

## Fitting basis sets for the RI-MP2 approximate second-order many-body perturbation theory method

David E. Bernholdt and Robert J. Harrison

Citation: *J. Chem. Phys.* **109**, 1593 (1998); doi: 10.1063/1.476732

View online: <http://dx.doi.org/10.1063/1.476732>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v109/i5>

Published by the AIP Publishing LLC.

---

### Additional information on J. Chem. Phys.

Journal Homepage: <http://jcp.aip.org/>

Journal Information: [http://jcp.aip.org/about/about\\_the\\_journal](http://jcp.aip.org/about/about_the_journal)

Top downloads: [http://jcp.aip.org/features/most\\_downloaded](http://jcp.aip.org/features/most_downloaded)

Information for Authors: <http://jcp.aip.org/authors>

## ADVERTISEMENT



**Goodfellow**  
metals • ceramics • polymers • composites  
70,000 products  
450 different materials  
small quantities **fast**

[www.goodfellowusa.com](http://www.goodfellowusa.com)

# Fitting basis sets for the RI-MP2 approximate second-order many-body perturbation theory method

David E. Bernholdt<sup>a)</sup>

Northeast Parallel Architectures Center, Syracuse University, 111 College Place,  
Syracuse, New York 13244-4100

Robert J. Harrison<sup>b)</sup>

Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory,  
P.O. Box 999, Mail Stop K1-90, Richland, Washington 99352-0999

(Received 19 December 1997; accepted 27 April 1998)

The “resolution of the identity” integral approximation applied to second-order many-body perturbation theory, or RI-MP2, method offers improved computational performance compared to traditional (exact) second-order perturbation theory calculations, but introduces a new auxiliary or “fitting basis set” into the method. We develop fitting basis sets for use with the correlation consistent cc-pVDZ and cc-pVTZ atomic orbital basis sets for the atoms H-Ne. These fitting sets are designed to reproduce exact second-order results for a set of 32 test cases, including a variety of reaction energies, weak interactions, and electrostatic properties, to better than 1% error averaged across all tests and less than 2% error in any individual case. Although the RI-MP2 method is primarily targeted to large-scale calculations, it offers substantial performance improvements even for the small molecules used in these test cases. © 1998 American Institute of Physics.  
[S0021-9606(98)02829-3]

## I. INTRODUCTION

The “resolution of the identity” (RI) integral approximation<sup>1–3</sup> has attracted some attention recently as one approach to reduce the cost of electronic structure calculations. The RI approximation has been used in conjunction with the self-consistent field (SCF),<sup>2,4</sup> multiconfiguration self-consistent field (MCSCF),<sup>5–7</sup> density functional theory (DFT),<sup>8–10</sup> MP2,<sup>1,11–13</sup> and coupled cluster methods.<sup>14</sup> RI-MP2 calculations on the binding of K<sup>+</sup> to 12-crown-4 ether exhibit as much as 94% reduction in CPU time relative to exact MP2 without loss of accuracy in the resulting binding energy.<sup>11</sup>

The RI approximation can be expressed as

$$\int \chi_p(r_1) \chi_q(r_1) \frac{1}{|r_1 - r_2|} \chi_r(r_2) \chi_s(r_2) d^3r_1 d^3r_2 \\ = (pq|rs) \approx \sum_{\Delta, \Theta} (pq|\Delta) V_{\Delta\Theta}^{-1} (\Theta|rs), \quad (1)$$

where the indices  $p$ ,  $q$ ,  $r$ , and  $s$  refer to the usual atomic or molecular orbital (AO or MO) basis functions, while  $\Delta$  and  $\Theta$  refer to an auxiliary or “fitting” basis. The RI approximation amounts to representing the space of pairwise products of AO basis functions in terms of a linear combination of fitting basis functions. It is closely related to the Dunlap fit widely used in density functional theory<sup>15,16</sup> and to the multiplicative integral approximation.<sup>17</sup> The history and relationships among these approaches, as well as other integral ap-

proximations, are thoroughly discussed in a recent review by Challacombe, Schwengler, and Almlöf.<sup>18</sup> The pseudospectral approach, pioneered by Friesner and co-workers,<sup>19</sup> combines familiar basis sets with numerical grids. Fitting of AO products is also used in this approach, but only in order to transfer results between the grid and the basis sets.

Different forms of the RI approximation arise depending on how one chooses to measure the residual of the approximation. The particular form shown in Eq. (1), referred to as the “V approximation” by Vahtras and co-workers,<sup>2</sup> involves minimizing the Coulomb repulsion of the residual with itself, and can be thought of as a least squares fit of the electrostatic potential of the product space density as a linear combination of the electrostatic potential arising from the fitting density. This leads to all of the three- and two-center (V) integrals being electron repulsion integrals, involving the  $1/|r_1 - r_2|$  operator. The V approximation is the closest analog to the Dunlap fit, and based on some simple tests by Vahtras and coworkers,<sup>2</sup> has become the form of the RI approximation that is most widely used.

The choice of a measure of the residual, such as the V approximation, fixes the form of the RI approximation, but does not impose any requirements on the fitting basis itself. In principle, if the fitting basis were chosen to be the  $N^2$  product space of  $N$  AO basis functions, the “approximation” could be made exact. The advantage of the RI approximation comes from the fact that in practice, the AO product space is nearly degenerate and can be represented with a smaller basis<sup>20,21</sup>—to the extent that the size of the fitting basis can be kept smaller than  $N^2$ , there are fewer integrals to generate, manipulate, and store.

Clearly one of the requirements for RI methods to

<sup>a)</sup>Electronic mail: bernholdt@npac.syr.edu

<sup>b)</sup>Electronic mail: rj\_harrison@pnl.gov

achieve general acceptance is that appropriate fitting basis sets be available to produce reliable results, just as researchers now must choose AO basis sets appropriate to each problem. As RI methods are fairly new, relatively little work has so far been done toward understanding or defining appropriate fitting basis sets.<sup>8–10,13</sup> Indeed, many existing codes implementing the RI approximation are limited to using the *same* basis for both the AO and fitting sets.<sup>1,2,5–7,14</sup>

In this paper, we begin to address this important issue by developing fitting basis sets appropriate to the RI-MP2 method. As the simplest method to include dynamic electron correlation, MP2 occupies an important position in the hierarchy of *ab initio* electronic structure methods, and improving the computational performance of the method while retaining the familiar accuracy and behavior would be very beneficial. Early experience with the RI-MP2 method has already shown it to have significant promise in terms of accuracy and performance, as well as applicability to parallel computers, which is beneficial in pushing MP2 calculations to even larger chemical systems.<sup>11,12</sup> The development of general-purpose fitting basis sets will facilitate further development of this method as a viable alternative to traditional exact MP2 calculations.

## II. APPROACH

As with the development of AO basis sets, there is a wide range of possible approaches and criteria which can be employed in the design of fitting basis sets. We have elected to pursue a straightforward and systematic course intended to quickly produce some useful fitting basis sets and at the same time begin to understand the general behavior of the fitting basis.

Obviously, once fitting basis sets are defined, their generality can easily be tested. For this work, we have chosen the widely used cc-pVDZ and cc-pVTZ correlation consistent sets<sup>22–24</sup> to represent the AO space. The systematic nature of the correlation consistent basis sets, which range from double- to sextuple-zeta (for first-row atoms) and also offer well-defined diffuse and core-valence augmentations, is an important advantage in the early stages of understanding RI fitting basis sets because it facilitates the generalization of “rules” to apply to multiple basis sets.

Initial experience with the RI-MP2 method indicates that while it is possible to develop large fitting basis sets which reproduce the exact MP2 energy quite accurately, it is possible to accurately reproduce exact MP2 energy *differences* with much smaller fitting basis sets.<sup>25</sup> As energy differences rather than the total energies themselves are obviously the primary interest of nearly all electronic structure calculations, and because smaller fitting basis sets increase the performance advantage of the RI-MP2 approach, we have chosen the reproduction of exact MP2 energy differences as the criterion for assessing the quality of fitting basis sets.

Just as with AO basis sets, it is, of course, possible to produce fitting sets which are tailored to a particular type of reaction or property. However, our aim was to develop fitting basis sets of general utility. As shown in Table I, we have selected a set of 32 different tests involving molecules comprised of zeroth- and first-row atoms. These test cases were

chosen to represent a range of valence electron properties, including reaction energies involving single and multiple bonds, hydrogen bonding, van der Waals interactions, and electrostatic properties. Some 55 separate calculations contribute to these tests. All calculations are carried out at fixed geometries, identical for RI-MP2 and exact MP2; no attempt is made in this work to explore the behavior of these fitting sets for calculations of potential energy surfaces. Weak interactions are computed both with and without the Boys-Bernardi counterpoise correction for basis set superposition error. Dipole moment and polarizability calculations are computed as double-sided finite differences, with fields generated by distant point charges. Full details of calculations may be obtained from the authors upon request.

Our primary criterion in evaluating fitting basis sets is the relative error compared to the exact MP2, with a target of less than 1% error averaged across all 32 test “reactions,” with a maximum of no more than 2% error for any single test case. The error in the second-order correlation contribution to the reactions, as well as the total energies of the 55 separate “molecules” comprising the test are also reported, but were not considered as optimization criteria in this work. Calculations were carried out both correlating and freezing core electrons of the first-row atoms, and both cases are considered together in this analysis (i.e., effectively 64 reactions and 110 separate molecules). No substantive differences were observed between the all-electron and frozen core results at any stage in the fitting basis development process.

All calculations were carried out with the NWCHEM computational chemistry package,<sup>27</sup> and used the Cartesian (6d, 10f, ...) form of the basis sets due to limitations of the current release of NWCHEM.

## III. RESULTS AND DISCUSSION

### A. cc-pVDZ

The development of a fitting basis to match the cc-pVDZ AO set began with calculations employing the cc-pVDZ set itself, as well as an uncontracted version as the fitting basis. This approach arose from our experience with the development of fitting basis sets to reproduce the binding energy of  $K^+$  to 12-crown-4 ether,<sup>11,25</sup> which required little more than uncontracting the AO basis to accurately reproduce the exact result. As can be seen in Table II, the results for the test cases in Table I are not nearly as promising. This is not surprising given that the range of interactions the basis is required to reproduce is greatly expanded. In addition, all of the test cases involve very small molecules compared to the crown ether, while we would naturally expect the RI approximation to perform better for larger molecules because it relies on near linear dependencies in the AO product space which increase with basis set size.

Augmenting the uncontracted cc-pVDZ with *pd/sp3df* (H-He/Li-Ne), the basis that was successful with the  $K^+$ :12-crown-4, reduced the average and maximum errors to roughly 3% and 100%, respectively. More detailed examination of the results, especially the nine test cases having only first-row atoms, indicated that the Li-Ne basis was performing well, while the H-He basis needed further improvement.

TABLE I. Reactions and properties used in evaluating the quality of fitting basis sets. Refer to the text for more information about these calculations; full details are available from the authors.

Category	Description	References	Designation
Reaction energy	$C_2H_2 + 3H_2 \rightarrow 2CH_4$	...	bh01
	$C_2H_4 + 2H_2 \rightarrow 2CH_4$	...	bh02
	$C_2H_6 + H_2 \rightarrow 2CH_4$	...	bh03
	$LiH + HF \rightarrow LiF + H_2$	...	bh08
	$HOOH \rightarrow cis\text{-}HOOH$	14	rl01
	$HOOH \rightarrow trans\text{-}HOOH$	14	rl02
	$cis\text{-}FNNF \rightarrow trans\text{-}FNNF$	14	rl03
	$FNO_2 \rightarrow cos\text{-}FONO$	14	rl04
	$HCN \rightarrow HNC$	14	rl05
	$HCN \rightarrow CHN$	14	rl06
	$Be_3 + 3Be$	14	rl07
	$Be_4 + 4Be$	14	rl08
	$2HOF \rightarrow F_2O + H_2O$	14	rl09
	$2BH_3 \rightarrow B_2H_6$	26	rpb1
	$BH_3 + CO \rightarrow H_3BCO$	26	rpb2
Hydrogen bonding	$(HF)_2 \rightarrow 2HF$	...	bh04
	$(HF)_2 \rightarrow 2HF$ (counterpoise corrected)	...	bh06
	$(H_2O)_2 \rightarrow 2H_2O$	...	bh05
	$(H_2O)_2 \rightarrow 2H_2O$ (counterpoise corrected)	...	bh07
van der Waals dimers	$He_2 \rightarrow 2He$	...	bh09
	$Ne_2 \rightarrow 2Ne$	...	bh10
	$HeNe \rightarrow He + Ne$	...	bh11
	$He_2 + 2He$ (counterpoise corrected)	...	bh19
	$Ne_2 + 2Ne$ (counterpoise corrected)	...	bh20
	$HeNe \rightarrow He + Ne$ (counterpoise corrected)	...	bh21
Dipole moments	HF	...	bh12
	CO	...	bh13
	$trans\text{-}FNNF$ (both components)	...	bh14, bh15
Polarizabilities	He	...	bh16
	Ne	...	bh17
	CO	...	bh18

By expanding the H-He set to  $s3pd$  (in the notation of Table III, adding  $\frac{1}{2}s_1 \rightarrow s$ ,  $2s_1 \rightarrow p$ , and  $s_2 + s_3 \rightarrow p$  to the  $pd's$   $2p_1 \rightarrow p, d$ ) in analogy to the first-row fitting set, errors were reduced to 0.8% and 21%. It is worth noting that a  $\frac{1}{2}s_1$  function would not appear in the product space of an atom

with itself, according to the Gaussian product theorem (GPT), but experiments indicate its importance to obtaining an appropriate fitting basis set. Clearly this basis function in particular must be serving to help fit basis function products involving two centers.

TABLE II. Numerical performance of fitting basis sets for the cc-pVDZ AO set. "Reactions" refer to the error (compared to exact MP2) results for the 32 test cases shown in Table I. "Total energies" refer to the error (compared to exact MP2) of the 55 separate calculations which contribute to the test cases of Table I. Best and worst case identify the particular test case or calculation for which the extreme was obtained.

		Fitting basis		
		cc-pVDZ	Uncontracted cc-pVDZ	cc-pVDZ-fit2-1
Description	H-He (size)	$2s1p$ (5)	$4s1p$ (7)	$9s6p4d$ (51)
	Li-Ne (size)	$3s2p1d$ (15)	$9s4p1d$ (27)	$10s5p4d1f$ (59)
Reactions	Average error (%)	7575	20804	0.13
	Maximum error (%)	239 806	323 376	1.4
	Worst case	bh09	bh11	bh11
Correlation contribution to reactions	Average error (%)	140	37.8	0.36
	Maximum error (%)	2791	335	1.8
	Worst case	bh09	bh09	bh21
Total energies	Minimum error ( $mE_h$ )	0.4	0.6	0.0003
	Best case	$C_2H_4$	Be	He+ghost He <sup>a</sup>
	Average error ( $mE_h$ )	17.3	17.9	0.2
	Maximum error ( $mE_h$ )	50.2	49.5	0.6
	Worst case	$FNO_2$	$FNO_2$	$FNO_2$

<sup>a</sup>A calculation in the He dimer counterpoise-correction, test case bh19.

TABLE III. Definitions of the cc-pVXZ-fit2-1 fitting basis sets. Each basis consists of the uncontracted primitive exponents from the corresponding AO set augmented with additional functions described below. Subscripted angular momentum labels refer to the AO primitive exponents ordered from most diffuse to tightest.

Atoms	Augmenting function angular momentum	cc-pVDZ-fit2-1	cc-pVTZ-fit2-1
H-He	<i>s</i>	$\frac{1}{4}s_1, \frac{1}{2}s_1, \frac{1}{2}s_2, \frac{1}{2}s_3, \frac{1}{2}s_4$	$\frac{1}{4}s_1, \frac{1}{2}s_1$
	<i>p</i>	$2s_1, s_2 + s_3, \frac{1}{4}p_1, \frac{1}{2}p_1, 2p_1,$	$s_2 + s_3, \frac{1}{4}p_1, \frac{1}{2}p_1, \frac{1}{2}p_2$
	<i>d</i>	$\frac{1}{4}p_1, \frac{1}{2}p_1, p_1, 2p_1,$	$\frac{1}{4}p_1, \frac{1}{2}p_1, 2p_1, 2d_1$
	<i>f</i>	$\dots$	$\frac{1}{4}d_1, \frac{1}{2}d_1, d_1, 2d_1$
Li-Ne	<i>s</i>	$\frac{1}{2}s_1$	$\frac{1}{2}s_1$
	<i>p</i>	$\frac{1}{2}p_1$	$\frac{1}{2}p_1$
	<i>d</i>	$2p_1, p_2 + p_3, 2d_1$	$p_2 + p_3, \frac{1}{2}d_1$
	<i>f</i>	$2d_1$	$2d_1, 2f_1$
	<i>g</i>	$\dots$	$2f_1$

Further refinement of the H-He basis sets, focusing in particular on the He dimer test case, indicating that other “non-GPT” functions, such as  $\frac{1}{2}s_2 \rightarrow s$  and  $\frac{1}{2}p_1 \rightarrow p$ , contribute significantly to improving results. (Note that for the H and He cc-pVDZ basis sets,  $\frac{1}{2}s_2$  is nearly the same as  $2s_1$ , which would be present in the Gaussian product space of an atom’s basis with itself, but this is not the case for  $\frac{1}{2}p_1$ .) The final result is a H-He augmentation set of  $5s5p4d$  with average and maximum errors of 0.13% and 1.4%, satisfying our criteria for the numerical performance of the fitting basis. Table II gives details of the performance, and Table III provides the complete definition for both the H-He and Li-Ne fitting sets. We label this basis cc-pVDZ-fit2-1, indicating that it is an RI-MP2 fitting basis (“fit2”) designed for the cc-pVDZ AO basis. The final “-1” is added to distinguish it from other RI-MP2 fitting basis sets for the cc-pVDZ AO set which may be defined in the future based on different criteria or refined procedures.

This fitting set is approximately ten times the size of the cc-pVDZ AO set for H-He, but only about four times larger for Li-Ne, significantly smaller for nearly all molecules, than the  $N^2$  AO product space they are approximating. Since the RI-MP2 method is intended for large molecules or larger AO basis sets, the small ones which make up our test cases can hardly be considered a “fair” test of the computational performance of the RI-MP2 method relative to the exact approach; however, even here there is a factor of roughly 2.5 in favor of the RI-MP2 method. On a 200 MHz UltraSPARC processor, the 55 exact MP2 calculations which contribute to Table I (all electrons correlated) required a total of 1142 s, while the analogous RI-MP2 calculations with the cc-pVDZ-fit2-1 fitting set required just 450 s. The best single calculation is  $\text{Be}_4$  (60 AO functions, 236 fitting functions), for which the RI-MP2 is 4.7 times faster than the exact calculation.

The fact that the H-He fitting basis is, in relative terms, several times larger than that for Li-Ne, gives some indication of how hard it is to obtain an adequate fit for hydrogen and helium. Our experience indicates that the  $\text{He}_2$  and  $\text{HeNe}$  cases (bh09 and bh11) are the hardest to reproduce, though use of the counterpoise correction (as in bh19 and bh21) makes it much easier to obtain a good fit. Pending a more detailed understanding of the behavior of the RI-MP2 ap-

proximation for weak interactions, it seems that some care should be exercised in applying this method to van der Waals systems.

## B. cc-pVTZ

After the numerous calculations involved in the development of the cc-pVDZ-fit2-1 basis, creating a fitting set for the cc-pVTZ AO basis was relatively straightforward. As mentioned above, the RI approximation might be viewed as a way of “taking advantage” of a large AO basis, and is expected to perform better as the size of the AO set increases. This is clearly evident in comparing the cc-pVTZ results to the cc-pVDZ. Table IV shows that even with the cc-pVTZ basis itself, or its uncontracted primitives as the fitting basis, the average and maximum errors are orders of magnitude better than the corresponding DZ results (see Table II). With the uncontracted cc-pVTZ as the fitting basis, the average and maximum errors are already down to 3.5% and 58.8%.

Application of the DZ augmentation formula ( $5s5p4d/sp3df$ ) to the cc-pVTZ primitives lowers the average error to 0.9%, but does not appreciably change the maximum error. However, accounting for the higher angular momentum in the TZ AO set by “shifting” some of the augmentation to a higher level quickly produced an acceptable fitting set with  $2s4p4d4f/sp2d2fg$  augmentation of the TZ primitives. This basis, which we label cc-pVTZ-fit2-1, detailed in Table III, has errors of just 0.06% and 0.4%.

The final TZ fitting set is approximately six times (H-He) or three times (Li-Ne) the size of the AO basis. These ratios are significantly better than for the DZ basis, in accord with the idea that the RI approximation should work better for larger basis sets. Computational performance is also better for the TZ case—the exact MP2 calculations for the test cases in Table I required 20 058, but only 3408 s for the RI-MP2, a factor of nearly 6 compared to 2.5 in the DZ case. The best single calculation is once again  $\text{Be}_4$ , with the RI-MP2 calculation 12.4 times faster than the exact with the TZ basis sets (140 AO functions, 396 fitting functions).

TABLE IV. Numerical performance of fitting basis sets for the cc-pVTZ AO set. “Reactions” refer to the error (compared to exact MP2) results for the 32 test cases shown in Table I. “Total energies” refer to the error (compared to exact MP2) of the 55 separate calculations which contribute to the test cases of Table I. Best and worst case identify the particular test case or calculation for which the extreme was obtained.

Description		Fitting basis		
		cc-pVTZ	Uncontracted cc-pVTZ	cc-pVTZ-fit2-1
	H-He (size)	3s2p1d (15)	5s2p1d (17)	7s6p5d4f (95)
	Li-Ne (size)	4s3p2d1f (35)	10s5p2d1f (47)	11s6p4d3f1g (98)
Reactions	Average error (%)	5.2	3.5	0.06
	Maximum error (%)	94.3	58.8	0.4
	Worst case	bh09	bh09	bh11
Correlation contribution to reaction	Average error (%)	17.1	11.6	0.24
	Maximum error (%)	197	136	2.2
	Worst case	bh09	rl01	bh01
Total energies	Minimum error ( $\mu E_h$ )	<0.1	<0.1	<0.1
	Best case	Be <sub>3</sub>	Be <sub>3</sub>	Be <sub>3</sub>
	Average error ( $mE_h$ )	3.3	4.9	0.3
	Maximum error ( $mE_h$ )	29.6	20.4	1.5
	Worst case	Be <sub>4</sub>	<i>trans</i> -FNNF in field <sup>a</sup>	<i>cis</i> -FNNF

<sup>a</sup>A finite field calculation of test case bh15.

### C. Applicability to other AO basis sets and comparison with other fitting basis sets

Having developed these fitting basis sets, there are two obvious questions to ask about them. First, can they be used with AO sets other than those for which they were originally designed? Second, how do they compare with the variety of fitting basis sets which are already available in the literature (primarily for use in DFT calculations)? In order to study these questions, we evaluated the test calculations (Table I) with various combinations of AO and fitting basis sets.

Table V reports the results obtained when different AO basis sets are used with the cc-pVXZ-fit2-1 fitting sets defined here. The AO basis sets were chosen as widely used sets of roughly double- and triple-zeta quality which are readily available for the atoms H-Ne. While the average errors with the DZ fitting set are uniformly less than 1%, the maximum errors can be as large as 15%. In half of the cases,

the maximum error occurs for the HOOH isomerization reaction (rl02), while for the remainder, it is one of the rare gas interactions (bh09–bh11). There is no apparent pattern in which test case gives rise to the largest error, except that for the three best AO sets (6-311G\*\*, pVDZ, and cc-pVDZ) it is always a rare gas dimerization.

As might be expected, the TZ fitting set performs better than the DZ for all of the AO sets tried, with one exception. The average errors are generally less than 0.1% and the maximum errors are also generally good. The test cases for which the maximum errors occur are the same as for the DZ fitting set, but in this case only three (4-31G, 6-31G, and VDZ) are the HOOH isomerization (rl02).

According to the criteria originally used in their design, there are only three AO sets for which the cc-pVDZ-fit2-1 could be considered an adequate fitting basis (6-311G\*\*, pVDZ, and cc-pVDZ), while the cc-pVTZ-fit2-1 fitting set

TABLE V. Performance of the cc-pVXZ-fit2-1 fitting basis sets with some widely-used AO sets of roughly double- and triple-zeta quality. Errors are reported for the 32 test cases of Table I with both all-electron and frozen core calculations. All basis sets were obtained from the EMSL Basis Set Database world-wide web service (Ref. 28).

AO basis	Refs.	Description (H-He/Li-Ne or H-He/Li-Be/B-Ne)	cc-pVDZ-fit2-1 Error (%)		cc-pVTZ-fit2-1 Error (%)	
			Avg	Max	Avg	Max
4-31G	29	2s/3s2p	0.27	5.6	0.05	0.8
6-31G	30,31	2s/3s2p	0.53	15.4	0.08	2.2
6-31G**	30–32	2s1p/3s2p1d	0.61	13.7	0.07	1.4
6-311G	33	3s/4s3p	0.27	7.7	0.10	1.6
6-311G**	33	3s1p/4s3p1d	0.11	0.9	0.10	1.5
Ahlrichs VDZ	34	2s/3s/3s2p	0.21	5.4	0.05	0.5
Ahlrichs pVDZ <sup>a</sup>	34	2s/3s1p/3s2p1d	0.09	0.5	0.03	0.4
Ahlrichs VTZ	34	3s/6s/6s3p	0.16	3.0	0.07	1.2
cc-pVDZ	22–24	2s1p/3s2p1d	0.14	1.4	0.17	4.7
cc-pVTZ	22–24	3s2p1d/4s3p2d1f	0.91	10.1	0.06	0.4

<sup>a</sup>Referred to as “SVP” in some publications (Refs. 8 and 13).

TABLE VI. Performance of a variety of alternative fitting basis sets with the cc-pVXZ AO sets. Errors are reported for the 32 test cases of Table I with both all-electron and frozen core calculations.

Fitting basis	Ref.	Description (H-He/Li-Ne or H-He/Li-Be/B-Ne)	cc-pVDZ Error (%)		cc-pVTZ Error (%)	
			Avg	Max	Avg	Max
cc-pVDZ-fit2-1	...	9s6p4d/10s5p4d1f	0.14	1.4	0.91	10.1
cc-pVTZ-fit2-1	...	7s6p5d4f/11s6p4d3f1g	0.17	4.7	0.06	0.4
RI-DFT Coul.	8	2s2p/6s2p2d1f/6s3p3d1f	24.53	760.9	13.83	340.7
DeMon Coul.	28,35	4s1p/7s3p3d	84.88	2626.7	13.56	163.1
DGauss A1 Coulomb	28,35	4s/7s2p1d/7s3p3d	58.93	1637.2	19.37	163.1
DGauss A1 Exchange	28,35	4s/7s2p1d/7s3p3d	33.22	695.1	27.00	267.0
DGauss A2 Coulomb	28,35	3s1p1d/7s2p1d/8s4p4d	13.57	340.3	5.11	57.9
DGauss A2 Exchange	28,35	3s1p1d/7s2p1d/8s4p4d	567.69	16 865.6	70.36	1164.0

would be considered adequate for all but two (6-31G, cc-pVDZ) of the AO sets tested. It is worth noting, however, that since the TZ fitting set is nearly twice as large as the DZ set, the improved performance of the TZ fitting basis does come at a cost. Moreover, one might naively expect that for a given fitting set, a smaller AO basis would have smaller errors than a larger AO set. This is not the case, however, even with closely related sets such as 6-311G and 6-311G\*\*, which differ only in the addition of polarization functions on the latter. For the DZ fitting basis, the larger 6-311G\*\* AO set provides acceptable performance, while the smaller 6-311G set does not. Similarly for the TZ fitting set, the larger 6-31G\*\* AO set satisfies our criteria while the 6-31G does not. Most notable perhaps is the fact that for the cc-pVDZ AO set, the TZ fitting set performs noticeably worse than the DZ.

As observed earlier, while little work has been done on RI-MP2 fitting basis sets, DFT calculations have fit both Coulomb and exchange terms in a fashion similar to the RI approximation for some time. Table VI compares calculations using a number of common DFT fitting sets with the cc-pVXZ-fit2-1 sets defined here. Since the TZVPPP fitting sets used by Weigand and Häser in recent RI-MP2 calculations<sup>13</sup> do not include definitions for helium or neon atoms, we also present Table VII, which considers only those test cases which do not include He or Ne, in order to present the TZVPPP fitting basis results in an appropriate context.

It is clear from Table VI that none of the other fitting sets comes close to meeting the criteria used in this paper, with average errors in excess of 10% and worse case errors several *times* the size of the result itself in most cases. Table VII shows that the worst of the errors are eliminated by excluding all helium and neon test cases from consideration (bh09-bh11, bh16, bh17, and bh19-bh21), in accord with our earlier observations that the weak rare gas interactions are among the hardest to fit adequately. The average errors for both of the basis sets that were used specifically with RI methods (the RI-DFT Coulomb set and the TZVPPP for RI-MP2) are in fact fairly good, but the maximum errors remain large by the criteria used in this paper.

In fact, calculations have been performed for all AO basis sets in Table V with each of the fitting sets in Table VII, as well as using the AO basis itself and the uncontracted version of the AO basis as the fitting basis. These results show only specific cases of basis set pairs which would be considered to give adequate numerical performance according to the criteria we have outlined, completely consistent with the results we have detailed above, which focus on the cc-pVXZ AO sets and the cc-pVXZ-fit2-1 fitting sets. It is worth noting that when the TZVPPP fitting set is used with the pVDZ AO set, as in the original paper,<sup>13</sup> the average and maximum errors are 0.4% and 1.4% (excluding He and Ne test cases, of course). Also interesting is the fact that the DFT fitting sets do a rather poor job on our RI-MP2 test

TABLE VII. Performance of a variety of alternative fitting basis sets with the cc-pVXZ AO sets. Errors are reported for the 24 test cases of Table I which include neither He nor Ne atoms. Both all-electron and frozen core calculations are considered in tabulating the errors.

Fitting basis	Ref.	Description (H-He/Li-Ne or H-He/Li-Be/B-Ne)	cc-pVDZ Error (%)		cc-pVTZ Error (%)	
			Avg	Max	Avg	Max
cc-pVDZ-fit2-1	...	9s6p4d/10s5p4d1f	0.04	0.2	0.49	1.4
cc-pVTZ-fit2-1	...	7s6p5d4f/11s6p4d3f1g	0.02	0.1	0.05	0.2
RI-DFT Coul.	8	2s2p/6s2p2d1f/6s3p3d1f	0.29	1.9	1.09	6.0
DeMon Coul.	28,35	4s1p/7s3p3d	1.70	5.0	3.26	8.6
DGauss A1 Coulomb	28,35	4s/7s2p1d/7s3p3d	5.83	30.3	6.90	32.4
DGauss A1 Exchange	28,35	4s/7s2p1d/7s3p3d	7.65	32.4	7.72	28.0
DGauss A2 Coulomb	28,35	3s1p1d/7s2p1d/8s4p4d	1.64	5.1	2.30	7.7
DGauss A2 Exchange	28,35	3s1p1d/7s2p1d/8s4p4d	4.26	32.4	4.23	19.7
TZVPPP	13	3s2p1d/6s3p1d/6s4p2d1f	0.34	2.1	1.03	4.0

cases, even with very small AO sets such as 4-31G. When used in DFT calculations, such fitting basis sets are considered more or less “universal” and are used with little regard for the AO basis being used. Historically, these fitting basis sets have performed very well for DFT calculations, probably due at least in part to the fact that a variational principle applies to the problem, which is not true for RI-MP2.

#### IV. CONCLUSIONS

We have developed fitting basis sets (cc-pVXZ-fit2-1, X=D,T) for use in conjunction with the cc-pVDZ and cc-pVTZ AO basis sets in RI-MP2 calculations on systems containing the atoms H-Ne. These fitting basis sets reproduce exact MP2 results for a collection of 32 different test cases, including a variety of reaction energies, weak interactions, and electrostatic properties, to better than 1% error on average, and better than 2% in the worst case. Total energies are reproduced to better than a milliHartree. Although the RI-MP2 method favors large-scale calculations, calculations using these AO and fitting basis sets are 2.5 (for DZ) to 5.9 (for TZ) times faster than the exact calculations for the collection of small molecules used in these tests, and in individual cases up to 4.7 (for DZ) or 12.4 (for TZ) times faster.

The fitting basis sets are based on the primitive AO basis augmented with additional functions at each angular momentum, and may be generated using the recipe in Table III or obtained by contacting the first author. Most of the augmentation comes from functions that would be in the GPT product space of the atom's primitive basis with itself, but some “non-GPT” diffuse functions make important contributions, and must be involved in fitting multicenter products. Our experience indicates that weak interactions, such as rare gas dimers, are the hardest to reproduce accurately in our test set. This is particularly true of those involving helium, and results in H and He fitting sets being substantially larger (relative to the corresponding AO set) than for the first-row atoms. However, as the size of the AO set increases, the relative size of the fitting basis set required decreases, as demonstrated here by the change between double- and triple-zeta fitting sets relative to their AO sets.

We have examined the performance of these fitting sets in conjunction with other AO basis sets of comparable size, and conclude that while fitting basis sets designed according to the criteria used in this paper can obviously be used with other AO sets, there is no way to determine *a priori* if a given AO set and fitting set will work together as desired. We have likewise considered a number of other fitting basis sets, such as those commonly used in DFT calculations and found the numerical performance to be far worse than those developed specifically for RI-MP2 calculations. Consequently, we believe that the most computationally efficient and numerically accurate fitting basis sets will be obtained by tailoring them to work with a particular AO set for the method in question.

While these fitting basis sets are neither unique nor definitive, they should provide a good starting point for reliable calculations using the RI-MP2 method.

#### ACKNOWLEDGMENTS

The authors wish to thank Thom Dunning, Kirk Peterson, Angela Wilson, Rick Kendall, and Herbert Früchtl for numerous helpful discussions. NWChem Version 3.0, as developed and distributed by Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352, and funded by the U.S. Department of Energy, was used to obtain some of these results. This work was supported through the U.S. Department of Energy by the Mathematical, Information, and Computational Science Division phase II grand challenges of the Office of Computational and Technology Research. This work was performed under Contract No. DE-AC06-76RLO 1830 with Battelle Memorial Institute (Pacific Northwest National Laboratory). D.E.B. gratefully acknowledges additional support for this work from the Alex G. Nason Fellowship at Syracuse University.

- <sup>1</sup>M. Feyereisen, G. Fitzgerald, and A. Komornicki, *Chem. Phys. Lett.* **208**, 359 (1993).
- <sup>2</sup>O. Vahtras, J. Almlöf, and M. W. Feyereisen, *Chem. Phys. Lett.* **213**, 514 (1993).
- <sup>3</sup>R. A. Kendall and H. A. Früchtl, *Theor. Chim. Acta* **97**, 158 (1997).
- <sup>4</sup>H. A. Früchtl, R. A. Kendall, R. J. Harrison, and K. G. Dyall, *Int. J. Quantum Chem.* **64**, 63 (1997).
- <sup>5</sup>S. Ten-no and S. Iwata, *Chem. Phys. Lett.* **240**, 578 (1995).
- <sup>6</sup>S. Ten-no and S. Iwata, *Int. J. Quantum Chem.* **60**, 1319 (1996).
- <sup>7</sup>S. Ten-no and S. Iwata, *J. Chem. Phys.* **105**, 3604 (1996).
- <sup>8</sup>K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, *Chem. Phys. Lett.* **240**, 283 (1995).
- <sup>9</sup>K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, *Chem. Phys. Lett.* **242**, 652 (1995).
- <sup>10</sup>K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, *Theor. Chim. Acta* **97**, 119 (1997).
- <sup>11</sup>D. E. Bernholdt and R. J. Harrison, *Chem. Phys. Lett.* **250**, 477 (1996).
- <sup>12</sup>D. Feller, E. Aprà, J. A. Nichols, and D. E. Bernholdt, *J. Chem. Phys.* **105**, 1940 (1996).
- <sup>13</sup>F. Weigend and M. Häser, *Theor. Chim. Acta* **97**, 331 (1997).
- <sup>14</sup>A. P. Rendell and T. J. Lee, *J. Chem. Phys.* **101**, 400 (1994).
- <sup>15</sup>B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, *J. Chem. Phys.* **71**, 3396 (1979).
- <sup>16</sup>J. W. Mintmire and B. I. Dunlap, *Phys. Rev. A* **25**, 88 (1982).
- <sup>17</sup>C. van Alsenoy, *J. Comput. Chem.* **9**, 620 (1988).
- <sup>18</sup>M. Challacombe, E. Schwengler, and J. Almlöf, in *Computational Chemistry: Reviews of Current Trends*, edited by J. Leszczynski (World Scientific, Singapore, 1996), Vol. 1, pp. 53–107.
- <sup>19</sup>R. A. Friesner, *Annu. Rev. Phys. Chem.* **42**, 341 (1991).
- <sup>20</sup>N. H. F. Beebe and J. Linderberg, *Int. J. Quantum Chem.* **12**, 683 (1977).
- <sup>21</sup>D. W. O'Neal and J. Simons, *Int. J. Quantum Chem.* **34**, 673 (1989).
- <sup>22</sup>T. H. Dunning, Jr., *Chem. Phys.* **90**, 1007 (1989).
- <sup>23</sup>D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.* **100**, 2975 (1994).
- <sup>24</sup>D. E. Woon and T. H. Dunning, Jr. (in preparation).
- <sup>25</sup>D. E. Bernholdt (unpublished results).
- <sup>26</sup>L. T. Redmon, G. D. Purvis, III, and R. J. Bartlett, *J. Am. Chem. Soc.* **101**, 2856 (1979).
- <sup>27</sup>High Performance Computational Chemistry Group, Pacific Northwest National Laboratory, Richland, Washington 99352, NWChem, A Computational Chemistry Package for Parallel Computers, Version 3.0, 1997.
- <sup>28</sup>Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 1.0, as developed and distributed by the Molecular Science Computing Facility, Environmental Molecular Science Laboratory, which is part of the Pacific Northwest Laboratory, P.O. Box 999, Richland, Washington 99352, and funded by the U.S. Department of Energy. The Pacific Northwest Laboratory is a multiprogram laboratory operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DE-AC06-76RLO 1830. Contact David Feller, Karen Schuchardt, or Don Jones for further information. The database is accessible via the URL <http://www.emsl.pnl.gov:2080/forms/basisform.html>.



- <sup>29</sup>R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **54**, 724 (1971).  
<sup>30</sup>W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).  
<sup>31</sup>J. D. Dill and J. A. Pople, J. Chem. Phys. **62**, 2921 (1975).  
<sup>32</sup>P. C. Hariharan and J. A. Pople, Theor. Chim. Acta **28**, 213 (1973).  
<sup>33</sup>R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).  
<sup>34</sup>A. Schäfer, H. Horn, and R. Ahlrichs, J. Chem. Phys. **97**, 2571 (1992).  
<sup>35</sup>N. Godbout, D. R. Salahub, J. Andzelm, and E. Wimmer, Can. J. Chem. **70**, 560 (1992).