Eur. Phys. J. B **29**, 481–484 (2002)
DOI: 10.1140/epjb/e2002-00328-1

THE EUROPEAN PHYSICAL JOURNAL B

Optimized Monte Carlo analysis for generalized ensembles

J. Ferkinghoff-Borg[a]

Niels Bohr Institute Blegdamsvej 17, 2100 Ø, Denmark
and
Condensed Matter Physics and Chemistry Department, Risø National Laboratory, Denmark

Received 4 June 2002
Published online 14 October 2002 –

Abstract. We present a new, maximum-likelihood based method to combine data from a multiple number of Monte Carlo simulations performed within any type of ensemble. The method offers an efficient iterative scheme to obtain the density of states of a wide range of energies as well as of other macroscopic variables. It should in particular be useful for the study of systems with a rough energy landscape.

PACS. 02.50.Ng Distribution theory and Monte Carlo studies – 02.70.Rr General statistical methods – 05.10.Ln. Monte Carlo methods – 05.50.+q Lattice theory and statistics

The Monte Carlo (MC) technique has proven successful for studying equilibrium statistical mechanics as well as other fields [1,2]. The basic idea is to sample the phase space by generating a Markov chain of states through a fixed matrix of transition probabilities. These probabilities are chosen so the condition of detailed balance is fulfilled for the statistical ensemble in concern. For instance, the standard Metropolis algorithm [3] samples directly in the canonical ensemble by the choice of Boltzmann transition probabilities.

A generic deficiency of the Metropolis sampling technique is the *slow relaxation* of the Markov chain which typically appears at points of phase transitions or at low temperatures. For systems with a rough energy landscape, such as spin glasses [4], random heteropolymers [5] or even spin-interacting homopolymers [6], this problem effectively causes the chain to be *trapped* in the phase space around local energy minima. In any case, slow relaxation easily leads to results which are errorously sensitive to the initial states of the Markov chain.

In the past few decades, a variety of MC-methods, based on non-Boltzmann transition probabilities, have been developed to improve the phase space sampling. This includes different types of 'extended' or generalized ensembles (GE) [7] in which the canonical probability weights ω are replaced with an explicit function, f, of the density of states, g, *i.e.* $\omega = f(g)$. The most prominent examples of GE-ensembles are the multicanonical approach [8–10], the $1/k$-ensemble [11] and simulated tempering [12,13]. They differ in the choice of the function f. Since the density of states is not *a priori* known, the GE-methods involve some additional complications compared to canonical simulations. The standard approach is to use an iterative procedure, where the weights for the $i+1$-simulation are calculated from histogram, $N_i(E)$, of energies E, and weights, $\omega_i(E)$, of the ith simulation. Despite several successful applications of this iteration method (see *e.g.* [7–15]), it suffers from the *loss of statistics* inherited in the updating rule for the weights, because the information obtained from the previous $i-1$ simulations is neglected.

In this paper, we present a maximum-likelihood method for combining the data from an arbitrary number of GE-simulations. The method can be used with any weight-scheme and any simulation method that provides data for a system in equilibrium. In this respect, it serves as the generalization of the multi-histogram equations, originally proposed by Ferrenberg and Swendsen [16] to combine data from canonical simulations. Specifically, the method suggests an iterative procedure for the GE-weights, that can account for the statistical information obtained from in principle all previous iterations. This allows for a more 'adiabatic' updating of the weights to speed up the convergence, thus enabling simulations on system sizes beyond those tractable by GE-methods hitherto. We demonstrate these two points by simulating a $L \times L$ Ising square lattice system with nearest neighbor coupling up to $L = 256$ and compare the results with other MC-methods. Our method offers better accuracy with the same simulation effort, in particular for large systems. The 2D Ising system is generally perceived as an ideal benchmark for simulation algorithms [16,17].

Consider M independent simulations, $i = 1,..,M$, with some arbitrary weights $\{\omega_i(E)\}_{i=1}^M$ and resulting histograms $\{N_i(E)\}_{i=1}^M$. The probability, $p_i(E\ |g)$, for

[a] e-mail: borg@alf.nbi.dk

observing an energy E within a restricted energy domain, $E \in \mathcal{S}_i$, in the ith simulation is given by

$$p_i(E|g) = \frac{g(E)\omega_i(E)}{Z_i}, \qquad Z_i = \sum_{E\in\mathcal{S}_i} \omega_i(E)g(E) \ .$$

Here, $g(E)$ is the density of states for energy E, and Z_i is the partition sum restricted to the domain $\mathcal{S}_i$ in order to insure the proper normalization. We will assume, that the ith simulation is equilibrated within a region $\mathcal{S}_i$ defined by the support of the observed histogram N_i, *i.e.* $\mathcal{S}_i = \{E \mid N_i(E) > 0\}$. In this case, the histogram $N_i(E)$ will be a member of the multinomial [18] probability distribution P_i;

$$P_i(N_i|g) = n_i! \prod_{E\in\ \mathcal{S}_i} \frac{p_i(E|g)^{N_i(E)}}{N_i(E)!} \ ,$$

where $n_i = \sum_E N_i(E)$ is the total number of counts for simulation i. The *likelihood* $\mathcal{L}$ for observing the full set of histograms $\{N_i\}_{i=1}^M$ is given by

$$\mathcal{L}(\{N_i\}_{i=1}^M|g) = \prod_{i=1}^M P_i(N_i|g).$$

An efficient [18] estimate, $\hat{g}$, of the true density of states can now be obtained by maximizing $\mathcal{L}$ with respect to g. This leads to the expression,

$$\hat{g}(E) = \frac{\sum_{i=1}^M N_i(E)}{\sum_{i=1}^M \chi_i(E) n_i \omega_i(E) Z_i^{-1}} \ , \tag{1}$$

where $\chi_i(E) = 1$ if $E \in \mathcal{S}_i$ and zero otherwise.

The reason for introducing the restricted regions, $\mathcal{S}_i$, into the equations can be deduced from equation (1). Through the presence of the χ_i variables the result of simulation i will only influence the overall estimate of the density of states at a given energy, if this energy belongs to the visited part of the phase space for that simulation.

The partition functions Z_i must be estimated self-consistently from equation (1). This set of estimates $\{\hat{Z}_i\}$ can be expressed as the solution to the M equations, $i = 1, .., M$:

$$\sum_{E\in\mathcal{S}_i} \frac{\sum_{t=1}^M N_t(E)}{\sum_{t=1}^M \chi_t(E) n_t \frac{\omega_t(E)}{\omega_i(E)} \frac{\hat{Z}_i}{\hat{Z}_t}} - 1 = 0 \ . \tag{2}$$

Since an arbitrary constant can be multiplied to each $\hat{Z}_i$ without affecting the solution, one has to fix an absolute normalization constant to assure uniqueness, for example by setting $Z_1 = 1$ or choosing a reference density of states. The estimate of the density of states is obtained by inserting the solution, $\{\hat{Z}_i\}$, into equation (1). Two aspects of these equations should be noted. 1) If $\mathcal{S}_i$ is replaced with the set of all possible energies and the weights ω_i are identified with the canonical Boltzmann weights $\omega_i(E) = \exp(-\beta_i E)$ for a set of inverse temperatures $\{\beta_i\}_{i=1}^M$ the equations are reduced to the multi-histogram equations [16]. For this reason we refer to equations (1–2) as the generalized multi-histogram (GMH) equations. 2) The equations can straightforwardly be applied to multiparametric problems as well. Bivariate representations of the density of states have for instance successfully been applied to spin systems with total energy and magnetization as independent variables [21]. As a practical matter, the estimate of the density of states of a given energy E is only reliable if the accumulated statistics is high enough, say $\sum_t N_t(E) > \delta$. A reasonable choice of δ will depend on the degree of correlation between the configurations of the Markov Chain in the simulations.

Recently, MC-methods have been developed to estimate the density of states on the basis of transition observables [19] rather than on the histogram information. This includes the flat histogram [20], the broad histogram method (BHM) [21–23] and the transition matrix approach (TM) [2,24], These methods are ideal for discrete systems, for which the MC-dynamics and the Hamiltonian are local, such as the 2D Ising system with single flip dynamics. However, transition observables can not in general be obtained as $O(1)$ in computer time, if the Hamiltonian or the MC-dynamics are non-local (*e.g.* cluster flip algorithms for spin systems [25] or pivot-moves for polymer systems [26]). The BHM-method has recently been extended to the XY-model [27], but for a broader class of continuous systems the implementation and robustness of the transition observable based methods remains to be demonstrated. Since the method proposed here is based on histogram-analysis, it involves no further computational cost at each MC-move, besides the move itself, and is thus directly compatible with any choice of Hamiltonian and dynamics. The same advantage is offered by the random walk (RW) method, proposed by Wang and Landau [17].

For the study of the 2D-Ising system, we will –for comparative reasons only– restrict ourselves to the local single spin dynamics. Thus, we choose a spin at random and flip it with the probability $p = \min\{1, \omega(E')/\omega(E)\}$, where E' and E is the energy of the new and the old state respectively. In this respect, the configurations along the Markov chain will be strongly correlated, therefore a high cut-off value for the statistics has been chosen ($\delta \sim 100$–300).

The GE-iteration scheme is performed in the following manner. Initially, we set $\omega_{i=1}(E) = 1$ for all energies and perform only a few number of MC-sweeps. We use $n_1 = 50L^2$ (50 sweeps), but any small choice will do. The weights for iteration $i+1$ are calculated from the estimate of the density of states using the GMH-equations with the previous $\min\{i, M\}$ histograms. The convergence generally improves with increasing 'memory', M, of the iteration scheme, so a rather high value, $M = 20$, is used. Note, that the $M = 1$ case corresponds to the standard iteration procedure. In the multicanonical approach the new weights will be given by $\omega_{i+1}(E) = 1/\hat{g}(E)$, whereas for the $1/k$-ensemble, $\omega_{i+1}(E) = (\sum_{E'\leq E} \hat{g}(E'))^{-1}$. For energies where no reliable estimate of the density of states is available ($\{E| \sum_t N_t(E) < \delta\}$) we will choose normal

Table 1. Error for the random walk method (RW), transition matrix method (TM), broad histogram method (BHM) and the generalized multihistogram method (GMH) after 10^6 MC-sweeps on the $L \times L$ Ising square lattice. The errors of the GMH-method represent averages over 10 independent runs. The values are taken from (a) [2], (b) [23] and (c) [17] ; this value represents the best result of the RW-method using only 7×10^5 MC-sweeps.

	$L = 32$		$L = 50$	
Error	$\epsilon(\hat{g})$	$\epsilon(\log(\hat{g}))$	$\epsilon(\hat{g})$	$\epsilon(\log(\hat{g}))$
RW	$0.36^{(a)}$	$0.035\%^{(c)}$	$9(\pm 4)^{(a)}$	
TM	$0.07^{(a)}$		$0.22(\pm 0.02)^{(a)}$	
BHM		$\approx 0.08\%^{(b)}$		
GMH	0.07	0.08%	0.14	0.14%

Boltzmann weights with the inverse temperature $\beta = \frac{\partial \log \hat{g}}{\partial E}$, evaluated numerically at the border of the region, $\mathcal{O}_i = \{E | \sum_{t=1}^{i} N_t(E) > \delta\}$. This merely corresponds to a linear extrapolation of the function $-\log(\hat{g}(E))$. The simulation time for the $i+1$ iteration is increased with a factor γ compared to the previous iteration, $n_{i+1} = \gamma n_i$, as long as the ith simulation has not revealed any new energies (*i.e.* $\mathcal{S}_i \subseteq \mathcal{O}_{i-1}$). Otherwise, we leave the simulation time unchanged, $n_{i+1} = n_i$ [28]. This strategy represents a compromise between two facts. On one hand, the probability of visiting new energies in iteration $i+1$ decreases exponentially away from the borders of $\mathcal{O}_i$, due to the exponential nature of the density of states and the fact that the weights here are based on an extrapolation. Therefore, convergence is more efficiently obtained by an 'adiabatic' change of the weights through short simulations, rather than by performing long simulations at each iteration step. On the other hand, the histogram N_i will fail to represent an equilibrium distribution if the simulation time is too short. We will simply assume this to be the case, whenever $\mathcal{S}_i \subseteq \mathcal{O}_{i-1}$. We present results with the choice $\gamma = 2^{1/10}$. Other values have been tested (in the range $2^{1/15} \leq \gamma \leq 2^{1/8}$) and the results compares well.

J.-S. Wang and R.H. Swendsen have recently carried out extensive comparisons between the RW-method and the best transition matrix based method (TM) [2], which combines the flat histogram weights and the N-fold algorithm for the dynamics. The accuracy of these two methods was tested on the 2D Ising system on system sizes up to $L = 50$, by measuring the average error, $\epsilon(\hat{g}) = \langle |\hat{g}(E)/g(E) - 1| \rangle_E$, of the estimated density of states compared to the exact calculation of Beale [29]. Here, the absolute values for $\hat{g}$ were determined by using the condition that the number of ground states for the Ising model is 2 (all spins up or down). Another measure of the accuracy is given by the error of the entropy, $\epsilon(\log(\hat{g})) = \langle |\log(\hat{g}(E))/\log(g(E)) - 1| \rangle_E$. The value of $\epsilon(\log(\hat{g}))$ averaged over 10 independent runs was measured with the broad histogram method (BHM) [23] for $L = 32$ using 10^6 MC-sweeps. We have tested the GMH-method for $L = 32$ and $L = 50$ with the same simulation time and averaging procedure, using the multicanonical weight scheme. The results are summarized in Table 1.

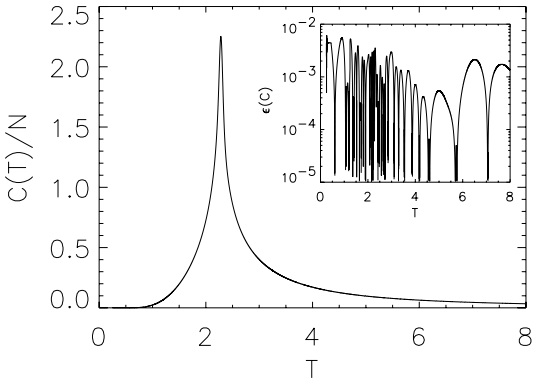

Fig. 1. The estimated and exact specific heat capacity $C(T)/N$ as function of temperature T for $L = 64$ 2D Ising system. Since no difference is visible, the relative error $\epsilon(\hat{C}(T)) = |1 - \hat{C}(T)/C(T)|$ is shown in the inset.

The GMH-method presented here is comparable in accuracy for $L = 32$ with the transition observable based methods (TM and BHM) and noticeable better for $L = 50$. However, keeping in mind that the MC-move for the transition based methods is least a factor of two slower, the values compare favorable for the GMH-method in both cases.

The combination of the TM approach with the flat histogram dynamics has not been extended beyond $L = 50$, but the TM reweighting technique has been applied for the 2D Ising system with $L = 64$, using 25 different temperatures and a total simulation time of 2.5×10^7 sweeps [24]. The calculated specific heat, $\hat{C}$, as function of the temperature T, was compared with the exact result, C, of Fisher and Ferdinand [30] in the region $T \in [0, 8]$. The relative error, $\epsilon(\hat{C}(T)) = |\hat{C}(T)/C(T) - 1|$, was shown to have a maximum $\epsilon(\hat{C}(T_{cr})) \approx 7 \times 10^{-3}$ around the critical temperature, T_{cr} (except for a higher value $\epsilon(\hat{C}) \approx 10^{-2}$ at low temperatures, which is mainly due to small values of the specific heat capacity here). The same precision was quoted with the RW-method using a total simulation time of 2×10^7 sweeps [17]. In Figure 1 we show the error obtained by the GMH-equations using the multicanonical ensemble with the same number of sweeps as for the RW-method. A maximum error, $\epsilon(\hat{C}(T)) \leq 6 \times 10^{-3}$, is found in the low temperature region, whereas the error around the critical region, $\epsilon(\hat{C}(T_{cr})) \leq 3.5 \times 10^{-3}$, is a factor of two smaller than the two other methods.

The full density of states for the $L = 256$ 2D Ising system has so far only been calculated by the RW-method [17]. Here, convergence was achieved by performing 15 independent simulations in different restricted part of the energies [31] with a total simulational effort of 6.1×10^6 MC-sweeps. However, albeit the fact that this parallelization procedure significantly reduces the required simulation time, a supplementing recipe for exchanging states between the different energy parts is generally needed. In lack of such recipe, the simulation will potentially suffer from the same type of deficiencies as the standard Metropolis algorithm for systems with a more rugged energy landscape. It should also be noted, that all

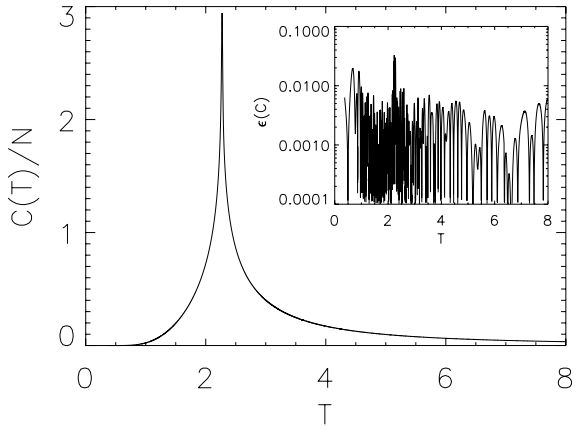

Fig. 2. The estimated and exact specific heat capacity $C(T)/N$ as function of temperature T for $L = 256$ 2D Ising system. Since no difference is visible, the relative error $\epsilon(\hat{C}(T)) = |1 - \hat{C}(T)/C(T)|$ is shown in the inset.

of the above mentioned MC-methods are directly compatible with the parallelization strategy used in [17]. Consequently, applying this strategy does not seem to benchmark any of these methods as to their robustness towards large systems. To demonstrate the scalability properties of the present method, we have therefore chosen to simulate the $L = 256$ 2D Ising system without any changes to the calculation scheme, except for the replacement of the multicanonical ensemble with the $1/k$-ensemble, only to enforce a faster search towards low energies. More precisely, as for the smaller systems we carry out the simulation in a *single* process, allowing all possible energies to be visited in each step of the iteration. For future comparison, we obtain the same precision in the specific heat capacity as quoted in [17] ($\epsilon(\hat{C}(T_{cr})) = 4.5\%$ and an average error of $\bar{\epsilon}(\hat{C}) = 0.39\%$) after 2.3×10^7 MC-sweeps. Naturally, a better estimate is obtained by extending the simulation time. At 3×10^7 MC-sweeps the precision yields $\epsilon(\hat{C}(T_{cr})) = 3.3\%$ and $\bar{\epsilon}(\hat{C}) = 0.27\%$. The result shown in Figure 2.

In conclusion, we have presented a new and general method to estimate the density of states. It is based on a maximum-likelihood approach to combine the histogram information from a multiple number of MC-simulations performed within any type of generalized ensemble. The method is robust towards large systems and it can directly be applied to non-local MC-dynamics, to complicated Hamiltonians (with or without continuous degrees of freedom) and to multivariate representations of the density of states. These last points are successfully demonstrated in a forthcoming publication on 3D off-lattice homopolymers with hydrogen-bonds, where the thermodynamic behavior is reconstructed from a bi-variate representation of the density of states [32].

It is a pleasure for me to acknowledge the helpful discussions with C. Rischel, M.H. Jensen, K. Sneppen, P.-A. Lindgaard, G. Tiana and in particular A. Trovato. I would also like to thank P.D. Beale for the MATHEMATICA code to calculate the exact density of states of the Ising systems up to $L = 50$.

References

1. *The Monte Carlo Method in Condensed Matter Physics, Topics in Applied Physics*, Vol. 71, 2nd edn., edited by K. Binder (Springer-Verlag, 1995). D.P. Landau, K. Binder, *A Guide to Monte Carlo Methods in Statistical Physics* (Cambridge University Press, Cambridge, England, 2000)
2. J.-S. Wang, R.H. Swendsen, J. Stat. Phys. **106**, 245 (2002)
3. N. Metropolis *et al.*, J. Chem. Phys. **21**, 1087 (1953)
4. M. Mezard, G. Parisi, M. Virasoro, *Spin Glasses and beyond* (World Scientific, New York, 1988)
5. E.I. Shakhnovich, A.M. Gutin, J. Phys. A **22**, 1647 (1989)
6. J. Borg, M.H. Jensen, K. Sneppen, G. Tiana, Phys. Rev. Lett. **86**, 1031 (2001)
7. U.H.E. Hansmann, Y. Okamoto, J. Comput. Chem. **18**, 920 (1997)
8. B.A. Berg, T. Neuhaus, Phys. Lett. B **267**, 249 (1991); Phys. Rev. Lett. **68**, 9 (1992)
9. J. Lee, Phys. Rev. Lett. **71**, 211 (1993); Phys. Rev. Lett. **71**, 2353(E) (1993)
10. B.A. Berg, U.H.E. Hansmann, Y. Okamoto, J. Phys. Chem. **99**, 2236 (1995)
11. B. Hesselbo, R.B. Stinchcombe, Phys. Rev. Lett. **74**, 2151 (1995)
12. A.P. Lyubartsev *et al.*, J. Chem. Phys. **96**, 1776 (1992)
13. E. Marinari, G. Parisi, Europhys. Lett. **19**, 451 (1992)
14. A. Irbäck, E. Sandelin, J. Chem. Phys. **110**, 12256 (1999)
15. N.A. Alves, U. Hansmann, Phys. Rev. Lett. **84**, 1836 (2000)
16. A.M. Ferrenberg, R.H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989), Phys. Rev. Lett. **63**, 1658(E) (1989)
17. F. Wang, D.P. Landau, Phys. Rev. Lett. **86**, 2050 (2001); Phys. Rev. E **64**, 056101 (2001)
18. A. Stuart, S.K. Old, *The Advanced Theory of Statistics* (John Wiley and Sons Inc., New York), Vol. 1, 6th edn. (1994), Vol. 2a (5th edn. 1991)
19. M. Kastner, M. Promberger, J.D. Muñoz, Phys. Rev. E **62**, 7422
20. J.-S. Wang, Eur. Phys. J. B **8**, 287 (1999)
21. A.R. Lima, P.M.C. de Oliveira, T.J.P. Penna, Solid State Commun. **114**, 447 (2000); *ibid.* **115**, 395 (2000)
22. P.M.C. de Oliveira, T.J.P. Penna, H.J. Herrmann, Eur. Phys. J. B **1**, 205 (1998); P.M.C. de Oliveira, Braz. J. Phys. **30**, 195 (2000)
23. A.R. Lima, P.M.C. de Oliveira, T.J.P. Penna, J. Stat. Phys. **99**, 691 (2000)
24. J.-S. Wang, T.K. Tay, R.H. Swendsen, Phys. Rev. Lett. **82**, 476 (1999)
25. U. Wolff, Phys. Rev. Lett. **62**, 361 (1989)
26. N. Madras, A.D. Sokal, J. Stat. Phys. **50**, 109 (1988)
27. J.D. Muñoz, H.J. Herrmann, Comp. Phys. Comm. **121-122**, 13 (1999)
28. private communication with A. Trovato
29. P.D. Beale, Phys. Rev. Lett. **76**, 78 (1996)
30. A.E. Ferdinand, M.E. Fisher, Phys. Rev. **185**, 832 (1969)
31. Additional 15 separate random walks was performed in the energy range around the critical temperature at a later stage of the simulation [17].
32. A. Trovato, J. Ferkinghoff-Borg, M.H. Jensen, `cond-mat/0112076`, submitted to Phys. Rev. E