Entropic sampling of simple polymer models within Wang–Landau algorithm

2004 J. Phys. A: Math. Gen. 37 1573

(http://iopscience.iop.org/0305-4470/37/5/008)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **37** (2004) 1573–1588 PII: S0305-4470(04)65229-2

Entropic sampling of simple polymer models within Wang–Landau algorithm

P N Vorontsov-Velyaminov, N A Volkov and A A Yurchenko

Faculty of Physics, St Petersburg State University, 198504, St Petersburg, Russia

E-mail: voron.wgroup@pobox.spbu.ru

Received 24 June 2003
Published 19 January 2004
Online at stacks.iop.org/JPhysA/37/1573 (DOI: 10.1088/0305-4470/37/5/008)

Abstract

In this paper we apply a new simulation technique proposed in Wang and Landau (WL) (2001 *Phys. Rev. Lett.* **86** 2050) to sampling of three-dimensional lattice and continuous models of polymer chains. Distributions obtained by homogeneous (unconditional) random walks are compared with results of entropic sampling (ES) within the WL algorithm. While homogeneous sampling gives reliable results typically in the range of 4–5 orders of magnitude, the WL entropic sampling yields them in the range of 20–30 orders and even larger with comparable computer effort. A combination of homogeneous and WL sampling provides reliable data for events with probabilities down to 10^{-35}.

For the lattice model we consider both the athermal case (self-avoiding walks, SAWs) and the thermal case when an energy is attributed to each contact between nonbonded monomers in a self-avoiding walk. For short chains the simulation results are checked by comparison with the exact data. In WL calculations for chain lengths up to $N = 300$ scaling relations for SAWs are well reproduced. In the thermal case distribution over the number of contacts is obtained in the N-range up to $N = 100$ and the canonical averages—internal energy, heat capacity, excess canonical entropy, mean square end-to-end distance—are calculated as a result in a wide temperature range.

The continuous model is studied in the athermal case. By sorting conformations of a continuous phantom freely joined N-bonded chain with a unit bond length over a stochastic variable, the minimum distance between nonbonded beads, we determine the probability distribution for the N-bonded chain with hard sphere monomer units over its diameter a in the complete diameter range, $0 \leqslant a \leqslant 2$, within a *single* ES run. This distribution provides us with excess specific entropy for a set of diameters a in this range. Calculations were made for chain lengths up to $N = 100$ and results were extrapolated to $N \to \infty$ for a in the range $0 \leqslant a \leqslant 1.25$.

PACS numbers: 05.40.Fb, 05.50.+q, 05.10.Ln, 82.35.−x

1. Introduction

The Monte Carlo (MC) method introduced fifty years ago by Metropolis *et al* [1] proved to be a powerful tool in studying a large variety of highly nonideal molecular systems [2, 3]. It appeared though that the conventional MC procedure becomes ineffective or even fails in a number of important physical situations. In order to calculate free energy or entropy of a condensed system, to study systems having rough multiminima potential landscapes (clusters, glasses, protein molecules), to treat phase transitions and other phenomena occurring at low temperatures, high densities or in the presence of complicated molecular components, one had to modify standard approaches. Such modifications proposed about ten years ago led to methods known now under the common name of generalized ensembles MC (see reviews [4, 5]). They are expanded ensemble MC or simulated tempering [6, 7] and multicanonic or entropic sampling [8, 9]. These methods proved to be efficient in solving problems mentioned above though they have their own common drawback, i.e. they require a preliminary adjustment of a set of parameters ('balancing factors' [6]) playing a key role in simulations. The parameters are initially unknown and are usually obtained iteratively in preliminary MC runs. This iterative stage sometimes is referred to as 'tedious and time consuming'. A replica exchange method suitable for parallel calculations [5, 10] provided one way to avoid this preliminary stage. Another way recently suggested by Wang and Landau (WL) in [11, 12] allows us to alleviate this drawback in the entropic sampling MC with the aid of a special self-adjusting procedure. The authors applied their method to simulate lattice spin systems (Ising and Pott's models) though the WL algorithm can be readily implemented in a large variety of molecular systems, thus being of general importance. Since its formulation in 2001, the WL algorithm has been promptly applied to simulation of a lattice model of proteins [13], of a polymer film [14], of fluids [15, 16] and in path-integral Monte Carlo [17].

The aim of this paper is to extend the WL algorithm to entropic sampling (ES) simulations of polymers starting with simple three-dimensional models. In order to explore the facilities of the method we consider a chain on a simple cubic (SC) lattice and a freely joined continuous model with rigid unit length bonds. The lattice model is considered both in the athermal case (the self-avoiding walk (SAW)) and the thermal case (when interactions of nonbonded monomers occurring in contact are accounted for). For the athermal case our numerical results are compared with scaling relations for SAWs [18, 19]. In the thermal case we compare our results with the data of Douglas *et al* [20] and Grassberger and Hegger [21]. For short chains ($N \sim 10$) the numerical result can be compared with the exact data obtained by consecutive enumeration of all possible lattice conformations. Obtaining the distribution of SAWs over the number of contacts enables us to calculate the temperature dependences of conformational internal energy, heat capacity, entropy, mean square end-to-end distance and the expansion factor for different chain lengths N.

The continuous model is studied only in the athermal case when monomer units are represented by hard spheres. By sorting conformations of a phantom chain over a specially introduced stochastic variable, the minimum distance between nonbonded beads, we determine the distribution for a free joined chain with hard sphere monomers over their diameter a within a single MC run.

It is important to stress here that a huge amount of analytical and simulation work has already been done so far in studying various types of polymer systems and particularly lattice models (e.g., [22–24] and references therein). A lot of simulation techniques were proposed mostly stemming from the early work of Rosenbluth and Rosenbluth [25]. Thus, the pruned-enriched Rosenbluth method developed by Grassberger [26] allows us to simulate very long

lattice chains—up to $N = 10^6$. The approach presented below in this paper could hopefully add some new opportunities to simulation techniques suggested and used so far.

The plan of the paper is as follows. In section 2 we discuss the general idea of the entropic sampling within the WL algorithm. An approach to the simulation of the lattice model in both cases, the obtained results and their treatment is given in section 3. In section 4 the WL algorithm is applied to the continuous model. A method providing calculation of excess entropy as a function of the diameter of monomer units in a single MC run is suggested and carried out. Section 5 contains final remarks.

2. ES method within Wang–Landau algorithm

Consider a canonical configurational partition function in the form:

$$Z = \int \mathrm{d}E\,\Omega(E)\exp(-\beta E) \qquad \Omega(E) = \int \mathrm{d}q\,\delta(E - H(q)) \tag{1}$$

where $\Omega(E)$ is the density of energy states, $\beta = T^{-1}$ is the inverse temperature (energy units) and q is the set of coordinates.

A step within the conventional metropolis MC method includes a uniform choice of a trial microstate in the coordinate space (q) which provides energy distribution $\Omega(E)$ and imposing an additional condition which accounts for the canonical factor, $\exp(-\beta H(q))$. As a result of a MC run based on this procedure one finally gets the distribution in energy: $p(E) = \mathrm{const}\cdot\Omega(E)\exp(-\beta E)$. Its particular case is that obtained in a purely unconditional random walk (URW) yielding $p(E) = \Omega(E)$; formally it corresponds to the one above at infinite temperature, $\beta = 0$.

On the other hand, if we arrange the random walk so that the additional factor (instead of the canonical probability $\exp(-\beta H(q))$) is chosen to be the inverse of $\Omega(E)$, i.e. $w(E(q)) = [\Omega(E(q))]^{-1}$, then the 'natural' weight $\Omega(E)$ is being completely compensated and the distribution $p(E)$ becomes homogeneous ('flat') [9].

As long as $\Omega(E)$ is initially unknown, being the aim of the ES calculations, an iterative procedure becomes necessary. Obtaining nearly flat $p(E)$ dependence testifies that the iterated function $w(E)$ is close to the required function. An effective means of obtaining flat distribution $p(E)$ was suggested recently by Wang and Landau (WL) in papers [11, 12]. The WL algorithm can be considered as a self- (or auto-) adjusting procedure for obtaining $\Omega(E)$ in ES simulations.

The energy range of interest, $E_{\min} \leqslant E \leqslant E_{\max}$, is divided into a finite number, N_b, of equal intervals ('boxes'), $\Delta E = \frac{(E_{\max}-E_{\min})}{N_b}$, and all the initial values of Ω_i corresponding to these boxes ($1 \leqslant i \leqslant N_b$) are taken to be equal (e.g., in [11], initially $\Omega_i = 1$). In order to avoid processing with large numbers it is convenient to introduce entropy distribution $S_i = \ln\Omega_i$ (in [11] initially $S_i = 0$). Two sets of counters of length N_b are introduced: one accumulates S_i (entropy boxes); another, n_i, counts visits to boxes yielding at the end of the run the normalized visit probabilities p_i (the histogram). Each configuration of the simulated system has its definite value of energy and hence it belongs to one of the N_b boxes.

A MC step includes a standard trial change of the state with a uniform coordinate distribution and further imposing of the following transition probability condition:

$$p(i \to i') = \min[1, \exp(S_i - S_{i'})]. \tag{2}$$

If this condition is fulfilled the trial state (i') is accepted; in the opposite case the accepted one is the initial state (i). Finally, the entropy of the accepted state (S_i or $S_{i'}$) is augmented by ΔS and the corresponding counter of visits (n_i or $n_{i'}$) is increased by 1.

A series of such elementary steps constitute a sweep. At the end of a current sweep ΔS is decreased: $\Delta S \to c\Delta S$ with an increment of $0 < c < 1$ (in [11], the initial value of ΔS is 1 and $c = 0.5$ though other values could also be used).

After several sweeps, the S_i dependence is formed and finetuned in the whole range of E. Further continuing the procedure results only in the addition of a constant to the S_i-dependence. It is accompanied by simultaneous formation of a flat histogram p_i.

This way the normalized density of states $\Omega(E) = \exp(S(E))$ can be calculated rather accurately in a very wide range of orders of magnitude, e.g., from 10^{-1} to 10^{-30} or even to lower orders (see below). It provides calculation of the canonical partition function (1) and averages, such as configurational energy and heat capacity, in a wide temperature range by numerical integration.

3. Lattice model

Within the lattice model a polymer chain is presented by a random walk (RW) on a d-dimensional lattice starting at the origin. The total number of conformations for a phantom chain with $N + 1$ monomers and N bonds is z^N, where z is the number of nearest neighbours of the lattice. For a 3D SC lattice which is considered below $z = 6$.

The effect of excluded volume for the *athermal* case implies excluding all RWs with intersections (overlaps) leaving only SAWs. The number of SAWs is determined by the asymptotic scaling relation [18, 20]:

$$W_N = A\mu^N N^{\gamma-1}. \tag{3}$$

For the SC lattice connectivity constant $\mu = 4.6838$, susceptibility $\gamma = 7/6$ and $A = 1.17$ [18, 27].

The *thermal* case also includes interactions of nonbonded monomers when they approach each other. The simplest scheme, which is usually considered, is an account of interactions of nonbonded monomers of a SAW occurring at closest contact.

In our simulations, a lattice polymer chain is presented by an array of integer numbers out of the set $(1, 2, \ldots, 6)$ pointing the direction of each segment in the SC lattice. One end of the chain (the 0th bead) is kept fixed at the origin and the initial directions of all consecutive segments are chosen randomly. In order to change the chain's conformation we choose homogeneously one of the beads from 0 to $N - l$ (e.g., the k_0th) and change randomly the directions of segments between beads k_0 and $k_0 + l$. We used values of l, the number of segments in the modified piece, in the range of $[N/5]$–$[N/20]$. The remaining piece, $l < k \leqslant N$, undergoes a parallel shift. Other ways of changing the conformation could also be possibly applied here, e.g., the pivot algorithm [23], but in this work we used only the above described method.

In most cases WL calculations included 25–30 sweeps with initial values of $S_i = 0$, $\Delta S = 1$ and the increment $c = 0.5$. The length of a sweep depended on the chain length N and varied from 1 to 5 million steps for $N = 12$–300. For different chain lengths the simulation time on Pentium III 1 GHz computer varied between 15 min and 30 h.

3.1. Athermal case

For the athermal case our general goal is to calculate the number of SAWs for different N using the WL algorithm and compare them with (3).

Our first concern in this context was to study convergence of the WL algorithm and to test its facilities when applied to the lattice polymer. It is worth pointing out here that it is not necessary to associate the entropy boxes introduced above with energy intervals. For

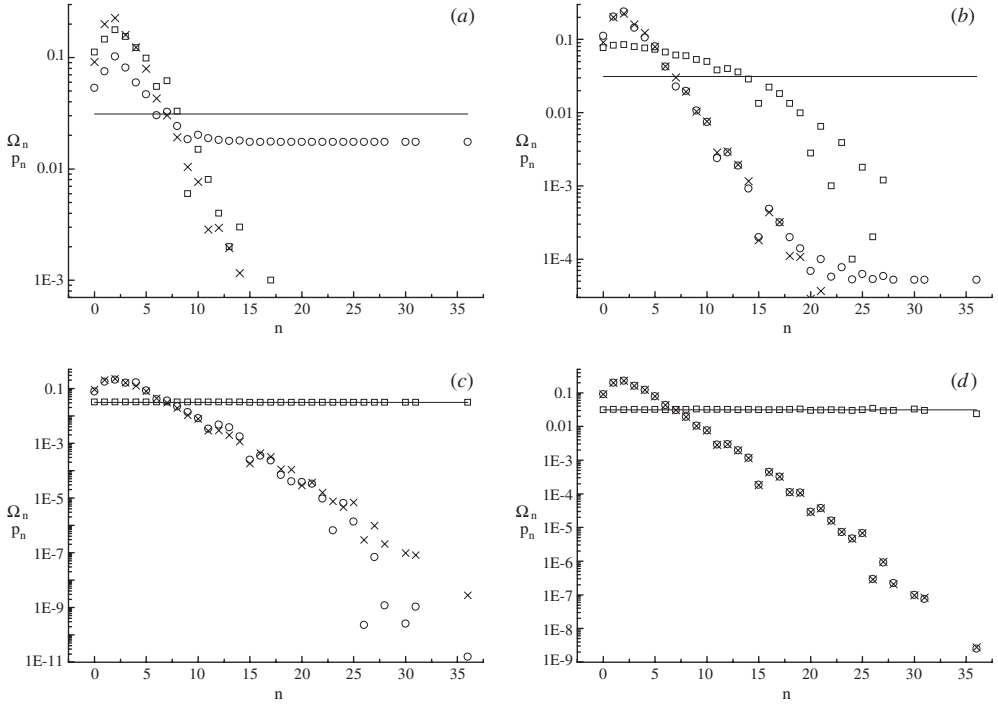

Figure 1. WL evolution of the normalized n-distribution Ω_n (circles) and visit rates p_n (squares) for a phantom lattice chain, $N = 12$, compared with exact values for Ω_n (crosses) and with prescribed level for p_n (horizontal line) after 1000 (*a*), 10 000 (*b*), 3 million (*c*), 240 million (*d*) steps; n—number of intersections.

instance, we can relate the entropy boxes to the number of intersections n for an N-bonded phantom chain, $0 \leqslant n \leqslant n_{\max}$. Its maximum number, $n_{\max}$, for given N can be determined exactly through binomial coefficients: $n_{\max}(N) = C^2_{\frac{N}{2}} + C^2_{\frac{N}{2}+1} = (N/2)^2$ for even N and $n_{\max}(N) = 2C^2_{\frac{N+1}{2}} = (N+1)(N-1)/4$ for odd N. The number of conformations with $n = n_{\max}$ is 6 irrespective of N, so the corresponding probability is $6^{-(N-1)}$.

The convergence of the WL algorithm for $N = 12$ is presented in figure 1. It is clearly seen how the normalized probabilities $\Omega_n, 0 \leqslant n \leqslant n_{\max}$, starting from equal initial values $\Omega_n = 1/N_b$, are gradually modified and finally attain values which do not change further and which coincide with the exact data. The number of boxes, N_b, in this case is equal to 32 though the maximum number of intersections is $n_{\max} = (N/2)^2 = 36$. This is due to the fact that five numbers of intersections, i.e. $n = 29, 32$–35, cannot occur at all for $N = 12$. Figure 1 illustrates the simultaneous evolution of the visit rates. After about 10^6 steps all the probabilities p_n are stabilized at the prescribed level, $1/N_b = 1/32 = 0.031\,25$, and hold at this level during further WL runs (mean deviation is about 3%).

We carried out calculations of normalized probabilities Ω_n within the WL algorithm for $N = 12, 30, 50$ and compared them with those obtained within URW, figure 2. First of all a very good agreement is observed between both results around the maximum of distributions within five orders of magnitude (from 10^{-1} to 10^{-6}), the discrepancy here does not exceed several per cent, figure 2(*b*). Out of this range URW results start displaying great scatter and then the URW completely fails while WL data steadily proceed to much

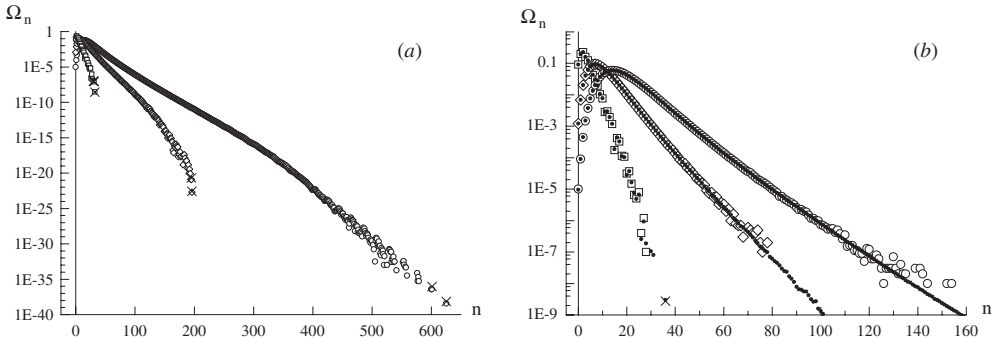

Figure 2. Normalized distributions Ω_n for phantom lattice chains (n—number of intersections). (*a*) WL data for N: 12 (squares), 30 (diamonds), 50 (circles). (*b*) Comparison of URW results for N: 12 (squares), 30 (diamonds), 50 (circles) with WL data (dots for all N). Crosses denote exact values.

lower orders which is clearly seen in figure 2(*a*). It is important to note that the WL algorithm reproduces sufficiently well values of n-probabilities for the maximum number of intersections $\Omega_{n_{\max}}$. For $N = 12, 30, 50$, $n_{\max} = 36, 225, 625$ and exact values are: $\Omega_{n_{\max}} = 6^{-11} = 2.76 \times 10^{-9}$; $6^{-29} = 2.71 \times 10^{-23}$; $6^{-49} = 7.42 \times 10^{-39}$, against the obtained WL values 2.99×10^{-9}; 1.92×10^{-23}; 3.61×10^{-39}, respectively. It should also be mentioned that two numbers of intersections preceding $n_{\max}$ can also be easily determined exactly, for even N they are: $n_1 = \left(\frac{N}{2}\right)^2 - \frac{N}{2} + 1$ and $n_2 = \left(\frac{N}{2}\right)^2 - \frac{N}{2}$. It yields 31, 30; 211, 210; 601, 600 for $N = 12, 30, 50$, respectively. Their exact probabilities are: $\Omega_{n_1} = 6^{(N-1)} \times 5 \times \frac{N}{2}$ and $\Omega_{n_2} = 6^{(N-1)} \times 5 \times \left(\frac{N}{2} + 1\right)$. These points are also reproduced sufficiently well in WL calculations, figure 2(*a*).

For short chains, e.g., $N = 12$, it is possible to compare the numerical results directly with exact data obtained by consecutive enumeration of all z^N conformations. Such comparison showed that the WL algorithm reproduces exact data for Ω_n in the whole range of n with the mean error about several per cent including rare events (e.g., ∼8% for $\Omega_n \sim 10^{-9}$). URW reproduces well values of Ω_n in the range $\sim 10^{-1}$–10^{-5}, with the scatter growing to 50% for $\Omega_n \sim 10^{-6}$–10^{-7} and being a complete failure when $\Omega_n \sim 10^{-8}$–10^{-9}.

For calculating probabilities of SAWs, Ω_0, there exist several opportunities. We can use the WL algorithm with all numbers of intersections taken as entropy boxes. That was the procedure in the cases described above. But with increasing N the number of boxes would increase approximately as $n_{\max} = (N/2)^2$ and can become prohibitively great. Another extreme possibility is to consider only two boxes: one for nonintersecting chains (SAWs) and the other for all the other conformations. This scheme was tested and proved to be reliable only for $N \leqslant 30$. For greater N, the WL scheme with two boxes became unstable: it was impossible to equilibrate the rates of visits of both boxes.

It appeared that the most reliable results can be obtained in a procedure combining WL and URW algorithms. The calculation starts with an URW during which the share of conformations with $n \leqslant n_b$, $\Pr(0, n_b)$, is calculated. If the boundary number of intersections, n_b, is chosen so that this share is several tens of per cent then it can be calculated in an URW rather accurately. The second step is a WL run within the restricted range of n, $0 \leqslant n \leqslant n_b$. If n in a trial state within the WL process exceeds the boundary value n_b this trial is rejected (as a rule the rate of such rejections did not exceed 5%). Thus, the whole WL process goes within the restricted range of n including $n = 0$ which is the item of our concern. Comparison of n_b with $n_{\max}$,

Table 1. Scaling and WL entropies S, specific excess scaling and WL entropies $\Delta S/N$ and their relative deviation δ, root mean square end-to-end distance $\sqrt{\langle R^2_{\mathrm{SAW}}\rangle}$ for self-avoiding conformations, expanding factor α, $\alpha/N^{\frac{1}{10}}$, n_b, $n_{\max}$ and $\Pr(0, n_b)$, for $N = 12, 30, 100, 150, 200, 300$.

N	S_{SC}	S_{MC}	$-\frac{\Delta S_{\mathrm{SC}}}{N}$	$-\frac{\Delta S_{\mathrm{MC}}}{N}$	δ (%)	$\sqrt{\langle R^2_{\mathrm{SAW}}\rangle}$	a	$\frac{\alpha}{N^{\frac{1}{10}}}$	n_b	$n_{\max}$	$\Pr(0, b_b)$
12	19.099	19.105	0.2002	0.1996	0.25	4.5688	1.3189	1.028 70	9	36	0.98
30	47.044	46.993	0.2236	0.2253	0.76	7.9000	1.4423	1.026 49	15	225	0.88
100	155.32	155.28	0.2385	0.2389	0.17				49	2 500	0.81
150	232.59	232.54	0.2411	0.2415	0.14	20.7167	1.6915	1.024 86	69	5 625	0.70
200	309.84	309.82	0.2426	0.2427	0.04				79	10 000	0.43
300	464.31	464.08	0.2441	0.2448	0.32	31.4031	1.8131	1.024 94	149	22 500	0.75

table 1, shows that the range of n used in our WL calculations is a minor part of the whole n-range, especially for greater chain lengths N. Including an optimal number of intersections n_b into the WL process makes the simulations both stable and computer time saving. The WL-procedure allows normalized probability of SAWs Ω'_0 to be obtained within this limited range of n while the initially obtained $\Pr(0, n_b)$, the probability of the whole area $0 \leqslant n \leqslant n_b$, provides the total statistical weight of SAWs, $\Omega_0 = \Omega'_0 \times \Pr(0, n_b)$.

Thus, results for $N = 12, 30, 100, 150, 200, 300$ were obtained, table 1. The first two columns, S_{SC} and S_{WL}, give entropy according to scaling relation (3) and to WL simulation, $S_N = \ln W_N$; here $W_N = 6^N \Omega_0$ and Ω_0 is the normalized probability of SAW for each N. The next two columns present excess specific entropies $\frac{1}{N}\Delta S_N = \frac{1}{N}\ln \Omega_0$. Good agreement between scaling- and WL data is observed: deviations for specific values (next column) do not exceed 1%.

Three further columns in table 1 present the root mean square end-to-end distance, $\langle R^2_{\mathrm{saw}}\rangle^{\frac{1}{2}}$, expansion factor $\alpha = \frac{1}{\sqrt{N}}\langle R^2_{\mathrm{saw}}\rangle^{\frac{1}{2}}$ and $\alpha/N^{1/10}$. It is seen that the values of the latter remain constant with a scatter about 0.5% throughout the whole range of N which corresponds to the scaling relation for SAW: $\sqrt{\langle R^2_{\mathrm{saw}}\rangle} \sim N^{3/5}$ [18, 19]. We tested other estimates of the factor ν (instead of $\nu = 3/5 = 0.6$), i.e. 0.592 from [23, 28] and 0.588 obtained in [29], but neither of them gave (in the limited range of N studied in our work) better convergence than for $\nu = 0.6$.

The last three columns in table 1 present n_b, $n_{\max}$ and $\Pr(0, n_b)$.

3.2. Thermal case

In order to account for interactions of nonbonded monomers, we attribute an energy ε to each monomer pair occurring at closest contact. The problem now is to distribute all SAWs among boxes corresponding to various numbers of contacts, m, which can vary between zero and a certain finite number $m_{\max}$ dependent on N. As long as the energy of a configuration E_m is the product $\varepsilon \times m$, this distribution yields the probability of energy states $\Omega(E_m) \equiv \Omega_{0m}$. The maximum value of m cannot be obtained as exactly as it was in the case of maximum number of intersections discussed above. In [20] the upper bound of $m_{\max}$ for the SC lattice is estimated in the range of chain lengths up to $N = 50$.

We used two procedures to sort SAWs due to the number of contacts m. In both of them the whole normalized contribution of SAWs, Ω_0, was determined within the WL algorithm. In the first procedure all the occurring SAWs were sorted over m unconditionally. In another

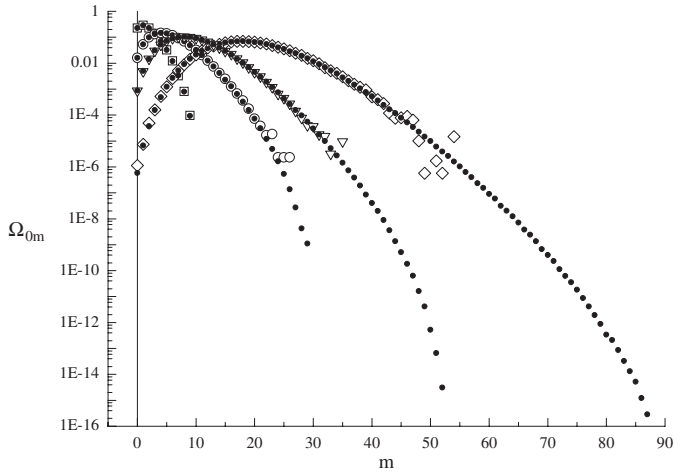

Figure 3. Normalized distributions for SAWs over the number of contacts m, Ω_{0m}, obtained by the URW algorithm for $N = 12$ (squares), 30 (circles), 50 (down triangles), 100 (diamonds) and by the WL algorithm (dots for all N).

approach the distribution over m was carried out within a WL procedure. The results of both approaches for normalized probabilities, Ω_{0m}, are presented in figure 3. Analogous to the case of intersections in figure 2, we observe good coincidence of both procedures around the distribution maxima in the range of five orders of magnitude. Out of this range the unconditional sorting exhibits great scatter and then completely fails while the WL results steadily follow down to considerably lower orders, for $N = 100$—down to 10^{-16}. It should also be stressed here that these distributions should be additionally multiplied by the probabilities of SAWs Ω_0: for $N = 12, 30, 100$ $\Omega_0 = 0.091, 0.0012, 3.9 \times 10^{-11}$, respectively. It could also be mentioned here that values for which Ω_{0m} have maxima for each N, figure 3, coincide with those of [20] (figure 13(a) in [20]). For $N = 12, 30, 50$ they are $m = 1, 4, 8, 16$.

For a short chain, $N = 12$, we again check our numerical results by comparing them with the exact data from enumeration of all z^N configurations, figure 4. Good coincidence is observed both for probabilities Ω_{0m} and for mean square end-to-end distances $\langle R\rangle_m^2$ (see below). We also compared our WL data obtained for $N = 18$ with those of [20] for exact numbers of SAWs with m contacts (table 1 of [20]). The accuracy of our WL calculations is comparable with that of the Rosenbluth MC data [20] in the whole m-range.

The obtained distributions Ω_{0m}, figure 3, can now be used for calculating canonical averages according to standard relations. For internal energy it reads

$$\langle E\rangle(\beta) = \frac{\sum_{m=0}^{m_{\max}} m\varepsilon \exp^{-\beta m\varepsilon}\, \Omega_{0m}}{\sum_{m=0}^{m_{\max}} \exp^{-\beta m\varepsilon}\, \Omega_{0m}} \equiv \varepsilon\langle m\rangle_{\text{can}} \tag{4}$$

$\langle E^2\rangle$ is calculated in the same way yielding the heat capacity as a function of temperature:

$$C(T) = \frac{1}{T^2}(\langle E^2\rangle(T) - (\langle E\rangle(T))^2). \tag{5}$$

In the expression for the energy (4), the normalization of Ω_{0m} is not important since Ω_{0m} enters both numerator and denominator in (4) and any constant cancels. In order to calculate free energy and entropy we must use instead the quantity $z^N\Omega_0\Omega_{0m}$ which is equal

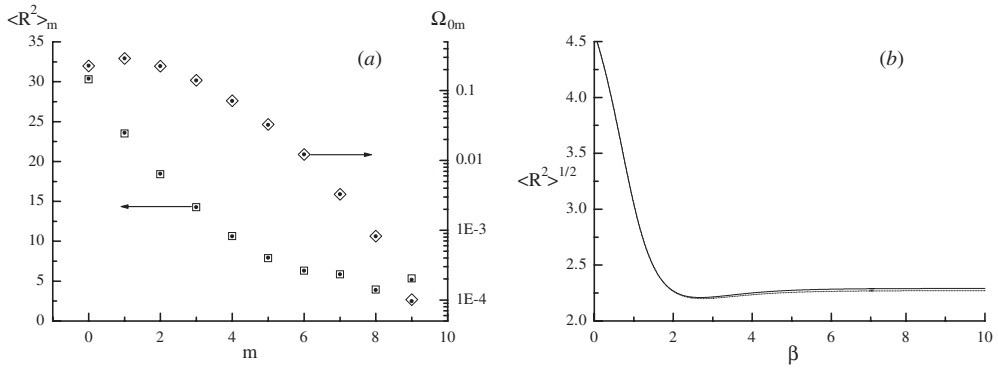

Figure 4. Comparison of WL data with exact values for SAWs, $N = 12$. (*a*) $\langle R^2\rangle_m$ and Ω_{0m} dependences on the number of contacts m; squares and diamonds are for WL data and dots are for exact values. (*b*) Root mean square end-to-end distance $\sqrt{\langle R^2\rangle}$ as a function of inverse temperature β (ε^{-1}-units) according to equation (9); WL data—full curve, exact—dots.

to the complete number of SAWs with m contacts for a chain with N bonds (both Ω_0 and Ω_{0m} depend on N). So for free energy F we have

$$-\beta F = \ln \sum_{m=0}^{m_{\max}} \exp(-\beta\varepsilon m)\Omega_0\Omega_{0m} z^N = N \ln z + \ln \Omega_0 + \ln \sum_{m=0}^{m_{\max}} \exp(-\beta\varepsilon m)\Omega_{0m}. \tag{6}$$

Now the canonical entropy is expressed as

$$S(\beta) = \beta\langle E\rangle - \beta F = S_0 + \Delta S_{\text{at}} + \Delta S(\beta) \tag{7}$$

where $S_0 = N \ln z$ is the entropy of a phantom chain; $\Delta S_{\text{at}} = \ln \Omega_0$ is the excess entropy for the athermal case (table 1)

$$\Delta S(\beta) = \beta\langle E\rangle + \ln \sum_{m=0}^{m_{\max}} \exp(-\beta\varepsilon m)\Omega_{0m} \tag{8}$$

is the excess canonical entropy (it vanishes for $\varepsilon = 0$ or $\beta = 0$).

If a certain quantity is not uniquely determined by m we can average it over m for each value of the set $0 \leqslant m \leqslant m_{\max}$. Thus for the square end-to-end distance of the chain, R^2, we calculate averages $\langle R^2\rangle_m$ within the WL procedure and finally get canonical averages in a standard way similar to (4):

$$\langle R^2\rangle(\beta) = \langle\langle R^2\rangle_m\rangle_{\text{can}}. \tag{9}$$

The canonical mean square radius of inertia, $R_I^2 = \frac{1}{2N^2}\sum_{(i,j)} r_{ij}^2$, can be calculated in the same way: $\langle R_I^2\rangle(\beta) = \langle\langle R_I^2\rangle_m\rangle_{\text{can}}$.

Averaging data according to (4), (5), (8) and (9) for $N = 12, 30, 50$ are presented in figures 5–7. Specific energies and heat capacities as functions of T are shown for attractive ($\varepsilon < 0$) and repulsive ($\varepsilon > 0$) interaction in figures 5(*a*) and (*b*). For repulsive interaction, the specific energy approaches zero as $T \to 0$ (neighbour avoiding walks) and tends to $\langle m\rangle_{\text{at}}(N)/N$ of the athermal case at $T \to \infty$ $\left(\langle m\rangle_{\text{at}}(N) = \sum_{m=0}^{m_{\max}} m\Omega_{0m}\right)$. For $\varepsilon < 0$ and $T \to \infty$ the specific energy tends to $-\langle m\rangle_{\text{at}}(N)/N$ while for $T \to 0$ it approaches the ground states, $-m_{\max}(N)/N$. Levels of $\langle m\rangle_{\text{at}}(N)/N$ are: 0.142, 0.174, 0.183 for $N = 12$, 30, 50. These values evidently tend to some limit with increasing N. In [20] this limit is estimated as 0.193 (figure 10(*a*) in [20]) and it conforms to our results. In figure 5(*a*) we also compare

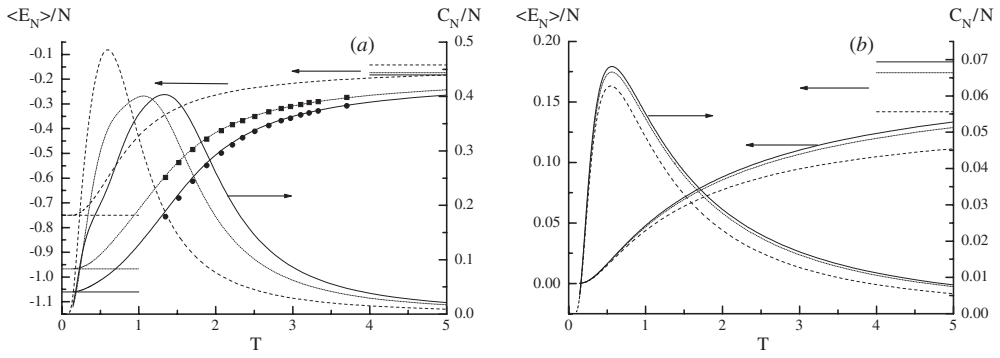

Figure 5. Specific energies (ε-units), equation (4), and heat capacities, equation (5), as functions of T (ε-units) for $N = 12$ (broken curve), 30 (dotted curve), 50 (full curve); (*a*) attractive case $\varepsilon < 0$; (*b*) repulsive case $\varepsilon > 0$. Asymptotes for energy at $T \to \infty$ are shown on the right-hand side of (*a*) and (*b*) and for energy ($\varepsilon < 0$) at $T \to 0$—on the left-hand side of (*a*) (the notation is the same). Full squares and circles are energy data for $N = 30, 50$ derived from [21].

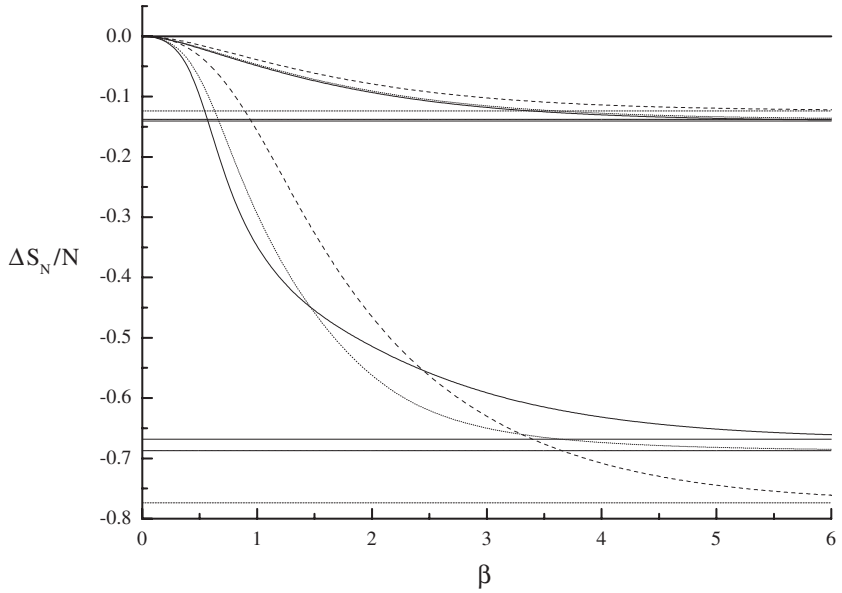

Figure 6. Dependences of specific excess canonical entropy $\Delta S_N/N$ on inverse temperature β (ε^{-1}-units) and their asymptotes at $\beta \to \infty$ for $\varepsilon > 0$ and $\varepsilon < 0$; $N = 12$ (broken curve), 30 (dotted curve), 50 (full curve).

our energy data for $N = 30, 50$ with those scanned from figure 4 of [21]. Almost complete coincidence of both results is observed with a slight deviation at lowest temperatures for $N = 50$.

Specific heat capacities determined for $\varepsilon > 0$, figure 5(*b*), clearly approach some limiting curve with increasing N while for $\varepsilon < 0$ such tendency is not so clear from the presented data.

Figure 6 presents specific excess canonical entropy (8) as a function of β both for $\varepsilon > 0$ and $\varepsilon < 0$. All curves start at zero for $\beta = 0$ since infinite temperature corresponds to the athermal case. Excess entropy is negative as far as interactions augment ordering of the chain in both

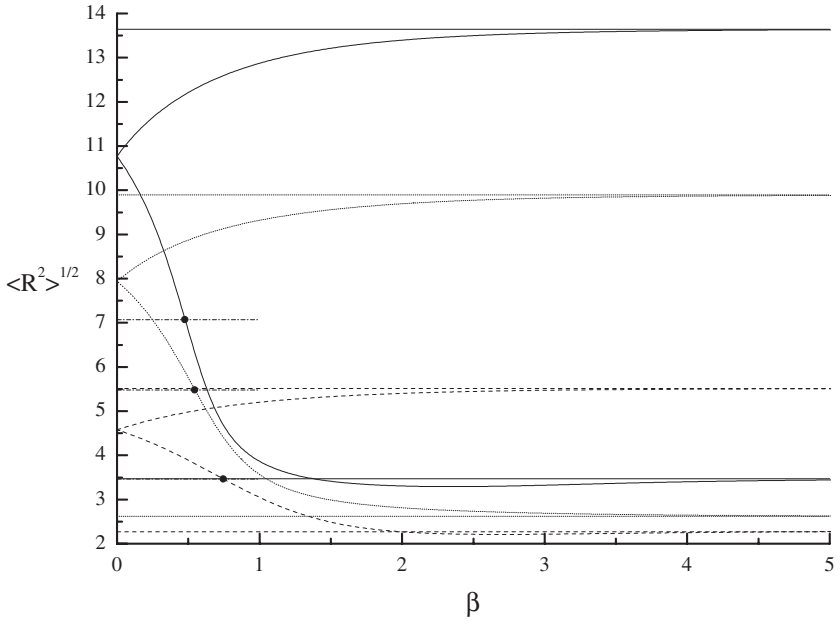

Figure 7. Dependences of root mean square end-to-end distance $\sqrt{\langle R^2\rangle}$, equation (9), on the inverse temperature β (ε^{-1}-units) and their asymptotes at $\beta \to \infty$ for $\varepsilon > 0$ and $\varepsilon < 0$; $N = 12$ (broken curve), 30 (dotted curve), 50 (full curve). Square roots of 12, 30, 50 (chain lines) are shown on the left-hand side.

cases. Increase of β lowers the entropy and this fall is much greater for the case of attraction due to the predominance of compact conformations; for $\varepsilon > 0$ it corresponds to comparatively weak swelling of the polymer coil. All curves have horizontal asymptotes at $\beta \to \infty$ which can be determined as (see (8)) $\frac{1}{N}\Delta S_N = \frac{1}{N}\ln\Omega_{00}(N)$ for $\varepsilon > 0$ and $\frac{1}{N}\Delta S_N = \frac{1}{N}\ln\Omega_{0m_{\max}}(N)$ for $\varepsilon < 0$. For $N = 12, 30, 50$ these asymptotic levels are -0.1236, -0.1379, -0.1407 for $\varepsilon > 0$ and -0.7736, -0.6872, -0.6680 for $\varepsilon < 0$.

Figure 7 shows β-dependences $(\langle R^2\rangle(\beta))^{1/2}$. For each N both curves ($\varepsilon > 0, \varepsilon < 0$) start at $\beta = 0$ from the point of the athermal regime and approach horizontal asymptotes for $\beta \to \infty$. For $\varepsilon > 0$ the dependences are monotonically increasing due to swelling while for $\varepsilon < 0$ a strong decrease with β corresponds to compactization of chains. In the latter case the approach of asymptotes for $N = 12, 50$ is accompanied by slight minima. First we found this behaviour when comparing our results for a short chain ($N = 12$) with exact data, figure 4. This effect is caused by the fact that conformations having maximum number of contacts can give values of $\langle R^2\rangle_m$ appreciably (up to 25%) greater than for the preceding number. So for $N = 12, m_{\max} = 9$ was observed in several spiral conformations of a kind presented in [30] with $\langle R\rangle_9$ exceeding $\langle R\rangle_8$. Accordingly at intermediate temperatures conformations with a smaller number of contacts but with slightly lower $\langle R^2\rangle$ can yield a greater contribution. The same effect in figure 7 for $N = 50$ corresponds to the nonmonotonic behaviour of $\langle R^2\rangle_m$ in the range $47 \leqslant m \leqslant 52$. At the same time no trace of a minimum is observed for $N = 30$. It should be pointed out here that this slight effect completely vanishes for the averaged square radius of inertia $\langle R_I^2\rangle(\beta)$.

According to scaling concepts [18, 19], the low temperature asymptotes for $(\langle R^2\rangle(\beta))^{1/2}$ should scale as $N^{3/5}$ for $\varepsilon > 0$ (neighbour avoiding walks) and as $N^{1/3}$, for $\varepsilon < 0$ (formation

of dense globules). Though our limited amount of data is insufficient for making serious tests of such a kind, it is possible to find out that they do not contradict these dependences. In the case $\varepsilon > 0$ we get values of $\sqrt{\langle R_N^2 \rangle}/N^{3/5}$ ($\beta \to \infty$): 1.241, 1.285, 1.305 for $N = 12, 30, 50$ (scatter about 4%). In the case $\varepsilon < 0$ values of $\sqrt{\langle R_N^2 \rangle}/N^{1/3}$ ($\beta \to \infty$) for the same set of N are: 0.992; 0.842; 0.941 (scatter about 15%).

4. Continuous model

We consider the simplest continuous model,—a freely joined chain of $N+1$ monomers (beads) with N bonds of fixed (unit) length, and study only the athermal case when each monomer unit is a hard sphere of diameter a. Our aim is to determine the excess entropy for the system with finite a, $\Delta S(a) = S(a) - S_0$, relative to the phantom chain. The entropy of the latter can be determined as $S_0 = \ln W_0$, where $W_0 = (4\pi)^N$. For chains with finite a, the conformation space volume decreases from W_0 to $W(a) = W_0 \times \nu(a)$ with $\nu(a) < 1$, and so $\Delta S(a) = \ln \nu(a) < 0$.

For calculation of $\nu(a)$ we suggest the procedure of a random walk in the conformation space of a *phantom* chain. Let us introduce a stochastic quantity: $\xi = \min(r_{ij})$ where r_{ij} is the distance between all pairs of its nonbonded beads. It is clear that ξ lies between 0 and 2 (2 corresponds to stiff rods). If we divide this area of ξ into a finite number, N_b, intervals (boxes) of length $2/N_b$ we can calculate normalized probabilities of corresponding conformational states, Ω_i, $1 \leqslant i \leqslant N_b$. By integrating (summing up) these probabilities from a certain i up to the boundary value, N_b, we determine ν_i for the corresponding a_i. This way we can calculate the excess entropy for given N in the whole range of a within a single MC run.

In order to obtain the initial chain conformation, one end of the chain (the 0th bead) is kept fixed at the origin while the bead which follows is chosen homogeneously on a sphere of a unit radius with the centre in the preceding one. The positions of all the following beads are chosen in the same way. In other words, the initial configuration is an N-step purely random walk in the continuous 3D space. A homogeneous modification of the conformation is carried out within one of the two following procedures. The first one is analogous to that for the lattice polymer: we choose homogeneously one of the N segments (e.g., the k_0th) and change its direction within the full spherical angle, 4π, with following parallel shift of the rest of the pieces of the chain ($k_0 < k \leqslant N$). Another procedure includes a choice of two arbitrary nonbonded beads and subsequent random rotation of the piece between them within the full angle, 2π. Tests showed the equivalence of both procedures. Most of the calculations were carried out with the number of boxes $N_b = 40$ which corresponds to the width of intervals $2/N_b = 0.05$. In the URW algorithm, the trial configuration is always accepted while in the WL procedure condition (2) is additionally imposed.

Each WL calculation included 25 sweeps with initial value of $S_i = 0$, $\Delta S_0 = 1$, the increment in all cases was $c = 0.75$ and the length of a sweep was taken as $\frac{1000}{c^{t-1}}$, where t is the sweep number. The total number of MC steps was about 4 million. For different chain lengths, the simulation time on the Pentium III 1 GHz computer varied between 5 min and 20 h.

For test purposes, comparative calculations of Ω_i within the URW and WL procedures were carried out for $N = 12, 20$, figure 8(c). The situation is similar to those presented in figures 2 and 3: in the vicinity of maxima within 4–5 orders we see good coincidence of both methods; out of this range the URW completely fails while the WL yields smooth dependence within 10 orders down. Unfortunately, we cannot compare our numerical results with exact data which, in contrast to the lattice case, do not exist for the continuous model.

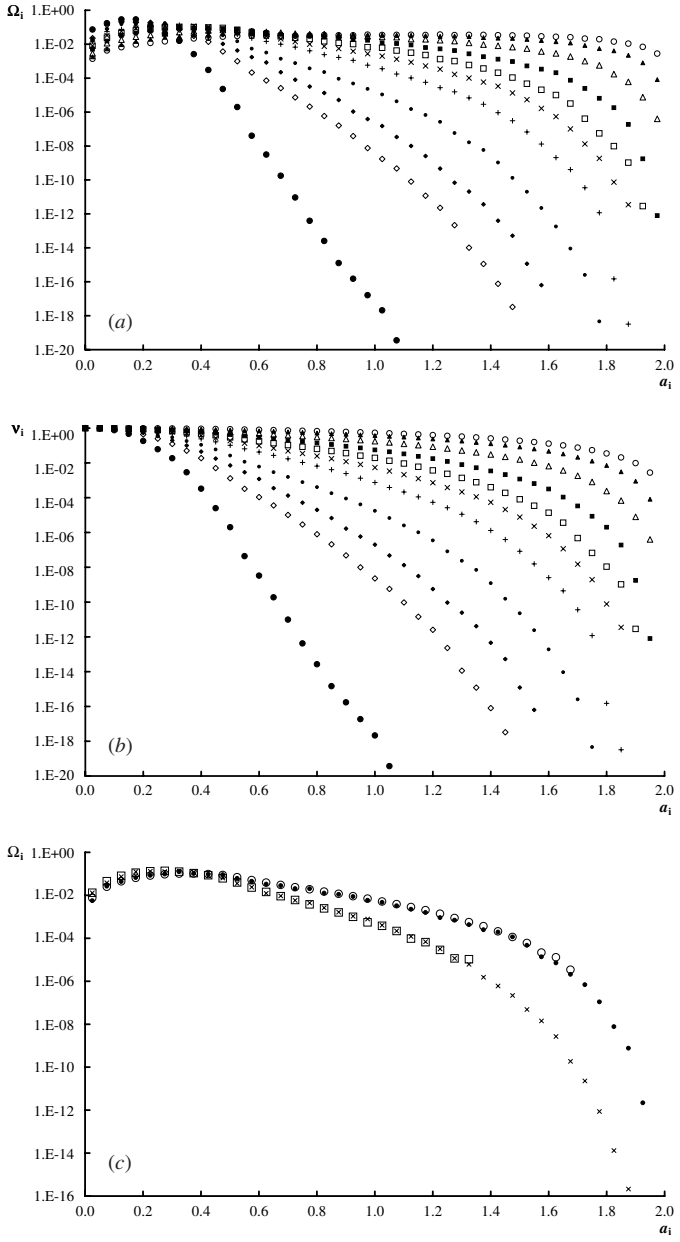

Figure 8. Normalized probabilities Ω_i for a freely joined chain (*a*) and values of ν_i (*b*) as functions of monomer diameter $a_i = \frac{2}{N_b}(i-1)$ for different chain lengths N: 3 (open circles), 4 (full triangles), 6 (open triangles), 9 (full squares), 12 (open squares), 15 (crosses), 20 (pluses), 30 (dots), 40 (full diamonds), 50 (open diamonds), 100 (full circles). (*c*) Comparison of Ω_i obtained by the WL and URW algorithms for $N = 12$ and 20 (WL—dots, crosses, URW—circles, squares, respectively).

Figures 8(*a*) and (*b*) present our WL results for Ω_i (a) and its integral ν_i (b) at a set of chain lengths: $N = 4, 6, 9, 12, 15, 20, 30, 50$. These data enable us to calculate excess specific entropy as $\Delta s_i(N) \equiv \Delta S_{Ni}/N = \ln \nu_i / N$. In figure 9 the latter is presented for each

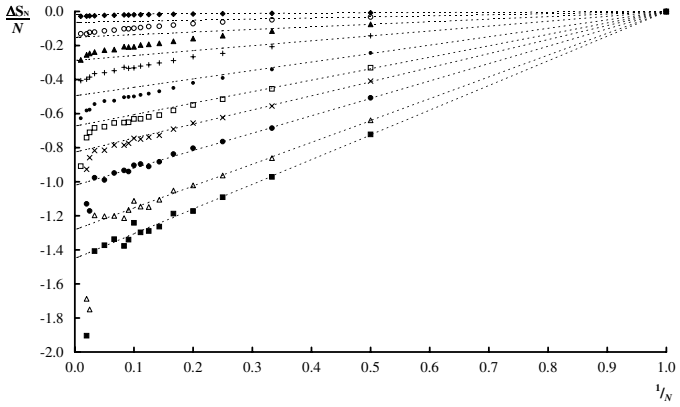

Figure 9. Dependence of specific excess entropy, $\frac{\Delta S_N}{N}$, on the inverse number of bonds, $\frac{1}{N}$, for freely joined chains with different monomer diameters a: 0.25 (full diamonds), 0.50 (open circles), 0.75 (full triangles), 1.00 (pluses), 1.25 (dots), 1.40 (open squares), 1.50 (crosses), 1.60 (full circles), 1.70 (open triangles), 1.75 (full squares) in the range $0 \leqslant a \leqslant 1.75$; straight lines are functions $(1 - \frac{1}{N}) \ln(1 - (\frac{a}{2})^2)$ for corresponding a.

a_i as a function of N in the inverse N-scale. For comparison we also draw in figure 9 straight lines approximating the excess specific entropy $\Delta S_N(a)/N$ as

$$\frac{1}{N}\Delta S_N(a) = \left(1 - \frac{1}{N}\right) \ln\left(1 - \left(\frac{a}{2}\right)^2\right). \tag{10}$$

It is based on the following scheme. If we create a free joined N-bonded chain conformation with the diameter of a spherical monomer a, the first monomer can move around the origin (the fixed zeroth monomer) within the full spatial angle, 4π; the second monomer can move around the first one only within a limited spatial angle, $\omega_a < 4\pi$, which can be expressed as $\omega_a = 4\pi\left(1 - \left(\frac{a}{2}\right)^2\right)$ (for $a = 0, \omega_a = 4\pi$, for $a = 2, \omega_a = 0$). If we neglect correlations in positions of subsequent monomers, we can assume that the third monomer can move around the second one also within the spatial angle ω_a, etc, up to the Nth monomer. In this approximation the entropy is $S_N(a) = \ln\left(4\pi\omega_a^{(N-1)}\right)$ and the excess specific entropy is determined by (10) as a linear function of N^{-1} with the slope dependent on a.

This scheme is exact only for $N = 2$. For longer chains true values of $\Delta S_N(a)/N$ must go below these straight lines in the N^{-1} plane as a result of correlations existing between distant nonbonded spherical monomer units due to bending of the chain. In figure 9 it is seen that for comparatively small a ($a \leqslant 1.25$), the calculated N^{-1} dependences actually go below the related straight lines (10). They are rather smooth and it is possible to extrapolate them to the infinite limit for specific entropy. For greater a ($a > 1.25$) another kind of dependence in figure 9 is gradually established: the MC data almost strictly follow the appropriate straight lines (10) up to $N = 30$. For greater N a steep jump to lower values is observed. This behaviour is evidently caused by growing stiffness of chains with increasing monomer diameter a; this stiffness holds for comparatively short chains and is broken only when the chain length exceeds a certain threshold, e.g., $N = 50$ for $a = 1.6$–1.7.

Figure 10 shows the obtained dependences of the mean square end-to-end distance on the number of bonds N, $4 \leqslant N \leqslant 50$, for a set of monomer diameters in the range $0 \leqslant a \leqslant 1.25$ approximated for each a by a power function. It is seen that for $a = 0$ (a phantom chain) the N-dependence of $\langle R^2 \rangle$ is practically linear with the unit slope. It confirms the correctness of our conformation change procedure since in this case the relation $\langle R^2 \rangle = Nl^2$ should hold.

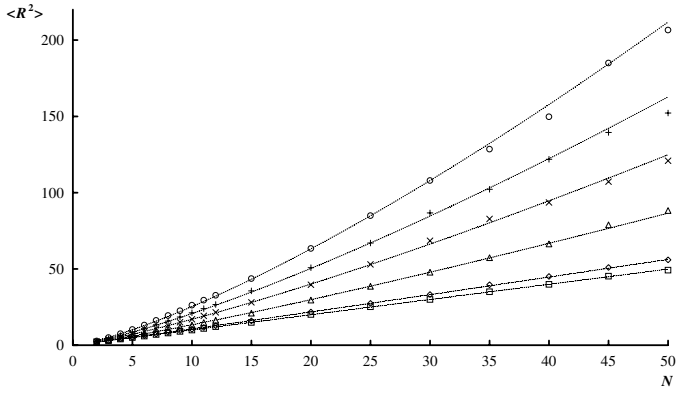

Figure 10. Dependence of the mean square end-to-end distance on the chain length, N, of freely joined chains with different monomer diameters a: 0.00 (squares), 0.25 (diamonds), 0.50 (triangles), 0.75 (crosses), 1.00 (pluses), 1.25 (circles). Lines are power approximations with the power: 1.00, 1.03, 1.13, 1.23, 1.31, 1.39 respectively.

For finite diameter the power increases with growth of a. Though for greater N it is not clear whether these power dependences would hold, it is evident that for $a = 2$, i.e. in the trivial case of stiff rods, the square law is valid for all N.

5. Conclusion

We extended the WL algorithm to ES-simulation of polymers starting with the simplest three-dimensional models. For short lattice chains our numerical results were tested by comparing them with the exact data. For the athermal lattice model we reproduced well the existing scaling relations—the N-dependence of the number of SAWs and the mean square end-to-end distance for SAWs. In the thermal case we calculated distributions over the number of contacts and used them for obtaining a set of canonical averages—conformational energy, heat capacity, entropy and mean square end-to-end distance. Some of our results were compared with MC data from [20, 21]. For a freely joined continuous model in the athermal case we suggested and carried out a variant of the WL algorithm providing us with excess entropy as a function of the monomer diameter.

The method proved to be an effective tool for studying equilibrium properties of simple models. It seems to be promising in further simulations including more complicated cases, e.g., continuous models with fixed bond–bond angles and taking into account interactions, closed and stretched chains, hetero-polymer molecules including polypeptides and cases of polymer confinement.

It seems that the WL self-adjusting procedure could also be helpful in the enhancement of various variants of the expanded ensemble method [31, 32] by optimizing (or complete elimination of) the preliminary stage. It should also be pointed out that the WL algorithm itself could become an object of further study and optimization in order to enlarge its facilities.

Acknowledgment

This work has been supported by the Russian Foundation for Fundamental Research (RFFI), grant 02-02-16618a, and the authors express their gratitude for this support.

References

[1] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087
[2] Binder K 1979 *Monte Carlo Methods in Statistical Physics* (Berlin: Springer)
[3] Allen M P and Tildesley D J 1987 *Computer Simulation of Liquids* (Oxford: Clarendon)
[4] Iba Y 2001 *Int. J. Mod. Phys.* C **12** 623
[5] Mitsutake A, Sugita Y and Okamoto Y 2001 *Biopolymers* **60** 96
[6] Lyubartsev A P, Martsinovskii A A, Shevkunov S V and Vorontsov-Velyaminov P N 1992 *J. Chem. Phys.* **96** 1776
[7] Marinari E and Parisi G 1992 *Europhys. Lett.* **19** 451
[8] Berg B A and Neuhaus T 1992 *Phys. Rev. Lett.* **68** 9
[9] Lee J 1993 *Phys. Rev. Lett.* **71** 211
[10] Hukushima K and Nemoto K 1996 *J. Phys. Soc. Japan* **65** 1604
[11] Wang F and Landau D P 2001 *Phys. Rev. Lett.* **86** 2050
[12] Wang F and Landau D P 2001 *Phys. Rev.* E **64** 056101
[13] Rathore N and de Pablo J J 2002 *J. Chem. Phys.* **116** 7226
[14] Jain T S and de Pablo J J 2002 *J. Chem. Phys.* **116** 7238
[15] Yan Q L, Faller R and de Pablo J J 2002 *J. Chem. Phys.* **116** 8745
[16] Shell M S, Debenedetti P G and Panagiotopoulos A Z 2002 *Phys. Rev.* E **66** 056703-1
[17] Vorontsov-Velyaminov P N and Lyubartsev A P 2003 *J. Phys. A: Math. Gen.* **36** 685
[18] de Gennes P G 1979 *Scaling Concepts in Polymer Science* (Ithaca, NY: Cornell University Press)
[19] Grosberg A Yu and Khokhlov A R 1994 *Statistical Physics of Macromolecules* (New York: American Institute of Physics)
[20] Douglas J, Guttman C M, Mah A and Ishinabe T 1997 *Phys. Rev.* E **55** 738
[21] Grassberger P and Hegger R 1995 *J. Chem. Phys.* **102** 6881
[22] Kremer K and Binder K 1988 *Phys. Rep.* **7** 259
[23] Madras N and Sokal A D 1988 *J. Stat. Phys.* **50** 109
[24] Duplantier B 1987 *J. Chem. Phys.* **88** 4233
[25] Rosenbluth M N and Rosenbluth A W 1955 *J. Chem. Phys.* **23** 356
[26] Grassberger P 1997 *Phys. Rev.* E **56** 3682
[27] Zhao D, Huang Y, He Z and Qian R 1996 *J. Chem. Phys.* **104** 1672
[28] Rapoport D C 1985 *J. Phys. A: Math. Gen.* **18** 113
[29] Le Guillou J C and Zinn-Justin J 1985 *J. Phys.* **46** L137
[30] Kloczkowski A and Jernian R L 1998 *J. Chem. Phys.* **109** 5147
[31] Lyubartsev A P and Vorontsov-Velyaminov P N 2003 *Recent Res. Chem. Phys.* **4** 63
[32] Vorontsov-Velyaminov P N, Broukhno A V, Kuznetsova T V and Lyubartsev A P 1996 *J. Phys. Chem.* **100** 1153