

# Density expansion of the radial distribution and bridge functions of the hard sphere fluid

Jiří Kolafa and Stanislav Labík

(Received 00 Month 200x; In final form 00 Month 200x)

The bridge function and the background correlation function (and consequently the radial distribution function) of the pure hard sphere fluid are expanded up to the sixth power in density. The calculations are based on the Ree–Hoover representation of the diagrams and Monte Carlo integration. The coefficients as functions of the particle–particle separation are fitted to splines taking into account discontinuities in higher derivatives up to the term of the order of  $(r - \text{const})^5$ .

**Keywords:** hard sphere fluid, radial distribution function, bridge function, diagrammatic expansion, Ree–Hoover diagram

## 1 Introduction

The structure of fluids is described by various distribution functions [1]. Among them, the radial distribution function (RDF)  $g(r)$ , also called the pair correlation function, and the bridge function  $B(r)$  are most important. The RDF expresses the normalized probability of finding a pair of particles  $r$  apart. It can be generalized to the background correlation  $y(r)$  function which describes a probability of finding a pair of cavities of sphere size  $r$  apart; it is nonzero even when the spheres overlap. The bridge function is the “most elementary” function in the following sense: (i) it is given by the smallest set of diagrams, (ii)  $y(r)$ ,  $g(r)$ , and other functions can be calculated from it via the Ornstein–Zernike equation, and (iii)  $g(r)$  is relatively insensitive to the bridge function, i.e., it is enough to know the bridge function with a low precision to obtain an accurate RDF. All these functions are used in perturbation theories of fluids.

The terms of the expansion of both functions up to the second order in density are known analytically [2]. The third term in the RDF was calculated by Monte Carlo (MC) integration [3] and for the bridge function later [4, 5]. In addition, the bridge function can be expressed by the  $h$ -bond expansion [7] ( $h = g - 1$  is the total correlation function).

In this work we first review the diagrammatic techniques used, provide details on the diagrammatic and numerical algorithms, and finally fit the calculated data to splines with variable order. Extensive tables of results, technical details, and computer code for calculation of the functions, are available [6].

## 2 Theory

### 2.1 Mayer expansion

The density expansions of the background correlation and bridge functions read as [1]

$$y(r) = 1 + H_1(r)\rho + H_2(r)\rho^2 + \dots, \quad (1)$$

$$B(r) = E_2(r)\rho^2 + E_3(r)\rho^3 + \dots, \quad (2)$$

where functions  $H_n$  and  $E_n$  are given by integrals represented by diagrams with  $f$ -bonds (see [1] for details and combinatorial factors). Here  $f$  is the Mayer function,  $f = e - 1 = \exp[-u(r)/kT] - 1$ , and  $u$  is the

particle–particle potential. From the background correlation function, the RDF  $g(r)$  can be calculated by

$$g(r) = e(r)y(r).$$

The background correlation function is more general because  $e(r)$  is small or zero at short separations and information in  $g(r)$  in this region is lost.

We consider a fluid of hard spheres of diameter  $\sigma$ . This diameter will be used as a unit length; e.g., the reduced number density,  $\rho = N\sigma^3/V$ , will be referred to as “number density”. Then  $f(r) = -1$  for  $r < 1$  and  $f(r) = 0$  otherwise.

Functions  $H_n(r)$  are given by the sums over all connected diagrams with two white points not connected by a bond and  $n$  black points. The white points, called also root points, represent particles with fixed positions of distance  $r$ . The integration is performed over the black points, called also  $\rho$ -points or field points. None of these  $n + 2$  points is a black-articulation point; the black-articulation point is defined here as a point which, if removed, leaves a disjoint diagram of only black points.

Diagrams in  $H_n(r)$  can be classified into chain diagrams (at least one black point is a white-articulation point; if removed, the diagram splits into two parts, each containing one white point; in other words, both root points make an articulation pair), parallel diagrams (if both white points are removed, the diagram splits into at least two parts), and elementary diagrams (neither chain nor parallel).

Function  $E_n(r)$  is given as the sum over all elementary diagrams, in other words, all diagrams of  $H_n$  containing at least two paths of points connecting the root points which remain connected even if both root points are removed.

The first terms in the expansions are known analytically [2]:

$$H_1(r) = \gamma_1(r), \tag{3}$$

$$H_2(r) = \frac{1}{2}\gamma_1^2(r) + \epsilon_2(r) + \gamma_2(r) + 2\delta_2(r), \tag{4}$$

$$E_2(r) = \epsilon_2(r), \tag{5}$$

$$\tag{6}$$

where

$$\gamma_1 = \text{diagram} = \begin{cases} \frac{\pi}{12}(r+4)(r-2)^2 & \text{for } 0 \leq r \leq 2, \\ 0 & \text{for } r \geq 2, \end{cases} \tag{7}$$

$$\gamma_2 = \text{diagram} = \begin{cases} \frac{\pi^2}{630}(r^6 - 63r^4 + 315r^2 - 525) & \text{for } 0 \leq r \leq 1, \\ -\frac{\pi^2}{1260r}(r^3 + 12r^2 + 27r - 6)(r-3)^4 & \text{for } 1 \leq r \leq 3, \\ 0 & \text{for } r \geq 3, \end{cases} \tag{8}$$

$$\delta_2 = \text{diagram} = \begin{cases} \frac{\pi^2}{1260r}(r^5 + 4r^4 - 51r^3 - 10r^2 + 479r - 81)(r-2)^2 & \text{for } 1 \leq r \leq 2, \\ 0 & \text{for } r \geq 2 \end{cases}$$

$$\epsilon_2 = \text{diagram} = \begin{cases} -\frac{1}{2}\gamma_1^2(r) + \frac{\pi}{2}X(r) & \text{for } 1 \leq r \leq \sqrt{3}, \\ -\frac{1}{2}\gamma_1^2(r) & \text{for } \sqrt{3} \leq r \leq 2, \\ 0 & \text{for } r > 2 \end{cases}$$

$$X(r) = \left( -\frac{3}{280}r^4 + \frac{41}{420}r^2 \right) (3-r^2)^{1/2}$$

$$\begin{aligned}
& + \left( -\frac{23}{15}r + \frac{36}{35r} \right) \arccos \frac{r}{(12 - 3r^2)^{1/2}} \\
& + \left( \frac{3}{560}r^6 - \frac{r^4}{15} + \frac{r^2}{2} + \frac{2r}{15} - \frac{9}{35r} \right) \arccos \frac{r^2 + r - 3}{(12 - 3r^2)^{1/2}} \\
& + \left( \frac{3}{560}r^6 - \frac{r^4}{15} + \frac{r^2}{2} - \frac{2r}{15} + \frac{9}{35r} \right) \arccos \frac{-r^2 + r + 3}{(12 - 3r^2)^{1/2}}.
\end{aligned}$$

For completeness, we define  $H_0(r) = 1$  and  $E_0(r) = E_1(r) = 0$ .

## 2.2 Topological analysis and the Ree–Hoover expansion

In our computer code, a diagram (more precisely a labeled diagram) is represented by a binary number in which 1 marks a presence of an  $f$ -bond and 0 its absence. In a diagram consisting of two white and  $n$  black points, there are  $b = (n + 2)(n + 1)/2 - 1$  possible bonds because there is no white-white bond. For  $n = 6$  (our maximum), this is 27 bit which can be coded in one four-byte integer; storing all possible  $2^{27}$  diagrams thus requires moderate half a gigabyte of memory.

**2.2.1 Labeled Mayer diagrams.** The algorithm takes all  $2^b$  diagrams one by one and first excludes diagrams which are not connected or which contain a black-articulation point; for obtaining elementary diagrams, white-articulation points and existence of two parallel paths connecting both white points are checked as well. This procedure results in a table of labeled Mayer diagrams with attached weights 0,  $-1$ , or  $+1$ . Weight 0 means that the diagram is not to be included. Weights  $+1$  ( $-1$ ) mean that the diagram should be included and it contains an even (odd) number of  $f$ -bonds, respectively.

**2.2.2 Labeled Ree–Hoover diagrams.** In the second step, the Mayer diagrams are converted into Ree–Hoover (RH) [8] diagrams in which missing bonds are replaced by  $e$ -bonds,  $e = f + 1$ . A RH diagram is coded by binary 1 for an  $f$ -bond and binary 0 for an  $e$ -bond. The conversion proceeds in a loop over all pairs of points (but white-white). For each Mayer diagram not containing an  $f$ -bond between this pair (i.e., containing  $1 = e - f$ ), the weight of the corresponding diagram with this  $f$ -bond is added. This procedure results in a table of labeled RH diagrams with attached weights.

The number of RH diagrams (of nonzero weight) is smaller than the number of Mayer diagrams. This original reason [8] for introducing the RH diagrams is now irrelevant. However, as we shall see in Section 2.4, the computer code base on RH diagrams is simpler.

**2.2.3 Unlabeled Ree–Hoover diagrams.** In the third step, the labeled RH diagrams of nonzero weight are classified into classes of topologically equivalent (isomorphic) RH diagrams. The equivalence test is based on generating all labelings (renumberings) of the diagram: labeled diagram  $S_1$  is topologically equivalent to labeled diagram  $S_2$  if there exists a renumbering  $L(S_1)$  of  $S_1$  such that  $L(S_1) = S_2$ . These classes are often referred to as “unlabeled” diagrams. We will loosely interchange terms “unlabeled diagram  $S$ ” and “class of labeled diagrams topologically equivalent to (labeled) diagram  $S$ ”. In the computer code, an unlabeled diagram is represented by the labeled diagram with the smallest binary number, called the canonical representation. The number of RH diagrams topologically equivalent to diagram  $S$ , multiplied by the diagram weight, will be called the RH weight and denoted by  $w(S)$ .

## 2.3 Spanning diagrams

Values of diagrams are calculated by random sampling of spanning diagrams.














$T$	$S$	$w(S)$	$L(S)$
		3	
		12	 
		6	  
		12	
		-4	

Figure 1. Spanning diagram  $T = \text{“1230+24”}$ , unlabeled Ree–Hoover diagrams for calculation of  $E_5$ , and labelings  $L(S) \supset T$ . See also the bold line in Table 1.

A spanning diagram (SD; called also spanning graph or framework)  $T$  of diagram  $S$  is a diagram of two white points and  $n$  black points such that its value can be calculated analytically and  $T \subset S$ . ( $T \subset S$  if each  $f$ -bond of diagram  $T$  is also an  $f$ -bond of diagram  $S$ .) We use one or more chains between the white points, optionally decorated by trees of black points. An example of a SD is shown in figure 1.

For convenience, a symbolic code is assigned to a SD. White points are labeled 0 and 1, black points 2,  $\dots$ ,  $n+1$ . Each chain composing the SD is represented by a string of digits. Adjacent digits correspond to an  $f$ -bond. The value  $\Gamma(T, r)$  of SD  $T$  is a product of terms corresponding to these strings. The value of “021” (or “031”, etc.) is  $\gamma_1$ , see eqn (7), the value of string “0321” is  $\gamma_2$ , see eqn (8), for higher  $\gamma_n$  see [5]. The value of each black point of a tree decoration (e.g., “4” in figure 1) is  $4\pi/3$ .

**2.3.1 Unlabeling factors.** The next step in topological analysis is finding unlabeling factors. The unlabeling factor  $\lambda(S, T)$  of an unlabeled RH diagram  $S$  with respect to labeled SD  $T$  is the number of labelings  $L(S)$  such that  $T \subset L(S)$ . The result of this analysis is a table  $\mathcal{R}(T)$  organized by canonical representations  $S$  of RH diagrams. At each table entry  $S$ , the following information is attached: RH weight  $w(S)$ , the unlabeling factor  $\lambda(S, T)$ , the list of labeled diagrams  $L(S)$  matching  $T$  (i.e., their binary numbers), and a histogram bin.

**2.3.2 Example.** An example of the topological analysis is shown in figure 1. The value of  $E_3(r)$  is given by five unlabeled RH diagrams listed in column  $S$  of the figure. One of four SDs used, “1230+24”, is shown in column  $T$ ; the value of this SD is  $\Gamma(\text{“1230+24”}; r) = 4\pi/3 \cdot \gamma_2(r)$ . Only four unlabeled RH diagrams can be sampled by this SD (see the bold line of Table 1, column unl). Some of the RH diagrams can be sampled by the SD  $T$  in several ways and therefore there are seven labeled RH diagrams matching given SD  $T$  (column lbl).

Spanning diagram “1230+24” is not sufficient for obtaining  $E_3(r)$  because the first unlabeled RH diagram of figure 1 is not sampled. In this case SD “12340” is sufficient (but combining several SDs improves precision). For  $n > 3$  several SDs must be combined to sample all unlabeled RH diagrams.

Table 1. The used spanning diagrams. lbl and unl denote the number of labeled and unlabeled RH diagrams spanned, respectively, and  $10^8$  MC is the number of MC configurations. In the “total” line, lbl and unl stand for the number of labeled and unlabeled RH diagrams, respectively, and  $10^8$  MC is the total number of MC configurations. The diagrams of the bold line are shown in figure 1. The full tables of 18 and 31 SDs for  $n = 5$  and  $n = 6$  are available [6].

SD	$E_3$			$H_3$		
	lbl	unl	MC	lbl	unl	MC
120+130+140	1	1	20	8	4	30
1230+140	6	4	20	10	7	30
<b>1230+24</b>	<b>7</b>	<b>4</b>	20	11	7	30
12340	9	5	20	13	8	30
total	9	5	80	59	12	120

SD	$E_4$			$H_4$		
	lbl	unl	MC	lbl	unl	MC
120+130+140+150	4	2	20	64	11	10
12340+150	103	37	40	139	53	20
1230+140+150	39	14	20	94	30	20
12340+25	149	46	20	202	67	20
1230+140+25	90	37	20	127	55	20
12340+35	163	34	20	188	44	20
1230+1450	112	32	40	154	42	20
123450	171	47	40	225	67	40
1230+24+25				142	36	20
total	1348	55	220	2276	95	190

total	$E_5$			$H_5$		
	138648	868	246	190939	1243	498

total	$E_6$			$H_6$		
	27870727	25149	439	33745015	31040	614

## 2.4 Monte Carlo integration

In the Monte Carlo integration algorithm, a look-up-table of all  $2^b$  diagrams is created. The index is the diagram binary number. If this number is a number from the list of labeled diagrams  $L(S)$  matching  $T$ , the look-up-table entry contains a pointer to the  $S$ -entry of table  $\mathcal{R}(T)$ , otherwise it contains a pointer to a void entry.

Random configurations of the SDs are generated by standard Metropolis Monte Carlo procedure (see [5,9] for details) at constant separation of white points  $r$ . All  $f$ -bonds (i.e., sphere overlaps) are calculated and the binary number of this configuration composed. Because  $f$  is nonzero if and only if  $e$  is zero, this number describes a diagram with nonzero  $f$ -bonds and at the same time zero  $e$ -bonds. From the look-up-table, a histogram bin (of table  $\mathcal{R}(T)$ ) is determined and incremented by 1. (The bin of the void table entry is incremented, too, but never used.)

The MC integration is sensitive to the quality of random numbers. We use double-precision numbers with all random bits based on four-tap shift-register generator [10]. Details are discussed in [11].

**2.4.1 Sums over spanning diagrams.** If we had just one SD  $T$ , the value of  $H_n(T; r)$  would be

$$H_n(r) = \Gamma(T; r) \sum_S w(S) \frac{H(S; r)}{M\lambda(S, T)}, \quad (9)$$

where  $M$  is the number of MC configurations and  $H(S; r)$  is the histogram value of unlabeled RH diagram  $S$ . The formula for  $E_n$  is analogous.

Unfortunately, one SD is not able to span all RH diagrams and several spanning diagrams have to be combined. For instance, for the calculation of  $H_6$  we used 31 SDs (this is not the minimum set). On the other hand, one RH diagram may be spanned by several spanning diagrams (on average by 19 for  $H_6$ ). “Strategy I” described in [9] thus becomes inefficient. More sophisticated “data mining” is required. We therefore adopted and extended “Strategy II” which makes use of all available information. The value of  $H_n(r)$  (and similarly  $E_n(r)$ ) is then given by a weighed sum over all used SDs

$$H_n(r) = \sum_T H_n(T; r), \quad (10)$$

where

$$H_n(T; r) = \Gamma(T; r) \sum_S W(S, T) w(S) \frac{H(S; r)}{M(T)\lambda(S, T)}. \quad (11)$$

Here  $W(S, T)$  is a weight (which is to be determined) and  $M(T)$  denotes the number of MC configurations generated in a run with SD  $T$ . If we assume that all contributions (for each MC configuration as well as each labeling out of  $\lambda(S, T)$  labelings) are independent, we get

$$W(S, T) = \frac{1}{N(S)} \frac{M(T)\lambda(S, T)}{p(S, T)[1 - p(S, T)]}, \quad (12)$$

where  $N(S)$  is the normalization constant (so that  $\sum_T W(S, T) = 1$ ) and  $p(S, T)$  is the probability of finding diagram  $S$  in samples of SD  $T$ . Because we do not know these probabilities in advance and estimate them, this estimate is biased.

The simplest estimate of  $p(S, T)$  is based on samples of  $T$  only,

$$p(S, T) = \frac{H(S; r)}{M(T)\lambda(S, T)}. \quad (13)$$

This choice uses different  $p(S, T)$  for each  $T$ . Although in the limit of infinitely many MC steps the probabilities  $p(S, T)$  do not depend on  $T$ , this is not the case of finite number of steps. The differences in  $p(S, T)$  for several  $T$  and lead to an estimator with a considerable bias; in our typical arrangement for  $n = 6$  the systematic error is of the order of one per cent and bigger than the standard error. To reduce the bias, we propose a uniform estimate of  $p(S, T)$  which uses all available information,

$$p(S, T) = \min \left\{ 1, \frac{H_n(r)}{w(S)\Gamma(T; r)} \right\}, \quad (14)$$

where  $H_n(r)$  is given by eqn (10). Formulas (10), (11), (12), and (14) thus represent an implicit equation for  $p(S, T)$  which is iterated until a self-consistent solution is obtained; approximation (13) serves as the initial estimate. The iterations converge very fast.

The minimum condition in (14) avoids problems with  $p(S, T)$  slightly above 1 which may happen if  $p(S, T) = 1$ , i.e., all configurations of certain SD  $T_1$  are for steric reasons subsets of diagram  $S$  while for

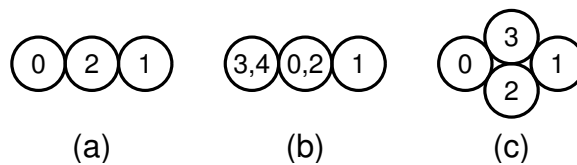


Figure 2. Configurations of spheres leading to: (a) a discontinuity in  $y(r)''$  at  $r = 2$ , (b) a discontinuity in  $B(r)^{(4)}$  at  $r = 1$ , and (c) term  $\propto (r - 3^{1/2})^{9/2}$ .

some other diagram  $p(S, T') < 1$ . In this case  $W(S, T') = 0$  (all data for  $S$  based on  $T'$  are discarded) and  $W(S, T) = 1$ . It never happens  $p(S, T_1) = p(S, T_2)$  for two different diagrams  $T_1, T_2$ .

**2.4.2 Optimization of the number of MC steps.** The final value (10) is a weighed sum of many contributions, generated by several SDs. To distribute, at least partly, the available CPU resources to calculations of individual SDs, we for  $n > 3$  adopted the following algorithm.

A test calculation was made (for  $r = 1.1$ ) with the same number of MC steps for each SD. Statistical errors of all RH diagram contributions were estimated. From this set (of 592509 data for  $H_6$ ) we selected, for each RH diagram, the SD which gives the smallest error. From this set of minimum errors we, for each SD, selected the RH diagram which has the greatest error. As a result, we get a table of min-max errors assigned to SDs. The numbers of MC steps were then set to be approximately proportional to these min-max errors.

## 2.5 Discontinuities

Because of the discontinuous nature of the hard sphere potential, functions  $E_n$  and  $H_n$  contain discontinuities in higher derivatives. These discontinuities have to be taken into account if the functions are to be expressed by analytic formulas. If data with a higher-order discontinuity are correlated by an analytical or “more continuous” function, the deviations of the fit from the MC data oscillate in the vicinity of the discontinuity point and the number of constants needed to fit the data within standard deviations increases. For instance, omitting the  $(r - 3^{1/2})^{9/2}$  term can be compensated by increasing the total number of adjustable constants by 2 or 3. On the other hand, imposing less continuity conditions may lead to functions unsuitable for some applications (e.g., if derivatives are needed).

We analyzed and used the continuity conditions up to the fifth derivative.

**2.5.1  $r = 2$ .** The most pronounced is a jump in the second derivative at  $r = 2$ . This jump arises from diagrams containing a path with one black point as in  $\gamma_1$ , eqn (7), at the configuration depicted in figure 2a. All functions but  $E_2(r)$  contain this discontinuity;  $E_2(r)$  contains its square (jump in the fourth derivative at  $r = 2$ ) because the only diagram for  $E_2 = \epsilon_2$  contains two such paths.

**2.5.2  $r = r_{\text{range}}$ .** Functions  $H_n(r)$  and  $E_n(r)$  are zero at large distances,  $r > r_{\text{range}}$ . The range is  $r_{\text{range}} = n + 1$  for  $H_n(r)$  and  $r_{\text{range}} = [n/2] + 1$  for  $E_n(r)$  ( $[\cdot]$  denotes the integer part). It holds  $H_n(r) \propto (n + 1 - r)^{2n}$  for  $r \rightarrow (n + 1)^- = r_{\text{range}}^-$  because  $H_n(r) = \gamma_n(r)$  for  $r \geq n$ . More generally,  $\gamma_k(r)$ ,  $k \leq n$ , may contribute to  $H_n(r)$  only if  $r < k + 1$  and thus  $H_n(r)$  has a jump in  $2k$ -th derivative at  $r = k + 1$ .

Function  $E_n(r)$  contains two paths between the root points. For even  $n$  the leading term at the range is given by two paths of the same length (with  $n/2$  black points) and reads as  $\gamma_{n/2}^2(r) \propto (n/2 + 1 - r)^{2n}$ . For odd  $n$  the leading term is given by one path (with  $n/2$  black points—the other path is longer) and accounts for term  $\gamma_{(n-1)/2} \propto ((n + 1)/2 - r)^{n-1}$ .

**2.5.3  $r = 1$ .** The background correlation function expresses a probability of finding a pair of cavities (of unit diameters)  $r$  apart. Therefore it is continuous also at  $r = 1$ . Discontinuities are caused by configura-

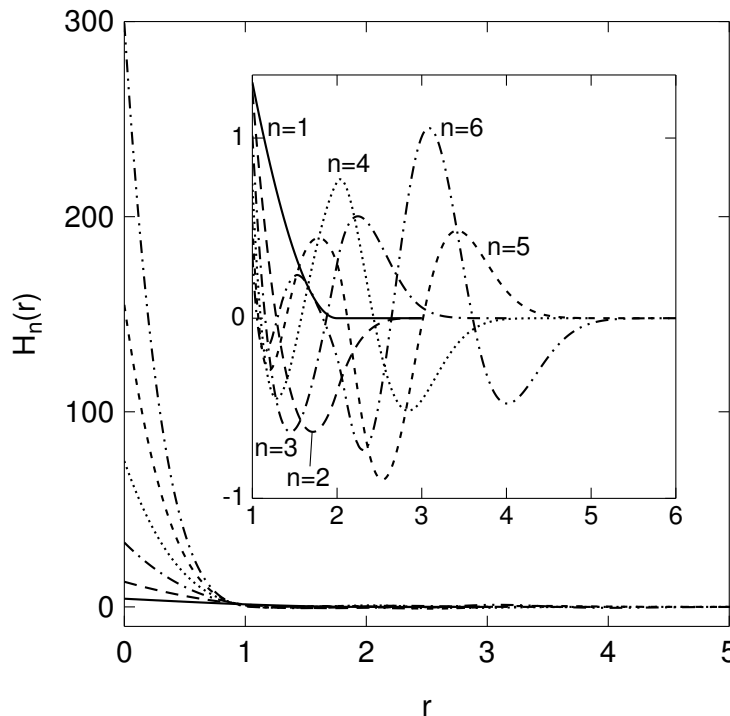


Figure 3. The expansion coefficients  $H_n(r)$  of the background correlation function.

tions of touching but not overlapping (because we have real fluid around the cavities) spheres. The above mentioned jump in  $y(2)''$  is caused by a sphere lying between a pair of cavities  $r = 2$  apart, see figure 2a. Only configurations of many spheres of this type are possible for  $r = 1$ . Therefore the first discontinuity at  $r = 1$  must be of a very high order. We assumed that the fifth derivative is continuous.

The above probabilistic argument cannot be used for the bridge function. Compact linear configurations as in figure 2b are possible for  $n > 2$ , leading to a jump in the fourth derivative at  $r = 1$ .

**2.5.4  $r = 3^{1/2}$ .** In addition, the fifth derivative  $\epsilon_2(r)^{(5)}$  diverges at  $r \rightarrow 3^{1/2}-$  due to configurations depicted in figure 2c. Detailed analysis of function (5) shows that the leading term is  $\propto (r - 3^{1/2})^{9/2}$ . Therefore a separate term proportional to  $(r - 3^{1/2})^{9/2}$  was added to splines. The nearest higher term,  $(r - 3^{1/2})^{11/2}$ , was not included.

### 3 Results

The values of functions  $H_n(r)$  and  $E_n(r)$  for  $n \leq 6$  were calculated with resolution  $\Delta r = 0.01$  in the full range  $[0, r_{\text{range}}]$ ; values of  $H_n(r)$  for  $r \in [r_{\text{range}} - 1, r_{\text{range}}]$  were calculated directly from  $H_n(r) = \gamma_n(r)$ . The functions are drawn in figure 3 and 4.

Since the data are correlated, the estimates of standard errors were determined from data blocked by  $10^8$  MC steps. The error estimates are then almost unbiased, but subject to uncertainties of about 10%. To decrease the influence of these “errors of errors” to the fitting procedure, we smoothed the standard errors; the smoothing was performed separately in  $[0, 1)$  and  $(1, r_{\text{range}})$ . In addition, too small errors close to  $r_{\text{range}}$  were enlarged. Typical standard errors in the range  $[1, 2]$  are less than  $10^{-4}$  for  $E_3$  and  $H_3$  and increase to  $10^{-3}$  for  $E_6$  and  $H_6$ . The absolute errors increase for  $r < 1$ , but the absolute values increase, too, so that the relative errors are in the range  $10^{-3}$  to  $10^{-5}$ . For increasing  $r$  the absolute errors fast decrease; the relative errors are not well defined because of null points of the functions, but near  $r_{\text{range}}$  they decrease, too.

Tables of these functions are available on the web [6].



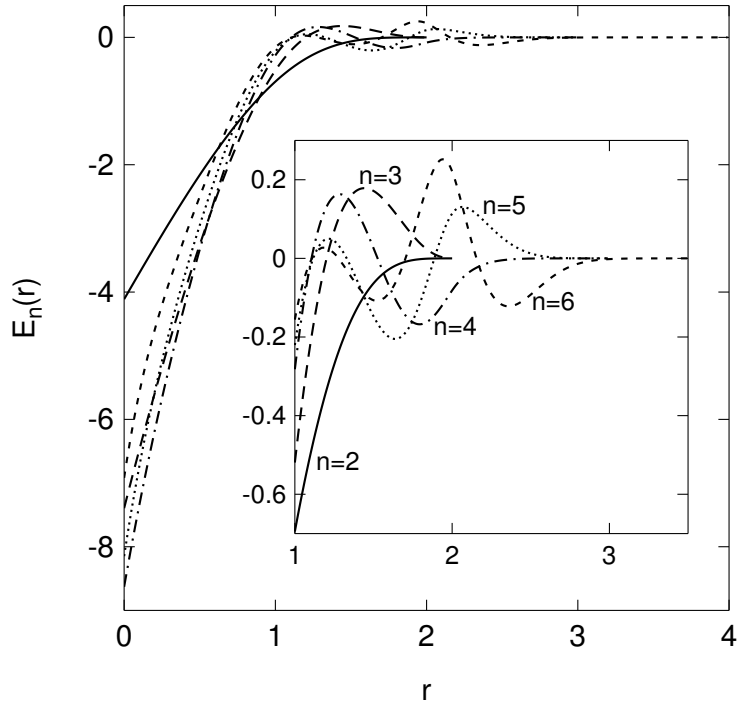


Figure 4. The expansion coefficients  $E_n(r)$  of the bridge function.

### 3.1 Tests

As a test, we investigate convergence of partial series

$$g_n(r) = \sum_{i=0}^n H_i(r) \rho^i$$

(and similarly  $B_n$ ) to the RDF and bridge functions [12] for  $r \geq 1$ . An example for  $\rho = 0.5$  is shown in figures 5 and 6 in the form of differences from the respective theoretical limit,  $g(r)$  or  $B(r)$ . It is seen that the differences converge to zero. The values of  $g_6(r)$  and  $B_6(r)$  for  $\rho = 0.2$  are indistinguishable from simulation results.

Another test is based on the relation for the compressibility factor  $Z$  [1],

$$Z = \frac{p}{\rho kT} = 1 + \sum_{n=1}^{\infty} B_{n+1} \eta^n = 1 + 4\eta g(1+),$$

where  $\eta = \pi\rho/6$  is the packing fraction. By comparing terms one gets the virial coefficients of the  $g$ -expansion (i.e., reduced by the sphere volume)

$$B_{n+2} = 4H_n(1) \left(\frac{6}{\pi}\right)^n.$$

Our results are in agreement with literature data [9, 13], see Table 2. Because the literature data are more accurate, we used these more accurate data in the correlation.

Finally, the results match within the error bars our previous data [5] for  $E_3$ ,  $E_4$ , and  $E_5$ , obtained by Mayer expansion. The new data are by about one order of magnitude more accurate and also include the  $r < 1$  region.

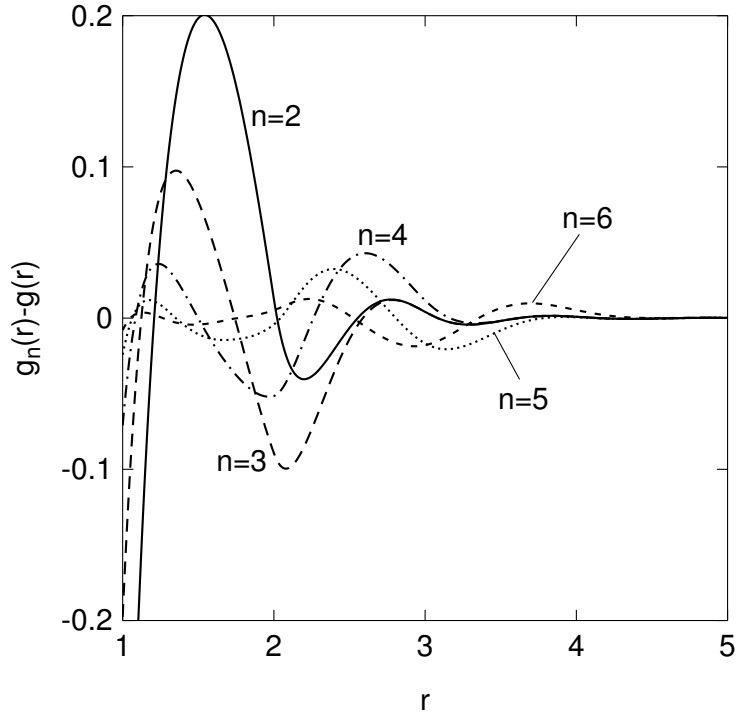


Figure 5. Convergence of the expansion (terms up to  $\rho^n$ ) to the RDF at  $\rho = 0.5$ . The differences of the partial sums from the simulation RDF are shown.

Table 2. Comparison of the virial coefficients obtained from  $g(1+)$  with literature values

$n$	this work	Ref. [9]
5	28.2302(33)	28.22445(10)
6	39.819(13)	39.81550(36)
7	53.342(19)	53.3413(16)
8	68.65(11)	68.540(10)

### 3.2 Correlation of data

In order to facilitate usage of these functions in computer code, we fitted the data to splines of variable order in intervals of lengths 1. In addition, the leading term of the discontinuity at  $r = 3^{1/2}$  was added. The final formula reads as

$$Y(r) = \sum_{i=1}^{r_{\text{range}}} (r \in [i-1, i]) \sum_{j=0}^{k_i} C_{ij} (r-i)^j + C_{9/2} (r - 3^{1/2})^{9/2}, \quad (15)$$

where symbol  $(r \in [i-1, i])$  is defined as 1 if  $r \in [i-1, i]$  and 0 otherwise and  $Y$  stands for either  $H_n$  or  $E_n$ . Coefficients  $C$  are determined by minimization of the objective function

$$\sigma^2 = \frac{1}{100r_{\text{range}} + 1 - N} \sum_{k=0}^{100r_{\text{range}}} \Delta^2(0.01k), \quad (16)$$

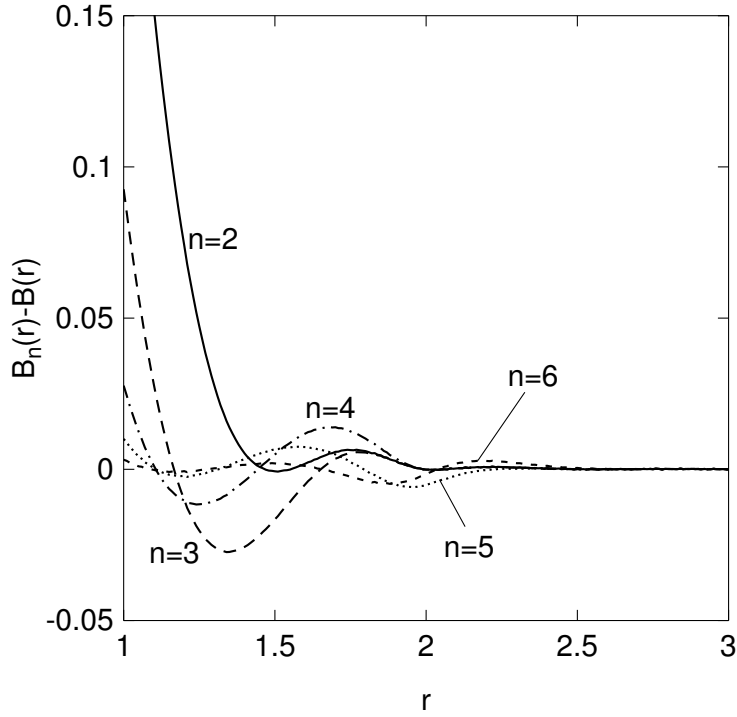


Figure 6. Convergence of the expansion (terms up to  $\rho^n$ ) to the bridge function at  $\rho = 0.5$ . The differences of the partial sums from the bridge function obtained by numerical inversion of simulation RDF are shown.

Table 3. Spline continuity conditions. The functions exhibit a jump in the derivative order indicated. For  $r = 3^{1/2}$ , see the text.

$r$	$E_3$	$E_4$	$E_5$	$E_6$	$H_3$	$H_4$	$H_5$	$H_6$
1	4	4	4	4	6	6	6	6
$3^{1/2}$	5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.5
2	2	2	2	2	2	2	2	2
3	-	8	4	4	4	4	4	4
4	-	-	-	12	$6^a$	6	6	6
5	-	-	-	-	-	$8^a$	$6^b$	$6^b$
6	-	-	-	-	-	-	$10^a$	$6^b$

<sup>a</sup> function  $\gamma_n$  at  $r = r_{\text{range}}$

<sup>b</sup> the actual order is higher

where the residual function is

$$\Delta(r) = \frac{Y_{\text{MC}}(r) - Y(r)}{\sigma_{\text{MC}}(r)} \quad (17)$$

and  $N$  is the number of independent constants  $C$ .  $Y_{\text{MC}}$  and  $\sigma_{\text{MC}}$  stand for the MC result of variable  $Y$  and its standard error, respectively.

Coefficients  $C_{ij}$  are subject to constraints to guarantee the continuity conditions, see section 2.5. Functions  $H_n$  are sewed with  $\gamma_n$  at  $r = n = r_{\text{range}} - 1$ . All continuity conditions used in splines are summarized in Table 3.

The correlation required determination of the orders of polynomials (number of constants) in all subintervals. Therefore we observed a graph of function  $\Delta(r)$  which should exhibit random scattering with normal distribution and no cumulation of errors at any subinterval. Since systematic trends are difficult to

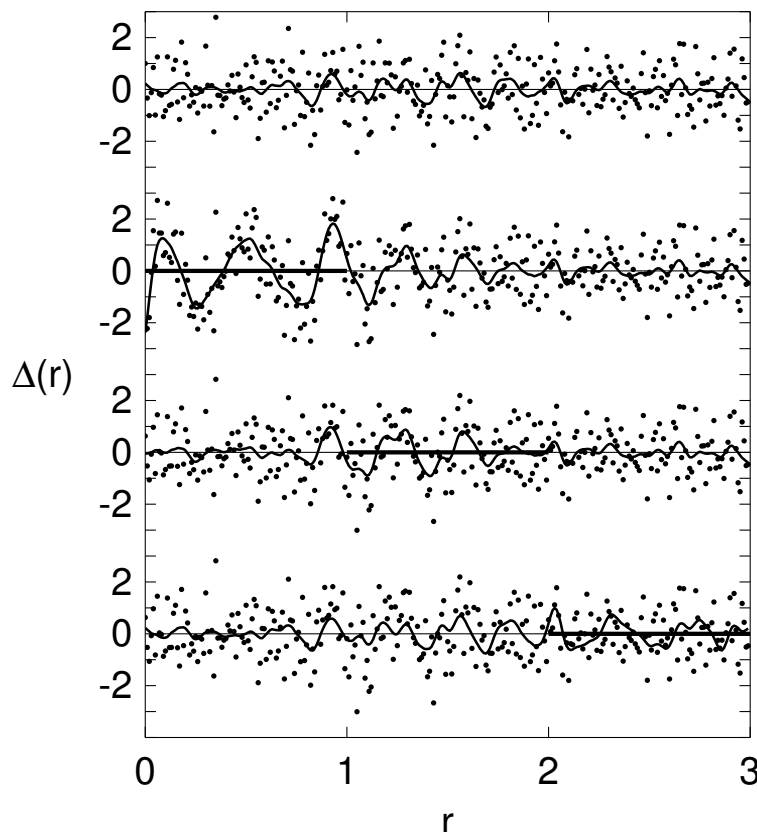


Figure 7. The residual function  $\Delta(r)$  (dots) and its smoothed version (line) of the  $E_5(r)$  fit (top curve). Second from top:  $\Delta(r)$  with the order of the polynomial in interval  $[0, 1]$  (marked by thick line) decreased from 7 to 6, third from top: in  $[1, 2]$  decreased from 8 to 7, bottom: in  $[2, 3]$  decreased from 12 to 11.

see, we also used a smoothed function (a five-point formula [14] repeated three times). The best correlation gives the value of  $\sigma$  around unity; in addition, decreasing the number of constants leads to a significant increase in  $\sigma$  while increasing the number of constants leads to only a marginal improvement.

An example of the correlation is shown in figure 7. It is seen that decreasing the order (number of constants) of the polynomials in the respective subintervals leads to an apparently smaller precision of the fit.

Functions  $H_n$  and  $E_n$ ,  $n \leq 6$ , are available [6] in the form of computer code, both in FORTRAN 77 and ANSI C version.

#### 4 Concluding remarks

The density expansion coefficients of the radial distribution and bridge functions of the pure hard sphere fluid were calculated up to  $\rho^6$  and finally expressed via general polynomial splines. The results for  $\rho^6$  are new, lower-order results are more accurate than previous data. These functions may be useful in perturbation theories of liquids.

Approaches from many branches of up-to-date computer science had to be used until final formulas (computer code) could be written: symbolic algebra, diagram analysis, high-precision random numbers, data mining, correlation by splines tailored to various discontinuities, etc.

#### Acknowledgement

This work was supported by the The Ministry of Education, Youth and Sports of the Czech Republic under the project LC512 (Center for Biomolecules and Complex Molecular Systems).

## References

- [1] J. P. Hansen, I. R. McDonald. *Theory of simple liquids*, Academic Press, London (1976).
- [2] B. R. A. Nijboer, L. van Hove. *Phys. Rev.*, **85**, 777 (1952).
- [3] Y. T. Lee, F. H. Ree, T. Ree. *J. Chem. Phys.*, **26**, 3506 (1968).
- [4] P. Attard, G. N. Patey. *J. Chem. Phys.*, **49**, 4970 (1990).
- [5] S. Labík, H. Gabrielová, J. Kolafa, A. Malijevský. *Mol. Phys.*, **101**, 1139 (2003).
- [6] <http://www.vscht.cz/fch/software/hsmid/>
- [7] S. K. Kwak, D. A. Kofke. *J. Chem. Phys.*, **122**, 104508 (2005).
- [8] F. H. Ree, W. G. Hoover. *J. Chem. Phys.*, **40**, 939 (1964).
- [9] S. Labík, J. Kolafa, A. Malijevský. *Phys. Rev. E*, **71**, 021105 (2005).
- [10] R. M. Ziff. *Comput. Phys.*, **12**, 385 (1998).
- [11] J. Kolafa, S. Labík, A. Malijevský. *Phys. Chem. Chem. Phys.*, **6**, 2335 (2004).
- [12] S. Labík, J. Kolafa, A. Malijevský. *Mol. Phys.*, **100**, 2629 (2002).
- [13] (a) N. Clisby, B. M. McCoy. E-print cond-mat/0410511; (b) N. Clisby and B. M. McCoy, *Pramana*, **64**, 775 (2005).
- [14] A. Ralston. *A first course in numerical analysis*, McGraw-Hill, NY (1965).