Stochastic Approximation Monte Carlo with a Dynamic Update Factor

Jordan K. Pommerenck, Tanner T. Simpson, Michael A. Perlin, and David Roundy
Department of Physics, Oregon State University, Corvallis, OR 97331

We present a new Monte Carlo algorithm based on the Stochastic Approximation Monte Carlo (SAMC) algorithm for directly calculating the density of states. The proposed method is Stochastic Approximation with a Dynamic update factor (SAD) which dynamically adjusts the update factor \( \gamma_t \) during the course of the simulation. We test this method on the square-well fluid and the 31-atom Lennard-Jones cluster and compare the convergence behavior of several related Monte Carlo methods. We find that both the SAD and \( 1/t \)-Wang-Landau (1/t-WL) methods rapidly converge to the correct density of states without the need for the user to specify an arbitrary tunable parameter \( t_0 \) as in the case of SAMC. SAD requires as input the temperature range of interest, in contrast to 1/t-WL, which requires that the user identify the interesting range of energies. The convergence of the 1/t-WL method is very sensitive to the energy range chosen for the low-temperature heat capacity of the Lennard-Jones cluster. Thus, SAD is more powerful in the common case in which the range of energies is not known in advance.

I. INTRODUCTION

Over the past several decades, a number of flat histogram Monte Carlo simulation algorithms have been developed which calculate the thermodynamic properties of various systems over a range of temperatures. This development began with the original histogram method, which used a single canonical Monte Carlo simulation to predict properties for nearby temperatures [1]. For large systems, this approach is limited to a narrow temperature range because a single canonical simulation explores only a small range of energies. Berg and Neuhaus developed multicanonical methods that introduced a weight function to enable flat histogram sampling which improved the exploration of configuration space and allowed the system to overcome free energy barriers [2, 3]. These works led to an explosion in the interest of “flat” (or “broad”) histogram methods [4]–[7], which explore a wider range of energies. In addition to simulating a range of temperatures, in contrast with low-temperature canonical Monte Carlo, these approaches avoid being trapped in a local energy minimum.

Wang and Landau introduced the most widely used flat histogram algorithm (WL) that uses an update factor and a statistical histogram to compute the density of states of a given system [7, 8]. While the method is incredibly powerful, it has a few disadvantages. The most significant disadvantage is that the method requires the user to select the range of energies to be studied [8–11]. This requirement adds an additional hurdle to its application to systems for which the interesting range of energies is not known \textit{a priori}. The simulation violates detailed balance, albeit briefly as the size of the violation decreases with time, which complicates convergence analysis. In fact, the error in a WL computation has been demonstrated to saturate at a non-zero value [11], i.e. the method does not converge to the true density of states [13, 15].

Belardinelli and Pereyra demonstrated that allowing the update factor to decrease faster than \( 1/t \) leads to nonconvergence [13], where \( t \) corresponds to the number of moves. This leads to their \( 1/t \)-WL algorithm which ensures that the error continues to decrease asymptotically as \( 1/\sqrt{t} \), which they demonstrated avoids error saturation and asymptotically approaches the true density of states [15]. Zhou et al. further confirmed that the original WL algorithm never converges exponentially and successfully bounded the statistical error between \( t^{-\frac{1}{2}} \) and \( 1/t \) [15]. Schneider et al. outline minor refinements to the \( 1/t \)-WL algorithm including scaling the update factor with the number of energy bins [15].

Liang began to consider Stochastic Approximation as a mathematical generalization of the WL approach, with convergence that could be mathematically proven [20]. In 2007, Liang et al. developed Stochastic Approximation Monte Carlo (SAMC) [21, 22], and proved its convergence, although the method still has a system specific user-defined parameter which must be tuned when applying this algorithm to a new system. In contrast to the WL-based methods, the SAMC method does not require users to determine the energy range of interest \textit{a priori}. Werlich et al. proposed the introduction of an additional tuning parameter into SAMC [23].

Kim et al. introduced Statistical Temperature Monte Carlo (STMC) and the related Statistical Temperature Molecular Dynamics (STMD) which is an adaption of the WL method that approximates the entropy (or natural logarithm of the density of states) as a piecewise linear function, which improves convergence for systems with a continuously varying energy [24, 25]. STMC applied to WL requires a temperature range be specified rather than an energy range. Kim et al. extended this work as Replica Exchange Statistical Temperature Monte Carlo (RESTMC), which uses replica exchange of multiple overlapping STMC simulations to improve convergence [26]. More recently, Junghans et al. have demonstrated a close connection between metadynamics, which was introduced by Laio and Parinello [28], and WL-based Monte Carlo methods, with STMD forging the connection [27].

In this work, we have developed an improved algo-
algorithm based on SAMC that does not require an array of non-physical, user-defined inputs and therefore should be easily applicable to any system. The method (like STMC above) does require the user to define a temperature range of interest \( (T_{\text{min}} \text{ to } T_{\text{max}}) \) which we explain in Section II. We call this method SAD (Stochastic Approximation with a Dynamic update factor), and will discuss it in detail in the methods section. We compare its convergence properties with three existing flat histogram methods: WL, \( 1/t \)-WL, and SAMC.

In this work, we compare four flat histogram methods. We outline the general workings of each algorithm that we developed in detail while summarizing algorithms that were developed in other works. The following methods are discussed and simulated for the square-well fluid and the 31-atom Lennard-Jones cluster: Wang-Landau (WL), \( 1/t \)-Wang-Landau (\( 1/t \)-WL), Stochastic Approximation Monte Carlo (SAMC), and SAD.

II. FLAT HISTOGRAM METHODS

The goal of flat histogram methods (also called broad histogram or multicanonical methods) is to simulate each energy with similar accuracy so as to accurately determine the density of states over a broad range of energies—and thus to determine the thermodynamic quantities such as heat capacity or internal energy over a broad range of temperatures. Properties that require more information—such as a spatial correlation function or a response function—can still be computed for any temperature, provided statistics are collected for each individual energy, which can then be reweighted for any temperature.

All the flat histogram Monte Carlo methods begin with randomly chosen “moves” which change the state of the system and must satisfy detailed balance. Each algorithm differs in how it determines the probability of accepting a move and in what additional statistics must be collected in order to decide on that probability.

Flat histogram methods calculate the density of states \( D(E) \) for a discrete set of energies \([7, 12, 32, 33]\). Therefore, energy binning becomes an important consideration for systems with a continuum of possible energies. Energy bins are typically of uniform size for the entire energy continuum \([34]\). Some methods such as AdaWL \([35]\) employ a tunable mechanism for controlling the binning for low entropic states in order to ensure the exploration of all energies. The method introduced in this paper is designed to scale appropriately as bin size is changed.

In this section we will introduce four closely related flat histogram methods each of which rely on a weight function \( w(E) \). In these algorithms, the probability of accepting a move is given by

\[
P(E_{\text{old}} \rightarrow E_{\text{new}}) = \min \left[ 1, \frac{w(E_{\text{old}})}{w(E_{\text{new}})} \right] \tag{1}
\]

which biases the simulation in favor of energies with low weights. A set of weights that are proportional to the density of states \( D(E) \) of the system will result in an entirely flat histogram, which means that the weights should converge to being proportional to the density of states. The natural logarithm of the weight is typically stored, since the density of states will often vary over a few hundred orders of magnitude. In the microcanonical ensemble, the entropy is defined as \( S(E) = \ln(D(E)) \) (where \( k_B = 1 \)), the logarithm of the weight is an approximation of the entropy.

Each approach uses a random walk in energy space to estimate the density of states. The core of these approaches is to update the weights at each step of the simulation

\[
\ln w_{t+1}(E) = \ln w_t(E) + \gamma_t \tag{2}
\]

where \( t \) is the number of the current move, \( \gamma_t \) is an update factor that varies over the simulation, and \( E \) is the current energy. This update causes the random walk to avoid energies that have been frequently sampled, enabling a rapid exploration of energy space. This approach violates detailed balance, due to the acceptance probabilities changing with each move, but the severity of this violation decreases as we decrease \( \gamma_t \). The four methods differ primarily in how they schedule the decrease of \( \gamma_t \).

A. Wang-Landau

The Wang-Landau approach \([7, 8, 36]\) begins with \( \gamma_{t=0} = 1 \), and then decreases \( \gamma_{\text{WL}} \) in discrete stages. When the energy histogram is sufficiently flat, \( \gamma_{\text{WL}} \) is decreased by a specified factor of \( \frac{1}{2} \). The flatness is defined by the ratio between the minimum value of the histogram and its average value. When this flatness reaches a specified threshold (typically 0.8), the \( \gamma_{\text{WL}} \) value is decreased and the histogram is reset to zero. This approach requires that the energy range of interest be known in advance, and difficulties can occur with this flatness criteria due to the fact that some energies in this energy range might never be sampled \([37]\). The entire process is repeated until \( \gamma_{\text{WL}} \) reaches a desired cutoff.

The Wang-Landau approach thus has three parameters that need be specified: the factor by which to decrease \( \gamma_{\text{WL}} \) when flatness is achieved, the flatness criterion, and the cutoff that determines when the computation is complete. In addition, an energy range (or in general, a set of energies) must be supplied, so that the flatness criterion can be defined.

B. \( 1/t \)-Wang-Landau

The \( 1/t \)-WL algorithm ensures convergence by preventing the \( \gamma_t \) factor from dropping below \( N_S/t \) \([15, 19]\). The method follows the standard WL algorithm with two modifications. Firstly, the histogram is considered flat,
and $\gamma_t$ is decreased by a factor of two, when every energy state has been visited once, i.e. when the WL “flatness” becomes nonzero. Secondly, when $\gamma_{WL}^{WL} < N_S/t$ at time $t_0$, the algorithm switches to use $\gamma_t = N_S/t$ for the remainder of the simulation:

$$\gamma_t^{1/t_{WL}} = \begin{cases} \gamma_{WL}^{WL} & \gamma_{WL}^{WL} > N_S/t \\ \frac{t_0}{N_S} \gamma_t & t \geq t_0 \end{cases}$$

(3)

where $t$ is the number of moves, $\gamma_t^{WL}$ is the Wang-Landau update factor at move $t$, and $N_S$ is the number of energy bins.

C. SAMC

The Stochastic Approximation Monte Carlo (SAMC) algorithm addresses the lack of convergence of Wang-Landau’s approach with a simple schedule by which the update factor $\gamma_t$ is continuously decreased [19, 21, 23]. The update factor is defined in the original implementation [21] in terms of an tunable parameter $t_0$,

$$\gamma_t^{SA} = \frac{t_0}{\text{max}(t_0, t)}$$

(4)

where as above $t$ is the number of moves that have been attempted. SAMC offers extreme simplicity, combined with is proven convergence. Provided the update factor satisfies

$$\sum_{t=1}^{\infty} \gamma_t = \infty \quad \text{and} \quad \sum_{t=1}^{\infty} \gamma_t^\zeta < \infty$$

(5)

where $\zeta > 1$, Liang has proven that the weights converge to the true density of states [20, 22]. In addition, the energy range need not be known a priori. The time to converge depends only on the choice of parameter $t_0$. Unfortunately, there is no prescription for finding an acceptable value for $t_0$, and while the algorithm formally converges, for a poor choice of $t_0$ that convergence can be far too slow to be practical. Liang et al. give a rule of thumb in which $t_0$ is chosen in the range from $2N_S$ to $100N_S$ where $N_S$ is the number of energy bins [21]. Schneider et al. found that for the Ising model this heuristic is helpful for small spin systems, but that larger systems require an even higher $t_0$ value [19]. We will describe below one case we examined, in which $t_0$ needs to be as much as two orders of magnitude higher than the rule of thumb of $100N_S$ in order to converge in $10^{12}$ moves.

Werlich et al. proposed scaling the SAMC $\gamma_t^{SA}$ by a factor $\gamma_0$ [24]. While this may result in an improved rate of convergence, it adds yet another parameter that must be empirically determined, and we have not explored this additional degree of freedom.

D. SAMC convergence time

A primary difficulty in using the SAMC method lies in identifying an appropriate value for $t_0$. Although SAMC is proven to formally converge regardless of the $t_0$ value, a choice that is either too high or too low will result in prohibitively slow convergence to the true entropy of the system. It is instructive to consider separately values of $t_0$ that are too low or too high.

We can place a rigorous lower bound $t_{min}$ on the number of moves required to find the true entropy by considering the total change that needs to be made to the entropy:

$$\Delta S_{tot} \equiv \sum_E S(E) - S_{min}.$$

(6)

This connects with $\gamma_t$ because the $\ln w(E)$ (our approximation for entropy) has $\gamma_t$ added to it on each move. Thus a minimum number of moves that could possibly result in the true entropy $S(E)$ starting with a flat set of weights is determined by the total entropy change required. We can estimate this number of moves required by summing the $\gamma_t$, which we can approximate using an integral:

$$\Delta S_{tot} = \sum_{t=0}^{t_{min}} \gamma_t^{SA}$$

(7)

$$\approx t_0 + \int_{t_0}^{t_{min}} \frac{t_0}{t} dt$$

(8)

$$= t_0 \left( 1 + \ln \left( \frac{t_{min}}{t_0} \right) \right)$$

(9)

Solving for $t_{min}$ we find that

$$t_{min} = t_0 e^{\frac{\Delta S_{tot}}{t_0} - 1}$$

(10)

which means that the minimum time to converge grows exponentially as $t_0$ is made smaller. You seriously don’t want to underestimate $t_0$!

One might reasonably choose to err by selecting a large $t_0$. The rate of convergence is harder to estimate when $t_0$ is large, but in general $\gamma_t^{SA}$ itself forms a lower bound on the accuracy with which the entropy may be known, with an unknown prefactor which is related to the coherence time ($t = t_0$) of the Monte Carlo simulation. Since $\gamma_t^{SA}$ is given by $t_0/t$, the time to converge to a given accuracy is increased in proportion to the ratio by which we overestimate $t_0$. Thus, while it is exponentially painful to underestimate $t_0$, overestimating by several orders of magnitude is also not acceptable. We should note that these extreme limiting cases do not preclude the possibility that there is a wide range of $t_0$ values that lead to an acceptable convergence rate.

III. SAD ALGORITHM

The Stochastic Approximation with Dynamic update factor (SAD) method is a variant of the SAMC Algorithm that attempts to dynamically choose the modification
While for SAMC, the update factor is defined in the original implementation (see equation (11)), for SAD the update factor $\gamma_{t}^{\text{SAD}}$ is thought of as $\frac{dS}{dt}$. This tells us that the SAMC parameter $t_0$ should have dimensions of entropy. We begin with an estimate of the average value of the entropy (relative to the lowest entropy at $T_{\text{min}}$). If we assume a quadratic dependence on energy (see Fig. 1), this is given by

$$\langle S \rangle \approx \frac{1}{3} \left( E(T = \infty) - E(T_{\text{min}}) \right)$$

We approximate this energy difference by $E_H - E_L$ where $E_H$ and $E_L$ are defined below. The entropy numerator of the update factor in general should scale with the total number of interesting energy states $N_S$, since updates to the weights are distributed between that many energy states. The product $N_S \langle S \rangle$ is the total change of entropy required (starting from constant weights) to find the true entropy, and puts a lower bound on the convergence time. After long times, when all the energies have been explored a long time ago, we wish for a lower update factor in order to more rapidly refine the remaining error in entropy. We track the time at which we first visited each possible energy. We define $t_L$ to be the last time that we encountered an energy that we currently believe is in the energy range of interest, so a $t \gg t_L$ we feel confident that we have established the true entropy range of interest. We gradually transition to a lower update factor (but still asymptotically scaling as $\gamma_t \propto 1/t$ to ensure eventual convergence). Finally, we wish for an update factor that is never greater than 1, because a very large update factor could introduce very large errors in entropy that may take many iterations to remove. The SAD expression for $\gamma_t$ which incorporates these ideas is:

$$\gamma_{t}^{\text{SAD}} = \frac{E_H - E_L}{E_H - E_L} + \frac{t}{t_L}$$

where $E_H$ and $E_L$ are the current estimates for the highest and lowest energies of interest as defined below. This factor asymptotically has the same $1/t$ behavior as the original SAMC algorithm and with the same $N_S$ prefactor used by the 1/t-WL method; however for earlier values of $t$, the update factor drops as $1/t^2$ and jumps every time a new energy is determined to be of interest. This behavior allows SAD to dynamically prevent the update factor from decreasing too rapidly.

Figure 2 compares $\gamma_t$ for the related methods SAD, WL, 1/t-WL, and SAMC. For SAMC, $\gamma_t$ remains constant before dropping as $1/t$. WL $\gamma_t$ remains at 1 for many iterations, and then decreases very rapidly, with 1/t-WL behaving similarly but decreasing more aggressively before transitioning to a more conservative 1/t behavior. The update factor for SAD fluctuates dynamically around a value less than 1 for early MC moves, and then decreases as approximately $1/t$ while continuing to fluctuate as new energies are found to be important. At intermediate times, the SAD $\gamma_t$ decreases as $1/t^2$ before asymptoting to $N_S/t$, which is the same as 1/t-WL.
we can define the interesting energy range as $E(T_{\text{min}}) < E < E(T = \infty)$ where $E(T)$ is the energy that maximizes $S - E/T$. During the course of the simulation, this precise energy is challenging to evaluate accurately. In order to ensure that we sample this entire energy range adequately, we define two energy limits: a high energy $E_H$ and a low energy $E_L$, which define the range over which the energy histogram is made flat. At move $t$, $E_H$ and $E_L$ are the greatest and lowest energy that prior to that move had the highest histogram value (i.e. been visited the most times) at some point during the course of the simulation. This definition results in a “ratcheting” effect, in which $E_H$ may only increase, while $E_L$ may only decrease over the course of the simulation, which results in a conservative estimate of the range of energies that need be sampled.

During the simulation when considering a move inside the energy range of interest $E_L \leq E \leq E_H$, the weights are used as in the three methods already described. If $E \geq E_H$, the weight is taken to be

$$w(E > E_H) = w(E_H),$$

which corresponds to an infinite temperature. This choice ensures that if the maximum in entropy is at an energy $E_{\text{max}} > E_H$, then the energy $E_{\text{max}}$ will eventually have the highest number of counts and the ratcheting will result in $E_H \geq E_{\text{max}}$. At lower energies, Boltzmann weights corresponding to the minimum temperature are used:

$$w(E < E_L) = w(E_L)e^{-E/L}.\quad (14)$$

This choice has the result that if the energy $E_{\text{min}}$ at which the free energy at $T_{\text{min}}$ is minimized is less than $E_L$, the lower energy limit will ratchet down to include $E_{\text{min}}$. Each time we change the value of $E_H$ or $E_L$, the weights outside the new portion of the interesting energy range are set to the expressions in Equations 13 and 14.

A significant advantage of SAD over SAMC—which the 1/t-WL and WL methods share after they have discovered all the energies—is that the schedule for $\gamma_t$ automatically responds to the choice of bin size. SAD should perform similarly over a reasonable range of bin sizes because $\gamma_t \propto N_S/t$. As the number of energy states $N_S$ found increases (fine binning), the time spent $t$ in each bin will decrease with the effect that the convergence should be roughly independent of the bin size chosen. SAMC could be used with a prefactor $\gamma_0$ to aid in a similar way but this adds yet another parameter for the user to choose.

IV. RESULTS

A. Square-well fluid

As our first test case, we consider the square-well fluid i.e. a system of particles whose interactions are governed by a square-well potential [38, 39]. The square-well potential is an ideal test-bed as it is a simple model for a liquid, which includes both attractive and repulsive interactions [40, 41]. To date, there have not been any published direct convergence comparison tests for flat histogram methods applied to the square-well fluid. The potential $U(r)$ for such a system is given by

$$U(r) = \begin{cases} 
\infty & |r| < \sigma \\
-\epsilon & \sigma < |r| < \lambda \sigma \\
0 & |r| > \lambda \sigma 
\end{cases} \quad (15)$$

where $\sigma$ is the hard-sphere diameter of the particle, $\lambda$ is the reduced range of the potential well, and $\epsilon$ is its depth. This model has the further advantage that binning is not required because the energy is discrete.

We tested the algorithms on two square-well fluid systems. The first is a smaller simulation with a particle number of 50, a well-width of $\lambda = 1.3$, and a volume corresponding to a filling fraction (defined as the fraction of volume filled by atoms) of $\eta = 0.3$. The second system is larger, with a particle number of 256, a well-width of $\lambda = 1.5$, and a volume corresponding to a filling fraction of $\eta = 0.17$. For each system we use a reasonable root-mean-square displacement distance $\delta_0 = 0.05\sigma$ for proposed moves, and for the smaller system we also use an unreasonably small displacement distance of 0.005$\sigma$.

The simulations explore the energy space of the systems
with minimum reduced temperatures of $T_{\text{min}} = 1/3$ for simulations of the smaller system, and $T_{\text{min}} = 1$ for the larger system. All simulations lead to the minimum important energy $E_{\text{min}}$ and maximum entropy energy $E_{\text{max}}$ being calculated (with the exception of the WL methods where both of these parameters are needed \textit{a priori}).

The SAMC simulations computed the density of states for the entire range of possible energies. The SAD simulations determined the energy range of interest dynamically as described above, based on a specified $T_{\text{min}}$. For the WL and $1/t$-WL simulations, we constrained the simulation to remain in the energy range corresponding to $T_{\text{min}} < T < \infty$, as determined by a previous SAMC simulation. Thus the WL and $1/t$-WL simulations were given extra information that in practice would not be available without additional computational effort, and the SAMC simulations computed the entropy over the entire range of possible energies, which required more effort.

We use the average entropy error versus moves as a metric to compare simulation runtimes and overall convergence. The overall accuracy is determined by examining the fractional error of a particular method to a precise reference system. For each simulation, the reference system is chosen to be the final output of a SAMC simulation with a fixed energy range corresponding to the temperature range of interest. Although SAMC does not require an energy range as an input parameter, we find that by limiting the simulation to this energy range, we can achieve much faster convergence with a smaller $t_0$. We compute an average of the error by averaging the error in the entropy over the interesting energy range, and then averaging this error over several simulations run with different random number seeds.

FIG. 3. (a) The average entropy error for each MC method for $N = 50$, $\delta_0 = 0.05\sigma$, $\eta = 0.3$, and $T_{\text{min}} = 1/3$ as a function of number of iterations run. The error is averaged over 8 independent simulations, and the best and worst simulations for each method are shown as a semi-transparent shaded area, and (b) the update factor $\gamma_t$ versus iteration number for the same simulations. (c) The average entropy error for each MC method for the same physical system with a smaller displacement distance $\delta_0 = 0.005\sigma$, as a function of number of iterations run, and (d) the update factor $\gamma_t$ versus iteration number for the same simulations.
1. 50 square-well atoms

For this 50 atom simulation, we chose a minimum reduced temperature of 1/3, which corresponds to an interesting energy range from $-248$ to $-120$. The number of important energy states for this system is therefore $N_S = 129$. The entropy of this system is shown in Fig. [1] above, which shows that over this energy range the entropy differs by 198, corresponding to a ratio of $10^{86}$ between the highest and lowest density of states.

In order to explore the effect of simulation details on convergence, we consider two values for the displacement distance by which atoms are moved during a Monte Carlo step. We began with a reasonable displacement distance of $\delta_0 = 0.05 \sigma$, which corresponds to an acceptance rate of proposed moves of 38%. We further ran simulations with a much smaller displacement distance of $\delta_0 = 0.005 \sigma$, which resulted in an acceptance rate of 86%, which converged more slowly.

Figure [3] shows the average error in the entropy as a function of time for this system with the reasonable displacement distance of $\delta_0 = 0.05 \sigma$. The solid/dashed lines represent the average of the absolute value of the error in the entropy averaged over eight simulations using different random number seeds. The range of average errors for each simulation is shown as a shaded region around its mean error. By the time $10^8$ moves have been made all but the SAMC simulation with the shortest $t_0$ have begun to converge as $1/\sqrt{t}$. We then see the WL error saturate around $10^{10}$ moves.

Figure [2] shows the average error in the entropy as a function of time for this system with the unreasonably small displacement distance of $\delta_0 = 0.005 \sigma$. The smaller translation scale causes all methods to take additional time to explore all energies. Based on random walk scaling, the convergence time of an ideal method should scale roughly as $\delta_0^{-2}$ in the limit of small $\delta_0$, that is, one order of magnitude decrease in the displacement distance should result in two order of magnitude increase in convergence time. SAMC simulations with a $t_0$ value that rapidly converged for $\delta = 0.05 \sigma$ do not converge at all in $10^{12}$ moves for a translation scale of $\delta = 0.005 \sigma$. It is also worth noting that for the smaller displacement distance, the SAMC rule of thumb of choosing $t_0$ to be approximately $100N_S$ is no longer valid. SAD, WL, and 1/$t$-WL handle the shift in displacement distance and converge roughly as expected.

The methods SAD, WL, and 1/$t$-WL compensate for the smaller displacement distance by reducing $\gamma_t$ more slowly, as can be seen from Figure [3] and [4]. The update factors take approximately $10 \times$ longer to reach steady-state for the smaller displacement distance. Because of this update behavior, these methods are less sensitive to the choice of displacement distance than SAMC is.

2. 256 square-well atoms

Next we introduce a considerably larger simulation containing 256 atoms which has a maximum entropy about 1500 greater than its minimum. This makes exploring the entire range of energies extremely expensive, and strongly favors the methods that restrict the energy (or temperature) range of interest. For this simulation, we chose a much higher minimum reduced temperature of 1.0, which corresponds to an interesting energy range from $-915$ to $-509$. The number of important energy states for this system is then $N_S = 407$. The minimum entropy over this energy range is just 395 less than the maximum, corresponding to a ratio of only $10^{118}$ between the highest and lowest density of states.

Figure [4] shows the average error in the entropy as a function of moves for this system with the reasonable displacement distance of $\delta_0 = 0.05 \sigma$. The solid lines represent the average of the absolute value of the error in the entropy averaged over eight simulations using different random number seeds. The range of average errors for each simulation is shown as a shaded region around its mean error. By the time $10^8$ moves have been made all but the SAMC simulation with the shortest $t_0$ shown have begun to converge as $1/\sqrt{t}$. We then see the WL error saturates around $10^{10}$ moves. Once again, the convergence of SAD is essentially the same as that of the 1/$t$-WL method.
FIG. 5. The heat capacity of the LJ31 cluster after $10^{12}$ moves as calculated with each method. Each simulation used a bin size of $\Delta E = 0.01\epsilon$.

B. 31-atom Lennard-Jones cluster

We also examined convergence of each algorithm on the 31-atom Lennard-Jones cluster (LJ31), which is the smallest Lennard-Jones cluster that exhibits broken ergodicity and a solid-solid phase transition at a low temperature [42, 43]. Clusters of Lennard-Jones atoms have been frequently used for testing Monte Carlo algorithms [44–47]. The LJ31 cluster is commonly used for testing convergence of algorithms because it features a two-funnel energy landscape with a significant barrier between the two low-energy states [43, 48–50]. In particular, the LJ31 cluster features a low-temperature solid-solid phase transition that is challenging to converge [47]. Poulain et al. tested a number of variants on the WL algorithm on LJ31 and failed to converge the spike in heat capacity corresponding to the solid-solid phase transition [50].

Figure 5 shows the heat capacity of the LJ31 cluster over the temperature range from $k_B T = 0.01\epsilon$ to $k_B T = 0.40\epsilon$. There are two peaks, a large peak around $k_B T \approx 0.3\epsilon$ corresponding to melting temperature, and a small peak around $k_B T \approx 0.027\epsilon$ for the solid-solid transition from a Mackay-to-anti-Mackay (M → aM) transition [47, 51, 52]. We constrain our atoms within a spherical box of radius $2.5\sigma$, as is common in the literature [47, 50]. We note that using a larger box (with radius 5\sigma) has a large effect on the melting peak, and a smaller but still significant effect on the solid-solid transition. Since our focus is on the convergence rate, we will restrict ourselves to the smaller box size. In addition, we further restricted all our simulations to negative total potential energies.

In order to test the low-temperature convergence, we focused on the heat capacity in the temperature range $0.01\epsilon \leq k_B T \leq 0.05\epsilon$ (see inset in Fig. 5), which contains the solid-solid phase transition. We compute the heat capacity from the entropy using a standard canonical analysis, which gives

$$C_V = 31 \times \frac{3}{2} k_B + \frac{1}{k_B T^2} \langle (E - \langle E \rangle)^2 \rangle \quad (16)$$

where the first term represents the kinetic energy contribution, and the second term is the potential energy contribution, with canonical averages using the microcanonical entropy to compute multiplicities. For our reference heat capacity, we ran a 1/t-WL simulation with the energy constrained to a range $-133.53\epsilon \leq E \leq -110\epsilon$. We chose this energy range through experimentation, as we found the convergence rate to be very sensitive to the minimum energy chosen.

We ran simulations with several bin widths using each algorithm we studied and show results for $\Delta E = 0.1\epsilon$ and $0.01\epsilon$. When examining the converged results, we saw that each bin size yielded a measurably different low temperature heat capacity. The largest bin size of 0.1\epsilon introduced a maximum error in heat capacity of more than 20k_B compared with the smaller bin size. We will focus on the smaller bin size, and used a bin size of 0.01\epsilon for our converged reference heat capacity.

Figure 6 shows the maximum error in heat capacity over the temperature range $0.01\epsilon \leq k_B T \leq 0.05\epsilon$ as a function of the number of Monte Carlo moves performed. We explored the convergence behavior of each algorithm at two energy bin widths ($\Delta E = 0.01\epsilon$ and $\Delta E = 0.1\epsilon$). We were unable to converge the WL or 1/t-WL algorithms with the larger bin size. The rate of convergence of SAD for the heat capacity seems to be roughly inde-
pendent of the bin width. In contrast (but not shown) the entropy itself shows much larger statistical fluctuations with smaller bin sizes. The canonical averaging process in Eq. [16] averages out those statistical fluctuations, which reduces the penalty of using a smaller bin size.

The convergence behavior of the WL and 1/t-WL is incredibly sensitive to the minimum energy chosen, which complicates the process of obtaining a workable result. The known ground state energy of this cluster is $-133.586422e$ [44, 53]. With a minimum energy of $-133.55e$, none of our WL or 1/t-WL simulations converged in $10^{12}$ moves. With minimum energies of $-133.54e$, $-133.53e$, and $-133.52e$, the simulations do seem to converge, although the lowest energy case was significantly slower. This result highlights the advantage of being able to specify a temperature range rather than an energy range when computing properties as a function of temperature. It is not only more convenient, but also far more efficient due the the fact that only a single simulation need be performed.

Finally, we want to mention our results for SAMC, which are not shown in our plots in order to reduce clutter. With a well-chosen power of ten value for $t_0$, SAMC can converge to the correct result, but less rapidly than either 1/t-WL or SAD. However, with a poor $t_0$ parameter, the method cannot converge in practice. In addition, the optimal $t_0$ parameter depends on the bin width as well as the energy range chosen, further complicating matters.

V. CONCLUSIONS

We have introduced a new algorithm, a variant of the SAMC method, which effectively samples the energy space corresponding to a desired range of temperatures for a few systems. We find that both SAD and 1/t-WL converge more rapidly than SAMC, and unlike WL consistently converge to the correct density of states. SAD requires the user to specify a temperature range of interest rather than an energy range of interest as 1/t-WL does. For use cases in which a range of desired temperatures is known, this will make the SAD method considerably more convenient.

We find that SAMC converges for a reasonable choice of $t_0$ but this parameter can be difficult to tune especially across significantly differing systems. We find that even simple changes to the Monte Carlo moves can have a dramatic effect on the range of practical $t_0$ values. Additionally, SAMC does not converge as rapidly as either SAD or 1/t-WL even for the best choice of $t_0$, when a relatively small range of energies is required, because it always simulates all possible energies.

Finally, we find that for examining a Lennard-Jones cluster the WL and 1/t-WL algorithms—versions of which had previously been found to be unsuccessful on this problem [50]—have convergence properties that are highly dependent on the choice of energy range to be examined. This is particularly problematic when examining the low-temperature heat capacity, where it is difficult to determine the lowest energy that will have a significant impact at the temperatures of interest. This is an example of how SAD is more powerful when the range of energies is not known in advance.

[1] A. M. Ferrenberg and R. H. Swendsen, Physical review letters 61, 2635 (1988).
[2] B. A. Berg and T. Neuhaus, Physics Letters B 267, 249 (1991).
[3] B. A. Berg and T. Neuhaus, Physical Review Letters 68, 9 (1992).
[4] T. Penna and H. Herrmann, arXiv preprint cond-mat/9610041 (1996).
[5] P. De Oliveira, T. Penna, and H. Herrmann, The European Physical Journal B-Condensed Matter and Complex Systems 1, 205 (1998).
[6] R. H. Swendsen, B. Diggs, J.-S. Wang, S.-T. Li, C. Genovesi, and J. B. Kadane, International Journal of Modern Physics C 10, 1563 (1999).
[7] F. Wang and D. P. Landau, Physical Review E 64, 056101 (2001).
[8] F. Wang and D. P. Landau, Physical review letters 86, 2050 (2001).
[9] D. P. Landau, S.-H. Tsai, and M. Exler, American Journal of Physics 72, 1294 (2004).
[10] B. J. Schulz, K. Binder, M. M"uller, and D. P. Landau, Physical Review E 67, 067102 (2003).
[11] Q. Yan and J. J. de Pablo, Physical review letters 90, 035701 (2003).
[12] S. Trebst, D. A. Huse, and M. Troyer, Physical Review E 70, 046701 (2004).
[13] R. E. Belardinelli and V. D. Pereyra, The Journal of chemical physics 127, 184105 (2007).
[14] R. E. Belardinelli and V. D. Pereyra, Physical Review E 75, 046701 (2007).
[15] R. E. Belardinelli, S. Manzi, and V. D. Pereyra, Physical Review E 78, 067701 (2008).
[16] R. E. Belardinelli, V. D. Pereyra, R. Dickman, and B. J. Lourenço, Journal of Statistical Mechanics: Theory and Experiment 2014, P07007 (2014).
[17] S. Singh, M. Chopra, and J. J. de Pablo, Annual Review of Chemical and Biomolecular Engineering 3, 369 (2012).
[18] C. Zhou, J. Su, et al., Physical Review E 78, 046705 (2008).
[19] S. Schneider, M. Mueller, and W. Janke, Computer Physics Communications 216, 1 (2017).
[20] F. Liang, Journal of statistical physics 122, 511 (2006).
[21] F. Liang, C. Liu, and R. J. Carroll, Journal of the American Statistical Association 102, 305 (2007).
[22] F. Liang et al., The Annals of Statistics 37, 2626 (2009).
[23] B. Werlich, T. Shakirov, M. Taylor, and W. Paul, Com-
Computer Physics Communications **186**, 65 (2015).

[24] J. G. Kim, J. E. Straub, and T. Keyes, Physical review letters **97**, 050601 (2006).

[25] J. G. Kim, J. E. Straub, and T. Keyes, The Journal of chemical physics **126**, 04B601 (2007).

[26] J. G. Kim, T. Keyes, and J. E. Straub, The Journal of chemical physics **130**, 124112 (2009).

[27] C. Junghans, D. Perez, and T. Vogel, Journal of chemical theory and computation **10**, 1843 (2014).

[28] A. Laio and M. Parrinello, Proceedings of the National Academy of Sciences **99**, 12562 (2002).

[29] A. Z. Panagiotopoulos, V. Wong, and M. A. Floriano, Macromolecules **31**, 912 (1998).

[30] A. Z. Panagiotopoulos, Journal of Physics: Condensed Matter **12**, R25 (2000).

[31] J. R. Errington, The Journal of chemical physics **118**, 9915 (2003).

[32] P. Dayal, S. Trebst, S. Wessel, D. Wurtz, M. Troyer, S. Sabhapandit, and S. N. Coppersmith, Physical Review Letters **92**, 097201 (2004).

[33] M. Troyer, S. Wessel, and F. Alet, Physical review letters **90**, 120201 (2003).

[34] M. Fasnacht, R. H. Swendsen, and J. M. Rosenberg, Physical Review E **69**, 056704 (2004).

[35] Y. W. Koh, H. K. Lee, and Y. Okabe, Physical Review E **88**, 053302 (2013).

[36] D. P. Landau and K. Binder, *A guide to Monte Carlo simulations in statistical physics* (Cambridge university press, 2014).

[37] A. Tröster and C. Dellago, Physical Review E **71**, 066705 (2005).

[38] J. K. Singh, D. A. Kofke, and J. R. Errington, The Journal of Chemical Physics **119**, 3405 (2003).

[39] J. A. Barker and D. Henderson, The Journal of Chemical Physics **47**, 2856 (2004).

[40] J. A. Barker and D. Henderson, The Journal of Chemical Physics **47**, 2856 (1967).

[41] L. Vega, E. de Miguel, L. F. Rull, G. Jackson, and I. A. McLure, The Journal of Chemical Physics **96**, 2296 (1992).

[42] J. P. K. Doye and D. J. Wales, Physical review letters **80**, 1357 (1998).

[43] S. Martiniani, J. D. Stevenson, D. J. Wales, and D. Frenkel, Physical Review X **4**, 031034 (2014).

[44] D. J. Wales and J. P. Doye, The Journal of Physical Chemistry A **101**, 5111 (1997).

[45] J. Neirotti, F. Calvo, D. L. Freeman, and J. Doll, The Journal of Chemical Physics **112**, 10340 (2000).

[46] P. A. Frantsuzov and V. A. Mandelshtam, Physical Review E **72**, 037102 (2005).

[47] V. A. Mandelshtam and P. A. Frantsuzov, The Journal of chemical physics **124**, 204511 (2006).

[48] F. Calvo and J. P. K. Doye, Physical Review E **63**, 010902 (2000).

[49] F. Calvo, J. Neirotti, D. L. Freeman, and J. Doll, The Journal of Chemical Physics **112**, 10350 (2000).

[50] P. Poulain, F. Calvo, R. Antoine, M. Broyer, and P. Dugourd, Physical Review E **73**, 056704 (2006).

[51] J. P. K. Doye and D. J. Wales, Zeitschrift für Physik D Atoms, Molecules and Clusters **40**, 466 (1997).

[52] S. C. Hendy and B. D. Hall, Physical Review B **64**, 085425 (2001).

[53] J. Northby, The Journal of chemical physics **87**, 6166 (1987).