Dynamical traps in Wang-Landau sampling of continuous systems: Mechanism and solution

Yang Wei Koh,* Adelene Y. L. Sim,† and Hwee Kuan Lee‡
Bioinformatics Institute, 30 Biopolis Street, No. 07-01, Matrix, Singapore 138671
(Received 11 May 2015; published 26 August 2015)

We study the mechanism behind dynamical trappings experienced during Wang-Landau sampling of continuous systems reported by several authors. Trapping is caused by the random walker coming close to a local energy extremum, although the mechanism is different from that of the critical slowing-down encountered in conventional molecular dynamics or Monte Carlo simulations. When trapped, the random walker misses the entire or even several stages of Wang-Landau modification factor reduction, leading to inadequate sampling of the configuration space and a rough density of states, even though the modification factor has been reduced to very small values. Trapping is dependent on specific systems, the choice of energy bins, and the Monte Carlo step size, making it highly unpredictable. A general, simple, and effective solution is proposed where the configurations of multiple parallel Wang-Landau trajectories are interswapped to prevent trapping. We also explain why swapping frees the random walker from such traps. The efficacy of the proposed algorithm is demonstrated.

DOI: 10.1103/PhysRevE.92.023306 PACS number(s): 07.05.Tp, 02.70.Uu, 02.70.Tt, 87.15.ak

I. INTRODUCTION

Wang-Landau sampling (WLS) [1,2] is increasingly becoming an important tool for simulating a wide range of slow-relaxing, glassy systems which are not tractable by more conventional techniques such as molecular dynamics (MD) or Monte Carlo (MC) sampling using the Metropolis algorithm. To date, WLS has been applied to a wide variety of systems ranging from random bond and random field systems [2–4], atomic fluid [5] and clusters [6], liquid crystals [7], polymers [6,8,9], and proteins [10–12] to lipid membranes in explicit solvent [13,14]. In MD and MC sampling, one frequently encounters the problem that the system gets trapped in some local minimum due to the presence of large barriers on the energy landscape [11,15,16,29]. WLS (and related entropic sampling techniques [17–19]), on the other hand, circumvents this problem by sampling from the density of states (DOS) instead of the Boltzmann distribution. Since there are no energy barriers to overcome, the system readily moves back and forth between the high- and the low-energy regions of phase space, thereby ensuring efficient sampling. Due to the effectiveness of such multicanonical techniques, related methods such as the replica-exchange multicanonical algorithm [20–23] and statistical-temperature MC [24,25] are very efficient for simulating systems with rough energy landscapes.

Although WLS is initially believed to be free of the problem of slow relaxation, there is increasing evidence that WLS is subjected to another kind of slowing-down when applied to continuous systems. In an early study, Brown and Schultess [26] reported that for the ferromagnetic Heisenberg model, WLS is expensive when sampling from rare configurations close to the ground state. Jayasri *et al.* [7], in their study of liquid crystals, reported that WLS becomes extremely slow even for moderately large systems because the system often gets stuck in certain regions of configuration space. In their study of polypeptides and Lennard-Jones clusters using WLS, Poulain *et al.* [6] also reported similar trappings of their systems.

II. MODELS AND THE WANG-LANDAU METHOD

To examine the nature of these trappings, we consider two systems. The first is the two-dimensional $L \times L$ square lattice, frustrated XY model,

$$H = -\sum_{\langle i,j \rangle} J_{ij} \cos(\theta_i - \theta_j), \tag{1}$$

where $\theta_i \in (-\pi,\pi)$ $(i = 1,\ldots,L^2)$ are XY spins, and $\langle i,j \rangle$ denotes summation over nearest-neighbor pairs. The periodic boundary condition is used for both lattice dimensions. J_{ij} are identical independent random variables sampled from a Gaussian distribution with 0 mean and unit variance. Single-spin MC moves are used. The second system is an 8-mer poly-alanine molecule in a medium with dielectric constant $\epsilon = 2$. We used the Amber99 force field [27] and simulated using a modified version of the MOSAICS package [28]. Torsion angle moves were used. We choose these two systems as illustrations because they are simple enough yet possess rough energy landscapes with many local minima. WLS studies of the poly-alanine molecule have also been reported by Poulain *et al.* [6].

Simulations were performed using the standard WLS algorithm [1,2]. In WLS, the quantity of interest is the DOS, $g(E)$, or the number of all possible states at energy E. Once $g(E)$ is known, the partition function can be calculated,

$$Z = \sum_{\{\text{configurations}\}} e^{-E/k_B T} = \sum_{E} g(E) e^{-E/k_B T}, \tag{2}$$

where k_B is the Boltzmann constant, T is the temperature, and most thermodynamic quantities can be obtained from Z. WLS estimates $g(E)$ via a random walk in the energy space. Trials for a state with energy E_i to a state with energy E_f are accepted according to the transition probability,

$$p(E_i \to E_f) = \min\left(1, \frac{g(E_i)}{g(E_f)}\right). \tag{3}$$

*Corresponding author: patrickkyw@gmail.com
†adelenes@bii.a-star.edu.sg
‡leehk@bii.a-star.edu.sg

If $g(E)$ is the true DOS, Eq. (3) obeys detailed balance. However, $g(E)$ is initially unknown, and one makes a guess at it. Whenever an energy E is visited according to the transition rule, Eq. (3), the DOS is updated as $g(E) \leftarrow f \cdot g(E)$, where f is called the modification factor. The role of the modification factor is to gradually refine the initially inaccurate DOS. One begins the simulation with a large f (usually $f = f_0 = e^1 \approx 2.718\,28$) and gradually lowers it, in stages, to 1. When $f = 1$ is reached, detailed balance for Eq. (3) is recovered, and one obtains the true DOS for $g(E)$. In practice, one works with $\ln g(E)$ to prevent numerical overflow, and the update rule is $\ln g(E) \leftarrow \ln g(E) + \ln f$. In our simulations, we start from $\ln f_0 = 1$ and lower the modification factor according to $\ln f_{i+1} \leftarrow \frac{1}{2} \ln f_i$ upon completing the ith stage of simulation.

In the original formulation of WLS, one computes an energy histogram and uses its "flatness" as a criterion to proceed to the next stage of simulation. However, as this criterion has been found to be quite arbitrary by several subsequent studies [29–33], we use the criterion discussed in [33] to lower $\ln f$. Another modification we adopted here is the one suggested by Zhou *et al.* [34], that for continuous systems the DOS should be linearly interpolated if the energy falls between the centers of two energy bins. Hence in our analysis the DOS is piecewise linear. The case of a piecewise constant DOS is discussed at the end of this paper.

III. TRAPPINGS DURING WANG-LANDAU SAMPLING

Figures 1(a) and 1(b) show examples of the DOS where the system exhibits signs of being trapped for the XY model and poly-alanine molecule, respectively. The main feature is the high spikes, which are formed because the random walker stays for an inordinate amount of time in a single energy bin, resulting in most of the modification factor $\ln f$ being accumulated there. We have checked that the system is capable of proposing new configurations with fairly large changes in energy at the MC step size we have used [cf. Fig. 3(b) and discussion below]. Nevertheless, during the period when the random walker is stuck in a spiked bin, we found that all its MC moves result in small energy changes. The reason is not the choice of a small MC step size, but the fact that the random walker has wandered close to a stationary point in configuration space. To ascertain that the random walker was close to a stationary point, we computed its energy gradient and found it to be very small indeed.

Figure 2 is a schematic illustrating the mechanism of trapping and how a spike develops. Suppose, as shown in the inset, that the system is at a local minimum with configuration C and energy E. Let E lie within the bin b as shown. As the random walker is at a local minimum, the proposed energy E' must be $> E$, and the change in energy is small because the first derivative vanishes. Due to the steepness of the DOS, the acceptance ratio $\exp[\ln g(E) - \ln g(E')]$ is very small, the move is mostly rejected, and $\ln f$ is added to the bin b. This process repeats itself, with the DOS at bin b constantly increasing but the configuration hardly changing. If the system can propose a move to an energy, say E'' or higher, then the move may be accepted and the random walker can escape the trap. But this is highly improbable due to the vanishing of the gradient. The situation becomes more serious

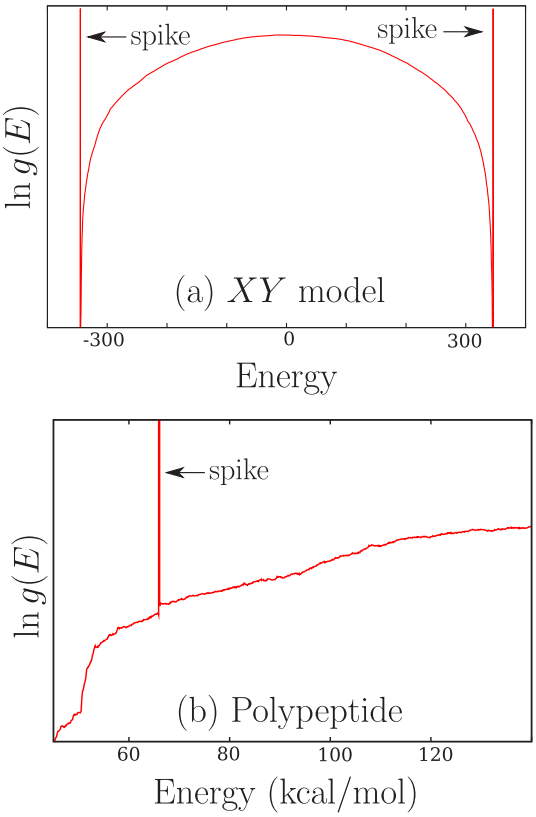

FIG. 1. (Color online) Spikes in the DOS for (a) the frustrated XY model ($L = 16$) and (b) the 8-mer poly-alanine. For both, we performed 10 stages of WLS, where each stage is simulated for 5×10^6 updates per spin/angle, and the modification factor $\ln f$ is halved at at each stage. Shown are the DOS values from a single trajectory at the end of the tenth stage.

as the system size increases because one usually increases the bin width in order to cover a wider energy range and reduce the computational cost; however, the change in energy

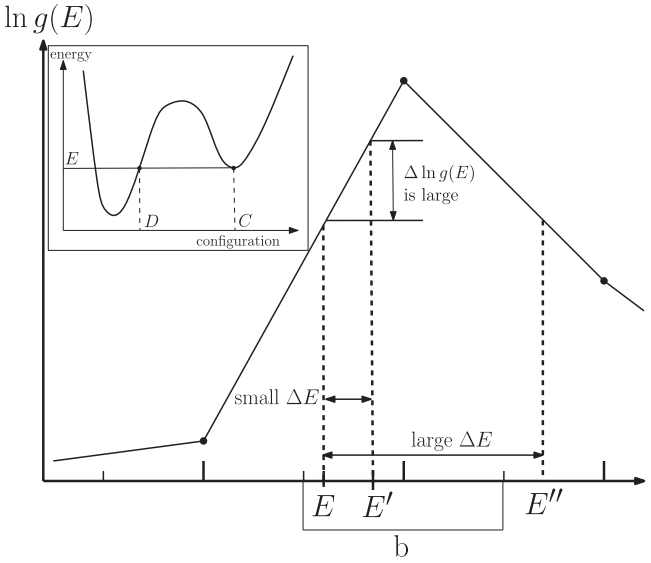

FIG. 2. Schematic illustrating how the random walker gets trapped inside a spiked bin.

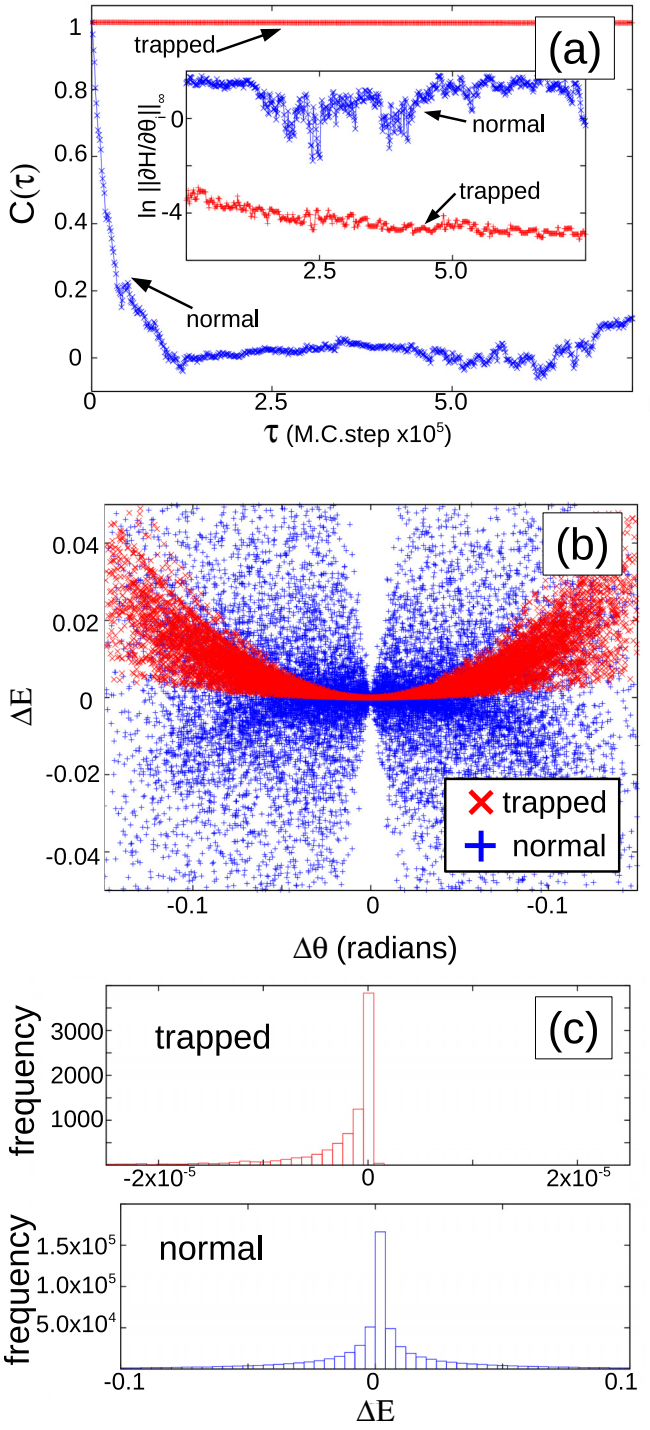

FIG. 3. (Color online) Comparison of behaviors between a trapped and normal segments of the trajectory for the XY model ($L = 16$). (a) Autocorrelation function and infinity norm of force along the segments. (b) Scatterplot of proposed energy change ΔE versus proposed angle change $\Delta\theta$. (c) Energy distribution of accepted moves.

produced by single-spin or a set of small torsion angle moves remains similar and, therefore, becomes smaller relative to the increasing bin width.

Figures 3(a)–3(c) illustrate, for the XY model ($L = 16$), the typical behavior of a trapped trajectory, compared to "normal" (i.e., not trapped) behavior. The trapped segment is taken from a continuous period of time when the random walker is inside a spiked bin ($\approx 8 \times 10^5$ steps in this case). The normal segment is of a similar length and taken from along the same trajectory but starting at a slightly earlier time, before the random walker gets trapped. To monitor the configuration of the system, we computed the autocorrelation function,

$$C(\tau) = \frac{1}{L^2} \sum_{i=1}^{L^2} \cos\left[\theta_i(0) - \theta_i(\tau)\right], \tag{4}$$

where τ denotes the number of MC steps that have elapsed since the start of the segment. The results are shown in Fig. 3(a). For the trapped segment, $C(\tau)$ remains close to unity throughout, meaning that its configuration hardly changes; for the normal segment, $C(\tau)$ decays to 0 very quickly. The inset in Fig. 3(a) plots the component of the force $\partial H/\partial\theta_i$ with the largest magnitude (i.e., infinity norm) versus τ. The force on the trapped segment is almost 0 throughout. Figure 3(b) is a scatterplot showing the proposed energy change ΔE versus the proposed angle change $\Delta\theta$. As mentioned above, with the MC step size we are using, the random walker is able to, under normal conditions, propose new configurations with fairly large changes in energy, as shown by the large dispersion of blue points (i.e., normal segment). However, when trapped, ΔE is very small compared to the normal segment and is distributed asymmetrically with very few energy-lowering moves. The acceptance rate is $\approx$1% and $\approx$60% for the trapped and normal segments, respectively. The energy distributions of the accepted moves are shown in Fig. 3(c). For the trapped segment, the histogram is very narrow and asymmetric, with almost only energy-lowering moves being accepted. We also examined other cases of trappings, for both the XY model (varying over a range of MC step sizes) and the alanine peptide molecule, and found the above behavior to be general in both systems.

Spikes are usually formed during the early stages of the simulation when the DOS is still rough. A small random bump on the DOS like the one in Fig. 2 can serve as a "seed" for spike growth if the random walker accidentally visits a local minimum (or maximum, where similar arguments apply) whose energy happens to lie within the same energy bin as the bump.

Spikes do not necessarily have to form in the same bin every time. This is because there are other local minima and maxima that reside in other energy bins. In addition, as shown in the inset in Fig. 2, the random walker can sample other configurations, such as D, with the same energy E. This means that spiking does not always occur in an energy bin where a spike was previously observed. Therefore spiking events are highly unpredictable.

A large amount of simulation time is wasted when the random walker is trapped because it spends less time sampling other parts of the configuration space. When a spike is discovered at the end of one's simulation one should discard the results knowing that it suffers from bad sampling, as shown later in our calculation of the specific heat capacity. A simple fix to the problem would be to vary the energy bin width and/or sampling step size. Unfortunately, it is not possible to know *a priori* the appropriate bin width and step size to choose to

avoid spiking (if at all possible), unless inordinately large step sizes are used in conjunction with small bin widths. Even then, this is not a feasible general solution, as large moves make it difficult to sample low-energy basins, while using small bin widths is impractical, especially for systems with large energy ranges.

IV. PROPOSED SOLUTION: TRAJECTORY SWAPPING

Here, we introduce a general solution. We propose running multiple trajectories of WLS in parallel, each starting from a different random seed (or initial condition). Periodically, one randomly pairs the configuration of each trajectory with a new DOS from another trajectory. If the random walker happens to be trapped in a spike, this swapping resets the configuration to a "safe" region of energy space and stops the spike from growing further. The idea is similar to that of the replica-exchange MC [35], where a low-temperature state trapped in a local minimum is swapped with a higher temperature one, enabling it to escape and sample other regions of configuration space. Swapping among different WL random walkers also ensures that the sampling is uniform in energy space. Our method is simple to implement and completely general, applicable to any energy binning or sampling step size. The DOS values from all the trajectories can be averaged at the end of the simulation so no computational resources are wasted.

Recently, Vogel *et al.* [13] proposed a parallel WL method to extend WLS to large-scale problems. In their method, different random walkers in overlapping energy windows are run in parallel and occasionally exchanged so as to allow the entire energy range to be sampled simultaneously and quickly. The purpose of our current work, on the other hand, is to use parallelization and swapping as a way to prevent trapping. In addition, each of our random walker samples the entire energy range, and we do not define a separate transition probability for the exchange of two random walkers. The method of Vogel *et al.* may also be effective in spike prevention. However, the general focus and actual implementation of the two works are different.

We tested our algorithm on the frustrated XY model. Let $\Theta^\alpha = (\theta_1^\alpha, \ldots, \theta_{L^2}^\alpha)$ and $g^\alpha(E)$ denote, respectively, the configuration and DOS of the αth trajectory, where $\alpha = 1, \ldots, N$ denotes the random seed of each trajectory. Every $(T \times L^2)$th MC step, a shuffling algorithm randomly permutates the trajectory index, $\alpha \to \alpha'$, and assigns $\Theta^{\alpha'}$ as the new configuration of $g^\alpha(E)$. The main factors affecting the efficacy of the algorithm are the time interval between swaps $T \times L^2$ and the total number of trajectories N. Swapping should be as frequent as possible so that spikes do not have the chance to build up. Using more trajectories reduces the probability of swapping two trajectories which are simultaneously trapped in the same energy bin. On the other hand, as we implemented the algorithm using message passing interface, the swapping process requires communication between different processors and incurs computational overhead, so one should also try to keep T large and N small.

As our swapping between trajectories is accepted with probability 1, strictly speaking, it violates the transition rule of WLS. However, as the number of swaps we performed is very small compared to the entire length of the trajectory, no serious error is introduced, as shown by our numerical experiments below.

V. NUMERICAL SIMULATIONS

In our simulations, we consider one set of J_{ij} for each L [36] and study the effects of T, N, and random shuffling on the prevention of spike occurrences. The energy binning for the DOS is chosen such that the bins are narrow enough to support the energy distribution $g(E)e^{-\beta E}$ near zero temperature. Such narrow bin widths easily induce spiking in the original WLS algorithm. Admittedly, the original WLS might not induce spiking for other choices of bin widths, but we did not perform an exhaustive study of this issue, due to the unpredictability of spiking, as previously discussed. All N trajectories are launched from the same initial configuration. Each $\ln f$ stage is simulated for $5 \times 10^6 \times L^2$ MC steps, and $\ln f$ is halved after each stage. We then count the percentage of the N trajectories that have a spiked DOS after the tenth $\ln f$ stage.

The results are summarized in Table I. For each L, we studied $N = 8$, 16, and 32 and $T = 250$, 1000, 5000, and ∞, where $T = \infty$ means no swapping (i.e., original WLS). Each entry in the table shows the percentage of spiked DOS in the order $L = 16/32$. For $T = \infty$, more than half of the N trajectories exhibit spiking. For each of the rest of the $N - T$ entries, the percentage is averaged over five runs, where each run uses a different seed for the random shuffler. As expected, a large N (32) and small T (250) are the most effective in preventing spike formation. Between N and T, the latter is more important in preventing spikes. We found that is it important to keep $T < 1000$ because otherwise swapping is ineffective. Increasing N has a comparatively more gradual, albeit still significant effect in preventing spikes.

We checked the accuracy of the DOS of our swapping algorithm by computing the specific heat capacity per spin, c_v, for $L = 16$. From the five runs at $N = 32$ and $T = 250$, we took one run and continued the swapping simulation until $\ln f \approx 5.96 \times 10^{-8}$ [37]. The c_v values of all 32 trajectories were calculated and averaged, giving $\langle c_v \rangle$. The value of $\ln g(E)$ in any spiked bin is replaced by the average of its two neighboring bins. The same is repeated for $T = \infty$, and results of replica-exchange MC calculations are used as the benchmark. The results are shown in Fig. 4(a). $T = 250$ and replica-exchange MC agree very well, even at low temperatures as shown in the inset. On the other hand, the

TABLE I. Reducing spikes in the DOS by swapping. N, number of trajectories; T, number of MC steps (per spin) between swaps. Each N-T entry is ordered $L = 16/32$ and shows the percentage of the spiked DOS averaged over five runs of the random shuffler. Values in parentheses are standard deviations. For $T = \infty$ (no swapping), only one run was performed for each N and L.

	T			
N	250	1000	5000	∞
8	13(10)/0(0)	58(10)/5(6)	70(10)/48(20)	88/75
16	9(8)/0(0)	49(10)/1(3)	64(10)/25(9)	69/69
32	1(1)/0(0)	31(10)/0(0)	66(6)/38(10)	75/69

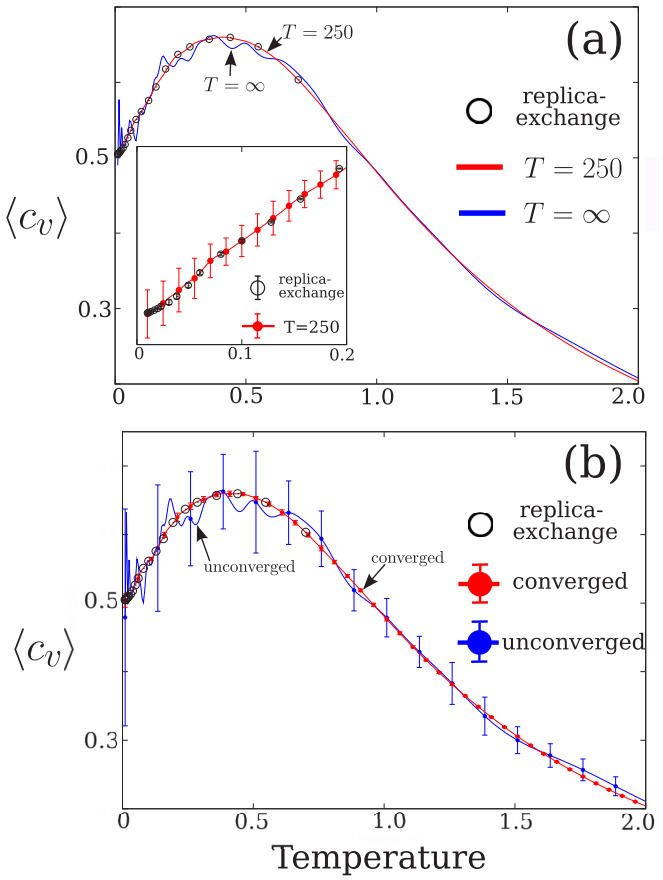

FIG. 4. (Color online) (a) Specific heat capacity per spin of the XY model for a fixed set of J_{ij} ($L = 16$). The $\langle c_v \rangle$ for $T = 250$ and $T = \infty$ is calculated by averaging over 32 trajectories. For replica-exchange Monte Carlo (open circles), $\langle c_v \rangle$ is computed by averaging over 25 independent runs. Inset: Closeup view near zero temperature of $T = 250$ (filled circles) and replica-exchange (open circles). Error bars indicate the standard deviation. The bars for replica exchange are smaller than the circles. (b) For $T = \infty$. Converged: $\langle c_v \rangle$, averaged over 14 converged trajectories. Unconverged: $\langle c_v \rangle$, averaged over 18 unconverged trajectories. Error bars indicate the standard deviation.

$\langle c_v \rangle$ of $T = \infty$ suffers from large fluctuations. We found that the trajectories of $T = \infty$ can be separated into two types: (i) Those with no spikes in their DOS give a $\langle c_v \rangle$ that has successfully converged, as shown by the red ("converged") curve in Fig. 4(b); and (ii) trajectories with spikes in their DOS sometimes spend entire or even a few ln f stages confined either within the spiked bin itself or "bouncing" between the spike and the edge of the DOS. When the random walker manages to escape, the ln f has already been lowered so much that the current modification factor does not effectively improve the DOS, and thus the DOS is effectively frozen at an earlier ln f stage. The $\langle c_v \rangle$ values computed from these trajectories are therefore unconverged, as shown by the blue ("unconverged") curve in Fig. 4(b). Our observation also explains why the annealing method of Poulain *et al.* [6] was successful. By reincreasing the ln f, their random walker is allowed to redo those earlier ln f stages when it was trapped.

VI. DISCUSSION

As mentioned above, the phenomenon of trapping is highly dependent on how the simulation is being set up, i.e., the kind of system, the energy binnings, the MC step size, etc. We would like to highlight two scenarios in which we think there is an increased likelihood of encountering trapping. The first is when it is preferable to use a small MC step size to achieve a good acceptance rate. In our study of the poly-alanine molecule, we found that the molecule sometimes folds into very tight conformations where small MC step sizes are necessary to enable the molecule to gently unfold (large moves will give rise to the "lever-arm" effect resulting in steric clashes). If small step sizes are used, then one needs to be aware of the trapping mechanism we have discussed above.

The second scenario is when one is interested in low-temperature behavior, which in WLS means sampling from low-lying energy states. To illustrate, we ran 200 independent, nonswapping trajectories for the XY model ($L = 16$) discussed above. At the end of the 15th ln f stage, we checked whether there was a spike in their DOS and, if there was, the energy bin at which the spike occurs. The energy distribution of the spike locations is shown in Fig. 5. Spiking occurs predominantly at an energy range near the ground state at $E \approx -346$. This energy range corresponds roughly to the temperature $T = 0.01$. In the inset in Fig. 5, the energy probability distribution $P(E) = g(E)e^{-E/T}$ at $T = 0.01$ is shown, and we see that the distribution is supported in an energy region where spiking is dominant. Hence, one should be aware of the possibility of trapping when one investigates low-temperature phenomena.

On the other hand, trapping might not be so important when one studies high-temperature behavior. We use the same XY model as an example. In Fig. 4, we see that the transition temperature is around $T = 0.4$. In the inset in Fig. 5, we plot the $P(E)$ at $T = 0.4$. It is shown that the distribution lies quite far away from the energy region where spiking occurs.

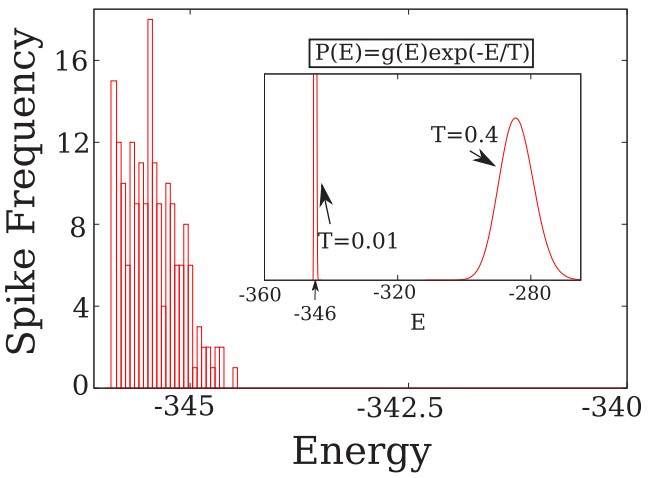

FIG. 5. (Color online) Distribution of spike locations with respect to energy for the frustrated XY model ($L = 16$). Inset: Energy probability distribution $P(E) = g(E)e^{-E/T}$ at temperature $T = 0.01$ (corresponding to $E \approx -346$, where spiking is dominant) and at $T = 0.4$ (phase transition temperature).

Hence, one might impose a low-energy threshold to prevent the random walker from visiting low-energy states if one is only interested in obtaining the transition temperature.

In our algorithm, swapping between trajectories occurs at probability 1. As mentioned, this violates detailed balance. One way to rectify this is to treat the swapping as an MC step by itself. The acceptance probability for the swap has been proposed by Vogel *et al.* [13],

$$P_{\text{accept}} = \min\left[1, \frac{g_i(E[X])}{g_i(E[Y])}\frac{g_j(E[Y])}{g_j(E[X])}\right], \tag{5}$$

where X and Y are the configurations of trajectories i and j before the swap, and $E[X]$ and $E[Y]$ are their energies, respectively. $g_i(E[X])$ is the DOS of trajectory i at energy $E[X]$. Hence, every few steps, a swap can be attempted and accepted according to Eq. (5). This way, detailed balance is obeyed and the swapping interval T can be made quite small.

If the system under study is small enough for many multiple trajectories to be simulated on a single processor, one should indeed adopt Eq. (5). However, for large systems where one assigns different trajectories to different processors, one needs to take into account the computational overhead coming from the communication among the processors during swapping. It may then be too expensive to use Eq. (5) to swap since frequent swapping will slow the computation down. On the rare occasion when a swap is attempted, it is best to let the swap succeed as we have done in this paper. Although theoretically less rigorous, our method is faster. Bouzida *et al.* have also shown within the context of the acceptance ratio method that occasional and infrequent violation of detailed balanced does not incur serious error [38]. Our calculation of the specific heat capacity also shows that there is no discernible error incurred from this violation of detailed balance during swapping.

Finally, we discuss the case of a piecewise constant DOS alluded to earlier. The original WLS algorithm takes the value of $\ln g(E)$ as constant within each energy bin. This will not give rise to any spiking, as without linear interpolation the mechanism in Fig. 2 no longer holds. However, for piecewise constant DOS the energy bins at the very edge of the DOS will usually only be visited after the first few $\ln f$ stages have been completed. By the time the random walker drops into these bins, the difference in $\ln g(E)$ between these bins and the neighboring, higher energy ones is already so large that $\exp[\Delta \ln g(E)]$ is effectively 0, and the random walker stays a long time in these bins to let the DOS "catch up" before it can escape. This leads to another form of trapping. The swapping algorithm presented here is also applicable to this problem encountered by a piecewise constant DOS.

ACKNOWLEDGMENTS

We thank P. Poulain for useful discussions. This work was supported (in part) by the Biomedical Research Council of A*STAR (Agency for Science, Technology and Research), Singapore.

[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] N. G. Fytas and A. Malakis, Eur. Phys. J. B **79**, 13 (2011).
[4] P. E. Theodorakis and N. G. Fytas, Eur. Phys. J. B **81**, 245 (2011).
[5] M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos, Phys. Rev. E **66**, 056703 (2002).
[6] P. Poulain, F. Calvo, R. Antoine, M. Broyer, and Ph. Dugourd, Phys. Rev. E **73**, 056704 (2006).
[7] D. Jayasri, V. S. S. Sastry, and K. P. N. Murthy, Phys. Rev. E **72**, 036702 (2005).
[8] P. Poulain, F. Calvo, R. Antoine, M. Broyer, and Ph. Dugourd, Europhys. Lett. **79**, 66003 (2007).
[9] D. T. Seaton, T. Wüst, and D. P. Landau, Phys. Rev. E **81**, 011802 (2010).
[10] T. Nagasima, A. R. Kinjo, T. Mitsuiand K. Nishikawa, Phys. Rev. E **75**, 066706 (2007).
[11] A. D. Swetnam and M. P. Allen, J. Comput. Chem. **32**, 816 (2010).
[12] S.Æ. Jónsson, S. Mohanty, and A. Irbäck, J. Chem. Phys. **135**, 125102 (2011).
[13] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013).
[14] L. Gai, T. Vogel, K. A. Maerzke, C. R. Iacovella, D. P. Landau, P. T. Cummings, and Clare McCabe, J. Chem. Phys. **139**, 054505 (2013).
[15] R. Faller and J. J. de Pablo, J. Chem. Phys. **119**, 4405 (2003).
[16] R. E. Belardinelli and V. D. Pereyra, J. Chem. Phys. **127**, 184105 (2007).
[17] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
[18] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
[19] J. Lee, Phys. Rev. Lett. **71**, 211 (1993).
[20] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **329**, 261 (2000).
[21] A. Mitsutake, Y. Sugita, and Y. Okamoto, J. Chem. Phys. **118**, 6664 (2003).
[22] A. Mitsutake, Y. Sugita, and Y. Okamoto, J. Chem. Phys. **118**, 6676 (2003).
[23] A. Mitsutake, Y. Mori, and Y. Okamoto, Methods Mol. Biol. **924**, 153 (2013).
[24] J. Kim, J. E. Straub, and T. Keyes, Phys. Rev. Lett. **97**, 050601 (2006).
[25] J. Kim, T. Keyes, and J. E. Straub, J. Chem. Phys. **130**, 124112 (2009).
[26] G. Brown and T. C. Schulthess, J. Appl. Phys. **97**, 10E303 (2005).
[27] J. Wang, P. Cieplak, and P. A. Kollman, J. Comp. Chem. **21**, 1049 (2000).
[28] http://www.cs.ox.ac.uk/mosaics/.
[29] C. Zhou and R. N. Bhatt, Phys. Rev. E **72**, 025701(R) (2005).
[30] Q. Yan and J. J. de Pablo, Phys. Rev. Lett. **90**, 035701 (2003).
[31] A. N. Morozov and S. H. Lin, Phys. Rev. E **76**, 026701 (2007).
[32] A. N. Morozov and S. H. Lin, J. Chem. Phys. **130**, 074903 (2009).

[33] H. K. Lee, Y. Okabe, and D. P. Landau, Comput. Phys. Commun. **175**, 36 (2006).

[34] C. Zhou, T. C. Schulthess, S. Torbrügge, and D. P. Landau, Phys. Rev. Lett. **96**, 120201 (2006).

[35] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).

[36] Although one should, in principle, average the results over the distribution of J_{ij}, as the system is approximately self-averaging at the large L we are considering, we choose instead to focus calculations on other aspects.

[37] The ln f schedule is as follows. From ln $f = (0.5)^0$ to $(0.5)^{13}$, 5×10^6 MCS per spin; ln $f = (0.5)^{14}$ to $(0.5)^{17}$, 4×10^7 MCS per spin; ln $f = (0.5)^{18}$ to $(0.5)^{20}$, 10^8 MCS per spin; ln $f = (0.5)^{21}$ to $(0.5)^{22}$, 2×10^8 MCS per spin; and ln $f = (0.5)^{23}$ to $(0.5)^{24}$, 4×10^8 MCS per spin.

[38] D. Bouzida, S. Kumar, and R. H. Swendsen, Phys. Rev. A **45**, 8894 (1992).