The Journal of Chemical Physics

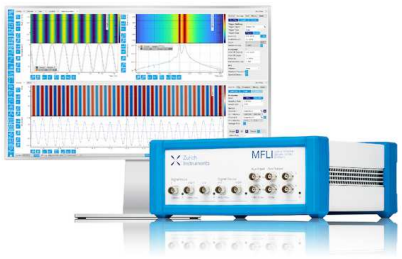
RESEARCH ARTICLE | OCTOBER 18 2024

Methodical evaluation of Boyle temperatures using Mayer sampling Monte Carlo with application to polymers in implicit solvent

Special Collection: Monte Carlo methods, 70 years after Metropolis et al. (1953)

Andrew J. Schultz ; David A. Kofke

J. Chem. Phys. 161, 154108 (2024)
https://doi.org/10.1063/5.0227411



The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Methodical evaluation of Boyle temperatures using Mayer sampling Monte Carlo with application to polymers in implicit solvent

Cite as: J. Chem. Phys. 161, 154108 (2024); doi: 10.1063/5.0227411
Submitted: 9 July 2024 • Accepted: 3 October 2024 •
Published Online: 18 October 2024

Andrew J. Schultz and David A. Kofke[a)]

AFFILIATIONS

Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, New York 14260-4200, USA

Note: This paper is part of the JCP Special Topic on Monte Carlo Methods, 70 years after Metropolis *et al.* (1953).
[a)]**Author to whom correspondence should be addressed:** kofke@buffalo.edu

ABSTRACT

The Boyle temperature, T_B, for an n-segment polymer in solution is the temperature where the second osmotic virial coefficient, A_2, is zero. This characteristic is of interest for its connection to the polymer condensation critical temperature, particularly for $n \to \infty$. T_B can be measured experimentally or computed for a given model macromolecule. For the latter, we present and examine two approaches, both based on the Mayer-sampling Monte Carlo (MSMC) method, to calculate Boyle temperatures as a function of model parameters. In one approach, we use MSMC calculations to search for T_B, as guided by the evaluation of temperature derivatives of A_2. The second approach involves numerical integration of an ordinary differential equation describing how T_B varies with a model parameter, starting from a known T_B. Unlike general MSMC calculations, these adaptations are appealing because they neither invoke a reference for the calculation nor use special averages needed to avoid bias when computing A_2 directly. We demonstrate these methods by computing T_B lines for off-lattice linear Lennard-Jones polymers as a function of chain stiffness, considering chains of length n ranging from 2 to 512 monomers. We additionally perform calculations of single-molecule radius of gyration R_g and determine the temperatures T_θ, where linear scaling of R_g^2 with n is observed, as if the polymers were long random-walk chains. We find that T_θ and T_B seem to differ by 6% in the $n \to \infty$ limit, which is beyond the statistical uncertainties of our computational methodology. However, we cannot rule out systematic error relating to our extrapolation procedure as being the source of this discrepancy.

I. INTRODUCTION

Macromolecules and polymers have long been of scientific and engineering interest due to their technological and biological importance. Of the many ways to characterize polymer structure and behavior, the "θ point" receives considerable attention. The θ point is the temperature at which an n-segment polymer in solution behaves as being ideal in some prescribed sense. This occurs not because the segments are non-interacting, rather because the temperature-weighted attractive and repulsive interactions among the polymer segments and the solvent molecules cancel on average. Two definitions of this temperature are in common use. One matches the scaling of the radius of gyration, R_g, of a single polymer molecule to that of an ideal chain, and in particular, the θ point is the temperature where R_g scales as $n^{1/2}$ (for a linear chain in three dimensions); we will use T_θ to represent this temperature. The other definition is based on the second osmotic virial coefficient, A_2, for a pair of molecules, and it specifically identifies the θ point as the polymer's Boyle temperature, T_B, defined as the temperature where $A_2 = 0$. In either case, at temperatures above the θ point ("good-solvent conditions"), the entropy is able to swell the polymer chains due to favorable polymer–solvent interactions. Below the θ point ("poor-solvent conditions"), the chains contract and may precipitate out of solution. At the θ point, the polymer behaves, by some measure, as if it were in an ideal solvent. The θ point can be determined experimentally through various methods, such as light scattering, osmotic pressure measurements, or viscometry. Understanding the θ point is important in

polymer science for designing and optimizing polymer processing and applications.

While the θ point depends on molecular weight, for $n \to \infty$ it is expected to be asymptotically independent of molecular weight. Moreover, it is hypothesized[1] and observed[2–6] to strongly correlate with the critical temperature, T_c, of the polymer in solution, with T_θ, T_B, and T_c becoming mutually equal asymptotically for $n \to \infty$. Accordingly, the θ point (using either definition, T_θ or T_B) helps in predicting the solubility of polymers, understanding their phase behavior and tailoring materials for specific uses such as drug delivery, coatings, and polymer blends. In addition, a class of biological macromolecules known as intrinsically disordered proteins is believed to exhibit a condensation transition to form membrane-less intracellular structures that are important to biological function, and there is interest in employing calculations of their θ point to understand conditions where this occurs.[5,6]

The Boyle temperature can be studied through numerical calculation of A_2, and for finite n, there is reason to believe that T_B provides a better characterization of T_c than does the single-molecule, R_g-based T_θ, particularly for heteropolymers.[5] Mayer and Mayer[7,8] showed that gas-phase virial coefficients B_N can be expressed as integrals over configurations of $N = 2, 3, 4$ molecules, for the second, third, and fourth virial coefficients, respectively. McMillan and Mayer[9] pointed out that osmotic virial coefficients can be expressed likewise, where A_2 involves integrating configurations of two solutes (macromolecules) in a solvent at fixed chemical potential. However, theory and calculations involving explicit solvent can be very difficult.[10] Sampling of polymer conformations can be slow, and details involving adjustments of the solvent to different polymer conformations are difficult to capture and moreover are largely not needed to understand the solute's behavior. Instead, McMillan–Mayer's treatment adopts models based on an implicit solvent, with the interactions between and among polymer segments interpreted as a solvent-mediated potential of mean force, which, in principle, is temperature-dependent. This greatly simplifies both theoretical and computational methods.

As with many treatments of polymer behavior, lattice models have been employed in many studies reporting calculations of A_2 from a molecular model,[3,11–16] although these studies often do not have a focus on the Boyle temperature. Of those that *do*,[11–14] one significant outcome is an observed $n^{-1/2}$ approach of T_B to its $n \to \infty$ value,

$$T_B(n) - T_B(\infty) = cn^{-1/2}. \tag{1}$$

All the studies found that T_B decreases monotonically with increasing n [$c > 0$ in (1)] for the entire range of n examined, which in one case extended up to $n = 512$.[14]

There have also been several studies of off-lattice chains of hard-sphere beads.[17–22] Our focus is on the Boyle temperature, which requires attractive and repulsive interactions, so we do not examine these results. Instead, we turn briefly to summarizing some relevant results for off-lattice models having intermolecular attraction:

Sheng *et al.*[2] computed both A_2 and the phase behavior of an LJ bead-spring model for $n = 20$, 50, and 100. Their Boyle temperatures varied little with n and were not obtained with precision sufficient to perform an extrapolation. Their three values of T_c varied considerably with n, and followed the Schultz–Flory form,[1,23] allowing for a credible $n \to \infty$ extrapolation. The so-obtained extrapolated T_c value coincided with their n-invariant value of T_B, within uncertainty.

Harismiadis and Szleifer[24] looked at flexible LJ chains of up to $n = 40$ atoms joined by rigid bonds. They too found that Boyle temperatures varied little with n, with data insufficiently precise to allow a meaningful extrapolation to $n \to \infty$.

Wichert and Hall[25] computed A_2 and T_B for chains of square-well beads up to $n = 50$, considering different well ranges extending up to twice the hard-core diameter; notably, intramolecular interactions were modeled as hard spheres (no attraction) in one case, or as square well in another (but only for n up to 16). They found that the slope of the approach of $T_B(n)$ to $T_B(\infty)$ depended on the range of the attractive potential. For short-range potentials, i.e., narrow wells, they found $T_B(n)$ to decrease with increasing n [$c > 0$ in Eq. (1), consistent with most of the lattice-model results], while for longer-range potentials (wider wells), they observed T_B to increase with n ($c < 0$). However, in both cases, the data appear to level off or even go through a maximum with respect to n for large n. Their power-law fit yielded, instead of the usual −0.5, rather large exponents (ranging from −1.23 to −2.79 for different well widths), which resulted from attempting to stretch the plateau into a straight line. The presence or absence of intramolecular attraction was found to have a surprisingly small effect on the Boyle temperatures.

Vega and López Rodríguez[26] examined models that are representative of n-alkanes, with up to 200 united-atom methylene groups represented by LJ sites with fixed bond lengths and employing a three-state rotational isomeric state approximation. They used their calculations of the Boyle temperature to estimate a critical point for the polymethylene melt. Vega and López Rodríguez observed for their data that T_B increases linearly with $n^{-1/2}$ ($c < 0$) for small n, but levels off at $n = 100$. Their comparison of the plateau T_B with short-alkane T_c values from the literature extrapolated to $n \to \infty$ supports the hypothesis that T_B and T_c coincide in this limit.

Rubio and Freire[27] performed calculations of A_2 for LJ beads joined by Gaussian bonds, examining both linear chain molecules, as well as star topologies. In a subsequent work,[28] they evaluated Boyle temperatures for the model, considering chains of length up to $n = 217$. They found that extrapolation to $n \to \infty$ yielded a coincidence of T_B with extrapolated T_θ data taken previously[29] for the same model.

Chiew and Sabesan[30] performed calculations of A_2 for freely jointed LJ chains with rigid bonds, up to $n = 48$ atoms. They found that the Boyle temperature increases with n [$c < 0$ in (1)] and plateaus for $n > 32$.

Withers *et al.*[31] examined a similar system up to $n = 16$, considering both rigid-length bonds and FENE springs for adjacent atoms. They observed the expected $n^{-1/2}$ approach to the long-chain limit, with T_B increasing with n ($c < 0$). However, they found that this n dependence is significantly diminished in going from rigid to weak FENE bonds. Otherwise, they did not observe a plateau with

increasing n, and this could be due to the relatively small values of n examined in their study.

Ida and Yoshizaki[32] examined the effect of torsional flexibility on A_2 and the radius of gyration of freely rotating chains with fixed bond angles, considering both linear and three-arm star topologies. They used an LJ bead model with n of up to 900 and focused on the effect of torsional flexibility on the ratio of A_2 for the two topologies. They did not examine or report Boyle temperatures.

Schultz and Kofke[33] reported values of virial coefficients for normal alkanes up to $n = 20$, in some cases, reporting coefficients up to A_6. Again, these data were presented without an attempt to focus on the Boyle temperature or reporting T_B values specifically.

Mohammadi *et al.*[34] (2012) computed A_2 for chains of atoms interacting with an exp-6 non-bonded potential, up to $n = 48$. They studied the effect of the strength of the core repulsion and the chain length on the Boyle temperature, finding also that $c < 0$.

A recent study by Liu *et al.*[35] applied restricted self-consistent field theory to examine (among other things) the n dependence of T_B. The results of Liu *et al.* again indicated $c < 0$ for the approach to of Boyle temperature to its value at $n \to \infty$.

Finally, we mention the recent work of Mittal and co-workers[5,6] in application to proteins. These are heteropolymers formed from amino acid subunits, and their function is governed by the specific sequence of these residues. Accordingly, they are not amenable to a process of extrapolation of $n \to \infty$. Nevertheless, working with model proteins of 50–150 (coarse-grained) residues, they showed that, while T_B and T_c do not coincide for finite n, they follow a strong linear correlation, sufficient to allow the former to reliably estimate the latter. They also demonstrated that in some systems, a second, lower-temperature zero in A_2 can be used to estimate a lower-critical solution temperature (LCST), in addition to the usual connection to the upper-critical solution temperature (UCST).[6]

Given the usefulness of the Boyle temperature as a means to understand polymer behavior, it is worthwhile to consider methods specifically formulated to compute it for a given molecular model and in particular methods to examine how it changes with various features of the model. The Mayer sampling Monte Carlo (MSMC)[36] technique has appealing features for this purpose. It is an importance-sampling Monte Carlo method that is tailored for the evaluation of virial coefficients and their derivatives. It can be applied to compute virial coefficients of arbitrary order N, and (unlike the most commonly used methods) it requires no tabulation or interpolation of histograms. These features make the approach particularly well suited for determination of the Boyle temperature. The conventional application of MSMC for the calculation of A_2 invokes a reference system of known A_2 (typically the hard-sphere model), and this requirement introduces sampling issues that must be handled using an established method. In this paper, we show that when computing the temperature at which A_2 is specifically zero, we can avoid this complication, opening up possibilities to evaluate a range of Boyle temperatures efficiently. We examine approaches based on a guided, direct search for the Boyle temperature of a given system and another that involves tracing T_B lines of a polymer as some parameter describing or defining it is varied.

The present work aims to demonstrate a proof of concept for evaluating Boyle points and lines with simple application and to uncover strengths and weaknesses in alternative MSMC-based approaches. In Sec. II, we provide a summary of the MSMC method and propose techniques for evaluating Boyle points using it. Then, in the section following that, we introduce the model used for this study, which is given by linear LJ chains with a fixed bond length and variable bond-angle stiffness. We present results from our calculations of T_B for varying stiffness and for different chain lengths. Finally, we close with concluding remarks and suggestions for extensions and applications of the methods.

II. METHODS

We begin this section with a review of MSMC methods for the calculation of A_2 and then turn the focus to two ways that MSMC can be adapted specifically for the calculation of the Boyle temperature, where $A_2 = 0$.

A. Evaluating $A_2(T)$

The key expression for the classical second virial coefficient in terms of a molecular model is[7,8,37,38]

$$A_2 = -\frac{1}{2q^2}\int f(\boldsymbol{r_{12}},\boldsymbol{\omega}_1,\boldsymbol{\omega}_2)e(\boldsymbol{\omega}_1)e(\boldsymbol{\omega}_2)d\boldsymbol{r_{12}}d\boldsymbol{\omega}_1 d\boldsymbol{\omega}_2, \tag{2a}$$

$$\equiv \frac{I}{q^2}, \tag{2b}$$

where f is the Mayer function for the pair of molecules,

$$f(\boldsymbol{r_{12}},\boldsymbol{\omega}_1,\boldsymbol{\omega}_2) = \exp\left(-\beta U(\boldsymbol{r_{12}},\boldsymbol{\omega}_1,\boldsymbol{\omega}_2)\right) - 1, \tag{2c}$$

e is the intramolecular Boltzmann factor,

$$e(\boldsymbol{\omega}_i) \equiv e_i \equiv \exp\left(-\beta u_a(\boldsymbol{\omega}_i)\right), \tag{2d}$$

$q = \int e(\boldsymbol{\omega})d\boldsymbol{\omega}$ is the single-molecule conformational integral, and $\beta \equiv 1/k_B T$ with T being the temperature and k_B being the Boltzmann constant; to simplify some equations presented in the following, we define the integral I via (2b). In (2c), U is the intermolecular energy, obtained by summing over all pairs of atoms with one from each molecule, and in (2d), u_a is the total intramolecular energy, comprising both bonded and non-bonded components. The integrals are taken over $\boldsymbol{\omega}$, which includes all internal conformation coordinates and overall orientations for each molecule, and $\boldsymbol{r_{12}}$ is the vector distance between the centers of mass of the pair.

Methods for computing A_2 have almost always been based upon an integration of the center-of-mass distribution of an interacting pair of molecules, averaged over their orientations and internal conformations. Sometimes, this entails sampling each molecule in its ideal-gas state, governed only by its intramolecular potential u_a and performing quadrature of the Mayer function for a sample of conformations and orientations (e.g., Refs. 24, 26, and 30). Another approach involves the tabulation of histograms of the center-of-mass separations as the molecules sample positions and conformations in a finite box, followed by integration of the resulting distribution (e.g., Refs. 10 and 31); this approach is preferred if working with an explicit solvent. Alternatively, A_2 can be evaluated by regressing simulation data for the osmotic pressure at low concentrations.[2]

In this study, we examine the suitability of MSMC for the evaluation of A_2 of macromolecules. The approach is much like a conventional Monte Carlo simulation of two molecules

(or N molecules for A_N), but sampling is based on the intermolecular Mayer function f, as well as the intramolecular Boltzmann factors e_i. In this manner, sampling is based on the importance to the A_2 integral, allowing both inter- and intra-molecular considerations to influence the molecules' conformations.

MSMC introduces a reference system, and the integral in Eq. (2a) is expressed as an ensemble average with respect to it; in particular,

$$A_2 = \Gamma_0 \left\langle \frac{f}{\gamma_0} \right\rangle_{\gamma_0 e_1 e_2}, \tag{3}$$

where γ_0 is the sampling weight and Γ_0 is A_2 as given by Eq. (2a) but with f replaced by γ_0; the angle brackets in (3) represent an ensemble average, weighted by $\gamma_0 e_1 e_2$, as indicated by the subscript on the brackets. Typically, $\gamma_0 \equiv |f_{ref}|$, where f_{ref} is the Mayer function for a reference model, which is chosen to be simple enough to allow easy evaluation of Γ_0. Almost always, this reference is selected to be a simple hard-sphere (HS) model. Such a reference can provide accurate results for any target system, but precision may become poor if there is great disparity in the size (which is easily adjusted by changing the HS diameter) or shape of the target and reference models. For a chain-molecule target, the HS diameter would typically be selected to approximate the radius of gyration of a polymer molecule. If the target is highly anisotropic (e.g., a long rigid polymer chain), it may become advantageous to select an anisotropic reference, such as hard ellipsoids or spherocylinders.

While Eq. (3) provides a rigorous expression for A_2, in practice, it can provide incorrect results, reflecting a bias due to systematically inadequate sampling of f/γ_0. An alternative formula may be written with the roles of the target (A_2, f) and reference (Γ_0, γ_0) switched, but this too may lead to biased results. An effective remedy is found with overlap sampling,[39,40] which has roots in Bennett's method[41] and is based on ensemble averages collected in both the target and reference systems,

$$A_2 = \Gamma_0 \left\langle \frac{\gamma_{OS}}{\gamma_0} \right\rangle_{\gamma_0 e_1 e_2} \frac{\langle f/|f| \rangle_{|f| e_1 e_2}}{\langle \gamma_{OS}/|f| \rangle_{|f| e_1 e_2}}, \tag{4a}$$

where γ_{OS} is an overlap function, defined as

$$\gamma_{OS} = \frac{\gamma_0 |f|}{\alpha \gamma_0 + |f|}, \tag{4b}$$

and α is a parameter chosen to optimize the calculations. Complete details of the implementation of the overlap sampling form of MSMC are given in Ref. 42.

An appealing feature of the MSMC method is that it neither requires *a priori* determination of a system volume for conducting the calculations nor does it require that any truncation of the potential be applied. Configurations and center-of-mass separations are sampled naturally in proportion to their importance to the average, and large separations do not have to be artificially restricted. Instead of selecting a volume, a reference-system hard-sphere diameter must be chosen, which affects γ_0, γ_{OS}, and sampling of the reference system, but has no effect on sampling of the target, which again occurs in an unrestricted infinite-volume space. In addition, with overlap sampling, the choice of the reference diameter does not affect the accuracy of the computed average, although very inappropriate values can impact efficiency (precision).

B. Tracing T_B lines

It is of interest to understand the effect of major macromolecular features on the location of the Boyle temperature. Such features, which we will label generally as λ, could include molecular weight, architecture (e.g., chains, stars, and brushes), electrostatics, rigidity, block structure, and more. Accordingly, it would be useful to have an efficient means to evaluate T_B across a range of parameter values. One approach of course is to evaluate $A_2(T; \lambda)$ for a range of T and observe where it crosses zero. This may be wasteful if the $A_2 \neq 0$ conditions are not of primary interest. Still, it can be effective to jump from one value of λ to the next and iterate intelligently to locate $A_2 = 0$ for the new λ. We employ this approach in some of our calculations and provide details in the following.

Alternatively, for cases where λ represents a continuous variable, we can formulate an ordinary differential equation (ODE) for $T_B(\lambda)$ (to simplify the equations that follow, we will work with β rather than T itself),

$$\left(\frac{\partial \beta_B}{\partial \lambda} \right)_{A_2} = m(\beta, \lambda), \tag{5}$$

and apply conventional methods for numerical solution of ODEs to trace out the T_B line. This approach is reminiscent of the Gibbs–Duhem integration method for tracing lines of phase coexistence.[43–45] In the present case, the necessary ODE is given by the "−1 rule," which states

$$m(\beta, \lambda) = -\frac{(\partial A_2/\partial \lambda)_\beta}{(\partial A_2/\partial \beta)_\lambda} \equiv -\frac{A_2^\lambda}{A_2^\beta}, \tag{6}$$

where we introduce the notation of using superscripts to represent derivatives. Therefore, if we start at a condition where $A_2 = 0$ (i.e., Boyle temperature), integration of the ODE will, in principle, trace a line of constant $A_2 = 0$ or, equivalently, a line of Boyle temperatures.

The necessary derivatives are obtained via Eq. (2),

$$A_2^\lambda = \frac{I^\lambda}{q^2} - 2A_2 (\ln q)^\lambda, \tag{7a}$$

$$A_2^\beta = \frac{I^\beta}{q^2} - 2A_2 (\ln q)^\beta. \tag{7b}$$

These can be evaluated via MSMC calculations, in terms of the following ensemble averages (writing $u_1 + u_2$ as u):

$$A_2^\lambda = \frac{\beta}{2} \Gamma_0 \left\langle \frac{e^{-\beta U}}{\gamma_0} U^\lambda + \frac{f}{\gamma_0} u^\lambda \right\rangle_{\gamma_0 e_1 e_2} - \beta \Gamma_0 \left\langle u_1^\lambda \right\rangle_{e_1} \left\langle \frac{f}{\gamma_0} \right\rangle_{\gamma_0 e_1 e_2}, \tag{8a}$$

$$A_2^\beta = \frac{1}{2} \Gamma_0 \left\langle \frac{e^{-\beta U}}{\gamma_0} U + \frac{f}{\gamma_0} u \right\rangle_{\gamma_0 e_1 e_2} - \Gamma_0 \langle u_1 \rangle_{e_1} \left\langle \frac{f}{\gamma_0} \right\rangle_{\gamma_0 e_1 e_2}. \tag{8b}$$

The latter term in each expression includes a simple Boltzmann-weighted (e_1) single-molecule average; they multiply A_2, but for

integration along the T_B line, $A_2 = 0$, and in principle, these terms drop out. Still, we retain these contributions because the integration is not precisely on $A_2 = 0$, and moreover, the ODE integration algorithm necessarily samples points off of this line by some amount. Now, crucially, as only the ratio of the A_2 derivatives is required by Eq. (6), Γ_0 cancels and does not need to be evaluated. Accordingly, we are free to select the reference γ_0 to be whatever provides the best sampling, without regard to whether it corresponds to a known reference integral Γ_0. We select $\gamma_0 = |f|$, which should ensure that the sampling covers all important contributions to the averages; hence, overlap sampling is not required.

In summary, we evaluate the T_B-line derivative $m(\beta,\lambda)$ via the ratio of averages,

$$\left(\frac{\partial \beta_B}{\partial \lambda}\right)_{A_2} = -\beta \frac{\left\langle \frac{e^{-\beta U}}{|f|} U^\lambda + \frac{f}{|f|}\left(u^\lambda - 2\langle u_1^\lambda\rangle_{e_1}\right)\right\rangle_{|f|e_1e_2}}{\left\langle \frac{e^{-\beta U}}{|f|} U + \frac{f}{|f|}\left(u - 2\langle u_1\rangle_{e_1}\right)\right\rangle_{|f|e_1e_2}}. \tag{9}$$

The single-molecule averages can be evaluated via Monte Carlo sampling of e_1 before starting calculation of the MSMC averages.

C. Stability

Error in the calculated β_B gives rise to error in the computed slope m, and as the integration proceeds, this error can contribute to further inaccuracy in β_B (instability) or it can attenuate it (stability). When integrating in the direction of increasing λ, stability requires that[46]

$$m^\beta \equiv \left(\frac{\partial m}{\partial \beta}\right)_\lambda < 0. \tag{10}$$

If integration is stable in one direction, it will be unstable in the other. Equation (10) applies at a given β,λ, and the condition may be satisfied at one point and not necessarily at the next. Even if (10) is satisfied, the integration will not necessarily be stable and can also depend on the integration algorithm and step size $\Delta\lambda$ relative to the magnitude of m^β. Conversely, an integration can succeed if (10) is not satisfied if the magnitude of $m^\beta\Delta\lambda$ is not too large.

While evaluation of the stability measure in Eq. (10) is not needed to perform the integration, it is of interest to examine it as we study the integration method. The slope derivative is given in terms of the A_2 derivatives as follows:

$$m^\beta = -\left(\frac{A_2^{\lambda\beta}}{A_2^\beta} + m\frac{A_2^{\beta\beta}}{A_2^\beta}\right), \tag{11}$$

where the double superscript indicates a second derivative. These can be obtained starting from Eq. (7),

$$\frac{A_2^{\lambda\beta}}{A_2^\beta} = \frac{I^{\lambda\beta}}{A_2^\beta q^2} - 2(\ln q)^\lambda + 2m(\ln q)^\beta - 2\frac{A_2}{A_2^\beta}\left(\frac{q^{\lambda\beta}}{q} + (\ln q)^\lambda(\ln q)^\beta\right), \tag{12a}$$

$$\frac{A_2^{\beta\beta}}{A_2^\beta} = \frac{I^{\beta\beta}}{A_2^\beta q^2} - 4(\ln q)^\beta - 2\frac{A_2}{A_2^\beta}\left(\frac{q^{\beta\beta}}{q} + (\ln q)^\beta(\ln q)^\beta\right). \tag{12b}$$

All these terms involve ratios of integrals, so they may be evaluated using MSMC (and single-molecule Boltzmann sampling for the q derivatives) without invoking a reference or imposing $A_2 = 0$. In particular,

$$\frac{I^{\lambda\beta}}{A_2^\beta q^2} = -\frac{1}{\alpha}\left\langle \frac{e^{-\beta U}}{|f|}\left[(\beta u - 1)U^\lambda + \beta U(u^\lambda + U^\lambda)\right] + \frac{f}{|f|}u^\lambda(\beta u - 1)\right\rangle_{|f|e_1e_2}, \tag{13a}$$

$$\frac{I^{\beta\beta}}{A_2^\beta q^2} = -\frac{1}{\alpha}\left\langle \frac{e^{-\beta U}}{|f|}(u+U)^2 - \frac{1}{|f|}u^2\right\rangle_{|f|e_1e_2}, \tag{13b}$$

$$\frac{A_2}{A_2^\beta} = -\frac{1}{\alpha}\left\langle \frac{f}{|f|}\right\rangle_{|f|e_1e_2}, \tag{13c}$$

$$\alpha = \left\langle \frac{e^{-\beta U}}{|f|}U + \frac{f}{|f|}\left(u - 2\langle u_1\rangle_{e_1}\right)\right\rangle_{|f|e_1e_2}, \tag{13d}$$

$$(\ln q)^\beta = -\langle u_1\rangle_{e_1}, \tag{13e}$$

$$(\ln q)^\lambda = -\beta\left\langle u_1^\lambda\right\rangle_{e_1}, \tag{13f}$$

$$\frac{q^{\lambda\beta}}{q} = \left\langle(\beta u_1 - 1)u_1^\lambda\right\rangle_{e_1}, \tag{13g}$$

$$\frac{q^{\beta\beta}}{q} = \left\langle u_1^2\right\rangle_{e_1}. \tag{13h}$$

D. Direct search for Boyle temperature

For a given λ, it is possible to locate the Boyle temperature through a straightforward root-finding procedure, solving for the values of β that yields $A_2(\beta) = 0$. Given that each test for the root involves an MSMC (or other method) calculation of A_2, this might not be initially appealing. However, the specific requirement that A_2 be zero introduces a simplification that puts this approach at a par with the integration method just described. In particular, according to Eq. (3), solving for $A_2 = 0$ is equivalent to solving

$$\left\langle \frac{f}{|f|}\right\rangle_{|f|e_1e_2} = 0. \tag{14}$$

The reference integral Γ_0, which is non-zero, cancels out and need not be evaluated, so we can again choose γ_0 to optimize sampling, selecting it to be $|f|$ as shown here. Again, no overlap sampling or other complications need be introduced.

The search for the Boyle temperature can be guided using temperature derivatives computed via the formulas presented above. A first-order Newton's method prescribes an update to β according to

$$\beta_{B,\text{next}} = \beta_B - A_2/A_2^\beta. \tag{15}$$

We adopt a second-order correction, which can accelerate convergence,

$$\beta_{B,\text{next}} = \beta_B - \left(\frac{A_2^\beta}{A_2^{\beta\beta}} - \sqrt{\left(\frac{A_2^\beta}{A_2^{\beta\beta}}\right)^2 - 2\frac{A_2}{A_2^{\beta\beta}}}\right). \tag{16}$$

Notably, both expressions require only ratios involving A_2 and its derivatives and may be evaluated via direct-sampling MSMC using Eqs. (12) and (13).

The final estimate of β_B is taken from Eq. (15) using quantities computed from the last simulation in the search, so that we can also compute an uncertainty,

$$\sigma(\beta_B) = \sigma\left(\frac{A_2}{A_2^{\beta}}\right), \tag{17}$$

where $\sigma(\cdot)$ indicates the uncertainty of the quantity in parentheses, which should be evaluated with consideration of the correlation between the two terms in the ratio.

III. APPLICATION: BOND-ANGLE FLEXIBILITY

A. Model

We consider an application to off-lattice linear LJ chain polymers and examine the effect of bond-angle flexibility on the Boyle temperature. In particular, atoms on different molecules interact with the LJ potential,

$$u_{LJ}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{18}$$

and the intermolecular energy is $U = \sum_i \sum_j u_{LJ}(r_{ij})$, where i and j each sum over the n atoms of one of the chains. Within each chain, adjacent atoms are connected by rigid bonds and non-bonded atoms (defined as those atoms separated by two or more bonds in the chain) interact with the same LJ potential given in Eq. (18). The flexibility of the chain is governed by a bond-bend potential of the form

$$u_{bend}(\phi) = k_b(1 + \cos\,\phi), \tag{19}$$

where ϕ is the angle formed by each interior atom with its two neighbors. By varying k_b from zero to infinity, we go from the limit of flexible chains, to rigid, rod-like molecules. We note that the "1" in (19) has no effect on configuration sampling and just adds a fixed constant to the total intramolecular energy. In principle, it can affect the precision of averages where it multiples other quantities, but we have not investigated this. We retain it because it was used in another study.[47]

Our study examines chains across a range of lengths, from $n = 2$ (which has no bend movement) to 512, so the results can also be used to examine the effect of chain length on T_B for fixed bond flexibility.

In all that follows, quantities are expressed in units such that $\sigma = 1$ and $\epsilon = 1$ and temperatures are in units of ϵ/k_B.

B. Computation details

We calculated the Boyle temperature for chains of length $n = 2, 4, 8, 16, 32, 64, 128, 256$, and 512 atoms, respectively. We represent the bond-bending parameter k_b of Eq. (19) via a surrogate variable, λ,

$$\lambda(k_b) = \frac{2}{\sqrt{4 + k_b}}. \tag{20}$$

This variable takes values from zero (corresponding to $k_b \to \infty$, rigid chain) to unity ($k_b = 0$, flexible chain). For this choice of λ,

$$u_1^{\lambda} = -\frac{8}{\lambda^3} \sum_{angles,i} 1 + \cos\,\phi_i, \tag{21}$$

$$U^{\lambda} = 0. \tag{22}$$

MSMC calculations are performed for a pair of chains to collect the averages described above. Sampling is weighted by $|f|e_1e_2$. Hence, the chains are internally self-avoiding, but may overlap each other (for which $f = -1$). Single-molecule simulations are performed every time a new β and/or k_b is encountered, to evaluate $\langle u_1 \rangle$ and $\langle u_1^{\lambda} \rangle$ needed by Eq. (13); we also measure the average radius of gyration. These single-molecule calculations are completed before initiating MSMC sampling to provide averages needed for some of the MSMC averages. All the calculations are performed using the molecular-modeling software framework *etomica*.[48,49]

The Metropolis Monte Carlo algorithm is employed,[50,51] and sampling of configurations is accomplished using five Monte Carlo trials, selected at random (the first two are not used for the single-molecule calculations):

1. rigid translation of a molecule;
2. rigid rotation of a molecule;
3. reptation, in which an atom at one end of the chain is placed at the other end, at a position selected uniformly on a sphere one bond length from the last atom there;
4. pivot, which proceeds as follows:
   (a) an atom is selected at random;
   (b) an atom adjacent to it is selected at random;
   (c) a rotation axis through the first atom is selected at random on the unit sphere; and
   (d) the adjacent atom and all atoms beyond it undergo a rigid rotation about the axis by an angle selected at random within a maximum rotation angle that is adjusted in preliminary runs to yield a 50% rate of acceptance of the trial;
5. bond shuffle, in which a 3- to 5-atom segment of the chain is selected at random, and the bond directions joining the atoms within the segment are shuffled among the pairs, reassigning them at random. The result is that the end atoms of the segment are at the same relative locations (because the vector sum of the bonds between them is unchanged), but the interior atoms are moved.

In order to prevent the molecules from drifting away from the origin, the first molecule is never translated, and when performing reptation, pivot and bond-shuffle moves, the first molecule is translated such that its geometric center does not move.

The maximum rotation angle for the pivot moves is different for different parts of the chain for $n \geq 8$, reflecting that it is easier to complete a larger pivot near the end of the chain compared to the center. For $n = 8$, the middle half of the chain takes larger steps than the outer half. For $n > 8$, the first and last $n/8$ atoms ($n/4$ total) in the chain have the largest move, then the next $n/8$ on each side toward the center, then $n/8$ after that, and then the $n/4$ in the middle have the smallest. Each maximum angle is adjusted independently

to achieve at 50% acceptance of pivot moves in the range where it applies.

We run simulations to search for $A_2 = 0$ for six values of λ: 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0. To begin each, we run 10^6-step simulations starting at $T = 4$, where $A_2 < 0$ for all models and increase the temperature to 5, 8, 12, and 16 until $A_2 > 0$. We take the temperature with the smallest A_2 magnitude as an initial guess for the Boyle temperature for the next part of the search. In the second part, we run a simulation at the guess for T_B and the necessary ratios of A_2, A_2^{β}, and $A_2^{\beta\beta}$ are computed. The next guess for the Boyle temperature comes from taking a second-order Taylor series about the current guess and solving for $A_2 = 0$, as given by Eq. (16) with the derivatives given by Eqs. (12) and (13). If the guess for T_B is much too low, the second-order series either does not go through zero or does so at a much higher temperature. If the discriminant is less than zero or if $\beta_{B,next} > 1.4\beta_B$, we use the first-order series instead, Eq. (15). Similarly, if the guess for T_B is much too high, the new guess would be much too low. If $\beta_{B,next} < \beta_B/1.4$, then we take $\beta_{B,next} = \beta_B/1.4$. When we use the second-order series for the next guess, we also increase the number of simulation steps fourfold. In other cases, we maintain the number of steps until the search is close enough to use the second-order series. The search ends once a 1024×10^6-step simulation is performed and thus typically examines six values of β before concluding. The final estimate for β_B is obtained by the first-order series, and the uncertainty in the estimate is as given by Eq. (17).

We also evaluate the Boyle lines via integration of Eq. (6), which is performed using the fourth-order Runge–Kutta method (RK4). We conduct simulations starting from $\lambda = 0.2$ ($k_b = 96$) at the Boyle temperature given by the direct-search calculations described in the previous paragraph and continue to $\lambda = 1$ ($k_b = 0$) in steps of $\Delta\lambda = 0.1$. Each simulation runs for 10^7 samples after 10^6 samples of equilibration. Integration in the reverse direction is also performed, starting at $\lambda = 1$ and proceeding toward $\lambda = 0$ in steps of -0.1. Each step in the RK4 method evaluates $\beta_B(\lambda)$ via a weighted sum of $m(\beta, \lambda)$ evaluated at four values of (β, λ). The stochastic error in the resulting value of β_B is obtained by summing in quadrature the errors in the four $m(\beta, \lambda)$ values. This error estimate does not include any contribution associated with the approximate nature of the finite-step RK4 solution of the ODE.

Uncertainties in the ensemble averages are computed as one standard deviation of the mean (68% confidence level) of 100 block averages for the single-molecule calculations and 1000 block averages for the two-molecule MSMC calculations. Correlation between adjacent block averages is monitored to ensure that the blocks are independent (less than 10% correlation).

From studies of single-molecule block correlations, we determined that pivot moves are not sufficiently effective at generating

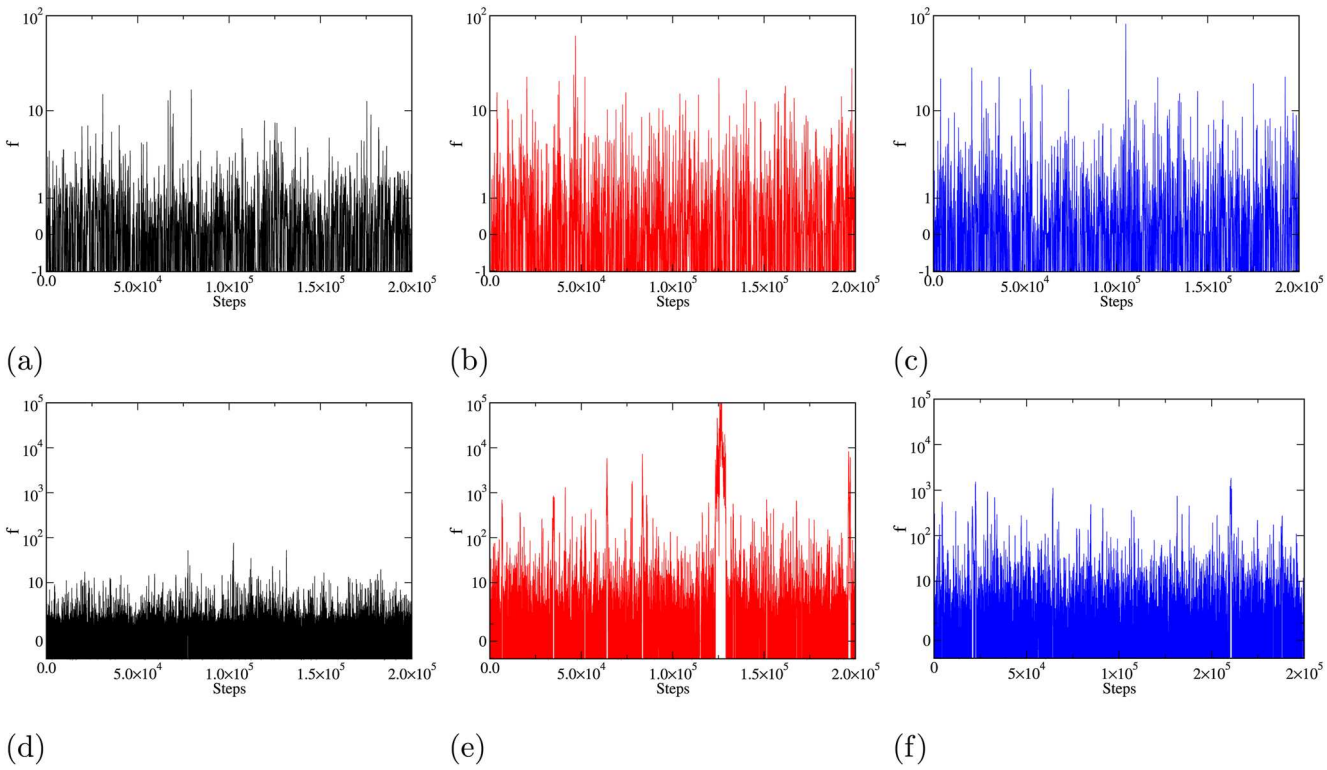


FIG. 1. Trace of Mayer function f for the representative segment of sampling. For each figure, the following are the number of atoms in each chain, n, and bond rigidity parameter, λ (and in parentheses, k_b): (a) 16, 0.2 (96). (b) 16, 0.6 (7.1). (c) 16, 1.0 (0). (d) 128, 0.2 (96). (e) 128, 0.6 (7.1). (f) 128, 1.0 (0). The ordinate uses an arcsinh scale, which for large magnitudes is similar to a log scale while accommodating negative numbers.

independent configurations by themselves, most likely because they have difficulty when the pivot atom is in the central region of the chain. We introduced reptation moves, and this has a significantly favorable effect on the block correlation, but again does not directly affect the mid-chain. We introduced the bond-shuffle move to address this concern and obtained good acceptance rates when it was limited to involving a random four-atom segment.

C. Results and discussion

1. f trace

In the Mayer-sampling method (and unlike Boltzmann sampling), configurations in which the two molecules overlap each other are not prohibited (overlap being where one or more atoms in one chain substantially overlap one or more atoms of the other) and in fact are expected to occur frequently. The second virial coefficient involves the average of $f/|f|$, which equals +1 when there is no overlap and −1 when there is overlap. For $A_2 = 0$, these contributions average to zero, hence overlap and non-overlap configurations should occur with equal likelihood. This can be helpful for sampling, as the molecules can get into and out of tangled configurations without having to untangle in a physically realistic manner or completely separating via a molecule-translation move.

While equal samples of overlap and non-overlap should occur on average, it is of interest to determine whether configurations easily transition from an overlap to non-overlap condition and vice versa. To this end, we can examine plots of f as it fluctuates through a short segment of the simulation. Such plots are shown in Fig. 1. The figures show that the configurations frequently transition back and forth between overlap ($f < 0$) and non-overlap ($f > 0$) arrangements. For longer chains, larger positive values of f are encountered, resulting from favorable arrangements of many atoms at once; such arrangements are favored by stiffer bond angles. Still, even though the weight of the configurations is given by $|f|$, these highly favorable arrangements do not persist unduly long and do not hinder sampling overall. It should be noted that large-f configurations do not contribute to the averages proportionate to their magnitude, as all the averages involve the ratio of f or $f + 1$ to its magnitude $|f|$.

2. Configuration snapshots

Some representative images of configurations sampled during the MSMC sampling are shown in Fig. 2. Examples are presented for a relatively short chain ($n = 16$) and the longest one ($n = 128$), each for three values of k_b. Each configuration is a snapshot for the calculation at the respective Boyle temperature. The configurations show how the persistence length of the chains increases with bond rigidity and also demonstrate that configurations with chain–chain overlap are encountered in the sampling.

3. Boyle-temperature lines

The results for the Boyle points and lines are shown in Fig. 3 and presented in Tables I–IV. Data are obtained from (1) direct search for the solution of Eq. (3), (2) from integration of Eq. (9) in the direction of increasing λ starting from the direct-solution result at $\lambda = 0.2$, and (3) in the direction of decreasing λ starting from $\lambda = 1.0$. Data from the three sources are largely consistent with each

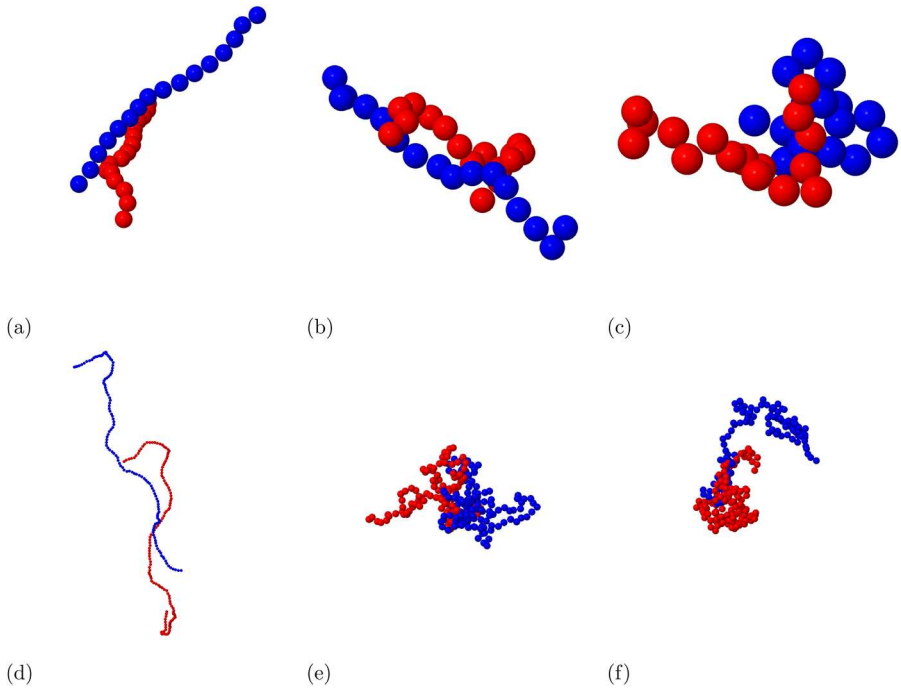


FIG. 2. Snapshots of configurations from MSMC calculations. The values listed with each figure are the number of atoms in each chain, n, and bond rigidity parameter, λ (and in parentheses, k_b). (a) 16, 0.2 (96). (b) 16, 0.6 (7.1). (c) 16, 1.0 (0). (d) 128, 0.2 (96). (e) 128, 0.6 (7.1). (f) 128, 1.0 (0).

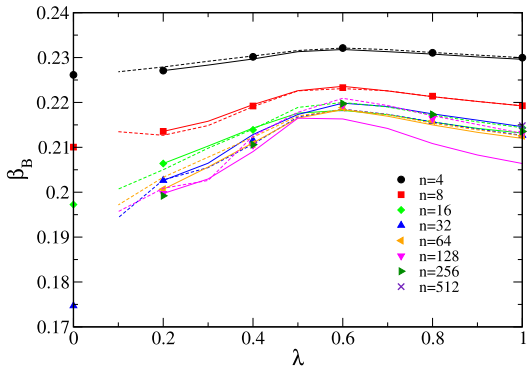


FIG. 3. Inverse Boyle temperature (β_B) as a function of bond-angle stiffness (rigid on left of scale, flexible on the right) for chain lengths in powers of 2–512. The symbols are from direct solution, solving for $\langle f/|f|\rangle_{|f|e_1e_2} = 0$. The solid line is from the integration of Eq. (9) in the direction of increasing λ, and the dotted line is from the integration in the reverse direction. The points (not visible) are placed at the end of each integration step ($\Delta\lambda = 0.1$), and the lines simply join the points. Uncertainties on all displayed data are smaller than the symbol size. The values for $n \geq 64$ at $\lambda = 0$ are off the bottom scale of the figure.

TABLE I. Boyle temperatures (in Lennard-Jones units) as computed using direct search, forward integration (increasing λ), and backward integration (decreasing λ) for different chain lengths n (2, 4, and 8) and bond-angle rigidity parameter λ [Eq. (20)]. Forward integration starts using the search value at $\lambda = 0.2$, and backward integration starts from the search value at $\lambda = 1$. The numbers in parentheses indicate the uncertainty (68% confidence) in the last digit(s) of the tabulated values.

n	λ	Search T_B	Forward integration T_B	Backward integration T_B
2	1.00	3.9808(10)		
4	0.00	4.4221(7)		
4	0.10			4.409(19)
4	0.20	4.4035(11)	4.4035(11)	4.388(13)
4	0.30		4.380(9)	4.364(9)
4	0.40	4.3448(11)	4.354(11)	4.341(7)
4	0.50		4.323(11)	4.318(5)
4	0.60	4.3082(10)	4.314(12)	4.308(3)
4	0.70		4.323(12)	4.315(2)
4	0.80	4.3279(10)	4.333(12)	4.3265(16)
4	0.90		4.345(12)	4.3375(13)
4	1.00	4.3483(11)	4.355(12)	4.3483(11)
8	0.00	4.7609(6)		
8	0.10			4.68(3)
8	0.20	4.6829(9)	4.6829(9)	4.70(2)
8	0.30		4.634(13)	4.654(14)
8	0.40	4.5620(9)	4.556(16)	4.565(10)
8	0.50		4.492(17)	4.493(7)
8	0.60	4.4783(8)	4.472(18)	4.483(5)
8	0.70		4.493(18)	4.493(3)
8	0.80	4.5168(9)	4.519(19)	4.519(2)
8	0.90		4.541(19)	4.5408(14)
8	1.00	4.5608(9)	4.561(19)	4.5608(9)

TABLE II. Same as presented in Table I but for $n = 16$ and 32.

n	λ	Search T_B	Forward integration T_B	Backward integration T_B
16	0.00	5.0697(6)		
16	0.10			4.98(5)
16	0.20	4.8451(9)	4.8451(9)	4.88(3)
16	0.30		4.756(18)	4.77(2)
16	0.40	4.6755(7)	4.67(2)	4.672(14)
16	0.50		4.60(2)	4.568(10)
16	0.60	4.5541(7)	4.58(2)	4.546(6)
16	0.70		4.60(3)	4.566(4)
16	0.80	4.6061(7)	4.64(3)	4.604(3)
16	0.90		4.67(3)	4.6367(18)
16	1.00	4.6648(8)	4.69(3)	4.6648(8)
32	0.00	5.7244(15)		
32	0.10			5.15(7)
32	0.20	4.9351(8)	4.9351(8)	4.93(4)
32	0.30		4.84(3)	4.87(3)
32	0.40	4.7266(7)	4.70(3)	4.75(2)
32	0.50		4.60(3)	4.613(14)
32	0.60	4.5745(7)	4.55(3)	4.577(9)
32	0.70		4.56(3)	4.601(6)
32	0.80	4.6347(7)	4.60(3)	4.637(4)
32	0.90		4.63(3)	4.671(2)
32	1.00	4.7020(7)	4.66(4)	4.7020(7)

TABLE III. Same as presented in Table I but for $n = 64$ and 128.

n	λ	Search T_B	Forward integration T_B	Backward integration T_B
64	0.00	8.331(5)		
64	0.10			5.07(10)
64	0.20	4.9840(8)	4.9840(8)	4.92(6)
64	0.30		4.86(4)	4.81(4)
64	0.40	4.7455(7)	4.74(4)	4.71(3)
64	0.50		4.62(4)	4.61(2)
64	0.60	4.5694(7)	4.58(5)	4.578(15)
64	0.70		4.61(5)	4.602(10)
64	0.80	4.6320(8)	4.65(5)	4.643(6)
64	0.90		4.69(5)	4.675(4)
64	1.00	4.7043(8)	4.72(5)	4.7043(8)
128	0.00	14.11(2)		
128	0.10			5.11(16)
128	0.20	5.0072(8)	5.0072(8)	4.98(9)
128	0.30		4.93(5)	4.94(7)
128	0.40	4.7498(7)	4.78(6)	4.71(5)
128	0.50		4.62(6)	4.59(4)
128	0.60	4.5577(10)	4.62(7)	4.53(3)
128	0.70		4.67(8)	4.560(17)
128	0.80	4.6209(10)	4.74(8)	4.608(11)
128	0.90		4.80(8)	4.649(6)
128	1.00	4.6923(11)	4.85(8)	4.6923(11)

TABLE IV. Same as presented in Table I but for n = 256 and 512.

n	λ	Search T_B
256	0.20	5.0199(15)
256	0.40	4.7485(17)
256	0.60	4.551(4)
256	0.80	4.602(3)
256	1.00	4.683(3)
512	1.00	4.661(10)

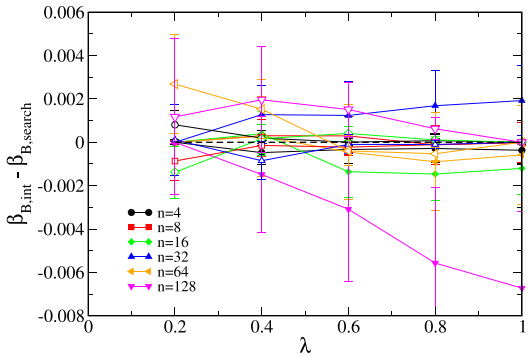


FIG. 4. Boyle temperature (given as its reciprocal, β_B) as a function of bond-angle rigidity, as computed using the integration method and presented as a difference from the direct solution results. Filled symbols: forward integration direction (increasing λ, less stiffness); open symbols: reverse integration direction. The lines join the points as a guide to the eye. Uncertainties represent contributions from both the direct-search calculation and the integration, and are at the 68% confidence level.

other, agreeing within their statistical uncertainties. The conclusion is shown more clearly in Fig. 4, which presents the integration data as a difference from the direct-search values. We first note that the scale of the differences and the uncertainties are both small, relative to the Boyle temperatures themselves—about 0.1%. Most of the uncertainty is ascribable to the integration, as the uncertainties in the direct-search values (which employed 100 times more sampling) are about 0.05%. There is no persistent systematic discrepancy, although we do note possible positive bias in the difference, independent of the integration direction. However, the integration errors do not accumulate, showing no consistent growth with distance along the integration; any inherent instability in the integration has not manifested itself. The agreement of these largely independent calculations suggests that they are accurate.

Lattice-model calculations for short chains ($n \le 32$) by Floriano *et al.*[52] found that chain stiffness pushes the critical temperature monotonically upward relative to the case of more flexible chains. Our calculations find that increasing stiffness (right to left in Fig. 3) at first produces a decrease in T_B (increase in β_B) and then after going through a minimum, T_B increases with stiffness, corresponding to the Floriano *et al.*[52] behavior for T_c.

4. Asymptotic behavior for large n

The Boyle temperature is shown as a function of chain length in Fig. 5. In all cases, T_B increases with n [$c < 0$ in Eq. (1)] for

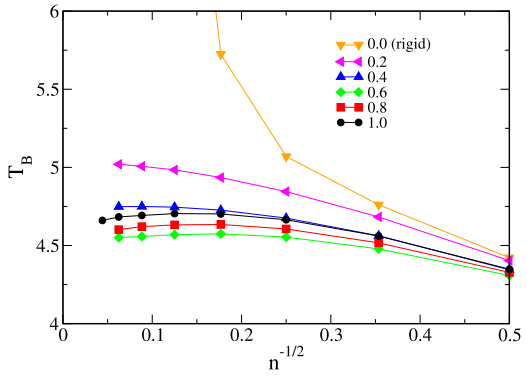


FIG. 5. Boyle temperature as a function of chain length, presented as $1/n^{1/2}$; hence, the y axis represents $n \to \infty$. The legend indicates the value of λ. Uncertainties are smaller than the symbol sizes.

small n and then appears to level off at larger n, as seen in other studies reviewed above. However, the behavior does not plateau, but instead exhibits a shallow maximum, with T_B beginning to decrease for $n > 64$ for $\lambda = 1$. The scale of the decrease—of order 0.04 in LJ units—may not have been discernible in prior off-lattice studies, which had uncertainties in T_B often larger than this. This $c > 0$ behavior was seen previously only for short-range potentials,[25] which suggests that it is the range of the potential relative to the size of the polymer that is relevant to determination of the slope. The existence of a shallow maximum in T_B with increasing molecular weight has been observed experimentally, for example, with polystyrene in decalin.[53]

The rigid-bond case $\lambda = 0$ is qualitatively different. T_B increases rapidly with n for large n, as strongly attractive interactions are easily accessed by the pair of straight chains, requiring a large increase in T to maintain $A_2 = 0$.

5. Stability

The stability measure m^β defined by Eq. (10) is shown in Fig. 6 as a function of λ for each n. It is mostly positive, indicating that the backward integration direction is the stable one, but its magnitude is small, suggesting that stability in either direction should not be

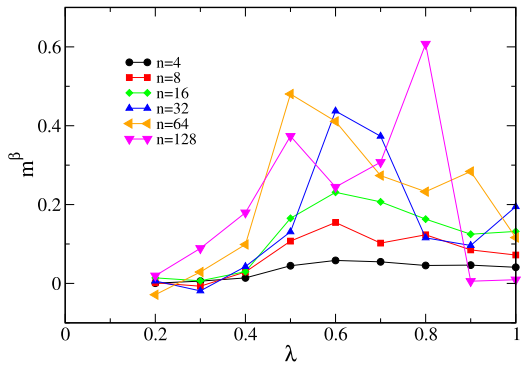


FIG. 6. T_B-line integration stability measure, m^β as computed according to Eq. (10) for the chain-length (n) values examined in this paper.

a significant concern. In fact, the key quantity is the product $m^{\beta}\Delta\lambda$, where $\Delta\lambda = \pm 0.1$ is the integration step size. This product should not be positive, but its magnitude is relevant as well. For example, RK4 is unstable if this product is less than about −3. The small magnitude of this product is consistent with the observation that neither integration direction shows any stability issues. This outcome is specific to the model and integration path; hence, in other applications, one may find that stability matters more.

6. Radius of gyration

A freely jointed ideal (non-self-avoiding) chain will exhibit a squared radius of gyration R_g^2 that scales linearly with the number of monomers, n. The idea of the θ temperature is that the steric repulsion that tends to increase R_g is perfectly offset by the temperature- and solvent-modulated attraction that tends to decrease it. However, this connection cannot be exact, as the Boyle temperature computed here is based on interactions of two molecules, while R_g is computed for a single molecule. Still, there is a balance of repulsion and attraction at play in both situations, so a connection can be expected.

We performed single-molecule simulations to compute R_g^2 for various values of n, λ, and T, to observe the scaling behavior overall and to evaluate T_θ in particular, for comparison with T_B. Scaling is defined by the form

$$R_g^2(n;T) = an^b, \tag{23}$$

such that an ideal chain has $b = 1$. We performed a simple check of our sampling and averaging methods by computing R_g^2 for freely jointed (ideal) chains of varying n, and observe that the expected scaling, $R_g^2 \sim n$ ($b = 1$), is obeyed. For non-ideal chains, the exponent b is a function of n and T, and the points where $b = 1$ locally define $T_\theta(n)$.

First, Fig. 7 shows the results for R_g^2 as a function of chain length, for various λ, each evaluated at their respective Boyle temperatures. Presented on a log–log scale, the ideal-chain behavior is given by a line of unit slope. The "linear" line demonstrates this, and we see that the $\lambda = 1$ chain exhibits behavior that is close to it. Differences from ideal scaling can have several sources: (1) the finiteness of n, which is particularly apparent for the small-n results; (2) the temperatures are T_B, and not T_θ, so they are not tuned to mimic the ideal-chain scaling; and (3) T is varying with n in the plot, which is not how scaling is defined. Apart from the $\lambda = 1$ behavior, we note that as the bond angles become increasingly rigid, the chains approach the a rigid-rod limit, where the scaling of R_g^2 is quadratic. The approach to the corresponding reference line can be observed in the figure.

Second, in Fig. 8, we show the scaling behavior for the $\lambda = 1$ system (zero bond-stiffness potential) as a function of chain length n for a set of isotherms from $T = 3$–5. These are single-molecule calculations performed independently of the Boyle-temperature study. We evaluate b for each n according to

$$b\left(\frac{n}{\sqrt{2}}\right) = \frac{1}{\ln 2}\ln\left(\frac{R_g^2(n)}{R_g^2(n/2)}\right). \tag{24}$$

Superimposed on these data is a line tracing (n, T_B) pairs. This allows one to directly examine the scaling of R_g^2 at the Boyle temperature for each chain length. For long chains, it does not appear that b is approaching the ideal-chain value of 1.0, but instead seems

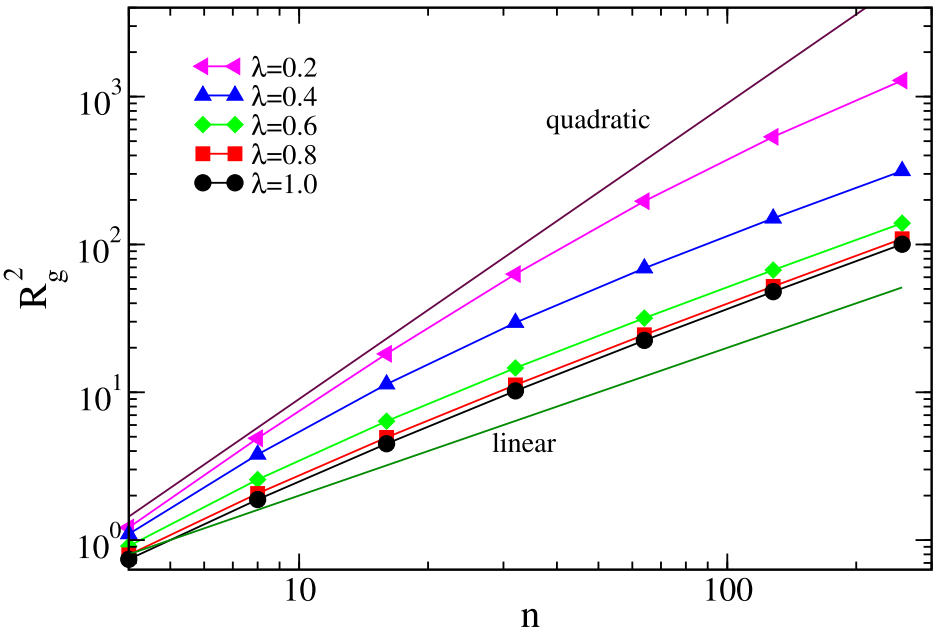


FIG. 7. Squared radius of gyration, R_g^2, as a function of chain length n, evaluated at the Boyle temperature for each (n, λ). The lines for linear (ideal-chain) and quadratic scaling are presented for reference. The legend indicates values of the chain stiffness parameter λ, such that stiffness increases as λ decreases.

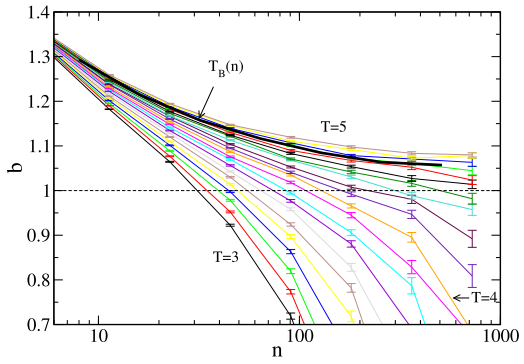


FIG. 8. Scaling exponent, b, for squared radius of gyration, R_g^2, as defined in Eq. (24). All data are for $\lambda = 1$, corresponding to zero bond-stiffness potential. Each narrow line presents b as a function of n for fixed T, ranging from $T = 3$ to $T = 5$ in steps of 0.1. The error bars indicate uncertainties at 68% confidence. The thick black line labeled $T_B(n)$ traces the (n, T_B) pairs, allowing reading of b at the Boyle temperature for each n. The horizontal dashed line at $b = 1$ shows the ideal-scaling exponent, and the value of n at the intersection with each T defines T_θ for that n. These are shown as a function of n in Fig. 9.

to be leveling off at about 1.05, suggesting that T_B and T_θ do not coincide at $n \to \infty$.

For another way to look at this, the intersection of each isotherm of temperature T with $b = 1$ locates the n for which T is the θ temperature, and from this, we can construct $T_\theta(n)$; we compute uncertainties using 100 bootstrap samples of $R_g(n)$ from Gaussians consistent with their uncertainties. These results are shown in Fig. 9, along with $T_B(n)$. Plotted against $n^{-1/2}$, the y-intercept yields the $n \to \infty$ limit for each dataset, and we see that T_B and T_θ again do not appear to coincide in this limit. Instead, we have $T_B(n \to \infty) = 4.64 \pm 0.02$ and $T_\theta(n \to \infty) = 4.916 \pm 0.018$. This disagreement is

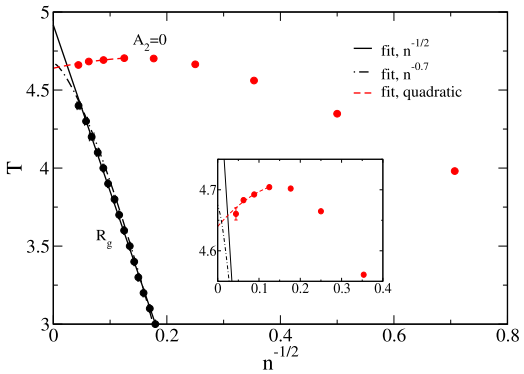


FIG. 9. Boyle and θ temperatures as a function of length of chain, n, for $\lambda = 1$ (zero bond stiffness) system. The black points are based on the definition of R_g, using the construction shown in Fig. 7, and the red points are Boyle temperatures, T_B. The dashed–dotted black line is a fit of T_θ to $a_0 + a_1 n^{-0.7}$, and the dashed red "quadratic" line is fit of T_B to $a_0 + a_1 n^{-0.5} + a_2 n^{-1}$. The inset expands the region near the Boyle-temperature intercept. Where not visible, uncertainties (68% confidence) on both sets of points are smaller than the symbol sizes.

at odds with previous findings, as summarized in the Introduction. This may be due to the increased precision of the data presented here, which is better able to reveal differences between the limit values of T_B and T_θ. Alternatively, we should point out that the extrapolation for T_θ is sensitive to the choice of exponent of n used to extrapolate T_θ. An exponent of −0.7, for example, yields an intercept (4.675 ± 0.005) that is in statistical agreement with the extrapolated T_B without producing a marked disagreement with the data, as shown in Fig. 9. So, while stochastic errors cannot explain the difference in T_B and T_θ at $n \to \infty$, we cannot rule out a systematic error related to the extrapolation procedure. Hence, it is still plausible from these data that T_B and T_θ coincide at $n \to \infty$, although the views shown in Figs. 8 and 9 strongly suggest that they do not. More sophisticated methods for extrapolation[28,54] might be applied to help resolve this question.

A notable observation to take from Fig. 9 is the large difference in sensitivity of T_θ and T_B to chain length n. From about $n = 25$ to ∞, T_B varies by about 2%, whereas over the same range, T_θ covers a range of about 40% of the extrapolated value. From this standpoint alone, the Boyle temperature provides a more practical measure of the infinite-chain behavior and most likely the critical temperature T_c, than does T_θ.

IV. CONCLUDING REMARKS

The Boyle temperature, defined as the point where the second osmotic virial coefficient A_2 is zero, is a key quantity for the understanding of the behavior of a given polymer system. It provides important information about the polymer in solution while requiring the study of only two molecules. Mayer-sampling Monte Carlo methods provide an appealing route to the calculation of A_2 and from this, the Boyle temperature. The method involves importance sampling with respect to configurations contributing most to the value of A_2. Among other outcomes, this allows the molecules to explore configurations where the pair overlaps, which can reduce problems with trapping in the subregions of configuration space due to polymer entanglement. In addition, the method requires no specification of a system volume, instead relying on importance sampling to restrict sampling to relevant configurations within an infinite volume. Related to this, the approach does not require arbitrary truncation of the intermolecular potential.

Postulating a specific interest in the Boyle temperature rather than the entire A_2 dependence on temperature, and considering how T_B changes with features of the polymer molecules, in this paper, we have examined two approaches to methodically evaluate the Boyle temperature with respect to the macromolecule's features or parameters. One approach is direct search for T_B using MSMC to evaluate whether $A_2 = 0$ at each candidate temperature and guiding the search with temperature derivatives of A_2. This is an effective and appealing approach, significantly simplified by the focus on $A_2 = 0$. An alternative that we examine is to integrate a differential equation that describes how T_B varies with the model parameter and particularly bond-angle stiffness. This method requires knowledge of a Boyle temperature for one parameter value and prescribes tracing of the T_B line using an ODE integrator. This calculation does not make particular use of the focus on $A_2 = 0$ and could be applied to trace lines for other constant values of A_2, if desired.

In comparison with established methods, the proposed techniques for evaluating T_B have the primary advantage of simplicity and arguably effectiveness and efficiency. Established methods usually involve evaluation of the radial distribution function. This means an evaluation of histograms or, alternatively, an in-place quadrature, both necessitating selection of bin widths or integration steps and an explicit or implicit truncation of the intermolecular potential. Furthermore, in these methods, importance sampling is not applied to the center-of-mass separations and molecule conformations are weighted only by using the intramolecular Boltzmann factor, which does not account for how the intermolecular interactions affect the conformations. All these issues are handled in MSMC, which in the present application requires evaluation of only a small number of single-molecule and pair ensemble averages. In principle, established methods could be extended to compute temperature and λ derivatives, to implement the methodical T_B search described here, but no such attempts have been reported previously. Apart from these issues of simplicity, one can argue that the MSMC-based method should be more efficient, providing more precise results with less computation. It has advantages due to the central role importance sampling, and the accommodation of configurations with overlaps, which should ease concerns about non-ergodicity. However, we have not attempted to prove such a claim in this study.

The example presented here is given as a proof of concept. There are several ways that one might extend the method to other problems. For example, there is also interest in the temperature at which the third osmotic virial coefficient, A_3 is zero, as it relates to the importance of non-pairwise contributions to the potential of mean force, and it perhaps possesses an even stronger connection to T_c. Extension of the methods here to this quantity is possible, although some complications would be introduced by the need to handle the flexible correction.[15,55] It would also be of interest to be able to trace the T_B lines for varying polymer architectures. The derivative with respect to the addition of another atom could be captured using a variant of Widom insertion and/or deletion, but it is unclear whether such an approach would be effective. Most likely, it would be better to apply the direct-search method, which requires only temperature derivatives, and this approach may be simpler and more robust in general. Finally, extension to mixtures of different polymer molecules would be easy and involve very little modification to the methods presented here.

SUPPLEMENTARY MATERIAL

All single-molecule and MSMC ensemble averages and uncertainties computed as part of this work are tabulated in a machine-readable form and provided in the supplementary material.

ACKNOWLEDGMENTS

The authors thank Jack Douglas of NIST for spurring our interest in this topic and providing a critical reading of the manuscript. This work was performed with the support from the National Science Foundation Grant No. CBET-2152946. D.A.K. acknowledges the support from the Walter E. Schmid Family Foundation. Computing resources were supported by the University at Buffalo Center for Computational Research.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Andrew J. Schultz: Conceptualization (supporting); Formal analysis (lead); Funding acquisition (supporting); Methodology (equal); Software (lead); Visualization (lead); Writing – review & editing (equal). **David A. Kofke**: Conceptualization (lead); Formal analysis (supporting); Funding acquisition (lead); Methodology (equal); Visualization (supporting); Writing – original draft (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material.

REFERENCES

[1] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, 1953).
[2] Y. J. Sheng, A. Z. Panagiotopoulos, S. K. Kumar, and I. Szleifer, Macromolecules **27**, 400 (1994).
[3] A. Rodriguez, J. Freire, and A. Horta, J. Phys. Chem. **96**, 3954 (1992).
[4] A. Z. Panagiotopoulos, V. Wong, and M. A. Floriano, Macromolecules **31**, 912 (1998).
[5] G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, Proc. Natl. Acad. Sci. U. S. A. **115**, 9929 (2018).
[6] G. L. Dignon, W. Zheng, Y. C. Kim, and J. Mittal, ACS Cent. Sci. **5**, 821 (2019).
[7] J. Mayer and M. Mayer, *Statistical Mechanics*, 1st ed. (Wiley, New York, 1940).
[8] J. Mayer and M. Mayer, *Statistical Mechanics*, 2nd ed. (Wiley, New York, 1977).
[9] W. G. McMillan and J. E. Mayer, J. Chem. Phys. **13**, 276 (1945).
[10] D. J. Ashton and N. B. Wilding, J. Chem. Phys. **140**, 244118 (2014).
[11] M. Janssens and A. Bellemans, Macromolecules **9**, 303 (1976).
[12] W. Bruns, Macromolecules **17**, 2826 (1984).
[13] I. Szleifer, E. M. O'Toole, and A. Z. Panagiotopoulos, J. Chem. Phys. **97**, 6802 (1992).
[14] P. Grassberger and R. Hegger, J. Chem. Phys. **102**, 6881 (1995).
[15] S. Caracciolo, B. Mognetti, and A. Pelissetto, J. Chem. Phys. **125**, 094903 (2006).
[16] S. Caracciolo, B. M. Mognetti, and A. Pelissetto, J. Chem. Phys. **128**, 065104 (2008).
[17] A. Yethiraj, K. G. Honnell, and C. K. Hall, Macromolecules **25**, 3979 (1992).
[18] T. Boublík, J. Chem. Phys. **119**, 7512 (2003).
[19] C. Vega, C. McBride, and L. G. MacDowell, Phys. Chem. Chem. Phys. **4**, 853 (2002).
[20] C. Vega, Mol. Phys. **98**, 973 (2000).
[21] L. G. MacDowell and C. Vega, J. Chem. Phys. **109**, 5670 (1998).
[22] C. Vega, J. M. Labaig, L. G. MacDowell, and E. Sanz, J. Chem. Phys. **113**, 10398 (2000).
[23] A. Shultz and P. Flory, J. Am. Chem. Soc. **74**, 4760 (1952).
[24] V. I. Harismiadis and I. Szleifer, Mol. Phys. **81**, 851 (1994).
[25] J. M. Wichert and C. K. Hall, Macromolecules **27**, 2744 (1994).
[26] C. Vega and A. López Rodríguez, J. Chem. Phys. **105**, 4223 (1996).
[27] A. M. Rubio and J. J. Freire, Macromolecules **29**, 6946 (1996).
[28] A. M. Rubio and J. J. Freire, J. Chem. Phys. **106**, 5638 (1997).
[29] C. W. Yong, J. H. R. Clarke, J. J. Freire, and M. Bishop, J. Chem. Phys. **105**, 9666 (1996).

[30]Y. Chiew and V. Sabesan, Fluid Phase Equilib. **155**, 75 (1999).
[31]I. M. Withers, A. V. Dobrynin, M. L. Berkowitz, and M. Rubinstein, J. Chem. Phys. **118**, 4721 (2003).
[32]D. Ida and T. Yoshizaki, Polym. J. **40**, 1074 (2008).
[33]A. J. Schultz and D. A. Kofke, J. Chem. Phys. **133**, 104101 (2010).
[34]A. Mohammadi, A. Ramazani Saadatabadi, and M. Khanpour, Chem. Phys. **397**, 26 (2012).
[35]L. Liu, C. Duan, and R. Wang, Polymer **258**, 125312 (2022).
[36]J. K. Singh and D. A. Kofke, Phys. Rev. Lett. **92**, 220601 (2004).
[37]E. Mason and T. Spurling, *The Virial Equation of State* (Pergamon Press, Oxford, 1969).
[38]J.-P. Hansen and I. McDonald, *Theory of Simple Liquids*, 4th ed. (Academic Press, London, 2013).
[39]K. M. Benjamin, A. J. Schultz, and D. A. Kofke, Ind. Eng. Chem. Res. **45**, 5566 (2006).
[40]K. M. Benjamin, J. K. Singh, A. J. Schultz, and D. A. Kofke, J. Phys. Chem. B **111**, 11463 (2007).
[41]C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[42]A. Bansal, A. J. Schultz, J. F. Douglas, and D. A. Kofke, J. Chem. Phys. **157**, 224801 (2022).
[43]D. A. Kofke, J. Chem. Phys. **98**, 4149 (1993).
[44]D. A. Kofke, Mol. Phys. **78**, 1331 (1993).
[45]D. A. Kofke, Adv. Chem. Phys. **105**, 405 (1999).
[46]D. Sanchez, "Ordinary differential equations and stability theory," in *Dover Books on Mathematics* (Dover Publications, 2019).
[47]F. Vargas-Lara, F. W. Starr, and J. F. Douglas, Soft Matter **13**, 8309 (2017).
[48]A. J. Schultz and D. A. Kofke, J. Comput. Chem. **36**, 573 (2015).
[49]See https://www.etomica.org/ for more information about downloading and using etomica molecular simulation software, 2024.
[50]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[51]D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 3rd ed. (Academic Press, San Diego, 2023).
[52]M. A. Floriano, V. Firetto, and A. Z. Panagiotopoulos, Macromolecules **38**, 2475 (2005).
[53]Y. Nakamura, N. Inoue, T. Norisuye, and A. Teramoto, Macromolecules **30**, 631 (1997).
[54]A. M. Rubio, J. J. Freire, C. W. Yong, and J. H. R. Clarke, J. Chem. Phys. **111**, 1302 (1999).
[55]K. R. S. Shaul, A. J. Schultz, and D. A. Kofke, J. Chem. Phys. **135**, 124101 (2011).