This is a repository copy of *Approaching the Hartree-Fock Limit Through the CABS Singles Correction and Auxiliary Basis Sets*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/113044/

Version: Accepted Version

Article:

Shaw, R.A. and Hill, J.G. orcid.org/0000-0002-6457-5837 (2017) Approaching the Hartree-Fock Limit Through the CABS Singles Correction and Auxiliary Basis Sets. Journal of Chemical Theory and Computation, 13 (4). pp. 1691-1698. ISSN 1549-9618

https://doi.org/10.1021/acs.jctc.7b00140

This document is the Accepted Manuscript version of a Published Work that appeared in final form in Journal of Chemical Theory and Computation, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see https://doi.org/10.1021/acs.jctc.7b00140.

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Approaching the Hartree-Fock Limit Through the CABS Singles Correction and Auxiliary Basis Sets

Robert A. Shaw and J. Grant Hill*

Department of Chemistry, University of Sheffield, Sheffield S3 7HF, U.K.

E-mail: grant.hill@sheffield.ac.uk

Abstract

Auxiliary basis sets for use in the resolution of the identity (RI) approximation in explicitly correlated methods are presented for the elements H–Ar. These extend the cc-pVnZ-F12/OptRI (n = D–Q) auxiliary basis sets of Peterson and coworkers by the addition of a small number of s- and p-functions, optimized so as to yield the greatest complementary auxiliary basis set (CABS) singles correction to the Hartree-Fock energy. The new sets, denoted OptRI+, also lead to a reduction in errors due to the RI approximation and hence an improvement in correlation energies. The atomization energies and heats of formation for a test set of small molecules, and spectroscopic constants for 27 diatomics, calculated at the CCSD(T)-F12b level, are shown to have improved error distributions for the new auxiliary basis sets with negligible additional effort. The OptRI+ sets retain all of the desirable properties of the original OptRI, including the production of smooth potential energy surfaces, whilst maintaining a compact nature.

1 Introduction

Electronic structure methods rely heavily on the expansion of the wavefunction in a finite basis of one-electron functions,[1] usually Gaussian-type orbitals.[2,3] In doing so, a basis set incompleteness error is introduced, whereby the basis does not completely span the first Sobolev space associated with the Hamiltonian.[4,5] Whilst it has only been proven that it is possible to form a complete basis,[4] not that any given basis will tend to completion, it is generally assumed that the electronic energy will converge with increasing basis size.[6] The fundamental problem therein is that this convergence is very slow for correlated methods[7] and the steep scaling[1] of such methods precludes reaching convergence on any reasonable timescale. This is largely attributed to the poor description of the electron cusp, where the wavefunction derivative should be linear in the interelectronic distance.[8,9] Explicitly correlated (F12) methods[10] solve this problem by including geminals that explicitly depend upon this distance, thereby substantially increasing the rate of convergence. However, doing so comes at the cost of the need to evaluate many 3-, 4-, and even 5-centre integrals. To avoid this potential bottleneck, the resolution of the identity (RI) approximation was introduced.[11] This rests on the fact[12] that, for any complete, not necessarily orthogonal, basis $\{|\lambda\rangle\}$,

$$\hat{1} = \sum_{\lambda\mu} |\lambda\rangle \langle\lambda|\mu\rangle^{-1} \langle\mu| \tag{1}$$

and so multi-centre integrals can be expanded over this basis, reducing their dimensionality. For example, the four-centre integral for some operator $\hat{A}$ becomes

$$\langle pq|\hat{A}|rs\rangle = \sum_{\lambda\mu} \langle pq|\lambda\rangle \langle\lambda|\mu\rangle^{-1} \langle\mu|\hat{A}|rs\rangle \tag{2}$$

where p, q, r, s denote orbitals. Further details of F12 methods can be found in a recent review.[13]

As already noted, in practice any finite basis used will not be complete, thus the above equality becomes an approximation. The original formulations of the RI approach utilised the orbital basis as the expansion basis.[11] Such a choice was found to be inefficient,[14] as these bases are generally optimized to best describe the occupied subspace; the virtual subspace often has significant overlap with the orthogonal complement of the space spanned by the orbital basis.[15] As such, auxiliary bases were introduced specifically for the RI approximation.[14] The most commonly used approach is that of the complementary auxiliary basis set[15] (CABS, also sometimes referred to as CABS+). In this, an auxiliary basis is formed so as to span the orthogonal complement of the space spanned by the orbital basis, by purification of the overlap matrix between the two sets such that their resulting overlap is zero. The CABS is then constructed as the union of the orbital and auxiliary bases.

The improvements in convergence gained by the above scheme, together with several other adaptations such as density fitting techniques,[3,16] are such that highly accurate electron correlation energies can be calculated using relatively small basis sets. At this point, the incompleteness error in the Hartree-Fock (HF) energy often outweighs that of the correlation energy.[17] In an attempt to rectify this, a perturbative correction scheme - the CABS singles correction - was proposed.[18,19] Effectively, the Fock matrix is expanded to include the auxiliary basis, and the HF orbitals allowed to relax in the larger space. The new Fock matrix $\mathbf{F}$ is constructed using the density calculated in the smaller basis, essentially augmenting the molecular orbital (MO) virtual space with the auxiliary set. The usual Roothan equations $\mathbf{FC} = \mathbf{CE}$ (assuming canonical HF orbitals) can then be solved for coefficients $\mathbf{C}$ and energies $\mathbf{E}$. Each matrix is blocked into the MO occupied orbitals (subscript O) and the MO+ABS virtual orbitals (subscript V). The energy lowering due to this single Roothan step is then given by

$$\delta E = \mathrm{Tr}\left\{\mathbf{E}_{OO} - \mathbf{F}_{OO}\right\} = \mathrm{Tr}\left\{\mathbf{F}_{OV}\mathbf{X}_{VO}\right\} \tag{3}$$

where $\mathbf{X}_{VO} = \mathbf{C}_{VO}\mathbf{C}_{OO}^{-1}$ can be found through iterative solution of

$$\mathbf{F}_{VO} + \mathbf{F}_{VV}\mathbf{X}_{VO} - \mathbf{X}_{VO}\mathbf{F}_{OO} - \mathbf{X}_{VO}\mathbf{F}_{OV}\mathbf{X}_{VO} = 0 \tag{4}$$

Most implementations, including that of MOLPRO used here,[20,21] neglect the quadratic term in equation 4, as in most cases this will be small and doing so permits non-iterative solution without diagonalization. Using the fact that the occupied-occupied block of the Fock matrix is diagonal with orbital energies ϵ_i, and transforming the virtual-virtual block to also be diagonal with energies ϵ_α, the above becomes

$$\left(\mathbf{X}_{VO}\right)_{\alpha i}\epsilon_i - \epsilon_\alpha \left(\mathbf{X}_{VO}\right)_{\alpha i} = F_{\alpha i} \tag{5}$$

Rearranging and inserting into Equation 3 gives the CABS singles correction:

$$E_{\mathrm{CABS}} = \sum_{i\alpha} \frac{\left|F_{i\alpha}\right|^2}{\epsilon_i - \epsilon_\alpha} = \sum_{i\alpha} F_{i\alpha} t_i^\alpha \tag{6}$$

where t_i^α is the single excitation amplitude from occupied orbital i into virtual orbital α. Therefore this is equivalent to the inclusion of single excitations perturbatively, possible due to the Brillouin condition-violating occupied-virtual mixings in the expanded Fock matrix. The inclusion of the CABS correction leads to a greatly improved reference energy for the subsequent correlated treatment.

Auxiliary basis sets for use in the RI approximation have been specifically matched to correlation consistent basis sets by optimization against the objective function

$$\delta\mathrm{RI} = \sum_{ij} \frac{\left(V_{ij,ij}^{\mathrm{RI}} - V_{ij,ij}^{\mathrm{RI_{ref}}}\right)^2}{\left|V_{ij,ij}^{\mathrm{RI_{ref}}}\right|} + \frac{\left(B_{ij,ij}^{\mathrm{RI}} - B_{ij,ij}^{\mathrm{RI_{ref}}}\right)^2}{\left|B_{ij,ij}^{\mathrm{RI_{ref}}}\right|} \tag{7}$$

where the $\mathbf{V}$ and $\mathbf{B}$ matrices are taken from MP2-F12 calculation determined from the approximated integrals,[20] and the ref script refers to values from a calculation performed in a very large (so as to be near-complete) reference basis.[22,23] The resulting auxiliary sets are suffixed OptRI and were designed to be compact, minimize errors due to the RI approximation, and result in smooth potential energy surfaces.[22,24,25] The choice of δRI was inspired by the 2*A MP2-F12 energy, and unlike direct minimization of the error in the MP2-F12 energy compared to the reference calculation, it led to well-behaved optimizations. However, as noted by Köhn and Tew,[26] when the OptRI auxiliary sets are used in the CABS approach the resulting basis is not sufficient to reach the implicit accuracy of the CABS singles correction.

The aim of the present contribution is to extend the cc-pVnZ-F12/OptRI (n=D, T, Q) auxiliary basis sets of Peterson and coworkers[22,25] with additional functions in order to maximize the magnitude of ΔE_{CABS}. The number of functions added is small in order to preserve the OptRI design philosophy of compact sets with well-controlled RI errors. Improvements in the CABS singles correction and correlation energies are verified for a number of molecules comprising first and second row elements, and the suitability of the resulting sets for generating smooth potential energy surfaces tested.

2 Basis set optimization

All auxiliary basis set optimizations were carried out in a locally modified version of the MOLPRO 2012.1 suite of programs,[27,28] with all other calculations carried out in MOLPRO 2015.1.[29] For simplicity, basis sets from the correlation consistent series shall be shortened to VnZ herein. Exclusively pure spherical harmonic basis functions were used and only valence electrons were correlated in the post-HF calculations. For the explicitly correlated calculations, a geminal beta exponent of 1.0 a_0^{-1} was used, with the 3C(FIX) ansatz.[30] The orbital basis sets used are the correlation consistent VnZ-F12 sets optimized specifically for use in F12 calculations.[25,31] Additionally, the Fock and exchange integrals in these were density-fitted using the aV5Z/JKFit or general-purpose def2-QZVPP/JKFit auxiliary basis sets of Weigend.[32,33] As the fitting does not impact on the calculation of the CABS singles correction, this held no relevance to the basis set optimization. It was instead chosen so as to match the reference calculations done by Köhn and Tew[26] for the molecular tests presented later. For the remaining two-electron integrals, the matching VnZ-F12/MP2Fit sets[23,34] for each zeta were used. In the reference calculations needed for determining the RI errors (Equation 7), the large even-tempered[35] MP2Fit auxiliary basis sets of Hill *et al.*[36] were extended by an additional three tight s functions. As these reference sets did not exist for some elements, the exponents for helium were taken from hydrogen and multiplied by 1.56, beryllium and magnesium used those of boron and aluminum divided by 1.56, respectively, as did lithium and sodium, but divided by a further factor of 1.56. The reference RI set was formed in the same way, taking the most diffuse exponent of each angular momenta from the relevant aV6Z[37,38] set (aV5Z[39] for group 1 and 2 elements) and dividing by 1.3. Even-tempered expansions with base 1.8 were used in all instances. The compositions for both sets are given in Table S1; all are uncontracted.

The V*n*Z-F12/OptRI sets, specifically matched to the V*n*Z-F12 orbital basis sets, were used as a starting point. The exponents of this parent set were held fixed while additional functions were added. The value of δRI (Equation 7) was calculated for all sets optimized in this work, but it was found that including it (or rather its square root, due to its negligible size) in a weighted objective function with the CABS singles energy (E_{CABS}, Equation 6) made no significant difference to the resulting exponents. In all cases, the magnitude of the change in δRI was orders smaller than the change in E_{CABS}. As this process is simply adding additional functions and not removing any, it is mathematically impossible[12] for the new basis to be 'less complete' than the previous sets, and so the error due to the RI approximation should only ever decrease. The possible exception is the case when there is linear dependency present in the basis, where most implementations will delete functions; this did not occur in any of the optimizations. Spuriously, in some instances it was found that δRI slightly increased (by, at most, tens of nE$_h$), suggesting that it is not in fact a concrete measure of the error due to the RI approximation. As such, the exponents were instead optimized solely by maximizing the magnitude of E_{CABS}. To mitigate the aforementioned possibility of linear dependency, a steep penalty function was applied such that the ratio of any new exponent to that of any existing exponent (in the union of the orbital and auxiliary bases) must be greater than or equal to 1.4. It should be noted that the default MOLPRO implementation of equation 6 does not include excitations from core orbitals, under the assumption that the amplitudes for these excitations are negligibly small. Inclusion of these 'core singles' contributions for the molecular tests described later resulted in changes to E_{CABS} of on average two percent, and changes to heats of formation of approximately 0.1 kJ mol^{-1}, regardless of which ABS was used. As such, we have chosen not to include them in calculations presented herein.

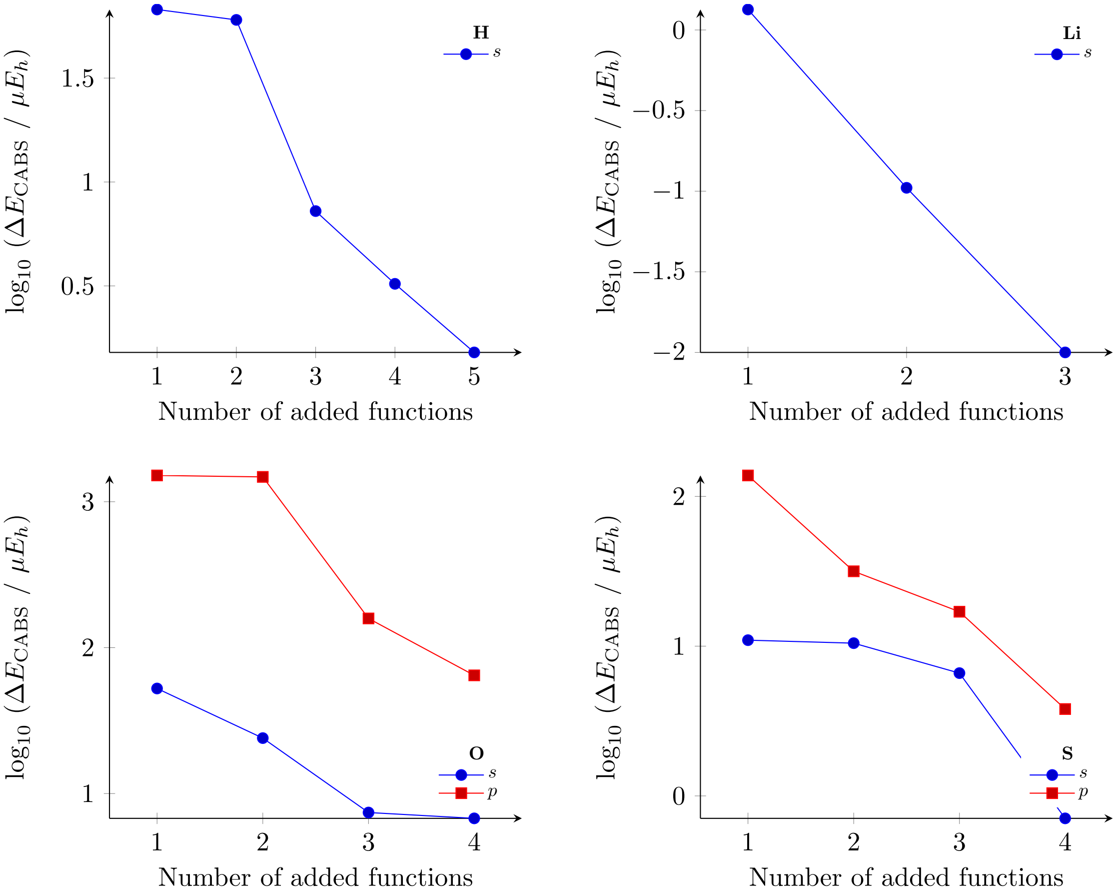

Figure 1: Contributions of additional auxiliary basis set functions to the CABS singles correction using the cc-pVDZ-F12 orbital basis for the atoms hydrogen (top left), lithium (top right), oxygen (bottom left), and sulfur (bottom right)

Exponents were optimized using the general purpose Nelder-Mead simplex algorithm in MOLPRO, with a convergence criterion of 10^{-8} E_h for the double- and triple-zeta basis sets. This was reduced to 10^{-10} for the quadruple-zeta sets as the energy changes were significantly smaller. In order to determine the number of functions to add to the parent OptRI bases, MP2-F12 calculations were carried out on neutral ground-state atoms to calculate the value of ΔE_{CABS} as the number of functions and their angular momentum shell was incremented. Figure 1 shows the incremental CABS singles correction recovered by the addition of extra auxiliary basis set functions for H, Li, O, and S with a cc-pVDZ-F12 orbital basis. The pattern of the incremental improvements in CABS singles with angular momentum was surprisingly similar for all zeta-levels of orbital basis, with only the total magnitude of the contributions changing; the relative contributions of different shells remained roughly constant.

For the atoms H–Be, Na, and Mg the only significant gains came from adding two s-type functions to the VnZ-F12/OptRI sets. For the remaining first and second row elements it is clear that p-type functions are the most important in terms of increasing the CABS singles correction, and that adding a third set of p-functions recovers substantially less energy than the first two. For the first row p-block elements a single additional s-function also contributes a relatively large amount. This s-function is significantly less important for the second row p-block elements, but it was included in order to preserve the same auxiliary basis composition for both first and second row elements. The inclusion of higher angular momentum functions, or further s- or p-functions, generally resulted in changes to the CABS singles energy of less than 1%, usually on the order of $1\mu\mathrm{E}_h$. The exception is in the case of symmetry-broken solutions, discussed in the next section, where a single f-function was often found to be mildly significant. In the case of sulfur, this was approximately 10 $\mu\mathrm{E}_h$, the same as for the first additional s-function. However, this is an artefact of the symmetry-broken nature of these cases (as evidenced by the lack of such contributions for the non-symmetry broken atoms), and in general it is not expected that higher angular momentum functions will contribute significantly at the HF level. Hence one s-function and two sets of p-type functions are added for the atoms B–Ne and Al–Ar. It is noted that this should also result in the least possible extra computational cost, as integrals over s- and p-shells are the least time consuming, and agrees with the suggestion that the addition of a few low angular momentum functions should make the biggest difference to the CABS singles correction.[26] The resulting total compositions of the new auxiliary basis sets are listed in Table S1 in the supporting information, and herein these sets are suffixed OptRI+ to signify the addition of extra functions to the existing OptRI sets.

2.1 Approaching the Hartree-Fock limit

Table 1 shows the mean percentage of total possible E_{CABS} recovered by both the old and new basis sets across all atoms. The limiting E_{CABS} value was determined by performing the calculation in the large even-tempered reference RI set described above (assuming that this is large enough to be essentially complete), but using the same VnZ-F12 orbital basis. In addition, it shows the average deviation of both the HF and HF+CABS energies from the HF limit, as calculated by Froese-Fischer.[40] For some atoms (B, C, O, F and their second-row equivalents), broken symmetry solutions are possible,[41] in which case HF calculations using an uncontracted aV6Z orbital basis were used to approximate the HF limit. The difference between the reference HF+CABS and HF limit is then entirely due to deficiencies in the orbital basis. It is common in high-accuracy calculations to try and reach the complete basis set limit via an ad hoc extrapolation, but attempts to do so here are essentially unnecessary for the HF energy: the QZ/OptRI+ HF+CABS values are already on average 32 μE$_h$ better than aV5Z and within 10 μE$_h$ of aV6Z. The latter values are themselves within 5 μE$_h$ of the Froese-Fischer limit (where available), suggesting that any attempts at extrapolation of HF energies would lead to overestimates.

Most notably, the new sets recover well over 50% of the total possible CABS energy in all cases, compared to considerably under 50% for the old sets; in fact, for each zeta, the percentages more than double. As can be seen from the deviations from the HF limit, this corresponds to absolute improvements in the energy of on average approximately 1130, 210, and 20 μE$_h$ (for double-, triple-, and quadruple zeta, respectively). This substantially

Table 1: Mean Percentage of Total Possible CABS Energy Recovered, Mean δRI, and Mean Absolute Deviations from HF Limit of HF (Δ_1) and HF+CABS (Δ_2) Energies for the Elements H-Ar

OptRI Basis	% E_{CABS}	Δ_1 (μE$_h$)	Δ_2 (μE$_h$)	δRI (μE$_h$)
DZ/OptRI	39	3325.37	2692.60	0.31
DZ/OptRI+	88	3325.37	1558.14	0.33
TZ/OptRI	35	673.72	550.46	0.04
TZ/OptRI+	78	673.72	336.24	0.08
QZ/OptRI	29	84.99	78.66	0.02
QZ/OptRI+	67	84.99	61.86	0.03

reduces the incompleteness errors present in the HF energy. Generally, the performance is uniform across all elements considered. For example, the percentage CABS singles energy recovered for the DZ sets is between 81.1% (Mg) and 96.6% (Si) for all elements except hydrogen (72.6%) and helium (46.1%). These are expected to be problem cases, as their function space is already saturated, and so only small gains are made in absolute energy terms. This is to be compared with the original OptRI sets, which have a vastly wider spread: from 21.6% for B up to 94.4% for Si (with 0.7 and 5% for hydrogen and helium). As such, the sets developed in this paper yield better reference energies for explicitly correlated calculations, and in a more consistent fashion. It can also be seen that the change in δRI, whilst paradoxically positive, is minimal and approximately four orders of magnitude smaller than the change in E_{CABS}.

3 Molecular tests

To test the performance of the new basis sets, a series of calculations were carried out on small molecules. These are split into three groups, A, B and C, listed in Table 2. The choices for group A were taken from Köhn and Tew,[26] which form a small subset of the molecules that were used originally to test explicitly correlated methods.[17,42] As these only include first row elements, group B was chosen by taking second row equivalents of those in group A. In keeping with the reference paper, equilibrium geometries were optimized at the CCSD(T)/cc-

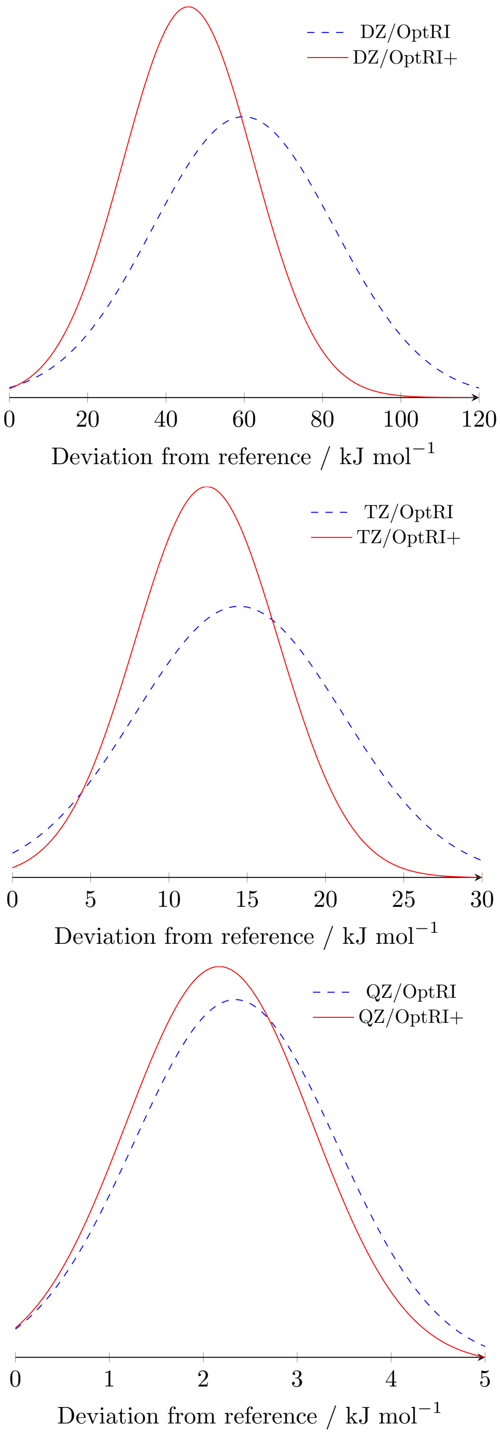

Figure 2: Normal distributions for the deviations (from the reference value) in the total CCSD-F12b energy of the molecules in groups A and B, for both the new OptRI+ (red) and existing (blue) OptRI auxiliary sets, at the DZ (top), TZ (middle) and QZ (bottom) levels.

Table 2: Groups of Small Molecules Used in Test Calculations

Group A^a	$\mathrm{C_2H_2, CF_2, CFN, CH_2O, CH_2, CH_4, CHF, CHN, cis{-}H_2N_2, CO_2, CO, F_2O, F_2, FHO, FH, FNO, H_2O_2, H_2, HNO, iso{-}CFN, iso{-}CH_2O, iso{-}CHN, iso{-}H_2N_2, N_2O, N_2, NH_3, O_3, trans{-}H_2N_2}$
Group B^b	$\mathrm{CCl_2, CClN, CH_2S, SiH_2, SiH_4, SiHF, SiHN, cis{-}H_2P_2, CS_2, CS, Cl_2O, Cl_2, ClHO, ClH, ClNO, H_2S_2, H_2S, HPO, iso{-}CClN, iso{-}CH_2S, iso{-}SiHN, iso{-}H_2P_2, P_2O, PH_3, trans{-}H_2P_2, SiO, SiO_2, P_2}$
Group C^c	$\mathrm{AlN, AlP, Ar_2, BeO, BeS, BN, Cl_2, CO, CS, F_2, FCl, H_2, HCl, He_2, HF, LiCl, LiF, MgO, MgS, N_2, NaCl, NaF, Ne_2, O_2, P_2, S_2, SiO}$

[a] Heats of formation and atomization energies calculated.
[b] Only heats of formation calculated, as reference atomization energies were not available.
[c] Spectroscopic constants calculated.

pCVTZ level[43,44] and are listed in the supporting information. For the heats of formation and atomization energies, CCSD-F12 calculations were performed using both the existing VnZ-F12/OptRI ABS, and the OptRI+ from the current work, matched with the corresponding VnZ-F12 orbital and MP2Fit bases. Benchmark values were determined for the former using an uncontracted aV6Z orbital basis, with density fitting using the aV5Z/JKFit and large even-tempered MP2Fit bases described in the previous section, whilst the RI auxiliary basis was chosen to match the aforementioned paper as aV5Z/MP2Fit. For the spectroscopic constants, the same approach was used but with added perturbative triples. The benchmark calculations used a contracted aV6Z orbital basis for all atoms except lithium, beryllium, sodium and magnesium, for which aV5Z sets were used instead, as the former do not exist. The RI basis for these was the even-tempered reference RI described earlier and overall this combination will be referred to as aV6Z for simplicity.

Normal distributions of the errors in the total energies of the molecules from groups A and B are shown in Figure 2 for the OptRI and OptRI+ sets, whilst summary statistics for the errors in the correlation energy are given in Table 3. These correlation energies do not include the CABS singles correction and thus act as a guide to the improvement in the RI

Table 3: Errors (and Standard Deviations of Errors, σ) due to Resolution of the Identity Approximation (Relative to Reference ABS) in the CCSD-F12b Correlation Energy (μE_h, per Correlated Electron), and Basis Set Incompleteness Errors (Relative to aV6Z)

Basis	Error	Mean	σ	Max
VDZ-F12	BSIE	1133.90	339.72	1927.59
	RI (OptRI)	12.14	9.03	39.23
	RI (OptRI+)	7.37	8.34	29.57
VTZ-F12	BSIE	295.11	90.22	462.12
	RI (OptRI)	2.86	4.58	14.82
	RI (OptRI+)	0.70	3.11	8.81
VQZ-F12	BSIE	43.01	24.15	100.94
	RI (OptRI)	1.71	2.03	9.91
	RI (OptRI+)	0.29	2.20	7.06

approximation itself. It is clear from Figure 2 that the change in CABS energy is yielding a significant improvement, with energies being both closer to the benchmark aV6Z energies and being a more consistent distance away. This is most drastic for $n = D, T$ - which are the most likely to be of practical use - whilst for $n = Q$, there is a much smaller, but still significant, difference. Interestingly, whilst the primary gain comes from the new HF+CABS reference energy, it can be seen from Table 3 that there are also improvements in the pure correlation energy, suggesting that the new sets have reduced the error associated with the RI approximation. Both the mean and spread of the RI errors decrease substantially in all cases, on the order of 42% (70%) and 11% (33%), respectively, for the double (triple) zeta sets. This is what should be expected, as the larger basis should be at least as complete as the smaller one, but it is in contrast with the results noted earlier with respect to the change in δRI (see Table 1). Either way, Table 3 demonstrates that the RI errors in the correlation energy are consistently two orders of magnitude smaller than the basis set incompleteness errors.

3.1 Heats of formation

The heats of formation for groups *A* and *B* were determined, with respect to H_2, CO, CO_2, N_2, F_2, Cl_2, P_2, CS, SiO, and SiO_2, using the reactions listed in the supporting information. Note that, for group *A*, these are the same as used in Ref.,[26] and the current benchmark energies agree with those in the literature to within the precision of the printed values. Normal distributions of the deviations of the enthalpies calculated using both the OptRI and OptRI+ sets compared to the benchmark values are shown in Figure 3, whilst summary statistics are given in Table 4. In most cases, at both the HF+CABS and CCSD-F12b levels, the overall distribution of errors moves closer to the datum and narrows. In particular, for the double zeta sets, the maximum error decreases by around 1 – 2 kJ mol^{-1}. However, as might be expected, the quadruple zeta sets show negligible improvement, as the CABS basis is increased by a much smaller percentage, and they were already much closer to completion. *F*-tests indicate that the error distributions for the OptRI and OptRI+ sets (both DZ and TZ) are different at the 5% significance level (under the assumption that they are normally distributed), whilst for QZ, there is no significant difference.

Table 4: Mean Absolute (MAE) and Maximum Absolute (Max) Basis Set Incompleteness Errors in HF+CABS and CCSD-F12b Heats of Formation (kJ mol^{-1}) for Molecules in Groups *A* and *B*, and Standard Deviations (σ)

Level	RI basis	MAE	σ	Max
HF+CABS	DZ/OptRI	0.417	0.562	2.263
	DZ/OptRI+	0.339	0.409	1.491
	TZ/OptRI	0.199	0.258	1.171
	TZ/OptRI+	0.194	0.268	1.332
	QZ/OptRI	0.200	0.250	1.071
	QZ/OptRI+	0.199	0.250	1.092
CCSD-F12b	DZ/OptRI	3.556	3.190	13.489
	DZ/OptRI+	3.272	2.959	11.759
	TZ/OptRI	0.768	0.760	3.931
	TZ/OptRI+	0.678	0.665	3.301
	QZ/OptRI	0.132	0.130	0.610
	QZ/OptRI+	0.130	0.130	0.632

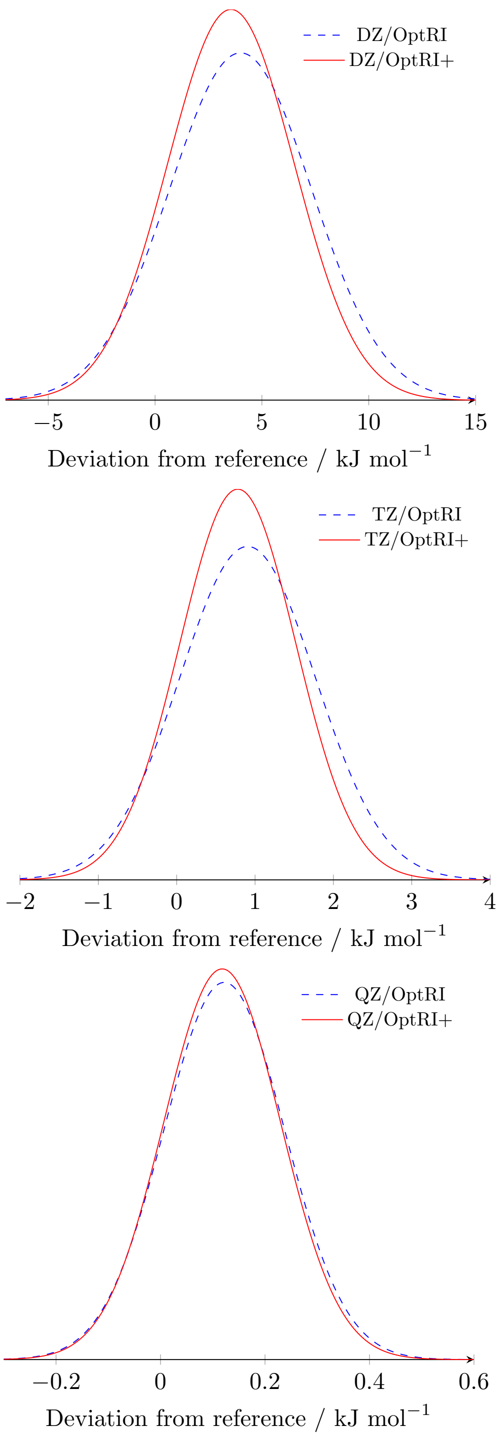

Figure 3: Normal distributions of the deviations (from benchmark values) in the heats of formation for the molecules in groups A and B, calculated at the CCSD-F12b level. Those for the OptRI (blue) and OptRI+ (red) sets are shown, for DZ (top), TZ (middle), and QZ (bottom).

3.2 Atomization energies

Atomization energies were determined for all of the molecules in group A, as benchmark-quality active thermochemistry table (ATcT) results were available for these from Ref.;[42] molecules with second row elements were not considered there, and so no data for group B was available. The errors compared to these benchmark values are summarized in Table 5. From this, it is clear that improvements on the order of 1 (0.2) kJ mol^{-1} are achieved for the DZ (TZ) quality sets presented here. It appears that there is again no significant difference for the QZ sets. Figure 4 demonstrates something stronger than this, however. The new DZ sets gave atomization energies closer (in absolute terms) to the benchmark value for all but four (out of 29) molecules: F_2, F_2O, FNO, and O_3. In the case of TZ, all were closer (or the same) except for F_2 and O_3, whilst for QZ the only exception was F_2. This indicates that the errors for VnZ-F12/OptRI and the equivalent OptRI+ set are not drawn from the same distribution, at each n. This is also confirmed by F-tests at the 5% significance level. Thus, modest improvements are achieved consistently, as was also the case with the heats of formation. The key point is that this comes at negligible extra cost, as will be shown in the next section.

Table 5: Mean Absolute (MAE) and Maximum Absolute (Max) Errors in CCSD-F12b Atomization Energies (kJ mol^{-1}, Relative to ATcT Values from Ref.[42]) for Molecules in Group A, and Standard Deviations (σ)

RI basis	MAE	σ	Max
DZ/OptRI	49.81	24.61	118.76
DZ/OptRI+	49.08	24.70	119.26
TZ/OptRI	44.37	24.03	118.95
TZ/OptRI+	44.12	24.04	119.02
QZ/OptRI	42.26	23.77	117.72
QZ/OptRI+	42.24	23.76	117.71

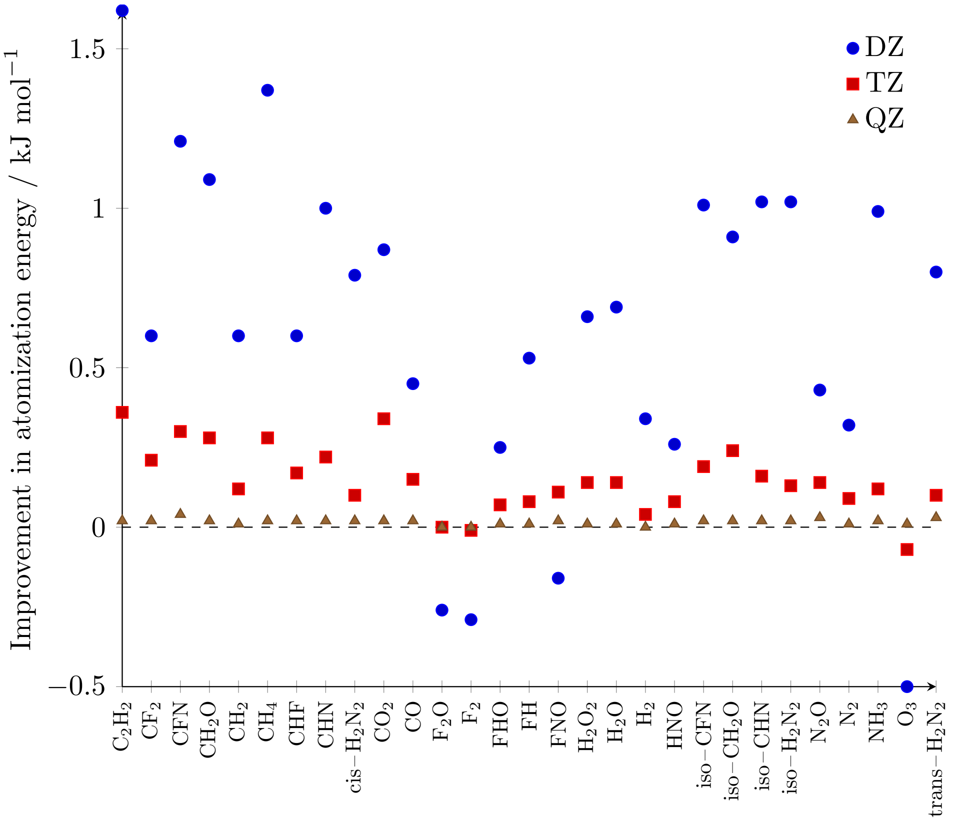

Figure 4: Errors in atomization energies for the molecules in group A calculated at the CCSD-F12b level using both OptRI and OptRI+ auxiliary sets, at DZ (blue dots), TZ (red squares), and QZ (brown triangles) levels. Shown are the differences, Δ, in absolute distance from the benchmark ATcT values (taken from Ref.[42]), determined as $\Delta_{\text{old}} - \Delta_{\text{new}}$. A positive value indicates OptRI+ was closer to the benchmark, and conversely, a negative value favors the existing OptRI set. Zero is marked with a dashed line.

3.3 Spectroscopic constants

One of the key qualities of the original OptRI sets is that they produce smooth potential energy surfaces.[22] As mentioned above, the CABS procedure usually involves purification of the overlap matrix, via application of a singular value decomposition and discarding the functions with near-zero singular values.[15] In general, this deletion may not be the same at each point on the potential surface, resulting in small discontinuities. The way to avoid this is to keep the ABS compact (which is in conflict with the desire for it to be complete) and by constraining the exponents such that they are not too close. As OptRI+ adds additional functions, albeit very few, it is important to check that the resulting calculated surfaces are smooth. This is most reasonably achieved through the calculation of spectroscopic constants for a set of diatomics, group C in Table 2. The CCSD(T)-F12b energies were calculated at seven points centered around the calculated equilibrium bond distance, followed by a sixth-order polynomial fit with the derivatives of this fit used in the usual second-order perturbation theory calculations.[45] From this, the bond dissociation energy, equilibrium bond distance, and principal harmonic frequency are determined. Table 6 shows the errors, compared to the reference calculations described above, in these constants. There are no substantial differences between the values calculated with the original OptRI ABS compared to those presented here, although slight improvements are seen, particularly in the frequencies. However, the fact that they are almost the same or marginally better demonstrates that linear dependencies are not being introduced, and the potential surfaces are smooth; this is an essential quality in terms of practical use.

Equally important is that the addition of functions has not introduced significant computational cost. Table 7 shows the average processing time required for a point on a potential energy curve, and the percentage change in time taken. Even at the double zeta level - where the percentage increase in basis size is at its largest - this difference equates to on average less than a second, or just under 3%. For the triple zeta sets, the calculations taken on average roughly 2 seconds *less*, which is remarkable. A possible explanation is that the proportional change in basis size is small enough to be outweighed by the improvements in reference energy, which as has been noted earlier are still substantial at this level. As such, the new basis can result in fewer coupled-cluster iterations (roughly half an iteration on average), the time-savings from which are greater than the penalty associated with additional integrals and marginally larger matrices. For reference, the mean wall time for calculations in the large reference basis used as a benchmark was approximately 22,000 seconds.

Table 6: Mean Absolute Errors[a] (with Standard Deviations in Parentheses) for HF+CABS and CCSD(T)-F12b Bond Dissociation Energies (D_e), Equilibrium Bond Distances (R_e), and Principal Harmonic Frequencies (ω_e), for Molecules in Group C

	D_e error (kJ mol^{-1})		R_e error (10^{-3}Å)		ω_e error (cm^{-1})	
RI basis	HF	CC	HF	CC	HF	CC
DZ/OptRI	1.27 (2.43)	8.32 (8.21)	4.1 (11.0)	6.6 (18.0)	1.26 (1.19)	3.58 (4.36)
DZ/OptRI+	1.35 (2.94)	8.23 (7.91)	2.8 (9.2)	6.0 (16.0)	0.76 (0.92)	3.34 (4.25)
TZ/OptRI	0.22 (0.22)	2.74 (1.96)	1.2 (3.3)	7.3 (26.0)	0.42 (0.64)	1.87 (2.11)
TZ/OptRI+	0.19 (0.19)	2.70 (1.94)	1.2 (3.6)	7.3 (26.0)	0.36 (0.55)	1.78 (2.16)
QZ/OptRI	0.17 (0.30)	1.61 (1.58)	0.5 (1.5)	3.0 (9.0)	0.17 (0.28)	0.71 (1.04)
QZ/OptRI+	0.16 (0.30)	1.61 (1.57)	0.4 (1.7)	3.0 (9.0)	0.17 (0.29)	0.71 (1.10)

[a] Relative to those determined using the even-tempered reference RI basis.

Table 7: Mean Wall Time [a] ($\bar{t}$, seconds) for a Single CCSD(T)-F12 Calculation on the Diatomics in Group C, and Mean Number of Coupled-Cluster Iterations[b] ($\bar{n}_{\text{iter}}$)

	OptRI		OptRI+		
	$\bar{t}$	$\bar{n}_{\text{iter}}$	$\bar{t}$	$\bar{n}_{\text{iter}}$	% Change
DZ	29.8	22.5	30.6	22.5	2.7
TZ	70.6	22.5	68.7	22	−2.7
QZ	226.4	22	226.2	22	0.0

[a] Determined from the average over 7 steps in a potential energy surface scan. All calculations were done on a single processor.

[b] Rounded to the nearest half iteration.

4 Conclusions

The VnZ-F12/OptRI (n =D–Q) auxiliary basis sets for the atoms H–Ar have been augmented with a small number of s- and p-functions in such a way as to maximize the CABS singles correction to the reference energy in explicitly correlated calculations. The resulting OptRI+ sets produce HF+CABS reference energies that are substantially closer to the Hartree-Fock limit. Additionally, the error in the correlation energy due to the RI approximation of the three- and four-centre integrals can be seen to be reduced, with errors compared to the large basis limit both closer to zero and more narrowly distributed. The improvements for relative energies are less striking but still noticeable, with the new DZ and TZ sets resulting in more accurate heats of formation and atomization energies, quantities that are traditionally very difficult to determine accurately. These improvements are modest but significant due to their consistency and overall distribution. Crucially, the small number of additional functions means that any increases in calculation time are negligible, and in some cases compute times decrease due to a reduction in the number of coupled-cluster iterations required for convergence. The new bases share with their predecessors the ability to generate smooth potential energy surfaces, offering slightly superior spectroscopic constants.

The benchmarking of the new OptRI+ sets indicates that they retain all of the desirable properties of the parent OptRI sets; compact, well-controlled errors, and smooth surfaces. When this is combined with the improvements in absolute and relative energies, at essentially no additional computational cost, there is no reason why they should not be routinely used in place of the original OptRI. The new auxiliary basis sets are provided in machine readable form as part of the supporting information, and will also be made available to download from ccRepo, a correlation consistent basis set repository.[46]

Acknowledgement

The authors thank the Engineering and Physical Sciences Research Council (UK) for a postgraduate studentship awarded to R.A.S. and project funding (EP/N02253X/1).

Supporting Information Available

The V*n*Z-F12/OptRI+ auxiliary basis sets in MOLPRO format, compositions of auxiliary basis sets, reactions used to determine the heats of formation, and optimized structures and heats of formation for the molecules in group *B*. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic Structure Theory*; John Wiley & Sons, Inc., 2000; pp 315–327.

(2) Boys, S. F. Electronic Wave Functions. I. A General Method of Calculation for the Stationary States of Any Molecular System. *Proc. R. Soc. A* **1950**, *200*, 542–554.

(3) Hill, J. G. Gaussian basis sets for molecular applications. *Int. J. Quantum Chem.* **2013**, *113*, 21–34.

(4) Klahn, B.; Bingel, W. A. The convergence of the Rayleigh-Ritz Method in quantum chemistry. *Theor. Chim. Acta* **1977**, *44*, 27–43.

(5) Kutzelnigg, W. Expansion of a wave function in a Gaussian basis. I. Local versus global approximation. *Int. J. Quantum Chem.* **2013**, *113*, 203–217.

(6) Jensen, F. Estimating the Hartree-Fock limit from finite basis set calculations. *Theor. Chem. Acc.* **2005**, *113*, 267–273.

(7) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N_2, and H_2O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(8) Kato, T. On the eigenfunctions of many-particle systems in quantum mechanics. *J. Chem. Phys.* **1957**, *10*, 151–177.

(9) Pack, R. T.; Brown, W. B. Cusp Conditions for Molecular Wavefunctions. *J. Chem. Phys.* **1966**, *45*, 556–559.

(10) Klopper, W.; Manby, F. R.; Ten-No, S.; Valeev, E. F. R12 methods in explicitly correlated molecular electronic structure theory. *Int. Rev. Phys. Chem.* **2006**, *25*, 427–468.

(11) Klopper, W.; Kutzelnigg, W. Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. II. Second-order Møller-Plesset (MP2-R12) calculations on closed-shell atoms. *J. Chem. Phys.* **1991**, *94*, 2002–2019.

(12) Kato, T. *Perturbation Theory for Linear Operators*, 2nd ed.; Springer-Verlag, 1980; pp 353–356.

(13) Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. Explicitly Correlated Electrons in Molecules. *Chem. Rev.* **2012**, *112*, 4–74.

(14) Klopper, W.; Samson, C. C. M. Explicitly correlated second-order Møller-Plesset methods with auxiliary basis sets. *J. Chem. Phys.* **2002**, *116*, 6397–6410.

(15) Valeev, E. F. Improving on the resolution of the identity in linear R12 ab initio theories. *Chem. Phys. Lett.* **2004**, *395*, 190–195.

(16) Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496–4501.

(17) Bischoff, F. A.; Wolfsegger, S.; Tew, D. P.; Klopper, W. Assessment of basis sets for F12 explicitly-correlated molecular electronic-structure methods. *Mol. Phys.* **2009**, *107*, 963–975.

(18) Adler, T. B.; Knizia, G.; Werner, H.-J. A simple and efficient CCSD(T)-F12 approximation. *J. Chem. Phys.* **2007**, *127*, 221106.

(19) Noga, J.; Šimunek, J. On the one-particle basis set relaxation in R12 based theories. *Chem. Phys.* **2009**, *356*, 1–6.

(20) Werner, H.-J.; Adler, T. B.; Manby, F. R. General orbital invariant MP2-F12 theory. *J. Chem. Phys.* **2007**, *126*, 164102.

(21) Knizia, G.; Werner, H.-J. Explicitly correlated RMP2 for high-spin open-shell reference states. *J. Chem. Phys.* **2008**, *128*, 154103.

(22) Yousaf, K. E.; Peterson, K. A. Optimized auxiliary basis sets for explicitly correlated methods. *J. Chem. Phys.* **2008**, *129*, 184108.

(23) Kritikou, S.; Hill, J. G. Auxiliary Basis Sets for Density Fitting in Explicitly Correlated Calculations: The Atoms H-Ar. *J. Chem. Theory Comput.* **2015**, *11*, 5269–5276.

(24) Yousaf, K. E.; Peterson, K. A. Optimized complementary auxiliary basis sets for explicitly correlated methods: aug-cc-pVnZ orbital basis sets. *Chem. Phys. Lett.* **2009**, *476*, 303–307.

(25) Hill, J. G.; Peterson, K. A. Correlation consistent basis sets for explicitly correlated wavefunctions: valence and core-valence basis sets for Li, Be, Na, and Mg. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10460.

(26) Köhn, A.; Tew, D. P. Towards the Hartree-Fock and coupled-cluster singles and doubles basis set limit: A study of various models that employ single excitations into a complementary auxiliary basis set. *J. Chem. Phys.* **2010**, *132*, 024101.

(27) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. MOLPRO, version 2012.1, a package of *ab initio* programs. 2012.

(28) Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. Molpro: A general-purpose quantum chemistry program package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 242–253.

(29) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Györffy, W.; Kats, D.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. MOLPRO, version 2015.1, a package of *ab initio* programs. 2015.

(30) Ten-no, S. Initiation of explicitly correlated Slater-type geminal theory. *Chem. Phys. Lett.* **2004**, *398*, 56–61.

(31) Peterson, K. A.; Adler, T. B.; Werner, H.-J. Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, B-Ne, and Al-Ar. *J. Chem. Phys.* **2008**, *128*, 084102.

(32) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

(33) Weigend, F. Hartree-Fock exchange fitting basis sets for H to Rn. *J. Comput. Chem.* **2008**, *29*, 167–175.

(34) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.

(35) Ruedenberg, K. In *Energy, Structure, and Reactivity*; Smith, D. W., McRae, W. B., Eds.; Wiley-Interscience: New York, 1973; pp 164–169.

(36) Hill, J. G.; Peterson, K. A.; Knizia, G.; Werner, H.-J. Extrapolating MP2 and CCSD explicitly correlated correlation energies to the complete basis set limit with first and second row correlation consistent basis sets. *J. Chem. Phys.* **2009**, *131*, 194105.

(37) van Mourik, T.; Wilson, A. K.; Dunning, T. H. Benchmark calculations with correlated molecular wavefunctions. XIII. Potential energy curves for He_2, Ne_2 and Ar_2 using correlation consistent basis sets through augmented sextuple zeta. *Mol. Phys.* **1999**, *96*, 529–547.

(38) van Mourik, T.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. VIII. Standard and augmented sextuple zeta correlation consistent basis sets for aluminum through argon. *Int. J. Quantum Chem.* **2000**, *76*, 205–221.

(39) Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning, T. H.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. VII. Valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg. *Theor. Chem. Acc.* **2011**, *128*, 69–82.

(40) Fischer, C. F. Hartree-Fock Calculations for Atoms with Inner-Shell Vacancies. *Phys. Rev. Lett.* **1977**, *38*, 1075–1077.

(41) Löwdin, P.-O. Quantum Theory of Many-Particle Systems. I. Physical Interpretations by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configurational Interaction. *Phys. Rev.* **1955**, *97*, 1474–1489.

(42) Klopper, W.; Ruscic, B.; Tew, D. P.; Bischoff, F. A.; Wolfsegger, S. Atomization energies from coupled-cluster calculations augmented with explicitly-correlated perturbation theory. *Chem. Phys.* **2009**, *356*, 14–24.

(43) Woon, D. E.; Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon. *J. Chem. Phys.* **1995**, *103*, 4572–4585.

(44) Peterson, K. A.; Dunning Jr., T. H. Accurate correlation consistent basis sets for molecular core-valence correlation effects. The second row atoms Al - Ar, and the first row atoms B - Ne revisted. *J. Chem. Phys.* **2002**, *117*, 10548.

(45) Murrell, J. N.; Carter, S.; Farantos, S. C.; Huxley, P.; Varandas, A. J. C. *Molecular Potential Energy Functions*; John Wiley & Sons, Inc.: Chichester, 1984; pp 1–11.

(46) Hill, J. G., Correlation consistent basis set repository. `http://bit.ly/ccBasis`, 2017; Accessed: 2017-02-28.

Graphical TOC Entry