



A new representation of the dispersion interaction

MATTHEW P. HODGES and ANTHONY J. STONE*

University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK

(Received 2 June 1999; accepted 13 October 1999)

The conventional formulation of the dispersion energy as a power series in R^{-1} with damping at small R is re-examined. Symmetry-adapted and intermolecular perturbation theory calculations have been performed and compared for the argon dimer. We show that conventional Tang–Toennies damping functions fit the data better when exchange as well as charge-overlap effects are included. We also construct a new representation of the dispersion energy which provides a more accurate, but still quite compact, description of the dispersion energy, and suggest several simpler forms which would be suitable for describing dispersion more accurately in larger systems. Some preliminary results for the HF dimer are also described.

1. Introduction

For simulations of molecular clusters and larger systems, a simple but accurate representation of the dispersion energy is an important requirement. However many potentials commonly used incorporate nothing more elaborate than the simple London form, $-C_6R^{-6}$, usually as part of an atom–atom description. Sometimes a damped form, $-f_6(R)C_6R^{-6}$, is used, where $f_6(R)$ is a damping function which suppresses the singularity at $R=0$. Our purpose here is to explore alternative forms that are potentially more accurate without being unduly complicated.

We begin by examining the simpler case of spherical atoms. For atomic systems, which lack first-order electrostatic and second-order induction interactions, the second-order dispersion energy provides the major component of the attractive potential. For large values of the interatomic separation, R , the dispersion energy between spherical atoms can be written as a series in even powers of $1/R$, so that

$$E_{\text{disp}}^{\text{mult}} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} - \dots \quad (1)$$

Here, the label ‘mult’ indicates that this form of the dispersion energy results from a multipolar expansion of the perturbation operator, and it is appealing because the long-range dispersion coefficients, C_n , are independent of R and can be related to the atomic polarizabilities at imaginary frequencies. (This follows from a second-order Rayleigh–Schrödinger perturbation theory treatment coupled with use of the Casimir–Polder identity [1].) However, equation (1) is only applicable in the limit of negligible overlap of the

charge distributions, and at short range the dispersion energy will be in error, diverging as $R \rightarrow 0$. In fact the series (1) is only asymptotically convergent [2]—that is, it is formally divergent for any finite value of R . The convergence can be improved by the use of Padé approximants, as was shown for interactions involving one-electron and two-electron atoms by Pan and Meath [3], but their approach was applied to the series (1) and so is still valid only in the limit of negligible overlap.

The invocation of zero overlap is not necessary in the formulation of a long-range perturbation theory, and in a *non-expanded* treatment which properly accounts for the charge overlap, the dispersion energy is finite for all R . The additional approximation that there is no intermolecular electron exchange is implicit in the zero-overlap model, and if standard Rayleigh–Schrödinger perturbation theory is to be used, this approximation must be retained. Methods which treat the exchange include InterMolecular Perturbation Theory [4, 5] (IMPT) and Symmetry-Adapted Perturbation Theory [6, 7] (SAPT). It is possible within a non-expanded framework to separate the various components of the dispersion into terms which correspond to the long-range multipolar contributions. By selecting excited states with particular angular momenta [8], the non-expanded dispersion energy can be written as

$$E_{\text{disp}}^{\text{non-exp}} = \sum_{l_a=0}^{\infty} \sum_{l_b=0}^{\infty} E(l_a, l_b). \quad (2)$$

At long range each $E(l_a, l_b)$ with $l_a \neq 0$ and $l_b \neq 0$ varies as $R^{-2(l_a+l_b+1)}$, and if the modification of this behaviour at short range is described by the damping function $X(l_a, l_b; R)$, then for any R we can write

* Author for correspondence. e-mail: ajs1@cam.ac.uk

$$E(l_a, l_b) = -X(l_a, l_b; R)C(l_a, l_b)R^{-2(l_a+l_b+1)}. \quad (3)$$

From equations (1)–(3) and the long-range behaviour of the $E(l_a, l_b)$ it follows that

$$\begin{aligned} C_6 &= C(1, 1), \\ C_8 &= C(1, 2) + C(2, 1), \\ C_{10} &= C(1, 3) + C(2, 2) + C(3, 1), \end{aligned} \quad (4)$$

and for identical atoms $C(l_a, l_b) = C(l_b, l_a)$. The terms in equation (2) with $l_a = 0$ or $l_b = 0$ contribute to the ‘spherical’ dispersion energy. They decay exponentially and therefore do not correspond to any long-range multipolar terms. For $\text{H} \cdots \text{H}$ these terms prove to be quite significant at small R [9].

In an alternative formulation, which is commonly used, the spherical dispersion energy is ignored, and the dispersion energy is written as

$$E_{\text{disp}}^{\text{non-exp}} = -f_6(R) \frac{C_6}{R^6} - f_8(R) \frac{C_8}{R^8} - f_{10}(R) \frac{C_{10}}{R^{10}} - \cdots, \quad (5)$$

where the f_n are also known as damping functions. Evidently, by comparison with equations (3) and (4),

$$\begin{aligned} f_6(R) &= X(1, 1; R) \\ f_8(R) &= X(1, 2; R) + X(2, 1; R), \\ f_{10}(R) &= X(1, 3; R) + X(2, 2; R) + X(3, 1; R) \end{aligned} \quad (6)$$

and so on.

Various forms have been proposed for the $f_n(R)$; some have been fitted to the results of accurate computations [8, 10], usually on $\text{H} \cdots \text{H}$ [9, 11–15]. The most widely used damping functions are probably those of Tang and Toennies, who developed a damping model which is based on coupling the short-range potential (usually a Born–Mayer fit of the Hartree–Fock repulsion) with the dispersion interaction which is approximated by a semi-classical, harmonic Drude model [16]. The original Tang–Toennies model was later modified [17] and the damping function for each R^{-n} term was determined as an incomplete gamma function of order $n + 1$, i.e.

$$f_n(R) = 1 - \exp(-bR) \sum_{k=0}^n \frac{(bR)^k}{k!}, \quad (7)$$

where b is the scale parameter from the Born–Mayer fit of the Hartree–Fock data. The $f_n(R)$ vary as R^{n+1} as $R \rightarrow 0$ so that each damped term tends to zero for small R , which is formally incorrect, but is not expected to introduce a significant error except at very small R . The damping functions for $\text{H} \cdots \text{H}$ were compared with those calculated by Koide *et al.* [9] for $n = 6$ –20 and good agreement was found up to $n = 10$ over a wide range of R . The Tang–Toennies model has been widely

used and provides an effective way to obtain a short-range dispersion potential from the long-range form of equation (1). The damped series (5) is often truncated after the R^{-8} or R^{-10} term, but in recent accurate work on the He dimer it was found necessary to include terms up to R^{-16} to obtain satisfactory agreement with the calculated dispersion energy [18, 19].

This approach is useful if the C_n are known. However it is becoming common, especially for molecules, to calculate dispersion energies at a number of geometries and to fit an analytical function to the results. In this case it is not clear that (5) is the most suitable form to use.

In this paper we investigate a different formulation of the dispersion energy. Using E_{disp} generically to refer to the dispersion energy with or without corrections for exchange-dispersion effects, we write

$$E_{\text{disp}} = -G(R) \times \frac{C_6}{R^6}, \quad (8)$$

where $G(R)$, defined by this equation, describes both the effects of the terms in higher powers of R^{-1} at intermediate range and the damping due to overlap at short range, as well as exchange effects if they are included in E_{disp} . $G(R)$ tends to 1 at large R , where the R^{-6} term in equation (1) dominates. At intermediate distances, in the region of the energy minimum, the terms in R^{-8} , R^{-10} , etc. become important and $G(R) > 1$. Near $R = 0$, E_{disp} remains finite, so $G(R) \sim R^6$. In the absence of any fundamental theory leading to an analytical expression for $G(R)$, we look here for empirical representations. Ideally, a $1/R$ expansion of $G(R)$ will recover the conventional long-range expansion in the form of equation (1), but we do not view that as an important requirement since our main concern is with the region near the potential-energy minimum.

We begin by examining the argon dimer. We are interested in the possibility of describing the dispersion energy more accurately than conventional methods allow, as well as in the possibility of finding simple expressions for the dispersion energy that provide moderate accuracy with only a small number of adjustable parameters. The latter objective is particularly important for applications to large molecules, where simple atom–atom dispersion potentials comprising only the undamped $-C_6R^{-6}$ term are still widely used. We consider the case of molecules briefly at the end of this paper, using the example of the HF dimer.

2. The argon dimer

We have chosen the argon dimer as a test case. Our objective is not to provide a definitive dispersion potential, but to explore the usefulness of various analytical

functions for describing the calculated dispersion energy. First we describe the perturbation theory calculations that we have carried out on this system. We then briefly discuss weighting schemes for the fitting procedures followed by the results of fitting the long-range and short-range data.

2.1. Perturbation theory calculations

We have performed calculations on the argon dimer using both the IMPT and SAPT methods with an aug-cc-pVTZ basis set [20] for $2 \leq R \leq 30$ bohr. We see this as an opportunity to test the accuracy of the IMPT results by comparison with SAPT, which can account for additional intramolecular correlation corrections. The terms available from the SAPT program [21] are: $E_{\text{disp}}^{(20)}$, the dispersion energy calculated with Hartree-Fock monomer wavefunctions; $E_{\text{disp}}^{(21)}$ and $E_{\text{disp}}^{(22)}$, intramolecular correlation corrections to the dispersion energy; and $E_{\text{exch-disp}}^{(20)}$, the modification of the dispersion energy due to a coupling with the exchange. The IMPT dispersion energy, $E(\text{IMPT})$, has no intramolecular correlation corrections, but since the dimer wavefunction is fully antisymmetrized, it does account for exchange effects. We consider four combinations of the SAPT data, which we define as:

$$E(\text{SAPT-20}) = E_{\text{disp}}^{(20)}, \quad (9)$$

$$E(\text{SAPT-20X}) = E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)}, \quad (10)$$

$$E(\text{SAPT-2N}) = E_{\text{disp}}^{(20)} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)}, \quad (11)$$

$$E(\text{SAPT-2NX}) = E_{\text{disp}}^{(20)} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)} + E_{\text{exch-disp}}^{(20)}. \quad (12)$$

These cover the inclusion and exclusion of both intramolecular correlation and exchange effects, although note that the exchange corrections to the $E_{\text{disp}}^{(2n)}$ terms for $n = 1$ and 2 are absent from $E(\text{SAPT-2NX})$. We can compare each of the four combinations with the IMPT results, which we expect to agree best with $E(\text{SAPT-20X})$ since it includes exchange effects but not intramolecular correlation. Log plots of the IMPT and the four SAPT combinations are presented in figure 1, which indeed shows that at long range, where the exchange effects are negligible, $E(\text{IMPT})$ is in agreement with $E(\text{SAPT-20})$ and $E(\text{SAPT-20X})$. Furthermore, at short range $E(\text{IMPT})$ follows the $E(\text{SAPT-20X})$ data closely to separations as small as about 5 bohr. Note that the equilibrium separation of the dimer is about 7.1 bohr, and here $E_{\text{disp}}^{(20)} = -2.40$ and $E_{\text{exch-disp}}^{(20)} = 0.144 \text{ kJ mol}^{-1}$, so the exchange contribution to the dispersion is not insignificant. It becomes more important as R decreases, and at 3 bohr, $E_{\text{disp}}^{(20)} = -299$ and $E_{\text{exch-disp}}^{(20)} = 70.7 \text{ kJ mol}^{-1}$.

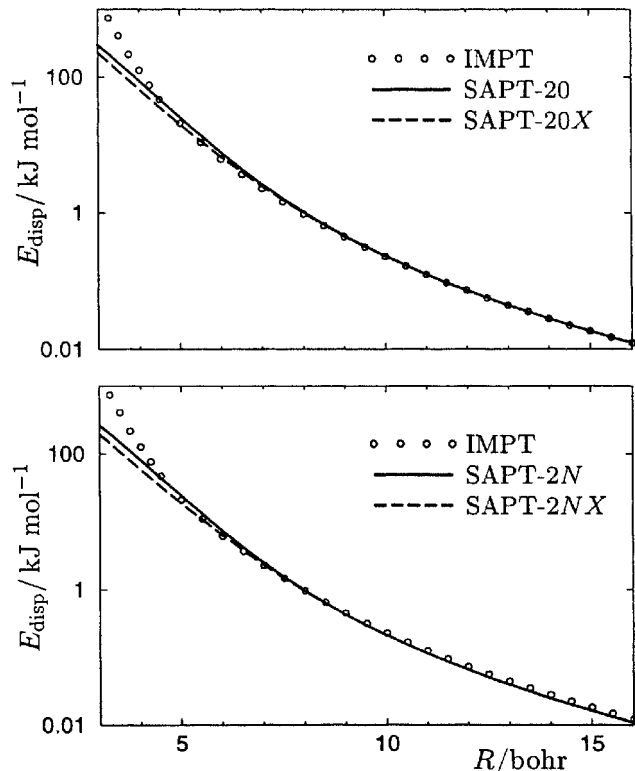


Figure 1. Plot of the dispersion energy calculated with perturbation theory methods against interatomic separation; the perturbation theory methods are described in the text.

2.2. Weighting schemes

One of the general problems associated with fitting dispersion energy data is that the energies span a large range of values—in our case, about six orders of magnitude. A uniform weighting scheme does not seem appropriate here, and we have opted, based on trial fittings, to adopt a ‘relative’ weighting scheme when fitting the energies. In this case if the data point E_i^{calc} is approximated in the fit by E_i^{fit} we minimize the root-mean-square (rms) error, defined for N points as

$$\left[\frac{1}{N} \sum_{i=1}^N \left(\frac{E_i^{\text{fit}} - E_i^{\text{calc}}}{E_i^{\text{calc}}} \right)^2 \right]^{1/2}. \quad (13)$$

This is equivalent to weighting each data point in the fit by $(E_i^{\text{calc}})^{-2}$. We have also used the pseudo-damping function, $G(R)$, defined in equation (8), which implicitly accounts for all higher C_n than C_6 and all the $f_n(R)$. We fit this function directly, using a uniform weighting scheme; this gives a distribution of weights of the energy similar to the relative weighting scheme described above. The $G(R)$ function also proves to be useful for visualizing fits of the dispersion energy, as discrepancies between the calculated and fitted data show up more clearly.

Table 1. Dispersion coefficients C_n and rms errors for fits of the long-range IMPT and SAPT dispersion energy data. The number of terms used to fit the data is governed by the value of n_{\max} . The C_n are given in atomic units.

n_{\max}	C_6	C_8	C_{10}	rms error/%
IMPT				
6	76.752			0.93
8	74.540	1188.6		0.0090
10	74.618	1103.7	20810	0.00014
SAPT-20				
6	76.753			0.93
8	74.543	1187.5		0.0096
10	74.626	1097.3	22122	0.00049
SAPT-2N				
6	67.514			1.1
8	65.202	1243.9		0.013
10	65.300	1137.2	26181	0.00076

2.3. Long-range fits

We have fitted the IMPT, SAPT-20 and SAPT-2N data for $R \geq 18$ bohr to the long-range form of the dispersion energy given in equation (1), and we have varied the maximum value of n , n_{\max} , up to $n = 10$. The fitted dispersion coefficients and rms errors are given in table 1. Several points emerge from these fits. First, the IMPT and SAPT-20 results are very similar, and the differences in the values of the C_6 coefficient are very small for each value of n_{\max} . For $n_{\max} = 10$, values of C_8 and C_{10} vary more, but in each case the three-parameter fit results in a very small rms error. Not surprisingly, the quality of the fit is much more sensitive to the C_6 value than to either C_8 or C_{10} . Changing the value of C_6 in the three-parameter IMPT fit by 1% yields a rms error of about 1.0%, whereas an equivalent fractional change in either C_8 or C_{10} yields rms errors of about 0.030%.

Secondly, the intramolecular correlation effects are quite significant. For the fit with $n_{\max} = 10$, the SAPT-2N value of C_6 , which includes intramolecular correlation, is 12.5% smaller than the SAPT-20 value. Recent literature values for C_6 are 63.6–70.8 [22] and 64.30 [23] where the latter is estimated to be accurate to about 1%. Our best estimate of 65.30 lies slightly outside these error limits, but it is apparent that if quantitative accuracy is to be approached, then intramolecular correlation effects must be considered. Our best estimates for the higher dispersion coefficients are $C_8 = 1137.2$ and $C_{10} = 26181$ which can be compared with the bounds of 1180–1880 for C_8 and 34900–60900 for C_{10} given by Standard and Certain [22]. Calculations by Knowles and Meath [8] with their ‘Basis II’ yielded values of $C_8 = 1538$ and $C_{10} = 46613$, while Thakkar *et al.* [24], using many-body perturbation theory,

obtained values of $C_8 = 1623$ and $C_{10} = 49063$. The reason for the large discrepancies between our value of C_{10} and the other literature values is that, because of basis set limitations, we can only recover the $C(2,2)$ component of C_{10} (see equation (4)). For this reason it is more instructive to compare our C_{10} with $C(2,2)$ values from the literature where available. Tang *et al.* [25] reported lower and upper bounds for $C(2,2)$ of 9900 and 22300 respectively. We note that their predicted range of values for C_{10} was 30450–55070 which can be compared with the later values of Standard and Certain given above. Knowles and Meath’s value of $C(2,2)$, again using their ‘Basis II’, was 18079. Clearly a larger basis with more high angular momentum basis functions is needed if a good account of the C_{10} coefficient is to be obtained. However, the SAPT program [21] is interfaced in our implementation with CADPAC [26] which at present cannot deal with the g -functions which form part of the aug-cc-pVQZ basis, for example.

This comparison confirms, as one would expect, that fitting of long-range *ab initio* dispersion energy data to equation (1) does not give accurate dispersion coefficients for basis sets of moderate quality. As a method for determining dispersion interactions between molecules it is not likely to be useful. We shall show, however, that a good description is possible if we adopt a different approach.

2.4. Short-range fits

We first fit the data by damping the dispersion using the incomplete gamma functions proposed by Tang and Toennies [17]. We have decided to keep the value of C_6 fixed at the values determined from the corresponding long-range fitting procedures and to treat the C_8 and C_{10} coefficients as parameters of the fit. We do not consider the IMPT results further, but concentrate on the SAPT quantities defined in equations (9)–(12). We have also considered the effect of excluding data points for which R is less than a minimum intermolecular separation, R_{\min} . This is important because it is difficult to assess the reliability of the SAPT results for small R , where the perturbation is necessarily larger. Values of R_{\min} in the range $2.0 \leq R_{\min} \leq 4.0$ bohr have been used, and the results are gathered in table 2 and in figures 2 and 3.

First we consider just those fits which have $R_{\min} \geq 3.0$. For these results, it is quite clear that the Tang–Toennies damping functions are able to describe the SAPT-20X and SAPT-2NX data significantly better than the SAPT-20 and SAPT-2N data. It is generally considered that the Tang–Toennies model cannot be expected to account for exchange effects, but we see that if it is treated simply as a model with parameters to be fitted to dispersion energy data it performs better when exchange is included. We do not consider a

Table 2. Dispersion coefficients and rms errors for fits of the SAPT dispersion energies using Tang–Toennies damping functions for a range of R_{\min} . The C_n are given in atomic units, R_{\min} in bohr and b in bohr $^{-1}$.

R_{\min}	C_8	C_{10}	b	rms error/%
SAPT-20				
2.0	62.443	158 310	1.865	3.6
2.5	293.95	131 940	1.914	2.4
3.0	436.33	115 920	1.952	1.8
3.5	551.82	102 720	1.992	1.4
4.0	640.90	92 303	2.036	1.1
SAPT-20X				
2.0	165.27	146 750	1.737	5.9
2.5	764.30	75 295	1.870	1.5
3.0	905.50	59 719	1.913	0.57
3.5	880.26	62 597	1.903	0.53
4.0	846.98	66 599	1.888	0.45
SAPT-2N				
2.0	86.172	172 330	1.813	3.9
2.5	311.26	145 080	1.859	2.7
3.0	449.88	128 500	1.894	2.0
3.5	564.97	114 520	1.932	1.6
4.0	655.10	103 350	1.973	1.2
SAPT-2NX				
2.0	-55.547	198 230	1.636	7.7
2.5	689.50	99 297	1.779	1.8
3.0	836.85	81 418	1.817	0.75
3.5	813.79	84 301	1.809	0.72
4.0	793.49	86 957	1.801	0.71

detailed analysis of the fitted values of C_8 and C_{10} to be particularly helpful since they might account in some average way for such factors as varying proportions of the $E(l_a, l_b)$ being captured at different values of R . We note, however, that more consistent sets of fitted values of C_8 and C_{10} are apparent for the SAPT-20X and SAPT-2NX data.

We also note that the fitted values do not correspond at all to the values obtained by an undamped fit to the long-range data, or to the values obtained by direct calculation [8, 24]. The reason for this is that the function being fitted, equation (5), does not provide an exact representation of the dispersion energy, even with the exact C_n , because the Tang–Toennies damping functions do not describe the damping exactly, nor is it likely that the exact C_n provide the best approximation to the data. There would be no advantage in constraining the C_n to the exact long-range values, because the important region of the dispersion energy for most purposes is the region near the potential minimum, where the long-range terms are substantially modified by damping. Accurate values of C_8 and C_{10} are not prerequisites for an accurate fit, nor is this fitting procedure a valid way to obtain the long-range values.

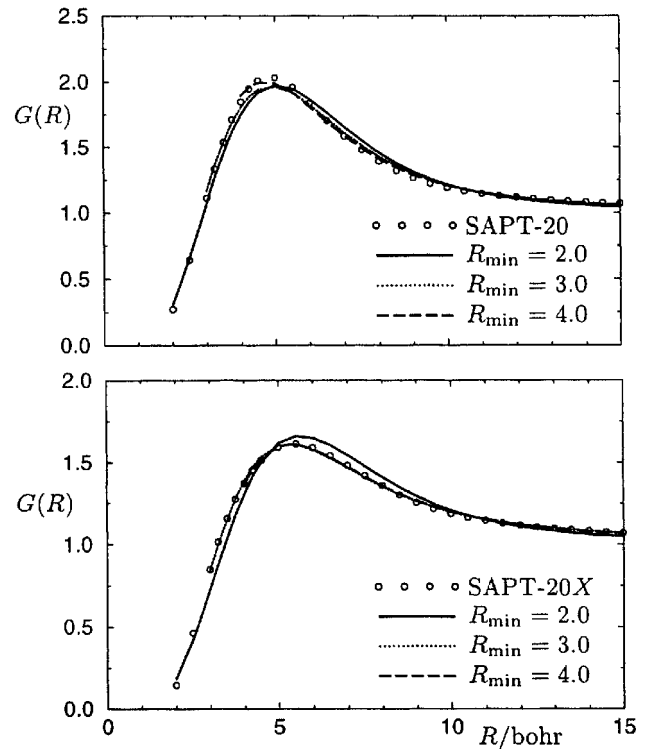


Figure 2. Fits of the SAPT-20 and SAPT-20X data using Tang–Toennies damping functions. C_8 , C_{10} and b are treated as parameters in the fitting procedure. R_{\min} is given in bohr.

For $R_{\min} < 3.0$, the quality of the fits is decidedly worse. This might be attributable to a breakdown of the model, and one reason could be that for very small values of R the Tang–Toennies damping functions are not physically realistic, damping each term in the dispersion series to zero at $R = 0$. Another possible reason, as mentioned above, is that the SAPT results are likely to be less reliable for small R .

To try to find more accurate models for the dispersion energy, we return to the pseudo-damping function, $G(R)$, introduced above. Comparison of equations (5) and (8) shows that $G(R)$ can be related to the conventional damping functions, $f_n(R)$, by the expression

$$G(R) = f_6(R) + \frac{C_8/C_6}{R^2} f_8(R) + \frac{C_{10}/C_6}{R^4} f_{10}(R) + \dots \quad (14)$$

This function should tend to 1 as $R \rightarrow \infty$ and be of order R^6 for small R if the dispersion energy is to remain finite at $R = 0$. To provide approximations to the $G(R)$ function, we turn to a series of rational functions of the general form

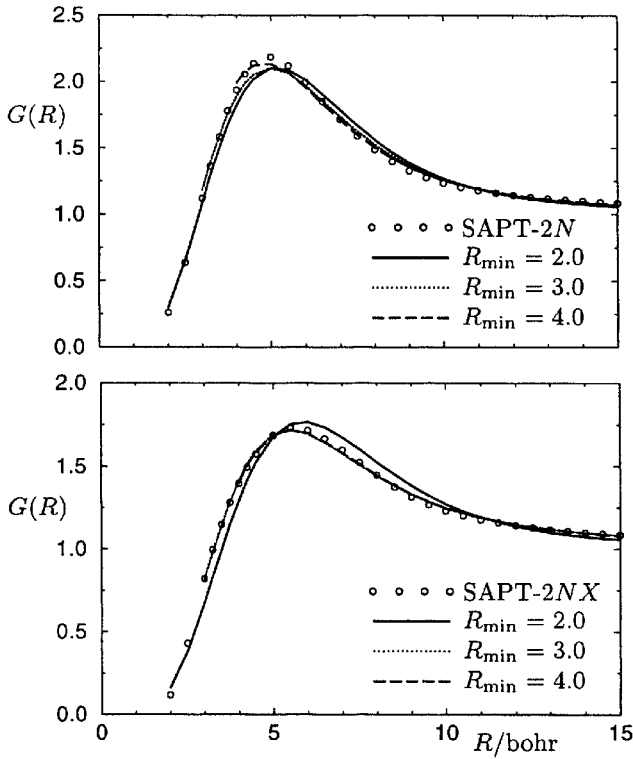


Figure 3. Fits of the SAPT-2*N* and SAPT-2*NX* data using Tang–Toennies damping functions. C_8 , C_{10} and b are treated as parameters in the fitting procedure. R_{\min} is given in bohr.

$$g_n^m(R) = \frac{\sum_{i=0}^m a_i R^{-i}}{\sum_{j=0}^n b_j R^{-j}}. \quad (15)$$

To get the correct long-range behaviour, we set both a_0 and b_0 to 1, and to get the desired short-range behaviour we specify that $m = n - 6$ and define

$$g_n(R) = \frac{\sum_{i=0}^{n-6} a_i R^{-i}}{\sum_{j=0}^n b_j R^{-j}}, \quad (16)$$

where the a_i and b_j are zero for odd values of i and j respectively.

We have performed trial fits of $G(R)$ with a uniform weighting scheme (see section 2.2) and have found that the first three rational functions, g_6 , g_8 and g_{10} are prone to singularities for $R < R_{\min}$, i.e. outside the range of the data that we include. We therefore concentrate on the g_{12} function, which is well behaved for all values of R_{\min} which we discuss. Problems with the g_{12} function are encountered for values of R_{\min} greater than 3.5 bohr. We present fits of the dispersion energy for all four combinations of the SAPT data at $R_{\min} = 2.0, 2.5$ and 3.0 bohr. The parameter values used in the g_{12} function are listed in table 3 for $R_{\min} = 2.5$ bohr.

The rms errors are collected in table 4 and we give the Tang–Toennies errors for comparison. The fits with $R_{\min} = 2.0$ bohr (see figures 4 and 5) are significantly worse than for the other values of R_{\min} and this may be because our limiting behaviour as $R \rightarrow 0$ is not appropriate for the data that we are fitting. Our functions provide considerably better approximations to the SAPT data than the Tang–Toennies functions, and the improvement is more marked for the SAPT-20 and SAPT-2*N* data where the Tang–Toennies functions perform less well. The g_{12} function may seem to involve an unacceptable proliferation of parameters, but we stress that it can be evaluated many times more quickly than the incomplete gamma function. The quality of our fits is quite consistent for all the combinations of the SAPT data, and in general the rms error is about 0.1%, when $R_{\min} \geq 2.5$ bohr; given that the data being represented span six orders of magnitude this is very encouraging.

Table 3. Parameter values used in the g_{12} rational approximations to the SAPT dispersion energies. (+ n) represents 10^n . Values are in atomic units: bohr n for a_n and b_n . $R_{\min} = 2.5$ bohr.

	SAPT-20	SAPT-20 <i>X</i>	SAPT-2 <i>N</i>	SAPT-2 <i>NX</i>
a_2	−3.888 (+1)	−5.178 (+1)	−5.593 (+1)	−6.469 (+1)
a_4	2.707 (+3)	2.827 (+3)	3.317 (+3)	3.305 (+3)
a_6	5.625 (+4)	5.383 (+4)	5.642 (+4)	4.478 (+4)
b_2	−5.500 (+1)	−6.757 (+1)	−7.501 (+1)	−8.344 (+1)
b_4	3.893 (+3)	4.069 (+3)	5.007 (+3)	5.009 (+3)
b_6	−6.626 (+4)	−5.929 (+4)	−9.719 (+4)	−9.729 (+4)
b_8	1.309 (+6)	1.655 (+6)	1.749 (+6)	2.268 (+6)
b_{10}	−5.860 (+6)	−1.092 (+7)	−8.561 (+6)	−1.582 (+7)
b_{12}	2.340 (+7)	4.915 (+7)	3.074 (+7)	6.308 (+7)

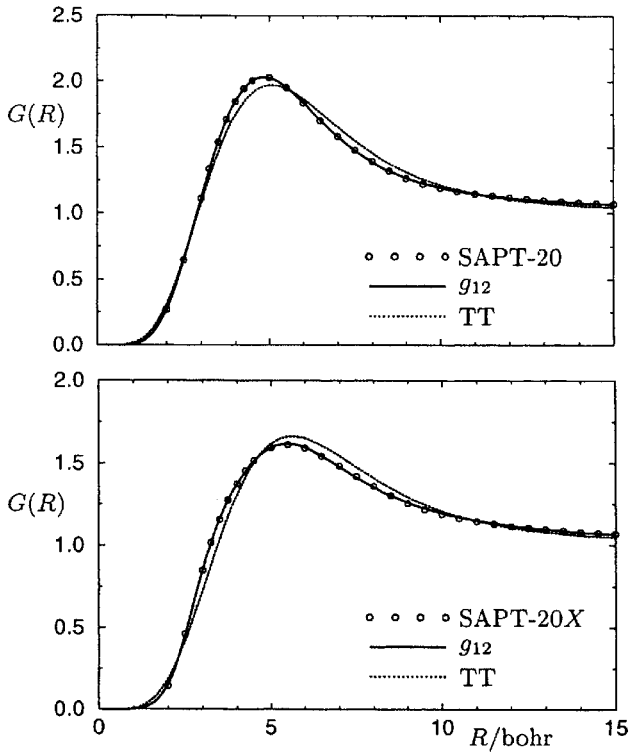


Figure 4. Fits of the SAPT-20 and SAPT-20X $G(R)$ functions using the g_{12} rational approximation for $R_{\min} = 2.0$ bohr. The Tang–Toennies (TT) results are shown for comparison.

We can assess whether fits to $G(R)$ using $g_n(R)$ are physically meaningful at long range by expanding equation (16) as a power series in R^{-1} . For example, for $n = 6$

$$\begin{aligned} g_6(R) &= 1 - (b_2R^{-2} + b_4R^{-4} + b_6R^{-6}) \\ &\quad + (b_2R^{-2} + b_4R^{-4} + b_6R^{-6})^2 - \dots \\ &= 1 - b_2R^{-2} + (b_2^2 - b_4)R^{-4} + \dots \end{aligned} \quad (17)$$

Comparing with equation (14) and remembering that each $f_n(R) \rightarrow 1$ as $R \rightarrow \infty$, we can derive the following relationships:

$$\begin{aligned} C_8/C_6 &= -b_2, \\ C_{10}/C_6 &= b_2^2 - b_4. \end{aligned} \quad (18)$$

More generally, i.e. for any $g_n(R)$, these quantities are given by

$$\begin{aligned} C_8/C_6 &= a_2 - b_2, \\ C_{10}/C_6 &= a_4 + b_2^2 - b_4 - a_2b_2. \end{aligned} \quad (19)$$

Applying this approach, we obtain the values of C_8 and C_{10} shown in table 4. It is apparent that meaningful values of C_{10} cannot be obtained from this dataset using this model, although the values of C_8 seem to be quite reasonable for $R_{\min} \geq 2.5$ bohr and in good agreement

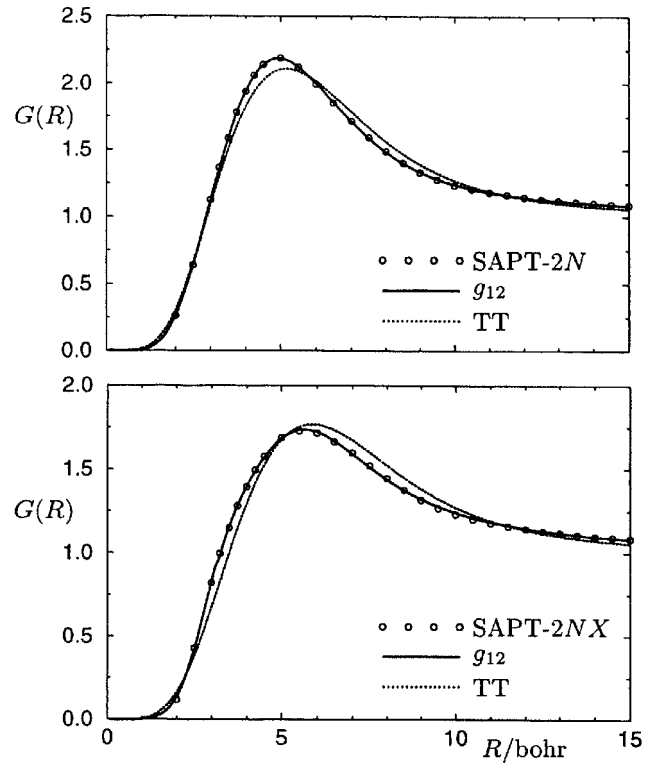


Figure 5. Fits of the SAPT-2N and SAPT-2NX $G(R)$ functions using the g_{12} rational approximation for $R_{\min} = 2.0$ bohr. The Tang–Toennies (TT) results are shown for comparison.

with those calculated from the long-range data in table 1. As with the Tang–Toennies model, however, this fitting procedure cannot be regarded as a useful method for finding the true long-range C_n . It nevertheless provides a very accurate fit to the dispersion energy data over the whole range fitted—better than 0.1% if the points at very small R are excluded.

It is possible that the poor values for C_{10} reflect the quality of the data that we are fitting, the deficiencies of which have already been mentioned. To test this assumption, we have fitted Koide *et al.*'s data for H \cdots H. We have included just those components corresponding to terms up to C_{10} . For $R_{\min} = 2$ bohr, we obtain values of 130.7 and 130.4 for C_8 with and without the spherical dispersion contribution respectively. The asymptotic limit of C_8 for this system is about 124.4 au. For C_{10} , the asymptotic limit is about 3286 au and our corresponding fitted values are 1234 and 1247. These values of C_8 and C_{10} are not too unreasonable, and we also note that they are not very sensitive to the inclusion of the spherical dispersion contribution. We see once again, however, that although the g_{12} representation describes the data very accurately, it is not sufficiently sensitive to the long-range behaviour to give accurate values for C_8 and C_{10} . The problem is

Table 4. Dispersion coefficients and rms errors for fits of the SAPT dispersion energies using a g_{12} rational approximation for a range of R_{\min} . The corresponding Tang–Toennies rms errors are given for comparison. The C_n are given in atomic units and R_{\min} in bohr.

R_{\min}	g_{12}			Tang–Toennies
	C_8	C_{10}	rms error/%	rms error/%
SAPT-20				
2.0	1225.9	−38 033	0.14	3.6
2.5	1202.4	−22 390	0.10	2.4
3.0	1189.8	−14 813	0.086	1.8
SAPT-20X				
2.0	918.05	51 468	0.83	5.9
2.5	1178.4	−13 071	0.080	1.5
3.0	1165.8	−6 127.2	0.066	0.57
SAPT-2N				
2.0	1272.2	−32 651	0.20	3.9
2.5	1246.1	−16 854	0.12	2.7
3.0	1231.5	−9 095.5	0.10	2.0
SAPT-2NX				
2.0	926.74	61 136	1.7	7.7
2.5	1224.2	−9 182.5	0.10	1.8
3.0	1210.1	−2 394.0	0.079	0.75

that the g_{12} function cannot be expected to give an exact description of the dispersion energy at all distances, and in order to get a good fit to the short-range data it is necessary to push the values of C_8 and C_{10} away from their true long-range values. We emphasize, however, that the quality of the fit in the long range region, where the dispersion energy is very small, is much less important than the fit in the region of the minimum. In fact the fit is very good throughout the range of values of R .

It is possible to constrain the fit so that the values of C_8 and C_{10} given by equation (19) are the same as those given by the long-range fit. For Ar₂, this increases the rms error in the SAPT-2NX case from about 0.1% to about 0.125% when the C_8 constraint is applied, and to about 0.5% when the C_{10} constraint is also applied. However, there seems to be little advantage in doing this; the improved description at long range, which is not very important, is obtained at the expense of the fit in the more important short-range region.

3. Simpler models

So far we have been concentrating on an accurate description of the SAPT data for the argon dimer. This led to the g_{12} function described above, which gives an excellent account of our data over the whole range of R . However, for many purposes a much more modest objective is more appropriate. We recall that many calculations on molecular systems, especially for larger molecules, are carried out using very simple

potentials. Lennard-Jones and exp-6 potentials are still widely used, and in both of these forms the dispersion term is simply $-C_6R^{-6}$. The Lennard-Jones form is no longer a serious contender for calculations that attempt to achieve reasonable accuracy, but the exp-6 form has a reasonable theoretical basis and is a sensible starting point. How can we improve on it? We use the Ar data as a convenient test-bed for exploring this issue, bearing in mind however that we are concerned with applications to molecules, not to Ar₂ itself.

Note that $-C_6R^{-6}$ corresponds to a $G(R)$ that is equal to unity everywhere. This undamped expression tends to $-\infty$ as $R \rightarrow 0$, and since the exponential repulsive term remains finite at $R = 0$ there is an unphysical singularity which often leads to difficulties. We can avoid this and also improve on the treatment near the minimum by using the simplest $g_n(R)$, i.e. $g_6(R)$. This cannot describe the behaviour of $G(R)$ accurately at the shortest distances, and attempts to fit it there have been unsuccessful. However, with $R_{\min} = 5.0$ bohr a satisfactory fit can be obtained with a rms error of 1.15%. We have fitted $g_6(R)C_6R^{-6}$ to the SAPT-2NX data, optimizing C_6 and the parameters b_2 , b_4 and b_6 . In figure 6 the circles denote $G(R) = E(\text{SAPT-2NX})/C_6^0R^{-6}$, where $C_6^0 = 65.3$ is the long-range value. The short dashes show the fitted $g_6(R) \times C_6/C_6^0$, where C_6 is the fitted value and $C_6^0 = 65.3$ as before. We see that g_6 gives a good account of the behaviour of $G(R)$ down to $R = 5$ bohr, but damps the dispersion too severely at shorter distances. At $R = 5$ bohr the Ar...Ar interaction energy is more than 40 kJ mol⁻¹ [27], so a good descrip-

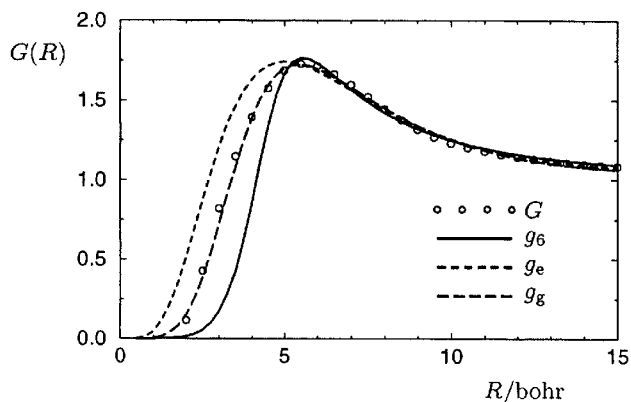


Figure 6. Fits of the SAPT-2NX $G(R)$ points using the g_6 , g_e and g_g approximations. $R_{\min} = 5.0$ bohr.

tion at smaller R is not required for most ambient-temperature phenomena.

The values of C_6 , C_8 and C_{10} that correspond to this fit, according to equation (18), are 64.07, 1525 and 36280. These are quite respectable values, but we emphasize once again that they are irrelevant; the important feature is that this very simple function gives a much better account of the dispersion energy than the simple $-C_6R^{-6}$ expression. We note that it has the same number of parameters as a Tang–Toennies treatment, if the distance scaling parameter is treated as adjustable.

Other functions may be used to approximate $G(R)$. One simple form that has the correct behaviour as $R \rightarrow 0$ and tends to 1 as $R \rightarrow \infty$ is

$$g_e = (1 + c \exp(-bR) - (1 + c) \exp(-tbR))^6. \quad (20)$$

This approaches 1 exponentially as $R \rightarrow \infty$, so it has no $1/R$ expansion at large R and we cannot use it to extract values of C_8 and C_{10} . We emphasize once again that this is not an important limitation. The important property is that it fits $G(R)$ with a rms error of about 1.25%—again not particularly good in comparison with the g_{12} function discussed above, but comparable with the Tang–Toennies fits discussed above, and very much better than no damping at all. It turns out that the fit is not sensitive to the value of the coefficient c , and if this is fixed at a suitable value (we used $c = 3$) there are only two parameters apart from C_6 . The result is shown in figure 6; again C_6 has been fitted, so the function plotted is $g_e \times C_6/C_6^0$. In this case the dispersion at small R is under-damped, but again this is a region of little importance for many phenomena, and one where other terms in the potential are likely to be in error also.

Another similar function uses gaussians instead of exponentials:

$$g_g = (1 + c \exp(-bR^2) - (1 + c) \exp(-tbR^2))^3. \quad (21)$$

Here c is well determined by the data, so this has three effective parameters besides C_6 . It fits the SAPT-2NX data for $R \geq 5.0$ bohr with a rms error of 1.29%. Figure 6 shows that this function gives a good account of $G(R)$ for $R < 5$ bohr, even though it was not fitted to those points.

In both of the latter two functions the parameter b plays the rôle of a distance scaling parameter, so it is easy to arrange for the maximum to occur at an appropriate distance. The height of the maximum can also be adjusted, although this involves more trial and error. Consequently these functions are well suited for use as empirical damping functions for atom–atom potentials in molecular systems.

4. Molecules

Atoms are relatively easy to handle. Molecules pose much more difficult problems. The dispersion energy becomes a function of the relative orientation of the molecules—5 angular coordinates in the general case—as well as the distance. In the long-range limit this is handled by allowing the dispersion coefficients in equation (1) to depend on orientation. Note that the sum will include odd powers of R if either of the interacting molecules is not centrosymmetric. The orientation dependence can be expanded in terms of S -functions [28] or other equivalent functions of orientation, and the coefficients in the expansion have been calculated for a number of systems by Wormer’s group [29–31]. This however provides a single-site treatment, and for all but the smallest molecules the expansion will fail to converge for small R . In any case the description requires a large number of expansion coefficients (HF dimer requires 20 different coefficients to describe the orientation dependence up to C_{10}) and the evaluation of complicated orientational functions. A site–site treatment has proved much more satisfactory for describing the electrostatic, induction and repulsion interactions, and the same is likely to be true for dispersion. The advantage of a site–site description is that the sites describe regions of the molecule (usually atoms) that are much more nearly spherical than the molecule as a whole.

Furthermore, the standard Tang–Toennies damping functions are isotropic, and we can be sure that this is not correct for molecules, especially in a single-site treatment, although they have been used, in conjunction with the Wormer anisotropic C_n coefficients, in the accurate ASP potentials for water [32, 33]. Isotropic damping functions are likely to be more successful if used in a site–site treatment. Fellers *et al.* [34] found that the ASP-S potential of Millot and Stone [32], which incorporates a three-parameter isotropic atom–atom dispersion function due to Szczęśniak *et al.* [35], gives a better

account of the vibration–rotation–tunnelling spectrum of the water dimer than the ASP-W version, which uses the largest 67 of Wormer’s anisotropic dispersion coefficients.

A site–site version of $-G(R)C_6R^{-6}$ must have the following properties. $G(R) \rightarrow 1$ as the site–site distance R tends to infinity, and $G(R) \sim R^6$ as $R \rightarrow 0$, as for the atomic case. The orientation dependence of C_6 for a molecule is known [36]; in addition to the isotropic component, it includes contributions that involve the functions S_{202}^{k0} , $S_{022}^{0k'}$, $S_{220}^{kk'}$, $S_{222}^{kk'}$ and $S_{224}^{kk'}$. The first two of these terms involve the polarizability anisotropy of one molecule and the mean polarizability of the other, and are larger than the last three terms, which involve the polarizability anisotropy of both molecules. These are the only orientational dependences that can survive at large R , where $G(R) \rightarrow 1$, so any other orientation dependence must be described by the $G(R)$ function.

We have explored the use of our new $G(R)$ approach for molecules with the example of HF. We have calculated the SAPT-2*NX* dispersion energy for 1205 configurations of the HF dimer, keeping the HF bond length fixed. The basis set was aug-cc-pVTZ, and we used the mono+ method of Williams *et al.* [37]. The magnitude of the dispersion energy ranged from about 60 to about 2.5×10^{-4} kJ mol⁻¹. We have fitted these points to a 2-site model, with one site at the F nucleus and the other at a position X in the bond, arbitrarily placed 1.1 bohr from the F atom. (The bond length is 1.732 bohr.) As has been observed for Ar··HCl and Ar··HF [38], the polarizability of the halogen site is expected to be larger in the direction perpendicular to the bond, while the polarizability of the bond site is expected to be larger in the direction parallel to the bond. If we include only the larger anisotropic terms in C_6 , this leads to the following model:

$$E_{\text{disp}} = \sum_{ab} G_{ab}(R_{ab}) (C_{ab}^{00} + S_{022}C_{ab}^{02} + S_{202}C_{ab}^{20})R_{ab}^{-6}, \quad (22)$$

where the sum runs over sites a in one molecule and b in the other. The indices k and k' are zero for linear molecules and have been dropped. The coefficients C^{00} , C^{02} and C^{20} are the isotropic C_6 coefficient and its anisotropic components; the ‘6’ has been dropped. Symmetry considerations imply that $C_{ab}^{02} = C_{ba}^{20}$ and $C_{ab}^{00} = C_{ba}^{00}$. Consequently there are 7 C_6 coefficients: C_{FF}^{00} , C_{FF}^{02} , C_{FX}^{00} , C_{FX}^{02} , C_{FX}^{20} , C_{XX}^{00} and C_{XX}^{02} . In the exponential model, there are 2 parameters for each $G_{ab}(R_{ab})$ function. However, we used the same t for all three, so there are 4 parameters. In the Gaussian model, using the same c and t for each G_{ab} , there are five, so the total number of adjustable parameters is 11 or 12. While

Table 5. Parameters for the fit of the SAPT dispersion energy in the HF dimer to equation (22). Values are in atomic units.

Coefficient	Model for $G(R)$		Wormer/TT
	g_c	g_g	
C_{FF}^{00}	10.5643	10.6060	
C_{FF}^{02}	-0.3498	-0.2485	
C_{FX}^{00}	3.7084	3.8336	
C_{FX}^{02}	1.2653	1.2997	
C_{FX}^{20}	-0.0324	-0.0565	
C_{XX}^{00}	1.4214	1.4670	
C_{XX}^{02}	0.4925	0.4543	
b_{FF}	0.4393	0.0196	
b_{FX}	0.3897	0.0201	
b_{XX}	0.4010	0.0173	
c	(3.0)	0.2519	
t	1.2245	8.2056	
b			1.877
rms error	1.765%	1.225%	10.125%
C_6	19.403	19.740	20.478

this is quite a large number, it should be remembered that we are describing a function of four coordinates, and that we are fitting this function to 1205 data points.

The parameters obtained in the fit are given in table 5, together with the rms error. Both models perform reasonably well, with the Gaussian version giving a slightly better result. The table also shows the long-range isotropic C_6 , which is equal to $C_{\text{FF}}^{00} + 2C_{\text{FX}}^{00} + C_{\text{XX}}^{00}$. This may be compared with the value derived by Kumar and Meath from dipole oscillator strength distributions, which was 19.0 ± 0.2 au, and with the value obtained by Rijks and Wormer, which was 20.48 au. Note also that C_{FF}^{02} and C_{FX}^{20} are negative and C_{FX}^{02} and C_{XX}^{02} are positive, as expected from the polarizability anisotropies.

Also shown in the table is the result of using the anisotropic dispersion coefficients of Rijks and Wormer, later corrected [30], with isotropic Tang–Toennies damping. There is only one adjustable parameter, the damping scale factor b , and moreover the basis set used by Rijks and Wormer was different, so it is unreasonable to expect a good fit. On the other hand there are 20 isotropic and anisotropic dispersion coefficients in the Rijks and Wormer description, so the present model is considerably simpler in form as well as more accurate.

5. Discussion and conclusions

We believe that the formulation of the dispersion energy as a product of the long-range form $-C_6R^{-6}$

and a correction function $G(R)$ provides a useful description, avoiding the use of the power series in $1/R$. It has promise both for an accurate description of dispersion for small systems such as the Ar dimer studied here, and for improved descriptions of dispersion in larger molecules, where the simple undamped $-C_6R^{-6}$ form is still widely used.

We consider our model to be particularly useful if the dispersion energy is calculated point by point using IMPT or SAPT. Fitting the C_n and $f_n(R)$ of equation (5) to the total dispersion energy is likely to be a difficult numerical problem, involving a large number of strongly-correlated parameters, and by considering the $G(R)$ function, these problems are generally avoided. Moreover, the use of a power series in R^{-1} at short range is artificial, and the undamped series is only asymptotically convergent, i.e. divergent at any finite R .

The *ab initio* data for $\text{Ar} \cdots \text{Ar}$ can be fitted accurately using a rational function approximation for $G(R)$. This gives a description of the dispersion energy with a relative accuracy of about 0.1% from 30 bohr down to 2.5 bohr. Values for the higher coefficients C_8 , C_{10} , etc., of the conventional multipole series can be derived in principle from the rational-function form for $G(R)$, but this is not a recommended procedure, because the fit is very insensitive to their values. This emphasizes the fact that the coefficients of the long-range R^{-1} expansion are irrelevant to a good description of the dispersion energy in the important region around the energy minimum.

The rational function form is not ideal; low-order forms of it are prone to unphysical singularities near $R = 0$, and higher-order forms, such as g_{12} , which we have used here, require more adjustable parameters than is convenient for most purposes. We suggest several simpler functions that can be used to approximate $G(R)$ in situations where very high accuracy is not a realistic objective, for instance for site-site dispersion functions in molecules. In the case of the HF dimer, a site-site description of this sort, including site anisotropy, has a rms error of less than 1.25% over a set of 1205 points calculated by SAPT.

MPH acknowledges financial support from the EPSRC.

References

- [1] CASIMIR, H. B. G., and POLDER, D., 1948, *Phys. Rev.*, **73**, 360.
- [2] AHLRICHS, R., 1976, *Theor. Chim. Acta*, **41**, 7.
- [3] PAN, Y. H., and MEATH, W. J., 1971, *Molec. Phys.*, **20**, 873.
- [4] HAYES, I. C., and STONE, A. J., 1984, *Molec. Phys.*, **53**, 69.
- [5] HAYES, I. C., and STONE, A. J., 1984, *Molec. Phys.*, **53**, 83.
- [6] RYBAK, S., JEZIORSKI, B., and SZALEWICZ, K., 1991, *J. chem. Phys.*, **95**, 6576.
- [7] SZALEWICZ, K., and JEZIORSKI, B., 1997, *Molecular Interactions*, edited by S. Scheiner (New York: Wiley), p. 3.
- [8] KNOWLES, P. J., and MEATH, W. J., 1986, *Chem. Phys. Lett.*, **124**, 164.
- [9] KOIDE, A., MEATH, W. J., and ALLNATT, A. R., 1981, *Chem. Phys.*, **58**, 105.
- [10] KREEK, H., and MEATH, W. J., 1969, *J. chem. Phys.*, **50**, 2289.
- [11] HEPBURN, J., SCOLES, G., and PENCO, R., 1975, *Chem. Phys. Lett.*, **36**, 451.
- [12] AHLRICHS, R., PENCO, P., and SCOLES, G., 1977, *Chem. Phys.*, **19**, 119.
- [13] NG, K.-C., MEATH, W. J., and ALLNATT, A. R., 1978, *Chem. Phys.*, **32**, 175.
- [14] NG, K.-C., MEATH, W. J., and ALLNATT, A. R., 1979, *Molec. Phys.*, **37**, 237.
- [15] DOUKETIS, C., SCOLES, G., MARCHETTI, S., ZEN, M., and THAKKAR, A. J., 1982, *J. chem. Phys.*, **76**, 3057.
- [16] TANG, K. T., and TOENNIES, J. P., 1977, *J. chem. Phys.*, **66**, 1496.
- [17] TANG, K. T., and TOENNIES, J. P., 1984, *J. chem. Phys.*, **80**, 3726.
- [18] WILLIAMS, H. L., KORONA, T., BUKOWSKI, R., JEZIORSKI, B., and SZALEWICZ, K., 1996, *Chem. Phys. Lett.*, **262**, 431.
- [19] KORONA, T., WILLIAMS, H. L., BUKOWSKI, R., JEZIORSKI, B., and SZALEWICZ, K., 1997, *J. chem. Phys.*, **106**, 5109.
- [20] WOON, D. E., and DUNNING, T. H., 1993, *J. chem. Phys.*, **98**, 1358.
- [21] JEZIORSKI, B., MOSZYNSKI, R., RATKIEWICZ, A., RYBAK, S., SZALEWICZ, K., and WILLIAMS, H. L., 1996, *SAPT: a program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies* (Warsaw: Department of Chemistry, University of Warsaw; Delaware: Department of Physics and Astronomy, University of Delaware).
- [22] STANDARD, J. M., and CERTAIN, P. R., 1985, *J. chem. Phys.*, **83**, 3002.
- [23] KUMAR, A., and MEATH, W. J., 1985, *Molec. Phys.*, **54**, 823.
- [24] THAKKAR, A. J., HETTEMA, H., and WORMER, P. E. S., 1992, *J. chem. Phys.*, **97**, 3252.
- [25] TANG, K. T., NORBECK, J. M., and CERTAIN, P. R., 1976, *J. chem. Phys.*, **64**, 3063.
- [26] AMOS, R. D., ALBERTS, I. L., ANDREWS, J. S., COLWELL, S. M., HANDY, N. C., JAYATILAKA, D., KNOWLES, P. J., KOBAYASHI, R., LAMING, G. J., LEE, A. M., MASLEN, P. E., MURRAY, C. W., PALMIERI, P., RICE, J. E., SIMANDIRAS, E. D., STONE, A. J., SU, M.-D., and TOZER, D. J., 1995, *CADPAC, The Cambridge Analytic Derivatives Package* (Cambridge, UK: University of Cambridge).
- [27] MAITLAND, G. C., RIGBY, M., SMITH, E. B., and WAKEHAM, W. A., 1981, *Intermolecular Forces: Their Origin and Determination* (Oxford: Clarendon).
- [28] STONE, A. J., 1996, *The Theory of Intermolecular Forces, International Series of Monographs in Chemistry* (Oxford: Clarendon).
- [29] RIJKS, W., and WORMER, P. E. S., 1988, *J. chem. Phys.*, **88**, 5704.

- [30] RIJKS, W., and WORMER, P. E. S., 1989, *J. chem. Phys.*, **90**, 6507; note subsequent Erratum, correcting many of the tabulated values, 1990, *J. Chem. Phys.*, **92**, 5754.
- [31] WORMER, P. E. S., and HETTEMA, H., 1992, *J. chem. Phys.*, **97**, 5592.
- [32] MILLOT, C., and STONE, A. J., 1992, *Molec. Phys.*, **77**, 439.
- [33] MILLOT, C., SOETENS, J.-C., MARTINS COSTA, M. T. C., HODGES, M. P., and STONE, A. J., 1998, *J. phys. Chem. A*, **102**, 754.
- [34] FELLERS, R. S., BRALY, L. B., SAYKALLY, R. J., and LEFORESTIER, C., 1999, *J. chem. Phys.*, **110**, 6306.
- [35] SZCZĘŚNIAK, M. M., BRENSTEIN, R. J., CYBULSKI, S. M., and SCHEINER, S., 1990, *J. phys. Chem.*, **94**, 1781.
- [36] STONE, A. J., and TOUGH, R. J. A., 1984, *Chem. Phys. Lett.*, **110**, 123.
- [37] WILLIAMS, H. L., MAS, E. M., SZALEWICZ, K., and JEZIORSKI, B., 1995, *J. chem. Phys.*, **103**, 7374.
- [38] DOUKETIS, C., HUTSON, J. M., ORR, B. J., and SCOLES, G., 1984, *Molec. Phys.*, **52**, 763.