Virial coefficients of hard spheres and hard disks up to the ninth

Stanislav Labík, Jiří Kolafa,* and Anatol Malijevský
Department of Physical Chemistry, Institute of Chemical Technology, 166 28 Prague 6, Czech Republic
(Received 24 September 2004; published 28 February 2005)

A technique for topological analysis of the Ree-Hoover diagrams is developed with the aim to calculate the Ree-Hoover weights up to the ninth order with moderate demands on computer storage and CPU time. The ninth virial coefficients of hard spheres and disks are calculated, and the lower virial coefficients are accurately recalculated. The calculations require several spanning diagrams; the most important spanning chains are generated by reptation, other spanning diagrams by the standard Metropolis Monte Carlo algorithm. The tenth and eleventh virial coefficients for hard spheres are estimated.

DOI: 10.1103/PhysRevE.71.021105 PACS number(s): 64.10.+h

I. INTRODUCTION

Virial coefficients, the coefficients in the density expansion of the compressibility factor, are cornerstones of the statistical thermodynamics of fluids at low and medium densities. They are defined by exact formulas (free from any approximation) in terms of integrals whose integrands depend on intermolecular potential energy [1,2]. They can serve for testing approximate theories of fluids and for developing equations of state.

In principle, the virial coefficients can be calculated for any order. Unfortunately, calculation of high-order coefficients becomes increasingly difficult because both the number of integrals and their dimensionality rapidly increase. The increase in complexity of the calculation is demonstrated in Table I where the numbers of integrals and their dimensionality for virial coefficients up to the ninth are summarized.

The virial coefficients have been best explored for the systems of hard spheres and two-dimensional hard disks for which they are known up to the eighth. The second, third, and fourth virial coefficients are known analytically. For hard spheres it holds that [3,4]

$$\tilde{B}_2 = 4, \quad \tilde{B}_3 = 10,$$

$$\tilde{B}_4 = \frac{2707\pi + [438\sqrt{2} - 4131 \arccos(1/3)]}{70\pi} = 18.364\,768\,4,$$

and for hard disks [5,6]

$$\tilde{B}_2 = 2, \quad \tilde{B}_3 = \frac{16}{3} - \frac{4\sqrt{3}}{\pi} = 3.128\,017\,75,$$

$$\tilde{B}_4 = 16 - 36\frac{\sqrt{3}}{\pi} + \frac{80}{\pi^2} = 4.257\,854\,46,$$

where $\tilde{B}_i$ are the virial coefficients reduced by the molecule "volume" $\mathcal{V}$ (sphere volume or disk area),

$$\tilde{B}_i = B_i/\mathcal{V}^{i-1}.$$

These virial coefficients appear in the expansion of the compressibility factor in powers of the packing fraction $\eta = N\mathcal{V}/V$, where N denotes the number of particles and V system volume.

The higher virial coefficients must be calculated numerically. The fifth virial coefficients for hard spheres and disks were calculated by Rosenbluth and Rosenbluth [7] and by Kratky [8–11], the sixths by Ree and Hoover [12], the sevenths also by Ree and Hoover [13], by Kim and Henderson [14], and by Janse van Rensburg and Torrie [15], and the eighths by Janse van Rensburg [16]. As a rule, when the higher virial coefficients were evaluated, the lower ones were more accurately recalculated. Recently Vlasov, You, and Masters [17] recalculated the seventh and the eighth virial coefficients.

The aim of this work is to determine the ninth virial coefficients of hard spheres and hard disks and to provide improved values of the lower virial coefficients. We use the algorithm originally developed in Refs. [12,13] (Ree-Hoover diagrams) and further extended in Refs. [16,17]. Within this general algorithm, we propose a technique for topological analysis of the diagrams and a technique for calculation of the diagrams.

II. THEORY

A. Basic formulas

The nth virial coefficient, B_n, of particles interacting via a spherically symmetric pair potential, $u(r_{ij})$, is given by the sum of all cluster integrals corresponding to labeled irreducible f-bond diagrams with n points [18],

$$B_n = \frac{1-n}{n!}\sum_R I_M(R). \tag{1}$$

We note that there are a number of synonymous expressions for these diagrams in the literature, such as Mayer diagrams, diagrams without articulation points, double connected diagrams, Mayer stars, cluster diagrams, blocks, etc. We will use term "Mayer diagrams."

The cluster integral $I_M(R)$ of Mayer diagram R is

*URL: http://www.vscht.cz/fch/en/people/

TABLE I. Number of unlabeled and labeled Mayer and Ree-Hoover diagrams (see Sec. II B for explanation) and dimensionality d of corresponding cluster integrals for hard spheres.

	Mayer		Ree-Hoover		
n	unlabeled	labeled	unlabeled	labeled	d
2	1	1	1	1	1
3	1	1	1	1	3
4	3	10	2	4	6
5	10	238	5	68	9
6	56	11 368	23	3 053	12
7	468	1 014 888	171	297 171	15
8	7 123	166 537 616	2 606	56 671 216	18
9	194 066	50 680 432 112	81 564	21 286 987 064	21

$$I_{\mathrm{M}}(R) = \int \cdots \int \prod_{\langle i,j\rangle \in R} f_{ij} \mathrm{d}\boldsymbol{r}_2 \cdots \mathrm{d}\boldsymbol{r}_n, \tag{2}$$

where the product is over all bonds in diagram R and the Mayer function f_{ij} is given by

$$f_{ij} = \exp[-\beta u(r_{ij})] - 1 = e_{ij} - 1. \tag{3}$$

Here $\beta = 1/(k_B T)$ and e_{ij} is the Boltzmann factor. For hard spheres and hard disks (generally for hard-body systems) f_{ij} is either -1 if the particles overlap or 0 if they do not.

Many Mayer diagrams in Eq. (1) differ in the numbering of the nodes (particles) only. Thus Eq. (1) can be written in a more compact form,

$$B_n = \frac{1-n}{n!} \sum_S w_{\mathrm{M}}(S) I_{\mathrm{M}}(S), \tag{4}$$

where now the sum is over unlabeled Mayer diagrams and $w_{\mathrm{M}}(S)$ denotes the Mayer weight. Each unlabeled diagram thus corresponds to the class (diagram group) of $w_{\mathrm{M}}(S)$ labeled diagrams which are topologically equivalent (isomorphic). For calculation purposes one labeled diagram from the class is chosen as its canonical representation.

Ree and Hoover in their pioneering work [12] replaced the irreducible Mayer diagrams by the generalized diagrams with f bonds and e bonds,

$$\prod_{\langle i,j\rangle \in S} f_{ij} = \prod_{\langle i,j\rangle \in S} f_{ij} \prod_{\langle i,j\rangle \notin S} 1 = \prod_{\langle i,j\rangle \in S} f_{ij} \prod_{\langle i,j\rangle \notin S} (e_{ij} - f_{ij}). \tag{5}$$

There are again a lot of synonyms used in the literature for these diagrams: Ree-Hoover diagrams, modified stars, Ree-Hoover complements, complement blocks, etc. We will use the abbreviation "RH diagrams."

By expanding the last product in Eq. (5), one derives the following formula for the RH weights:

$$w_{\mathrm{RH}}(S) = \sum_{S'} (-1)^{|S'|-|S|} w_{\mathrm{M}}(S') h(S,S'), \tag{6}$$

where the sum is over all canonical representation S', $|S|$ denotes the number of f bonds in diagram S, and $h(S,S')$ is the number of distinct labeled RH diagrams S having the canonical representation of Mayer diagram S' as a subdiagram. Finally,

$$B_n = \frac{1-n}{n!} \sum_S w_{\mathrm{RH}}(S) I_{\mathrm{RH}}(S), \tag{7}$$

where

$$I_{\mathrm{RH}}(S) = \int \prod_{\langle i,j\rangle \in S} f_{ij} \prod_{\langle i,j\rangle \notin S} e_{ij} \mathrm{d}\boldsymbol{r}_2 \cdots \mathrm{d}\boldsymbol{r}_n. \tag{8}$$

The RH expansion leads to a considerable reduction of the number of diagrams that must be evaluated, as many Mayer diagrams cancel out identically, see Table I. In addition, the computer code to implement Eq. (8) is much simpler and more efficient than for Eq. (2).

B. Calculation of the Mayer and Ree-Hoover weights

It is easy to determine the number of irreducible diagrams and their Mayer weights for $n \leqslant 6$ because this can be done using "pencil and paper." The higher n is, the more difficult it becomes to keep relevant information under control. Fortunately, state-of-the-art algebra computing facilities allow us to determine the numbers of relevant diagrams without a "human factor" error.

The analysis is based on n-point diagrams with nodes labeled by numbers $(1,2,\ldots,n)$. Any diagram S can be uniquely characterized by binary code number $F(S)$ with $n_b = \binom{n}{2}$ binary digits. Each bit of $F(S)$ corresponds to some bond f_{ij} where digit 1 denotes a presence of f_{ij} bond and digit 0 either its absence (for Mayer diagrams) or the presence of an e_{ij} bond (for RH diagrams). The choice of numbering of the bonds appearing in diagrams is in principle arbitrary. In this paper, the bits from the most significant, 2^{n_b-1}, to the least significant, 2^0, correspond to bonds $f_{12}, f_{23}, f_{34}, \ldots, f_{n1}, f_{13}, f_{24}, \ldots, f_{n2}, f_{14}, \ldots$. It means that

$$F(S) = -(2^{n_b-1} f_{12} + 2^{n_b-2} f_{23} + 2^{n_b-3} f_{34} + \cdots). \tag{9}$$

This choice differs from numbering used in all previous papers. As shown later, it enables us to reduce considerably both the computer time and memory requirements. Moreover, we define the canonical representation using this num-

bering: *The canonical representation of a class of isomorphic labeled diagrams is the diagram S with the greatest value of* $F(S)$.

The algorithm starts by selecting the irreducible diagrams and determining their Mayer weights. The following algorithm is repeated for all diagrams starting from code number $2^{n_b}-1$ (corresponding to diagram with all f_{ij} bonds).

(i) Has the given diagram been analyzed? If so, continue to the next diagram with the code number decreased by 1.

(ii) Is the diagram connected and irreducible? If not, continue to the next diagram in the same manner.

(iii) In this case, the chosen diagram is the canonical representation of the next isomorphic class and its Mayer weight is calculated. This is done as follows. We generate all $n!$ labelings of the diagram, i.e., all permutations of numbers $(1,2,\ldots,n)$ assigned to the nodes. For all of them, the code number $F(S)$ is calculated and the number of different values of $F(S)$ gives the Mayer weight of the class. Simultaneously, all these diagrams are marked as already analyzed. In this way, it is assured that all diagrams are taken into account once and only once.

The proposed algorithm avoids usage of the so-called numeration invariants (called also graph determinants) and different diagrammatic theorems which have been used in previous papers [12,13,16,17] as characteristics enabling us to distinguish between different isomorphic classes. We also remark that the numeration invariants defined in [13] are not sufficient to distinguish between all classes for diagrams with $n>7$.

There are no memory problems with this analysis for the virial coefficients up to B_8. However, there are $\binom{9}{2}=36$ possible bonds for B_9 and 2^{36} diagrams are to be analyzed. With a single bit needed for each diagram (as a flag marking diagram status in the algorithm), this corresponds to 2^{33} bytes (8 GiB; 1 Gi$=1024^3$) of memory, which is more than is addressable on 32-bit computers. The trick which allowed us to proceed is based on the following statement: *The canonical representation of any n-point,* $n\geqslant 5$, *Mayer diagram contains path* $f_{12}f_{23}f_{34}f_{45}$. The proof of this statement is given in the Appendix.

Consequently, in the above algorithm it is sufficient to consider only diagrams with the four most important bits of $F(S)$ set to 1. The number of diagrams to analyze thus reduces from 2^{36} to 2^{32} and the memory requirements to moderate half a gigabyte. At the same time, the computer time of the diagram analysis is considerably reduced.

Resulting CPU times for the analysis of Mayer diagrams were about 2s, 4 min, and 11 h for $n=7$, 8, and 9, respectively, on a PC with a Pentium 4/2.4 GHz processor. In [16], 30 min and 250 h were reported for $n=7$ and 8, respectively, on an unspecified fast workstation. It implies that the usage of our algorithm along with the progress in computer speed enabled us to speed up the calculations by more than three orders of magnitude.

The second part of the topological analysis is calculation of the RH weights from the Mayer weights. For all classes of diagrams, we start with the canonical representation chosen in the previous analysis. Then we generate all its distinct labelings in the same manner as previously and calculate $h(S,S')$. The RH weights are finally calculated from Eq. (6). The algorithm is further optimized by sorting the diagrams in a given class and using the property that if $S'\subseteq S$, then $F(S')\leqslant F(S)$.

As noted earlier, the diagrammatic analysis becomes substantially complicated with increasing n and the number of diagram classes rises very rapidly. To verify the results of the analysis is therefore rather problematic. We nevertheless compared our results with those of previous authors [12,13,16,17] and we got complete agreement. The numbers of the Mayer diagrams and RH diagrams also agree with values from the literature where they were published for $n\leqslant 9$ [19,20]. The condition which states [13]

$$\sum_S w_{\mathrm{RH}}(S) = 1$$

is another independent test which was also successfully verified.

C. Monte Carlo integration

The integrals in Eq. (8) are calculated by Monte Carlo integration. To do this, a certain diagram (spanning diagram) T formed as a subset of f_{ij} bonds of S is selected. The spanning diagram must locate positions of all particles of S and simultaneously must be simple enough to enable analytical evaluation of $I_{\mathrm{M}}(T)$. The simplest example of such a spanning diagram can be the linear or nonlinear tree [17]. We note that it is not necessary to use only trees as the spanning diagrams and in some cases it even can be more efficient to use more complicated diagrams [21].

Positions of particles for which the product of f bonds in T is nonzero, i.e., with overlaps of given bonds, are sampled by the Monte Carlo method. We used two methods in dependence on the spanning diagram. For the linear chain we used reptation [22]. In one reptation step a particle is added at random at the head of the chain so that it overlaps with the head particle; this particle becomes the new head while the tail particle is removed. For nonlinear trees we used the standard Metropolis Monte Carlo method [5,22]. In one step, each particle was subject to a trial move which was accepted if all f bonds in T were preserved and rejected otherwise. The size of the move was adjusted (for each particle independently) to an acceptance ratio of about 0.4. Technical details about the used random number generator are given in Appendix B of Ref. [21].

The value of $I_{\mathrm{RH}}(S)$ is then calculated using the formula

$$I_{\mathrm{RH}}(S) = I_{\mathrm{M}}(T)\left\langle \prod_{\langle i,j\rangle\in S\setminus T} f_{ij} \prod_{\langle i,j\rangle\notin S\cup T} e_{ij} \right\rangle_T , \qquad (10)$$

where the brackets are mean values over all MC samples of T. Generally, for an unlabeled diagram S there may exist several isomorphic labeled RH diagrams $U(S,T)$ sharing the same labeled spanning diagram T. Therefore, the average value

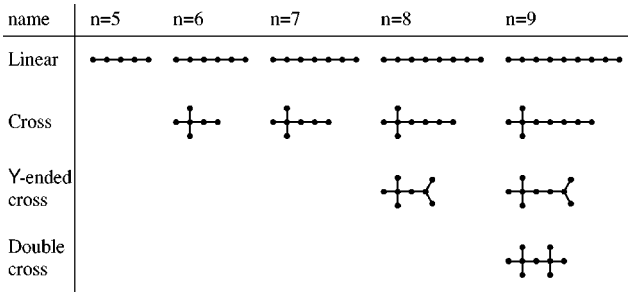

FIG. 1. Spanning trees and their mnemonic names.

$$\bar{I}_{\mathrm{RH}}(S) = \frac{1}{\gamma(S,T)} \sum_{U(S,T)} I_{\mathrm{RH}}(U), \tag{11}$$

can substantially improve the efficiency of the MC method. The "unlabeling factor" $\gamma(S,T)$ counts the number of these labelings.

For $n \leqslant 5$, the simplest linear spanning diagram can be used for all unlabeled RH diagrams. For $n \geqslant 6$, it is not possible (or useful) to cover all S by one spanning diagram. The spanning diagrams used are depicted in Fig. 1. In all cases $I_{\mathrm{M}}(T) = (-V_{\mathrm{exc}})^{n-1}$, where V_{exc} is the excluded volume (or area) of the particle.

Using these sets of spanning diagrams, two different strategies might be used to calculate values of all RH diagrams.

In strategy I, used already in previous papers [12,13,15–17], $\bar{I}_{\mathrm{RH}}(S)$ are calculated from Eqs. (10) and (11) for all diagrams having the first (linear) spanning diagram T_1 as a subset, i.e., $\gamma(S,T_1) > 0$. For other RH diagrams with $\gamma(S,T_1) = 0$, the second spanning diagram T_2 is chosen and $\bar{I}_{\mathrm{RH}}(S)$ for $\gamma(S,T_2) > 0$ is calculated. This procedure is repeated until all $\bar{I}_{\mathrm{RH}}(S)$ are obtained. The numbers of unlabeled and labeled RH diagrams calculated in this way by particular spanning diagrams are listed in Table II. We note that the number of labeled RH diagrams for $n=8$ generated by a linear spanning tree (1 106 208 isomorphs) in [17] is incorrect, the error being caused by an insufficient number of rotational invariants in diagrammatic analysis [23]. These errors affected only a small portion of diagrams with marginal impact on the final results. The virial coefficient obtained from strategy I is given by

$$B_n = \sum_T \Delta B_n(T), \tag{12}$$

where the sum is over spanning trees depicted in Fig. 1 in the order defined in this figure and $\Delta B_n(T)$ is a contribution to the virial coefficient due to spanning diagram T. The explicit formula for each contribution is

$$\Delta B_n(T) = \frac{1}{M(T)} \sum_{i=1}^{M(T)} \frac{1-n}{n!} (-V_{\mathrm{exc}})^{n-1} \times \sum_S \frac{(-1)^{|S|-|T|} w_{\mathrm{RH}}(S)}{\gamma(S,T)} \sum_{j=1}^{\gamma(S,T)} H(i,j), \tag{13}$$

where $M(T)$ is a number of configurations of a given spanning diagram generated during MC sampling, Σ_S is a sum over all unlabeled RH diagrams contributing to the given $\Delta B_n(T)$ (as defined above), and $H(i,j)$ is 1 if the ith generated configuration has the same bonds as the jth labeled diagram and 0 otherwise.

In the actual computer code, all remaining (not already in T) f bonds are determined and then used to compose a binary number F similarly to Eq. (9) but now with the f bonds of the spanning tree omitted because they are all 1. This number is used as an index of a look-up-table containing indices of all $N(T)$ unlabeled diagrams spanned over T. A unity is then

TABLE II. Numbers of unlabeled and labeled RH diagrams used in calculation for different spanning diagrams. n is the number of nodes (particles), T is the type of spanning diagram (see Fig. 1), $\Delta N(T)$ is the number of RH diagrams generated by the given spanning diagram that have not been sampled by any previous spanning diagram, and $N(T)$ is the overall number of RH diagrams generated by a given spanning diagram.

		$\Delta N(T)$		$N(T)$	
n	T	unlabeled	labeled	unlabeled	labeled
5	Linear	5	16	5	16
6	Linear	22	318	22	318
6	Cross	1	1	16	175
7	Linear	168	13 506	168	13 506
7	Cross	3	16	146	9 715
8	Linear	2 576	1 099 116	2 576	1 099 116
8	Cross	29	453	2 517	885 899
8	Y-ended cross	1	1	2 466	831 866
9	Linear	81 302	173 778 502	81 302	173 778 502
9	Cross	255	15 594	80 812	150 339 039
9	Y-ended cross	6	23	77 709	143 458 475
9	Double cross	1	1	72 170	131 532 025

TABLE III. Reduced virial coefficients $\tilde{B}_n$ of hard spheres calculated by different spanning trees T. $\Delta\tilde{B}_n(T)$ and $\tilde{B}_n(T)$ denote the results of Eqs. (13) and (17), respectively. The total number of MC configurations for each spanning diagram is denoted by $M(T)$. Values in parentheses are estimated standard errors.

n	T	$\Delta\tilde{B}_n(T)$	$\tilde{B}_n(T)$	$10^{-9}M(T)$
5	Linear	28.22438(31)	28.22438(31)	807
6	Linear	39.81591(111)	36.41260(104)	1318
6	Cross	−0.0000063(4)	3.40318(26)	152
6	sum	39.81590(111)	39.81578(107)	1470
7	Linear	53.34130(162)	49.36955(158)	16463
7	Cross	0.000129(3)	3.97175(26)	819
7	sum	53.34143(162)	53.34130(160)	17282
8	Linear	68.5516(101)	65.4293(99)	13909
8	Cross	−0.01174(12)	2.0011(14)	624
8	Y-ended cross	0	1.1096(5)	102
8	sum	68.5398(101)	68.5400(100)	14635
9	Linear	85.679(83)	111.173(76)	7611
9	Cross	0.129(1)	16.087(16)	870
9	Y-ended cross	0.00066(4)	−35.175(13)	926
9	Double cross	0	−6.286(7)	399
9	sum	85.809(83)	85.799(79)	9806

added to the corresponing element of an array accumulating thus the last sum of Eq. (13). This efficient implementation is based on the f and e bonds in RH diagrams and is not possible with the Mayer diagrams. For B_9, the size of the look-up-table (of 4-byte integers) is 1 GiB.

In strategy II, $\overline{I_{\mathrm{RH}}}(S)$ are calculated as weighted averages from all spanning diagrams having their $\gamma(S,T)>0$,

$$\overline{I_{\mathrm{RH}}}(S) = \sum_T \frac{W(S,T)}{\gamma(S,T)} \sum_{U(S,T)} I_{\mathrm{RH}}(U), \tag{14}$$

where the Σ_T is over all spanning diagrams T and $W(S,T)$ is the weight (of diagram S in the MC set of configurations for T). It is proportional to the product of the number of configurations generated for given T, $M(T)$, and the unlabeling factor $\gamma(S,T)$,

$$W(S,T) = \frac{M(T)\gamma(S,T)}{\sum_{T'} M(T')\gamma(S,T')}. \tag{15}$$

The virial coefficient obtained from this strategy is given by

$$B_n = \sum_T B_n(T), \tag{16}$$

where $B_n(T)$ is a part of the virial coefficient calculated by spanning diagram T,

$$B_n(T) = \frac{1}{M(T)} \sum_{i=1}^{M(T)} \frac{1-n}{n!}(-V_{\mathrm{exc}})^{n-1} \times \sum_S \frac{W(S,T)(-1)^{|S|-|T|} w_{\mathrm{RH}}(S)}{\gamma(S,T)} \sum_{j=1}^{\gamma(S,T)} H(i,j). \tag{17}$$

We remark in passing that both Eqs. (13) and (17) are in the form of MC averages and therefore their errors can be easily estimated even though the consecutive MC configurations are correlated.

Both proposed strategies can be implemented simultaneously within the same Monte Carlo simulation. Generally speaking, the second strategy is better than the first one. The main reason is that Eq. (17) uses all available configurations of the spanning diagrams to obtain average values of $\overline{I_{\mathrm{RH}}}(S)$, while Eq. (13) ignores the configurations belonging to the diagrams covered by previous spanning trees. For instance, for $n=6$ the linear spanning diagram generates only 22 RH diagrams out of a total 23 needed. The missing one with bonds $f_{12}, f_{23}, f_{24}, f_{25}, f_{26}, f_{34}, f_{35}, f_{36}$ is sampled by a cross spanning diagram; this spanning tree generates 15 other diagrams.

On the other hand, the first strategy can be used to distinguish between the importance of the groups of diagrams mapped by individual spanning diagrams. A more detailed comparison of both strategies is given in the following section.

III. RESULTS AND DISCUSSION

A. Virial coefficients up to B_9

Virial coefficients B_5 to B_9 for hard spheres and hard disks were calculated using both strategies described in the previ-

TABLE IV. Reduced virial coefficients $\tilde{B}_n$ of hard disks calculated by different spanning trees T. The notation is the same as in Table III.

n	T	$\Delta\tilde{B}_n(T)$	$\tilde{B}_n(T)$	$10^{-9}M(T)$
5	Linear	5.3368943(69)	5.3368943(69)	12843
6	Linear	6.3630277(111)	6.1856656(108)	27465
6	Cross	0	0.1773607(17)	796
6	sum	6.3630277(111)	6.3630263(109)	28261
7	Linear	7.352083(28)	7.272598(28)	24532
7	Cross	0	0.079482(3)	258
7	sum	7.352083(28)	7.352080(28)	24790
8	Linear	8.318669(62)	8.216161(62)	32435
8	Cross	0	0.093210(5)	356
8	Y-ended cross	0	0.0092971(14)	33
8	sum	8.318669(62)	8.318668(62)	32824
9	Linear	9.27235(29)	9.12516(29)	11006
9	Cross	0	0.09931(2)	116
9	Y-ended cross	0	0.024431(8)	31
9	Double cross	0	0.023461(10)	26
9	sum	9.27235(29)	9.27236(29)	11179

ous section. In Tables III and IV, we summarize the numbers of configuration generated for each spanning diagram and values of $\Delta\tilde{B}(T)$ and $\tilde{B}(T)$ contributing to virial coefficients according to Eqs. (12) and (16).

It is seen that substantially more configurations were generated on the basis of the linear spanning diagrams. There are two reasons for this. First, the linear spanning diagrams were generated by reptation while the nonlinear ones by a less efficient standard Monte Carlo simulation. Second, the sets of the RH diagrams covered by the linear spanning diagrams are much more important as seen from their contributions to B_n's as well as their standard errors.

For hard disks, all spanning trees but the linear one give zero contributions and can be omitted. For hard spheres, the second non-negligible contributions are based on the cross spanning diagrams but they have only a small influence on the final results. Our preliminary calculations show that for more dimensional systems, the other spanning diagrams have greater importance: the diagrams spanned over cross trees not sampled by linear trees contribute by about 30% to the total value of B_9 for $D=4$. This behavior is mainly due to the steric hindrances which occur in two and three dimensions [12,13] but are much less important in higher dimensions.

Both strategies described in the previous section are compared in Tables III and IV. Is is seen that the second strategy gives only a minor decrease in standard errors. Is is mainly due to a much smaller number of configurations generated for all spanning diagrams but the linear one. The accuracy gain by using all information available for hard sphere B_6 (including the above-mentioned 15 diagrams) is marginal 4% and similarly for higher virial coefficients. Total efficiency analysis shows that the optimum numbers of configurations to reach the smallest overall error with the same CPU time would be even more for the linear spanning trees at the expense of cross trees and especially Y-ended and double cross trees.

An important by-product of these considerations is that the final values do not depend, within statistical errors, on the method even though the partial contributions to the virial coefficients differ a lot. This demonstrates the consistency of calculations using different spanning trees.

The calculations were performed on a PC-cluster based on a dual Pentium 3 1 GHz processors and one 3 GHz hyperthreading Pentium 4. The total CPU time per one virial coefficient (sixth to ninth) was in the range 4–6 month processors.

The recommended values of B_n for $n=5$ to $n=9$ for hard spheres and hard disks together with their uncertainty estimates and with the older literature results are shown in Table V. It follows from the table that the lower virial coefficients, $n\leqslant 8$, calculated in this work are more precise by more than one order than the older ones. In addition, in all cases the new and literature values match within doubled combined standard error (i.e., at the 95% confidence level).

B. Higher-order virial coefficients

A question may be posed as to whether the tenth and higher virial coefficients can be calculated using the technique proposed in this work and the state-of-the-art computer facilities. Let us consider $n=10$. There are $\binom{10}{2}=45$ possible bonds for B_{10}, and 2^{45} diagrams are to be analyzed, which can be reduced to 2^{41} using the trick described in this work. It can be easily addressed on 64-bit computers both for the analysis (1 bit needed per diagram) and for the Monte Carlo run (we used a 4-byte integer for one diagram; this number can be, however, reduced by more sophisticated programming).

TABLE V. Summary of virial coefficients $\tilde{B}_n$ for hard spheres and hard disks.

n	hard spheres		hard disks	
	recommended	literature	recommended	literature
5	28.22445(10)[a]	28.22451(26)[d]	5.33689664(64)[d]	
6	39.81550(36)[b]	39.739(56)[e]	6.363026(11)[c]	6.36256(32)[e]
7	53.3413(16)[c]	53.44(9)[f]	7.352080(28)[c]	7.35213(70)[f]
8	68.540(10)[c]	68.2(3)[f]	8.318668(62)[c]	8.3238(40)[f]
9	85.80(8)[c]		9.27236(29)[c]	

[a]Weighted average of 28.22451(26) by [8–11] and 28.224444(102) by [21].
[b]Weighted average of 39.81546(38) by [21] and this work.
[c]This work.
[d]References [8–11].
[e]Reference [16].
[f]Reference [17].

Besides computer memory, there is a computer time problem. For $n=9$, the CPU time needed to perform topological analysis is about a day. The estimated time for $n=10$ is a year. In addition, the expected error in B_{10} estimated by extrapolating the data of Table III is about ±1, provided that a several-month processor of CPU time is available. To summarize, calculation of B_{10} for hard spheres and disks is at the edge of current computer technology while B_{11} is beyond.

Methods to obtain the values of higher virial coefficients generally lie in their extrapolation from the known values of the lower virial coefficients using computer simulation equation-of-state (EOS) results. Using the present virial coefficients and the EOS molecular-dynamics data from Ref. [21], we estimate $\tilde{B}_{10}=106.5\pm0.5$, $\tilde{B}_{11}=130\pm2$. This is fully consistent with the estimates based on the virial coefficients up to B_6 only and the same EOS data: $\tilde{B}_{10}=106\pm2$, $\tilde{B}_{11}=130\pm5$ [21]. It is important to realize that these extrapolations (and consequently error estimates) are based on an implicit assumption that the series of higher ($n>9$) virial coefficients behaves "regularly" as a function of n, which need not be the case.

For a hard-disk system, there are no sufficiently precise simulation data of compressibility factors to perform similar simultaneous correlation.

IV. CONCLUSIONS

A method of calculation of the virial coefficients consisting in a safe determination of the Ree-Hoover weights and effective evaluation of the cluster integrals has been proposed. The method was applied to calculation of the ninth virial coefficients of hard spheres and hard disks and accurate recalculation of the lower virial coefficients using standard PC computers.

The method can be extended in several ways. It can be used to evaluate the virial coefficients of D-dimensional hard hyperspheres, $D>3$, which are of some theoretical interest [24]. It can also be used to calculate the ninth virial coefficients of hard-body fluids (spherocylinders, diatomics, ellipsoids, etc.) Finally, it can be modified to calculate virial coefficients of hard-body mixtures.

Note added. Recently, a preprint appeared [25] reporting the virial coefficients up to B_{10} of hard spheres in dimensions 2 to 8 and showing that the calculation of B_{10} (which we considered "at the edge of current computer technology") is feasible.

The comparison with our results for disks and spheres up to B_9 shows an excellent agreement within combined error bars. Our results are on average two to three times more accurate with the exception of B_9 for hard spheres where the accuracy is about the same. Also the exact value $\tilde{B}_{10}=105.8\pm0.4$ of [25] compares well with our EOS-based estimate 106.5 ± 0.5. In addition, our four-dimensional results obtained meanwhile [$\tilde{B}_5=146.2461(13)$, $\tilde{B}_6=253.399(12)$, $\tilde{B}_7=375.09(13)$, $\tilde{B}_8=608.1(16)$, and $\tilde{B}_9=746(19)$] also agree with those of Ref. [25].

Using these new virial data along with the hard-sphere EOS data shifts the extrapolated $\tilde{B}_{11}$ to a bit lower value than obtained in Sec. III B, 129 ± 2, while $\tilde{B}_{12}=155\pm10$ is inaccurate. It cannot be reliably determined by this method whether some higher virial coefficients become negative or not.

ACKNOWLEDGMENTS

Financial support by the Ministry of Education, Youth and Sports of the Czech Republic, project CB–MSM 22234 00008, is acknowledged.

APPENDIX

In this appendix, we prove the following statement: *The canonical representation of any n-point, $n\geqslant5$, Mayer diagram contains path $f_{12}f_{23}f_{34}f_{45}$.*

Let us first consider $n=5$. All unlabeled Mayer diagrams are listed, e.g., in [8–11] and it is easy to number them so that such a path is obtained.

Let us therefore assume $n>5$. We will build the required path for the given unlabeled Mayer diagram. First, let us choose a bond and call it f_{12}. Because of double connectivity, there exists a bonded node (a neighbor) connected to 2 and different from 1; let us call it 3. In addition, there exists a neighbor of 3 different from 2; if it differs from 1, we call it 4 and have a chain 1-2-3-4. If it does not, there must exist a neighbor of either 1, 2, or 3 and (after renumbering) we arrive at chain 1-2-3-4 again.

Now we have at least two spare (unassigned) nodes to add one node to chain 1-2-3-4. Two possibilities may happen: (i) There exists a bonded pair in the group of the spare nodes. Since this bond must be connected, directly or indirectly (via another spare node), to (any of four nodes of) the chain 1-2-3-4, we arrive at a chain of at least five nodes (four bonds) which can be renumbered to 1-2-3-4-5. (ii) No bonded pair exists in the group of the spare nodes. If there is at least one bond from node 1 or 4 (of chain 1-2-3-4) to any of the spare nodes, we have the fifth node for the chain. If there is no such bond, then all the spare nodes must be connected to nodes 2 and 3. Then chain 1-2-(spare node)-3-4 has the required properties.

[1] T. L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956).
[2] G. E. Uhlenbeck and G. W. Ford, *Studies in Statistical Mechanics*, edited by J. de Boer and G. E. Uhlenbeck (North-Holland, Amsterdam, 1962).
[3] B. R. A. Nijboer and L. van Hover, Phys. Rev. **85**, 777 (1952).
[4] I. Lyberg, e-print cond-mat/0410080.
[5] M. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[6] N. Clisby and B. M. McCoy, J. Stat. Phys. **114**, 1343 (2004).
[7] M. N. Rosenbluth and A. W. Rosenbluth, J. Chem. Phys. **33**, 1439 (1953).
[8] K. W. Kratky, Physica A **85**, 607 (1976).
[9] K. W. Kratky, Physica A **87**, 548 (1977).
[10] K. W. Kratky, J. Stat. Phys. **27**, 533 (1982).
[11] K. W. Kratky, J. Stat. Phys. **29**, 129 (1982).
[12] F. H. Ree and W. G. Hoover, J. Chem. Phys. **40**, 939 (1964).
[13] F. H. Ree and W. G. Hoover, J. Chem. Phys. **46**, 4181 (1967).
[14] S. Kim and D. Henderson, Phys. Lett. **27A**, 379 (1968).
[15] E. J. J. van Rensburg and G. M. Torrie, J. Phys. A **26**, 943 (1992).
[16] E. J. J. van Rensburg, J. Phys. A **26**, 4805 (1993).
[17] A. Y. Vlasov, X.-M. You, and A. J. Masters, Mol. Phys. **100**, 3313 (2002).
[18] J.-P. Hansen and I. R. McDonald, *Theory of Simple Fluids* (Academic Press, Amsterdam, 2003).
[19] N. J. A. Sloane, *A Hanbook of Integer Sequences* (Academic Press, New York, 1973), p. 107.
[20] N. Clisby and B. M. McCoy, J. Stat. Phys. **114**, 1361 (2004).
[21] J. Kolafa, S. Labík, and A. Malijevský, Phys. Chem. Chem. Phys. **6**, 2335 (2004).
[22] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1986).
[23] A. Y. Vlasov (private comunication).
[24] J. G. Loeser and Z. Zhen, J. Chem. Phys. **95**, 4525 (1991).
[25] N. Clisby and B. M. McCoy, e-print cond-mat/0410511.