Article:

Shaw, R.A. and Hill, J. orcid.org/0000-0002-6457-5837 (2018) Midbond basis functions for weakly bound complexes. Molecular Physics, 116 (11). pp. 1460-1470. ISSN 0026-8976

https://doi.org/10.1080/00268976.2018.1440018

This is an Accepted Manuscript of an article published by Taylor & Francis in Molecular Physics on 26 Feb 2018, available online: http://www.tandfonline.com/10.1080/00268976.2018.1440018.

Midbond basis functions for weakly bound complexes

Robert A. Shaw and J. Grant Hill

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

ARTICLE HISTORY

Compiled February 1, 2018

ABSTRACT

Weakly bound systems present a difficult problem for conventional atom-centred basis sets due to large separations, necessitating the use of large, computationally expensive bases. This can be remedied by placing a small number of functions in the region between molecules in the complex. We present compact sets of optimised midbond functions for a range of complexes involving noble gases, alkali metals, and small molecules for use in high accuracy coupled cluster calculations, along with a more robust procedure for their optimisation. It is shown that excellent results are possible with double-zeta quality orbital basis sets when a few midbond functions are added, improving both the interaction energy and the equilibrium bond lengths of a series of noble gas dimers by 47 and 8%, respectively. When used in conjunction with explicitly correlated methods, near complete basis set limit accuracy is readily achievable at a fraction of the cost that using a large basis would entail. General purpose auxiliary sets are developed to allow explicitly correlated midbond function studies to be carried out, making it feasible to perform very high accuracy calculations on weakly bound complexes.

KEYWORDS

Midbond functions, explicitly correlated, noncovalent interactions

1. Introduction

It is common practice in quantum chemistry to expand the wavefunction in a finite set of atom-centred Gaussian basis functions[1, 2]. The advantage of this particular choice of basis lies in the ease of evaluation of the numerous integrals required, allowing for accurate calculations on larger systems than could be achieved with other choices, such as Slater-type functions. However, it is well-known that correlation energies converge slowly with the size of the basis set[3, 4],

---

CONTACT J. Grant Hill Email: grant.hill@sheffield.ac.uk

with large numbers of high angular momentum and diffuse functions necessary to approach the complete basis set (CBS) limit. This is particularly problematic in weakly bound systems, where intermolecular separations can be large and dispersion, i.e. electron correlation, can be very important. The latter dependence, combined with the small energy differences involved, mean that only the most expensive correlated wavefunction methods suffice, with coupled cluster with single, double and perturbative triple excitations [CCSD(T)][5] being the standard. Given the N^7 scaling of CCSD(T), where N is the number of basis functions, benchmark calculations are only possible for relatively small systems. Density functionals, which would be considerably cheaper to use, perform especially poorly on non-covalent interactions[6–8], although considerable effort has recently gone in to improving this situation[9, 10].

The reasons for the slow convergence are twofold. Fundamentally, a basis spans a function space, and is said to be 'complete', if any member of that space can be written as a linear combination of functions in the basis. In practice, only a compact region of the space need be described, allowing for a finite basis to be adequate and with enough functions become near complete. Firstly, electronic wavefunctions have a cusp where the inter-electronic separation tends to zero, where they must satisfy certain conditions[11]; Gaussians are everywhere continuous, however, and therefore the cusp cannot be represented with a finite number of such functions. Secondly, the basis set is atom centred, meaning that if two molecules are well separated, the 'bonding' region in between is unevenly described compared to the molecules themselves. The problem is exacerbated by the overly rapid e^{-r^2} decay of the Gaussians. This is why a large number of diffuse functions become necessary when studying weakly-bound complexes.

Explicitly correlated (F12) methods[12–14] correct for the first of these issues by including geminals with factors linear in the inter-electronic separation. It has been shown repeatedly that these F12 approaches can achieve convergence in the energy much more efficiently, for example allowing calculations using double-zeta quality basis sets to achieve between triple- and quadruple-zeta accuracy[15, 16]. This does however come at a considerable computational cost, as it introduces a significant number of difficult many-electron integrals that would not be required in a canonical calculation. These can be circumvented by using resolution-of-the-identity and density-fitting approaches[17–20], whereby a large auxiliary basis is used to expand the many-body integrals in terms of simpler, fewer-body integrals.

F12 methods perform remarkably well for a broad range of systems, but by focusing on the cusps they do not significantly improve the long-range description. As a result, slow convergence can still be observed in intermolecular interactions, particularly in difficult cases where the separations are very large or the interaction very weak, such as the noble gas dimers[21]. There has been some success in applying simple extrapolations of energies to approximate the complete basis set limit[3, 22–24], but these are generally quite situational and method specific. An alternative approach is to use additional basis functions centred in the middle of bonds, or more precisely, in regions where electron density is anticipated to accumulate but where there are no atomic centers[25–30]. Such midbond functions (MBF) have been shown to provide a greater improvement to interaction energies in many weakly-bound systems than using explicitly correlated methods[31]. The idea behind it is intuitive, as the simplest way to provide a better description of a region in space is to place functions in that region.

There are several potential problems, however, not least of which is due to basis set superposition errors[32]. If care is not taken to account for these, for example by using the counterpoise procedure of Boys and Bernardi[33], spuriously large interaction energies would be observed. This is due to an uneven improvement in the description of the interaction compared to the monomers, which, while potentially still problematic, would not be as significant if the MBFs were not present. An additional issue pertains to the computation of potential energy surfaces, where the uncertainty in the optimum position of the midbond functions could result in subtle differences in the quality of results at points on the surface with substantial changes in geometry. This would in turn affect any calculated forces and force constants. This is not so much a problem with the MBFs themselves, however, but rather the choice to use exclusively atom-centred functions in most calculations. This inherently means that the basis functions must move as the molecules move, creating these discrepancies. A more ideal, but considerably more expensive, approach would be to use fixed grids of functions, decoupling the nuclear positions from the mathematical description of the space. Similarly, alternative types of basis functions could be used, such as the elliptical Gaussians proposed by Mester *et al.*[30], or a variety of others[34, 35]. One sensible solution would be to use a mixed Gaussian and plane-wave basis[36], with the former allowing for compact representations of the molecules and the latter giving improved densities in between. Such sets are used in several other types of calculation[37, 38]; the difficulty here, however, is that the codes used to perform high-accuracy studies on non-covalent interactions are generally not set up to calculate the necessary integrals. Therefore MBFs are an efficient compromise.

Previous studies into the use of bond functions have focused almost exclusively on small, dimeric systems using canonical coupled-cluster methods[28, 29]. Detail on what constitutes a sufficient number of MBFs, and the importance of different angular momentum shells, is sparse in the literature. There is a prevalence in the assumption that these functions are not particularly situation specific, with exponents only roughly optimised, and that the results seem insensitive to where they are placed[26]. It is not *a priori* obvious that this should be the case when systems with more than two molecules are considered, or where substantially different atoms are involved in the interaction. Preparing such functions for use with F12 methods also presents a set of difficulties, due to the need for auxiliary basis sets (ABSs); as such, a previous study simply used a pre-existing hydrogen basis with its matching auxiliary sets[31]. In the current study, we demonstrate a procedure for the robust optimisation of midbond functions for use in both canonical and explicitly-correlated calculations. The properties of these are then investigated with regards to the placement and system-specificity of the functions, and how they can be used effectively to reach convergence rapidly in weakly-bound systems.

2. Methods

The optimisation of Gaussian exponents and positions is a difficult, highly non-linear problem. Traditionally, such optimisations are carried out using general purpose algorithms, predominantly Nelder-Mead simplex or quasi-Newton methods[39], with gradients estimated if used at all. The objective function is typically chosen to be the total energy. There is an underlying assumption that the problem is far from convex, with several shallow minima close to each other. This is necessarily the case as the basis set approaches completeness and numerical constraints mean that linear dependencies are introduced. However, this assumption has not been investigated in cases far from completeness, as especially applies to midbond functions. MBFs present the added difficulty that their position needs to be determined as well as their exponent. As such, more robust optimisation procedures are desirable. In the first part of this section, we briefly present analytical gradients for restricted Hartree-Fock and MP2 energies with respect to basis function exponent. These can then be used in full Newton-step optimisations, for which convergence is guaranteed if the problem is convex[40]. The second part of the section details the overall optimisation procedure.

2.1. Analytical gradients

Gaussian-type functions take the form

$$\phi_{l_x l_y l_z}(\mathbf{r}; \alpha, \mathbf{A}) = x_A^{l_x} y_A^{l_y} z_A^{l_z} \exp(-\alpha r_A^2) \tag{1}$$

where $q_A = q - A_q$, with $q = x, y, z$, and $r_A^2 = x_A^2 + y_A^2 + z_A^2$. The total angular momentum of ϕ_{ijk} is $l = l_x + l_y + l_z$. The derivative with respect to the exponent, α, is trivially given by

$$\partial_\alpha \phi_{l_x l_y l_z} = -r_A^2 \phi_{l_x l_y l_z} = -\sum_{q=x,y,z} \phi_{l_q+2} \tag{2}$$

Using equation 2, the derivative of any integral over the Gaussians can be found by simply shifting the angular momenta as shown and using the pre-existing integration procedures.

Integrals become more expensive to evaluate with increasing angular momentum, however, especially when considering the two-electron integrals. An alternative, particularly when gradients with respect to the positions are also needed, is to use the following result:

$$\partial_{A_q A_q} \phi_{l_x l_y l_z} = \phi_{l_q+2} - (2l_q + 1)\phi_{l_x l_y l_z} + l_q(l_q - 1)\phi_{l_q-2} \tag{3}$$

A simple rearrangement of equation 3 allows us to rewrite equation 2 as

$$\partial_\alpha \phi_{ijk} = -\left[\nabla_A^2 + 2l + 3\right]\phi_{ijk} + \sum_{q=x,y,z} l_q(l_q - 1)\phi_{l_q-2} \tag{4}$$

Equation 4 can then be used to write the gradient integrals in terms of the second-derivative integrals with respect to position (via the Laplacian) and the undifferentiated integrals (for the $2l + 3$ term). The final term above requires additional integrals, but with smaller total angular momenta - this means that such integrals are not required for functions with $l < 2$, and are relatively cheap to evaluate.

Once the derivative integrals have been determined, the energy gradients can be calculated in the same way as they would be with respect to nuclear coordinates. The restricted Hartree-Fock energy is given by

$$E_{\mathrm{HF}} = \frac{1}{2}\sum_{\mu\nu} \left\{h_{\mu\nu} + f_{\mu\nu}\right\} P_{\mu\nu} \tag{5}$$

where $\mathbf{h}$, $\mathbf{f}$, and $\mathbf{P}$ are the core Hamiltonian, Fock, and density matrices, respectively, and μ, ν denote basis functions. The Fock matrix elements are as per usual given by

$$f_{\mu\nu} = h_{\mu\nu} + \sum_{\lambda\sigma} \langle \mu\nu || \lambda\sigma \rangle P_{\lambda\sigma}$$

where $\langle \mu\nu || \lambda\sigma \rangle$ are the antisymmetrised two-electron integrals. The derivative of equation 5 with respect to alpha is thus as follows:

$$\partial_\alpha E_{\mathrm{HF}} = \sum_{\mu\nu} \left\{ \partial_\alpha h_{\mu\nu} + \frac{1}{2} \sum_{\lambda\sigma} P_{\lambda\sigma} \partial_\alpha \langle \mu\nu || \lambda\sigma \rangle \right\} P_{\mu\nu} + \sum_{\mu\nu} f_{\mu\nu} \partial_\alpha P_{\mu\nu} \tag{6}$$

The final term in equation 6 can be avoided by differentiating the orthonormality relation for the molecular orbitals and the definition of the density matrix, $P_{\mu\nu} = \sum_i C_{\mu i} C_{\nu i}$, where $\mathbf{C}$ is the tensor of molecular orbital coefficients and i denotes an occupied molecular orbital. This gives

$$\sum_{\mu\nu} f_{\mu\nu} \partial_\alpha P_{\mu\nu} = - \sum_{\mu\nu} W_{\mu\nu} \partial_\alpha S_{\mu\nu} \tag{7}$$

where $\mathbf{S}$ is the overlap matrix, and $W_{\mu\nu} = \sum_i \epsilon_i C_{\mu i} C_{\nu i}$ is the weighted density matrix, with ϵ_i the eigenvalue associated with molecular orbital i.

The MP2 contribution to the energy gradient is considerably more complicated[41], as the density matrix derivatives can no longer be avoided; the details are deferred to the supplementary material. The density derivatives, found through solution of coupled-perturbed Hartree-Fock equations[42], are also necessary to determine the Hartree-Fock Hessian analytically[43]. As equation 4 necessitates the second derivatives with respect to positions of the integrals, the Hessian can be evaluated at little extra cost, allowing the positions of the MBFs to be optimised using a full Newton-Raphson procedure. The Hessian with respect to the exponents is calculable in exactly the same way, although additional derivative integrals are necessary; these can be found by shifting angular momenta as per a reapplication of equation 2, or by finite differences.

2.2. Optimisation procedure

For any given system X···Y, midbond functions were initially placed halfway between the points of least separation on X and Y; for example, in H_2O···Ar this would be the midpoint between the oxygen and the argon atoms. Functions were added incrementally in angular momentum shells, with all functions of the same angular momentum optimised simultaneously. The analytic gradients and Hessians described above were used to optimise the exponents and positions at the MP2 level via a standard trust-region Newton-Raphson procedure[44]. The exponents of the rest of the orbital basis were constrained to be fixed, with the geometries taken from the reference values in Tables 1 and 2 for the group 1 and 18 complexes, and from Ref. [31] for the

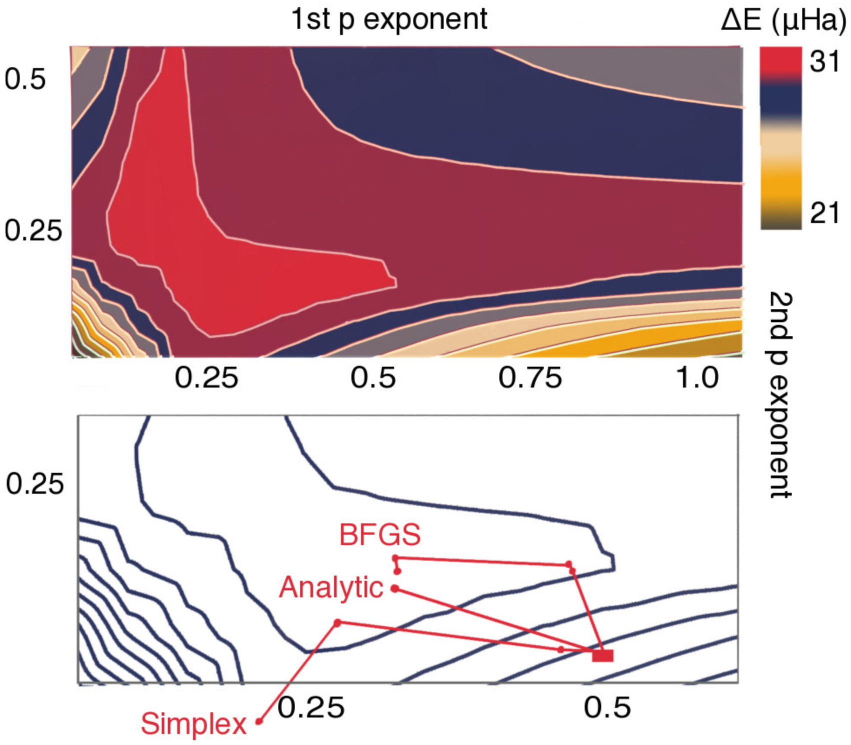

Figure 1. Contours showing how the interaction energy of the helium dimer depends on the first two p-exponents of midbond functions located halfway between the two helium atoms, at the CCSD(T)/aVDZ level. The results of three different optimisation methods from a starting guess of (0.1, 0.5) are shown as red lines on the magnified lower plot; one point corresponds to five steps in the optimisation.

small molecular systems. The optimisation was attempted on the positions and exponents both separately and simultaneously. In general, it was found that the MBFs would move towards the heavier atom in the interaction - for example argon in the system mentioned above. While this lowered the total energy, it did not improve the interaction energies, strongly suggesting that the functions were being used to preferentially improve the description of one of the monomers, rather than the description of the interaction itself. This effect was far less pronounced when triple-zeta quality basis sets were used instead of double-zeta sets, with the midbond functions moving very little from the centre position in the case of the former. As such, the MBFs were fixed to be at the midpoint for all subsequent calculations. The total energy was optimised with a convergence threshold of 10^{-7}E$_h$ in all cases, using in-house electronic structure software. This was done separately using diffuse augmented correlation consistent, aug-cc-pVxZ (aVxZ), basis sets for all atoms,[45–50] with $x = \mathrm{D}, \mathrm{T}$. The evaluation of spectroscopic constants for the group 1 elements was carried out using augmented weighted core-valence sets, aug-cc-pwCVDZ-PP (awCVDZ),[50], with the core electrons correlated in the coupled cluster calculations. All sets for the elements K, Rb and Cs were paired with the small-core relativistic pseudopotentials of Lim *et al.*[51]

Figure 1 shows the interaction energy landscape for the helium dimer as a function of the first two MBF p-exponents. This demonstrates a butterfly-like bowl shape suggestive of a degree of convexity to the problem; there are however substantially flat regions. The efficacy of the analytic optimisation is demonstrated, taking only five steps to reach a minimum, as compared with 20 for the standard BFGS procedure. The popular simplex algorithm takes 15 steps without converging, resulting in an erroneous negative exponent. This is symptomatic of the sensitivity of the MBFs to small changes in exponent in this instance, where the difference between (0.1, 0.5) and (0.2, 0.5) is over 30% of the interaction energy.

All further interaction energy calculations were performed at the CCSD(T) or CCSD(T)-F12b[52] levels in the MOLPRO suite of programs[53], using the Boys and Bernardi full counterpoise correction[33] to account for basis set superposition errors. It should be noted that superposition errors are greatly exacerbated by midbond functions, such that correcting for them is essential. Firstly, the MP2-optimised exponents were further optimised at this level using MOLPRO's general purpose BFGS optimiser, with a tighter energy threshold of 10^{-9} E_h; in all cases, the exponents changed by less than 10^{-4}, essentially corresponding to no change. For the explicitly correlated calculations, the 3C(Fix) ansätz was used,[54] with a geminal beta of 1.0 a_0^{-1}. This requires three auxiliary basis sets, for density-fitting of the Coulomb and exchange matrices (JKFit), the MP2 integrals (MP2Fit), and the resolution-of-the-identity approximation (OptRI). The matching auxiliary sets[55–59] for the aVxZ bases were used, but as these were not available for the noble gases, the def2-QZVPP/JKFit sets were used instead[58]. For the MBFs, auxiliary functions were optimised by minimizing the error in the total energy compared to results obtained using large, near-complete reference sets comprising the aV5Z ABSs for carbon[55, 57–59]; this was done only for the $H_2O\cdots Ar$ complex, and as will be explained later, the results transferred to the other systems.

3. Results

Midbond functions should be useful in complexes with large separations, whereby small orbital bases are unlikely to be able to describe the intermediate regions well. Good examples of such systems are diatomics involving group 1 or group 18 elements. In the latter case particularly, the very weakly-bound, dispersive interactions are described poorly by conventional basis sets, typically requiring very large sets to be used to achieve reasonable accuracy. For the alkali metals, which are especially interesting in the context of ultracold chemistry[60, 61], the interaction energies are much larger (on the order of 10 kcal mol^{-1}), but the dispersion

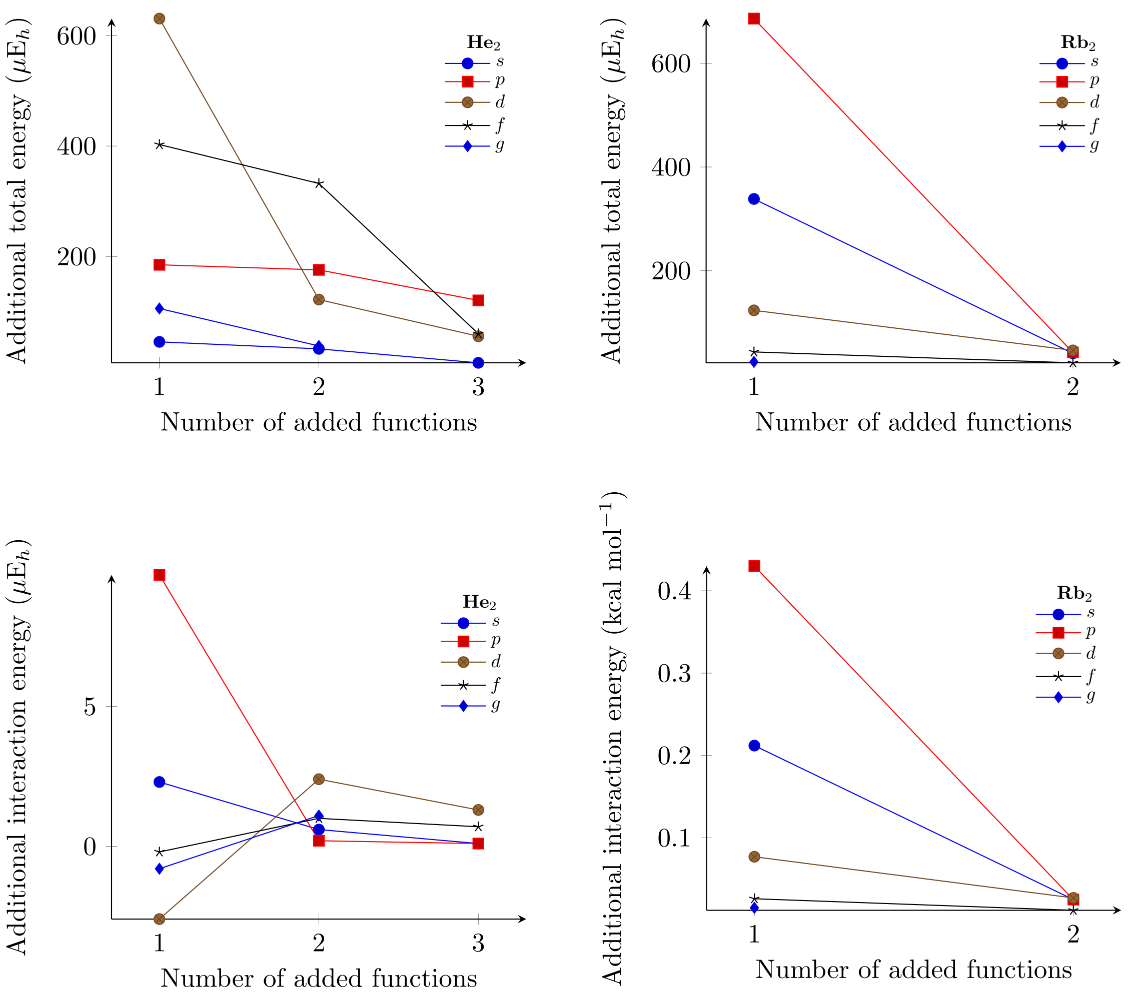

Figure 2. Incremental total (top) and interaction (bottom) energy gains from using additional midbond functions of varying angular momenta for the helium (left) and rubidium (right) dimers. CCSD(T)/aVDZ results are shown, with interatomic separations of 2.9695 and 4.4 Å for helium and rubidium respectively.

effects are anomalously long-range. In addition, the high accuracy necessitated by experimentally motivated studies into these complexes requires consideration of core-valence effects and excitations beyond perturbative triples. In such cases, the use of very large, saturated orbital bases is unfeasible. Previous studies looking at the use of MBFs have often focused on the noble gas diatomics, and typically use several high angular momentum functions: the most frequently cited include those of Tao and Pan with configuration $[3s3p2d1f]$[26], or the $[3s3p2d1f1g]$ set of Partridge *et al.*[29] It was suggested that these functions, not necessarily optimised, would be applicable across a broad range of systems without change.

Figure 2 shows the energy changes from incrementally adding angular momentum shells of MBFs in a manner similar to correlation-consistent basis sets. For the helium dimer, the higher angular momentum functions contributed greatly to the total energy of the system, with the first d and f function being the most important. However, the lower left plot shows that these two functions are not improving the interaction energy, but making it slightly worse - that is,

Table 1. Interaction energies (ΔE) and equilibrium bond lengths (R_e) for a series of noble gas diatomics, calculated at the CCSD(T)/aVDZ level with and without midbond functions. Reference values from Refs. [62, 63] are given.

	ΔE (μE$_h$)			R_e (Å)		
System	aVDZ	aVDZ+mb	Ref.	aVDZ	aVDZ+mb	Ref.
He_2	−18.6	−31.5	−35.1	3.1791	3.0330	2.9695
HeNe	−28.3	−59.2	−65.3	3.2949	3.0669	3.0309
HeAr	−46.1	−87.2	−90.8	3.7500	3.5497	3.4804
Ne_2	−46.0	−112.0	−133.9	3.4005	3.1464	3.091
NeAr	−76.5	−190.1	−213.5	3.8268	3.5370	3.4889
Ar_2	−198.4	−393.1	−454.2	4.0626	3.8420	3.7565
Kr_2	−240.7	−544.8	−637.4	4.3959	4.1290	4.008

they decrease the interaction energy. This suggests that they are improving the description of the monomers rather than the interaction, similar to the effect seen with regards to position of the midbond functions earlier; this is also reflected in the relatively diffuse exponents that result (0.005064 and 0.037598, respectively). That these d and f functions present a problem is therefore an artefact of the need to correct for BSSE. In fact, it is the first s and p functions that give the greatest improvement to the description of the interaction. Similar results are seen for the other noble gas dimers, but with the angular momenta incremented by one. The double-zeta quality orbital basis sets being used contain only s and p functions for helium, and s, p, and d functions for neon through krypton. Thus, the additional MBFs beyond these angular momenta appear to essentially be augmenting the orbital basis from double- to triple-zeta polarization quality.

For the alkali metals, the total and interaction energy plots in Figure 2 are essentially identical, with the first s, p, and d functions being the most important, and additional functions contributing essentially nothing. This could either be due to the much larger interaction energies involved as compared to the noble gases, or the somewhat greater separations, as both would encourage optimisation for the interaction region rather than the monomers. The overall choice of configuration then agrees with that for the noble gases, suggesting $[2s2p1d]$ for all except helium, where the d function can be excluded. Strictly, $[1s1p1d]$ would give broadly similar results, but the cost of the additional s and p is minimal. The resulting set is considerably smaller than any previously used midbond functions, allowing for much greater efficiency. All of the optimised exponents can be found in the supplementary material.

Tables 1 and 2 demonstrate that even with such a small set of MBFs, considerable improvements can be achieved. Energies and equilibrium geometries were determined by a sixth-order polynomial fit to seven near-equilibrium points. Across the noble gas diatomics, the interaction

Table 2. Interaction energies (ΔE) and equilibrium bond lengths (R_e) for alkali metal dimers, calculated at the CCSD(T)/awCVDZ level with and without midbond functions. Experimental reference values are shown for comparison.

	ΔE (kcal mol^{-1})			R_e (Å)		
System	awCVDZ	awCVDZ+mb	Expt.	awCVDZ	awCVDZ+mb	Expt.
Na_2	−16.2	−16.9	−17.2[a]	3.1231	3.1031	3.0788[e]
K_2	−11.2	−12.3	−12.7[b]	4.0008	3.9802	3.9244[b]
Rb_2	−10.0	−10.9	−11.4[c]	4.3845	4.3248	4.2099[f]
Cs_2	−8.7	−9.7	−10.4[d]	4.8995	4.8032	4.6499[g]

[a]Ref. [64], [b]Ref. [65], [c]Ref. [66], [d]Ref. [67], [e]Ref. [68], [f]Ref. [69], [g]Ref. [70].

energy is underestimated by on average 57% without midbond functions, as opposed to 10% when they are used. More importantly, the average discrepancy in the position of the equilibrium bond length is 0.3 Å(10%) compared with 0.07 Å(2%), a considerable improvement. For the alkali metals, there is a more modest improvement in the geometries, but there are still substantial gains in the interaction energies of on average 0.9 kcal mol^{-1}. In all instances, the aVDZ+mb results are better than the corresponding aVTZ and aVQZ results, and for the noble gases they are roughly equivalent to aV5Z values. For example, the helium dimer CCSD(T)/aV5Z interaction energy is -31.2 μE_h, compared with -31.5 μE_h for aVDZ+mb. The increase in cost from using the midbond functions is minimal, whereas that from moving from double- to triple-zeta or higher is considerable.

3.1. Transferability

The considerable gains reaped from the addition of such a small set of functions highlight the importance of a robust optimisation. As noted earlier, Figure 1 suggests that the optimisation is highly sensitive, with small changes in exponent leading to large differences in interaction energy. This challenges the assumption that such functions are easily transferable between different systems. Figure 3 shows how the $2p$ exponents change going down groups 1 and 18, and show that in general they become considerably more diffuse. This could be due to the larger sizes of the elements involved, or the larger separations involved - as Tables 1 and 2 show, the separation systematically increases with each period. The lightest elements are more drastically different from the heavier elements, with for example sodium having much larger exponents than potassium, but the latter having broadly similar exponents to rubidium. In addition, the group 1 exponents are consistently more diffuse than those for group 18. Cumulatively, this suggests that there is unlikely to be much transferability between distinct systems. This might

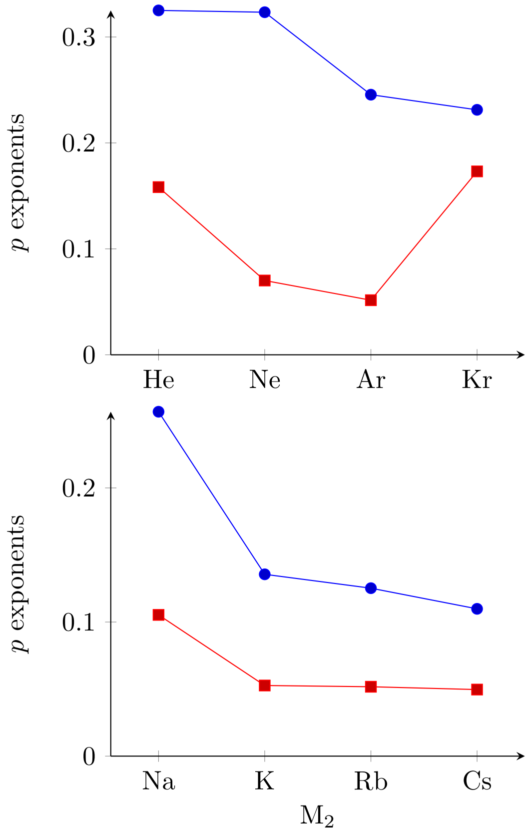

Figure 3. Optimised exponents of the p-type midbond functions for the noble gas (top) and alkali metal (bottom) dimers, showing how the functions, with one exception, become systematically more diffuse going down each group.

Table 3. Interaction energies (μE_h) for heteronuclear diatomics X···Y calculated using CCSD(T)/aVDZ, with varying sets of midbond functions: the MBFs from X_2, Y_2, the average of the two paired in ascending order), and MBFs optimised specifically for the system.

X	Y	mb(X)	mb(Y)	mb(avg.)	mb(opt.)
He	Ne	−54.1	−56.9	−57.5	−59.2
He	Ar	−79.5	−85.4	−86.2	−87.2
Ne	Ar	−177.9	−182.1	−184.3	−190.1

be a result of using a very compact set of functions, as opposed to the larger, more complete sets used in previous studies.

Further evidence is given in Table 3, where the MBFs from the noble gas homodimers are used in calculations on heterodimers. In general, it can be seen that the functions of the lighter element perform poorly compared to the set optimised specifically for that system (a roughly 10% discrepancy). Improvements can be found by using the functions from the heavier element, or better still, by taking the arithmetic mean of the two; this is done by placing each set of exponents in ascending order and taking pairwise averages. The result is still not as good as a fully optimised set, but the difference is only approximately 3% of the interaction energy. As it would be impractical to optimise new exponents for every system, this provides a useful way of applying the midbond functions to new complexes.

3.2. Explictly correlated methods

Convergence towards the complete basis set limit can be accelerated greatly by using explicitly correlated methods, but they have not been widely applied to the weakly-bound, long range interactions considered here. Patkowski[31] compared the efficacy of CCSD(T)-F12b to canonical coupled cluster with added midbond functions, and found that for smaller basis sets and weaker interactions, the latter gives better results, whereas for larger basis sets and stronger interactions, explicit correlation offers the biggest improvements. However, the midbond functions used were just those of a hydrogen atom in the orbital basis placed at the midpoint, so that auxiliary basis sets were automatically available. As discussed above, the results are sensitive to the choice of exponents, and as can be seen from Figure 3 and the supplementary material, the exponents of the MBFs are quite different to those of the orbital basis of a hydrogen atom; specifically, the former are considerably more diffuse, as would be expected given the delocalised nature of the interaction.

Figure 4 shows the incremental effects of adding midbond functions for two of the systems studied by Patkowski, $H_2O\cdots He$ and $H_2O\cdots CH_4$. The geometries and CBS limit results for these were taken from Ref. [31]. As before, a sensible configuration appears to be $[2s2p1d]$, which in both cases gives considerable improvement on both CCSD(T)/aVDZ and CCSD(T)-F12b/aVDZ results. Table 4 gives details for two further systems, the water dimer and $H_2O\cdots Ar$. For all complexes except the water dimer, the addition of midbond functions yields greater improvements than using explicit correlation. For the two weaker interactions, involving noble gases, CCSD(T)/aVDZ+mb is essentially as good as CCSD(T)-F12b/aVTZ. It should be noted that the explicitly correlated calculations are considerably more expensive than their canonical equivalents, even for the same basis set size, due to the numerous additional integrals the former requires. Especially striking is the result for the water-methane complex, where Patkowski concluded that explicit correlation performs considerably better than midbond functions. The optimised MBFs used here, however, give an interaction energy of $-1483.1\ \mu\mathrm{E}_h$, compared to $-1307.7\ \mu\mathrm{E}_h$ in the aformentioned paper. This again emphasises the importance of a robust optimisation and the relative lack of transferability.

For the water dimer, which is by far the strongest interaction of the four in Table 4 and the one with the shortest separation, explicit correlation is considerably more effective than the addition of midbond functions. This is perhaps unsurprising, as the orbital basis is better

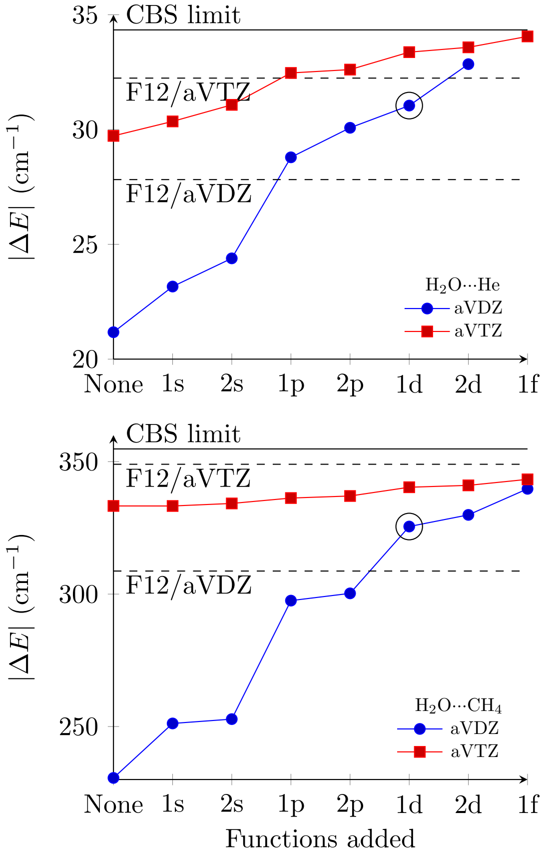

Figure 4. Incremental CCSD(T) absolute interaction energy ($|\Delta E|$) contributions from adding midbond functions in the systems $H_2O{\cdots}He$ (top) and $H_2O{\cdots}CH_4$ (bottom), using the aVDZ and aVTZ orbital basis sets. The CCSD(T)-F12b results for the same sets and in the CBS limit are shown, taken from Ref. [31], and the results for the recommended midbond configurations are circled.

Table 4. A comparison of interaction energies (in μE_h) calculated with midbond functions or explicit correlation, or both. Complete basis set (CBS) limits are taken from Ref. [31].

		No midbond		With midbond		
System	Method	aVDZ	aVTZ	aVDZ	aVTZ	CBS
$H_2O{\cdots}Ar$	CCSD(T)	−362.4	−552.4	−584.9	−618.3	−635.7
	CCSD(T)-F12b	−504.6	−601.6	−629.5	−632.2	
$H_2O{\cdots}CH_4$	CCSD(T)	−1050.7	−1518.5	−1483.1	−1550.7	−1616.6
	CCSD(T)-F12b	−1406.5	−1590.2	−1539.8	−1592.7	
$(H_2O)_2$	CCSD(T)	−6931.2	−7565.9	−7411.2	−7723.2	−7950.8
	CCSD(T)-F12b	−7648.3	−7876.1	−7791.1	−7880.8	
$H_2O{\cdots}He$	CCSD(T)	−96.5	−135.4	−141.4	−152.0	−156.5
	CCSD(T)-F12b	−126.4	−146.9	−148.9	−153.3	

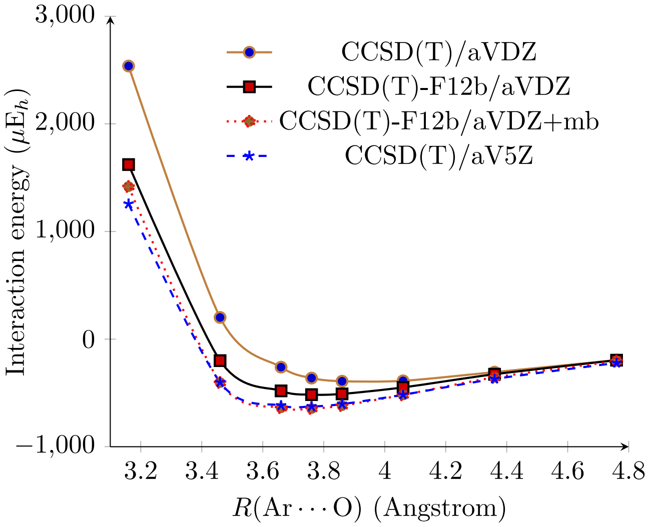

Figure 5. Interaction energy curves for $H_2O\cdots Ar$ as a function of the oxygen-argon separation, calculated with and without additional midbond functions, or using explicitly correlated methods. Using midbond functions with aVDZ gives results almost indistinguishable from the aV5Z, large basis set limit.

Table 5. Errors in the interaction energy (nE_h) solely from the auxiliary basis sets for the midbond functions at the CCSD(T)-F12b/VDZ-F12 level, relative to results using a large, even-tempered reference ABS. The same auxiliary bases were used for all systems.

System	JKFit	MP2Fit	OptRI	Total %
He_2	65.26	0.11	0.88	0.22
Ne_2	48.35	1.33	1.93	0.05
Ar_2	−166.73	0.08	2.79	0.05
Na_2	4.52	0.003	9.45	0.00
$H_2O\cdots Ar$	41.66	33.16	387.03	0.07
$H_2O\cdots CH_4$	1.68	−2.71	223.12	0.01
$(H_2O)_2$	31.48	−42.34	−338.09	0.01
$H_2O\cdots He$	32.04	430.27	18.32	0.32

able to describe a more tightly-bound system. Without exception, however, the combination of midbond functions with CCSD(T)-F12b gives the results closest to the CBS limit, and this improvement is noticeable even for the water dimer. Figure 5 further demonstrates this for the case of $H_2O\cdots Ar$, showing that both approaches lead to improvements not only at the minimum but across the whole potential energy curve. In particular, CCSD(T)/aVDZ can be seen to severely overestimate the equilibrium bond distance as compared to the aV5Z result, while all other methods agree well with the latter. In fact, the CCSD(T)-F12b/aVDZ+mb curve is essentially indistinguishable from the reference, despite being roughly a thousand times faster (as determined from wall times using MOLPRO on a single processor, averaged across all geometries).

The complication with using midbond functions with F12 methods is, as already noted, the need for three auxiliary basis sets. This increases the difficulty of developing system-specific

MBFs considerably. However, while the interaction energies are sensitive to changes in the exponents of the midbond functions themselves, the overall range of exponents is very similar across all the sets given here. The purpose of the auxiliary sets is to span the orthogonal complement of the space spanned by the orbital (or in this case, midbond) basis functions, and the closeness of the exponents means that this complement should be fairly consistent. In keeping with the analogy between the choice of $[2s2p1d]$ for the MBFs and the use of a double-zeta basis, the configurations of the ABSs were chosen to match those for the VDZ-F12 equivalents for a first-row atom: $[11s8p6d3f]$ (JKFit)[56]; $[7s5p4d2f]$ (MP2Fit)[59]; and $[5s6p4d3f1g]$ (OptRI)[71, 72]. These sets can be found in the supplementary material. Table 5 shows the errors due to using these ABSs compared to the aV5Z auxiliary sets for carbon. For these calculations, the VDZ-F12 orbital basis sets[73] were used for the molecules with the matching JKFit[56], MP2Fit[59], and recent OptRI+ auxiliary sets[72], demonstrating that while the MBFs are system specific, they are not basis specific. Despite the same auxiliary sets being used for all the midbond functions, all of the absolute errors were less than 1 $\mu\mathrm{E}_h$ and the overall percentage errors were on average less than 0.1%. The largest errors were to be found in the OptRI sets for the more strongly bound systems at the bottom of Table 5. This is in part due to the use of the OptRI basis in the complementary auxiliary basis set (CABS) singles correction to the Hartree-Fock energy, which directly affects the energy rather than indirectly through the accuracy of the density-fitted integrals, as can be seen in the supplementary material by the decrease in error when the CABS correction is removed. The general applicability of the ABSs demonstrated should make the use of midbond functions in F12 calculations on other systems considerably easier.

4. Conclusions

We have presented a series of small and efficient sets of midbond basis functions for use in high accuracy calculations on complexes with long-range interactions. While these are considerably smaller than previously used sets, they they perform at least as well, converging calculations on dimers of alkali metals and noble gases to very near the complete basis set limit using only double-zeta orbitals basis sets. Considerable improvements are seen both in terms of the interaction energies and the geometries, with negligible additional computational cost. In the case of the noble gas dimers, aV5Z quality results are obtainable three orders of magnitude faster. The compact nature of the combined atom-centred and midbond basis set makes it feasible to perform very high accuracy calculations on weakly bound systems, particularly interesting in the context of understanding experiments on ultracold molecules[60], which would otherwise be prohibitively expensive. Constraining the positions to be exactly halfway between the interacting substituents, and the angular momenta to be small, simplifies the process of developing and using midbond functions in other systems; care needs to be taken in ensuring that it is the description of the interaction, and not of one of the monomers, that is being improved. The price of using such compact sets of functions is that they are system specific, and the interaction energies are highly sensitive to small changes in the exponents. However, reasonable results can be achieved when transferring between chemically similar contexts, for example in the heteronuclear group 18 diatomics where using the average of the two sets of exponents from the relevant homonuclear systems gives accuracy similar to a fully optimised set. In addition, this sensitivity can be tempered by the use of more robust optimisation procedures using analytical gradients of the energy with respect to the exponents, the expense of which is mitigated by the relatively small number of functions and the efficiency of the procedure.

Convergence towards the basis set limit can be further accelerated by combining these sets of MBFs with explicitly correlated techniques, especially in more strongly bound systems. The quality of the functions do not show great dependence on the orbital basis being used. This is likely due to the fact that the exponents are all in a broadly similar range, roughly between 0.05 and 0.5. The similarities also mean that the auxiliary basis sets necessary for F12 calculations do not differ greatly, such that the same ABSs can be used for each of the sets of midbond functions with minimal errors. This expedites the use of such functions in future explicitly correlated studies on weakly bound systems, again allowing for high accuracy to be achieved efficiently.

The basis sets developed in this work are provided in the supplementary material and will be made available to download from the correlation consistent basis set repository.[74]

Supplementary material

See the supplementary material for the MP2 analytic gradients, additional data on the ABS errors, and the optimised midbond and auxiliary basis sets in MOLPRO format.

Acknowledgements

The authors thank the Engineering and Physical Sciences Research Council (UK) for a postgraduate studentship awarded to R.A.S. and project funding (No. EP/N02253X/1).

References

[1] J.G. Hill, Int. J. Quantum Chem. **113** (1), 21–34 (2013).

[2] B. Nagy and F. Jensen, in *Reviews in Computational Chemistry, Volume 30*, edited by Abby L. Parrill and Kenny B. Lipkowitz (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2017), pp. 93–149.

[3] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen and A.K. Wilson, Chem. Phys. Lett. **286** (3-4), 243–252 (1998).

[4] A. Halkier, W. Klopper, T. Helgaker, P. Jørgensen and P.R. Taylor, J. Chem. Phys. **111** (20), 9157–9167 (1999).

[5] K. Raghavachari, G.W. Trucks, J.A. Pople and M. Head-Gordon, Chem. Phys. Lett. **157** (6), 479–483 (1989).

[6] J. Witte, J.B. Neaton and M. Head-Gordon, J. Chem. Phys. **144** (19), 194306 (2016).

[7] A. Forni, S. Pieraccini, S. Rendine and M. Sironi, J. Comput. Chem. **35** (5), 386–394 (2014).

[8] S. Kozuch and J.M.L. Martin, J. Chem. Theory Comput. **9** (4), 1918–1931 (2013).

[9] S. Grimme, Wiley Interdiscip. Rev. Comput. Mol. Sci. **1** (2), 211–228 (2011).

[10] J. Hermann, R.A. DiStasio and A. Tkatchenko, Chem. Rev. **117** (6), 4714–4758 (2017).

[11] T. Kato, Comm. Pure Appl. Math. **10** (2), 151–177 (1957).

[12] W. Klopper, F.R. Manby, S. Ten-No and E.F. Valeev, Int. Rev. Phys. Chem. **25** (3), 427–468 (2006).

[13] L. Kong, F.A. Bischoff and E.F. Valeev, Chem. Rev. **112**, 75–107 (2012).

[14] C. Hättig, W. Klopper, A. Köhn and D.P. Tew, Chem. Rev. **112** (1), 4–74 (2012).

[15] G. Knizia, T.B. Adler and H.J. Werner, J. Chem. Phys. **130** (5), 054104 (2009).

[16] O. Marchetti and H.J. Werner, J. Phys. Chem. A **113** (43), 11580–11585 (2009).

[17] J.L. Whitten, J. Chem. Phys. **58** (10), 4496–4501 (1973).

[18] W. Klopper and C.C.M. Samson, J. Chem. Phys. **116** (15), 6397–6410 (2002).

[19] H.J. Werner, F.R. Manby and P.J. Knowles, J. Chem. Phys. **118** (18), 8149–8160 (2003).

[20] J.C. Womack and F.R. Manby, J. Chem. Phys. **140** (4), 044118 (2014).

[21] K. Patkowski, J. Chem. Phys. **137** (3), 034103 (2012).

[22] A.J.C. Varandas, J. Chem. Phys. **113** (20), 8880–8887 (2000).

[23] J.G. Hill, K.A. Peterson, G. Knizia and H.J. Werner, J. Chem. Phys. **131** (2009), 194105 (2009).

[24] D. Feller, K.A. Peterson and J.G. Hill, J. Chem. Phys. **135** (4), 044102 (2011).

[25] F. Tao and Y. Pan, J. Chem. Phys. **97** (7), 4989–4995 (1992).

[26] F. Tao, J. Chem. Phys. **98** (4), 3049–3059 (1993).

[27] J. van de Bovenkamp and F.B. van Duijneveldt, J. Chem. Phys. **110** (23), 11141–11151 (1999).

[28] G. Tasi and A.G. Császár, Chem. Phys. Lett. **438** (1-3), 139–143 (2007).

[29] H. Partridge and C.W. Bauschlicher, Mol. Phys. **96** (4), 705–710 (1999).

[30] D. Mester, J. Csontos and M. Kállay, Theor. Chem. Acc. **134** (6), 74 (2015).

[31] K. Patkowski, J. Chem. Phys. **138** (15), 154101 (2013).

[32] J.S. Wright and V.J. Barclay, J. Comput. Chem. **12** (6), 697–704 (1991).

[33] S. Boys and F. Bernardi, Mol. Phys. **19** (4), 553–566 (1970).

[34] E. Jakubikova, A.K. Rappé and E.R. Bernstein, J. Phys. Chem. A **110** (31), 9529–9541 (2006).

[35] L.K. McKemmish, A.T.B. Gilbert and P.M.W. Gill, J. Chem. Theory Comput. **10** (10), 4369–4376 (2014).

[36] P. Čársky, R. Čurík and Š. Varga, J. Chem. Phys. **136** (11), 114105 (2012).

[37] J. Hutter, M. Iannuzzi, F. Schiffmann and J. VandeVondele, Wiley Interdiscip. Rev. Comput. Mol. Sci. **4** (1), 15–25 (2014).

[38] G. Lippert, J. Hutter and M. Parrinello, Theor. Chim. Acta. **103** (2), 124–140 (1999).

[39] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, 2nd ed. (Camridge University Press, Cambridge, 1992).

[40] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. (Camridge University Press, Cambridge, 2004).

[41] J. Gauss and D. Cremer, in *Advances in Quantum Chemistry, Volume 23*, pp. 205–299.

[42] J. Gerratt and I.M. Mills, J. Chem. Phys. **49** (4), 1719–1729 (1968).

[43] J.A. Pople, R. Krishnan, H.B. Schlegel and J.S. Binkley, Int. J. Quantum Chem. **16** (S13), 225–241 (1979).

[44] P. Pulay, Wiley Interdiscip. Rev. Comput. Mol. Sci. **4** (3), 169–181 (2014).

[45] T.H. Dunning Jr., J. Chem. Phys. **90** (2), 1007–1023 (1989).

[46] R.A. Kendall, T.H. Dunning Jr. and R.J. Harrison, J. Chem. Phys. **96** (9), 6796 (1992).

[47] D.E. Woon and T.H. Dunning Jr., J. Chem. Phys. **98** (2), 1358–1371 (1993).

[48] A.K. Wilson, D.E. Woon, K.A. Peterson and T.H. Dunning Jr., J. Chem. Phys. **110** (16), 7667–7676 (1999).

[49] B.P. Prascher, D.E. Woon, K.A. Peterson, T.H. Dunning Jr. and A.K. Wilson, Theor. Chem. Acc. **128** (1), 69–82 (2011).

[50] J.G. Hill and K.A. Peterson, J. Chem. Phys. **147**, 244106 (2017).

[51] I.S. Lim, P. Schwerdtfeger, B. Metz and H. Stoll, J. Chem. Phys. **122** (10), 104103 (2005).

[52] C. Hättig, D.P. Tew and A. Köhn, J. Chem. Phys. **132** (23), 231102 (2010).

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

[54] S. Ten-No, Chem. Phys. Lett. **398** (1), 56–61 (2004).

[55] K.E. Yousaf and K.A. Peterson, Chem. Phys. Lett. **476** (4-6), 303–307 (2009).

[56] F. Weigend, Phys. Chem. Chem. Phys. **4** (18), 4285–4291 (2002).

[57] F. Weigend, Phys. Chem. Chem. Phys. **8** (9), 1057–1065 (2006).

[58] F. Weigend, J. Comput. Chem. **29** (2), 167–175 (2008).

[59] F. Weigend, A. Köhn and C. Hättig, J. Chem. Phys. **116** (8), 3175–3183 (2002).

[60] J.N. Byrd, R. Côté and J.A. Montgomery, J. Chem. Phys. **135** (24), 244307 (2011).

[61] M. Pawlak, Y. Shagam, A. Klein, E. Narevicius and N. Moiseyev, J. Phys. Chem. A **121** (10), 2194–2198 (2017).

[62] J. Ogilvie and F.Y. Wang, J. Mol. Struct. **273**, 277–290 (1992).

[63] J. Ogilvie and F.Y. Wang, J. Mol. Struct. **291** (2-3), 313–322 (1993).

[64] K.M. Jones, S. Maleki, S. Bize, P.D. Lett, C.J. Williams, H. Richling, H. Knöckel, E. Tiemann, H. Wang, P.L. Gould and W.C. Stwalley, Phys. Rev. A **54** (2), R1006–R1009 (1996).

[65] A. Pashov, P. Popov, H. Knöckel and E. Tiemann, Eur. Phys. J. D **46** (2), 241–249 (2008).

[66] J.Y. Seto, R.J. Le Roy, J. Vergès and C. Amiot, J. Chem. Phys. **113** (8), 3067–3076 (2000).

[67] C. Amiot and O. Dulieu, J. Chem. Phys. **117** (11), 5155–5164 (2002).

[68] K.P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure. IV. Constants of Diatomic Molecules* (Springer-Verlag, New York, 2013).

[69] C. Amiot, P. Crozet and J. Vergés, Chem. Phys. Lett. **121** (4), 390–394 (1985).

[70] G. Höning, M. Czajkowski, M. Stock and W. Demtröder, J. Chem. Phys. **71** (5), 2138–2149 (1979).

[71] K.E. Yousaf and K.A. Peterson, J. Chem. Phys. **129** (18), 184108 (2008).

[72] R.A. Shaw and J.G. Hill, J. Chem. Theory Comput. **13** (4), 1691–1698 (2017).

[73] K.A. Peterson, T.B. Adler and H.J. Werner, J. Chem. Phys. **128** (8), 084102 (2008).

[74] Correlation consistent basis set repository. http://bit.ly/ccBasis. Accessed: 2017-10-26.