A New Monte Carlo Algorithm for Free Energy Calculation

Sheng Bi\textsuperscript{1,2,} and Ning-Hua Tong\textsuperscript{1,2,}\
\textsuperscript{1}Department of Physics, Renmin University of China, 100872 Beijing, China\textsuperscript{}\textsuperscript{}
\textsuperscript{2}Beijing Key Laboratory of Opto-electronic Functional Materials and Micro-nano Devices (Renmin University of China)\textsuperscript{}\textsuperscript{}
(Dated: October 21, 2014)

We propose a new Monte Carlo algorithm for the free energy calculation based on configuration space sampling. We implement this algorithm for Ising model. Comparison with the exact free energy shows an excellent agreement. We analyse the properties of this algorithm and compares it with Wang-Landau algorithm which samples in energy space. This method is applicable to classical statistical models. The possibility of extending it to quantum systems is discussed.

I. INTRODUCTION

Monte Carlo simulation is one of the most important numerical methods for solving statistical problems in fields such as chemistry, biology and physics. In condensed matter physics, Monte Carlo is extensively used to study the properties of many statistical models, phase transitions, and quantum many-body systems. Often, besides evaluating expectation values of certain physical quantities, one needs to calculate the free energy of the system in thermal equilibrium. This is a difficult problem for traditional Metropolis algorithm since partition function plays the role of the normalization constant in the thermal probability density distribution of an ensemble, instead of an expectation value.

In the past three decades, various Monte Carlo algorithms have been proposed for the free energy calculation of statistical Hamiltonians. For classical systems, the frequently used ones are the histogram reweighting method, transition matrix Monte Carlo, entropic sampling, flat histogram method, and the Wang-Landau method. Recently a massive parallel version of the Wang-Landau method is realized. All these methods have their respective advantages and disadvantages. For examples, the histogram reweighting method produces the energy histogram $P_T(E)$ at a given temperature $T_0$ and employs the reweighting method to recover the distribution at a different temperature. As is shown below, usually the canonical distribution $P_T(E)$ is sharply peaked around $\langle E \rangle_T$ which is $T$-dependent. Thus the error of free energy becomes large when $|T - T_0|$ is large. For both the entropic sampling and the Wang-Landau method, the ensemble is created in the energy space instead of in the configuration space. This helps to obtain the density of states efficiently, but is less convenient if one hopes to study other physical quantities in a single simulation, especially when these quantities are not simple functions of energy $E$, such as the correlation function.

For quantum systems, the calculation of free energy is also a pertinent but harder problem. In this regard, the Wang-Landau algorithm has been combined with statistical series expansion method to calculate the free energy of quantum Hamiltonians such as the Heisenberg model. The idea of flat histogram has been applied to the diagrammatic Monte Carlo method to improve the long imaginary-time results, and to calculate the grand potential of a cluster system with electron bath. However, considering that the flat-histogram method or Wang-Landau algorithm have not been implemented under the path integral(PI) quantum Monte Carlo (QMC) methods such as the determinantal QMC and the continuous-time PI QMC method, which are based on the Metropolis sampling in configuration space, it is still desirable to develop a free energy calculation method which can calculate free energy by the configuration space sampling.

It is the purpose of this paper to propose such a new Monte Carlo algorithm that can calculate the free energy using the configuration-based sampling algorithm. The price that we have to pay is a sequential scan from low to high temperatures. We demonstrate the implementation of this algorithm using the two-dimensional Ising model whose exact free energy is known. Comparison with Wang-Landau method shows that both efficiency and accuracy of this method are satisfactory. The additional advantage of this method is that it is based on Metropolis algorithm and hence in principle it can be extended to quantum systems within determinantal or path integral methods.

II. METHOD AND RESULTS

In this paper, we demonstrate the implementation of our method and analyse its features using the two dimensional Ising model. For comparison purposes, here we use the equivalent Hamiltonian of the two-state Potts model

$$H = -J \sum_{\langle i,j \rangle} \delta_{s_i, s_j}.$$  \hspace{1cm} (1)

Here, $s_i$ is the spin degrees of freedom on site $i$ and it takes integer values from 0 to 1. $\delta_{s_i, s_j} = 0$ if $s_i \neq s_j$ and $\delta_{s_i, s_j} = 1$ if $s_i = s_j$. The summation is for pairs of nearest neighbour sites on a square lattice with $N \times N$ geometry. This model has been studied extensively as a basic statistical model. Its exact critical transition temperature on two-dimensional square lattice is $k_B T_c = 2J/\ln(1 + \sqrt{2})$. In the following we use $J = 1$ as the
energy unit and set the Boltzmann constant $k_B = 1$ for convenience.

One of the widely used Monte Carlo algorithms for studying classical statistical models is the Metropolis sampling in configuration space. In this algorithm, one starts by choosing a random spin configuration $S_0$ (here we use capital letter $S$ to denote the spin configuration of the whole lattice), and update the configuration $S_n \rightarrow S_{n+1}$ according to a given proposal probability $P(S_n \rightarrow S_{n+1})$ and an accepting probability $A(S_n \rightarrow S_{n+1})$. The transition probability $T(S_n \rightarrow S_{n+1}) = P(S_n \rightarrow S_{n+1})A(S_n \rightarrow S_{n+1})$ must satisfy the detailed balance condition $f(S_n)T(S_n \rightarrow S_{n+1}) = f(S_{n+1})T(S_{n+1} \rightarrow S_n)$ to guarantee that the resulting Markovian chain has the target distribution $f(S_n)$ in the equilibrium limit. For the thermodynamical calculation, we use the Boltzmann distribution as our target distribution,

$$f(S_n) = \frac{1}{Z} e^{-\beta H(S_n)}, \quad (2)$$

where $Z$ is the partition function $Z = \sum_S e^{-\beta H(S)}$. Here $\beta = 1/T$ is the inverse temperature. After the Markovian chain reaches equilibrium, sampling on this chain can produce the required statistical averages. However, the free energy $F$ cannot be calculated directly, because the partition function $Z$ is a normalization factor of the probability distribution $f(S)$ and hence cannot be treated as a statistical average. To calculate $F$, usually one either employs the concept of energy histogram within Monte Carlo method, the maximum entropy method, or by numerical integration over the derivative of free energies. In the following, we propose a new method which combines the idea of energy histogram and the configuration-space sampling to calculate the free energy over full temperature range.

In the Metropolis algorithm, the energy probability distribution produced by the Markovian chain is

$$p(E) = \frac{g(E)}{Z} e^{-\beta E}, \quad (3)$$

where $g(E)$ is the degeneracy of energy level $E$ of the given Hamiltonian. $p(E)$ can be estimated approximately from the energy histogram of the Markovian chain, that is,

$$p(E) \approx \frac{N(E)}{m}. \quad (4)$$

Here $N(E)$ is the number of spin configurations with energy $E$ and $m$ is the total number of sampled configurations in the Markovian chain. The precision of the above estimation increases as $m$ increases. In the limit $m \rightarrow \infty$, we get

$$\frac{g(E)}{Z} e^{-\beta E} = \frac{N(E)}{m}, \quad (5)$$

and hence

$$F(T) = -\beta \ln Z = -\beta \ln \left[ \frac{g(E) e^{-\beta E}}{m N(E)} \right]. \quad (6)$$

In principle, $F(T)$ can be calculated from $g(E)$ and the histogram $N(E)$ at any $E$. For a large variety of classical Hamiltonians, the ground state degeneracy $g(E_g)$ is easy to obtain. For the Ising model, for example, $g(E_g) = 2$. As a result, $F(T)$ is in principle obtainable from $N(E_g)$.

In practice, however, the above simple scheme does not work at arbitrary $T$ because the energy distribution $p(E)$ is sharply peaked at $E(T)$ (or at many different $E$’s for some models) which is an increasing function of $T$. In Fig.1, we show $p(E) \approx N(E)/m$ for the Ising model on the square lattice with $N = 32$, at different temperatures. Here $E$ is the energy per site and $E_g = -2$. $p(E)$ is plotted in logarithmic scale and it decays very fast away from the peak position. At high temperatures, $p(E)$ is peaked at high energy and $p(E_g)$ is so small that it is impossible to calculate it accurately from $N(E_g)/m$, because the latter is practically zero for a finite $m$.

Therefore, at finite $T$, $p(E)$ can be evaluated accurately by $N(E)/m$ only at $E \approx E_{peak}$, where $p(E)$ reaches its peak. To employ Eq.(6) for $F(T)$, one also needs the density of state $g(E)$ at the same energy $E$. A strategy to obtained $g(E)$ for $E > E_g$ is to transfer the knowledge of $g(E_g)$ to higher energies using the following equation,

$$Z = g(E) e^{-\beta E} \frac{m}{N(E)} = g(E') e^{-\beta E'} \frac{m}{N(E')} \quad (7)$$

FIG. 1: The normalized energy distribution probability in the Markovian chain for $N = 32$ square lattice Ising model, at (a) low temperatures $T < 0.6$, and at (b) high temperatures $T > 0.6$. 

Given the value of \( g(E) \) at energy \( E \), one obtains \( g(E') \) from \( g(E') = e^{-\beta(E'-E')/N(E')/m}g(E) \). In the actual calculation, Eq.(7) works only when both \( N(E)/m \) and \( N(E')/m \) have reasonably large magnitude. At temperature \( T \), this is the case for \( E \) and \( E' \) within the energy window \( W_T \) around the peak position of \( p(E) \). Eq.(7) then guarantees that knowing the value of \( g(E) \) at any \( E \) in \( W_T \), one can calculate \( g(E) \) at all \( E' \)’s in the same energy window. \( F(T) \) can then be calculated by Eq.(6) from the data \( N(E)/m \) and \( g(E) \) at \( E \in W_T \).

Since we only know the ground state degeneracy \( g(E_0) \), to obtain \( g(E) \) at higher energies \( E \in W_T \) for a finite temperature, we need to increase \( T \) from zero in small steps such that for adjacent temperatures \( T_i \) and \( T_{i+1} \), the corresponding energy windows \( W_i \) and \( W_{i+1} \), in which \( p(E) \) is large, have significant overlap. For temperature \( T_i \), suppose one knows \( g(E) \) value for an energy \( E \in W_i \), one calculates the histogram \( N(E')/m \) for a properly chosen \( E' \in (W_i \cap W_{i+1}) \). Using Eq.(7), \( g(E') \) can be calculated from the data of \( g(E) \). One can then calculate \( F(T) \) from \( g(E') \) using Eq.(6). For the next temperature \( T_{i+1} \), \( g(E') \) will be used to produce \( g(E'') \) for an energy \( E'' \in (W_{i+1} \cap W_{i+2}) \). This process goes on until the desired high temperature is reached. In this way, \( g(E) \) for larger and larger energies and \( F(T) \) at successively higher temperatures are obtained. Here, the practice is similar in spirit to the histogram reweighting method but used iteratively at successively higher temperatures.

In the implementation of the above algorithm, an important technical issue is how to select the common energy point \( E \in (W_i \cap W_{i+1}) \). In our calculation, we use the crossing energy \( E_c \), which is determined by \( p(E_c, T_i) = p(E_c, T_{i+1}) \). It is the energy where the \( p(E) \) curves of adjacent temperatures \( T_i \) and \( T_{i+1} \) cross each other. The \( E_c \) value chosen in this way has the largest value \( N(E)/m \) for both \( T_i \) and \( T_{i+1} \), hence guarantees the optimal precision. To avoid too small histogram \( N(E_c)/m \), which has a large error, we use equal distance temperature points \( T_i = T_0 + i\delta T \) and carefully controls the interval \( \delta T \). If \( \delta T \) is too large, \( p(E_c) \) will be too small and the error in \( N(E_c)/m \) will be large. If \( \delta T \) is too small, the number of temperature points will increase, going from the initial \( T = 0 \) to the final \( T \). This will lead to an increase of the accumulated error in \( g(E) \) transfer. Therefore, a suitable \( \delta T \) should be found by testing calculations. This issue is discussed below (Fig.5).

In our benchmark calculation, we use the cluster update scheme of Wolff. It has a relatively high updating speed and weak critical slowing down near the critical temperature. The whole free energy calculating algorithm described above can be used with general Metropolis local update algorithm without modification. In Fig.1, we show the normalized energy probability distribution obtained from the Markovian chain, for \( T < 0.6 \) (in Fig.1(a)) and for relatively high temperatures \( T \geq 0.6 \) (in Fig.1(b)). Here \( E \) is the energy per site. At low \( T \), \( p(T) \) has a peak at \( E_g = -2 \) and its width broadens with increasing \( T \). While for \( T \geq 0.6 \), the peak position begins to increase with temperature and its width gets saturated to about 0.2. In the high temperature limit, the peak position tends to \( E = 0 \). Significant overlap in the peak energy windows for adjacent temperatures \( T_i \) and \( T_{i+1} \) is crucial for our algorithm to work. In actual calculations, we use a uniform temperature mesh and choose \( \delta T = 0.02 \sim 0.15 \) such that the overlap of the peaks are large enough to guarantee the precision.

In Fig.2, we compare the free energy per site for \( N = 64 \) obtained by the present Monte Carlo algorithm with the exact result. Although we use only a modest \( m = 10^3 \) samples, our results are indistinguishable from the exact curve on the scale of Fig.1 for

![FIG. 2: The free energy per site of Ising model as function of \( T \), calculated from present method (black dots) and from the exact solution (red line) for \( N = 64 \) square lattice, using \( m = 10^3 \). The dots and the line are indistinguishable in the present scale. Inset: standard variance \( \sigma \) and the numerical error \( \epsilon \) of \( F/N^2 \). \( \sigma \) is estimated from 500 independent Markov chains.](image)

![FIG. 3: The free energy per site of Ising model as function of \( T \), calculated from present method (black dots) and from the exact solution (red line) for \( N = 128 \) square lattice, using \( m = 10^3 \). The dots and the line are indistinguishable in the present scale. Inset: standard variance \( \sigma \) and the numerical error \( \epsilon \) of \( F/N^2 \). \( \sigma \) is estimated from 500 independent Markov chains.](image)
independent Markov chains for \( \sigma \) lines are power law fit of \( \sigma = 0.004 \epsilon^0 \). FIG. 5: The standard variance \( \sigma \) as functions of linear size of the lattice \( N \). Inset: the peak width of \( p(E) \) curve as a function of \( N \). Other parameters are \( k_B T/J = 6.0 \) and \( m = 10^4 \). \( \sigma \) is estimated from 500 independent Markov chains.

0 < \( k_B T/J < 6.0 \). In the inset, the standard variance \( \sigma = \sqrt{(\langle F^2 \rangle - \langle F \rangle^2)/N^2} \) and the actual numerical error \( \epsilon = |\langle F \rangle - F_{exc}|/N^2 \) are shown as functions of \( T \). Both of them are measured from 500 independent runs. It is seen that \( \sigma \) and \( \epsilon \) are in same order of magnitude, with \( \epsilon \) less than \( 3\sigma \). Both are small in \( T < 1.15 J \) and increase linearly with temperature in \( T > 1.15 J \). This value corresponds to the finite temperature order-disorder transition at \( T_c \approx 1.181 J \) in the thermodynamic limit. The linear increase of error with \( T \) in the disordered phase is due to error accumulation in going from small \( T \) to larger ones, since the calculation at \( T_{i+1} \) uses the \( g(E) \) data of lower \( T_i \). For the highest temperature that we study \( T = 6.0 \), the actual error is smaller than \( 4 \times 10^{-3} \). The relative error \( \epsilon/|F_{exc}| \) first increases with \( T \) and then saturates to \( 6 \times 10^{-4} \) in \( T \geq 0.6 \).

The same comparison is made for larger lattice size \( N = 128 \) in Fig.3. The results are similar to \( N = 64 \) and the agreement between Monte Carlo and the exact result is excellent. The main difference from the \( N = 64 \) case is that the standard variance \( \sigma \) is larger than the actual error \( \epsilon \). As temperature increases, an abrupt increase of \( \sigma \) and \( \epsilon \) occurs at \( T \approx 1.15 \) and the linear \( T \) behavior occurs at \( T \approx 1.15 \). These features are same as \( N = 64 \) case. For the relative error, a similar saturation in \( \epsilon/|F_{exc}| \) is observed in high temperature to about \( 8 \times 10^{-4} \).

III. DISCUSSIONS

In this section, we discuss several issues about this new Metropolis-based algorithm for the free energy calculation. First, we discuss the size dependence of the calculation precision. In Fig.4, the \( N \)-dependences of \( \sigma \) and \( \epsilon \) are shown. It is seen that for all the calculated size \( 16 \leq N \leq 128 \), \( \sigma \) and \( \epsilon \) are on the same order, all smaller than \( 8 \times 10^{-3} \). This shows that the efficiency of our algorithm does not deteriorate with increasing \( N \), at least for \( N \leq 128 \). The relative magnitude of \( \sigma \) and \( \epsilon \) may vary for different \( N \)'s, but the actual error \( \epsilon \) is always within \( 3\sigma \). In the inset of Fig.4, we show the peak width of \( p(E) \) as a function of \( N \) at \( T = 6.0 J \). It is observed that the peak width scales as \( 1/N \). This reminds us that for very large \( N \), the \( p(E) \) curves will be very sharp and we have to use a denser \( T \)-mesh. At \( N = 128 \), however, this effect is not dramatic on the results, and we can still use the same \( T \)-mesh while keep the numerical precision, as shown by the main figure.

In Fig.5, we show the standard variance \( \sigma \) as functions of sampling number \( m \) for various \( \delta T \) values. Our results on \( N = 32 \) lattice shows that \( \sigma \approx m^{-x} \) with the average \( x = 0.44 \), close to the expected 1/2 from the central limit theorem. For fixed \( m \), \( \sigma \) has a weak dependence on \( \delta T \) in the range that we studied. As shown in the inset of Fig.5, \( \sigma \) is approximately a parabolic function of \( \delta T \), with the minimum reached at a \( m \)-dependent \( \delta T \). For all \( m \) values we used, we find that the smallest error is reached at \( \delta T \) values around \( \delta = 0.06 \sim 0.1 \).

Making quantitative comparison with Wang and Landau's results, we find that when scaled to the same \( m \) values, our result of \( F(T) \) has an error about one magnitude larger than that in Ref.2. This is not surprising because our algorithm is not the optimal method to produce \( F(T) \), but is a constrained one by the requirement that it
works in configuration space. Unlike the Wang-Landau algorithm which generates random walks in energy space, our algorithm samples directly in the configuration space. As a result, this method can be combined readily into certain Monte Carlo simulation of quantum systems. In the PI QMC and the determinantal QMC methods, the partition function of a given quantum Hamiltonian is expressed by the summation over configurations of classical auxiliary fields. The proposed algorithm can then be used to calculate $F(T)$, using the same Markovian chain as used for evaluating general expectation values.

One example of the application of $F(T)$ appears in the study of the Mott metal-insulator transition in half-filled Hubbard model using the dynamical mean-field theory (DMFT), which is exact in infinite spatial dimensions. The transition from the Fermi liquid state in small $U$ regime into the Mott insulator in large $U$ regime was found to be a special second order phase transition at $T = 0$ and a first order one at $T > 0$. To determine the actual transition line at $T > 0$, one needs to compare the free energy of the two coexisting phases within the two spinodal lines. Within DMFT, this task is reduced to the evaluation of free energy of the effective Anderson impurity model. This proves to be a difficult problem for QMC methods such as the Hirsch-Fye algorithm. Recently, the grand potential of the cluster problem is calculated by Wang-Landau method combined with the continuous time QMC, and it is used to calculate the grand potential of lattice model within cluster dynamical mean-field theory. It is an interesting topic to apply our algorithm in various QMC methods to handle similar problems. Work in this direction is under progress.

IV. ACKNOWLEDGEMENTS

We are grateful to the helpful discussions with Prof. You-Jin Deng. This work is supported by the 973 Program of China under Grant No. 2012CB921704 and by National Natural Science Foundation of China under Grant No.11374362.

* Electronic address: ruc.bs.plu@hotmail.com
† Electronic address: nhtong@ruc.edu.cn

1 D. P. Landau and K. Binder, *A Guid to Monte Carlo Simulations in Statistical Physics*, 2nd ed., Cambridge University Press, 2008.
2 A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
3 A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
4 J. S. Wang and R. H. Swendsen, J. Stat. Phys. **106**, 245 (2001).
5 J. Lee, Phys. Rev. Lett. **71**, 211 (1993).
6 B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
7 F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001); Phys. Rev. B **64**, 056101 (2001).
8 T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013).
9 M. Troyer, S. Wessel, and F. Alet, Phys. Rev. Lett. **90**, 120201 (2003).
10 N. G. Diamantis and E. Manousakis, arXiv: 1306.6320.
11 G. Li et al., Phys. Rev. B **80**, 195118 (2009).
12 E. Gull et al., Rev. Mod. Phys. **83**, 349 (2011).
13 M. Suzuki, Prog. Theor. Phys. **65**, 1454 (1976).
14 J. E. Hirsch, Phys. Rev. B **31**, 4403 (1985).
15 B. B. Beard and U. J. Wiese, Phys. Rev. Lett. **77**, 5130 (1996).
16 F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).
17 N. Metropolis et al., J. Chem. Phys. **21**, 1087 (1953).
18 C. Huscroft, R. Gass, and M. Jarrell, Phys. Rev. B **61**, 9300 (2000).
19 See, for an example, N. H. Tong, S. Q. Shen, and F. C. Pu, Phys. Rev. B **64**, 235109 (2001).
20 A. E. Ferdinand and M. E. Fisher, Phys. Rev. **185**, 832 (1969).
21 U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
22 F. Gebhard, *The Mott Metal-Insulator Transition*, Springer-Verlag Berlin Heidelberg, 1997.
23 W. Metzner and D. Vollhardt, Phys. Rev. Lett **62**, 324 (1989).
24 A. Georges, G. Kotliar, W. Krauth, M. Rozenberg, Rev. Mod. Phys. **68**, 13, (1996).
25 J. E. Hirsch and R. M. Fye, Phys. Rev. Lett. **56**, 2521 (1986).
26 N. Bluemer, *Mott-Hubbard Metal-Insulator Transition and Optical Conductivity in High Dimensions*, Shaker Verlag, Aachen, 2003.