Nonconvergence of the Wang-Landau algorithms with multiple random walkers

R. E. Belardinelli[1,2] and V. D. Pereyra[2]
[1]*Instituto de Física Aplicada (INFAP)–CONICET, San Luis, Argentina*
[2]*Departamento de Física, Universidad Nacional de San Luis, CONICET, Chacabuco 917, 5700 San Luis, Argentina*
(Received 19 November 2015; revised manuscript received 12 April 2016; published 19 May 2016)

This paper discusses some convergence properties in the entropic sampling Monte Carlo methods with multiple random walkers, particularly in the Wang-Landau (WL) and $1/t$ algorithms. The classical algorithms are modified by the use of m-independent random walkers in the energy landscape to calculate the density of states (DOS). The Ising model is used to show the convergence properties in the calculation of the DOS, as well as the critical temperature, while the calculation of the number π by multiple dimensional integration is used in the continuum approximation. In each case, the error is obtained separately for each walker at a fixed time, t; then, the average over m walkers is performed. It is observed that the error goes as $1/\sqrt{m}$. However, if the number of walkers increases above a certain critical value $m > m_x$, the error reaches a constant value (i.e., it saturates). This occurs for both algorithms; however, it is shown that for a given system, the $1/t$ algorithm is more efficient and accurate than the similar version of the WL algorithm. It follows that it makes no sense to increase the number of walkers above a critical value m_x, since it does not reduce the error in the calculation. Therefore, the number of walkers does not guarantee convergence.

DOI: 10.1103/PhysRevE.93.053306

I. INTRODUCTION

The Wang-Landau (WL) algorithm is currently one of the most widely used variations of the Monte Carlo simulation method introduced in recent years [1–3]. It belongs to the broader class of flat-histogram Monte Carlo simulations, aimed at obtaining an estimate of the density of states (DOS) $g(E)$ of a system with high accuracy [$g(E)$ represents the number of possible states or configurations with energy E].

Recent studies have proposed improvements and sophisticated implementations of the WL algorithm [4–19]. The convergence properties of the original WL formulation have been an issue of controversy. In fact, several studies show that the saturation of the final error persists [i.e., the difference between the simulation estimates for $g(E)$ and the exact values] regardless of the simulation effort employed. This problem was first pointed out by Yang and de Pablo in Ref. [15]. Several authors [5,7,8,16,17] have also analyzed the WL convergence. In particular, Zhou and Bhatt [5] presented an argument for its convergence.

It is well known that the exponential decrease of the modification factor $F = \ln f$ (which is defined below) with the number of iterations is the reason for the saturation of the error in the original WL algorithm, so that in the final sampling stages, the error to estimate $g(E)$ is essentially constant. To overcome this limitation, a new version of the WL algorithm has been introduced in Refs. [20–23], in which the modification factor is scaled down as $1/t$ instead of exponentially. The $1/t$ algorithm has been applied successfully to several statistical systems [24–33]. Very recently, a new version of the WL algorithm, the stochastic approximation Monte Carlo [34,35], which uses the $1/t$ strategy, was applied successfully to semiflexible polymers chains.

The convergence of the $1/t$ algorithm has been discussed in previous work [8,21,23]. In fact, it has been demonstrated analytically [21] that the entropy $S(E,t) = \ln g(E,t)$ can be expressed as a series in which $F(t)$ is the kernel; in those algorithms where $F_k = F_{k-1}/l$ (with any value of $l > 1$), the resulting series converges to a finite value, and then the error reaches a constant value (saturates in time). On the contrary, in those algorithms where the modification factor depends on time as $F(t) = t^{-\gamma}$ with $\gamma \leqslant 1$ (the optimum choice is $\gamma = 1$), the series is divergent and the calculated density of states approaches asymptotically the exact values as $\approx t^{-\gamma/2}$.

Recently, the tomographic sampling method was modified with use of the $1/t$ scheme. The tomographic algorithm was originally implemented using, in effect, a modification factor $F = \ln f$ that does not change with time [13]. It was shown that convergence is improved by using $F \sim 1/t$ in this method as well [23]. In addition, the authors demonstrated that there is convergence in the case in which $0 < \gamma \leqslant 1$ by using an analytical argument applied to the simple two-state model.

Nevertheless, numerical studies show that the error in the $1/t$ algorithm decays as $1/\sqrt{t}$, and to our knowledge this has not been improved upon. Therefore, whenever the modification factor F decreases exponentially with the number of iterations, the algorithm does not converge, regardless of any modification of the WL algorithm. That is, the error in calculating the DOS approaches a constant value (i.e., it reaches saturation), as was proved analytically in Refs. [21,23].

A general comment on the use of WL algorithm: despite the problems of convergence, it is well known nowadays that the WL method works very well in getting a first approximation of the density of states, and it is then used as an ingredient of a controlled numerical scheme (any type of multihistogram method). However, instead of being a critique to the original WL algorithm, any new contribution that helps to understand the behavior of the method and solve the problem of convergence should be considered.

In Refs. [36–39], Landau and co-workers introduced a massive parallel WL sampling based on the replica-exchange framework for Monte Carlo simulations. They introduced m random walkers in an energy subwindow. They emphasized that the estimated density of states converges to the true one with an increasing number of iterations, and the simulation

 ©2016 American Physical Society

is terminated when the modification factor reaches a minimal value $f_{\rm final}$. They demonstrated the advantages and general applicability of the method for the simulation of complex systems. They also showed that this algorithm is extremely efficient and that its parallel implementation is straightforward. This practice reduces the error during the simulation with $1/\sqrt{m}$, where m is the number of independent walkers in the energy sub-window.

A similar strategy to reduce the error with the number of walkers was implemented in Refs. [40,41]. Parallel implementation of other multihistogram methods was introduced in Ref. [42].

Although the method has been implemented in a massive parallel sampling in systems with multiple windows, it is easy to apply to a system with a single window. Thus, even when the error saturates for a single random walk, the average of m random walkers seems to converge to the exact value, i.e., the error seems to depend on $1/\sqrt{m}$.

In this paper, the validity of this assumption is discussed. Toward that end, a simple implementation of the algorithms is performed to calculate the DOS and other observables such as the critical temperature in the Ising model and the number π by numerical integration. The Ising model and numerical integration are used as a laboratory test for different reasons, i.e., (i) the DOS is known, at least for small systems; (ii) the observable can be obtained with high precision; and (iii) if it does not converge for trivial systems such as that mentioned above (which presents a relatively well-behaved energy landscape), it seems unlikely that it will converge in a more complex system. The remainder of this paper is organized as follows: In Sec. I, the algorithms and different quantities are introduced, as well as the definition of the errors for the different models. In Sec. II, the algorithms and their implementations are discussed. The results and the conclusions are presented in Secs. III and IV, respectively.

II. ALGORITHMS AND THEIR IMPLEMENTATIONS

The density of states in energy, $g(E)$, measures the energy degeneracy of the admissible states of a system, from which the partition function Z can be calculated:

$$Z(T) = \sum_{\rho} e^{-E[\rho]/k_B T} = \sum_{E} g(E) e^{-E/k_B T}, \tag{1}$$

where ρ stands for a state or configuration in which the system can reside, k_B is the Boltzmann constant, and T is the temperature. The first sum runs over all possible states of the system, whereas the second sum runs over all possible total energies and it can only be calculated once $g(E)$ is known. While $g(E)$ is temperature-independent and only depends on the definition of the Hamiltonian, Eq. (1) allows for the calculation of the temperature-dependent Z via the corresponding Boltzmann factors. One also defines a dimensionless entropy $S(E) \equiv k_B \ln g(E)$. An important consequence is the possibility of calculating the thermodynamic quantities at any temperature with the sole knowledge of $g(E)$. For example, the average energy $\langle E \rangle$ is

$$\langle E(T) \rangle = \frac{1}{Z} \sum_{E} g(E) E e^{-E/k_B T}, \tag{2}$$

and the heat capacity C_V can be calculated as

$$C_V(T) = \frac{\langle E^2 \rangle - (\langle E \rangle)^2}{k_B T^2}. \tag{3}$$

These thermodynamic observables provide a measure to identify and locate phase transitions, and hence to understand critical phenomena.

The standard WL algorithm [1–3] estimates the DOS using a single random walker in an energy range $[E_{\rm min}, E_{\rm max}]$. During the simulation, trial moves are accepted with a probability $P = \min[1, g(E_{\rm old})/g(E_{\rm new})]$, where $E_{\rm old}(E_{\rm new})$ is the energy of the original (proposed) configuration. The estimation of $g(E)$ is continuously adjusted and improved using a modification factor f [i.e., $g(E) \to f \times g(E)$], which starts with $f_0 > 1$ and progressively approaches unity as the simulation proceeds. A histogram, $H(E)$, keeps track of the number of visits to each energy E during a given iteration. When $H(E)$ is sufficiently $flat$ [43], the next iteration begins with $H(E)$ reset to zero but keeping the estimate of $g(E)$ from the previous iteration, and f is reduced by some predefined rule (e.g., $f \to \sqrt{f}$). The simulation ends when f reaches a sufficiently small value $f_{\rm stop}$, at which point the accuracy of $g(E)$ is proportional to $\sqrt{f_{\rm stop}}$ for sufficiently flat $H(E)$. The $1/t$ algorithm works as the original WL algorithm, but as soon as $F = \ln(f) \leqslant 1/t$, $F \to 1/t$; thereafter, $F(t) = 1/t$ is updated at each event (here, t is the Monte Carlo time defined as $t = n/N$, where n is the number of attempted changes of state, or steps, and N is the energy range). In other words, for a characteristic time, the modification factor F goes from exponential to power decay. For more details, see Ref. [23].

To assess its applicability, feasibility, and performance, the $1/t$ and the WL m-random walkers are applied to the two-dimensional Ising model on square lattices, as well as the calculation of the number π by numerical integration.

The two-dimensional Ising model on a square lattice with linear size $L = 8$ and periodic boundary conditions is used for the calculation of the DOS and the critical temperature. The size of the system is similar to the size of a single window in the replica exchange Wang-Landau sampling [36,37,39]. Despite the size, this is sufficient for the purposes of this study. However, in order to show the effect of the size on the behavior of the error, the study is also applied to a window of $N = 300$, which belongs to a larger system size $L = 64$. Monte Carlo multidimensional integration using the WL and $1/t$ algorithms is also implemented to calculate the number π [18,22,44].

Before discussing the results, it is necessary to explain how to calculate the errors of the quantities to be measured. In this paper, the error in the calculation of the DOS as a function of the number of walkers m and time t, is defined as:

$$\varepsilon_S(t,m) = \frac{1}{N} \sum_{E} \left| \frac{\overline{S_E(t,m)} - S_E^{\rm exc}}{S_E^{\rm exc}} \right|, \tag{4}$$

where $S_E^{\rm exc}$ is the exact value of the DOS for the energy E. The average over the number of walkers m is indicated by the overbar, which is given by

$$\overline{S_E(t,m)} = \frac{1}{m} \sum_{i}^{m} S_{E,i}(t). \tag{5}$$

Similarly, one proceeds with the error in the calculation of the observable $X(t,m)$:

$$\varepsilon_X(t,m) = \left| \frac{\overline{X(t,m)} - X^{\text{exc}}}{X^{\text{exc}}} \right|, \tag{6}$$

where

$$\overline{X(t,m)} = \frac{1}{m} \sum_i^m X_i(t), \tag{7}$$

and the corresponding standard deviation is

$$\sigma_m = \sqrt{\overline{X(t,m)^2} - (\overline{X(t,m)})^2}. \tag{8}$$

From the above equations, the errors as a function of time for m walkers can obtained. To smooth the curves, which usually present some noise, we perform an average over p independent realizations. The average is indicated by angular brackets and is defined as

$$\langle \varepsilon(t,m) \rangle = \frac{1}{p} \sum_i^p \varepsilon_i(t,m). \tag{9}$$

Similarly, for the mean value,

$$\langle \overline{X(t,m)} \rangle = \frac{1}{p} \sum_i^p \overline{X_i(t,m)}, \tag{10}$$

and the standard deviation,

$$\sigma_p = \sqrt{\langle \overline{X(t,m)}^2 \rangle - \langle \overline{X(t,m)} \rangle^2}. \tag{11}$$

III. DISCUSSION

First, the convergence properties of the WL and the $1/t$ algorithms as a function of time t, for a single walker, are discussed. The error in the calculation of the DOS as a function of time, t, for a two-dimensional Ising model, with linear size L is obtained from the above equations. Note that, for $m = 1$, the error in the DOS [Eq. (4)] and in the observable [Eq. (6)] is in agreement with the definition given in Refs. [20–23], and $\sigma_{m=1} = 0$. In this case, the range of the energy is $N = L^2 - 1$. The exact density of states, as well as the exact critical temperature for the Ising model, with a given system size is obtained by using the methodology developed by Beale in Ref. [45]. Similarly, one can obtain the error in the calculation of the number π, as discussed in Refs. [18,22,44].

In Fig. 1, different errors are shown as a function of time, using the WL and $1/t$ algorithms. As described in Ref. [23], the $1/t$ algorithm presents two temporal regimes: (i) the first stage of the simulation, where the $1/t$ algorithm coincides with the WL algorithm, and the error decreases sharply; and (ii) the second stage, when $F_i < 1/t$. Time $t = t_x$ separates the two regimes and coincides with the saturation time of the WL algorithm (where the error becomes constant). This time is shown in Fig. 1 with a vertical solid line. The saturation time for a two-dimensional Ising model with $L = 8$ is $t_x \approx 140\,000$ MCS, using an 80%-flatness criterion. As expected, for Figs. 1(a) and 1(b), t_x coincides, while the corresponding saturation time for the calculation of π with the WL algorithm is $t_x = 74\,000$ MCS. On the other hand, the slope of the curves corresponding to the $1/t$ algorithm goes as $1/\sqrt{t}$. The times t_1, t_2, t_3, t_4, and t_5 are characteristic times (indicated with vertical dashed lines) that will be used later.

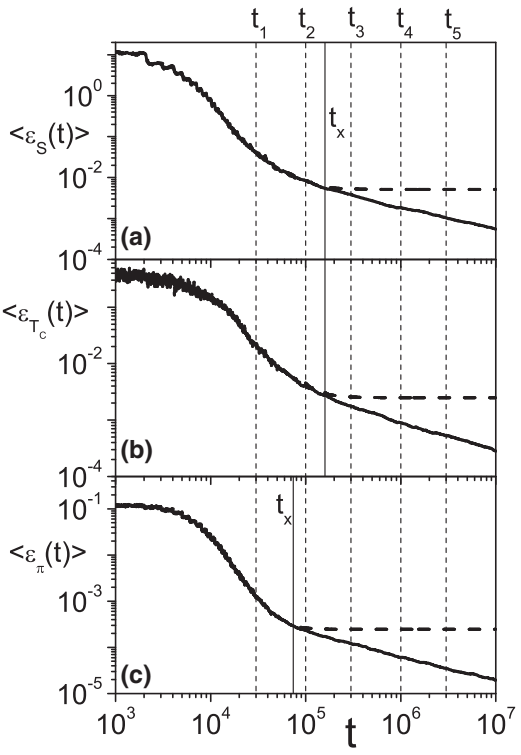

FIG. 1. Behavior of the error as a function of time using the WL algorithm (long dashed line) and the $1/t$ algorithm (solid line) for a single walker. The data correspond to (a) the DOS and (b) the critical temperature, obtained from the peak location of the specific heat for a two-dimensional Ising model; and (c) the calculation of the number π using multidimensional integrations. The critical time t_x, corresponding to the saturation of the error using the WL algorithm, is shown in the figures (vertical solid line). The times t_1, t_2, t_3, t_4, and t_5 correspond to the times at which the algorithm is stopped to start the m walkers; these are indicated with vertical dashed lines. The slope of the curves corresponding to the $1/t$ algorithm goes as $1/\sqrt{t}$. The data represent the average of 200 independent realizations ($p = 200$).

Next, let us discuss the effect of the flatness criteria in the measurement of the error. Figure 2 shows the behavior of the error as a function of the modification factor $\ln f$ for increasing values of the flatness criteria for the WL algorithm. From this figure, it is clear that no matter how flat the histogram is, the error always reaches a constant values, i.e., it saturates, even for a very high value of the flatness criterion (99.9%).

Figure 3 shows the behavior of the modification factor $F(t)$ as a function of time for two flatness criteria [Fig. 3(a)] and the corresponding error in the calculation of the critical temperature [Fig. 3(b)]. Note that for $t < t_x$, the error curve corresponding to WL 80% is below that corresponding to WL 90%. After that, for $t \geqslant t_x$, the behavior is reversed, that is, the curve corresponding to WL 80% is above that corresponding to WL 90%. This can lead to an erroneous evaluation of the accuracy and precision. In fact, if the error is calculated at $t < t_x$, it is found that the error of WL 90% is greater than that of WL 80%; however, if $t > t_x$, the behavior is the opposite.

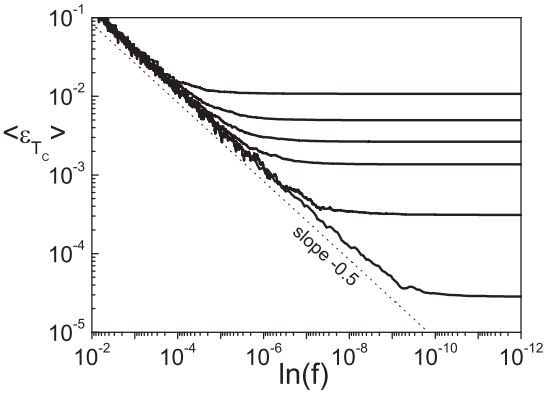

FIG. 2. The error as a function of the flatness criteria calculated for the WL algorithm. The curves are shown in decreasing order with 50%, 80%, 90%, 95%, 99%, and 99.9%, respectively. The data correspond to the critical temperature, obtained from the peak location of the specific heat for a two-dimensional Ising model with $L = 8$, and it is the average of 200 independent realizations ($p = 200$).

Next, we discuss the range of validity of the conjecture of Zhou and Batt [5], which assumes that the error is proportional to $\sqrt{\ln f_k}$, i.e., for a fixed value of f_k, the error will be the same for any flatness criteria.

To visualize this, the value of f_k is fixed in Fig. 3(a) (for example, $k = 13$, which corresponds to $\ln f_k = 1.2208 \times 10^{-4}$, horizontal dotted line) in such a way that the intersection

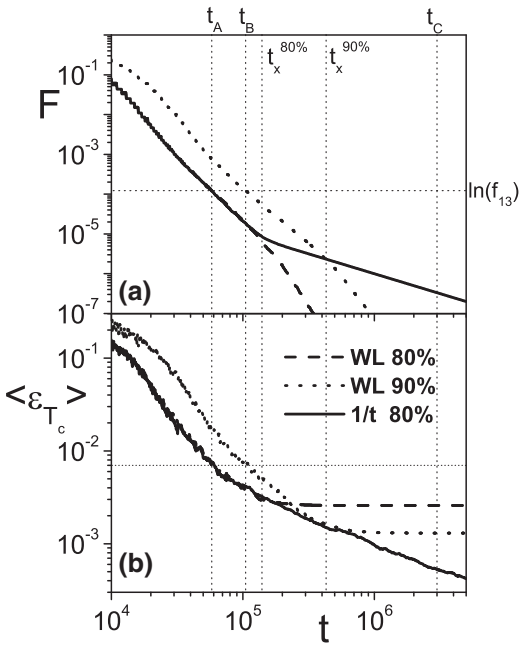

FIG. 3. (a) Behavior of $F(t) = \ln f$, and (b) the error in the calculation of the critical temperature for a two-dimensional Ising model $\epsilon_{T_c}(t)$, calculated by using the WL algorithm for 80% and 90% flatness criteria, and $1/t$ algorithms. Vertical solid lines represent the saturation times for the WL algorithm ($t_x^{80} \approx 140\,000$ MCS and $t_x^{90} \approx 430\,000$ MCS). The times $t_A \approx 58\,000$ MCS, $t_B \approx 105\,000$ MCS, and $t_C \approx 3 \times 10^6$ MCS as $F(t) = \ln f_{13} = 1.2208 \times 10^{-4}$ are described in the text. The data represent the average of 200 independent realizations ($p = 200$).

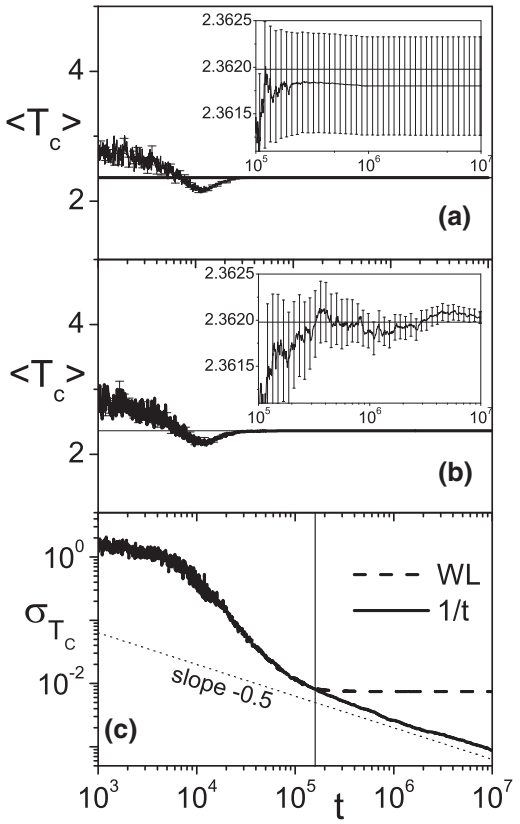

FIG. 4. Behavior of the mean value of the critical temperature as a function of time and the confidence interval for (a) the WL algorithm and (b) the $1/t$ algorithm (a magnification of the curve is shown in the inset). In (c), the behavior of the standard deviation for both algorithms is shown. The fitting of the $1/t$ curve gives a slope of 0.493(2). The data represent the average of 200 independent realizations ($p = 200$).

between the horizontal line and the $F(t)$ curves with 80% and 90% occurs at times t_A and t_B, respectively. These times are less than t_x. The errors corresponding to these times, in Fig. 3(b), are the same, confirming the conjecture. However, for $t > t_x$ [for example, t_C in Fig. 3(a)], the conjecture cannot be applied because of the saturation of the error in the WL algorithm. Then, one can say that, for the WL algorithm, the error is proportional to $\sqrt{\ln f_k}$, provided that the time $t < t_x$; in other words, the conjecture of Zhou and Bhatt is valid for $\ln f_k \geqslant 1/t$. In contrast, for the $1/t$ algorithm, the Zhou and Batt conjecture is valid for all time.

From the above, the determination of t_x is of fundamental importance, and it can be obtained not by using the WL algorithm but by using the $1/t$ algorithm instead.

To compare the statistical and systematic errors for a single walker, one proceeds to calculate the mean value of the observable over p independent realizations using Eq. (10) and the corresponding standard deviation σ_p using Eq. (11) (remember that $m = 1$).

Figure 4 shows the behavior of the mean value of the critical temperature as a function of time and the confidence interval ($\langle T_c \rangle \pm \frac{\sigma_p}{\sqrt{p}}$) for both WL [Fig. 4(a)] and $1/t$ [Fig. 4(b)] algorithms. In the inset of Figs. 4(a) and 4(b), a magnification

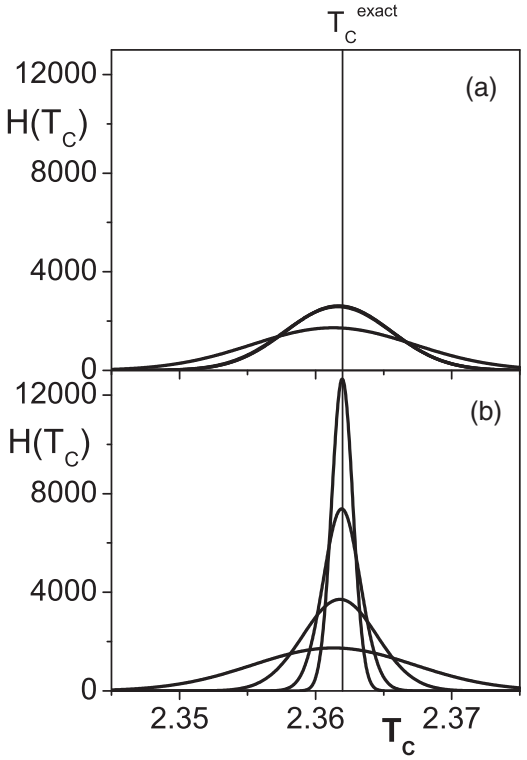

FIG. 5. Best-fit Gaussians for the histograms of the critical temperatures obtained using (a) the WL algorithm with the 80%-flatness criterion, and (b) $1/t$ algorithms. Each of the curves corresponds to $p = 100\,000$ independent runs. The curves are ordered from bottom to top according to the times defined in Fig. 1. Note that for the WL case, the curves collapse into each other for $t > t_x$.

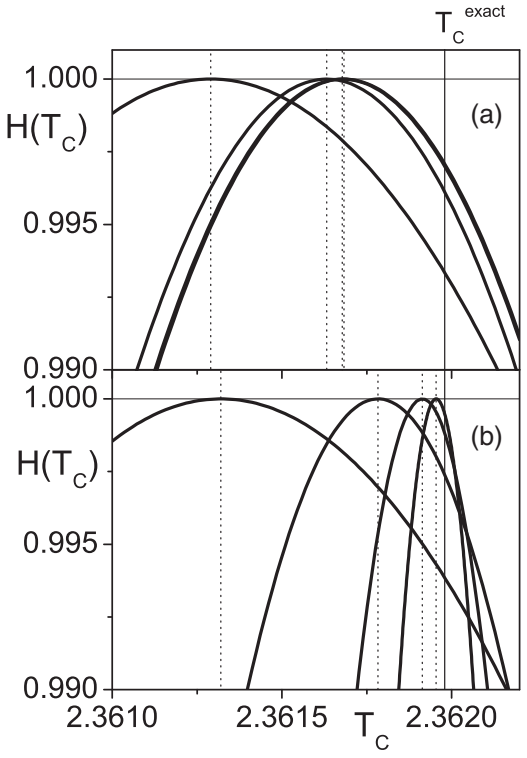

FIG. 6. Magnification of the curves described in Fig. 5. All the curves are normalized to 1 for comparison purposes.

of the curves is shown. As observed, the standard deviation, which is a measurement of the statistical error, presents different behavior according to the algorithm used. In fact, σ for the WL algorithm decreases for $t < t_x$; after that, for $t > t_x$ it reaches a constant value. On the contrary, σ for the $1/t$ algorithm always decreases, and for $t > t_x$ it decays as $1/\sqrt{t}$, as shown in Fig. 4(c). The statistical error decreases with time for the $1/t$ algorithm, while it remains constant for the WL algorithm. It is important to note that for the WL algorithm, both the error measured by Eqs. (4) and (9) [Fig. 1(b)] as well as the standard deviation, Eq. (11) [Fig. 4(c)], reach a constant value for $t > t_x$. The unusual behavior of the standard deviation as a function of time has been discussed in Ref. [23].

To confirm this effect, Fig. 5 shows the best-fit Gaussian for the histograms of the critical temperature obtained at times t_2, t_3, t_4, and t_5, which are defined in Fig. 1, for the sampling using (a) the WL algorithm using the 80%-flatness criterion, and (b) the $1/t$ algorithm; each histogram is obtained for $p = 100\,000$ independent runs. The vertical line corresponds to the exact temperature obtained with data from Ref. [45]. To ensure a proper comparison, the scales are the same in both figures.

In Figs. 6(a) and 6(b), an enlargement of the curves is shown. To compare them, the curves are adequately normalized.

It is observed that for $t > t_x$ (t_3,t_4,t_5), the curves corresponding to the WL algorithm are superimposed [Figs. 5(a) and 6(a)], which is in agreement with the discussion above, i.e., the standard deviation is constant. On the contrary, for the $1/t$ algorithm, the standard deviation decreases with time [Figs. 5(b) and 6(b)]. Note that at $t = t_2$, the curves coincide within statistical error. This is due to the fact that this takes place in stage (a) of the $1/t$ algorithm, which coincides with the WL algorithm. In the rest of this section, the dependence of the errors as a function of the number of walkers m, for a fixed value of time t, is discussed.

To get the DOS and the observables using the so-called m-random-walkers algorithm, one proceeds as follows: (a) the running time is fixed to a certain value, $t = t'$, and then $S(E,t = t') = \ln g(E,t = t')$ is obtained for all values of E; (b) the algorithm is executed by m independent random walkers; and (c) the quantities of interest are averaged adequately.

The error in the calculation of the DOS as a function of the number of walkers, using the WL and $1/t$ algorithms, is shown in Figs. 7(a) and 7(b), respectively, where t' takes the following values: $t_1 = 3 \times 10^4$ MCS, $t_2 = 1 \times 10^5$ MCS, $t_x = 1.4 \times 10^5$ MCS, $t_3 = 3 \times 10^5$ MCS, $t_4 = 1 \times 10^6$ MCS, and $t_5 = 3 \times 10^6$ MCS (indicated by vertical dotted lines in Fig. 1).

As observed, the error decreases with the number of walkers as $1/\sqrt{m}$, and for a certain value of m it loses this functionality, approaching a constant value, i.e., the error is saturated with the number of walkers. However, this behavior presents different characteristics depending on the algorithms used.

For $t < t_x$, the $1/t$ algorithm is still in the WL regime (see Fig. 1). Therefore, it should be expected that the errors are statistically the same. In fact, this is confirmed in Figs. 7(a) and 7(b), where the error curves corresponding to

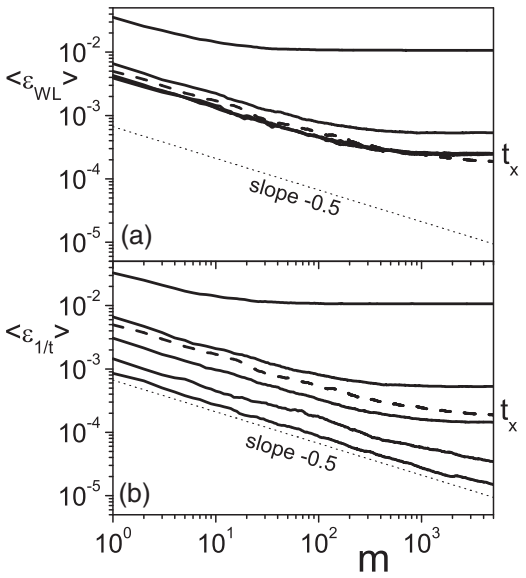

FIG. 7. Behavior of the error as a function of the number of walkers for the calculation of the DOS, using the WL algorithm (a) and the $1/t$ algorithm (b). In both figures, the curves are shown in decreasing order according to the times t_1, t_2, t_x, t_3, t_4, and t_5, which are defined in Fig. 1. The dotted line with slope $1/\sqrt{m}$ is shown for comparison. The data represent the average of 100 independent realizations ($p = 100$).

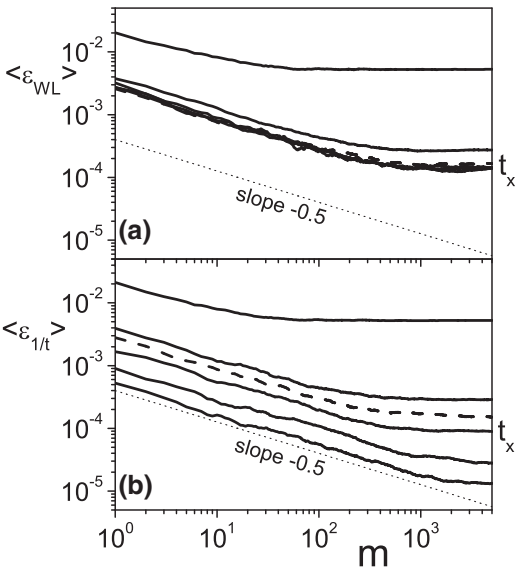

FIG. 8. Behavior of the error as a function of the number of walkers for the calculation of the critical temperature using the WL algorithm (a) and the $1/t$ algorithm (b). The parameters are the same as in Fig. 7. The data represent the average of 100 independent realizations ($p = 100$).

$t_1 = 3 \times 10^4$ MCS and $t_2 = 1 \times 10^5$ MCS (the first two curves from top to bottom) are the same. The error decreases with the number of walkers as $1/\sqrt{m}$ for $m \leqslant 100$ for the top curve, and for $m \leqslant 400$ for the next curve; then it loses this functionality, approaching a constant value. In other words, there is a critical number m_x that separates these two regimes, such that, for $m < m_x$, the error goes as $1/\sqrt{m}$, and for $m \geqslant m_x$, the error reaches a constant value (saturation value).

Although this behavior is observed in all cases, the error calculated by the WL algorithm for $t \geqslant t_x$ has a peculiar characteristic, i.e., all the error curves collapse into a single curve [see Fig. 7(a)]. That means that, for $t \geqslant t_x$, m_x is the same for all curves, while in the $1/t$ algorithm, the error curves do not collapse [see Fig. 7(b)], and therefore m_x increases with time.

It is important to note that the WL algorithm does not determine the saturation time, t_x, which is critical when running the algorithm properly, since for longer times it becomes an unnecessary calculation. Therefore, the WL algorithm could be inefficient, since for $t \geqslant t_x$ the error curves collapse into one, regardless of the number of walkers used. For example, in this particular system (a two-dimensional Ising model with size $L = 8$, periodic boundary conditions, and using the 80%-flatness criterion), the saturation time is $t_x \approx 140\,000$ MCS and the critical number of walkers is $m_x \approx 400$. By simple inspection of the figure, it seems to be impossible to obtain an error below 10^{-4} in the calculation of the DOS using the WL algorithm, either by increasing the running time ($t > t_x$) or the number of walkers ($m > m_x$).

This behavior is also observed in the calculation of the critical temperature T_c with the number of walkers m [see Figs. 8(a) and 8(b)], and in the continuum approximation, i.e., the multidimensional numerical integration to calculate the number π [see Figs. 9(a) and 9(b)].

The effect of the size of the system on the behavior of error in the calculation of the DOS can be important for real systems. However, the characteristics described above are expected to be the same for small systems. That is, if there is a number of walkers m_x to which the error saturates for a small system, this

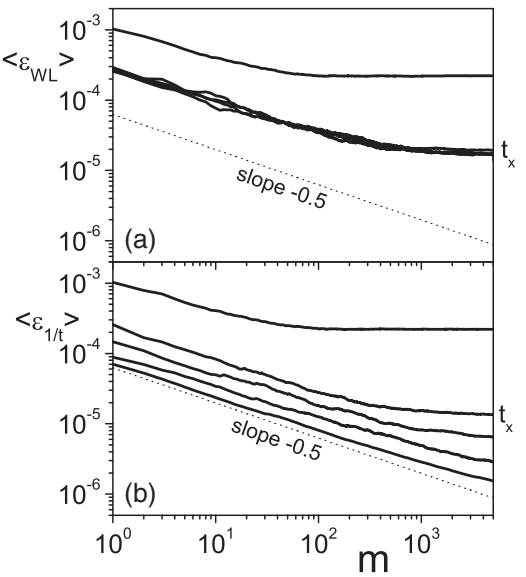

FIG. 9. Behavior of the error as a function of the number of the walkers for the calculation of number π, using the WL algorithm (Fig. 9a) and the $1/t$ algorithm (Fig. 9b). In both figures, the curves are shown in decreasing order according to the times t_1,t_2,t_x,t_3,t_4 and t_5, defined in Fig. 1. The parameters are the same as in Fig. 7. The data represent the average of 100 independent realizations ($p = 100$).

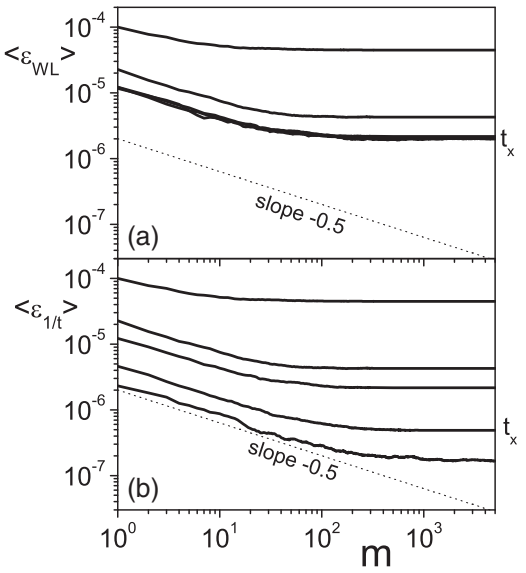

FIG. 10. Behavior of the error as a function of the number of walkers for the calculation of the DOS for a window of 300 energy sites, corresponding to a two-dimensional Ising model with size $L = 64$, using the WL algorithm (a) and the $1/t$ algorithm (b). The data represent the average of 100 independent realizations ($p = 100$).

must occur to a larger system. To confirm this, in Fig. 10 the behavior of the error in the calculation of the DOS is shown as a function of m for a window of $N = 300$ energy sites, corresponding to a two-dimensional Ising model with size $L = 64$. The energy range is between $(-0.29,0]$ (energy per sites). This is a usual size of window used in this particular case. As shown, the general behavior is the same as that described in the previous cases.

Figure 11(a) shows the behavior of the mean value of the critical temperature $\overline{T_c}$ as a function of m at a fixed time $t_5 = 3 \times 10^6$ MCS, using WL (open symbols) and $1/t$ (filled symbols) algorithms; the error bars ($\overline{T_c} \pm \frac{\sigma_m}{\sqrt{m}}$) are also shown. The data represent one realization ($p = 1$). The exact value of the critical temperature, T_c^{exc}, is indicated with a horizontal line.

The mean value calculated by the WL algorithm is farther away from the exact value of T_c than that calculated by the $1/t$ algorithm. The standard deviation remains constant in both cases. The value of σ for the WL algorithm is always greater than that corresponding to the $1/t$ algorithm; this is confirmed in Fig. 11(b). Looking at Fig. 11(a), one can observe that the systematic error, which is a measure of the distance between the average value of the calculated critical temperature and the exact one, is greater than the statistical error for increasing values of m for both algorithms.

With respect to understanding the reason for the saturation of the DOS to the number of walkers, in Fig. 12 we show the behavior of the normalized DOS as a function of E/L^2. Note that the DOS is normalized to the mean value. Different values of m are used (in the figure, the values of m increase from top to bottom at $E/L^2 = 0$) at fixed time $t' = 10\,000$. Note that the DOS approaches some limiting value $S(t) = S_{\text{lim}}(t')$ (thick black line), which differs from the exact one (dotted line). It is interesting to note that the growth of the DOS is skewed,

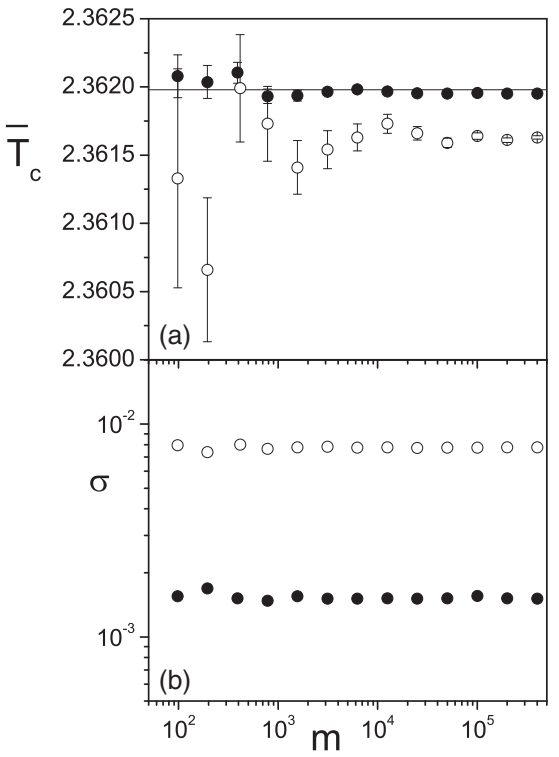

FIG. 11. (a) Behavior of the mean value of the critical temperature, $\overline{T_c}$, as a function of the number of walkers m for the WL algorithm and the $1/t$ algorithm, calculated at fixed time $t = t_5$; the confidence interval is also shown for both curves. (b) Behavior of the standard deviation as a function of m for both the WL and $1/t$ algorithms. Filled (empty) symbols represent $1/t$ (WL) procedures, respectively.

because it overestimates the most likely energy configurations (central part of the energy range) and underestimates the less likely energy configurations (left and right parts of the energy range). If one measures the DOS as a function of m, the value of $S(t',m)$ will approach $S_{\text{lim}}(t')$ and not the exact value, as $1/\sqrt{m}$ (thin black line in Fig. 12). For $m > m_x$, the mean

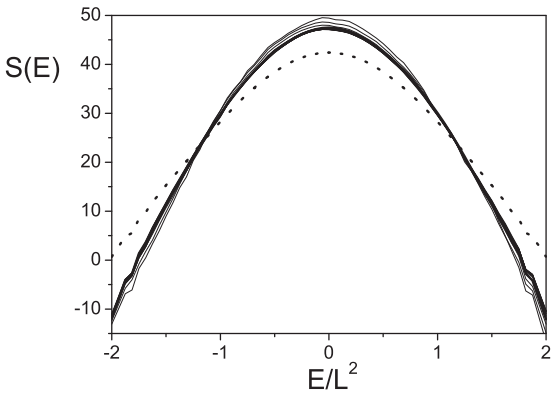

FIG. 12. Density of state (DOS) for a fixed value of $t' = 10\,000$ MCS, and different values of m ($m = 10,20,30,40,\ldots,200,2000$, from top to bottom in $E/L^2 = 0$). In this case, the critical value of the number of walkers is $m_x = 100$. The exact value of the DOS is denoted by a dotted line.

value will not be altered by new measurements because they will not differ from it. That is, the mean value is more accurate, but it differs from the exact one, i.e., the exactness does not change for $m > m_x$.

For $t' < t_x$, $S_{\rm lim}(t',m)$ changes with time for both algorithms. However, for $t' > t_x$ the behavior is very different. In fact, for the WL algorithm, $S_{\rm lim}(t',m)$ does not change in time (it is frozen). Then, regardless of the number of walkers used to calculate it, the mean value is also saturated. In other words, m_x does not change, and the error curves collapse into a single one [see Fig. 7(a)].

On the contrary, for the $1/t$ algorithm, for $t' > t_x$, $S_{\rm lim}(t',m)$ approaches asymptotically the exact value. The longer the time, the closer to the exact value. Therefore, m_x changes with the value of t'.

From the above, it is clear that by increasing the running time, the $1/t$ measurement can be improved, but not the WL measurement, since for $t > t_x$ any measurement will saturate. On the other hand, it makes no sense to increase the number of walkers above a critical value m_x, since it does not reduce the error in the calculation. Therefore, the number of walkers does not guarantee convergence.

Summarizing, for a single random walker one can write

$$\lim_{t\to\infty} \varepsilon_{\rm WL}(t) \to \text{const} \tag{12}$$

for the WL algorithm, and

$$\lim_{t\to\infty} \varepsilon_{1/t}(t) \to 0 \tag{13}$$

for the $1/t$ algorithm; while for m-random walkers,

$$\lim_{m\to\infty} \varepsilon_{\rm WL}(t,m)|_{t=t'} \to C_1, \tag{14}$$

and for the $1/t$ algorithm,

$$\lim_{m\to\infty} \varepsilon_{1/t}(t,m)|_{t=t'} \to C_2, \tag{15}$$

where C_1 and C_2 are constants that fulfill the following conditions: for $t < t_x$, $C_1 = C_2$, while for $t > t_x$, $C_1 > C_2$.

IV. CONCLUSIONS

Before presenting the conclusions about the convergence problem of the m-random walkers, it is convenient to revisit the convergence properties of a single random walker. It has been demonstrated analytically that the exponential decrease of the modification factor, $F = \ln f$, with the number of iterations is the reason for the saturation of the error in the WL algorithm with a single random walker.

One important conclusion that has not been mentioned before is that for the WL algorithm, the error is proportional to $\sqrt{\ln f_k}$, provided that the time $t < t_x$; in other words, the conjecture of Zhou and Bhatt is only valid for $\ln f_k \geqslant 1/t$. By contrast, for the $1/t$ algorithm, the conjecture is valid for all time.

An interesting feature of the single random walker is the comparison between the statistical and systematic errors. The first one is proportional to the standard deviation, while the second one is proportional to the distance between the mean value and the exact one. As discussed in the text, one can observe that the standard deviation for the WL algorithm for $t > t_x$ remains constant, while for the $1/t$ algorithm it decreases as $1/\sqrt{t}$. Thus, the statistical error for the WL algorithm is greater than that corresponding to the $1/t$ algorithm. On the other hand, the systematic error corresponding to the WL algorithm remains constant, while for the $1/t$ algorithm it decreases as $1/\sqrt{t}$ [see Fig. 1(b)].

There are convergence problems associated with the building of the density of states, and consequently the calculation of different observables, by using entropic sampling methods (WL and $1/t$ algorithms) with multiple random walkers. In fact, if the error is calculated using m experiments (walkers) at a fixed time, t, it decreases with the number of walkers as $1/\sqrt{m}$ until it reaches a certain value of walkers $m = m_x$ from which it saturates. This critical number m_x separates these regimes, such that, for $m < m_x$, the error goes as $1/\sqrt{m}$, and for $m \geqslant m_x$, the error reaches a constant value (saturation value). The critical value m_x depends on the characteristics of the system (size, interactions, connectivity, etc.).

The saturation of the error with the number of walkers is observed in the WL algorithm as well as in the $1/t$ algorithm. However, there are substantial differences for both algorithms.

As observed, the critical value m_x is lower in the WL algorithm than in the $1/t$ algorithm. Moreover, the saturation for the $1/t$ algorithm occurs at a very high number of walkers. On the other hand, the WL algorithm presents a peculiar behavior, that is, for $t \geqslant t_x$, all the error curves calculated by the m-walkers collapse into a single curve, while in the case of the $1/t$ algorithm the error curves do not; therefore, it makes no sense to run the WL algorithm for $t > t_x$.

It is shown that the statistical error is reduced with the number of walkers. However, the systematic error depends on the algorithm. It is also shown that the WL algorithm presents a systematic error that is not reduced with the number of walkers; the precision increases but not the exactness.

For a given system, the WL algorithm cannot calculate the DOS with greater accuracy than a certain value, even when the time is increased (t tends to infinity) or the number of walkers is increased (m tends to infinity). However, with the implementation of the $1/t$ algorithm, the error will always be reduced because there is no saturation in time.

Summarizing, one can claim that the $1/t$ algorithm is convergent. That is, the error in calculating the density of states versus time tends to zero as t tends to infinity. This was previously demonstrated numerically [20] and analytically [21,23]. In contrast, the WL algorithm is not convergent, i.e., the error saturates at a finite time.

The $1/t$ algorithm is always more efficient than the WL algorithm, even when it is as a function of the number of walkers. However, to calculate the DOS with high accuracy by using the $1/t$ algorithm, it is better to run the algorithm as a function of time and not as a function of the number of walkers.

In conclusion, it makes no sense to increase the number of parallel programs (number of walkers) above a critical value m_x, since it does not reduce the error in the calculation. The number of walkers does not guarantee convergence.

ACKNOWLEDGMENTS

The authors would like to thank A. P. Velasco for the useful discussion, and O. J. Furlong for reading the manuscript. This work is partially supported by the CONICET (Argentina).

[1] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[2] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).
[3] F. Wang and D. P. Landau, Comput. Phys. Commun. **147**, 674 (2002).
[4] P. Dayal, S. Trebst, S. Wessel, D. Wurtz, M. Troyer, S. Sabhapandit, and S. N. Coppersmith, Phys. Rev. Lett. **92**, 097201 (2004).
[5] C. Zhou and R. N. Bhatt, Phys. Rev. E **72**, 025701(R) (2005).
[6] Y. Okabe and H. Otsuka, J. Phys. A **39**, 9093 (2006).
[7] H. K. Lee, Y. Okabe, and D. P. Landau, Comput. Phys. Commun. **175**, 36 (2006).
[8] C. Zhou and J. Su, Phys. Rev. E **78**, 046705 (2008).
[9] C.-O. Hwang, Phys. Rev. E **80**, 042103 (2009).
[10] N. G. Fytas and A. Malakis, Phys. Rev. E **81**, 041109 (2010).
[11] A. Malakis, A. N. Berker, I. A. Hadjiagapiou, N. G. Fytas, and T. Papakonstantinou, Phys. Rev. E **81**, 041113 (2010).
[12] V. T. Ngo, D. T. Hoang, and H. T. Diep, Phys. Rev. E **82**, 041123 (2010).
[13] R. Dickman and A. G. Cunha-Netto, Phys. Rev. E **84**, 026701 (2011).
[14] G. Brown, Kh. Odbadrakh, D. M. Nicholson, and M. Eisenbach, Phys. Rev. E **84**, 065702 (2011).
[15] Q. Yan and J. J. de Pablo, Phys. Rev. Lett. **90**, 035701 (2003).
[16] D. J. Earl and M. W. Deem, J. Phys. Chem. B **109**, 6701 (2005).
[17] A. N. Morozov and S. H. Lin, Phys. Rev. E **76**, 026701 (2007).
[18] Y. W. Li, T. Wüst, D. P. Landau, and H. Q. Lin, Comput. Phys. Commun. **177**, 524 (2007).
[19] T. Wüst and D. P. Landau, J. Chem. Phys. **137**, 064903 (2012).
[20] R. E. Belardinelli and V. D. Pereyra, Phys. Rev. E **75**, 046701 (2007).
[21] R. E. Belardinelli and V. D. Pereyra, J. Chem. Phys. **127**, 184105 (2007).
[22] R. E. Belardinelli, S. Manzi, and V. D. Pereyra, Phys. Rev. E **78**, 067701 (2008).
[23] R. E. Belardinelli, V. D. Pereyra, R. Dickman, and B. J. Lourenço, J. Stat. Mech. (2014) P07007.
[24] N. G. Fytas, A. Malakis, and I. A. Hadjiagapiou, J. Stat. Mech. (2008) P11009.
[25] N. G. Fytas and A. Malakis, Physica A **388**, 4950 (2009).
[26] P. Ojeda, M. E. Garcia, A. Londoño, and N.-Y. Chen, Biophys. J. **96**, 1076 (2009).
[27] A. Malakis, A. N. Berker, I. A. Hadjiagapiou, and N. G. Fytas, Phys. Rev. E **79**, 011125 (2009).
[28] P. Ojeda and M. E. Garcia, Biophys. J. **99**, 595 (2010).
[29] N. G. Fytas and A. Malakis, Eur. Phys. J. B **79**, 13 (2011).
[30] A. D. Swetnam and M. P. Allen, J. Comp. Chem. **32**, 816 (2011).
[31] Y. Komura and Y. Okabe, Phys. Rev. E **85**, 010102(R) (2012).
[32] S. Inglis and R. G. Melko, Phys. Rev. E **87**, 013306 (2013).
[33] W. Janke and W. Paul, Soft Matter **12**, 642 (2016).
[34] B. Werlich, M. P. Taylor, and W. Paul, Phys. Proc. **57C**, 82 (2014).
[35] B. Werlich, T. Shakirov, M. P. Taylor, and W. Paul, Comput. Phys. Commun. **186**, 65 (2015).
[36] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013).
[37] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. E **90**, 023302 (2014).
[38] G. Shi, T. Vogel, T. Wüst, Y. W. Li, and D. P. Landau, Phys. Rev. E **90**, 033307 (2014).
[39] Y. W. Li, T Vogel, T. Wüst, and D. P. Landau, J. Phys.: Conf. Ser. **510**, 012012 (2014).
[40] A. A. Caparica and A. G. Cunha-Netto, Phys. Rev. E **85**, 046702 (2012).
[41] A. A. Caparica, Phys. Rev. E **89**, 043301 (2014).
[42] J. Zierenberg, M. Marenz, and W. Janke, Comput. Phys. Commun. **184**, 1155 (2013).
[43] In this paper, the flatness criteria means that histogram $H(j)$ for all possible j is not less than 80% of $\langle H\rangle = \frac{1}{L}\sum_{i=1}^{L} H(i)$ [2].
[44] W. Atisattapong and P. Maruphanton, Appl. Num. Math. **104**, 133 (2016).
[45] P. D. Beale, Phys. Rev. Lett. **76**, 78 (1996).