

**Wavelet treatment of radial distribution functions of solutes**

G. N. Chuev and M. V. Fedorov

*Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Puschino, Moscow Region 142290, Russia*

(Received 20 February 2003; published 29 August 2003)

Discrete wavelets are applied to parametrize the radial distribution functions of hydrated ions and hydrophobic solutes. The data on radial distribution functions are derived from the integral equation theory and neutron scattering experiment. The Coifman and the discrete Meyer basis sets are used for the wavelet approximation. The quality of the approximation is verified by calculations of the solvation energy, the coordination number, and the change in chemical potential of solutes.

DOI: 10.1103/PhysRevE.68.027702

PACS number(s): 02.60.-x, 61.20.-p

**I. MOTIVATION**

The solvation structure around the solutes is commonly described in terms of the radial distribution function (RDF), representing a probability distribution of interatomic distance. The RDF theory has been quite successful in describing the behavior of various liquids, but has been able to provide analytical expressions of RDF only for very simple fluids [1]. The development of a universal way for simple parametrization of RDF could promise a successful study of solvation. However, the RDF parametrization is a rather complex problem due to a complicated shape of RDF even for simple solutes (Fig. 1). The RDF reveals different features inside the two regions: it has pronounced peculiarities such as sharp peaks and slopes in the range from 0 to 5 Å (of about 2 solvent diameters  $\sigma$ ), while the function oscillates weakly in the range from  $2\sigma$  to infinity [2]. As a rule, the RDF parametrization is distinct for these ranges [3,4] and, hence, a special procedure is required to equate the approximated RDFs and their derivatives at the boundary. The quality of the RDF approximation cannot be controlled with a high accuracy due to arising additional slopes or peaks.

The compact and accurate RDF parametrization is believed to be the subject of multiresolution analysis [5,6]. At present there are special basis sets referred to as wavelets [7]. These basis sets consist of compactly supported wavelike functions, which can be shifted and scaled [6–8]. The special features of wavelets such as (bi)orthogonality and vanishing moments result in only several approximating coefficients. That is the reason why wavelets are actively used to construct distribution functions in calculations of electronic structure [9]. We want to address ourselves the question whether wavelets can be advantageous to approximate RDFs in the classical domain. There is an attempt to apply wavelet analysis for analyzing x-ray-diffraction patterns observed in glass [10], which indicates that the wavelet approach is promising for treating experimental data. The main goal of our Brief Report is to apply discrete wavelets for approximating the RDF of solutes. Although some information of the RDFs can be obtained by the extrapolation of the data on concentrated solutions [11,12], the most part of the information about RDFs of solutes is derived from simulations but not from experiments because of experimental limitations. Hence, additional efforts are required for modeling liquid structure. In this Brief Report, we will use models based on

integral equation theory [13] as well as data of RDF obtained from neutron experiments [11]. Since the main criterion of the RDF approximation concerns the calculations of various thermodynamic quantities of solvation, we will investigate the influence of the wavelet basis set not only on the quality of the RDF parametrization, but also on the thermodynamic parameters.

**II. METHOD**

We use the discrete wavelet transform [5,7] to parametrize the RDFs. Any function  $f(r) \in L^2(R)$  can be approximated with the required accuracy as a wavelet series [7]:

$$f(r) = \sum_b a_{m_0,b} \varphi_{m_0,b}(r) + \sum_{m \geq m_0}^M \sum_b d_{m,b} \psi_{m,b}(r), \quad (1)$$

where  $m \in R^+$ ,  $b \in R$ ,  $\varphi(r)$  is the scaling function, while  $\psi(r)$  is the mother wavelet of the basis set  $\{\varphi(r), \psi(r)\}$  [7]. In Eq. (1), the functions  $\varphi_{m_0,b}(r)$  and  $\psi_{m,b}(r)$  are generated by integer translation and stretching (or squeezing),  $b$  is the translation parameter, while  $m$  is the scaling factor indicating the wavelet stretching (squeezing). The first term in Eq. (1) gives the crudest approximation for  $f(r)$  at the chosen resolution (scale),  $m = m_0$ , while the second term characterizes details at various scales,  $m$  from  $m_0$  to the finest level  $M$ .

In numerical calculations of the  $a$  and  $d$  coefficients, we can avoid the direct integration using the algorithm of fast wavelet transform (FDWT) [5–7]. By choosing relevant basis functions and scales, we can nullify most part of the coefficients  $a$  and  $d$ . Then the studied function can be reconstructed with the use of only nonzero coefficients without a great loss of accuracy. The above feature of wavelet approximations is widely used to process signals and images, which data should be compressed with minimal losses.

The typical way of wavelet approximation is as follows [14]. The coefficients obtained by FDWT are sorted by decreasing their modules and only the number  $L$  of the largest coefficients is treated, while other coefficients are nullified. Then the inverse transform (recomposition) is applied. Note that the number  $L$  depends on the required accuracy. However, the conventional scheme can be hardly applied to RDFs, since the coefficients of details are to be considered at large scales due to the sharp behavior of RDFs. Apart from this, the quality of the approximation is not high because

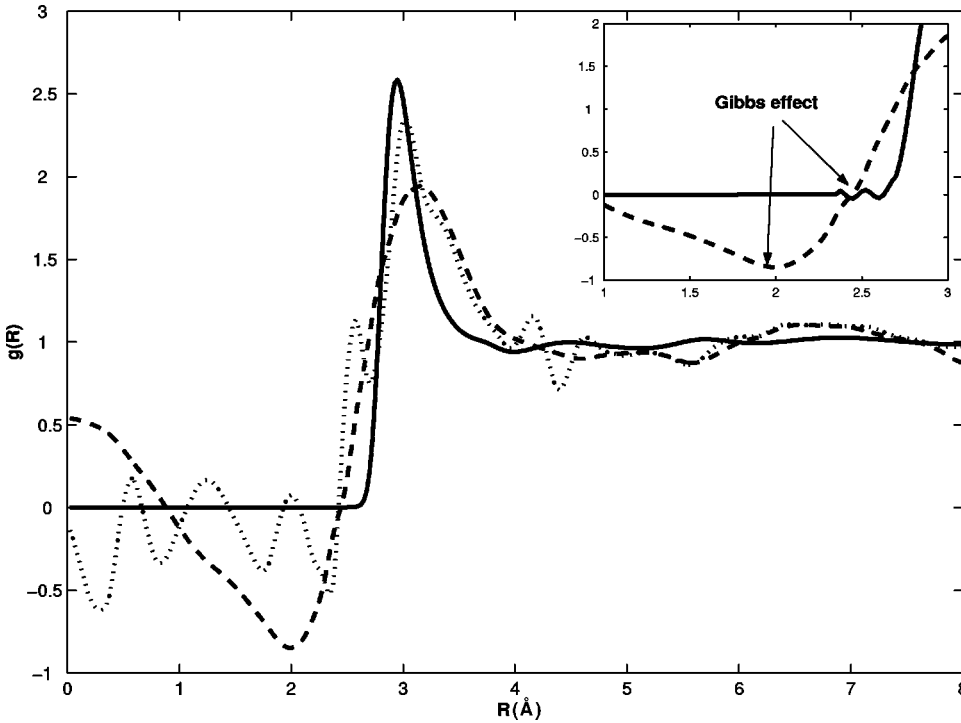


FIG. 1. Experimental potassium-water RDF [11] obtained with (dashed line) and without smoothing (dotted line), and its wavelet approximation (solid line). The inset shows the Gibbs effect: solid line corresponds to the RDF approximated by 25 FDWT coefficients, dashed line is derived by smoothing of the experimental data [11].

numerical boundary artifacts result in the Gibbs effect (inset in Fig. 1), i.e., false pulsations of the approximated RDF [7].

We have developed an advanced scheme of the wavelet approximation, which takes into account the peculiarities of RDF. First, we use a single basis set  $\{\varphi(r)\}$  by applying a nonuniform grid with a step decreasing in the vicinity of the RDF peaks. As a result, the norm of  $d$  coefficients is much less than that of  $a$  coefficients. Hence, all the  $d$  coefficients can be ignored and the RDF is approximated as a combination of scaling functions  $\varphi_{m_0,b}(r)$ . Taking into account the asymptotic behavior of RDF, we can decrease the number of required coefficients by an order. Considering the behavior  $g(R \rightarrow 0)$ , we can nullify all the wavelet coefficients corresponding to the range  $[0, 0.5R_{\max}]$ , where  $R_{\max}$  is the coordinate of the first RDF peak. On the other hand, all the  $a$  coefficients corresponding to the range  $R > 6\sigma$  are supposed to be constant depending only on the wavelet scale. Then, using the relation between functions and their wavelet coefficients [7,8], we replace the wavelet coefficients by their interpolated values. Summing all the facts, we propose the following scheme for RDF approximation: (1) we perform FDWT at the largest scale  $m_0$  satisfying the condition  $\sum_b |d_{m_0,b}| \leq 0.05 \sum_b |a_{m_0,b}|$ , then all the  $d$  coefficients are nullified; (2) all the coefficients corresponding to the range  $R \in [0, 0.5R_{\max}]$  are also nullified; (3) all the coefficients corresponding to  $R \in [6\sigma, R_\infty]$  are supposed to be constant  $C = 2^{m_0/2}$ ; (4) only the coefficients corresponding to the RDF extrema are treated, while the rest coefficients are neglected; (5) to reconstruct the RDF, we apply the linear or spline approximation to the  $a$  coefficients ignored at step (4), and use the coefficients corresponding to the RDF extrema as nodes for the approximation; (6) we perform the conventional inverse FDWT but for interpolated coefficients; (7) to suppress the Gibbs effect (inset in Fig. 1), the approximated

RDF is to be equal to zero up to  $R_{\text{cross}}$ , where  $R_{\text{cross}}$  is the largest zero point of the approximated RDF, i.e.,  $g_{\text{app}}(R_{\text{cross}}) = 0$ .

The above scheme is a combination of the conventional FDWT and the lifting procedure [8]. Due to the interpolation, the expenses are not high and can be compensated by the benefit of application of the single basis set instead of the double one. Concerning the choice of the wavelet basis set, we note that there are a lot of sets [7,8] to realize FDWT. We have used two basis sets, namely, the Coifman (C2), and the discrete Meyer (DM) sets. The main feature of the Coifman basis is that the scaling function  $\varphi(r)$  has the maximum number of vanishing moments at the fixed support. The DM basis is a discrete approximation of continuous Meyer set [6]. The advantage of the DM basis is that there are analytical expressions for basis functions in the reciprocal space [6,7]. Nevertheless, the choice of the basis set depends on the concrete problem.

### III. RESULTS AND DISCUSSION

To illustrate our scheme, we have investigated RDFs of neutral and charged hydrated solutes. We have treated the RDF of  $\text{K}^+$  taken from neutron data observed in a concentrated aqueous solution [11]. Figure 1 depicts RDF of hydrated  $\text{K}^+$  obtained with and without smoothing of the Fourier transform of the experimental data [11]. The coordination number  $N = 4\pi\rho \int_0^{r_m} g(r)r^2 dr$  (where  $r_m$  is the first minimum of the RDF) can be used for testing the accuracy of RDF data. It should be noted that both experimental RDFs predict incorrect value of the coordination number. The direct application of the Fourier transform (dashed line) to the experimental data results in the sufficiently decreased coordination number  $\approx 2$ , while the smoothed RDF yields

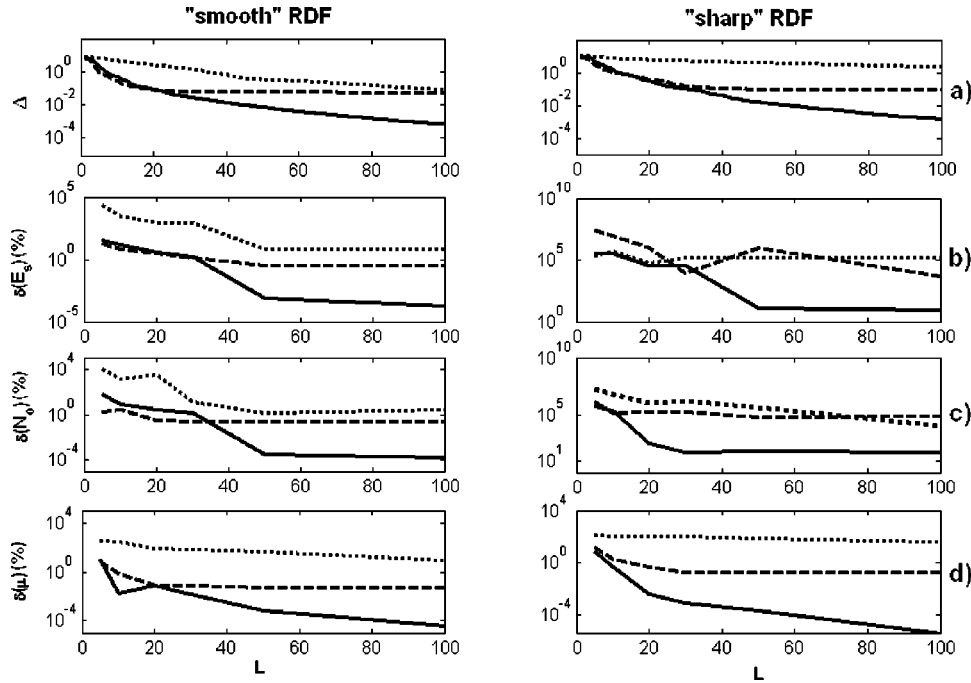


FIG. 2. The dependencies of  $\Delta$  (a) and thermodynamic parameters (b)–(d) on the number  $L$  of the coefficients. Dotted line represents cosine basis, dashed line represents DM basis, and solid line represents  $C2$  basis. The left icons correspond to the smooth RDFs, while the right ones to the sharp RDFs.

$N \approx 33$ . The latter is due to the fact that the high-frequency smoothing of RDF leads to the artificial broadening of the RDF peaks. In contrast to this, our wavelet approximation (solid line in Fig. 1) provides more adequate RDF with  $N \approx 7.5$ , which is comparable with other simulation data [15].

To reveal possibilities of wavelets for approximating RDFs derived from simulations we have calculated RDFs with the use of the integral equations based on the reference interaction site model (RISM) [16]. The correlation functions of pure water under the normal conditions were calculated by the RISM with the use of conventional procedure [13] on the grid with the number of points  $n=2048$  and step size  $\delta r = 0.025 \text{ \AA}$  for the SPC/E potential [17]. The solute-water site-site interactions are represented by the Lennard-Jones potential and the coulomb term. For hydrophobic atomic solutes, we use the size and energy parameters of Ref. [18], while for ions we apply the parameters of Ref. [19]. The RDFs obtained by the calculations are approximated by the above wavelet procedure. We calculated mean square norm  $\Delta = (1/n) \sqrt{\sum_{i=1}^n [g(r_i) - g_{app}(r_i)]^2}$ , where  $r_i = i \delta r$ ,  $g(r_i)$  is the “true” RDF and  $g_{app}(r_i)$  is the approximated RDF. Figure 2(a) depicts the dependence of the norm  $\Delta$  on the number  $L$  of the coefficients. For comparison, we also depict the corresponding dependencies for the conventional cosine approximation. Our study has revealed that the RDF under consideration can be classified in two types: smooth (for neutral solutes) and extremely sharp (for charged solutes). To control the difference between these classes, we consider the norm of numerical derivative  $\eta = (1/\delta r) \sqrt{\sum_{i=1}^{n-1} [g(r_{i+1}) - g(r_i)]^2}$ . The parameter  $\eta$  for sharp RDF is higher by a factor 2 or 3 than that for smooth functions. As it is seen from Fig. 2, the approximation of the sharp RDF includes more approximating coefficients than that of the smooth functions.

We have also evaluated thermodynamic characteristics of solvation such as the energy of solvation  $E_s$  and chemical potential  $\mu$  given by the following expressions:

$$E_s = 4\pi \sum_j \int_0^\infty u^j(r) g^j(r) r^2 dr, \quad (2)$$

$$\mu = 2\pi \sum_j \int_0^\infty \{ \gamma^j(r) [g^j(r) - 1] - 2[g^j(r) - 1] + 2\gamma^j(r) \} r^2 dr,$$

where  $j$  denotes site (hydrogen or oxygen) in water molecule,  $u^j(r)$  is the ion-oxygen (hydrogen) interaction potential, and  $\rho$  is the water density. The variable  $\gamma^j(r)$  is the indirect correlation function [1]. Figures 2(b) and 2(c) show the results of the calculations. As it is seen, the evaluation of thermodynamic characteristics is not good for the cosine approximation. The same situation takes place also for the conventional wavelet approximation. The main reason for these peculiarities is the Gibbs effect (see inset in Fig. 1), which is more intensive for the cosine basis set. The thermodynamic quantities are strongly affected by minor false pulsations of RDF at  $r \approx 0$ . Due to our scheme, we avoid the Gibbs effect and calculate accurately the thermodynamic parameters. Table I includes the obtained results for the RDFs of hydrated ions and atoms derived from the integral equation theory. In the table,  $\delta N_0$  is the relative error of the coordination number,  $\delta \mu$  is the same for the chemical potential,  $\delta E_s$  is for the solvation energy, and  $\Delta$  is the norm of inaccuracy. The symbols “*Sh*” and “*S*” denote sharp and smooth types of RDFs. We use the wavelet approximation with 5–7 coefficients for the smooth RDFs and 7–9 coefficients for the sharp functions. As it is seen, the quality of approximation is rather good for the combined scheme, while the number of approximating coefficients is small.

Our study indicates that the discrete wavelets are a suitable and powerful instrument to approximate distribution functions of classical solutes. Due to this the wavelets are well localized in the real and reciprocal space [7,14], they

TABLE I. The number  $L$  of the coefficients, relative errors for the RDFs approximated by the  $C2$  set. In parenthesis are the same values for the DM basis.

Type	Solute	$L$	$\delta N_0$ (%)	$\delta\mu$ (%)	$\delta E_s$ (%)	$\Delta$
$S$	$\text{CH}_4$	5	1.0 (2.1)	0.5 (5.0)	7.3 (3.9)	$0.8 \times 10^{-3}$ ( $0.5 \times 10^{-3}$ )
$S$	Ar	5	1.0 (1.8)	0.5 (4.7)	7.2 (3.8)	$0.8 \times 10^{-3}$ ( $0.5 \times 10^{-3}$ )
$Sh$	$\text{Cl}^-$	9	3.5 (7.0)	3.7 (5.0)	8.5 (5.0)	$1.1 \times 10^{-3}$ ( $1.2 \times 10^{-3}$ )
$Sh$	$\text{Br}^-$	9	5.1 (10.5)	4.0 (5.0)	10.5 (5.0)	$1.2 \times 10^{-3}$ ( $1.4 \times 10^{-3}$ )
$Sh$	$\text{Na}^+$	7	1.6 (3.2)	1.2 (3.7)	7.5 (4.5)	$1.1 \times 10^{-3}$ ( $1.2 \times 10^{-3}$ )
$Sh$	$\text{K}^+$	7	2.5 (4.0)	1.3 (5.5)	7.4 (4.7)	$1.2 \times 10^{-3}$ ( $1.4 \times 10^{-3}$ )
$S$	Ne	5	1.0 (1.7)	0.8 (3.5)	7.0 (3.8)	$0.8 \times 10^{-3}$ ( $0.5 \times 10^{-3}$ )
$S$	Kr	5	1.0 (1.9)	0.7 (3.4)	7.1 (3.9)	$0.8 \times 10^{-3}$ ( $0.5 \times 10^{-3}$ )
$S$	Xe	5	0.9 (1.5)	0.7 (3.7)	7.2 (3.9)	$0.9 \times 10^{-3}$ ( $0.7 \times 10^{-3}$ )

can provide accurate approximation of RDF without additional broadening of RDF peaks. Using wavelets we are able to obtain correct values of thermodynamic parameters of solvation. The applied scheme of the wavelet approximation allows us to treat RDF with small number of approximating coefficients. For example, our procedure provides the accuracy of approximation of about  $\Delta \sim 0.5 \times 10^{-3}$ , which is higher approximately by two orders of magnitude than that obtained within the approximation based on combination of polynoms, exponents, and cosines [3].

Since our procedure is based on conventional FDWT, only the coefficients of the corresponding wavelet filters should be replaced to perform the change of the basis set [7]. We have used the  $C2$  and DM wavelets taking into account the following objectives. The scaling function of the  $C2$  basis set has the maximum number of vanishing moments, but the larger is the number, the better is the approximation for singular points of the studied function [7]. Hence, using the  $C2$  wavelets we can treat accurately the sharp peaks of RDF. However, the DM wavelets are more preferable for the

calculations of thermodynamic parameters since the DM basis set is more regular than the  $C2$  basis set.

Because the real three-dimensional solvation structure is more interesting for the applications, a question arises that how the wavelet scheme can be extended to the three-dimensional case. At present there are several approaches to the problem. The most popular of them is based on the tensor product of the basis functions [20]. Another approach is to use continuous wavelets of the sombrero (Mexican hat) or Morle types [7], which have a explicit expression for the basis functions. In this case, all the coefficients can be obtained analytically. However, since the sombrero (Mexican hat) and Morle sets are not orthogonal, the obtained coefficients of the wavelet expansion do not provide exact recomposition of the studied function and can result in sufficient numerical artifacts. We believe that the search of new ways should be carried out in the reciprocal space with application of the Meyer set.

The authors are thankful to E. A. Arinstein and Gary N. Sarkisov for fruitful discussions. This work was supported by the Russian Foundation of Basic Research.

- 
- [1] J.-P. Hansen and I.R. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic, London, 1986).
- [2] The RDF is to be parametrized only inside a finite range up to  $R < R_\infty = 10\sigma$ , since the RDF is about of unity at long distances.
- [3] E. Matteoli and G.A. Mansoori, *J. Chem. Phys.* **103**, 4672 (1995).
- [4] S. Goldman, *J. Phys. Chem.* **83**, 3033 (1979).
- [5] S.G. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. (Academic Press, San Diego, 1999).
- [6] Y. Meyer, *Wavelets: Algorithms and Applications* (SIAM, Philadelphia, 1993).
- [7] I. Daubechies, *Ten Lectures on Wavelets*, CBMS/NSF Series in Applied Math. Vol. 61 (SIAM, Philadelphia, 1992).
- [8] W. Sweldens, *Appl. Comput. Harmon. Anal.* **3**, 86 (1996).
- [9] T.A. Arias, *Rev. Mod. Phys.* **71**, 267 (1999).
- [10] Y. Ding, T. Nanba, and Y. Miura, *Phys. Rev. B* **58**, 14279 (1998).
- [11] G.W. Neilson and N. Skipper, *Chem. Phys. Lett.* **114**, 35 (1985).
- [12] T. Head-Gordon and G. Hura, *Chem. Rev.* **102**, 2651 (2002).
- [13] P.A. Monson and G.P. Morriss, *Adv. Chem. Phys.* **77**, 451 (1990).
- [14] D.L. Donoho, *Appl. Comput. Harmon. Anal.* **1**, 100 (1993).
- [15] L.M. Ramanialaha, M. Bernasconi, and M. Parrinello, *J. Chem. Phys.* **111**, 1587 (1999).
- [16] D. Chandler and H.C. Andersen, *J. Chem. Phys.* **57**, 1930 (1972).
- [17] H.J.C. Berendsen, J.R. Grigera, and T.P. Straatsma, *J. Phys. Chem.* **91**, 6269 (1987).
- [18] T.P. Straatsma, H.J.C. Berendsen, and J.P.M. Postma, *J. Chem. Phys.* **85**, 6720 (1986).
- [19] B.M. Pettitt and P.J. Rossky, *J. Chem. Phys.* **84**, 5836 (1986).
- [20] H.-J. Flad, W. Hackbusch, D. Kolb, and R. Schneider, *J. Chem. Phys.* **116**, 9641 (2002).