rightarrow \infty$ limit. Possibly the strong L_t dependence can be eliminated by using tadpole improved actions for which Beinlich et al. [6] report

$$\sigma = 0.0155(16) T_c^3, (L_t = 3 \text{ and } L_t = 4).$$

5.4 Electroweak Phase Transition. Baryon violating processes are unsuppressed for $T > T_c$, where T_c is the electroweak critical temperature. It has been conjectured, that this may allow to explain the baryon asymmetry in nature. Models tie the nucleation rate to the interface tension of the transition. Using an effective scalar field theory [60] or the full theory [33], multicanonical and related techniques turn out to be useful for simulations at a Higgs mass

$$m_H \approx (35 - 37) GeV,$$

where one deals with a relatively strong first-order transitions, as needed to explain the baryon asymmetry. Unfortunately, it turn out that the transition weakens for higher Higgs masses, see ref.[69] for a concise review.

5.5 Weight factor estimates. For spin systems with first-order phase transitions the FSS behavior is relatively well-known. Provided the steps between system sizes are not too large, it is then possible to get working estimates of the $w_{mu}(E)$ weights by means of a FSS extrapolation from the already simulated smaller systems [8, 13]. Another method which works for these systems is patching of overlapping, constraint [22] MC simulations. This has been employed by various groups, but documentation has remained sketchy, presumably it is best in [33]. It should be emphasized that the purpose of the constrained MC simulations is here to get a working estimate of the multicanonical weights, whereas in [22] they were used for final estimates of physical observables. The advantage of the multicanonical simulation is that it has not the ergodicity problems from which microcanonical

and constraint MC simulations tend to suffer. While these approaches work for first order phase transitions, they fail as soon as ergodicity problems become severe what is normally the case for complex systems. Then it is necessary to employ a recursion like the one of section 3, which is a convenient choice for first order transitions too.

6 Complex Systems

Soon after the introduction of multicanonical methods, their potential relevance for investigations of spin glasses and other complex systems was pointed out [11]. In these systems one encounters large free energy barriers due to disorder and frustrations. Multicanonical simulations try to overcome the barriers through excursions into the disordered phase. Examples are spin glasses, proteins, hard optimization problems and others.

For an efficient application of the multicanonical idea one needs some physical insight: The configurations of interest have to be identified and a suitable parameterization has to be found which allows their actual enhancement. Whereas for first order phase transition the appropriate parameters (temperature, magnetization, etc.) are quite obvious, it is a major open problem whether suitable parameters exist for classes of complex systems.

6.1 Spin glasses. Multicanonical studies have so far been limited to the simplest, realistic prototypes, the $2d$ and $3d$ Edwards-Anderson $\pm J$ Ising⁷ spin glass [11, 15]. Significant progress has been achieved with respect to groundstate energy *and* entropy calculations. In $3d$ the groundstate energy and entropy per spin ($e_g = E_g/N$ and $s_g = S_g/N$) estimates of [15] are

$$e_g = -1.8389 \pm 0.0040 \text{ and } s_g = -1.7956 \pm 0.0042 .$$

However, on the algorithmic side the slowing down with volumes size is very bad, around V^4 or, possibly, exponential. Certain advantages of simulated tempering are claimed in [61] and [53], but these studies fail to present comparisons under identical conditions and presently there is no firm evidence that simulated tempering yields a significantly better slowing down. These studies weight spin glass configurations with the inverse spectral density (1.6). Recently a more focused application of the multicanonical method to spin glass simulations was developed [19]. The idea is to explore the free-energy structure in the Parisi order parameter such that the barriers in this variable can be mapped out. These investigations are in progress. Marinari, Parisi and collaborators perform major spin glass simulations using parallel tempering, for a review see [68].

6.2 Proteins. Proteins are linear polymers with the 20 naturally occurring amino acids as monomers. Chains smaller than a few tens of amino acids are called peptides. The problem is to predict the folded conformation of proteins and peptides solely from their amino acid sequence. For many years the emphasis of numerical investigations has been on finding the global minimum potential energy and the major difficulty encountered is the multiple minima problem. Molecular dynamics has been the numerical method of first choice, but the fraction of stochastic investigations shows an increasing trend. In particular simulated annealing has been frequently used.

⁷This is $J = 0$ or 1 in the 2-state Potts model notation.

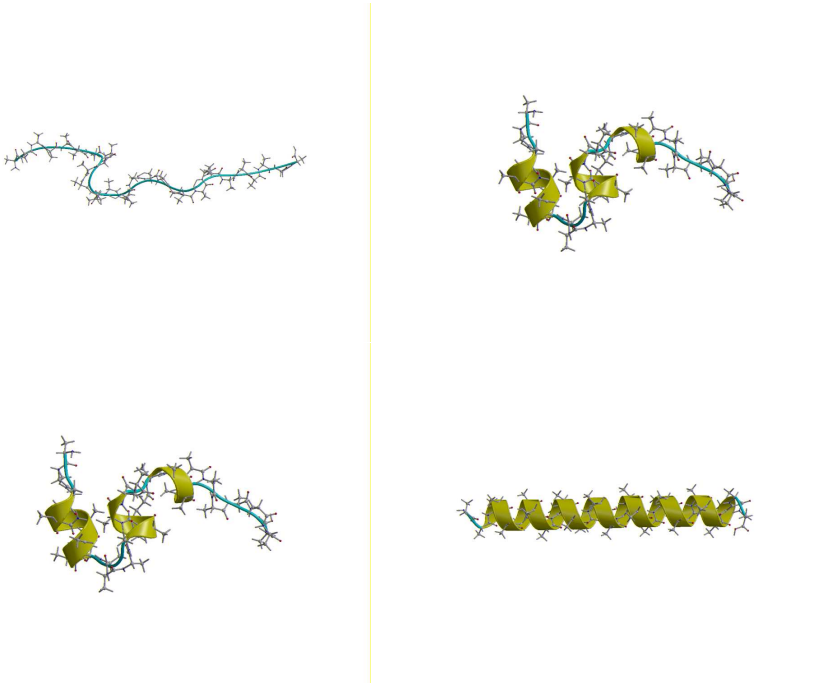


Figure 5 Example configurations from a multicanonical simulation of poly-alanine [45] (courtesy Ulrich Hansmann).

The major advantage of multicanonical and related methods in the context of proteins is that they allow for investigations of the *thermodynamics* of the entire free energy landscape of the protein. This was realized by Hansmann and Okamoto [43] when they introduced multicanonical sampling to the problem of protein folding and, slightly later, by Hao and Scheraga [49]. Since then numerous applications were performed and the simulations have been quite successful for peptides, but face tremendous problems concerning scaling to large systems. A particularly nice example is depicted in figure 5: The folding of poly-alanine into its α -helix coil [45]. No *a-priori* information about the groundstate conformation is used in these kind of simulations. By now a quite extensive literature exists which is compiled in [46].

6.3 Optimization problems. They occur in engineering, network and chip design, traffic control, and many other situations. General purpose algorithms for their solution are simulated annealing and genetic algorithms. To those, we may now like to add *multicanonical annealing* [64] and *random cost* [14].

Multicanonical annealing is a combination of multicanonical sampling with variable upper bounds and frequent adaption of parameters. A promising study [64] has been performed for the traveling salesman problem. Up to $N = 10,000$ cities, randomly distributed in the unit square, were considered and scaling of the path length as function of N was investigated. The reported algorithmic performance is so good, that an independent confirmation would be desirable.

Recently the random cost algorithm (4.5) has been applied to topology optimization in the engineering of trusses [5]. The performance was found to be competitive to that of evolutionary (genetic) algorithms.

7 Outlook and Conclusions

Sampling of broad energy distributions allows to overcome supercritical slowing down. This is well established for first-order phase transitions. Systems with conflicting constraints remain, despite some progress, notoriously difficult and for them most hope lies on achieving further algorithmic improvements. In addition, multicanonical methods may also be of interest when dealing with second order phase transitions, but so far nobody really cared to investigate this direction.

Multicanonical simulations have the potential to be replace canonical simulations as the method of first choice for exploratory MC studies. Their major advantage is that they give the thermodynamics for the entire temperature range of interest in one single MC run, whereas several to many (for each β -value one) canonical MC runs are needed. So far the major stumbling block against using multicanonical simulations in this way has been the lack of easily available implementations of recursion schemes and data analysis programs. The author thinks that these shortcomings can be overcome.

References

- [1] N. Alves, B.A. Berg and S. Sanielevici, *Spectral Density Study of the SU(3) Deconfining Phase Transition*, Nucl. Phys. **B376** (1992), 218–252.
- [2] G. Arnold, Th. Lippert and K. Schilling, *Multicanonical Hybrid Monte Carlo: Boosting Simulations of Compact QED*, Phys. Rev. **D59** (1999), 054509.
- [3] P. Bacilieri et al., *Order of the Deconfining Phase Transition in Pure-Gauge QCD*, Phys. Rev. Lett. **51** (1988), 1545–1548.
- [4] B. Baumann, *Noncanonical Path and Surface Simulation*, Nucl. Phys. **B 285** (1987), 391–409.
- [5] B. Baumann and B. Kost, *Topology Optimization of Trusses – Random Cost Method versus Evolutionary Algorithms*, Comp. Opt. and Appl., to appear.
- [6] B. Beinlich, F. Karsch and A. Peikert, *SU(3) Latent Heat and Surface Tension from Tree Level and Tadpole Improved Actions*, Phys. Lett. **B390** (1997), 268–274.
- [7] B.A. Berg and T. Neuhaus, *Multicanonical Algorithms for First Order Phase Transitions*. Phys. Lett. **B267** (1991), 249–253.
- [8] B.A. Berg and T. Neuhaus, *Multicanonical Ensemble: A New Approach to Simulate First-Order Phase Transitions*, Phys. Rev. Lett. **68** (1992), 9–12.
- [9] B.A. Berg, *The Multicanonical Ensemble: A New Approach to Computer Simulations*, Int. J. Mod. Phys. **C3** (1992), 1083–1098.
- [10] B.A. Berg, *Double Jackknife Bias Corrected Estimators*, Comp. Phys. Commun. **69** (1992), 7–15.
- [11] B.A. Berg and T. Celik, *A New Approach to Spin Glass Simulations*, Phys. Rev. Lett. **69** (1992), 2292–2295.
- [12] B.A. Berg, U.H. Hansmann and T. Neuhaus, *Simulation of an Ensemble with varying magnetic Field: A Numerical Determination of the Order-Order Interface Tension in the D=2 Ising Model*, Phys. Rev. **B47** (1993), 497–500.
- [13] B.A. Berg, U.H. Hansmann and T. Neuhaus, *Properties of Interfaces in the two and three dimensional Ising Model*, Z. Phys. **B90**, (1993), 229–239.
- [14] B.A. Berg, *Locating Global Minima in Optimization Problems by a Random-Cost Approach*, Nature **61** (1993), 708–710.
- [15] B.A. Berg, U.H. Hansmann and T. Celik, *Groundstate Properties of the 3d Ising Spin Glass*, Phys. Rev. **B50** (1994), 16444–16452.
- [16] B.A. Berg, *Multicanonical Recursions*, J. Stat. Phys. **82** (1996), 323–342.
- [17] B.A. Berg, *Algorithmic Aspects of Multicanonical Simulations*, Nucl. Phys. B (Proc. Suppl.) **63A-C** (1998), 982–984.
- [18] B.A. Berg and W. Janke, (*unpublished*).
- [19] B.A. Berg and W. Janke, *Multi-Overlap Simulations of the 3d Edwards-Anderson Ising Spin Glass*, Phys. Rev. Lett. **80** (1998), 4771–4774.

- [20] B.A. Berg and U.H. Hansmann, *Configuration Space for Random Walk Dynamics*, Eur. Phys. J. **B6** (1998), 395–398.
- [21] B.A. Berg, H. Meirovitch, M.A. Novotny and P.A. Rikvold (editors), *Monte Carlo and Structure Optimization Methods for Biology, Chemistry and Physics*, Electronic Proceedings, <http://www.scri.fsu.edu/MCatSCRI/proceedings>, Florida State University, April 1999.
- [22] G. Bhanot, S. Black, P. Carter and R. Salvador, *A New Method for the Partition Function of Discrete Systems with Application to the 3d Ising Model*, Phys. Lett. **B183** (1986), 331–336.
- [23] A. Billoire, T. Neuhaus and B. Berg, *Observation of FSS for a First Order Phase Transition*, Nucl. Phys **B396** (1993), 779–788.
- [24] A. Billoire, T. Neuhaus and B.A. Berg, *A Determination of Interface Free Energies*, Nucl. Phys. **B413** (1994), 795–812.
- [25] K. Binder, in *Phase Transitions and Critical Phenomena*, C. Domb and M.S. Green (editors), Academic Press, 1976, p.1–105.
- [26] K. Binder, *Monte Carlo Calculation of the Surface Tension for Two- and Three-Dimensional Lattice-Gas Models*, Phys. Rev. **A25** (1982), 1699–1709.
- [27] K. Binder, *Quadrupolar Glasses and Random Fields*, in *Spin Glasses and Random Fields*, edited by A.P. Young, World Scientific, Singapore, 1997.
- [28] C. Borgs, R. Kotecký and S. Miracle-Sole, *Finite Size Scaling for Potts Models*, J. Stat. Phys. **62** (1991), 529–551.
- [29] C. Borgs and W. Janke, *An Explicit Formula for the Interface Tension of the 2d Potts Model*, J. de Physique I, **2** (1992), 2011–2018.
- [30] C. Borgs and S. Kappler, *Equal Weight versus equal Height: A Numerical Study of an Asymmetric First-Order Transition*, Phys. Lett. **A171** (1992), 37–42.
- [31] E. Brezin and J. Zinn-Justin, *Finite Size Effects in Phase Transitions*, Nucl. Phys. **B257** (1985) 867–893.
- [32] E. Buffenoir and S. Wallon, *The Correlation Length of the Potts Model at the First-Order Phase Transition Point*, J. Phys. **A26** (1993), 3045–3062.
- [33] F. Csikor, Z. Fodor, J. Heim and J. Heitger, *Interface Tension of the Electroweak Phase Transition*, Phys. Lett. **B357** (1995), 156–162.
- [34] K. Ding and J.P. Valleau, *Umbrella-Sampling Realization of “Widom” Chemical Potential Estimation*, J. Chem. Phys. **98** (1993), 3306–3312.
- [35] S.F. Edwards and P.W. Anderson, *Theory of Spin Glasses*, J. Phys. **F5** (1975), 965–974.
- [36] B. Efron, *The Jackknife, the Bootstrap and other Resampling Plans*, SIAM, Philadelphia, 1982.
- [37] A.M. Ferrenberg and R.H. Swendsen, *New Monte Carlo Technique for Studying Phase Transitions*, Phys. Rev. Lett. **61** (1988), 2635–2638; **63** (1989), 1658.
- [38] M. Fitzgerald, R.R. Picard and R.N. Silver, *Canonical Transition Probabilities in Adaptive Monte Carlo*, Europhys. Lett., to appear.
- [39] M.P. Gelfand and M.E. Fisher, *Finite Size Effects in Fluid Interfaces*, Physica **A166** (1990), 1–74.
- [40] C.J. Geyer, *Markov Chain Monte Carlo Maximum Likelihood*, in *Computing Science and Statistics*, Proceedings of the 23rd Symposium on the Interface, E.M. Keramidas (editor), Interface Foundation, Fairfax, Virginia, 1991, pp.156–163.
- [41] B. Grossmann, M.L. Laursen, T. Trappenberg and U.-J. Wiese, *A Multicanonical Algorithm for SU(3) Pure Gauge Theory*, Phys. Lett. **B293** (1992), 175–180.
- [42] B. Grossmann and M.L. Laursen, *The Confined-Deconfined Interface Tension in Quenched QCD Using the Histogram Method*, Nucl. Phys. **B408** (1993), 637–656.
- [43] U.H. Hansmann and Y. Okamoto, *Prediction of Peptide Conformation by Multicanonical Algorithm: New Approach to the Multiple-Minima Problem*, J. Comp. Chem. **14** (1993), 1333–1338.
- [44] U.H. Hansmann, Y. Okamoto and E. Eisenmenger, *Molecular Dynamics, Langevin and Hybrid Monte Carlo Simulations in a Multicanonical Ensemble*, Chem. Phys. Lett. **259** (1996), 321–330.
- [45] U.H. Hansmann and Y. Okamoto, *Finite-Size Scaling of Helix-Coil Transitions in the Poly-Alanine Studied by Multicanonical Simulations*, J. Chem. Phys. **110** (1999) 1267–1276.
- [46] U.H. Hansmann and Y. Okamoto, *New Monte Carlo Algorithms for Protein Folding*, Current Opinion in Structural Biology **9** (1999) 177–183.

- [47] M. Hasenbusch and K. Pinn, *Comparison of Monte Carlo Results for the 3d Ising Interface Tension and Interface Energy with (Extrapolated) Series Expansions*, Physica **A203** (1994), 189–213.
- [48] N. Hatano, *Bivariate Multicanonical Monte Carlo of the 3D $\pm J$ Spin Glass*, in [21].
- [49] M.-H. Hao and H.A. Scheraga, *Monte Carlo Simulation of a First-Order Transition for Protein Folding*, J. Phys. Chem. **98** (1994), 4940–4948.
- [50] B. Hesselbo and R. Stinchcombe, *Monte Carlo Simulations and Global Optimization without Parameters*, Phys. Rev. Lett. **74** (1995), 2151–2155.
- [51] J. Higo, N. Nakajima, H. Shirai, A. Kidera and H. Nakumara, *Two-Component Multicanonical Monte Carlo Method for Effective Conformation Sampling*, J. Comp. Chemistry **18** (1997), 2086–2092.
- [52] K. Huang, *Statistical Physics*, John Wiley & Sons, 1987.
- [53] K. Hukusima and K. Nemoto, *Exchange Monte Carlo Method and Applications to Spin Glass Simulations*, J. Phys. Soc. Japan **65** (1996), 1604–1608.
- [54] J.E. Hunter III and W.P. Reinhardt, *Finite-Size Scaling Behavior of the Free Energy Barrier between Coexisting Phases: Determination of the Critical Temperature and Interfacial Tension of the Lennard-Jones Fluid*, J. Chem. Phys. **103** (1995), 8627–8637.
- [55] Y. Iwasaki, K. Kanaya, L. Kärkkäinen, K. Rummukainen and T. Yoshié, *Interface Tension in Quenched QCD*, Phys. Rev. **D49** (1994), 3540–3545.
- [56] W. Janke, B. Berg and M. Katoot, *Monte Carlo Calculation of the Surface Free Energy for the Two-Dimensional 7-State Potts Model, and an Estimate for Four-Dimensional SU(3) Gauge Theory*, Nucl. Phys. **B382** (1992) 649–661.
- [57] W. Janke and T. Sauer, *Multicanonical Multigrid Monte Carlo Method*, Phys. Rev. **E49** (1994), 3475–3479.
- [58] W. Janke and S. Kappler, *Multibondic Cluster Algorithm for Monte Carlo Simulations of First-Order Phase Transitions*, Phys. Rev. Lett. **74** (1995), 212–215.
- [59] K. Kajantie, L. Kärkkäinen and K. Rummukainen, *Interface Tension and Structure at T_c in the Potts Model*, Phys. Lett. **B223** (1989), 213–217.
- [60] F. Karsch, T. Neuhaus, and A. Patkós, *Effective Scalar Field Theory for the Electroweak Phase Transition*, Nucl. Phys. **B441** (1995), 629–657.
- [61] W. Kerler and P. Rehberg, *Simulated Tempering Procedure for Spin-Glass Simulations*, Phys. Rev. **E50** (1994), 4220–4225.
- [62] A. Klümper, A. Schadschneider and J. Zittartz, *Inversion Relations, Phase Transitions and Transfer Matrix Excitations for Special Spin Models in Two Dimensions*, Z. Phys. **B76** (1989), 247–258.
- [63] A. Klümper, *Investigation of Excitation Spectra of Exactly Solvable Models Using Inversion Relations*, Int. J. Mod. Phys. **B4** (1990), 871–893.
- [64] Y. Lee and M.Y. Choi, *Optimization by Multicanonical Annealing and the Traveling Salesman Problem*, Phys. Rev. **E50** (1994), R651–R654.
- [65] A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkanov and P.N. Vorontsov-Velyaminov, *New Approach to Monte Carlo Calculation of the Free Energy: Method of Expanded Ensembles*, J. Chem. Phys. **96** (1992), 1776–1783.
- [66] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, *Equation of State Calculations by Fast Computing Machines*, J. Chem. Phys. **21** (1953), 1087–1092.
- [67] E. Marinari and G. Parisi, *Simulated Tempering: A New Monte Carlo Scheme*, Europhys. Lett. **19** (1992), 451–458.
- [68] E. Marinari, G. Parisi and J.J. Ruiz-Lorenzo, in *Spin Glasses and Random Fields*, A.P. Young (editor), World Scientific, Singapore, 1997.
- [69] I. Montway, *Electroweak Phase Transition and Numerical Simulation in the SU(2) Higgs Model*, Proceedings of the 3rd Colloquium on Cosmology, Paris, June 7–9, 1995, H. de Vega and S. Sanchez (editors), World Scientific, Singapore 1995.
- [70] J.J. Morris, *Finite Size Scaling of the Interfacial Tension*, J. Stat. Phys. **69** (1991), 539–571.
- [71] T. Munakata and S. Oyama, *Adaption and Linear-Response Theory*, Phys. Rev. **E54** (1996), 4394–4398.
- [72] P.M.C. Oliveira, *Broad Histogram Relation is Exact*, Eur. Phys. J. **B6** (1998), 111–115.
- [73] J. Potvin and C. Rebbi, *Simple and Efficient Procedure for the Monte Carlo Calculation of the Surface Free Energy in Coexisting Phases*, Phys. Rev. Lett. **62** (1989) 3602–3065.
- [74] J. Propp and D. Wilson, *Coupling from the Past Users’s Guide*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science (AMS) **41** (1998), 181–192.

- [75] Z.W. Salsburg, J.D. Jackson, W. Fickett and W.W. Wood, *Application of the Monte Carlo Method to the Lattice-Gas Model. I. Two-Dimensional Triangular Lattice*, J. Chem. Phys. **30** (1959), 65–72.
- [76] G.R. Smith and A.D. Bruce, *A Study of the Multi-Canonical Monte Carlo Method*, J. Phys. **A28** (1995), 6623–6643.
- [77] G.R. Smith and A.D. Bruce, *Multicanonical Monte Carlo Study of Solid-Solid Phase Coexistence in a Model Colloid*, Phys. Rev. **E53** (1996), 6530–6546.
- [78] G.M. Torrie and J.P. Valleau, *Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling*, J. Comp. Phys. **23** (1977), 187–199.
- [79] C. Tsallis, *Possible Generalization of Boltzmann-Gibbs Statistics*, J. Stat. Phys. **52** (1988), 479–487.
- [80] J.P. Valleau, *Density-Scaling: a New Monte Carlo Technique in Statistical Mechanics*, J. Comp. Phys. **96** (1991) 193–216.
- [81] J.P. Valleau, *Density Scaling Monte Carlo Study of Subcritical Lennard-Jonesium*, J. Chem. Phys. **99** (1993) 4718–4728.
- [82] J. Villain, *Spin Glass with Non-Random Interactions*, J. Phys. **C10** (1977), 1717–1734.
- [83] J.-S. Wang, *Is the Broad Histogram Random Walk Correct?*, Eur. Phys. J. **B8** (1999), 287–291.
- [84] N.B. Wilding, *Critical-Point and Coexistence-Curve Properties of the Lennard-Jones Fluid: A Finite Size Scaling Study*, Phys. Rev. **E52** (1995), 602–611.
- [85] N.B. Wilding and M. Müller, *Liquid-Vapor Asymmetry in pure Fluids: A Monte Carlo Simulation Study*, J. Chem. Phys. **102** (1995), 2562–2573.
- [86] N.B. Wilding, F. Schmid and P. Nielaba, *Liquid-Vapor Phase Behavior of a Symmetrical Binary Fluid Mixture*, Phys. Rev. **E58** (1998), 2201–2212.
- [87] S.-Y. Zinn and M.E. Fisher, *Universal Surface-Tension and Critical-Isotherm Amplitude Ratios in Three Dimensions*, Physica **A226** (1996), 168–180.