Calculation of the Entropy of random coil polymers with the hypothetical scanning Monte Carlo Method

Ronald P. White and **Hagai Meirovitch**[*]
Department of Computational Biology, University of Pittsburgh School of Medicine, Biomedical Science Tower, 1058w, Pittsburgh PA 15261

Abstract

Hypothetical scanning Monte Carlo (HSMC) is a method for calculating the absolute entropy, *S* and free energy, *F* from a *given* MC trajectory developed recently and applied to liquid argon, TIP3P water and peptides. In this paper HSMC is extended to random coil polymers by applying it to self-avoiding walks on a square lattice – a simple but difficult model due to strong excluded volume interactions. With HSMC the probability of a given chain is obtained as a product of transition probabilities calculated for each bond by MC simulations and a counting formula. This probability is exact in the sense that it is based on all the interactions of the system and the only approximation is due to finite sampling. The method provides rigorous upper and lower bounds for *F*, which can be obtained from a very small sample and even from a *single* chain conformation. HSMC is independent of existing techniques and thus constitutes an independent research tool. The HSMC results are compared to those obtained by other methods, and its application to complex lattice chain models is discussed; we emphasize its ability to treat any type of boundary conditions for which a reference state (with known free energy) might be difficult to define for a thermodynamic integration process. Finally, we stress that the capability of HSMC to extract the absolute entropy from a given sample is important for studying relaxation processes, such as protein folding.

I. Introduction

In spite of progress achieved in the last 50 years, calculation of the entropy and free energy remains a central problem in computer simulation, which affects physics, chemistry, biology, and engineering.[1,2] Recently, we have developed a new technique for calculating the entropy - the hypothetical scanning Monte Carlo (HSMC) method and applied it to liquid argon, water, [3,4] and peptides in helical, extended, and hairpin states.[5] The aim of the present paper (as that of our preliminary study[6]) is to extend HSMC to lattice polymer models, and in particular to examine its applicability to random coil polymers.

It should be pointed out that lattice models have been utilized for studying a wide range of phenomena in polymer physics[7,8] as well as in structural biology, mainly related to protein folding and stability[9] (Refs. 7 and [8] present only very limited lists). Because of their simplicity, these models have been invaluable tools for understanding global properties that do not depend strongly on molecular details. Such models vary in complexity, ranging from self-avoiding walks on a square lattice to chain models on enriched 3D lattices with a large effective coordination number.

Commonly, these systems are simulated by variants of Metropolis Monte Carlo (MC) - a method that enables one to generate samples of chain configurations *i* distributed according to their Boltzmann probability, P_i^B, from which equilibrium information can be extracted.[10] In

---
*Corresponding author (email: hagaim@pitt.edu. phone : 412-648-3338. fax: 412-648-3163.

many cases where the simulation moves are local, MC is referred to as a dynamical method. [It is noted however that the technique does not need to (and often doesn't) map the physical dynamics of the system.] Using MC it is straightforward to calculate properties that are measured directly from i, such as the potential energy E_i (that is obtained by summing up the atom-atom interactions) or geometrical quantities such as the radius of gyration. On the other hand, the *value* of $P_i{}^B$ cannot be obtained in a straightforward manner, which makes it difficult to obtain the *absolute* entropy, $S \sim -\ln P_i{}^B$ *directly*, i.e., as a byproduct of the simulation (like E_i). There is a strong interest in S as a measure of order and as an essential ingredient of the free energy, $F = E - TS$, where T is the absolute temperature; F constitutes the criterion of stability, which is mandatory in structure determination of proteins, for example. Furthermore, because MC simulations constitute models for dynamical processes, one would seek to calculate changes in F and S during a relaxation process, by assuming local equilibrium in certain parts along the MC trajectory; a classic example is simulation of protein folding.[11]

S, and F are commonly calculated by thermodynamic integration (TI) and perturbation techniques[1,2] that do not operate on a given MC sample but require conducting a *separate set* of MC simulations. This is a robust approach that enables one to calculate differences, ΔS_{ab} and ΔF_{ab}, between two states **a** and **b** of a system; however, if the structural variance of such states is large (e.g., helical and hairpin states of a polypeptide) the integration from state **a** to **b** becomes difficult and in many cases unfeasible. On the other hand, if one could calculate the absolute F_a and F_b directly from two separate sets of simulations carried out at states **a** and **b**, $F_{ab} = F_a - F_b$ and the integration can be avoided. Still, the absolute F can also be obtained with TI provided that a reference state **r** is available, where the free energy is known exactly and the integration path between **r** and **a** (and **b**) is relatively short. A classic example is the calculation of F of liquid argon or water by integrating the free energy change from an ideal gas reference state. However, for non-homogeneous systems such integration might not be trivial, and in models of peptides and proteins defining reference states that are close to the state of interest is a standing problem. It should be noted that F and S can always be obtained *approximately* from a given sample by harmonic or quasi-harmonic methods,[12] or by the local states method (see below).[13]

Another type of simulation method has been developed for polymers, where a chain is constructed step-by-step with transition probabilities (TPs).[14-17] The product of these TPs leads to $P_i{}^B$, hence S is known. To this category belong the direct MC (DMC)[14] procedure, the enrichment, Rosenbluth & Rosenbluth, and the dimerization methods,[15] the scanning method,[16] and other techniques.[17] However, these build-up procedures are not always the methods of choice mainly because they lack the dynamical aspects (and simplicity) of MC, which thus has become the commonly used method. Hence, it is important to develop methods for calculating the entropy from MC trajectories. Nonetheless, a hybrid of one buildup procedure, the scanning method,[16] with the dynamical MC approach has led to two *approximate* techniques, the local states[13] and hypothetical scanning (HS) methods.[18,19] These methods enable one to calculate S and F directly from a *given* sample generated by *any* simulation technique, they are general, and have been applied successfully to polymers, peptides, proteins, magnetic systems, and lattice gas models.[2] Unlike the harmonic and quasi-harmonic methods mentioned above,[12] HS and LS, in principle, can handle any chain flexibility, i.e., local fluctuations of a stable state (e.g., around an α-helix structure of a peptide), random coil fluctuations, as well as mixtures of these two extreme cases.

Recently, the HS method has been extended to fluids and has been further developed by replacing the deterministic partial future scanning used to calculate the TPs with a stochastic but complete scanning based on MC simulations;[3-5] this HSMC method has been applied very successfully to liquid argon, TIP3P water, and polyglycine molecules in helical, extended and hairpin states.[3-5] HSMC is significantly more accurate than HS, it provides rigorous upper

and lower bounds for *F*, which can be calculated from a relatively small sample and even from a *single* conformation.

As stated earlier, the aim of this paper is to extend the scope of HSMC to lattice polymer models, in particular to random coil chains. For that we study self-avoiding walks (SAWs) on a square lattice - a difficult test case due to the strong excluded volume interactions occurring in 2D. This paper is an extension of our recent letter[6] in which part of Table 1 has been presented and discussed. We emphasize the generality of HSMC and discuss its unique aspects for lattice systems, which makes it a powerful research tool independent of existing techniques. The HSMC results are compared to those obtained some time ago by the scanning method,[20] to results obtained by us using TI, to series expansion values,[21] and to results obtained by the HS method.

II. Theory

II.1 Statistical mechanics of SAWs

Assume a single SAW of *N* steps (bonds), i.e., *N*+1 monomers starting from the origin on a square lattice. All the SAWs *i* are equally probable with Boltzmann probability

$$P_i^{B} = 1 \big/ Z_{SAW}, \tag{1}$$

where the partition function, Z_{SAW}, is the total number of different SAWs, and the free energy is thus

$$F \big/ k_{B} T = -S \big/ k_{B} = \sum_i P_i^{B} \ln P_i^{B} = -\ln Z_{SAW} = \ln P_j^{B} \tag{2}$$

where k_B is the Boltzmann constant, and *j* is *any* SAW. The summations (in *i*) in Eq. (2) and in the rest of the paper (accept in section II.7) are over the *ensemble* of SAWs. Eq. (2) demonstrates that *F* (and *S* for this particular model) has zero fluctuation, which is a general property of the *correct* free energy of any system. On the other hand, the fluctuation of a free energy functional based on an *approximate* probability distribution (see below) is expected to be finite.[22] Eq. (2) also shows that if the Boltzmann probability of *any* single SAW (*j*) is known, *F* (and *S* for this particular model) is known as well, which again is a general property satisfied by any system in equilibrium.

II.2 The Direct Monte Carlo method

An unbiased sample of SAWs on a square lattice can be obtained by the direct Monte Carlo (DMC) method.[14] With this method a non-reversal *random walk* (ideal chain) is generated step-by-step, where at step *k* the direction of the kth bond is chosen at random (i.e. blindly) out of three possible directions (immediate chain reversal is not allowed). If the chosen site is unoccupied the bond is added to the existing chain, and the process continues; in the other case, the partial chain is discarded and a new one is started. This process is very inefficient for generating long SAWs due to strong (exponential) sample attrition. However, the entropy can be obtained from the relation $Z_{SAW}/Z_{id} \approx n_{suc}/n_{start}$ where Z_{id} is the known partition function of ideal chains, $Z_{id} = 4 \times 3^{N-1}$ and n_{suc} and n_{strart} are the number of chains started and the number of SAWs of *N* steps succeeded, respectively; this leads to an estimation, S_{DMC} for the entropy of SAWs,

$$S_{DMC} \big/ k_B = \ln (4 \times 3^{N-1} n_{suc} \big/ n_{start}) \tag{3}$$

II.3 The scanning simulation method

The scanning method is a step-by-step construction (growth) procedure, where unlike DMC, the bonds are not selected blindly, but with transition probabilities (TPs) that are based on scanning possible SAWs in future steps;[16] thus, at step k of the process, $k-1$ directions (bonds), ν will have already been constructed [these bond directions at each step are denoted $\nu_1, \ldots, \nu_{k-1}$, where $\nu = 1, 4$]. To determine the direction ν_k (out of 3 possible directions, ν) one enumerates all the possible continuations $Z_k^{\nu}(f)$ of the chain in a limited number of f future steps (typically less than the remaining bonds) that start from ν of step k, where $Z_k^{\nu}(f)$ is a partial future partition function and f is the scanning parameter. $Z_k^{\nu}(f)$ enables one to define TPs for ν,

$$p(\nu \mid \nu_{(k-1)}, \ldots, \nu_1, f) = Z_k^{\nu}(f) \Big/ \sum_{\nu=1}^{4} Z_k^{\nu}(f), \tag{4}$$

where because immediate reversal is forbidden, the summation is only over the three allowed directions. Using these TPs, the kth step is determined by a random number and the process continues. The construction probability $P_i^0(f)$ of SAW i is the product of the TPs with which the steps have been chosen,

$$P_i^0(f) = \prod_{k=1}^{N} p(\nu_k \mid \nu_{(k-1)}, \ldots, \nu_1, f). \tag{5}$$

For $f \ll N$ the scanning is incomplete and $P_i^0(f)$ is approximate. Due to this "incomplete" scanning, the chain can get trapped in a dead end during construction, meaning that the number n of completed constructions is smaller than the number n_{start} of those started. In other words, $P_i^0(f)$ is normalized over a subgroup of the random walks that includes all the SAWs and part of the self-intersecting walks. Also, $P_i^0(f)$ is biased, i.e., unlike P_i^B, it is larger for the compact SAWs than for the open ones. This bias can be decreased *systematically* by increasing f, where for a complete future scanning, i.e., $f_{\text{max}} = N - k + 1$, the TPs [Eq. (4)] become exact and no trapping occurs.[16] In practical applications the bias is removed by an *importance sampling* procedure, which leads to an unbiased estimation S that is exact within the statistical error

$$S \Big/ k_{\text{B}} = \ln \frac{1}{n_{\text{start}}} \sum_{t=1}^{n} \frac{1}{P_t^0(f)} \tag{6}$$

The scanning method can easily be extended to a chain model with finite interactions; in this case the interaction energy $E_{j(\nu)}^{k}(f)$ of the future chain j that starts from ν with itself and with the rest of the chain is calculated and the corresponding Boltzmann factor contributes to $Z_k^{\nu}(f)$, rather than 1,

$$Z_k^{\nu}(f) = \sum_{j(\nu)} \exp[-E_{j(\nu)}^{k}(f) \Big/ k_B T]. \tag{7}$$

II.4 The hypothetical scanning (HS) method

The HS method (as well as the local states method) is based on the concept that two samples in equilibrium generated by different simulation methods are equivalent in the sense that they both lead to the same estimates (within the statistical errors) of average properties, such as the entropy, energy, and their fluctuations. Relying on this equivalence, one assumes that a given sample of SAWs constructed by *any* exact procedure (e.g., Metropolis MC[10]) has instead been generated with the scanning method. Thus, for each of the bonds [$\nu_k(i)$] of SAW i one calculates the TP [Eq. (4)] as if i had been generated with the scanning method (we call this process the *reconstruction* of i, essentially an analysis procedure for obtaining TPs). The product of these

TPs leads to $P_i^0(f)$ [Eq. (5)] and to a functional S^A, which can be shown *rigorously* (using Jensen's inequality) to be an upper bound for S,[19]

$$S^A = -k_B \sum_i P_i^B \ln P_i^0(f),$$ (8)

where i runs on the *complete* ensemble of SAWs and S^A is a function of f. Because the sample is generated with an exact simulation procedure, S^A is a statistical average defined with the Boltzmann probability, which is normalized over the ensemble of SAWs. Each SAW i is associated with the variable $\ln P_i^0(f)$, where $P_i^0(f)$ is normalized over a larger ensemble that also consists of self-intersecting walks. The fluctuation σ_A of $\ln P_i^0(f)$,

$$\sigma_A = \left\{ \sum_{SAWs\, i} P_i^B [S^A + k_B \ln P_i^0(f)]^2 \right\}^{1/2},$$ (9)

is expected to be larger than zero, decreasing with increasing f (i.e., with improving the approximation), which means that S^A and σ_A are positively correlated. This correlation has been found to exist for good enough approximations.[22]

One can define another approximate entropy functional denoted S^B[19]

$$S^B = -k_B \sum_{SAWs\, i} P_i(f) \ln P_i^0(f),$$ (10)

where $P_i(f) = P_i^0(f)/P_i^0(f)$. If $P_i^0(f)$ was replaced in Eq. (10) by $P_i(f)$, according to the free energy minimum principle,[23] S^B would become a lower bound for S; but for SAWs it can only be shown rigorously[19] that $S^B \leq S^A$. However, when reliably estimated for a good enough approximation, S^B has been found in most cases[19,24] to underestimate S, as is also shown in the present calculations. In practice, lower bound behavior can be verified if S^B increases as the approximation improves; one can then assume that this trend would continue for better approximations meaning that S^B would converge to S. S^B can be estimated from a sample of size n by importance sampling,

$$\bar{S}^B = -\frac{k_B}{n} \left[\sum_{t=1}^n P_{i(t)}^0(f) \ln P_{i(t)}^0(f) \right] \Big/ \sum_{t=1}^n P_{i(t)}^0(f),$$ (11)

where $i(t)$ is SAW i obtained at time t of the *correct* Boltzmann simulation and the bar above S^B denotes estimation. However, the statistical reliability of this estimation (unlike the estimation of S^A) decreases sharply with increasing chain length, because the overlap between the probability distributions P_i^B and $P_i(f)$ decreases exponentially.

If S^B is a lower bound for S and the deviations of S^A and S^B from S (in the absolute values) are approximately equal, their average S^M becomes a better approximation than either of them individually,

$$S^M = [S^A + S^B] \Big/ 2.$$ (12)

Typically, several approximations for S^A, S^B, and S^M are calculated as a function of f, and their convergence enables one to determine the correct entropy with high accuracy. While application of HS to SAWs has been found to be quite efficient, for structured molecules such as an α- helix of a peptide HS has failed because it is impossible to carry out the future scanning within the limited conformational space defined by the local fluctuations of this structure, hence to define appropriate TPs. As discussed below, this problem does not exist with HSMC.

II.5 The HSMC method

While the TPs defined by HS are deterministic (based on the *entire* conformational space defined by f at step k of the reconstruction process), for a large chain they are always approximate, i.e., $f << N$ due to the exponential growth (with f) of the number of future SAWs. The HSMC method overcomes this limitation by seeking to estimate the *exact* TP defined by Eq. (4) with $f_{max}=N-k+1$, i.e. the whole future is scanned at step k. This is achieved by replacing the exact enumeration of f future steps at k by an MC simulation of the *entire* future segment of the chain (i.e., steps $k,k+1, \ldots N$) in the presence of the "frozen past" $[v_1 \ldots, v_{k-1}]$. The TP, denoted p^{MC} of the actual direction, $v_k(i)$ in the reconstructed SAW i is calculated from the number of MC steps, $n_k^{v(i)}$ for which $v_k(i)$ was visited during the simulation of total n_{MC} steps at step k

$$p^{MC}(v_k(i) \,|\, v_{(k-1)}, \ldots, v_1) = n_k^{v(i)} \Big/ n_{MC}$$ (13)

and the reconstruction probability of chain i is

$$P_i^{MC} = \prod_{k=1}^{N} P^{MC}(v_k(i) \,|\, v_{(k-1)}, \ldots, v_1)$$ (14)

where, for simplicity, i has been omitted in the TPs. In Eqs. (13) and (14) and in the rest of this paper, for brevity, we denote by MC physical quantities calculated by HSMC; notice, however, that in previous publications these properties were denoted by HS, which in this paper is reserved to denote results obtained with the HS method. Unlike the deterministic $P_i^0(f)$ [Eq. (5)], P_i^{MC} is defined stochasticly. The fact that the entire future is considered is important for systems with strong long-range interactions such as SAWs, proteins, etc; also, unlike $P_i^0(f)$ that is defined over the ensemble of SAWs and part of the ensemble of self-intersecting walks, P_i^{MC} is defined *only* over the ensemble of SAWs. As discussed below, this property of P_i^{MC} distinguishes HSMC from HS in many respects. Still, p^{MC} hence P_i^{MC} are approximate (due to finite simulation lengths), but one can show that as the MC simulation is increased, $p^{MC} \rightarrow p^{exact}$ and $P_i^{MC} \rightarrow P_i^B$, meaning that S can be estimated by reconstructing a *single* SAW. In practice, however, P_i^{MC} is approximate leading to an upper bound for S, [compare with Eq. (8)]

$$S^A = -k_B \sum_i P_i^B \ln P_i^{MC} = \sum_i P_i^B S_i^{MC}$$ (15)

where $S_i^{MC} = -k_B \ln P_i^{MC}$. It can be shown (see Appendix of Ref. 4) that like S^A [Eq. (8)], S^A [Eq. (15)] defined with stochastic probabilities, P_i^{MC}, is a rigorous upper bound, which is expected to have non-zero fluctuation σ_A [Eq. (9)].

One can define the entropy functional, S^B [Eq. (10)] and thus S^M [Eq. (12)] also for HSMC, where S^B becomes a *rigorous* lower bound of S due to the fact that P_i^{MC} is defined only over the ensemble of SAWs. We express S^B as

$$S^B = -k_B \sum_i P_i^{MC} \ln P_i^{MC} = -k_B \frac{\sum_i P_i^B [P_i^{MC} \ln P_i^{MC}]}{\sum_i P_i^B P_i^{MC}} = \frac{\sum_i P_i^B \exp[-S_i^{MC} / k_B][S_i^{MC}]}{\sum_i P_i^B \exp[-S_i^{MC} / k_B]},$$ (16)

where it is estimated by Eq. (11). Eq. (16) emphasizes an *explicit* dependence of S^B on the variable $S_{iMC} = -k_B \ln P_i^{MC}$, that is directly related to the average, S^A [Eq. (15)], and its fluctuation, σ [defined in the same manner as in Eq. (9)]. Because of the stochastic A nature of S_i^{MC} it is plausible to assume that when configurations (i) are sampled from the Boltzmann

distribution (i.e. with P_i^B), their corresponding S_i^{MC} values occur with a Gaussian probability centered around S^A with standard deviation σ_A. Indeed, such Gaussian behavior has been observed in models for liquid argon and TIP3P water, which has led (see details in Ref. 4) to a Gaussian approximation, S_G^B for S^B,

$$S_G^B = -\frac{(\sigma_A)^2}{k_B} + S^A, \tag{17}$$

and to the corresponding S_G^M (see Eq. 12),

$$S_G^M = (S^A + S_G^B)\Big/ 2 = S^A - \frac{1}{2}\frac{(\sigma_A)^2}{k_B}. \tag{18}$$

The fact that S_G^B depends only on S^A and σ_A is an advantage because these quantities are typically easier to estimate than S^B (directly) from Eqs. (10), (11) or (17), meaning that S_G^B is expected to be statistically more reliable than S^B. Previous results have shown that this Gaussian distribution is a very good approximation as there is excellent agreement of F_G^B with F^B for cases where F^B is well converged (when finite interactions are defined F G replaces S). Again, several approximations for S, S^A, S_G^B, and S_G^M can be calculated, and their convergence leads to highly accurate free energy determination. It should be pointed out that formally one can calculate S_G^B also for S_i^{HS} defined by $P_i^0(f)$ of the HS method. However S_i^{HS} (unlike S_i^{MC}) is not stochastic and thus deviates from a Gaussian distribution, where this deviation increases as the approximation worsens, i.e., with increasing chain length N.

The entropy can be expressed *exactly* by S^D (see Ref. 4), which can also be estimated from a sample generated with P_i^B. One obtains

$$S^D = -k_B \ln \sum_i P_i^B P_i^{MC} = -k_B \ln \left[\sum_i P_i^B[\exp(-S_i^{MC}/k_B)]\right] \tag{19}$$

In practice, the efficiency of estimating S by S^D depends on the fluctuation of this statistical average, which is determined by the fluctuation of S_i^{MC} exponentiated. That is, if the fluctuations in S_i^{MC} are small, then the values for $\exp(-S_i^{MC}/k_B)$ do not vary drastically, and the averages for S^D (and S^B) can be estimated reliably from a relatively small sample. Still (as for S^B), the direct calculation of S through S^D will not be as statistically reliable as estimating S^A. Obviously, as $S_i^{MC} \to S$ (i.e., $P_i^{MC} \to P_i^B$) all fluctuations become zero and S can be obtained from a single configuration. We note additionally that to improve convergence, S^D (like S^B) can be approximated by the Gaussian distribution (for the S_i^{MC} values in Eq. (19)); applying this approximation leads to S_G^M defined in Eq. (18).

As for S_G^B, one can formally calculate S^D also for $P_i^0(f)$ defined by HS. However, because $P_i^0(f)$ is not defined only on the ensemble of SAWs, $S^D(HS)$ (unlike $S^D(MC)$ will not converge to the correct S even for a very large sample. Convergence could occur for $P_i(f) = P_i^0(f)/\sum P_i^0(f)$ which is normalized over the SAWs alone (i.e., $\sum P_i^0(f) < 1$); however, calculation of $\sum P_i^0(f)$ by HS is impossible. This suggests that S^D calculated by HS for a large chain will always be an upper bound for S.

While the theory above has been introduced for the entire ensemble of SAWs, it also applies to a set of reconstructions of a single chain conformation (see Appendix, Ref. 4). That is, the required averages can be obtained from a set of n independent reconstructions of the same chain, where each reconstruction contributes an estimation for S_i^{MC}. For S^A, for example, these estimations are arithmetically averaged; for S^D the arithmetic average of $\exp(-S_i^{MC}/k_B)$ is used, etc.

II.6 Calculation of the entropy by series expansion

For comparison we also present results obtained with a formula based on series expansion (exact enumeration) data.[21] The entropy, S_{series}, is obtained from the total number of SAWs, c_N,

$$S_{series}\bigg|\, k_B = \ln c_N \cong \ln \left\{ \begin{array}{l} \mu^N[a_1 N^{11/32} + a_2 N^{-21/32} + b_1 N^{-37/32} + \\ +(-1)^N d_1 N^{-3/2} + (-1)^N d_2 N^{-2}] \end{array} \right\} \tag{20}$$

where $a_1 = 1.1771(2)$, $a_2 = 0.554(2)$, $b_1 = -0.19(2)$, $d_1 = -0.19(2)$, $d_2 = 0.034(2)$, and $\mu = 2.6381585$ (10) (the error of the last digit appears in parenthesis).

II.7 Calculation of the entropy by thermodynamic integration

In order to calculate the partition function of a SAW via thermodynamic integration (TI)[25] the system must be linked with a calculable reference state, which in this case is the ideal chain. Samples of chains are generated where monomers are allowed to overlap each other. To effect this, a unitless energy function, E, is defined where

$$E = \sum_j \phi_j \tag{21}$$

ϕ_j is the "overlap value" at lattice site j, and the summation is carried out over all sites. The overlap value is defined as follows. A lattice site that is occupied by only a single monomer (or is unoccupied) contributes nothing to the energy ($\phi_j = 0$). A doubly occupied site (i.e. a single overlap) contributes $\phi_j = 1$; a triply occupied site (a double overlap) contributes $\phi_j = 2$, and so on. The value of E is thus always an integer. For a SAW, E must be zero.

The above defined energy function is used to describe a *general* chain which can exist at any arbitrary finite temperature. The partition function for the general chain ensemble is given by

$$Z = \sum_{id} \exp[-E_i / T] \tag{22}$$

where the sum is carried out over all *ideal chain* configurations i (the total configuration space), and where we have introduced a unitless temperature, T. We note that at high (infinite) T, the Boltzmann factor, $\exp[-E_i/T]$, is unity and the partition function approaches that of the ideal chain reference state (i.e. $Z_{id} = 4 \times 3^{N-1}$, where immediate reversal is forbidden). At low T ($T = 0$), only zero energy configurations will contribute to the summation and the partition function becomes that of the SAW, Z_{SAW}.

The difference, $\ln Z_{SAW} - \ln Z_{id}$, can be evaluated by integration over T, or over $1/T$, using the derivative relations,

$$\frac{d \ln Z}{dT} = \left(\frac{1}{Z}\right)\sum_{id} \frac{E_i}{T^2} \exp[-E_i / T] = \left\langle \frac{E}{T^2} \right\rangle \tag{23}$$

and

$$\frac{d \ln Z}{d(1 / T)} = \left(\frac{1}{Z}\right)\sum_{id} -E_i \exp[-E_i / T] = -\langle E \rangle. \tag{24}$$

The corresponding integrals are respectively

$$\ln\left[\frac{Z_{\mathrm{SAW}}}{Z_{\mathrm{id}}}\right] = \int\limits_{\infty}^{0}\frac{d\ln Z}{dT}dT \ \text{ and } \ \ln\left[\frac{Z_{\mathrm{SAW}}}{Z_{\mathrm{id}}}\right] = \int\limits_{0}^{\infty}\frac{d\ln Z}{d(1/T)}d(1/T).$$ (25)

We have chosen to use both of these relations where we conduct the integration in two stages as

$$\ln\left[\frac{Z_{\mathrm{SAW}}}{Z_{\mathrm{id}}}\right] = \int\limits_{0}^{1/T^*}\frac{d\ln Z}{d(1/T)}d\left(1\left|T\right.\right) + \int\limits_{T^*}^{0}\frac{d\ln Z}{dT}dT,$$ (26)

where T^* is an intermediate temperature. The left-hand term thus quantifies the change in ln$Z$ for going from an ideal chain to the general chain at T^*, and the right-hand term is the change from this point to the SAW.

In our implementation, the generalized chain was simulated at a total of 199 temperatures. The relevant result in these simulations is the average energy, $\langle E\rangle$; these values are used for derivative points [Eqs. (23) and (24)] in the numerical evaluation of Eq. (26). In the first stage/series (corresponding to the left-hand term in Eq. (26), 100 simulation temperatures were spaced evenly in $1/T$, ranging from $1/T = 0$ (the ideal chain) to $1/T^*$ where $T^* = 0.75757575$ ($1/T^* = 1.32$). In the second stage (for the right-hand term in Eq. (27), 100 simulation temperatures were spaced evenly in T, ranging from T^* to $T = 0$ (the SAW). With the finite limits in Eq. (26), a simple trapezium integration was adequate. It should also be noted that the number of simulation points employed in this work was actually well more than was necessary. We note further that the results are insensitive to the choice of T^* as long as there are enough points. One could drastically reduce the number of simulation points (thereby increasing the efficiency) by careful (optimized) choice of T^* and/or by implementing less simple-minded quadrature techniques; however the present performance is sufficient for our purposes. Details about the MC simulations appear below in III.1.

III. Results and discussion

We have calculated the entropy of SAWs consisting of $N= 29, 49, 99, 149, 249, 399$ and 599 bonds. The results of Table 1 were obtained by reconstructing a single chain conformation (see Appendix, Ref. 4), i.e., by n replicate reconstructions (based on different sets of random numbers) of a *straight* SAW of N bonds, while the results in Table 2 were obtained by reconstructing a sample of SAWs.

III.1 MC simulations and the HSMC reconstruction procedure

The efficiency of HSMC is affected considerably by the MC procedure employed in the reconstruction process. On a square lattice, "crankshaft" moves are in most cases rejected due to the strong excluded volume interactions while corner moves have somewhat higher acceptance rate.[8] Therefore, for the reconstruction process we have used an MC procedure based on 50% corner moves (that provide local conformational changes) and 50% "pivot" moves that have been shown to effectively induce global changes.[26] This procedure has been employed not only in the reconstruction process, but also for generating samples of SAWs (to be reconstructed by HSMC and HS), and for the TI simulations.

The HSMC calculations are based on the sample size n - the number of reconstructed SAWs and n_{future}, which is related to the number of future MC steps per bond applied during the reconstruction process as defined below. First we note that the first bond of the chain is not reconstructed; its probability is always ¼. The number of MC steps, n_{MC}, for bond k is scaled as $n_{\mathrm{MC}}=(N-k+1)n_{\mathrm{future}}$, meaning that the maximal number of future MC steps is applied for

the reconstruction of the second bond (to which corresponds the largest future segment of N −1 bonds), while the last bond (N) is allotted the minimal number of MC steps. Because each simulation at step k always starts from the structure of the reconstructed chain it is important to let the future SAW equilibrate, otherwise p^{MC} [Eq. (13)] would (on average) be too high; therefore, 300 MC steps per future bond are used for equilibration. As discussed earlier, the larger is n_{future} the better (i.e., smaller) is S^A [Eq. (15)], the larger is S^B [Eq. (16)] [and $S_G{}^B$ Eq. (17)] and the smaller is the fluctuation, σ_A [Eq. (9)]. To demonstrate this effect, the results for each chain length are presented in the Tables for $n_{future} = 500$, 5000, and 50000, where the corresponding sample size, n, is decreased, which results in approximately the same computer time for each calculation. Notice that for a single chain (Table 1), n is the number of reconstructions applied to the same straight chain, while for a sample of chains (Table 2), n is the number of different configurations reconstructed – one reconstruction is performed for each configuration.

For the TI process the chains were simulated as described above, by the 50/50 ratio of pivot and corner moves, where in this case the entire chain is moveable (except of the first bond). The total simulation length was the same at each temperature, however it varied depending on chain size. 6×10^7 MC moves were carried out at each temperature for $N = 29$, where run lengths of 10^8, 10^8, 6×10^7, 5×10^7, 3.2×10^7 and 2.4×10^7 steps were used for $N = 49$, 99, 149, 249, 399, and 599, respectively. All of these runs were replicated 9 times (i.e. 9 independent simulations were performed), thus yielding 9 independent integration results (trials) for each chain size. Our final reported result is the average of these trials, with the standard deviation of the mean being used as the uncertainty estimate.

III.2 Results by TI, series expansion, and the scanning method

To a large extent, we judge the performance of HSMC by comparing its results to those obtained by other techniques, such as the scanning method [Eq. (6), Ref. 20], series expansion (Eq. 20), thermodynamic integration (TI) [Eq. (26)] and HS (using f=8); therefore, we start by discussing the results of these methods which appear in both tables.

We first would like to point out the surprising accuracy for large N obtained by the series expansion formula [Eq. (20)] that is based on extrapolating exact enumeration data for relatively short chains. Thus, the results for S_{TI} and S_{series} are equal within the error bars for all N, with comparable errors for N=49, 99, and 149. However, for N=29 the error in S_{series} is significantly larger than that of S_{TI} and for N>149 error(S_{TI})−error(S_{series}) increases constantly with N. For N=29 the DMC and TI values are equal within comparable errors.

The results obtained with the scanning method long ago[20] (based on a relatively small scanning parameter, f=6) are also very good. They are equal to the TI and series results for all N accept for N=599, where S_{scan} is smaller due to a bias (for generating compact chains) that was not removed completely by the importance sampling procedure [Eq. (6)]. For $N \geq 249$ the statistical errors of S_{scan} are significantly larger than those of S_{TI}. In what follows, for comparison we shall consider the TI and series results to be exact.

III.3 Entropy by reconstructing straight chains

Results obtained by n replicate reconstructions of a straight chain appear in Table 1. Part of the data have already been provided in Ref. 6; however, the HS results and those for $S_G{}^B$, $S_G{}^M$, and for the chain length N=29 are new.

The table supports the expectations of the HSMC theory presented in section II. Thus, for all chain lengths, as n_{future} is increased from 500 to 50,000, the fluctuation decreases, S^A decreases and remains an upper bound, and S^B and $S_G{}^B$ increase remaining lower bounds. For $n_{future} =$

500 the $S_G{}^B$ results are slightly inferior (i.e., lower) than those of S^B. However, for n_{future} =5000 and 50000 $S_G{}^B$ and S^B are equal within error bars that are, however, 2–3 times smaller for $S_G{}^B$ than for S^B; therefore, the corresponding results for $S_G{}^M$ are equal to those of S^M but with slightly smaller errors.

In all cases S^M and $S_G{}^M$ are equal (within the error bars) to S^D, to the TI and series results, and for N<599 also to the scanning results. However, the error bars of TI are the smallest. The fact that for each N the S^M (and $S_G{}^M$) results for n_{future}=5000 and 50000 (and in some cases for n_{future}=500) are equal (and they are also equal to the TI values) demonstrates that the absolute values of S^A and S^B ($S_G{}^B$) deviate equally from the correct results. Overall the HSMC statistical errors are small (0.002–0.005%); however, much more computer time has been invested in the simulations of the longer chains.

We also obtained results with HS where the entropy was calculated from a generated sample of chains (see next section) with a limited but systematic scanning of f=8. Our main interest has been only to check how much are these results larger than those obtained by HSMC (based on the stochastic MC scanning of the *entire* future); therefore, the HS results were calculated only for several chain lengths of N=49, 99, 249, and 399. Indeed, the S^A(HS) values [Eq. (8)] are always larger than the exact ones, where the deviation increases with N; thus, for N=49 the HS value is relatively accurate, comparable to that of HSMC(n_{future}=5000), while for N=399 S^A(HS) worsens becoming close to S^A(HSMC) for n_{future}=500. Correspondingly, σ_A (HS) is always larger than σ_A (HSMC) obtained for n_{future}=5000 (accept for N=49). A similar trend is observed for S^B(HS) which is always a lower bound but smaller than S^B(HSMC) obtained for n_{future} =5000.

The results for S^M(HS) are very close to the correct ones for N=49 and 99, but overestimate the correct values as chain length increases, where for N=399 the error is of ~0.2%. As discussed earlier, $S_G{}^B$ (HS) is not well defined and indeed it constitutes a lower bound only for N=49 and 99 (where its values are larger than the corresponding S^B(HSMC) values), becoming larger than the exact value for larger N. The related average, $S_G{}^M$ is always larger than the exact value with the largest deviation of 0.36% occurring for N=399. As expected (see last paragraph of II.5) S^D(HS) is always an upper bound, which is slightly smaller than the corresponding $S_G{}^M$. These results demonstrate that the performance of HS is inferior to that of HSMC.

III.4 Entropy by reconstructing a sample of chains

In practice, however, one would apply HSMC to samples of chains of different conformations, therefore a second set of results has been obtained from thermodynamic samples of SAWs. To generate such samples we have carried out long MC runs (based on the pivot and corner moves described previously) starting from a straight chain, equilibrating for 300 MC steps per bond, where every 2300 MC steps per bond the current conformation (*i*) was selected for reconstruction as described earlier. (This same prescription was also used to generate samples for the HS method.)

Calculation of the entropy from a sample of chains is more difficult than for a straight chain. Because the frozen past of the chain is not straight, part of conformational space of the future SAWs might become unreachable with our dynamic pivot/corner MC procedure; this might affect in particular the movement of the treated bond k hence the corresponding transition probability. Because the reconstruction starts from configuration i, in an extreme case bond k (*i*) will be unable to move to another direction leading to p^{MC} =1; in another case it will change direction but may never return to its original direction in chain i leading to p^{MC}=0. Such TPs will affect significantly the probability $P_i{}^{MC}$ [Eqs. (13) and (14)] of the chain. Notice, however, that these cases do not demonstrate a drawback of the HSMC method but reflect the strong excluded volume interaction of SAWs on a square lattice and the inability inherent in our MC

procedure to search the entire conformational space (i.e., the procedure is non-ergodic). As discussed below, these problems can be alleviated by generating the future SAWs with more efficient MC techniques. Such problems have not been encountered in application of HSMC to fluid systems (argon and water) and peptides. Obviously, this problem will be weakened significantly for SAWs on a simple cubic lattice, for example, where the excluded volume interactions are less severe than on a square lattice.

To alleviate these problems we have taken several measures. First, before carrying out the future sampling at step k the program checks the nearest neighbor sites of monomer k (located at the end of bond $k-1$); if all four of them are already occupied by chain monomers (i.e., step k has only one choice) the future sampling is avoided, the TP(k) is defined as 1, and the next step ($k+1$) is treated. When p^{MC}=0 or 1 occur, p^{MC} is calculated by the (systematic) HS method, i.e., by an exact enumeration of the future SAWs of f=8 bonds and this value is considered in the calculation of P_i^{MC}. Still, the reconstruction probability of some chains might be affected significantly by similar problems (i.e., p^{MC} values that are incorrectly very small or close to 1). Because the Boltzmann probability of all chains is the same, one can ignore the contribution of such chains to the average entropy. In practice, a SAW i with $-k_B \ln P_i^{MC}$ beyond four standard deviations of the average is not considered in the averaging of the entropy.

Comparing the results in Tables 1 and 2 demonstrates the increase in sampling difficulty and decrease in accuracy involved in reconstructing a sample of chains. Thus, while S^A(sample) in Table 2 (as expected) is an upper bound that decreases as n_{future} increases, it is always larger (i.e., worse) than the corresponding S^A(straight) in Table 1; for n_{future} =50000 the deviations are small for $N \leq 149$ but increase for larger N. A similar trend is observed for σ_A (sample) that always decreases (as expected) with increasing n_{future} but it is larger than the corresponding σ_A (straight). Notice that for n_{future} =50000 S^A(sample) and σ_A (sample) are always better (i.e., smaller) than S^A(HS) and σ_A (HS), which again reflects the superior accuracy of HSMC. S^B and S_G^B always increase with n_{future} and for n_{future} =5000 and 50000 they are in most cases equal with slightly lower errors for S_G^B. Again, these values are always smaller (i.e., worse) than those in Table 1, where the deviations are small for $N \leq 149$ and increase for larger N. For $N \leq 99$ the (6) results for S^M, S_G^M, and S^D for n_{future} = 5000 and 50000 are all equal (within the error bars), whereas for larger N these functionals have slightly better values at 50000 than at 5000. For $N \leq 249$ the best results for S^M, S_G^M and S^D (i.e. for n_{future}= 50000) are equal to those of Table 1, while for N=399 and 599 the results for the straight chains are more accurate with errors that are ~6 times smaller.

The conclusion from the above comparison is that for any model studied it is more efficient to carry out a relatively large number of reconstructions (replicates) of a small number of "good" chain configurations than to reconstruct a thermodynamic sample of chains.

III.5 Discussion

The results of the two tables show that for a given amount of computer time it is preferable to increase n_{future} using relatively small values of sample size n; this leads to improved (smaller) S^A, larger S^B, hence better estimates S^M and S^D (in particular for large N). This effect is significant in particular for a sample of chains, where, for N=49, for example, S^A(n_{future}=50000) is equal in both tables, while for n_{future}=5000 and 500 the results in Table 2 are always worse (larger) than the corresponding results of Table 1. Thus, the best (lowest) S^A will be obtained in the extreme case, where only a single (good) chain is reconstructed with the maximal n_{future} for a given amount of computer time. However, this would come with a price that the information provided by the other functionals would be lost because S^A=S^B=S^M=S^D.

An inherent inefficiency of HSMC lies in the need to carry out $N-1$ simulations for an N-bond SAW. Still, performance of HSMC for a sample of SAWs can be improved by changing the scaling function discussed in section II.3, which controls the extent of simulation applied to each bond in the reconstruction process. However, the most significant factor affecting efficiency is the simulation method used for the chain reconstruction. Thus, our preliminary simulations based on corner moves alone have converged extremely slowly, and adding the pivot moves improved performance dramatically. In three dimensions, where the excluded volume effect is weaker, one can add crankshaft moves (and other moves, see Ref. 8) that are expected to increase efficiency further. To improve accuracy one can increase the scanning parameter used in the HS parts of the processes from f=8 to12 (and even to 14).

The pivot moves are very important for an open chain, but they become unsuitable for a SAW enclosed in a small volume, or for a highly compact SAW with attractive interactions at low temperature, where only local MC moves are applicable. Notice, however, that simulating these restricted models (on a square or a simple cubic lattice) with dynamic MC procedures based on local moves is generally non-ergodic and extremely inefficient, meaning that the corresponding HSMC reconstructions will be inefficient as well. On the other hand, restricted SAW models are better handled by step-by-step construction procedures.[14–17] The scanning method, for example, is ergodic and due to its "feelers" one can generate chains in restricted environments quite efficiently. Thus, the idea would be to implement within the framework of HSMC a suitable growth procedure, which will lead to exact results, unlike HS. Notice that growth procedures provide the entropy by themselves from their generated sample of chains; however, a suitable HSMC/growth procedure would enable estimating the entropy from a given trajectory.

An interesting test case is a model of multiple SAWs enclosed in a "box", studied previously by the scanning and HS methods,[24] where chains are added successively to an initially empty box. However, with HS only the *partial* future of a reconstructed chain is considered, whereas HSMC can take into account the entire future, including that of the reconstructed chain and the positions and conformations of the as yet unreconstructed chains. If the system is not extremely dense local dynamic MC moves would suffice. In the extreme case where all sites are populated (density=1) one can apply simulation methods as those implemented by Pakula and Reiter.[27] It should be emphasized that HSMC can handle volumes with any shape and boundary conditions, where defining a suitable reference state for TI is not trivial.

Chain models with finite interactions have been defined on enriched lattices (i.e., with a large coordination number, such as the bond fluctuating model) and have been simulated by dynamic MC procedures. All of these models can be treated by HSMC. Such models have been used to study protein folding trajectories, for example, where transitions between different conformational regions (microstates) occur, but their relative populations can be obtained only crudely from the trajectory. However, these populations can be calculated with high accuracy by applying HSMC locally to these microstates, in the same way it has been applied to the helical, extended, and hairpin microstates of polyglycine molecules.[5] It should be noticed that the entropy of microstates (i.e. local fluctuations) can also be obtained *approximately* by the harmonic and quasi-harmonic techniques[12] or the local states method,[13] while a similar calculation by TI is a standing problem. Returning to the present model of SAWs, it appears that the most efficient is the scanning method (where a run for generating SAWs of N=599 provides results for all intermediate N), followed by TI, where HSMC is the least efficient. For example, the tabulated TI value for a 399-bond SAW required ~100 h CPU, while for HSMC (n_{future}=50000), generating the 225 chains in Table 1 took 945 h CPU. It is stressed however that these levels of precision will often not be necessary in novel investigations on related polymer systems. A single reconstruction of a 399-bond SAW for n_{future}=50000 requires far less computational investment (~4.2 h CPU), and already gives a result of S=0.9757 (6).

In summary, calculation of S is a central problem in computer simulation, and HSMC with its unique features constitutes a new tool for obtaining S independent of other methods. With HSMC all interactions are considered, and its accuracy depends only on the amount of MC sampling. Furthermore, a "self checking" accuracy analysis is inherent in the method, based on verifying the increase and decrease of the rigorous upper and lower bounds, S^B, $S_G{}^B$, and S^A, and the decrease of σ_A, as the approximation improves. Finally, unlike other methods, HSMC is of general applicability, covering liquids (argon and water), microstates of polypeptide molecules, and in this work also random coil polymers. HSMC can be applied to any type boundary conditions, which is very difficult to handle by TI, and unlike most methods, enables one to extract the absolute entropy from a given sample, where only a small number of SAWs (and even a single chain) need to be reconstructed; this is important for studying relaxation processes, such as protein folding.

References

1. Beveridge DL, DiCapua FM. Annu Rev Biophys Biophys Chem 1989;18:431. [PubMed: 2660832] Kollman PA. Chem Rev 1993;93:2395.Jorgensen WL. Acc Chem Res 1989;22:184.

2. Meirovitch, H. Reviews in Computational Chemistry. Kenny B, Lipkowitz; Donald B, Boyd, editors. 12 . Wiley, New York: 1998. p. 1

3. Szarecka A, White RP, Meirovitch H. J Chem Phys 2003;119:12084.White RP, Meirovitch H. J Chem Phys 2003;119:12096.White RP, Meirovitch H. Proc Natl Acad Sci USA 2004;101:9235. [PubMed: 15197270]

4. White RP, Meirovitch H. J Chem Phys 2004;121:10889. [PubMed: 15634040]

5. Cheluvaraja S, Meirovitch H. Proc Natl Acad Sci USA 2004;101:9241. [PubMed: 15197271] Cheluvaraja S, Meirovitch H. J Chem Phys 2005;122:054903–1.

6. White RP, Funt J, Meirovitch H. Chem Phys Lett 2005;410:430. [PubMed: 16912812]

7. Carmesin I, Kremer K. Macromolecules 1988;21:2189.Deutsch HP, Binder K. J Chem Phys 1991;94:2294.Geisinger T, Müller M, Binder K. J Chem Phys 1999;111:5251.Müller M, Binder K, Schäfer L. Macromolecules 2000;33:4568.Xu G, Mattice WL. J Chem Phys 2002;117:3440.Chen D, Mattice WL. Polymer 2004;45:3877.Termonia Y. biomacromolecules 2004;5:2404. [PubMed: 15530057]Termonia Y. Text Res J 2003;73:74.

8. Sokal, A. Monte Carlo and Molecular Dynamics Simulations in Polymer Science. Kurt, Binder, editor. Oxford University Press; 1955. p. 47-124.

9. Taketomi H, Ueda Y, Gō N. Int J Pept Protein Res 1975;7:449.Lau KF, Dill KA. Macromolecules 1989;22:3986.Covell DJ, Jernigan RL. Biochemistry 1990;29:3287. [PubMed: 2334692]Hinds DA, Levitt M. Proc Natl Acad Sci USA 2004;89:2536. [PubMed: 1557356]Berriz GF, Shakhnovich EI. Curr Opin Colloid Interface Sci 1999;4:72.Kolinski A, Milik M, Rycombel J, Skolnick J. J Chem Phys 1995;103:4312.Zhang Y, Skolnick J. Biophys J 2004;87:2647. [PubMed: 15454459]Zuckerman DM. J Phys Chem B 2004;108:5127.

10. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. J Chem Phys 1953;21:1087.

11. Duan Y, Kollman PA. Science 1998;282:740. [PubMed: 9784131]

12. Gō N, Scheraga HA. J Chem Phys 1969;51:4751.Hagler AT, Stern PS, Sharon R, Becker JM, Naider F. J Am Chem Soc 1979;101:6842.Karplus M, Kushick JN. Macromolecules 1981;14:325.

13. Meirovitch H. Chem Phys Lett 1977;45:389.Meirovitch H, Koerber SC, Rivier J, Hagler AT. Biopolymers 1994;34:815. [PubMed: 8054467]

14. Wall FT, Hiller LA, Wheeler DJ. J Chem Phys 1954;22:1036.

15. Rosenbluth MN, Rosenbluth AW. J Chem Phys 1955;23:356.Wall FT, Erpenbeck JJ. J Chem Phys 1959;30:634.Alexandrowicz Z. J Chem Phys 1969;51:561.

16. Meirovitch H. J Phys A 1982;15:L735.Meirovitch H. J Chem Phys 1988;89:2514.

17. Bascle J, Garel T, Orland H, Velikson B. Biopolymers 1993;33:1843. [PubMed: 8268410] Grassberger P, Hegger R. J Phys A 1994;27:4069.Grassberger P. Phys Rev E 1997;56:3682.Kumar SK, Szleifer I, Panagiotopoulos AZ. Phys Rev Lett 1991;66:2935. [PubMed: 10043657]

18. Meirovitch H. J Phys A 1983;16:839.

19. Meirovitch H. Phys Rev A 1985;32:3709. [PubMed: 9896540]

20. Meirovitch H. Macromolecules 1985;18:563.

21. Guttmann AJ, Enting IG. J Phys A 1988;21:L165.Conway AR, Enting IG, Guttmann AJ. J Phys A 1993;26:L1519.

22. Meirovitch H, Alexandrowicz Z. J Stat Phys 1976;15:123.Meirovitch H. Chem Phys 1999;111:7215.

23. Hill, TL. Statistical Mechanics Principles and Selected Applications. Dover, New York: 1956.

24. Meirovitch H. J Chem Phys 1992;97:5803.5816

25. Muller M, Paul W. J Chem Phys 1994;100:719.

26. Madras N, Sokal AD. J Stat Phys 1987;47:573.

27. Pakula T. Macromolecules 1987;20:679.Reiter J. Macromolecules 1990;23:3811.

Table 1

HSMC results for the entropy per bond of N-bond SAWs[a]

n_{future}	S^A/k_B	σ_A	S^B/k_B	S_G^B/k_B	S^M/k_B	S_G^M/k_B	S^D/k_B	n
				$N = 29$ $S_{DMC} = 1.016147\ (5)$				
500	1.02084 (2)	0.01807 (2)	1.01148 (5)	1.01137 (3)	1.01616 (3)	1.01610 (2)	1.01614 (3)	1250000
5000	1.01662 (2)	0.00568 (2)	1.01568 (3)	1.01568 (3)	1.01615 (2)	1.01615 (2)	1.01615 (2)	125000
50000	1.01618 (2)	0.00181 (2)	1.01609 (3)	1.01609 (2)	1.01614 (2)	1.01614 (2)	1.01614 (2)	12500
S_{TI}	1.016145 (3)		1.016145 (3)	1.016145 (3)	1.016145 (3)	1.016145 (3)	1.016145 (3)	
S_{series}	1.01615 (1)		1.01615 (1)	1.01615 (1)	1.01615 (1)	1.01615 (1)	1.01615 (1)	
				$N = 49$ $S_{SCAN} = 1.000904\ (4)$				
500	1.00583 (1)	0.01424 (2)	0.99602 (5)	0.99590 (3)	1.00093 (3)	1.00086 (2)	1.00091 (3)	1250000
5000	1.00140 (1)	0.00448 (2)	1.00042 (5)	1.00042 (1)	1.00091 (2)	1.00091 (1)	1.00091 (2)	125000
50000	1.00095 (1)	0.00142 (2)	1.00085 (3)	1.00085 (1)	1.00090 (2)	1.00090 (1)	1.00090 (2)	12500
S_{HS}	1.00149 (1)	0.00434 (1)	1.00026 (2)	1.00057 (1)	1.00088 (2)	1.00103 (1)	1.00094 (1)	250000
S_{TI}	1.000897 (3)		1.000897 (3)	1.000897 (3)	1.000897 (3)	1.000897 (3)	1.000897 (3)	
S_{series}	1.000899 (5)		1.000899 (4)	1.000899 (4)	1.000899 (4)	1.000899 (4)	1.000899 (4)	
				$N = 99$ $S_{SCAN} = 0.987726\ (5)$				
500	0.99294 (2)	0.01030 (3)	0.9826 (6)	0.98243 (6)	0.98775 (5)	0.98769 (4)	0.98773 (5)	250000
5000	0.98826 (2)	0.00324 (3)	0.98722 (5)	0.98722 (3)	0.98774 (3)	0.98774 (2)	0.98774 (3)	25000
50000	0.98777 (2)	0.00101 (3)	0.98767 (4)	0.98767 (2)	0.98772 (2)	0.98772 (2)	0.98772 (3)	2500
S_{HS}	0.98994 (1)	0.00507 (1)	0.9856 (2)	0.9874 (1)	0.9878 (1)	0.9887 (1)	0.98817 (5)	250000
S_{TI}	0.987727 (3)		0.987727 (3)	0.987727 (3)	0.987727 (3)	0.987727 (3)	0.987727 (3)	
S_{series}	0.987730 (3)		0.987730 (3)	0.987730 (3)	0.987730 (3)	0.987730 (3)	0.987730 (3)	
				$N = 149$ $S_{SCAN} = 0.982740\ (3)$				
500	0.98806 (2)	0.00852 (3)	0.9774 (2)	0.97725 (8)	0.9827 (1)	0.98265 (4)	0.9827 (1)	250000
5000	0.98329 (2)	0.00267 (3)	0.98222 (5)	0.98223 (3)	0.98276 (3)	0.98276 (2)	0.98276 (3)	25000
50000	0.98281 (2)	0.00085 (3)	0.98270 (4)	0.98270 (2)	0.98275 (2)	0.98275 (2)	0.98275 (3)	2500
S_{TI}	0.982742 (3)		0.982742 (3)	0.982742 (3)	0.982742 (3)	0.982742 (3)	0.982742 (3)	
S_{series}	0.982740 (2)		0.982740 (2)	0.982740 (2)	0.982740 (2)	0.982740 (2)	0.982740 (2)	
				$N = 249$ $S_{SCAN} = 0.97836\ (2)$				
500	0.98391 (3)	0.00669 (4)	0.9727 (3)	0.9728 (1)	0.9783 (2)	0.97833 (7)	0.9783 (2)	63000
5000	0.97889 (2)	0.00208 (4)	0.97782 (8)	0.97782 (5)	0.97836 (4)	0.97836 (3)	0.97836 (5)	9100
50000	0.97840 (2)	0.00066 (4)	0.97829 (5)	0.97829 (2)	0.97835 (3)	0.97835 (2)	0.97835 (3)	930
S_{HS}	0.98306 (1)	0.00401 (1)	0.9745 (5)	0.9791 (3)*	0.9788 (3)	0.9811 (3)	0.9799 (1)	176000
S_{TI}	0.978358 (4)		0.978358 (4)	0.978358 (4)	0.978358 (4)	0.978358 (4)	0.978358 (4)	
S_{series}	0.978360 (1)		0.978360 (1)	0.978360 (1)	0.978360 (1)	0.978360 (1)	0.978360 (1)	
				$N = 399$ $S_{SCAN} = 0.97567\ (4)$				
500	0.98138 (6)	0.00540 (5)	0.9710 (5)	0.9697 (2)	0.9762 (3)	0.9756 (1)	0.9759 (3)	9500
5000	0.97625 (4)	0.00170 (5)	0.9751 (1)	0.97509 (8)	0.97567 (5)	0.97567 (5)	0.97567 (5)	2000
50000	0.97568 (4)	0.00053 (5)	0.97557 (7)	0.97557 (5)	0.97563 (4)	0.97563 (5)	0.97563 (5)	225
S_{HS}	0.98141 (5)	0.00335 (5)	0.9743 (5)	0.9769 (3)	0.9779 (3)	0.9792 (2)	0.9782 (2)	5500
S_{TI}	0.975655 (8)		0.975655 (8)	0.975655 (8)	0.975655 (8)	0.975655 (8)	0.975655 (8)	
S_{series}	0.975652(1)		0.975652 (1)	0.975652 (1)	0.975652 (1)	0.975652 (1)	0.975652 (1)	
				$N = 599$ $S_{SCAN} = 0.97395\ (5)$				
500	0.98003 (8)	0.00445 (7)	0.9706 (8)	0.9682 (4)	0.9753 (4)	0.9741 (2)	0.9748 (5)	3000
5000	0.97466 (7)	0.00139 (7)	0.9736 (2)	0.9735 (1)	0.9741 (1)	0.97408 (9)	0.9741 (1)	450
50000	0.97413 (5)	0.00036 (7)	0.9741 (1)	0.97405 (6)	0.97409 (6)	0.97409 (5)	0.97409 (5)	45
S_{TI}	0.97404 (1)		0.97404 (1)	0.97404 (1)	0.97404 (1)	0.97404 (1)	0.97404 (1)	
S_{series}	0.974025(1)		0.974025(1)	0.974025(1)	0.974025(1)	0.974025(1)	0.974025(1)	

[a]The results were obtained from n reconstructions of a straight chain. S^A [Eqs. (8) and (15)] is an upper bound, and σ_A is its fluctuation [Eq. (9)]. S^B [Eqs (10), (11), and (16)] and its Gaussian approximation, S_G^B [Eq. (17)] are lower bounds, and their averages with S^A are denoted S^M [Eq. (12)] and S_G^M [Eq. (18)], respectively. S^D [Eq. (19)] is an exact entropy functional. n_{future} is related to the number of MC steps per bond (see text). S_{TI}, S_{scan}, S_{series}, and S_{HS} were obtained by thermodynamic integration [Eq. (26)], the scanning method [Eq. (6), Ref. 20], a series expansion formula [Eq. (20)], and the HS method, respectively. The statistical error is defined by parentheses: $1.00(3) = 1.00 \pm 0.03$.

Table 2

HSMC results for the entropy per bond obtained from a sample of chain configurations. For details, see the · caption of Table 1.

n_{future}	S^A/k_B	σ_A	S^B/k_B	S_G^B/k_B	S^M/k_B	S_G^M/k_B	S^D/k_B	n
				$N = 29$ $S_{DMC} = 1.016147\ (5)$				
500	1.02366 (2)	0.02319 (2)	1.00866 (5)	1.00807 (3)	1.01616 (3)	1.01587 (2)	1.01609 (3)	1250000
5000	1.01689 (2)	0.00724 (2)	1.01537 (3)	1.01537 (3)	1.01613 (2)	1.01613 (2)	1.01613 (2)	125000
50000	1.01621 (2)	0.00237 (2)	1.01604 (3)	1.01605 (2)	1.01613 (2)	1.01613 (2)	1.01613 (2)	12500
S_{TI}	1.016145 (3)		1.016145 (3)	1.016145 (3)	1.016145 (3)	1.016145 (3)	1.016145 (3)	
S_{series}	1.01615 (1)		1.01615 (1)	1.01615 (1)	1.01615 (1)	1.01615 (1)	1.01615 (1)	
				$N = 49$ $S_{SCAN} = 1.000904\ (4)$				
500	1.00959 (2)	0.01923 (2)	0.99215 (7)	0.99146 (4)	1.00087 (4)	1.00053 (3)	1.00078 (4)	1248547
5000	1.00172 (2)	0.00600 (2)	0.99996 (5)	0.99995 (2)	1.00084 (2)	1.00084 (2)	1.00084 (3)	124763
50000	1.00094 (2)	0.00189 (2)	1.00077 (4)	1.00077 (2)	1.00086 (2)	1.00086 (2)	1.00086 (3)	12467
S_{HS}	1.00149 (1)	0.00434 (1)	1.00026 (2)	1.00057 (1)	1.00088 (2)	1.00103 (1)	1.00094 (1)	250000
S_{TI}	1.000897 (3)		1.000897 (3)	1.000897 (3)	1.000897 (3)	1.000897 (3)	1.000897 (3)	
S_{series}	1.000899 (5)		1.000899 (4)	1.000899 (4)	1.000899 (4)	1.000899 (4)	1.000899 (4)	
				$N = 99$ $S_{SCAN} = 0.987726\ (5)$				
500	0.99840 (3)	0.01539 (3)	0.9762 (1)	0.9750 (1)	0.9873 (1)	0.98668 (5)	0.9873 (1)	249621
5000	0.98883 (3)	0.00478 (3)	0.9866 (1)	0.98657 (4)	0.98770 (5)	0.98770 (3)	0.98770 (5)	24907
50000	0.98786 (3)	0.00153 (3)	0.98763 (5)	0.98763 (3)	0.98775 (3)	0.98775 (3)	0.98775 (3)	2476
S_{HS}	0.98994 (1)	0.00507 (1)	0.9856 (2)	0.9874 (1)	0.9878 (1)	0.9887 (1)	0.98817 (5)	250000
S_{TI}	0.987727 (3)		0.987727 (3)	0.987727 (3)	0.987727 (3)	0.987727 (3)	0.987727 (3)	
S_{series}	0.987730 (3)		0.987730 (3)	0.987730 (3)	0.987730 (3)	0.987730 (3)	0.987730 (3)	
				$N = 149$ $S_{SCAN} = 0.982740\ (3)$				
500	0.99460 (3)	0.01347 (5)	0.9688 (5)	0.9676 (2)	0.9817 (3)	0.9811 (1)	0.9818 (3)	249628
5000	0.98398 (3)	0.00428 (5)	0.9813 (1)	0.98126 (7)	0.98263 (5)	0.98262 (4)	0.98264 (5)	24860
50000	0.98293 (3)	0.00143 (5)	0.98262 (5)	0.98262 (4)	0.98277 (3)	0.98277 (3)	0.98277 (4)	2470
S_{TI}	0.982742 (3)		0.982742 (3)	0.982742 (3)	0.982742 (3)	0.982742 (3)	0.982742 (3)	
S_{series}	0.982740 (2)		0.982740 (2)	0.982740 (2)	0.982740 (2)	0.982740 (2)	0.982740 (2)	
				$N = 249$ $S_{SCAN} = 0.97836\ (2)$				
500	0.99188 (5)	0.01149 (7)	0.961 (2)	0.9590 (4)	0.976 (1)	0.9755 (2)	0.976 (1)	50451
5000	0.97977 (4)	0.00374 (7)	0.9760 (2)	0.9763 (1)	0.9779 (1)	0.97803 (8)	0.9780 (1)	7261
50000	0.97851 (4)	0.00129 (7)	0.9781 (1)	0.97809 (6)	0.97830 (5)	0.97830 (5)	0.97830 (5)	938
S_{HS}	0.98306 (1)	0.00401 (1)	0.9745 (5)	0.9791 (3)*	0.9788 (3)	0.9811 (2)	0.9799 (1)	176000
S_{TI}	0.978358 (4)		0.978358 (4)	0.978358 (4)	0.978358 (4)	0.978358 (4)	0.978358 (4)	
S_{series}	0.978360 (1)		0.978360 (1)	0.978360 (1)	0.978360 (1)	0.978360 (1)	0.978360 (1)	
				$N = 399$ $S_{SCAN} = 0.97567\ (4)$				
500	0.9908 (1)	0.0099 (1)	0.955 (3)	0.9517 (8)	0.973 (2)	0.9712 (4)	0.971 (2)	6670
5000	0.97729 (8)	0.0032 (1)	0.9727 (5)	0.9733 (3)	0.9750 (3)	0.9753 (2)	0.9752 (3)	1577
50000	0.9759 (1)	0.0012 (1)	0.9754 (2)	0.9754 (1)	0.9757 (1)	0.9757 (1)	0.9757 (1)	115
S_{HS}	0.98141 (5)	0.00335 (5)	0.9743 (5)	0.9769 (3)	0.9779 (3)	0.9792 (2)	0.9782 (2)	5500
S_{TI}	0.975655 (8)		0.975655 (8)	0.975655 (8)	0.975655 (8)	0.975655 (8)	0.975655 (8)	
S_{series}	0.975652(1)		0.975652 (1)	0.975652 (1)	0.975652 (1)	0.975652 (1)	0.975652 (1)	
				$N = 599$ $S_{SCAN} = 0.97395\ (5)$				
500	0.9904 (2)	0.0087 (2)	0.957 (5)	0.945 (2)	0.974 (3)	0.968 (1)	0.969 (3)	2540
5000	0.9760 (2)	0.0030 (2)	0.970 (2)	0.971 (2)	0.973 (3)	0.9733 (4)	0.973 (1)	316
50000	0.9743 (1)	0.0010 (2)	0.9738 (5)	0.9737 (3)	0.9741 (3)	0.9740 (2)	0.9741 (3)	60
S_{TI}	0.97404 (1)		0.97404 (1)	0.97404 (1)	0.97404 (1)	0.97404 (1)	0.97404 (1)	
S_{series}	0.974025(1)		0.974025 (1)	0.974025 (1)	0.974025 (1)	0.974025 (1)	0.974025 (1)	