FLAT HISTOGRAM MONTE CARLO FOR LOW TEMPERATURE SIMULATIONS

LEI-HAN TANG
Department of Physics, Hong Kong Baptist University, Kowloon Tong, Hong Kong
E-mail: lhtang@hkbu.edu.hk

The flat histogram Monte Carlo (FHMC) algorithm has been proposed as an efficient sampling scheme for problems with a complex free energy landscape. Its successful implementation requires fast and stable determination of the sampling weight function which can be a challenge for simulation at low temperatures. We describe here a polynomial parametrization of the sampling weight function which allows one to perform noise filtering and extrapolation at the same time. Efficiency of the scheme as compared to Berg's original iterative formula is demonstrated on the two-dimensional compass model for d-orbital ordering.

1 Why canonical sampling is not always good enough

In the study of systems at thermal equilibrium, Monte Carlo simulations have proven to be a powerful tool for extracting detailed and quantitative description of the properties of the system in the temperature range of interest. The basic idea of this approach, dating back to Metropolis and coworkers[1], is very simple in principle. At a given temperature T, the average value of a physical quantity X is given by,

$$\langle X \rangle = \frac{1}{Z} \sum_{\alpha} X_\alpha \exp(-E_\alpha/T), \tag{1}$$

where $Z = \sum_\alpha \exp(-E_\alpha/T)$ is the partition function and the sum is over all allowed states of the system. In a Monte Carlo simulation, instead of summing over all possible states, one generates a set of sample configurations $\{\alpha_1, \alpha_2, \ldots, \alpha_s\}$ according to a prescribed probability distribution function P_α. The formula

$$X_P = \frac{\sum_{i=1}^{s} X_{\alpha_i} P_{\alpha_i}^{-1} \exp(-E_{\alpha_i}/T)}{\sum_{i=1}^{s} P_{\alpha_i}^{-1} \exp(-E_{\alpha_i}/T)} \tag{2}$$

is then used to compute an approximate value for $\langle X \rangle$. The *naive* choice $P_\alpha = const.$, known as uniform sampling, gives rather poor performance when the number of degrees of freedom is large. The usual practice is to choose

$$P_\alpha = Z^{-1} \exp(-E_\alpha/T), \tag{3}$$

i.e., to pick samples according to their statistical importance in the partition sum.

Obviously, to obtain an accurate estimate of $\langle X \rangle$, we need to focus on configurations that give significant contributions to the sum in (1). In many problems of interest in physics, where the system is spatially uniform, such configurations also have very similar physical properties in the sense that their values of X are quite similar. In this case, computing $\langle X \rangle$ using the "canonical sampling weights" Eq. (3) is quite adequate. There are, however, other cases where configurations that make significant contributions to $\langle X \rangle$ (and other quantities of interest) have either distinct or a broad spectrum of physical properties. This is what happens at phase transitions, or in glassy systems with a complex energy landscape. In such circumstances, in order to achieve an accurate estimate of $\langle X \rangle$, one needs to not only visit all these configurations during sampling, but also make sure that their relative frequencies are in accordance with the correct weights. This type of systems have been the focus of novel Monte Carlo sampling schemes in recent years[2,3,4].

2 The flat-histogram Monte Carlo methodology

In usual implementation of Monte Carlo simulations, sample configurations are generated through a sequence of "local moves", each time modifying the current configuration at one or a few sites only. (Parallel updating in most applications can be viewed as a collection of independent local moves.) Such moves are usually easy to construct and can be readily made to satisfy the detailed balance condition. However, simulations based on local moves may fail to produce accurate values for the statistical averages if configurations which need to be sampled sufficiently are separated by large barriers, i.e., the path that connects such configurations needs to go through regions of very low probability in the configuration space, which may be loosely called "transition states".

The flat-histogram Monte Carlo (FHMC) methodology[2,5] tackles this problem by demanding that the transition states are visited as frequently as the "equilibrium states" that make important contributions to the statistical averages such as (1). Provided the local moves are ergodic, this can always be achieved through a suitable choice of the sampling probability P_α as discussed below. In the best scenario, the system will then be able to travel freely back and forth between the equilibrium states in the simulation to produce a fair sampling. The transition states may themselves be part of the equilibrium state at higher temperatures, in which case one gains an extra-advantage of obtaining statistics over a broad temperature range in a single Monte Carlo run.

The FHMC methodology has been applied to a number of computationally hard problems in statistical physics with varying degrees of success[6]. In many applications, energy is a natural "reaction coordinate" that can be used to parametrize the transition states. In a "multicanonical Monte Carlo" (MCMC) simulation, the sampling probability P_α is chosen to be a function of the energy E_α. In particular, the choice

$$P_\alpha = \rho^{-1}(E_\alpha) = \exp[-S(E_\alpha)] \tag{4}$$

yields a flat energy histogram when the simulation is run for a sufficiently long time. Here $\rho(E)$ is the density of states (or configurations) in energy, and $S(E) = \ln \rho(E)$ is the Boltzmann entropy. From the thermodynamic relation

$$T^{-1} = \partial S/\partial E, \tag{5}$$

we may write the sampling probability around a given state α_0 with energy E_0 and probability $P_0 = P(E_0)$ as

$$P_\alpha = P_0 \exp[S(E_0) - S(E_\alpha)] \simeq P_0 \exp(-\Delta E_\alpha / T_0). \tag{6}$$

Hence the sequence of configurations generated in a MCMC simulation resembles locally those of constant temperature simulations. However, as the simulation continues, the temperature of the system drifts with its energy in such a way that there is no bias or restoring force towards any particular energy value.

For problems where the (free) energy barrier is the main obstacle towards sampling the statistically important but well-separated configurations, one may expect the MCMC scheme to perform well. Its success has been well documented in several recent reviews[4,6,7,8]. On the other hand, it has been argued that MCMC is not so effective when applied to problems such as the Ising spin glass in three dimensions[9]. While one may speculate about the existence of other important "reaction coordinates"[10], this issue has not been investigated sufficiently so far. In the following we shall focus on a technical aspect regarding implementation of the MCMC scheme, and leave the discussion of "good coordinates" for flattening histograms for future study.

3 Achieving fast and stable convergence

Unlike the canonical sampling weights (3), Eq. (4) can not be used directly in most cases as it contains an unknown function $S(E)$. Berg proposed an iterative procedure to overcome this difficulty[6,11]. Starting with a trial sampling weight function $P_k(E)$ at $k = 0$, a set of configurations are generated. Their energies are recorded in the energy histogram $h_k(E)$. Once sufficient number of samples are collected, the sampling weight function is updated using the formula

$$P_{k+1}(E) = P_k(E)/h_k(E). \tag{7}$$

In theory, a perfect $h_k(E)$ yields the desired $P_{k+1}(E)$ in one iteration. In practice, however, $h_k(E)$ contains statistical fluctuations due to the finite length of the simulation. More seriously, when $P_k(E)$ is far from the weight given by (4), there are energy intervals which are not visited at all in the simulation. Therefore the key to reaching optimal convergence of the recursion (7) is to minimize the influence of statistical fluctuations on the one hand, and to expand the range of energies with sufficient number of counts on the other hand.

We have implemented the following three-step procedure which combines noise-filtering with extrapolation to regions with poor or no statistics in each iteration. At the end of each sampling run, we first compute the function

$$\Delta_k(E) = \ln[h_k(E) + 1] \tag{8}$$

from the raw histogram data $h_k(E)$. Next, an energy interval $(E_{\min}, E_{\max})$ is determined from $h_k(E)$, based on the condition that for any E within this interval, the count $h_k(E)$ exceeds certain threshold value h_c. Normally, h_c is chosen to be 2% of the maximum count over all bins. The data

$$\tilde{R}_{k+1}(E) = R_k(E) + \Delta_k(E), \tag{9}$$

with $R_k(E) \equiv -\ln P_k(E)$, is then used for a least-squares fit to either a single polynomial function over the entire interval $(E_{\min}, E_{\max})$, or several polynomial functions over a set of sub-intervals, with the requirement that the polynomials covering neighboring sub-intervals join smoothly at their meeting point. For the examples discussed in the next section, a fifth order polynomial is sufficient for our purpose. In the final step, the interval $(E_{\min}, E_{\max})$ is extended at one or both ends to $(E'_{\min}, E'_{\max})$, with $E'_{\min} \le E_{\min} < E_{\max} \le E'_{\max}$. The polynomial functions are then used to assign values to $R_{k+1}(E) \equiv -\ln P_{k+1}(E)$ on the expanded interval. For $E < E'_{\min}$, $R_{k+1}(E)$ is approximated by a linear function with the slope given by the slope of the polynomial at $E = E'_{\min}$. This is equivalent to setting a minimum temperature $T_{\min,k+1}$ in the new round of sampling.

Figure 1 shows the energy histograms and the corresponding sampling weight functions under three different iterative schemes. The system simulated is the two-dimensional (2D) compass model described in the next section, with $N = 32 \times 32$ lattice sites. In all cases, the initial sampling probability function is chosen to be a constant, corresponding to the infinite temperature ensemble. The sampling time grows by a factor 1.3 in successive iterations, but is identical in all three cases. Figures 1(a) and (d) correspond to a straightforward implementation of (7) without noise filtering and extrapolation. In this case, the interval with significant statistics expands quite slowly under the iteration. In addition, part of the ruggedness in the histogram can be attributed to the spurious variations in the function $R_k(E)$. After the noise-filtering and extrapolation step is introduced, expansion of the interval with good statistics is much accelerated, as seen in Fig. 1(b). In the final round, however, the sampling probability density function increases too fast at the low energy end. Consequently, only configurations in a narrow energy range are visited in the simulation.

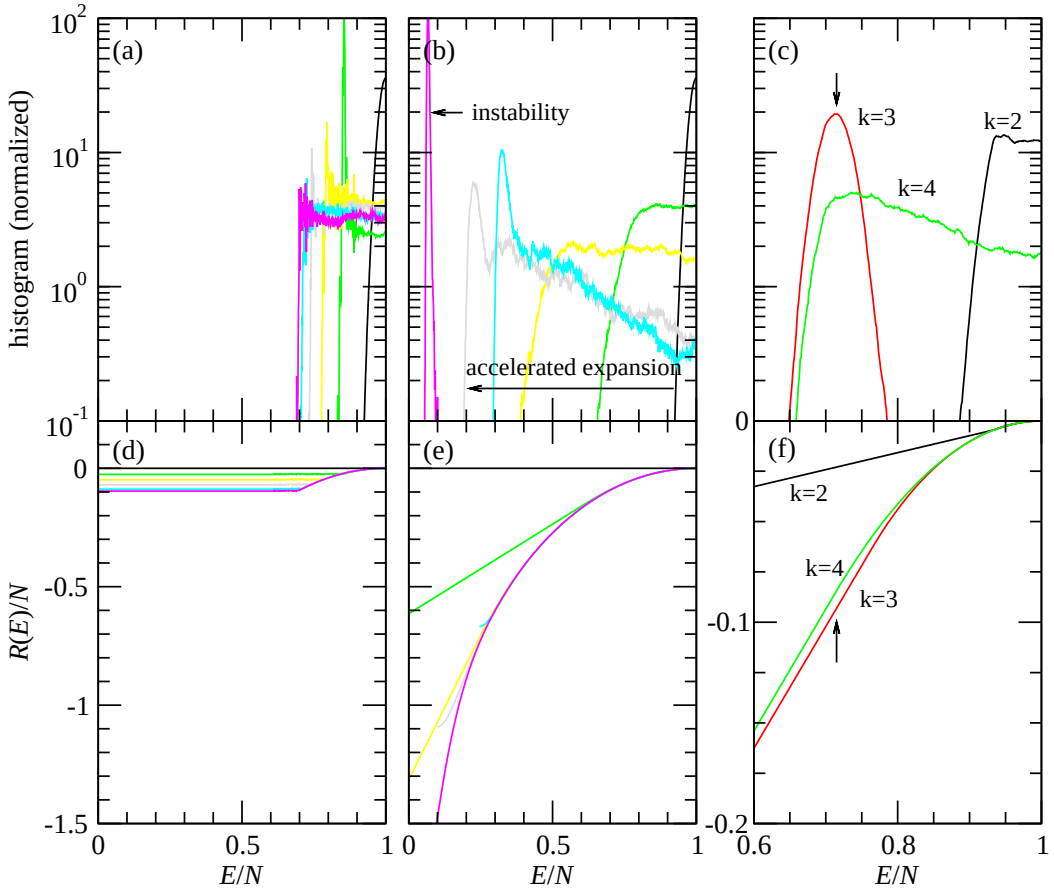

Figure 1. Energy histograms [(a)-(c)] and the corresponding sampling functions $R(E) = -\ln P(E)$ [(d)-(f)] under three different iterative schemes. See text for details.

The "instability" due to over-extrapolation is a serious obstacle against fast convergence to flat-histogram over the entire energy range, particularly for large systems and at low temperatures. Instead of reducing the range of extrapolation, which necessarily slows down convergence, one may find ways to restore the flat histogram by making use of the statistics already collected. In the example illustrated in Fig. 1(c), the problem appears at the iteration $k = 3$, where the histogram $h_3(E)$ runs away from the intended energy range that includes $E/N = 1$. Nevertheless, we may still use $h_3(E)$ to construct a good sampling function $R_4(E)$ in the interval $0.66 < E/N < 0.78$ where large number of counts are registered. On the other hand, from the previous run, we already have a good sampling function for $0.94 < E/N < 1$. Each of these functions, however, can be shifted up or down by a constant without affecting the sampling process. Taking this property into account, we first perform a least-squares fit to the derivatives of the two functions using a single polynomial. Integrating the polynomial allows us to determine how much one piece should be shifted against the other. The $k = 4$ curve in Fig. 1(f) shows the best polynomial that matches both pieces after the shift is performed. Using this sampling function, a flat histogram over the interval $(0.7, 1)$ is achieved in the next iteration. This procedure allows iteration to continue until the desired flat histogram is obtained.

4 Simulation results on the 2D compass model

We have implemented the above iteration scheme on a number of problems with success. In this section some results from the simulation of a classical 2D compass model are presented. The Hamiltonian of the model is given by[12,13],

$$H = -J \sum_i s_i^x s_{i+\hat{x}}^x + s_i^y s_{i+\hat{y}}^y. \tag{10}$$

Here $\vec{s}_i = (\cos\theta_i, \sin\theta_i)$ are classical XY spins on a square lattice. The problem arises in the study of the ordering of d-orbitals in certain transition metal oxides. An important aspect of the model is the directional coupling of the orbital states specified by $\vec{s}_i$. One is interested in the possibility of a directionally ordered state at low temperatures.

The classical ground state of (10) is highly degenerate due to the presence of continuous and discrete symmetries. The continuous $O(2)$ symmetry corresponds to a global rotation of all angles starting from a uniform state $\theta_i = \theta_0$, all i. This symmetry is restricted to the ground state. Indeed, self-consistent harmonic approximation applied to (10) yields an extra free energy gain of the order $T^{2/3}$ for the special orientations $\theta_0 = 0, \pm\pi/2, \pi$. This entropic advantage stabilizes orbital ordering at sufficiently low temperatures.

The one-dimensional (1D) coupling of the x and y components of $\vec{s}_i$ gives rise to a discrete symmetry. Take any row of spins, the transformation $s_i^x \to -s_i^x$ leaves the Hamiltonian (10) invariant. Similarly, the transformation $s_i^y \to -s_i^y$ for any column of spins leaves (10) invariant as well. Since this symmetry is 1D, it is not expected to be broken except at $T = 0$. A suitable order parameter for the low temperature phase is thus $q = N^{-1} \sum_i \cos(2\theta_i)$.

Figures 2(a)-(c) show the energy, entropy, and specific heat data against T from our MCMC simulation for seven system sizes $L = 8, 12, 16, 24, 32, 48$ and 64 under periodic boundary conditions (PBC). A background term due to classical harmonic motion of the angles has been taken out in the energy E/N and entropy S/N, respectively, so that change in these quantities in the transition region can be examined more closely. The discrete 1D symmetry gives rise to a 2×2^L degeneracy in the ordered phase at low temperatures. This translates to a finite-size correction $\Delta S_L = (L+1)\ln 2$ in the entropy of an $L \times L$ system under PBC. Such a contribution is evident in Fig. 2(b). It is quite remarkable that, in order to observe the effect, the 2^{L+1} degenerate configurations should all be visited with more or less the same frequency, which appears to be the case with the MCMC sampling.

The strong size-dependence of the specific heat curves shown in Fig. 2(c) makes it quite difficult to locate the transition temperature T_c. At first sight, both the energy and entropy seem to exhibit a finite jump at the transition, suggesting the transition is first order. More careful analysis shows that this is not the case. The first-order like behavior can be attributed to the ΔS_L term which gives extra stability to the low temperature phase in a finite system. Consequently, the specific heat peak acquires a strong size-dependence. The effect can be eliminated through the introduction of an "annealed" boundary condition, where the sign of the coupling constant J in (10) on a selected row and a selected column is allowed to fluctuate. In the low temperature ordered phase, the $-J$ configurations are energetically unfavorable and do not contribute to the equilibrium statistics. However, in the high temperature disordered phase, both $\pm J$ configurations contribute nearly equally. The ΔS_L term, which contributes only in the ordered phase under PBC, is now present in both phases under the annealed boundary condition. Figure 2(d) shows the specific heat curves in the latter case. The data can be well fitted to a logarithmic divergence at a $T_c \simeq 0.147$, as in the 2D Ising model. Details of our analysis, supported by an exact mapping to the 2D Ising model, can be found in Ref. 13.

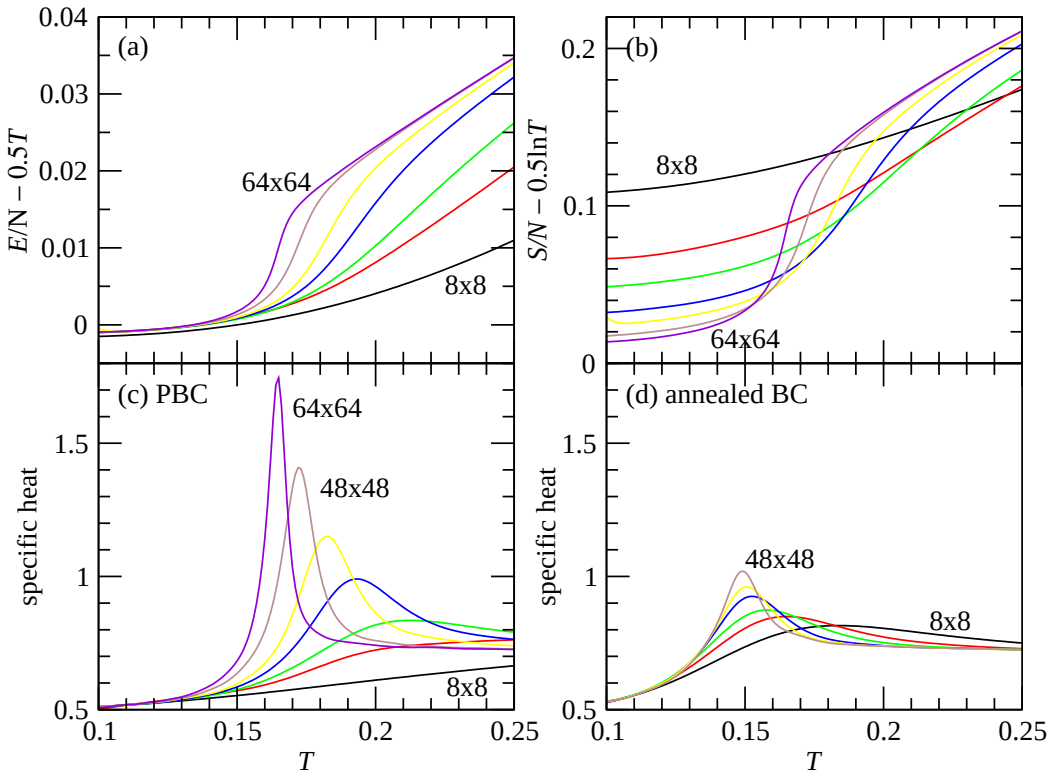

Figure 2. Multicanonical Monte Carlo simulation results on the two-dimensional compass model for various system sizes. (a) Energy against temperature; (b) Entropy against temperature; (c) Specific heat against temperature; (d) Specific heat against temperature under "annealed boundary condition". See text for details.

5 Conclusions

In simulation studies of equilibrium systems with a complex energy landscape, the flat histogram Monte Carlo scheme offers a generic and potentially powerful methodology to achieve efficient sampling of the statistically important configurations. Successful implementation of the idea, particularly in the case of low temperature simulations, requires improvements in the iterative determination of the sampling weight function. For many problems of interest, the optimal sampling weight function is smooth (except near phase transitions) and can be well-approximated by one or a few polynomials in the respective intervals. The use of polynomials to parametrize the sampling weight function allows one to achieve noise filtering and extrapolation at the same time in the iterative process. Success of this strategy is demonstrated in the simulation of the 2D compass model.

Acknowledgments

I would like to thank Mike Ma for introducing me to the d-orbital problem, and the Isaac Newton Institute, University of Cambridge for hospitality, where the manuscript is completed. The work is supported in part by the Research Grants Council of the HKSAR under grant HKBU 2076/02P. Computations were carried out at Hong Kong Baptist University's High Performance Cluster Computing Centre Supported by Dell and Intel.

References

1. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
2. B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
3. E. Marinari, G. Parisi, and J. J. Ruiz-Lorenzo, in *Spin Glasses and Random Fields*, A. P. Young Ed., World Scientific, Singapore (1998) pp. 59-98.
4. For a recent review, see Y. Okamoto, J. Mol. Graph Model. **22**, 425 (2004).
5. G. M. Torrie and J. P. Valleau, J. Comp. Phys. **23**, 187 (1977).
6. B. A. Berg, Fields Institute Communications **26**, 1 (2000).
7. U. H. Hansmann and Y. Okamoto, Ann. Rev. Comp. Phys. **6**, 129 (1999).
8. J. S. Wang and R. H. Swendsen J. Stat. Phys. **106**, 245 (2002).
9. P. Dayal, S. Trebst, S. Wessel, D. Würtz, M. Troyer, S. Sabhapandit, and S. N. Coppersmith, Phys. Rev. Lett. **92**, 097201 (2004); S. Alder, S. Trebst, A. K. Hartmann, and M. Troyer, cond-mat/0405409.
10. N. Hatano and J. E. Gubernatis, Phys. Rev. B **66**, 054437 (2002).
11. B. A. Berg and T. Celik, Phys. Rev. Lett. **69**, 2292 (1992).
12. D. I. Khomskii and M. V. Mostovoy, cond-mat/0304089.
13. A. Mishra, M. Ma, F.-C. Zhang, S. Gürtler, L.-H. Tang, and S. Wan, to be published.