



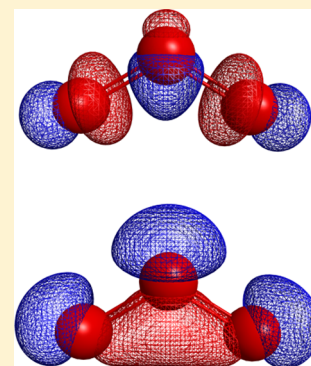
Optimal Composition of Atomic Orbital Basis Sets for Recovering Static Correlation Energies

Andrew J. Wallace and Deborah L. Crittenden*

Department of Chemistry, University of Canterbury, Christchurch, New Zealand

Supporting Information

ABSTRACT: Static correlation energies (E_{stat}) are calculated in a range of basis sets for a chemically diverse collection of atoms and molecules. The reliability of a basis set in capturing E_{stat} is assessed according to the following: mean and maximum absolute deviations from near-exact E_{stat} estimates, monotonic convergence to the complete basis set limit, and ability to capture E_{stat} accurately independent of changes in geometry, molecular size, and electronic configuration. Within the polarization and correlation-consistent basis set series, triple- ζ basis sets are the smallest that can reliably capture E_{stat} . The cc-pVTZ basis set performs particularly well, recovering E_{stat} to chemical accuracy for all atoms and molecules in our data set. A series of customized basis sets are constructed by stripping polarization functions from, and swapping polarization functions among, existing basis sets. Basis sets without polarization functions are incapable of accurately recovering E_{stat} . Basis sets with a near-complete set of s , p , and d functions can approach chemical accuracy in maximum absolute error. However, this may be achieved at lower computational cost by using a well balanced triple- ζ basis set including f functions, along with a smaller number of s , p , and d functions. Recommended basis sets for calculating E_{stat} with increasing accuracy at increasing computational cost are 6-311G(2d,2p), cc-pVTZ, and cc-pVQZ stripped of g functions.



INTRODUCTION

The correlation energy of a chemical system is defined¹ as the difference between its exact electronic energy and its numerically exact Hartree–Fock energy:

$$E_c = E^{\text{exact}} - E^{\text{HF}} \quad (1)$$

In practice, it is not generally possible to calculate E^{exact} exactly, but highly accurate approximations of E^{exact} may be obtained via full configuration interaction expansion (FCI) in a near-complete basis set (CBS).^{2–4} It is well-known that the computational cost of full CI energy calculations scales factorially with basis size,³ whereas the energy converges slowly to the CBS limit.⁵ Therefore, thermochemically accurate FCI/CBS calculations are only feasible for systems with up to three atoms from the first and second rows of the periodic table.^{6–9} Truncation of the CI expansion by excitation level where appropriate helps decrease computational cost, but it remains challenging to calculate chemically accurate total energies for molecules with more than five or six non-hydrogen atoms.¹⁰

The slow convergence of the FCI energy with respect to basis size may be traced back to the difficulties associated with capturing the interelectronic cusps in the true wave function in a basis of one-electron Slater determinants.⁵ Therefore, it is useful to consider partitioning the total correlation energy into “static” and “dynamic” components:

$$E_c = E_{\text{stat}} + E_{\text{dyn}} \quad (2)$$

where the static contribution arises from near-degeneracies between occupied and unoccupied orbitals, and the dynamic

term accounts for changes in short-range interactions due to the presence of interelectronic cusps in the true wave function that are poorly approximated by Hartree–Fock theory. Wave functions which capture the static correlation energy (e.g., CASSCF)^{11–13} are constructed quite differently to those designed primarily to efficiently recover the dynamic correlation energy, for example, R12-based schemes,^{14,15} and variational Monte Carlo wave functions incorporating Jastrow factors.¹⁶

Following a large body of work in the scientific literature,^{11–13,17–21} we define the static correlation energy as the exact correlation energy of the valence electrons in an optimal minimal molecular orbital basis:

$$E_{\text{stat}} = E^{\text{FCI(val)}/\text{MBS(molecular)}} - E^{\text{R(O)HF}/\text{MBS(molecular)}} \quad (3)$$

A number of different ways of defining and calculating E_{stat} have been proposed,^{22–28} of which the most general and widely used is:

$$E_{\text{stat}} = E^{\text{CASSCF(val)}/\text{CBS}} - E^{\text{R(O)HF}/\text{CBS}} \quad (4)$$

The factorial computational scaling of CASSCF limits its applicability to systems of up to 14 valence electrons in 14 valence molecular orbitals, regardless of the quality of the underlying atomic orbital basis. This restriction can be overcome by resorting to more approximate E_{stat} models, for

Received: January 20, 2014

Revised: February 17, 2014

Published: February 19, 2014



Table 1. Truncation Schemes for Polarization Consistent Basis Sets

name	description
pc-0[sp] = pc-0	unmodified pc-0 basis
pc-0[spd ₂]	pc-0 basis, plus <i>d</i> functions from pc-2 basis
pc-1[sp]	minimal pc-1 basis, stripped of all polarization functions
pc-1[spd] = pc-1	unmodified pc-1 basis
pc-1[spd ₂]	minimal pc-1 basis, plus <i>d</i> functions from pc-2 basis
pc-1[spd ₃]	minimal pc-1 basis, plus <i>d</i> functions from pc-3 basis
pc-2[sp]	minimal pc-2 basis, stripped of all polarization functions
pc-2[spd ₁]	minimal pc-2 basis, plus <i>d</i> functions from pc-1 basis
pc-2[spd]	pc-2 basis, stripped of <i>f</i> functions
pc-2[spd ₃]	minimal pc-2 basis, plus <i>d</i> functions from pc-3 basis
pc-2[spdf] = pc-2	unmodified pc-2 basis
pc-3[sp]	minimal pc-3 basis, stripped of all polarization functions
pc-3[spd ₂]	minimal pc-3 basis, plus <i>d</i> functions from pc-2 basis
pc-3[spd]	pc-3 basis, stripped of <i>f</i> and <i>g</i> functions
pc-3[spdf]	pc-3 basis, stripped of <i>g</i> functions
pc-3[spdfg] = pc-3	unmodified pc-3 basis
pc-4[sp]	minimal pc-4 basis, stripped of all polarization functions
pc-4[spd]	pc-4 basis, stripped of <i>f</i> and <i>g</i> and <i>h</i> functions
pc-4[spdf]	pc-4 basis, stripped of <i>g</i> and <i>h</i> functions
pc-4[spdfg] ≈ pc-4	pc-4 basis, stripped of <i>h</i> functions

example, valence orbital optimized coupled cluster doubles (VOO–CCD),²² the spin-flip method,²³ or full CI calculations within valence molecular orbitals constructed from a minimal molecule polarized atomic orbital basis.²⁸ These approximate E_{stat} models are expected to converge toward the complete basis set limit at the same rate as those calculated according to eq 4.

Irrespective of the E_{stat} model, substantial computational savings may be realized by first evaluating E_{stat} in a moderately sized but sufficiently large one-electron basis and then recovering E_{dyn} using explicitly correlated methods, incorporating functions of the interelectronic distance to reconstruct the wave function cusps. The above insight has motivated the recent development of a number of promising multi reference explicitly correlated methods (MR-R12)¹⁵ such as MRMP2-F12,²⁹ CASSCF[2]_{R12},³⁰ CASPT2-F12,³¹ MRCI-F12,³² and geminal-CASSCF.³³

In principle, it is also possible to recover E_{dyn} using density functional theory but, in practice,^{18,24,34–36} it is hard to derive and/or parametrize such functionals to avoid “double-counting” (i.e., unintentionally recovering some or all of E_{stat} concurrently with E_{dyn}).

Initial applications suggest that MR-R12 methods are so successful in capturing E_{dyn} that the largest remaining source of error in the total energy lies in the incompleteness of the one-electron basis used to calculate E^{HF} and E_{stat} .¹⁵

It is well-established that Hartree–Fock energies converge more rapidly to the complete basis set limit than exact energies. For the pc-*n* basis series, the leading term in the basis set truncation error decays exponentially with the number of primitive *s* functions,^{37,38}

$$E_n^{\text{HF}} = E_{\infty}^{\text{HF}} + A \exp(-B\sqrt{n_s}) \quad (5)$$

These empirical results are consistent with a recent theoretical analysis which demonstrates root-exponential convergence of the Hartree–Fock energy of atomic orbitals expanded in an even-tempered Gaussian basis:³⁹

$$E_{n_s}^{\text{HF}} = E_{\infty}^{\text{HF}} + A \exp(-\pi\sqrt{3n_s}) \quad (6)$$

A similar relationship has been observed for the cc-pVnZ basis series,⁴⁰

$$E_n^{\text{HF}} = E_{\infty}^{\text{HF}} + A \exp(-B\sqrt{n}) \quad (7)$$

where *n* is the cardinality of the basis set. For correlation-consistent basis sets, *n* gives the number of contracted Gaussian-type functions describing the valence orbitals, and the identity of the highest angular momentum functions present in the basis for second and third row atoms (*n* = *L*_{max} for atoms Li – Ar). The number of primitive Gaussians comprising each contracted valence orbital function also increases with *n*. Therefore, *n* is a general measure of basis set quality that encompasses the overall ability of the basis to recover the electron–nucleus cusps in the wave function, provide spatial flexibility in molecular orbitals, and capture atomic polarization effects within a molecular environment.

Despite this rapid convergence behavior, obtaining Hartree–Fock total energies to chemical accuracy typically requires basis sets of at least quadruple- ζ quality. To reliably recover chemically accurate HF energies, pentuple- ζ basis sets must be used. Fortunately, a range of highly accurate numerical approximations are available to reduce the complexity of this task, such as resolution of the identity, resolution of the operator, pseudospectral, and Cholesky decomposition.¹⁰ This makes large basis Hartree–Fock calculations feasible for most organic molecules of practical interest.

Basis set incompleteness errors in E^{HF} arise predominantly from the description of the chemically uninteresting nucleus–electron cusps in the wave function.³⁹ Relative energies (e.g., reaction enthalpies, atomization, and ionization energies) provide a much more chemically relevant metric of basis set convergence and typically only require double or triple- ζ basis sets to achieve chemical accuracy, as cusp errors cancel in these cases.

Having accounted or controlled for basis set incompleteness errors in E^{HF} , the largest potential source of remaining error in E^{exact} is basis set incompleteness error in E_{stat} . Much less is known about the basis set convergence behavior of E_{stat} than E^{HF} , although it has been empirically established that CASSCF

Table 2. Truncation Schemes for Correlation-Consistent Basis Sets

name	description
cc-pVDZ[<i>sp</i>]	minimal cc-pVDZ basis, stripped of all polarization functions
cc-pVDZ[<i>spd</i>] = cc-pVDZ	unmodified cc-pV(D+d)Z basis
cc-pVTZ[<i>sp</i>]	minimal cc-pVTZ basis, stripped of all polarization functions
cc-pVTZ[<i>spd</i>]	cc-pV(T+d)Z basis, stripped of <i>f</i> functions
cc-pVTZ[<i>spdf</i>] = cc-pVTZ	unmodified cc-pV(T+d)Z basis
cc-pVQZ[<i>sp</i>]	minimal cc-pVQZ basis, stripped of all polarization functions
cc-pVQZ[<i>spd</i>]	cc-pV(Q+d)Z basis, stripped of <i>f</i> and <i>g</i> functions
cc-pVQZ[<i>spdf</i>]	cc-pV(Q+d)Z basis, stripped of <i>g</i> functions
cc-pVQZ[<i>spdfg</i>] = cc-pVQZ	unmodified cc-pV(Q+d)Z basis
cc-pVSZ[<i>sp</i>]	minimal cc-pVSZ basis, stripped of all polarization functions
cc-pVSZ[<i>spd</i>]	cc-pV(5+d)Z basis, stripped of <i>f</i> and <i>g</i> and <i>h</i> functions
cc-pVSZ[<i>spdf</i>]	cc-pV(5+d)Z basis, stripped of <i>g</i> and <i>h</i> functions
cc-pVSZ[<i>spdfg</i>] ≈ cc-pVSZ	cc-pV(5+d)Z basis, stripped of <i>h</i> functions

energies also converge toward the CBS limit with the same root-exponential dependence on the number of primitive basis functions as Hartree–Fock.⁴¹

Intuition suggests that it should be possible to recover E_{stat} to chemical accuracy employing only small- to medium-sized atomic orbital basis sets. First, E_{stat} is much smaller in magnitude than E^{HF} or $E^{\text{CASSCF(val)}}$ so the same order of relative error will correspond to a much smaller absolute error. Second, static correlation energies arise from near-degeneracies between occupied and unoccupied orbitals, which are expected to be only weakly sensitive to the completeness of the underlying atomic orbital basis.

The aims of this paper, arising from the observations and hypotheses above, are the following: (1) determine the optimal size and composition of atomic orbital basis sets for recovering E_{stat} , that is, establishing the requirements for obtaining E_{stat} to chemical accuracy with minimal computational cost; (2) elucidate the convergence behavior of static correlation energies; and (3) establish thermochemically accurate static correlation energies for a chemically diverse range of atoms and molecules.

METHODS

To investigate the sensitivity of E_{stat} to basis set composition, a range of truncated polarization-consistent basis sets^{37,42,43} and correlation-consistent basis sets^{44–47} were constructed as detailed in Tables 1 and 2. Modifications to polarization functions are described for Li–Ar. Equivalent modifications were performed for H and He involving polarization functions one unit lower in angular momentum.

Hartree–Fock and full valence CASSCF calculations were carried out in the GAMESS suite of quantum chemical software⁴⁸ for all atoms and molecules in Pople's G1 data set,⁴⁹ plus all reactants, products, and transition states in the DBH reaction database,⁵⁰ and the notoriously multireference ozone molecule. This collection contains 82 systems with nonzero E_{stat} and will be referred to hereafter as our data set.

Static correlation energies were also calculated for a series of dissociating diatomics (H_2 , N_2 , and F_2) to further confirm the ability of each basis set to accurately recover E_{stat} during bonding breaking and formation.

For comparison, the same calculations were also done using a selection of Pople basis sets: STO-3G,^{51,52} 6-31G,^{53–55} 6-31G(*d*, *p*),^{53–55} 6-31G(2*d*, 2*p*),^{53–56} 6-311G,^{57,58} 6-311G(*d*, *p*),^{57,58} 6-311G(2*d*, 2*p*),^{56–58} and 6-311G(2*df*, 2*pd*).^{56–58}

RESULTS AND DISCUSSION

Finite Basis Estimates of E_{stat} . Static correlation energies calculated in the near-complete cc-pVSZ basis set provide thermochemically accurate correlation energies against which the performance of smaller basis sets are assessed:

$$\Delta E_{\text{stat}}^Z = E_{\text{stat}}^{\text{cc-pVSZ}} - E_{\text{stat}}^Z \quad (8)$$

where Z denotes the smaller basis set.

The ability of each basis set to recover E_{stat} is measured by its mean and maximum absolute values of ΔE_{stat} across the data set. The size of each basis set is quantified by the number of primitive basis functions describing the carbon atom ($n_{\text{prim}}^{\text{C}}$). The carbon atom has been chosen as a prototype to illustrate the relative sizes of different basis sets, as it is the one of the most common constituents of the molecules within our data set and intermediate in atomic size. Mean and maximum absolute ΔE_{stat} and $n_{\text{prim}}^{\text{C}}$ for all non-customized basis sets are given in Table 3. All raw data are available as Supporting Information.

From Table 3, we observe that split-valence basis sets without polarization functions (pc-0, 6-31G, 6-311G) cannot accurately capture the static correlation energy, regardless of the number of *s* and *p* functions. These basis sets produce E_{stat}

Table 3. Mean and Maximum Absolute Errors Across Our Data Set (mE_{h}), Plus Number of Primitive Functions Describing the Carbon Atom for Each Basis Set

Z	mean $ \Delta E_{\text{stat}}^Z $	max $ \Delta E_{\text{stat}}^Z $	$n_{\text{prim}}^{\text{C}}$
pc-0	5.8	34.9	16
pc-1	1.2	7.3	28
pc-2	0.6	2.9	51
pc-3	0.1	0.3	92
pc-4	0.04	0.2	129
cc-pVDZ	1.2	6.4	34
cc-pVTZ	0.3	1.3	50
cc-pVQZ	0.04	0.3	77
cc-pVSZ	0	0	107
STO-3G	11.4	92.2	15
6-31G	6.1	40.9	22
6-311G	6.5	41.7	26
6-31G(<i>d</i> , <i>p</i>)	1.3	7.5	27
6-311G(<i>d</i> , <i>p</i>)	1.2	10.3	31
6-31G(2 <i>d</i> , 2 <i>p</i>)	1.0	5.0	32
6-311G(2 <i>d</i> , 2 <i>p</i>)	0.6	4.1	36
6-311G(2 <i>df</i> , 2 <i>pd</i>)	0.4	2.6	43

Table 4. Thermochemically Accurate cc-pVSZ E_{stat} Values and Basis Set Truncation Errors^a

	E_{stat}					ΔE_{stat}			
	cc-pVSZ	pc-2	cc-pVTZ	6-311G (2df, 2pd)		cc-pVSZ	pc-2	cc-pVTZ	6-311G (2df, 2pd)
Be	43.8	−1.3	−0.2	−0.1	Mg	31.4	−0.6	−0.1	−0.1
B	34.7	−0.9	−0.3	−0.2	Al	24.2	−0.4	−0.2	−0.1
C	19.3	−0.5	−0.2	−0.2	Si	13.4	−0.2	−0.1	0.0
H ₂	18.5	0.0	−0.1	−0.1	Li ₂ ^b	8.8	0.5	0.0	0.0
LiH ^b	16.4	−0.2	−0.1	−0.1	Na ₂ ^b	10.6	0.0	0.1	0.0
LiF ^b	15.0	−0.3	−0.2	−0.2	NaCl ^b	8.4	0.0	−0.2	−0.5
BeH	27.7	−1.3	−0.3	−0.4	CH	43.1	−0.8	−0.2	−0.2
NH	26.1	−0.2	−0.1	0.0	OH	25.0	−0.2	−0.1	0.0
FH	24.4	−0.2	−0.1	0.1	ClH	17.4	−0.1	−0.1	0.0
CH ₂ (³ B ₁)	38.8	−0.4	−0.2	−0.1	SiH ₂ (³ B ₁)	32.9	−0.2	−0.2	−0.1
CH ₂ (¹ A ₁)	62.1	−1.3	−0.4	−0.2	SiH ₂ (¹ A ₁)	51.3	−2.8	−0.4	−0.5
CH ₃	59.4	−0.6	−0.3	−0.2	SiH ₃	42.7	−0.4	−0.3	−0.2
CH ₄	83.2	−0.8	−0.5	−0.5	SiH ₄	55.6	−0.7	−0.5	−0.4
NH ₂	49.1	−0.5	−0.2	−0.1	PH ₂	34.8	−0.2	−0.1	−0.1
NH ₃	75.0	−0.7	−0.6	−0.6	PH ₃	48.4	−0.4	−0.2	−0.1
OH ₂	53.7	−0.4	−0.4	−0.2	SH ₂	34.4	−0.2	−0.2	0.0
CH ₃ OH ^c	118.8	−1.0	−0.7	−0.5	CH ₃ SH	104.7	−0.8	−0.5	−0.3
CH ₃ Cl	87.2	−0.8	−0.5	−0.4	FCI	37.7	0.1	0.2	0.6
C ₂ H ₆	151.2	−1.2	−0.7	−0.9	Si ₂ H ₆	96.5	−1.0	−0.7	−0.4
O ₂ H ₂	107.5	−0.6	−0.4	−0.1	ClO	42.6	0.3	0.0	0.2
N ₂ H ₄	135.1	−1.0	−0.9	−0.8	HOCl	66.4	−0.2	−0.2	0.1
O ₂	104.3	0.1	0.1	0.6	S ₂	46.0	0.8	0.1	0.4
F ₂	79.1	−0.1	0.0	0.2	Cl ₂	23.3	0.2	0.1	0.3
CN	151.9	−0.5	−0.1	0.3	HCO	125.8	−0.4	−0.3	0.1
HCN	151.4	−1.3	−0.6	−0.7	H ₂ CO	147.1	−1.9	−1.3	−2.0
N ₂	148.5	−0.4	−0.3	0.0	P ₂	91.5	1.6	0.2	0.5
C ₂ H ₄	144.1	−1.0	−0.4	0.2	Si ₂	80.0	2.9	0.1	0.8
C ₂ H ₂	148.0	−0.9	−0.3	0.3	SiO	122.1	−0.8	−0.2	0.3
CO	131.8	−0.7	−0.4	−0.1	CS	104.6	0.1	−0.2	−0.1
NO	120.6	−0.2	0.0	0.4	SO	61.2	1.8	1.2	2.6
CO ₂	176.3	−0.2	−0.1	0.6	SO ₂	130.4	1.3	0.8	2.1
HOH [‡]	40.6	−0.2	−0.2	−0.3	O ₃	237.3	−0.1	0.3	1.1
HOCH ₄ ^{‡d}	124.1	−1.0	−0.9	−1.0	SH	18.7	−0.1	−0.1	0.0
CH ₃ CH ₂	128.9	−1.0	−0.6	−0.4	SH ₃ [‡]	41.1	−0.1	−0.1	0.0
HC ₂ H ₄ [‡]	148.3	−0.9	−0.4	0.1	HClH [‡]	37.4	0.5	0.2	0.4
N ₂ O	214.9	−0.2	0.0	0.6	CH ₃ F	93.8	−0.8	−0.5	−0.3
HN ₂ O [‡]	224.1	−0.2	0.3	1.0	CH ₃ FCI [‡]	112.3	−0.4	0.0	0.4
HN ₂	129.9	−0.3	0.0	0.4	OH [−]	22.5	−0.2	0.0	0.2
HN ₂ [‡]	158.2	−0.3	−0.1	0.3	FCH ₃ Cl ^{−‡}	77.1	−1.2	−0.2	0.5
HNC	140.8	−1.1	−0.3	0.0	HOCH ₃ F ^{−‡}	119.7	−1.2	−0.4	0.0
HCN [‡]	141.9	−0.7	−0.3	0.0	ClCH ₃ Cl ^{−‡}	80.2	−0.7	−0.1	−0.4

^aAll energies reported in mE_h. ^bCalculated using (2,2) active space, with unoccupied alkali metal atomic orbitals and doubly occupied halide atomic orbitals. ^cCalculated using (10,10) active space, constraining oxygen atomic orbitals to remain doubly occupied. ^dCalculated using (11,11) active space, constraining oxygen atomic orbitals to remain doubly occupied.

estimates with mean absolute deviations around 6 mE_h and worst-case errors up to 40 mE_h.

Including a single shell of polarization functions significantly improves basis set quality, reducing the mean absolute error to <1.6 mE_h = 1 kcal/mol, commonly accepted as “chemical accuracy”, except in the case of pc-0, for which no polarization functions are defined. Even supplementing pc-0 with polarization functions from pc-2 only slightly improves E_{stat} estimates, reducing the mean and maximum absolute errors to ~5 and 20 mE_h, respectively. Therefore, it is primarily the inability of this basis to recover appropriately sized molecular orbitals that limits E_{stat} accuracy.

Double- ζ basis sets of similar composition—pc-1, cc-pVDZ, 6-31G(*d*, *p*)—yield acceptable mean absolute errors. However,

maximum errors in individual cases can reach over 7 mE_h, which is clearly well outside chemical accuracy.

Triple zeta basis sets—pc-2, cc-pVTZ, 6-311G(2df, 2pd)—produce significantly more accurate E_{stat} values than their double- ζ counterparts, halving the mean and maximum errors. Notably, within this class of basis sets, cc-pVTZ outperforms both pc-2 and 6-311G(2df, 2pd). Indeed, cc-pVTZ is the smallest basis set that yields chemically accurate static correlation energies for all molecules in our data set, as listed in Table 4.

From this table, it also appears that the greatest flexibility in atomic orbital basis is required to describe oxygen in a range of molecular environments, with all triple- ζ basis sets struggling to recover E_{stat} to chemical accuracy for H₂CO, SO, and SO₂.

Table 5. Mean Absolute Errors, Maximum Errors (mE_h) in Static Correlation Energies and Primitive Basis Function Counts for the Carbon Atom for Truncated pc- n and cc-pV n Z Basis Sets

	mean $ \Delta E_{\text{stat}} $					mean $ \Delta E_{\text{stat}} $			
	[sp]	[spd]	[spdf]	[spdfg]		[sp]	[spd]	[spdf]	[spdfg]
pc-1	6.8	1.2			cc-pVDZ	6.4	1.2		
pc-2	6.1	0.8	0.6		cc-pVTZ	6.2	0.5	0.3	
pc-3	6.2	0.3	0.05	0.05	cc-pVQZ	6.4	0.3	0.05	0.04
pc-4	6.1	0.2	0.04	0.04	cc-pVSZ	6.2	0.3	0.00	0
	max $ \Delta E_{\text{stat}} $					max $ \Delta E_{\text{stat}} $			
	[sp]	[spd]	[spdf]	[spdfg]		[sp]	[spd]	[spdf]	[spdfg]
pc-1	38.8	7.3			cc-pVDZ	39.9	6.4		
pc-2	41.7	3.4	2.9		cc-pVTZ	41.4	2.7	1.3	
pc-3	40.5	1.7	0.3	0.3	cc-pVQZ	40.6	2.0	0.3	0.3
pc-4	39.7	1.6	0.2	0.2	cc-pVSZ	40.2	1.8	0.03	0
	n_{prim}					n_{prim}			
	[sp]	[spd]	[spdf]	[spdfg]		[sp]	[spd]	[spdf]	[spdfg]
pc-1	23	28			cc-pVDZ	29	34		
pc-2	34	44	51		cc-pVTZ	33	43	50	
pc-3	49	69	83	92	cc-pVQZ	39	54	68	77
pc-4	60	90	111	129	cc-pVSZ	48	68	89	107

The pc-2 basis is generally less accurate than cc-pVTZ and produces particularly poor estimates of E_{stat} for the second-row containing molecules Si_2 , P_2 , and SiH_2 ($^1\text{A}_1$), in addition to the oxygen-containing molecules discussed above. This result is surprising, as the molecular data set used in constructing the pc-2 basis contained all three of these molecules.⁴²

Quadruple and pentuple- ζ basis sets provide very accurate estimates of E_{stat} but at a correspondingly high computational cost. The rapid convergence of E_{stat} with respect to basis set size suggests that higher angular momentum functions do not contribute significantly to the accuracy of the results, and so they may be omitted to improve computational performance.

Optimal Basis Composition. To quantify the relationship between basis set composition, accuracy, and computational cost, mean and maximum absolute errors for a selection of truncated basis sets are reproduced in Table 5, along with primitive basis function counts.

As observed above, basis sets without polarization functions do not accurately capture the static correlation energy. Mean and maximum absolute errors for [sp] truncated polarization-consistent and correlation-consistent basis sets are very similar to those of pc-0, 6-31G and 6-311G. Therefore, the remainder of this discussion will focus on basis sets including at least one shell of polarization functions.

In the pc- n basis series, starting from pc-1, the greatest accuracy gains can be achieved for the least computational cost by first increasing the number of valence and first polarization shell functions, that is, moving down the [spd] basis series to pc-3[spd], before adding an additional shell of f functions to ensure sub- mE_h accuracy for all molecules and mean absolute errors less than $100 \mu E_h$.

The correlation-consistent basis sets tell a slightly different story. Starting from cc-pVDZ, it is again best to initially increase the number of valence and first polarization shell functions to cc-pVTZ[spd]. From there, the mean absolute error data suggest that it is almost equally advantageous to add an additional layer of polarization functions (cc-pVTZ) or increase the number of valence and first polarization shell functions (cc-pVQZ[spd]). However, adding f functions is more useful than increasing the number of s , p , and d functions in reducing the maximum absolute error. Physically, this means that atoms need the flexibility to adopt a range of different and highly

asymmetric shapes to fit the large range of molecular environments encountered within our data set. It is more important to allow a larger range of shapes than further refine the size of the cc-pVTZ[spd] orbitals.

Analogous to the pc- n series, sub- mE_h accuracy for all molecules is guaranteed by the cc-pVQZ[spdf] basis. Additional valence and polarization functions beyond those contained within pc-3[spdf] and cc-pVQZ[spdf] do not improve the accuracy of calculated E_{stat} values enough to justify the increase in computational cost.

For each column in Table 5, excluding [sp], the number of valence orbital and polarization functions increase simultaneously as the quality of the basis set improves. To decouple these effects, we have constructed a series of basis sets with swapped polarization functions, whose performance is summarized in Table 6.

Table 6. Mean and Maximum Absolute Errors (mE_h) in Static Correlation Energies for Polarization Function Swapped Truncated pc- n Basis Sets

	mean $ \Delta E_{\text{stat}} $		
	pc-1	pc-2	pc-3
[spd ₁]	1.2	1.3	
[spd ₂]	1.0	0.8	0.5
[spd ₃]	0.9	0.5	0.3
	max $ \Delta E_{\text{stat}} $		
	pc-1	pc-2	pc-3
[spd ₁]	7.3	7.7	
[spd ₂]	3.3	3.4	3.2
[spd ₃]	3.0	2.3	1.7

Basis sets which contain the same number of split valence functions and polarization functions, viz., pc-1[spd₂] and pc-2[spd₃], tend to provide the optimal balance between basis size and accuracy. From Table 6, it appears that improving the ability of basis sets to capture atomic polarization within the molecular environment plays a major role in increasing accuracy, particularly in terms of reducing maximum errors in E_{stat} .

Dependence of ΔE_{stat} on Molecule Size, Geometry, and Electronic Configuration. Mean and maximum absolute errors are a simple and robust metric for assessing the ability of

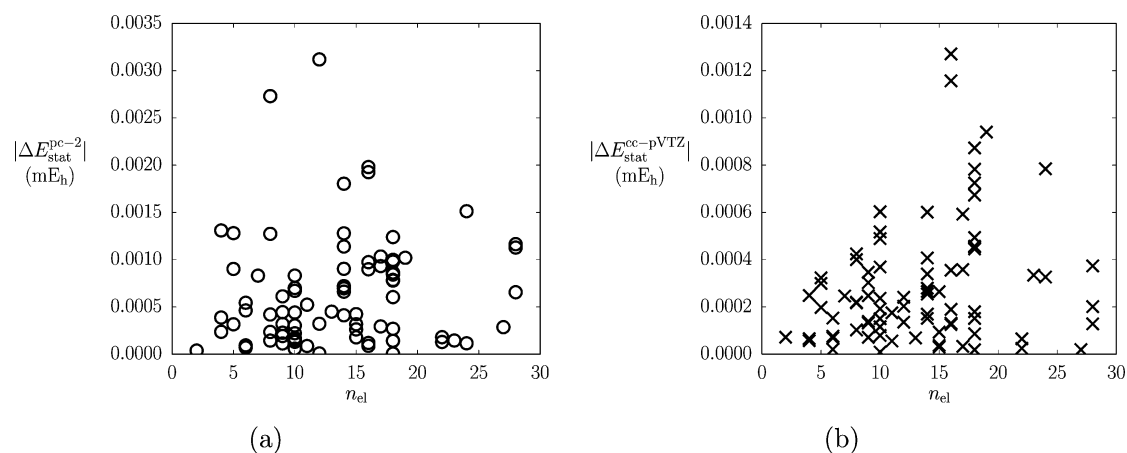


Figure 1. (a) $|\Delta E_{\text{stat}}^{\text{pc-2}}|$ and (b) $|\Delta E_{\text{stat}}^{\text{cc-pVTZ}}|$ as a function of number of electrons for each atom, molecule, and transition state complex in our data set.

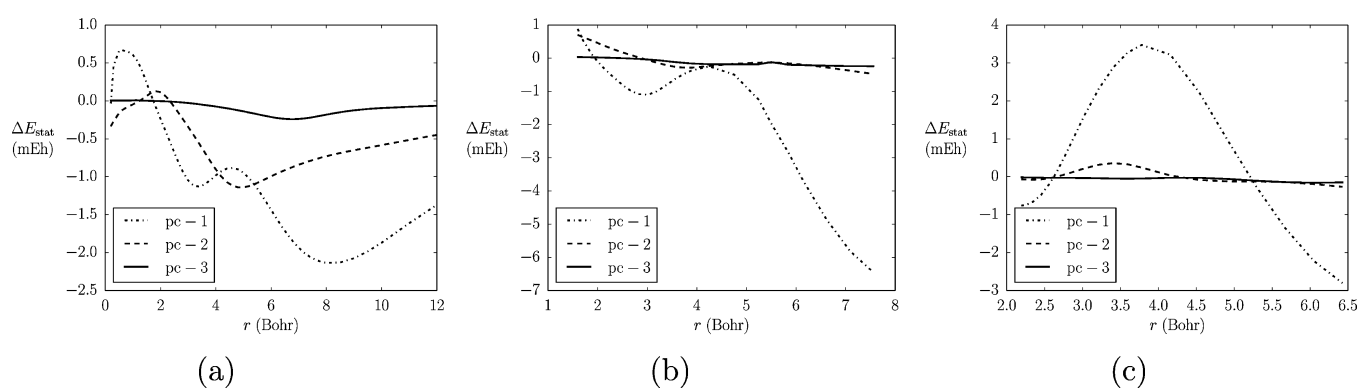


Figure 2. Deviation of smaller basis static correlation energies from pc-4 reference values, ΔE_{stat} , as a function of bond length for a series of dissociating diatomics: (a) H₂, (b) N₂, (c) F₂.

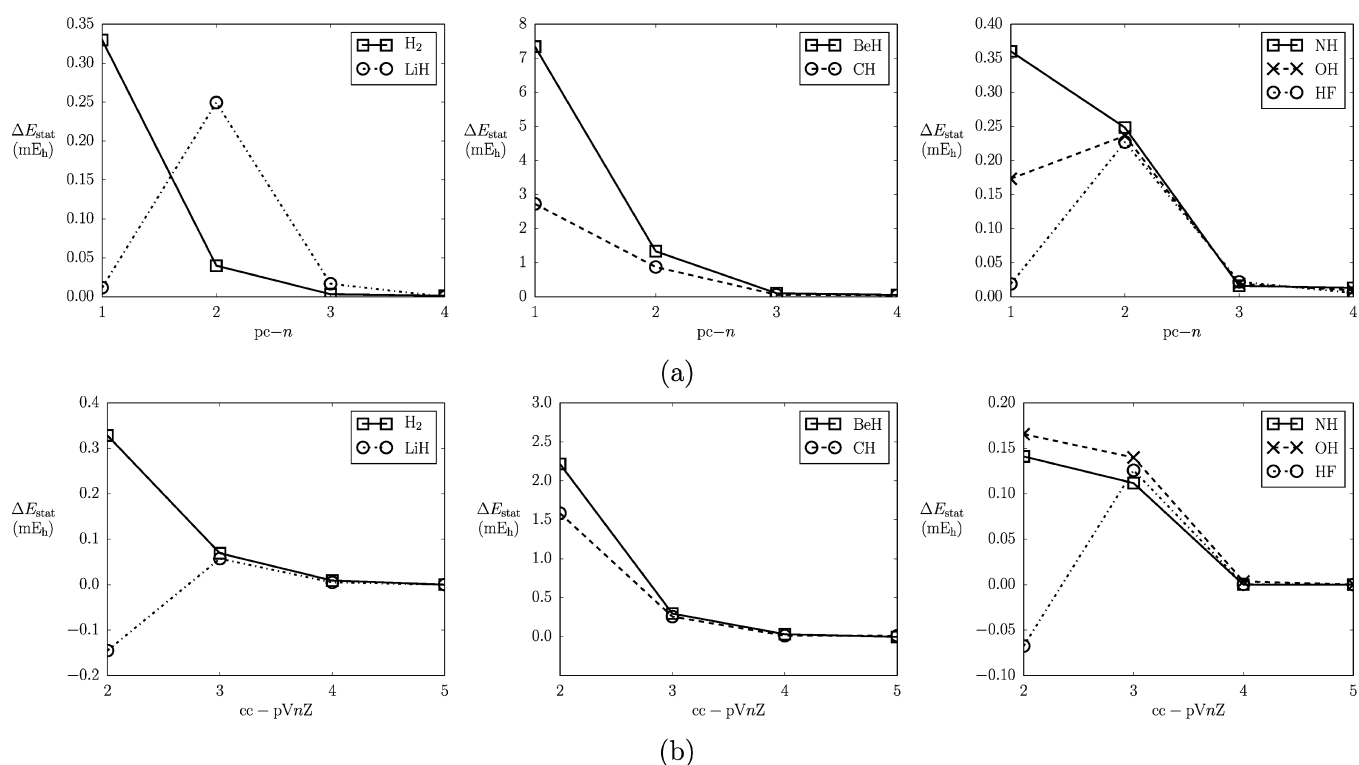


Figure 3. Deviation of static correlation energies from extrapolated complete basis set reference values for a series of hydrides, as a function of (a) polarization-consistent and (b) correlation-consistent basis size.

each basis set to recover static correlation energies for a range of atomic and molecular systems. However, a more physically motivated criterion for acceptable performance is that ΔE_{stat} should be independent of molecule size, geometry, and bonding patterns. In other words, the atomic orbital basis should be flexible enough to describe each atom in a range of molecular environments.

The ability of the pc-2 basis to recover E_{stat} is largely independent of molecule size (measured as the number of valence electrons), as illustrated in Figure 1. Linear regression analysis indicates that there is no statistically significant correlation between molecule size and error in E_{stat} values calculated in the pc-2 basis, with a near zero slope ($9 \pm 22 \mu E_{\text{h}}$ /valence electron) and small R^2 value (<0.01). This is also true of the cc-pVTZ basis (slope = $10 \pm 9 \mu E_{\text{h}}$ /valence electron and $R^2 = 0.05$). Reported uncertainties represent 95% confidence intervals.

In general, all basis sets of at least double- ζ quality show little to no correlation between ΔE_{stat} and molecular size. Hence, this is not as strict a criterion for acceptable performance as error threshold measures. However, it is reassuring that the errors calculated from our data set are independent of molecular size, so the conclusions about basis set quality drawn from our data set are expected to apply also to larger molecules. A complete set of linear regression parameters for all basis sets is provided as Supporting Information.

The ability of each basis set to capture E_{stat} in a range of molecular environments is most clearly illustrated by monitoring errors in E_{stat} as a function of bond length for a series of prototypical diatomics, as shown in Figure 2.

The pc-2 basis set provides the minimum flexibility required calculate E_{stat} to chemical accuracy across all bond lengths for H_2 , N_2 and F_2 .

On the basis of the dissociating diatomic curves in Figure 2, it is plausible to hypothesize that atomic orbital basis sets recover E_{stat} more effectively at their equilibrium geometries (~ 1.4 Bohr for H_2 , ~ 2.1 Bohr for N_2 , and ~ 2.7 Bohr for F_2). However, separate analysis of the G1 and DBH data sets reveals no significant difference in accuracy of E_{stat} estimates between equilibrium and transition state systems for all nonminimal basis sets. Therefore, triple- ζ basis sets provide an optimal compromise between accuracy and computational cost for molecules both at, and slightly displaced from, their equilibrium geometries.

The final physical test of basis set quality lies in the ability to capture E_{stat} accurately, independent of bonding patterns. As molecules with unsaturated bonding have larger E_{stat} than saturated systems, they may also have larger basis set incompleteness errors. Separate analysis of saturated and unsaturated systems within our data set confirms that this is only true for basis sets smaller than triple- ζ . For triple- ζ and larger basis sets, ΔE_{stat} is independent of molecular electronic structure.

Basis Set Convergence Behavior of E_{stat} . The convergence of E_{stat} toward the complete basis set limit is illustrated for a series of diatomic hydrides in Figure 3. E_{stat} converges monotonically from pc-2 (Figure 3a) and cc-pVTZ (Figure 3b) to the complete basis set limit. This behavior is consistent across our complete data set (data not shown, available as Supporting Information).

From smaller basis sets (pc-0, pc-1, cc-pVDZ), E_{stat} no longer converges monotonically to the CBS limit. In these cases, the CI expansion no longer primarily recovers the static correlation

energy from qualitatively correct CASSCF orbitals, but it also compensates for basis set incompleteness errors. This complicated nonlinear coupling between CI and MO coefficients and the competing incentives to optimize both the orbitals and static correlation energy leads to the erratic convergence at the small basis end of each plot in Figure 3.

Similar basis set convergence patterns have also been observed in high level CCSD(T) calculations of spectroscopic parameters.⁵⁹ This suggests that the inadequacy of double- ζ basis sets leading to competing effects in determining CI/CC coefficients will show up clearly in any metric more sensitive than the total CASSCF energy. Further, it has long been observed that double- ζ basis sets are too small to be used reliably in basis set extrapolation schemes, even for total energies.⁵⁹

Convergence of CASSCF Wave Functions and Energies. The basis set convergence of E_{stat} provides a useful diagnostic for monitoring convergence of individual CASSCF calculations within a basis set series. Convergence of all CASSCF calculations, not just full-valence, can be complicated by strong nonlinear coupling between configuration interaction (CI) and molecular orbital (MO) coefficients.⁶⁰ One surprising consequence of this is that convergence of the CASSCF energy with respect to basis set size for a given molecule and basis set series does *not* imply convergence of its configuration interaction and molecular orbital coefficients.

A less surprising consequence is that individual medium-to-large basis CASSCF calculations often converge to solutions that represent only local minima in CI and MO coefficient space.

The simplest solution to this problem is to use a smaller basis set. Indeed, CASSCF calculations in a minimal atomic basis always converge to the global minimum, as molecular orbitals are completely defined by the preceding Hartree–Fock procedure, so there is no coupling between CI and MO coefficients. Unfortunately, minimal atomic orbital basis sets (e.g., STO-3G) provide unreliable estimates of E_{stat} with mean and maximum absolute errors of ~ 12 and 93 mE_{h} , respectively.

However, the qualitatively correct valence molecular orbitals produced by minimal atomic orbital basis sets provide a useful starting point for finding global minimum CASSCF solutions in larger basis sets. To enable universal expansion of molecular orbitals in different atomic orbital bases, we constructed a small utility program⁶¹ to perform least-squares fitting of one atomic orbital basis to another and subsequent expansion of the original molecular orbitals in terms of the new fitted basis functions.

This approach almost exactly reproduces molecular orbitals when expanding from a small AO basis to a larger AO basis. The BasisFit utility may also be reliably used to interconvert molecular orbitals between basis sets of the same quality but different basis set families, for example, pc-2 \leftrightarrow cc-pVTZ \leftrightarrow 6-311G(2df, 2pd). However, it is less reliable when fitting from a larger basis to a smaller one, because the smaller basis lacks the flexibility to exactly reproduce the functions in the larger AO basis, and contributions to the molecular orbitals from higher angular momentum functions than available in the smaller AO basis must be neglected. The resulting molecular orbitals are therefore slightly distorted and no longer orthogonal.

In practice, this makes it easier to converge CASSCF calculations by starting from smaller basis molecular orbitals, even if converged molecular orbitals from a larger basis are available. Therefore, to reliably obtain converged CASSCF

molecular orbitals in medium to large basis sets, we advocate the following bootstrapping procedure:

- obtain molecular orbitals in minimal AO basis e.g. STO-3G (note that CASSCF and Hartree–Fock orbitals are equivalent for most molecules whose orbitals are constrained by symmetry).
- expand molecular orbitals from minimal AO basis in a slightly larger AO basis (e.g., pc-0, as described above).
- optimize valence molecular orbitals during CASSCF calculation.
- expand these molecular orbitals in a yet larger AO basis (e.g., pc-0 → pc-1, pc-1 → pc-2, etc).
- repeat last two steps until E_{stat} is obtained with desired accuracy.

This procedure may also be useful to expedite the convergence of VOO–CCD calculations, which also involve concurrent orbital optimization and determination of coupled cluster coefficients.²²

In this work, we have taken great care to ensure that CASSCF wave functions and energies are fully converged to their global minima. The CASSCF energies, available as Supporting Information, may therefore be used for validating approximate E_{stat} models, as well as testing CASSCF convergence algorithms and strategies.

Accuracy of Near-CBS E_{stat} Values. The fully converged CASSCF energies in near-complete atomic orbital basis sets pc-4 and cc-pVSZ are expected to produce highly accurate static correlation energies, such that they may be considered exact for most practical purposes. However, it is important to verify this assertion by estimating the remaining basis set incompleteness error.

A range of strategies were employed to assess the accuracy of the pc-4 and cc-pVSZ E_{stat} values including:

- comparison of pc-4 and cc-pVSZ results
- comparison between quadruple and pentuple- ζ basis results
- comparison with extrapolated estimates of E_{stat} at the complete basis set limit

The basis set convergence behavior of E_{stat} across both the pc-($n-1$) and cc-pVnZ basis set series is well-captured by a three-point extrapolation of the form:

$$E_{\text{stat},n} = E_{\text{stat},\infty} + A \exp(-B\sqrt{n}) \quad (9)$$

provided that all basis sets are of at least triple- ζ quality. RMS fitting errors are less than 21 m E_h across all molecules and basis sets in each basis set series. Here, n represents the number of split-valence basis functions, and B is a fixed exponent optimized for each basis set series. $E_{\text{stat},\infty}$ may also be denoted $E_{\text{stat}}^{\text{CBS}}$.

In general, pc- n E_{stat} values converge faster to the CBS limit than cc-pVnZ, with optimal exponents of 3.2 and 2.5, respectively. This reflects the basis set convergence behavior observed for E^{HF} by Jensen.^{37,42,43}

More sophisticated extrapolation schemes that depend on the number of primitive s -type basis functions are available for the pc- n basis series, but not for cc-pVnZ. Further, E_{stat} as a relative energy, appears to exhibit basis set convergence behavior that is less sensitive to choice of extrapolation variable than its parent E^{HF} and $E^{\text{CASSCF(val)}}$ absolute energies.

All quantities relevant to assessing the accuracy of pentuple- ζ results are presented in Table 7. Differences between various estimates of E_{stat} are quantified as:

$$\Delta E_{\text{stat}}^{X,Y} = E_{\text{stat}}^X - E_{\text{stat}}^Y \quad (10)$$

with X and Y chosen as shown in Table 7.

Table 7. Mean and Maximum Absolute Deviations in E_{stat} (μE_h), between Pentuple- ζ Correlation Consistent and Polarization Consistent Basis Sets, between Quadruple- ζ and Pentuple- ζ Basis Sets within Each Series, and between Pentuple- ζ Basis Sets and Estimates of E_{stat} at the Complete Basis Set Limit Extrapolated from the Correlation Consistent Basis Set Series (cc-CBS) and the Polarization Consistent Basis Set Series (pc-CBS)

X	Y	mean $ \Delta E_{\text{stat}}^{X,Y} $		maximum $ \Delta E_{\text{stat}}^{X,Y} $	
		first row molecules	second row molecules	first row molecules	second row molecules
cc-pVSZ	pc-4	14	79	84	228
cc-pVSZ	cc-pVQZ	40	43	229	309
pc-4	pc-3	24	46	177	247
cc-CBS	cc-pVSZ	14	17	75	103
pc-CBS	pc-4	9	12	74	57

Differences between quadruple and pentuple- ζ basis results should provide an upper limit on the remaining variation between the pentuple- ζ results and the complete basis set limit, but comparison of pc-4 and cc-pVSZ results reveals that these upper limits are exceeded for molecules containing second row atoms (Na–Ar). This implies that either the pc- n or cc-pVnZ basis set series for second row atoms possesses systematic deficiencies which prevent convergence to the true CBS limit.

This supposition can only be tested indirectly by comparing finite basis and numerically exact Hartree–Fock energies for a selection of atoms and diatomic molecules from our data set, per the results presented in Table 8. In all cases, the numerical

Table 8. Deviation of pc-4 and cc-pVSZ Hartree–Fock Energies (m E_h) from Their Numerically Exact Values

	ΔE^{HF}			ΔE^{HF}	
	pc-4	cc-pVSZ		pc-4	cc-pVSZ
H	0.000	0.005	He	0.005	0.055
Li	0.083	0.004	Na	8.311	0.236
Be	0.132	0.011	Mg	7.207	0.031
B	0.012	0.029	Al	6.386	0.081
C	0.039	0.051	Si	5.935	0.105
N	0.036	0.082	P	5.650	0.104
O	0.057	0.166	S	5.475	0.171
F	0.061	0.241	Cl	5.323	0.193
F [−]	0.213	4.028	Cl [−]	5.468	2.020
Ne	0.055	0.328	Ar	5.202	0.171
H ₂	0.002	0.022	Na ₂	16.605	0.578
HF	0.058	0.374	NaCl	13.854	0.961
LiF	0.167	0.333	FCI	5.348	0.737
F ₂	0.155	0.706	Cl ₂	10.619	0.603
N ₂	0.099	0.447	P ₂	11.411	0.528
CO	0.108	0.313	CS	5.631	0.483

HF energies were calculated to sub- μE_h accuracy. These values, along with the number of grid points and radial parameters used to generate them, are available as Supporting Information.

Overall, cc-pVSZ accurately recovers the exact HF energies, except for anions, where the cc-pVSZ energies are in error by up to 4 m E_h . These results reinforce the well-known conclusion

that diffuse functions are required when modeling anions using correlation-consistent basis sets.

On the other hand, the pc-4 basis performs equally well for anions and uncharged species. Providing this additional flexibility in the valence region comes at the expense of accurately modeling the chemically uninteresting core region for second row atoms. The pc-4 Hartree–Fock energies for second-row atoms differ from the exact values by up to 8 mE_h. This error is additive as second row atoms are incorporated into molecules, but is expected to cancel when calculating relative energies (e.g., reaction enthalpies, activation energies, static correlation energies).

Indeed, the mE_h basis set incompleteness errors in E^{HF} for second-row atoms cancel to become only μE_h errors in E_{stat}. Nonetheless, as pentuple- ζ basis sets are capable of recovering E_{stat} to μE_h accuracy, this represents a significant and systematic deviation of the pc-4 E_{stat} values from the true CBS limit.

CBS-extrapolated static correlation energies provide a more realistic estimate of differences between pentuple- ζ results and the complete basis set limit. These results suggest that the pc-4 and cc-pV5Z results lie, on average, within 20 μE_h of the complete basis set limit, although this does not account for the systematic deficiencies in the polarization-consistent basis sets for second-row atoms. Basis set extrapolation predicts a maximum deviation from the CBS limit of 0.1 mE_h, whereas more conservative approaches indicate that the maximum error could be up to 0.3 mE_h.

CONCLUSIONS

Split valence basis sets without polarization functions cannot accurately capture the static correlation energy, regardless of the number of *s* and *p* functions. On the other hand, functions of high angular momentum—*g* and above—can be excluded from the basis set without appreciable loss of accuracy in E_{stat}. The sweet spot for the inclusion of polarization functions lies in the *d* and *f* functions. Chemical accuracy for all molecules may be achieved using a near complete basis of *s*, *p*, and *d* functions. However, this may be achieved at lower computational cost by using a well-balanced basis set including *f* functions, along with a smaller number of *s*, *p*, and *d* functions.

For this reason, triple- ζ basis sets exhibit a near-optimal balance between accuracy and computational cost in recovering static correlation energies. In particular, the cc-pVTZ basis provides chemically accurate E_{stat} values for all molecules in our data set. The 6-311G(2d,2p) basis gives remarkably accurate results for its relatively small size but cannot guarantee chemical accuracy for all molecules.

Truncated quadruple- ζ basis sets stripped of their *g* functions provide the most cost-effective way of recovering E_{stat} to sub-mE_h accuracy and may be considered close enough to the complete basis set limit for most practical purposes.

ASSOCIATED CONTENT

Supporting Information

All geometries in our molecular data set, with their Hartree–Fock, CASSCF and static correlation energies in all basis sets. Linear regression statistics for the relationship between ΔE_{stat} and number of valence electrons, parameters from fitting to basis set convergence formulae, and numerical Hartree–Fock results for a selection of atoms and diatomics. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: deborah.crittenden@canterbury.ac.nz.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to thank the BlueFern Supercomputing Facility at the University of Canterbury for a generous allocation of computer time.

REFERENCES

- (1) Löwdin, P.-O. The Correlation Problem in Many-Electron Quantum Mechanics. 1. Review of Different Approaches and Discussion of Some Current Ideas. *Adv. Chem. Phys.* **1959**, *2*, 207–322.
- (2) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; McGraw-Hill: New York, 1989.
- (3) Sherrill, C. D.; III, H. F. S. In *The Configuration Interaction Method: Advances in Highly Correlated Approaches*; Per-Olov Löwdin, M. C. Z., John R. Sabin, Brändas, E., Eds.; Advances in Quantum Chemistry; Academic Press, 1999; Vol. 34; pp 143 – 269.
- (4) O'Neill, D. P.; Gill, P. M. W. Benchmark Correlation Energies for Small Molecules. *Mol. Phys.* **2005**, *103*, 763–766.
- (5) Kutzelnigg, W.; Morgan, J. D., III Rates of Convergence of the Partial Wave Expansions of Atomic Correlation Energies. *J. Chem. Phys.* **1992**, *96*, 4484–4508.
- (6) Feller, D. The Use of Systematic Sequences of Wave Functions for Estimating the Complete Basis Set, Full Configuration Interaction Limit in Water. *J. Chem. Phys.* **1993**, *98*, 7059–7071.
- (7) Csaszar, A. G.; Tarczay, G.; Leininger, M. L.; Polyansky, O. L.; Tennyson, J.; Allen, W. D. *Spectroscopy from Space*; NATO Science Series; Springer, 2001; Vol. 20; Chapter 19. Dream or Reality: Complete Basis Set Full Configuration Interaction Potential Energy Hypersurfaces, pp 317–339.
- (8) Dutta, A.; Sherrill, C. D. Full Configuration Interaction Potential Energy Curves for Breaking Bonds to Hydrogen: An Assessment of Single-Reference Correlation Methods. *J. Chem. Phys.* **2003**, *118*, 1610–1619.
- (9) Varandas, A. J. Extrapolation to the Complete Basis Set Limit without Counterpoise. The Pair Potential of Helium Revisited. *J. Phys. Chem. A* **2010**, *114*, 8505–8516.
- (10) Sherrill, C. D. Frontiers in Electronic Structure Theory. *J. Chem. Phys.* **2010**, *132*, 110902.
- (11) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M.; Elbert, S. T. Are Atoms Intrinsic to Molecular Electronic Wavefunctions? I. The FORS Model. *Chem. Phys.* **1982**, *71*, 41–49.
- (12) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M. Are Atoms Intrinsic to Molecular Electronic Wavefunctions? II. Analysis of FORS Orbitals. *Chem. Phys.* **1982**, *71*, 51–64.
- (13) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M.; Elbert, S. T. Are Atoms Intrinsic to Molecular Electronic Wavefunctions? III. Analysis of FORS Configurations. *Chem. Phys.* **1982**, *71*, 65–78.
- (14) 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.
- (15) Kong, L.; Bischoff, F. A.; Valeev, E. F. Explicitly Correlated R12/F12 Methods for Electronic Structure. *Chem. Rev.* **2012**, *112*, 75–107.
- (16) Braida, B.; Toulouse, J.; Caffarel, M.; Umrigar, C. J. Quantum Monte Carlo with Jastrow-Valence-Bond Wave Functions. *J. Chem. Phys.* **2011**, *134*, 084108.
- (17) Sinanoglu, O. Many-Electron Theory of Atoms, Molecules and their Interactions. *Adv. Chem. Phys.* **1964**, *6*, 315–412.
- (18) Mok, D. K. W.; Neumann, R.; Handy, N. C. Dynamical and Nondynamical Correlation. *J. Chem. Phys.* **1996**, *100*, 6225–6230.

- (19) Lee, M. S.; Head-Gordon, M. Extracting Polarized Atomic Orbitals from Molecular Orbital Calculations. *Int. J. Quantum Chem.* **2000**, *76*, 169–184.
- (20) Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg, K. Molecule Intrinsic Minimal Basis Sets. I. Exact Resolution of Ab Initio Optimized Molecular Orbitals in Terms of Deformed Atomic Minimal-Basis Orbitals. *J. Chem. Phys.* **2004**, *120*, 2629–2638.
- (21) Abrams, M. L.; Sherrill, C. D. Natural Orbitals as Substitutes for Optimized Orbitals in Complete Active Space Wavefunctions. *Chem. Phys. Lett.* **2004**, *395*, 227–232.
- (22) Krylov, A. I.; Sherrill, C. D.; Byrd, E. F. C.; Head-Gordon, M. Size-consistent Wave Functions for Nondynamical Correlation Energy: The Valence Active Space Optimized Orbital Coupled-Cluster Doubles Model. *J. Chem. Phys.* **1998**, *109*, 10669–10678.
- (23) Krylov, A. I.; Slipchenko, L. V.; Levchenko, S. V. In *Electron Correlation Methodology*; Wilson, A., Ed.; ACS Symposium Series; American Chemical Society: Washington, DC, 2007; Chapter 6. Breaking the Curse of the Non-Dynamical Correlation Problem: The Spin-Flip Method.
- (24) Becke, A. D.; Johnson, E. R. A Unified Density-Functional Treatment of Dynamical, Nondynamical and Dispersion Correlations. *J. Chem. Phys.* **2007**, *127*, 124108.
- (25) Tsuchimochi, T.; Scuseria, G. E. Strong Correlations via Constrained-Pairing Mean-Field Theory. *J. Chem. Phys.* **2009**, *131*, 121102.
- (26) Chan, G. K. L.; Sharma, S. In *Solving the Schrodinger Equation: Has Everything Been Tried?*; Popelier, P., Ed.; Imperial College Press: London, 2011; Chapter 3. Solving Problems With Strong Correlation Using The Density Matrix Renormalization Group (DMRG), pp 43–60.
- (27) Avella, A.; Mancini, F., Eds. *Strongly Correlated Systems: Numerical Methods*, 1st ed.; Springer Series in Solid-State Sciences; Springer, 2013; Vol. 176.
- (28) Crittenden, D. L. A Hierarchy of Static Correlation Models. *J. Phys. Chem. A* **2013**, *117*, 3852–3860.
- (29) Ten-No, S. A Simple F12 Geminal Correlation in Multi-Reference Perturbation Theory. *Chem. Phys. Lett.* **2007**, *447*, 175–179.
- (30) Torheyden, M.; Valeev, E. F. Universal Perturbative Explicitly Correlated Basis Set Incompleteness Correction. *J. Chem. Phys.* **2008**, *131*, 171103.
- (31) Shiozaki, T.; Werner, H.-J. Second-Order Multireference Perturbation Theory with Explicit Correlation CASPT2-F12. *J. Chem. Phys.* **2010**, *133*, 141103.
- (32) Shiozaki, T.; Knizia, G.; Werner, H.-J. Explicitly Correlated Multireference Configuration Interaction: MRCI-F12. *J. Chem. Phys.* **2011**, *134*, 034113.
- (33) Varganov, S.; Martinez, T. J. Variational Geminal-Augmented Multireference Self-Consistent Field Theory: Two-Electron Systems. *J. Chem. Phys.* **2010**, *132*, 054103.
- (34) Grimme, S.; Waletzke, M. A Combination of Kohn-Sham Density Functional Theory and Multi-Reference Configuration Interaction Methods. *J. Chem. Phys.* **1999**, *111*, 5645–5656.
- (35) Leininger, T.; Stoll, H.; Werner, H.-J.; Savin, A. Combining Long-Range Configuration Interaction with Short-Range Density Functionals. *Chem. Phys. Lett.* **1997**, *275*, 151–160.
- (36) Grafenstein, J.; Cremer, D. The Combination of Density Functional Theory with Multi-Configuration Methods - CAS-DFT. *Chem. Phys. Lett.* **2000**, 316.
- (37) Jensen, F. Polarization Consistent Basis Sets: Principles. *J. Chem. Phys.* **2001**, *115*, 9113–9125.
- (38) Jensen, F. Estimating the Hartree-Fock Limit from Finite Basis Set Calculations. *Theor. Chem. Acc.* **2005**, *113*, 267–273.
- (39) McKemmish, L. K.; Gill, P. M. W. Gaussian Expansions of Orbitals. *J. Chem. Theory Comput.* **2012**, *8*, 4891–4898.
- (40) Karton, A.; Martin, J. M. L. Comment on: "Estimating the Hartree-Fock Limit from Finite Basis Set Calculations" [Jensen F (2005) Theor Chem Acc 113:267]. *Theor. Chem. Acc.* **2006**, *115*, 330–333.
- (41) Petersson, G. A.; Malick, D. K.; Frisch, M. J.; Braunstein, M. The Convergence of Complete Active Space Self-Consistent-Field Energies to the Complete Basis Set Limit. *J. Chem. Phys.* **2005**, *123*, 074111.
- (42) Jensen, F.; Helgaker, T. Polarization Consistent Basis Sets. V. The Elements Si–Cl. *J. Chem. Phys.* **2004**, *121*, 3463–3470.
- (43) Jensen, F. Polarization Consistent Basis Sets. VI. The Elements He, Li, Be, B, Ne, Na, Mg, Al, Ar. *J. Phys. Chem. A* **2007**, *111*, 11198–11204.
- (44) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *80*, 1007–1023.
- (45) Woon, D. E.; Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum Through Argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (46) Woon, D. E.; Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. IV. Calculation of Static Electrical Response Properties. *J. Chem. Phys.* **1994**, *100*, 2975–2988.
- (47) Dunning, T. H., Jr.; Peterson, K. A.; Wilson, A. K. Gaussian Basis Sets for Use in Correlated Molecular Calculations. X. The Atoms Aluminum Through Argon Revisited. *J. Chem. Phys.* **2001**, *114*, 9244–9253.
- (48) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1363–1363.
- (49) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, A. Gaussian-1 Theory: A General Procedure for Prediction of Molecular Energies. *J. Chem. Phys.* **1989**, *90*, S622–S630.
- (50) Zheng, J.; Zhao, Y.; Truhlar, D. G. Representative Benchmark Suites for Barrier Heights of Diverse Reaction Types and Assessment of Electronic Structure Methods for Thermochemical Kinetics. *J. Chem. Theory Comput.* **2007**, *3*, S69–S82.
- (51) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *J. Chem. Phys.* **1969**, *51*, 2657–2664.
- (52) Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular Orbital Methods. IV. Use of Gaussian Expansions of Slater-Type Orbitals. Extension to Second-Row Molecules. *J. Chem. Phys.* **1970**, *52*, 2769–2773.
- (53) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (54) Dill, J. D.; Pople, J. A. Self-Consistent Molecular Orbitals Methods. XV. Extended Gaussian-Type Basis Sets for Lithium, Beryllium and Boron. *J. Chem. Phys.* **1975**, *62*, 2921–2923.
- (55) Franci, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XXIII. A Polarization-Type Basis Set for Second-Row Elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
- (56) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-Consistent Molecular Orbital Methods 25. Supplementary Functions for Gaussian Basis Sets. *J. Chem. Phys.* **1984**, *80*, 3265–3269.
- (57) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wave Functions. *J. Chem. Phys.* **1980**, *72*, 650–654.
- (58) McLean, A. D.; Chandler, G. S. Contracted Gaussian Basis Sets for Molecular Calculations. I. Second Row Atoms, Z=11–18. *J. Chem. Phys.* **1980**, *72*, S639–S648.
- (59) Temelso, B.; Valeev, E. F.; Sherrill, C. D. A Comparison of One-Particle Basis Set Completeness, Higher-Order Electron Correlation, Relativistic Effects, and Adiabatic Corrections for Spectroscopic Constants of BH, CH+, and NH. *J. Phys. Chem. A* **2004**, *108*, 3068–3075.
- (60) Sherrill, C. D.; Krylov, A. I.; Byrd, E. F. C.; Head-Gordon, M. Energies and Analytic Gradients for Coupled-Cluster Doubles Model Using Variational Brueckner Orbitals: Application to Symmetry Breaking in O4+. *J. Chem. Phys.* **1998**, *109*, 4171–4181.

(61) BasisFit (accessed 18 February, 2014). <http://sourceforge.net/projects/basisfit/>.