

Globule transitions of a single homopolymer: A Wang-Landau Monte Carlo study

Drew F. Parsons* and David R. M. Williams

*Research School of Physical Sciences and Engineering, Australian National University,
Canberra, Australian Capital Territory 0200, Australia*

(Received 1 September 2005; revised manuscript received 21 June 2006; published 6 October 2006)

The temperature-independent Wang-Landau Monte Carlo approach is implemented for an off-lattice model of flexible homopolymers and applied to the coil-globule and solidification transitions based on chain sizes up to $N=300$. An intermediate transformation from low-density liquid globule to high-density liquid globule is suggested. A scheme for identifying polymer structures representative of particular temperatures in the course of the simulation is presented and applied to illustrate intermediate states in the coil-globule transition. Transition temperatures are calculated and used to obtain a theta point of at least $\Theta=1.96$, distinctly higher than the solid-liquid transition temperature $T_M=1.26$.

DOI: [10.1103/PhysRevE.74.041804](https://doi.org/10.1103/PhysRevE.74.041804)

PACS number(s): 36.20.-r, 02.70.Uu, 61.43.Bn, 64.70.-p

I. INTRODUCTION

Monte Carlo techniques have complemented molecular dynamics (MD) simulations for almost as long as computer simulations of molecular or physical systems have been performed. The random manner in which new configurations of a system are generated by Monte Carlo methods leads to two main advantages over MD simulations. First, the potentially small correlations between configurations means there is a greater likelihood that all configurations are sampled, allowing for improved ergodicity of the simulation. Second, and as a consequence of the first, this freedom to explore outlying regions of configuration space means that Monte Carlo methods have the capability of studying behavior across configurations that would require prohibitively long time scales to reach using MD methods.

To an extent, however, this potential of superior ergodicity is diminished in traditional Monte Carlo methods, namely, the Metropolis algorithm, by explicitly tying a simulation to a specific temperature. The Metropolis algorithm accepts or rejects a new configuration in reference to the Boltzmann factor $\exp(-E/kT)$ and may lead to an unwelcome “trapping” of the system in one configuration, where a high-energy intermediate configuration required to reach an alternate low-energy configuration cannot be reached. The “trapping” problem may not necessarily be significant in studies within single-phase regions but can become prominent in studies of phase transitions [1].

An alternate Monte Carlo method has recently been formulated by Wang and Landau [2,3]. This method centers on the computation of the system’s density of states, $g(E)$, which describes the relative degeneracies of the system’s energies and is independent of temperature. In this approach, temperature becomes a secondary quantity used to obtain a canonical probability distribution of the system’s energies specific to that temperature: $p(E,T)=g(E)e^{-E/kT}$. This probability function is used to obtain thermodynamic averages:

$$\langle A(T) \rangle = \frac{\int dE A(E)p(E,T)}{\int dE p(E,T)} = \frac{\int dE A(E)g(E)e^{-E/kT}}{\int dE g(E)e^{-E/kT}}. \quad (1)$$

The Wang-Landau method was originally formulated for discrete systems, where the ensemble average is a sum over discrete energies, but is readily generalized [4] to a continuous energy domain where the average is calculated by an integral over a continuous energy domain, as shown in Eq. (1). Studies using the Wang-Landau approach in this manner have started to appear, including studies on the entropy of Lennard-Jones fluids [4], comparing simple lattice and off-lattice models of polymers [5], and calculating surface tensions of polymer films using lattice models in a grand-canonical ensemble [6]. A hybrid version of the density-of-states approach has been applied to amino-acid sequences which includes temperature in the acceptance criteria [7,8]. The method was recently applied by Rampf, Paul, and Binder, studying the coil-globule and solid-liquid transition temperatures of flexible polymers in the lattice model [9]. Their study concluded that the two temperatures coincided in the thermodynamic limit, so that the liquid globule phase is lost in favor of a direct coil–frozen globule transition. In the off-lattice model described by this paper, we find that the coil-globule transition temperature retains a value distinct from the solid-liquid melting point [10], so that the liquid globule phase is not lost in the thermodynamic limit.

We apply the Wang-Landau approach in order to extract geometrical information about off-lattice polymer systems. In order to demonstrate the applicability of the technique, we report here on its success in reproducing the coil-globule transition of flexible polymers in a solvent. After calculating the microcanonical ensemble average $\bar{m}(E)$ of a range of geometrical measures calculated for each energy reached by the system, we calculate the canonical ensemble average $\langle \bar{m} \rangle(T)$ of that measure over a range of temperatures using Eq. (1). The geometrical measures considered here are the end-to-end distance, the eigenvalues of the radius of gyration tensor (which provide an average radius of the polymer along its three main perpendicular directions), and the radius of gyration (the sum of the eigenvalues). We compare the

*Electronic address: Drew.Parsons@anu.edu.au

relative usefulness of these different measures in the analysis of the coil-globule and solidification transitions. We also use the measures as the basis on which to identify representative samples of the polymer at different temperatures.

II. SIMULATION DETAILS

The approach of the Wang-Landau algorithm has been described elsewhere [2,3]. We summarize its key features here and then describe the specific adaptations we applied to make the algorithm suitable for polymer studies.

The temperature-independent density of states in logarithmic form, $\ln[g(E)]$, is dynamically built up as new configurations are rejected or accepted. Initially $\ln[g(E)]=0$ for all E . A new configuration of the system with energy E_2 is accepted for replacing the old configuration with energy E_1 according to probability $p(E_2 \leftarrow E_1) = \exp\{\ln[g(E_2)] - \ln[g(E_1)]\}$. The density of states (logarithmic form) is updated for a given configuration with energy E by adding a value γ to $\ln[g(E)]$. If the new configuration was accepted, then the function is updated at E_2 , if it was rejected, then $\ln[g(E)]$ is updated again at E_1 . A histogram is kept of all the values of E sampled to date. When this histogram is sufficiently flat, the value of the incrementing factor γ is reduced by half, with γ initially set to 1. The algorithm is repeated until γ has been reduced to a sufficiently small value ($\gamma \leq 10^{-6}$).

A. Adaptations of the Wang-Landau approach

1. Continuous energy range

The Wang-Landau approach was initially formulated for systems with discrete energies. The generalization to a system with a continuous energy domain is relatively straightforward [4,8]. The domain is divided into a set of bins $\{E_i, i=0, \dots, M\}$ with energy interval dE such that a configuration with energy E is registered in bin i if $E_i \geq E < E_{i+1} = E_i + dE$. A typical value for the number of bins is $M=1000$.

In principle the energy domain for a polymer system is unbounded: infinite energies may be reached due to repulsion if two monomers approach too close to one another, or by bond stretching, if the distance between two bonded monomers is stretched beyond the maximum bond length. Such configurations, where $E > E_{\max} = E_M + dE$, are simply rejected and the previous configuration is recounted. The choice of the maximum energy handled relates to the maximum temperature that the calculation can describe, as discussed further in Sec. II G.

The case of the system's lowest energy is a more subtle one. Ideally we would like the range of our energy bins to encompass the system's global energy minimum, such that the smallest energy in the range, E_0 , lies at or below the system's global minimum E_{\min} . However, we find that due to the relatively low degeneracy of the global minimum, the number of samples in the bin corresponding to E_{\min} is noticeably smaller than the samples generated for other energies. This situation is in conflict with the basic requirement of the Wang-Landau approach that flat histograms be gener-

ated. While in principle the calculation may proceed until histogram flatness is achieved, in practice the calculation becomes prohibitively long (many months), as other energies are sampled in preference to the E_{\min} bin. We work around this difficulty by truncating the energy domain, keeping E_0 above the global minimum E_{\min} . The value of E_0 must be selected to accommodate the lowest temperature value required for the study in question. In our studies an energy range of $E \in [-4N, 0)$ corresponded to a temperature range of about $T \in (0.5, 3)$, sufficient for studying the coil-globule and solid-liquid transitions. We believe further theoretical work needs to be done for the Wang-Landau method to address this problem of sampling near the global minimum. Some recent papers [11,12] suggest, for instance, calculating the density of states in more than one variable.

We note that the Wang-Landau approach allows the calculation to be split into energy windows [3,8,13], so the density of states calculated within our truncated energy domain remains valid in its own right. Furthermore in each of the cases we have calculated it is a very smooth function. We suggest therefore that where it is impractical to explicitly perform the calculation down to the global energy minimum, it is reasonable to extrapolate from the calculated curve using a fitted function back towards that global minimum. This strengthens the ability of the Wang-Landau method to describe values at $T \rightarrow 0$ where the probability of meeting the global minimum is the greatest.

As mentioned above, the Wang-Landau approach may in theory be readily split over a number of energy windows, with the final density of states derived by stitching the separate windows together. In practice, we have found one window, with a suitable choice of minimum energy E_0 , energy interval dE , and number of bins M , to be sufficient for our calculations, after allowing for the problem of counting global minimum energy states as discussed in the previous paragraph.

2. Histogram flatness

Achieving a "sufficiently flat" histogram is a central idea in the Wang-Landau approach. The definition used by Landau and colleagues [2,3] for sufficiently flat is that the value of all nonempty bins in the histogram must be within some value (80%, say) of the average value of that histogram. They define [14] the average value as the number of entries in the given histogram divided by the number of bins.

We make three adaptations of this basic definition of histogram flatness. First, we permit the flatness level to be progressively relaxed as the simulation proceeds. The histogram is checked for flatness after 1000 Monte Carlo steps (each Monte Carlo step consists of N individual geometry displacements, where N is the number of monomers in the polymer). If the check for flatness has been made unsuccessfully $1000/N$ times, then the flatness criterion is reduced by 5%. For instance, the histogram for a 100-monomer polymer will be checked ten times (with 1000 Monte Carlo steps between each check) before reducing the flatness criteria. We start the flatness criterion at 85%, and permit it to be reduced to no less than 50%. When flatness is achieved and a new histo-

gram is started for a smaller $\ln[g(E)]$ modification factor, the flatness criterion is again started at 85%.

Second, we follow Shell *et al.* [4] in placing a lower bound on how many times the smallest bin in the histogram needs to be filled before the histogram may be declared flat. Shell *et al.* suggest a minimum value of 20. We have found this level to be prohibitively low, especially when sampling dense collapsed globules at lowest energies, and applied a minimum value of 4000 samples per bin per processor (i.e., 32 000 samples per bin in a calculation using eight processors). The minimum level criterion is not switched on until the flatness criterion has been reduced to its lowest permitted level according to the procedure in the previous paragraph. The justification for this step is that even when sampling has been relatively uneven so that the main flatness criterion is not met, we may still nevertheless say that the full energy range has been sufficiently sampled.

Lastly, perhaps least importantly, we only ignore empty bins in the low-energy region to the left of the lowest known nonempty bin. That is, any empty bins in the higher-energy regions will need to be filled before the histogram may be declared flat. This step is necessary to prevent the algorithm from becoming trapped in a previously unoccupied low-energy bin, where the algorithm will attempt to compensate by filling this bin to the exclusion of the higher-energy bin. In this case, were the empty higher-energy bins to be ignored, the histogram might be prematurely declared flat on the basis of only a small handful of occupied bins. This step of forcing higher bins to be filled assumes that there are no forbidden energy bands at higher energies, which is a reasonable assumption for the classical polymer models studied here. Only the empty region below the global energy minimum is allowed for.

We comment that the efficiency of the Wang-Landau approach compared to other expanded ensemble approaches has been criticized by Åberg *et al.* [13]. However, they have used Wang-Landau's original static definition of histogram. It would be interesting to learn if, in their experience, the Wang-Landau approach continues to exhibit inferior performance to their adaptive expanded ensemble approach after implementing the less rigid, dynamic definition of "flatness" described here.

B. Measures of polymer geometries

We assess the effectiveness of the Wang-Landau approach at characterizing polymer geometries by using it to calculate the end-to-end distance R_e and the radius of gyration R_g . Calculations are made using the ensemble averages of their squares $\langle R_e^2 \rangle$ and $\langle R_g^2 \rangle$, so that $R_e = \sqrt{\langle R_e^2 \rangle}$, $R_g = \sqrt{\langle R_g^2 \rangle}$.

The radius of gyration is calculated via the eigenvalues of the moment of inertia tensor $I_{ab} = (1/N) \sum_{i=1}^N (r_{i,a} - r_{c,b})^2$, $a, b = 1, 2, 3$, where \vec{r}_c is the center of mass of the monomers making up the polymer. If the eigenvalues of this tensor are $\lambda_1, \lambda_2, \lambda_3$, then the radius of gyration is $R_g^2 = \lambda_1 + \lambda_2 + \lambda_3$.

A microcanonical ensemble average of the end-end distance and the inertia tensor eigenvalues is made by averaging over each of the geometries whose energy fits in energy bin E . All geometries are counted in this average, regardless of

whether or not they are counted in the Wang-Landau Monte Carlo acceptance step through which $g(E)$ is constructed. This provides a set of geometrical measures of the polymer as a function of energy: $\langle R_e^2 \rangle(E)$, $\lambda_1(E)$, $\lambda_2(E)$, and $\lambda_3(E)$. The functions of energy are then passed through the density of states in Eq. (1) to provide the average geometric measures of the polymer as a function of temperature.

We calculate the Khalatur geometrical parameters [15,16], which use the inertia tensor eigenvalues to provide a measure of the shape of the polymer. In the current case of a flexible polymer, distinguishing between the largely spherical globule and the rodlike swollen coil is relatively straightforward. The Khalatur parameters will prove (and have proven [16,17]) particularly important in studies of semiflexible since they enable a disk- or toruslike globule to be distinguished from a spherical globule.

The Khalatur parameters are $K_1 = (\lambda_1 + \lambda_2) / (\lambda_2 + \lambda_3)$ and $K_2 = (\lambda_1 + \lambda_3) / (\lambda_2 + \lambda_3)$, where $\lambda_1 < \lambda_2 < \lambda_3$. A perfect sphere is characterized by $K_1 = K_2 = 1$, a perfect rod is characterized by $K_1 = 0, K_2 = 1$. A perfect disk would be given by $K_1 = K_2 = 1/2$, but is not typically found among flexible polymers.

Core density is measured as the number of monomers inside a sphere of radius 2.5σ of the monomer closest to the polymer's center of mass. Some structural insight is provided by measuring the average distance of each monomer to its closest neighbor.

C. Θ point and melting point

A key parameter for the coil-globule transition is the temperature at which the transition occurs. We compare various estimates of this temperature: the point at which the geometrical measures change value at the fastest rate (i.e., at which the derivative $d\langle m \rangle / dT$ is maximum, where m is the particular measure in question), and the corresponding peak in the heat capacity.

The power of the Wang-Landau approach is evident in these estimates of the transition temperature. When finding the maximum derivative, the temperature is treated as a free variable without the need to perform an additional Monte Carlo calculation. The density of states needs only to be calculated once, with the geometric measures $m(E)$ being collected at the same time alongside it.

We estimate the theta point by fitting the coil-globule transition temperatures to the functional form with logarithmic correction [10],

$$T_c(N) = \Theta - \frac{A}{\sqrt{N(\ln N)^{7/11}}}, \quad (2)$$

where T_c is the coil-globule transition temperature at each N and Θ is the theta point, the transition temperature of an infinitely long polymer.

We compare the coil-globule transition temperatures to the solid-liquid transition, determined as the highest peak in C_V . The melting point T_M is then calculated as the solid-liquid transition temperature of an infinitely long polymer, using the function

$$T_c(N) = T_M - \frac{B}{\sqrt[3]{N}}, \quad (3)$$

where T_c here are the solid-liquid transition temperatures determined from the peaks in the heat capacity.

D. Interaction potentials

Flexible polymers are modeled in this paper using a two-part energy potential comprised of a bond potential between neighboring bonded monomers and a two-body interaction potential between non-neighboring monomers. The solvent is modeled implicitly through the value of the constants in the potentials.

For the bond potential we use a standard finitely extensible nonlinear elastic (FENE) model [18]:

$$U_{\text{bond}}(l) = -\frac{K}{2}R^2 \ln\{1 - [(l - l_0)/R]^2\}. \quad (4)$$

Here l is the bond length, the distance between the two neighboring monomers. l_0 is the standard bond length (set to 0.7). K is a spring constant, set to 40, and $R = l_{\text{max}} - l_0$ with $l_{\text{max}} = 1$.

We use a shifted-truncated Lennard-Jones potential for two-body interactions between all monomers (bonded and nonbonded):

$$U_{\text{two body}}(l) = \begin{cases} U_{\text{LJ}}(l) - U_{\text{LJ}}(l_c), & l < l_c, \\ 0, & l \geq l_c, \end{cases} \quad (5)$$

where

$$U_{\text{LJ}}(l) = 4\varepsilon[(\sigma/l)^{12} - (\sigma/l)^6]. \quad (6)$$

Here l is the distance between the two monomers. We use $\varepsilon = 1$. The length parameter σ is arranged such that the Lennard-Jones minimum coincides with the FENE bond minimum, that is, $\sigma = l_0/2^{1/6}$. The cutoff distance l_c is set to 2.5σ [19].

E. List of neighboring monomers

The majority of two-body interactions between two distant monomers are effectively zero. We significantly improve the efficiency of the energy calculation, particularly for long polymers, by applying a link-cell algorithm in order to not count these interactions. We do this by dividing three-space into a grid of cubes of side d . Cubes are identified by an integer triplet $\{ijk\}$, corresponding to the coordinates of the “bottom left” corner of the cube, in units of d . For instance, the cube located at the origin and bounded by the positive x , y , and z axes has grid index $\{000\}$. The cube located one unit above it, moving along the z axis, is $\{001\}$.

We keep a record only of occupied cubes within the grid. Each record, identified by the grid index, contains a reference to the monomers contained within it.

Two-body interactions between monomers are only counted for monomers in neighboring cubes. That is, for a given cube, only the 27 cubes surrounding it (including itself) are considered. To avoid double counting we apply an

ordering to the grid index triplets, such that x -indices are considered smaller than y indices which are smaller than z indices. Under this ordering we have, for instance, $000 < 100 < 010 < 110 < 001 < 011 < 101 < 111$. When we consider two-body interactions between monomers inside two cubes $\{ijk\}$ and $\{lmn\}$, the interactions are only calculated if $\{ijk\} \leq \{lmn\}$ and if $|i-l| \leq 1$, $|j-m| \leq 1$, and $|k-n| \leq 1$ (the condition for the cubes being neighbors).

We align the size of the cubes with the Lennard-Jones cutoff distance [see Eq. (5)], taking $d = l_c = 2.5\sigma$. This ensures that calculated energies are consistent, so that interactions are not neglected and cut off prematurely.

Applying this grid algorithm we reduce the complexity of the energy calculation considerably down from $O(N^2)$.

F. Generation of configurations

We take a twofold approach to generating new configurations. For the first $N-1$ moves (where N is the number of monomers) we perform a local displacement. One monomer is chosen at random and shifted in a random direction by a distance between 0.3 and 0.7 of the standard bond length as used in the bond potential, Eq. (4) (2 length units here).

Then a large-scale pivot move is used for the N th move. A monomer i is chosen at random, where $i \in \{1, \dots, N-1\}$. The entire segment following it (monomers $i+1, \dots, N$) is rotated through a randomly selected angle. If the new configuration thus generated is rejected, a new pivot move is used starting from the old configuration. The algorithm returns to the local displacements only after a pivot move has been accepted.

In both types of moves, the random direction of the local displacements and the random angle of the pivot move are generated in such a way as to ensure a uniform distribution of the directions over the entire unit sphere. That is, we ensure the directions are not biased along the x , y , or z axes [20].

The rationale behind this two-prong approach is to facilitate a steady sampling of configuration space via the local displacements, while ensuring ergodicity, preventing the system from getting trapped in one portion of configuration space, via the pivot moves.

G. Applicable temperature range

In principle the Wang-Landau method is able to describe the limiting temperature behavior as $T \rightarrow 0$ if the energy domain of the density of states includes the global minimum E_{min} . However, as discussed above in Sec. II A 1, the global minimum is in practice difficult to describe accurately. Since we are primarily interested in the solid-liquid and coil-globule transitions, we restrict our energy range to $E/N \in [-4, 0)$.

The temperature range over which we may be confident that the system has been accurately determined is characterized by the probability function $p(E, T) = g(E)e^{-E/kT}$. At finite T this function typically forms a single-peaked distribution with the peak centered around some energy $E_p(T)$ and $p(E, T) \rightarrow 0$ as $E \rightarrow E_{\text{min}}$ or $E \rightarrow \infty$. At the temperature of a first-order transition the distribution would be double

peaked, but the following argument still applies using the larger of the two peaks (if they are the same size then it does not matter which peak is used). In practice we expect any transition temperature to be safely located in the middle of the temperature range, so at the bounding temperatures described here the probability distribution will be single peaked.

The minimum and maximum temperatures are the limits at which $p(E, T)$ forms a complete, peaked curve, such that the values of $p(E, T)$ at low E (at the global minimum) and at high E (at E_{\max} , the highest point for which data are collected) are significantly smaller than the value at the peak. We quantify this by requiring the peak to be 1000 times higher than the end points: $p(E_p, T)/p(E_{\min}, T) < 1000$, $p(E_p, T)/p(E_{\max}, T) < 1000$.

As the temperature is increased, the low- E criterion is easily met, but at a certain temperature we find that the high- E criterion is no longer met. The probability function becomes truncated at E_{\max} rather than reaching 0.

Note that as $T \rightarrow 0$, $p(E, T)$ becomes a δ function centered on the global minimum: $p(E, T=0) = \delta(E - E_{\min})$. It can be sufficient to numerically model this limit with a function forming a maximum at $E = E_{\min}$, and therefore we could argue for relaxing the criteria that $p(E_p, T)/p(E_{\min}, T) < 1000$ as $T \rightarrow 0$. Since we are interested in the behavior at finite T in the region near the transition temperature, corresponding to our energy range $E/N \in [-4, 0)$, we will use the same criteria given in the previous paragraph at high and low E for testing the valid temperature range of the calculated density of states. This is reasonable since we will see the system has effectively reached its low-temperature limiting values at a finite T midway between 0 and the transition temperature.

H. Implementation details

The Wang-Landau algorithm was implemented in C++, whose object-oriented nature facilitates the easy adaptation of the same code to different systems (such as block copolymers or even nonpolymeric systems). Random numbers, matrix eigenvalues, and curve minimization and root finding were obtained with the assistance of the GNU Scientific Library (GSL).

A multiple processor implementation of the algorithm was performed with Message Passing Interface (MPI) library for parallelization, applying the following master-slave architecture. A copy of the current density of states is sent to each slave processor, which independently generates new polymer configurations and applies the Monte Carlo acceptance step, building up the local copy of the density of states. Each processor operates independently for 250 Monte Carlo steps (250N new configurations) before sending the difference between its current density of states and the value it started with in this iteration to the master processor. The master collates each processor's difference and tests if the histogram of sampled states passes the flatness criteria. If convergence is not yet achieved the master sends the new combined value of the density of states to the slaves, starting the next iteration.

Calculations were performed on eight nodes of the APAC SGI Altix cluster situated at the Australian National Univer-

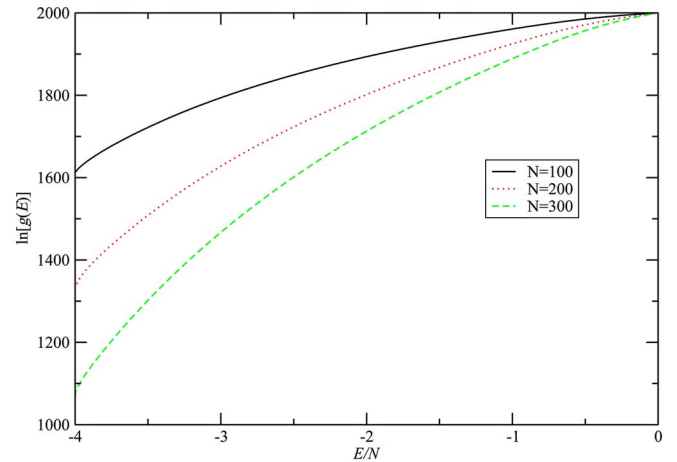


FIG. 1. (Color online) Density of states as a function of energy. The density of states is in logarithmic representation $\ln[g(E)]$ and energy is scaled against polymer size N for comparison. Sample curves are given for $N=100, 200$, and 300 only. The curves here have been shifted to a common $E=0$ value of 2000.

sity. The implementation was validated using a two-particle Lennard-Jones system, producing a density of states identical to the exact analytical result [4].

III. RESULTS FOR THE COIL-GLOBULE TRANSITION

Calculations were performed over a range of N from $N=50$ through to 300. One thousand energy bins were used over the range $E \in [-4N, 0)$. The average number of samples in total was of the order of 10^9 , or 10^6 per energy bin. The Monte Carlo acceptance ratio averaged around 0.2. A number of calculations were performed at each N using different initial geometries in order to reduce random errors in determining transition temperatures due to the starting configuration. The density of states calculated for these flexible homopolymers were smooth and monotonic (see Fig. 1). The density of states are defined up to a multiplicative constant (an additive constant in logarithmic representation); the curves in Fig. 1 have been shifted for the purposes of the diagram so that the $E=0$ value of each curve reaches 1000. We comment that these density of state curves are monotonic increasing, in contrast to those calculated by Rampf, Paul, and Binder for lattice polymer, which exhibit a maximum value in $\ln[g(E)]$ at a certain value of E , beyond which the curve decreases. We interpret the difference as being an artifact of the lattice approach used by Rampf *et al.*, where bond lengths (and therefore bond energies) between neighboring monomers can only be one of a handful of fixed values, whereas our bond lengths vary smoothly over a range of values between nearly 0 and several average bond lengths, depending on the action of the local displacement moves. Bond stretching in particular leads to high-energy samples. Our Lennard-Jones potential between non-neighboring monomers, which is not used in Rampf *et al.*'s model, also gives rise to high-energy states when monomers come in close to one another. Hence we would expect a greater number of higher-energy states in our model, accounting for the

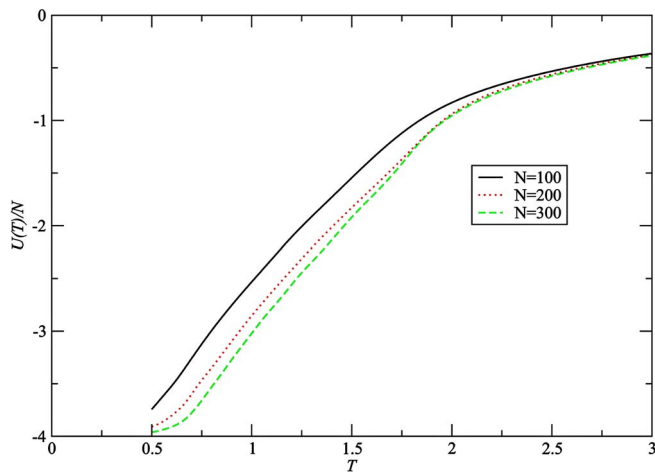


FIG. 2. (Color online) Average energy $U(T)=\langle E \rangle$, scaled against N for $N=100, 200, 300$.

difference between ours and Rampf *et al.*'s density of states. The off-lattice model seems to us to be significantly more realistic than the lattice model in this respect.

The average internal energy $U(T)=\langle E \rangle$ is calculated directly from the density of states and is presented in Fig. 2 for selected N . Its derivative $C_V=dU/dT$, shown in Fig. 3 by the solid line for $N=300$, is the heat capacity of the system at constant volume. We notice three regions in the heat capacity: we interpret the first, a peak at low temperature, to be the solid-liquid transition. The higher-temperature peak in region III corresponds to the coil-globule transition. We argue that the middle region corresponds to a transformation within the liquid globule. The transition temperatures of the two peaks at regions I and III for each N are displayed in Fig. 10 below.

We notice that a certain amount of noise remains in the heat capacity. Recalling that the original density of states $\ln[g(E)]$ follows a shallow curve over the range $E/N \in [-4, 0]$, we are able to smooth the data by fitting it to a

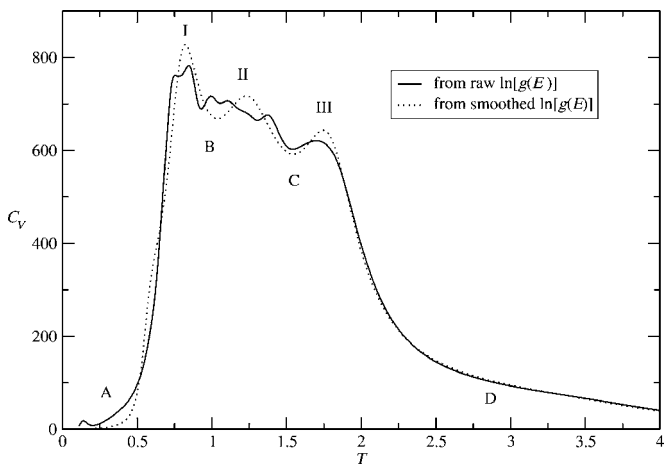


FIG. 3. Heat capacity $C_V=dU/dT$ for $N=300$. The solid curve corresponds to the raw density of states; the dotted curve is derived from the density of states fitted to a tenth-order polynomial. Stable phases are marked as regions A, B, C, and D; transition peaks between them are marked as peaks I, II, and III.

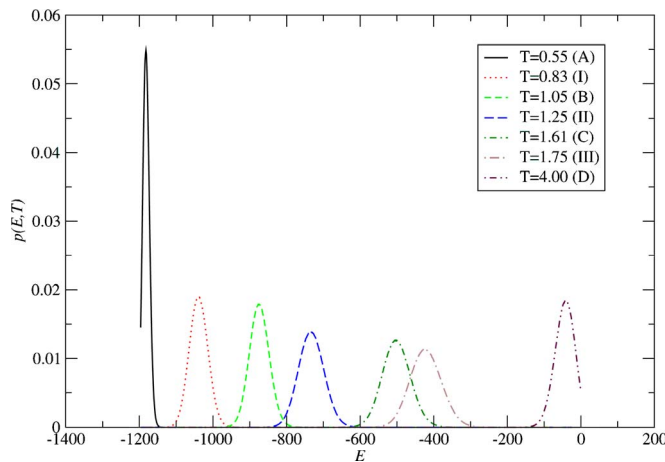


FIG. 4. (Color online) Probability curves $p(E, T)=g(E)e^{-E/KT}$, normalized to a total probability of 1. These curves correspond to $N=150$. Temperature points represent regions A, B, C, and D and transition peaks I, II, and III.

tenth-order polynomial. Since $\ln[g(E)]$ flattens off somewhat across energies higher than zero, this polynomial fitting would not be successful over a larger range, but it appears to be successful over the energy range we have used. The heat capacity calculated from the fitted density of states is shown by the dotted curve in Fig. 3. We see how the excess noise has been eliminated and the two peaks in regions I and III are made more distinct. Region II is reduced by this process to a single peak. An eighth-order or lower fit does not preserve all three peaks, at orders higher than 12 the noise from the raw curve starts to appear. However in order to not introduce biases due to this smoothing step we use the raw unsmoothed density of states in the following analyses.

The probability functions $p(E, T)=g(E)e^{-E/KT}$ for sample values of temperature T are shown in Fig. 4, scaled to a total probability of 1. The curves are shown for $N=300$; other polymer sizes gave similar probability functions. For convenience the probability functions have been calculated at temperatures representing the four phase regions A, B, C, and D and the transitions between them, the peaks in regions I, II,

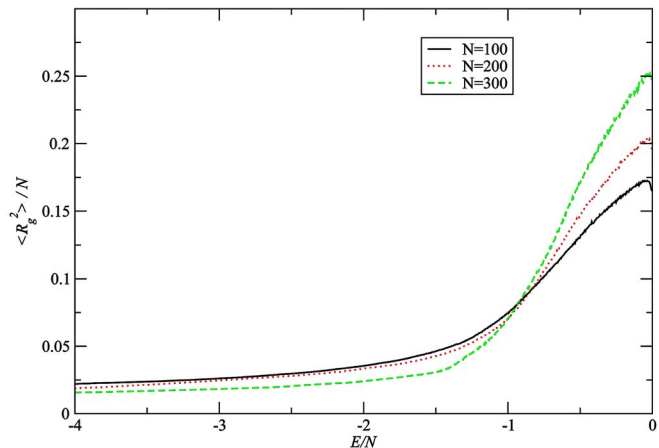


FIG. 5. (Color online) Microcanonical ensemble averages of the radius of gyration squared, shown for $N=100, 200, 300$.

TABLE I. Temperature bounds indicating the range of validity of the densities of states, calculated over the energy range $E/N \in [-4, 0)$.

N	50	75	100	125	150	175	200	225	250	300
Lower T	0.28	0.41	0.45	0.46	0.5	0.54	0.56	0.59	0.6	0.63
Upper T	1.89	2.14	2.35	2.51	2.62	2.71	2.8	2.88	2.94	3.03

and III. The figure illustrates the probability curve transforming into a δ function as $T \rightarrow 0$, and the truncated probability curve at both high and low T . The temperature bounds for each N between which the probability curves are considered untruncated (see Sec. II G) are presented in Table I. The probability functions at the transition temperatures show only one peak, consistent with higher-order rather than first-order transitions.

Polymer properties calculated include the end-end distance and the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of the moment of inertia tensor. We collate these eigenvalues into the radius of gyration, $R_g^2 = \lambda_1 + \lambda_2 + \lambda_3$. The values are collected as microcanonical averages at each value of energy E , and transformed into functions of temperature using Eq. (1). The microcanonical averages for the radius of gyration squared are plotted in Fig. 5, and their transformed values as a function of temperature are shown in Fig. 6. The curves for the end-end distance are similar. The coil-globule transition temperature is well defined in these curves, with the transition approaching first-order, where the derivative with respect to T becomes vertical, as the length of the polymer increases. In the Wang-Landau approach the temperature becomes a free variable, and we are therefore able to calculate derivatives to an effectively arbitrary precision. The derivative for the radius of gyration squared at $N=175$ is also shown in Fig. 6. The coil-globule transition temperature is identified as the point where the derivative is maximal and is presented in Fig. 10 below, calculated from a range of measures along with the heat capacity transition peaks. We notice that the transition temperature from the end-end distance and from the radius of gyration are essentially equal. We cite comments made by Ivanov, Paul, and Binder [17] with regard to the imprecision

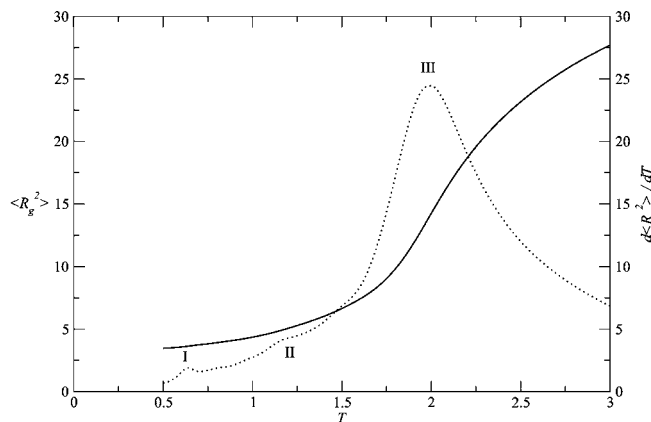


FIG. 6. Radius of gyration squared $R_g^2(T)$ as a function of temperature, shown for $N=175$. The solid line is $R_g^2(T)$, the dotted line is its derivative. Peaks in the derivative corresponding to transitions I, II, and III have been marked.

of the traditional Metropolis Monte Carlo approach in determining the transition temperature, where an estimate must be made from fluctuations $(\langle R_g^4 \rangle - \langle R_g^2 \rangle^2) / \langle R_g^2 \rangle^2$ using a finite number of temperature points. In contrast the ease of the Wang-Landau approach in identifying the transition temperature is remarkable. The radius of gyration or end-end distance predominantly identifies the coil-globule transition, corresponding to peak III in the heat capacity, but peaks I and II can also be located (see Fig. 6), indicating that a slight compactification of the globule, reducing its size by a small amount, occurs at these solid-liquid and liquid-liquid transitions.

The coil-globule transition temperature determined from the Khalatur parameter $K_1 = (\lambda_1 + \lambda_2) / (\lambda_2 + \lambda_3)$ is also shown in Fig. 10. Interestingly, the K_1 transition temperature is very close to the estimate of the coil-globule transition temperature provided by peak III from the heat capacity, while the transition temperatures from the radius of gyration and end-end distance are slightly higher. We would expect peak III to be the best estimate of the transition since it is directly related to the energy of the system. This suggests that K_1 provides a superior measure of the coil-globule transition than the radius of gyration. The Θ points calculated from these alternate measures of the coil-globule transition are presented in Sec. III B.

A sample plot for $N=175$ of the core density and its derivative is shown in Fig. 7. The three transition peaks are clearly evident, indicating a progressive compactification of the globule. The average distance from each monomer to its

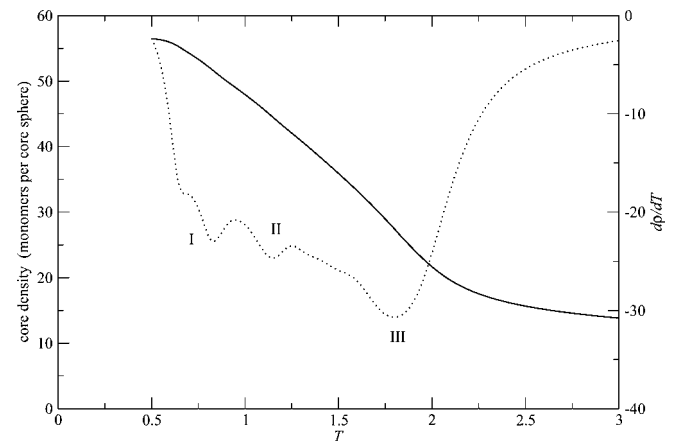


FIG. 7. Core density as a function of temperature, shown for $N=175$. The solid line is the core density, defined as the number of monomers in a sphere of radius 2.5σ from the monomer closest to the center of mass. The dotted line is its derivative with respect to temperature. Peaks in the derivative corresponding to transitions I, II, and III have been marked.

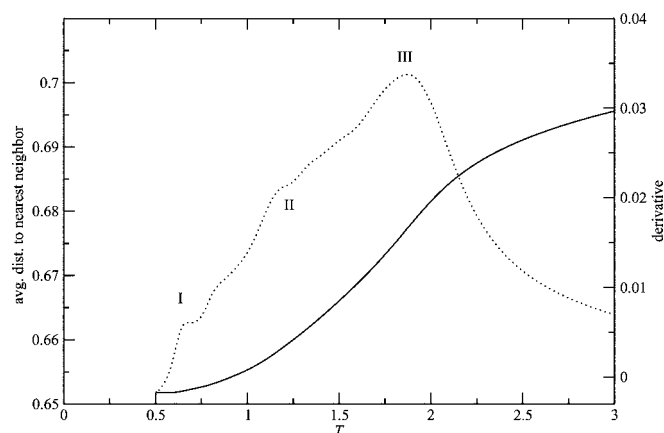


FIG. 8. Average distance from each monomer to its nearest neighbor, as a function of temperature. Shown for $N=175$. The solid line is the average density; the dotted line is its derivative with respect to temperature. Peaks in the derivative corresponding to transitions I, II, and III have been marked.

nearest neighbor is given in Fig. 8, also for $N=175$. In this figure the coil-globule transition is strong, with the other two transitions being less prominent but still visible as shoulders in the derivative.

A. Identifying representative samples

We use the Wang-Landau approach and the microcanonical ensemble averages to identify representative samples at temperature points of interest, filtering them from the thousands of samples taken at each energy. Our approach is to assume an approximate correspondence between the canonical ensemble averages as a function of T and the microcanonical ensemble averages at an energy equal to the canonical ensemble average energy at the given temperature, $E = U(T)$. Samples within or near to this microcanonical energy bin whose individual properties (radius of gyration, etc.) match the microcanonical ensemble average at that energy [$E=U(T)$] are considered representative of that ensemble, and hence of the temperature originally used.

This procedure for identifying representative samples strictly works best when there is approximately a one-one correspondence between temperature, internal energy and geometry. This is reasonable in the case of flexible homopolymers but in other circumstances, for instance the semiflexible case including chain stiffness, there may be competing structures such as discs or torii contributing equally to a given energy state. In this more complex situation we anticipate the use of histograms [16] describing the distributions of polymer properties within each microcanonical ensemble will become indispensable. Similarly at first-order transition temperatures there may be two equally dominant contributing energies [two peaks in the probability curve $p(E, T)$] and samples at both energies would need to be considered.

We implement this idea by calculating the density of states and simultaneously collecting samples on file, up to 10 000 per energy bin. The average value of a range of mea-

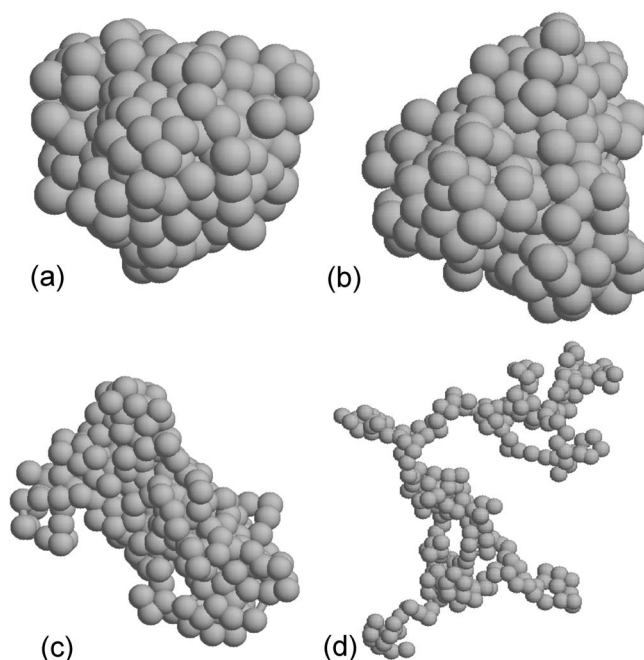


FIG. 9. Representative structures for $N=300$. Structure (a) represents a low-temperature frozen globule (region A, $T=0.6$), structure (b) represents the high-density liquid globule (region B, $T=1.05$), structure (c) represents a low-density liquid globule (region C, $T=1.61$), and structure (d) represents a high-temperature swollen coil (region D, $T=2.5$).

asures at a given temperature is then calculated, including core density, average distance to closest neighbor, radius of gyration squared, end-end distance squared, and K_1 . The sum of the square of the differences between these average values and their value for a particular sample is calculated. The samples with the smallest sum of square differences are accepted as representative samples.

We apply this procedure to identify representative samples for each of the four phases, regions A, B, C, and D, taking $N=300$ as an example. A representative state corresponding to the frozen globule, region A ($T=0.7$, $E=-1130.5$) is shown in Fig. 9(a) (its radius of gyration squared is 5.04 units and core density is 56.0 monomers, close to the microcanonical ensemble average of 4.98 and 55.9, respectively). Representatives of liquid globules from regions B ($T=1.05$, $E=-871.5$, $\langle R_g^2 \rangle = 6.30$, $\langle \rho \rangle = 48.2$) and C ($T=1.61$, $E=-513.6$, $\langle R_g^2 \rangle = 10.30$, $\langle \rho \rangle = 33.0$) are given in Fig. 9(b) ($R_g^2 = 6.34$, $\rho = 48.0$) and Fig. 9(c) ($R_g^2 = 10.70$, $\rho = 33.0$), respectively. The qualitative difference between the two liquid globules is readily seen, with the high-density region B sample appearing more compact than the more diffuse low-density region C sample. Finally, Fig. 9(f) ($R_g^2 = 43.71$, $\rho = 10.0$) shows a high-temperature swollen coil ($T=2.5$, $E=-174.3$, $\langle R_g^2 \rangle = 43.19$, $\langle \rho \rangle = 13.2$).

B. Θ point and melting point: Calculated values

We obtain a number of alternative estimates of the theta point Θ by fitting coil-globule transition temperatures to T_c

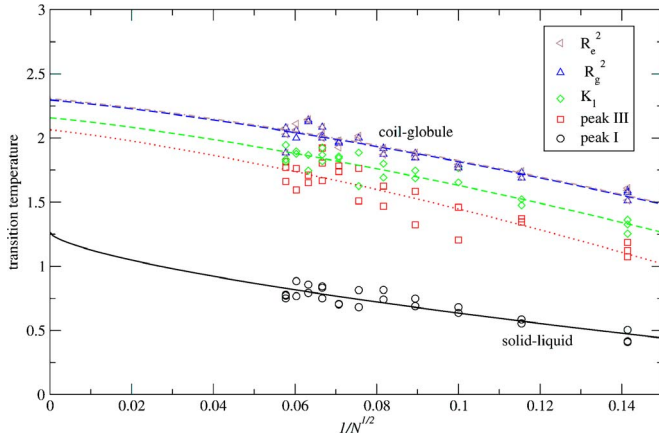


FIG. 10. (Color online) Scaling of transition temperatures and determination of Θ point and melting point. The points in the top four are the coil-globule transition temperatures for each N calculated using the radius of gyration, end-end distance, K_1 parameter, and heat capacity peak III, each shown with a curve of best fit to $T_c = \Theta - A/[\sqrt{N}(\ln N)^{7/11}]$ (see Table II). The lowest curve shows the solid-liquid transition temperatures (peak I) for each N , fitted to $T_{tr} = T_M - B/N^{1/3}$ with $T_M = 1.263$ and $B = 2.91$. The transition curves are plotted against $1/\sqrt{N}$ to illustrate the extrapolated behavior at large N .

$= \Theta - A/[\sqrt{N}(\ln N)^{7/11}]$ [Eq. (2)]. We take the transition temperatures from a range of sources, namely, the radius of gyration R_g^2 , the end-end distance R_e^2 , peak III in the heat capacity, and the Khalatur parameter K_1 . The fits are shown in Fig. 10 and the Θ point estimated given in Table II. Error bars in Θ were estimated from the error matrix generated during the χ^2 minimization used to perform the fit [21], with estimates of the standard deviation in each temperature point $T_c(N)$ derived from the scatter of the repeated simulations at each N .

As mentioned above the transitions from the radius of gyration and end-end distance essentially coincide but are situated at a slightly higher temperature than the K_1 and C_V peak III transitions. The R_g^2 and R_e^2 estimates are based on the temperature at which their derivatives are the greatest, at which the expanded coil is collapsing most rapidly to a globule. We assume that the C_V estimate is the truest estimate of the theta point since it is directly related to the energy of the system. It follows then that the “ideal coil” configuration belonging to the Θ point is located at slightly lower temperature from the temperature at which the expanded coil collapses most rapidly.

TABLE II. Alternate estimates of the Θ point, calculated by fitting coil-globule transition temperatures (see Fig. 10) to $T_c = \Theta - A/[\sqrt{N}(\ln N)^{7/11}]$.

Source	Θ	A
C_V peak III	2.064 ± 0.052	16.3
R_g^2	2.296 ± 0.017	12.7
R_e^2	2.305 ± 0.019	12.7
K_1	2.158 ± 0.034	13.9

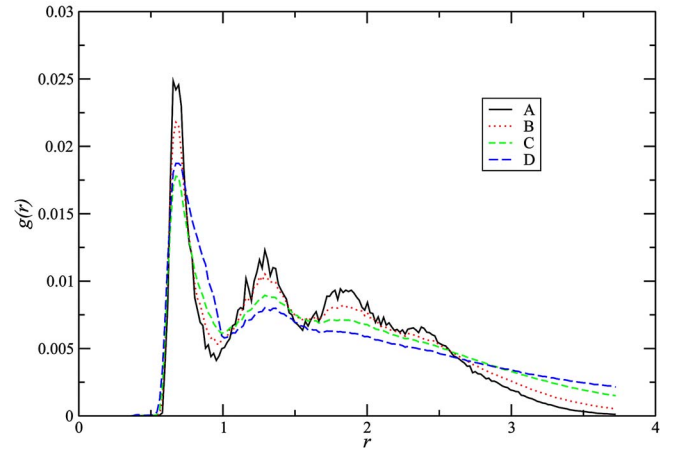


FIG. 11. (Color online) Radial distribution function $g(r)$ centered on monomer closest to center of mass, for $N=300$ in regions A, B, C, and D ($T=0.7, 1.05, 1.61, \text{ and } 2.5$, respectively).

The solid-liquid transition temperatures measured from the first peak (peak I) in C_V are fitted to $T_c = T_M - B/N^{-1/3}$ [Eq. (3)] and shown in the lower curve in Fig. 10. The fitting calculation yields a melting point of $T_M = 1.263 \pm 0.045$ (with $B = 2.91$). Our estimates of the Θ point are significantly higher than the solid-liquid transition temperature, with a ratio $\Theta/T_M = 1.6$ or higher, in contrast to Rampf’s model, where the two temperatures coincided ($\Theta/T_M = 1$). Rampf *et al.* concluded that the coil-globule transition may not be seen in real large-chain polymers, being preempted by the liquid-solid transition. Our results, reported briefly in [10], demonstrate that their conclusion does not generally apply to all flexible homopolymers.

C. Transformation within the liquid globule

If we have correctly identified the heat capacity peak I as the solid-liquid transition and peak III as the coil-globule transition, then it follows that region II must be a liquid-liquid transition inside the globule [10]. The lower-temperature region B corresponds to a high-density liquid globule; the higher-temperature region C is a low-density liquid globule.

The radial distribution functions for the two liquid regions for $N=300$ are shown in Fig. 11 and compared to examples from region A (solid) and region D (expanded coil). The radial distribution functions $g(r)$ [not to be confused with the density of states $g(E)$] were calculated relative to the monomer closest to the polymer’s center of mass. The radial distribution function was first calculated as a microcanonical ensemble average $g(r|E)$ over 10 000 samples for each energy bin E and then converted to $g(r|T)$ at a given temperature by applying Eq. (1) to the values at each point r . The temperatures used were $T=0.7, 1.05, 1.61, \text{ and } 2.5$ for regions A, B, C, and D respectively.

The region D radial distribution function has a strong peak at $r=0.7$ corresponding to bonded neighbors, followed by a relatively smooth slowly decaying form describing the gaslike behavior of a swollen coil, where the probability of

finding a monomer is evenly distributed along the entire reach of the polymer away from its central monomer. Regions B and C have an undulating distribution which we identify with the liquid-droplet structure of a condensed globule. The troughs and peaks in the distribution function for region B are more pronounced than for region C, indicating that the high-density liquid globule in region B is more structured than the low-density liquid globule in region C. The distribution function for region A is similar to that of region B, with slightly deeper troughs, suggesting that the solid globule in region A has been measured in a glasslike state.

IV. CONCLUSIONS

The Wang-Landau Monte Carlo method has been applied to the study of the coil-globule transition of flexible homopolymers using an off-lattice geometry and is particularly facile in determining a precise value for the transition temperature.

We consider the approach useful in identifying structures of particular interest (representative structures below, above, and at the transition point), by enabling geometric criteria to be calculated which match against representative structures.

The approach provides a broad temperature range easily encompassing transition temperatures. The valid temperature

range may be readily defined to a required degree of accuracy through analysis of the probability function $p(E, T) = g(E)e^{-E/kT}$. Care would need to be taken in order to apply the Wang-Landau method to the study of global minima (or equivalently, the limit $T \rightarrow 0$). Further algorithmic development [11,12] in this regard will no doubt prove useful.

We have suggested a means of improving the speed of the algorithm, by declaring the histogram flat when each energy bin has received a sufficient number of samples. This leads to a significant increase in computation speed can be obtained compared to the original flatness criteria used by Wang and Landau [2,3].

For the Hamiltonian used in this paper (FENE bond potential plus truncated Lennard-Jones two-body potential between monomers), the coil-globule transition led to a Θ point significantly higher than the solid-liquid transition temperature (melting point) with $\Theta/T_M = 1.6$ or higher. This contrasts with the Rampf-Paul-Binder model where these two temperatures (calculated in the limit of an infinitely long chain) are the same. We conclude that the liquid-droplet form of the collapsed globule can usually be observed before the frozen globule is obtained.

We also identify an intermediate liquid-liquid transition between the coil-globule and solid-liquid transitions. The lower-temperature high-density liquid globule is more structured than the higher-temperature low-density liquid globule.

-
- [1] N. Wilding and D. P. Landau, *Monte Carlo Methods for Bridging the Timescale Gap*, Lecture Notes in Physics Vol. 605 (Springer-Verlag, Berlin, 2002), Chap. 8, pp. 231–266.
 - [2] F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
 - [3] D. P. Landau, S.-H. Tsai, and M. Exler, *Am. J. Phys.* **72**, 1294 (2004).
 - [4] M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos, *Phys. Rev. E* **66**, 056703 (2002).
 - [5] P. N. Vorontsov-Velyaminov, N. A. Volkov, and A. A. Yurchenko, *J. Phys. A* **37**, 1573 (2004).
 - [6] T. S. Jain and J. J. de Pablo, *J. Chem. Phys.* **118**, 4226 (2003).
 - [7] N. Rathore and J. J. de Pablo, *J. Chem. Phys.* **116**, 7225 (2002).
 - [8] N. Rathore, T. A. Knotts IV, and J. J. de Pablo, *J. Chem. Phys.* **118**, 4285 (2003).
 - [9] F. Rampf, W. Paul, and K. Binder, *Europhys. Lett.* **70**, 628 (2005).
 - [10] D. F. Parsons and D. R. M. Williams, *J. Chem. Phys.* **124**, 221103 (2006).
 - [11] A. Tröster and C. Dellago, *Phys. Rev. E* **71**, 066705 (2005).
 - [12] C. Zhou, T. C. Schulthess, S. Torbrugge, and D. P. Landau, *Phys. Rev. Lett.* **96**, 120201 (2006).
 - [13] K. M. Åberg, A. P. Lyubartsev, S. P. Jacobsson, and A. Laaksonen, *J. Chem. Phys.* **120**, 3770 (2004).
 - [14] D. P. Landau, S.-H. Tsai, and M. Exler, sample Wang-Landau source code published in connection with Ref. [3], ftp://ftp.aip.org/epaps/am_j_phys/E-AJPIAS-72-006406/
 - [15] P. G. Khalatur, *Polym. Sci. U.S.S.R.* **22**, 2438 (1980).
 - [16] V. A. Ivanov, M. R. Stukan, V. V. Vasilevskaya, W. Paul, and K. Binder, *Macromol. Theory Simul.* **9**, 488 (2000).
 - [17] V. A. Ivanov, W. Paul, and K. Binder, *J. Chem. Phys.* **109**, 5659 (1998).
 - [18] A. Milchev, A. Bhattacharya, and K. Binder, *Macromolecules* **34**, 1881 (2001).
 - [19] I. M. Withers, A. V. Dobrynin, M. L. Berkowitz, and M. Rubinstein, *J. Chem. Phys.* **118**, 4721 (2003).
 - [20] E. W. Weisstein, *MathWorld*, <http://mathworld.wolfram.com/SpherePointPicking.html>
 - [21] P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 1969).