AIP | The Journal of Chemical Physics

Extrapolating MP2 and CCSD explicitly correlated correlation energies to the complete basis set limit with first and second row correlation consistent basis sets

J. Grant Hill, Kirk A. Peterson, Gerald Knizia, and Hans-Joachim Werner

Citation: The Journal of Chemical Physics **131**, 194105 (2009); doi: 10.1063/1.3265857
View online: http://dx.doi.org/10.1063/1.3265857
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/131/19?ver=pdfcov
Published by the AIP Publishing

Extrapolating MP2 and CCSD explicitly correlated correlation energies to the complete basis set limit with first and second row correlation consistent basis sets

J. Grant Hill,[1] Kirk A. Peterson,[1,a)] Gerald Knizia,[2] and Hans-Joachim Werner[2]
[1]*Department of Chemistry, Washington State University, Pullman, Washington 99164, USA*
[2]*Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany*

(Received 16 September 2009; accepted 29 October 2009; published online 20 November 2009)

Accurate extrapolation to the complete basis set (CBS) limit of valence correlation energies calculated with explicitly correlated MP2-F12 and CCSD(T)-F12b methods have been investigated using a Schwenke-style approach for molecules containing both first and second row atoms. Extrapolation coefficients that are optimal for molecular systems containing first row elements differ from those optimized for second row analogs, hence values optimized for a combined set of first and second row systems are also presented. The new coefficients are shown to produce excellent results in both Schwenke-style and equivalent power-law-based two-point CBS extrapolations, with the MP2-F12/cc-pV(D,T)Z-F12 extrapolations producing an average error of just 0.17 mE_h with a maximum error of 0.49 for a collection of 23 small molecules. The use of larger basis sets, i.e., cc-pV(T,Q)Z-F12 and aug-cc-pV(Q,5)Z, in extrapolations of the MP2-F12 correlation energy leads to average errors that are smaller than the degree of confidence in the reference data ($\sim$0.1 mE_h). The latter were obtained through use of very large basis sets in MP2-F12 calculations on small molecules containing both first and second row elements. CBS limits obtained from optimized coefficients for conventional MP2 are only comparable to the accuracy of the MP2-F12/cc-pV(D,T)Z-F12 extrapolation when the aug-cc-pV(5+*d*)Z and aug-cc-pV(6+*d*)Z basis sets are used. The CCSD(T)-F12b correlation energy is extrapolated as two distinct parts: CCSD-F12b and (T). While the CCSD-F12b extrapolations with smaller basis sets are statistically less accurate than those of the MP2-F12 correlation energies, this is presumably due to the slower basis set convergence of the CCSD-F12b method compared to MP2-F12. The use of larger basis sets in the CCSD-F12b extrapolations produces correlation energies with accuracies exceeding the confidence in the reference data (also obtained in large basis set F12 calculations). It is demonstrated that the use of the 3C(D) *Ansatz* is preferred for MP2-F12 CBS extrapolations. Optimal values of the geminal Slater exponent are presented for the diagonal, fixed amplitude *Ansatz* in MP2-F12 calculations, and these are also recommended for CCSD-F12b calculations. © *2009 American Institute of Physics*. [doi:10.1063/1.3265857]

I. INTRODUCTION

The extrapolation of the correlation energy to the complete basis set (CBS) limit in electronic structure calculations has been facilitated by the development of systematically convergent Gaussian basis sets, i.e., the family of correlation consistent basis sets. As correlation energies at the CBS limit reduce the error in a given computation to that intrinsic to the underlying method in which the expansion of the wave function in terms of Slater determinants has been carried out, extrapolation to this limit is routinely employed in a great number of quantum chemical applications such as composite thermochemistry,[1] generation of potential energy surfaces,[2] and obtaining accurate geometries,[3,4] among many others. The development of systematically improvable basis sets began in earnest with the atomic natural orbital basis sets of Almlöf and Taylor,[5] before the correlation consistent basis sets for the first row elements and hydrogen were reported by Dunning.[6] In their standard form these latter sets are denoted cc-pV*n*Z (where *n*=D, T, Q, 5,…) and have since been expanded upon to cover almost all of the periodic table (for a recent review see Ref. 7) along with application-targeted improvements such as diffuse-augmented[8] (aug-cc-pV*n*Z) and core-valence correlating[9] (cc-pCV*n*Z or cc-pwCV*n*Z) sets.

The problem of slow convergence of the correlation energy with respect to basis set size, which arises from the poor description of the two-electron correlation cusps in the wave function by just products of one-electron functions, can also be approached in methods such as second-order Møller–Plesset perturbation theory (MP2) or coupled cluster theory including single and double excitations (CCSD) by the use of wave functions that are explicitly correlated: those containing terms that depend explicitly on the interelectronic distances r_{ij}.[10] The origin of such methods was the realization by Hylleraas[11] that the convergence of calculations on helium could be accelerated by including two-electron basis functions that depended explicitly on the interelectronic co-

a)Electronic mail: kipeters@wsu.edu.

ordinate. Unfortunately this leads to complicated many-electron integrals, meaning that the problem is essentially computationally intractable for all but the smallest of systems. The recent popularity of such methods is largely due to the introduction of resolution of the identity (RI) methods that employ an auxiliary basis set to accurately and efficiently approximate the many-electron integrals.[12,13] Since this time, developments within the methodology have appeared at a great rate, with some important advances including density fitting (DF) approximations to increase the speed of integral evaluation[14] and the use of correlation factors that are nonlinear in the treatment of the interelectronic distances (the F12 methods).[15–17] A more extensive history of the development and use of explicitly correlated wave functions in quantum chemistry is available in a recently published review,[18] and interested readers are referred to that text and references therein. As a consequence of these advances, methods such as DF-MP2-F12 and CCSD(T)-F12 produce results that are much closer to the CBS limit than can be obtained with their conventional analogs using the same basis set.

Extrapolations to the CBS limit generally operate on the premise that by obtaining correlation energies with a series of correlation consistent basis sets starting with a cardinal number n and continuing with $n+1$, etc., a functional form can be applied that will produce a value close to that at the CBS limit for a given choice of wave function method. Following the first CBS extrapolations of Feller[19] and earlier formal analyses involving partial wave expansions of the correlation energy,[20,21] a power-law-based relation of the angular momentum quantum number ℓ, that involved the maximum value of ℓ in a given correlation consistent basis set in a two-point extrapolation, was proposed by Helgaker and co-workers[22,23] for conventional correlation methods

$$E_n^{\mathrm{corr}} = E_{\mathrm{CBS}}^{\mathrm{corr}} + A\ell_{\mathrm{max}}^{-3}, \tag{1}$$

where, conveniently, for first and second row elements ℓ_{max} is also equal to the cardinal number n of the correlation consistent basis sets. Many extrapolation schemes follow a similar, yet more complex, functional form and have been recently reviewed by Varandas.[24] Some notable contributions include scaling singlet and triplet-pair contributions individually[25] and treatments specifically targeting extrapolations from small basis sets.[26] It has been previously noted that the correlation energies obtained from explicitly correlated methods should formally converge as ℓ_{max}^{-7} rather than ℓ_{max}^{-3}.[13,21] Yamaki *et al.*[27] utilized extrapolations based on this much more rapid convergence behavior in F12 calculations, but these met with limited success due to the exceedingly stringent demands this places on the underlying basis set, i.e., each angular momentum level must be very well converged with each set, which is not sufficiently satisfied by correlation consistent basis sets. Schwenke has proposed[28] a generalized form of a two-point CBS extrapolation

$$E_{\mathrm{CBS}}^{\mathrm{corr}} = (E_{\mathrm{large}}^{\mathrm{corr}} - E_{\mathrm{small}}^{\mathrm{corr}})F + E_{\mathrm{small}}^{\mathrm{corr}}, \tag{2}$$

where $E_{\mathrm{large}}^{\mathrm{corr}}$ and $E_{\mathrm{small}}^{\mathrm{corr}}$ are the correlation energies evaluated with the larger and smaller of two basis sets, respectively, and F is some function. While F can potentially take any form, Schwenke used accurate estimates of the (conventional) correlation energies of a few small molecules along with a least-squares fit to produce a set of coefficients to use as F in the extrapolation of coupled cluster energies. Results of these extrapolations on closed-shell, first row atom-containing molecules produced correlation energies that were closer to the literature best estimates than l_{max}^{-3} power-law-based extrapolations.[28] The results of Eq. (2), however, can be easily related to functional forms such as Eq. (1), but with a noninteger exponent of ℓ_{max}. In fact, starting from a more general form of Eq. (1) recast as a two-point extrapolation

$$E_{\mathrm{CBS}}^{\mathrm{corr}} = \frac{E_{\mathrm{large}}^{\mathrm{corr}}\ell_{\mathrm{large}}^{\mathrm{pow}} - E_{\mathrm{small}}^{\mathrm{corr}}\ell_{\mathrm{small}}^{\mathrm{pow}}}{\ell_{\mathrm{large}}^{\mathrm{pow}} - \ell_{\mathrm{small}}^{\mathrm{pow}}}, \tag{3}$$

where pow is an arbitrary, in principle noninteger, exponent, it is reasonably straightforward to show that

$$F = \frac{\ell_{\mathrm{large}}^{\mathrm{pow}}}{\ell_{\mathrm{large}}^{\mathrm{pow}} - \ell_{\mathrm{small}}^{\mathrm{pow}}}. \tag{4}$$

Hence, as discussed previously be Schwenke,[28] Eq. (2) is a more general form of Eq. (1), and for particular values of F the two expressions give exactly the same results. Crawford and Valeev recently reported Schwenke-style CBS extrapolations of CCSD(T)$_{\overline{R12}}$ correlation energies,[29] but unlike the present work their study was limited to molecules containing just first row elements, all electrons were correlated, and the CBS coefficients were determined from an analysis of just five molecules.

It is also well known that basis set extrapolations produce more accurate results when larger basis sets are employed; the use of double-ζ basis sets in two-point power-law-based extrapolation schemes of the correlation energy using conventional methods introduces a significant source of error and it is desirable to include data only from basis sets with higher cardinal numbers—those equal to or greater than five produce the most accurate estimates of the CBS limit.[22] The present work will show that combining explicitly correlated MP2-F12 methods with coefficients optimized in a Schwenke-type generalized basis set extrapolation produces an average error in the correlation energy of less than half a millihartree (mE_{h}), relative to accurate estimates of the CBS limit, from just double- and triple-ζ basis sets. Increasing the basis set sizes by just one cardinal number reduces the errors to below 0.1 mE_{h}, which is smaller than the precision of the reference CBS MP2 estimates employed. It will also be shown that while the extrapolation of CCSD(T)-F12b correlation energies is less accurate than with the MP2-F12 method with smaller basis sets, the average errors still approach 0.1 mE_{h} when triple-ζ or larger basis sets are utilized. How to exploit these results for accurate energy differences, which is of extreme importance for *ab initio* thermochemistry methodologies, will the subject of a subsequent publication.

II. COMPUTATIONAL PROCEDURE

All explicitly correlated second-order perturbation theory calculations were carried out with the DF-RMP2-F12[30,31] or MP2-F12 (Ref. 30) implementations

in the MOLPRO (ref. 32) package of *ab initio* programs. Coupled cluster calculations utilized the approximate CCSD (T)-F12b implementation[33,34] in the same package. Three related explicitly correlated *Ansätze* have been investigated in the present work: (a) the general, orbital invariant 3C *Ansatz*,[30] (b) the diagonal 3C(D) *Ansatz*,[30,31] which is not orbital invariant (and size consistent only when localized orbitals are used) but does not suffer from geminal basis set superposition errors (GBSSE) as in the previous case, and (c) the diagonal, fixed amplitude *Ansatz* 3C(FIX) whereby in addition to the diagonal approximation the amplitudes of the explicitly correlated configurations are determined by the wave function cusp conditions.[16] This latter *Ansatz* is orbital invariant, size consistent, and free of GBSSE. Explicitly correlated calculations carried out with the 3C(D) *Ansatz* did so with orbitals localized via the Pipek–Mezey[35] procedure.

For the majority of calculations the orbital basis sets employed were either the cc-pV*n*Z-F12 basis sets[36] for both first and second row elements (and H), or the aug-cc-pV*n*Z basis sets[8] for first row elements (and H) along with the aug-cc-pV($n+d$)Z basis sets[37] for second row elements, where "+*d*" denotes the inclusion of an additional tight *d* function. Herein these basis sets will be referred to as V*n*Z-F12 and AV*n*Z, respectively, for the sake of brevity (basis sets used for reference correlation energies are described in the next section). For the V*n*Z-F12 and AV*n*Z orbital basis sets the OptRI auxiliary basis sets of Yousaf and Peterson[38] were used as the RI basis for the many-electron integrals within the CABS approach of Valeev[39] as implemented in MOLPRO,[30,31] while the DF of the Fock and exchange matrices employed the cc-pV*n*Z/JKFIT sets of Weigend.[40] The DF of the remaining integral quantities utilized the AV*n*Z/MP2FIT sets of Weigend *et al.*[41] in the AVDZ, AVTZ and AVQZ cases, and the AV5Z/MP2FIT basis of Hättig[42] for the AV5Z orbital basis. The values of the geminal Slater exponent β used with the 3C and 3C(D) *Ansätze* match those recommended in previous work,[36] namely, 0.9, 1.0, and 1.1 a_0^{-1} for VDZ-F12, VTZ-F12, and VQZ-F12, respectively, and 1.1 a_0^{-1} for AVDZ, 1.2 a_0^{-1} for AVTZ, and 1.4 a_0^{-1} for both AVQZ and AV5Z.

In order to be consistent with other benchmark studies, all geometries used corresponded to those optimized using the CCSD(T) method with the cc-pCVQZ basis sets[9,43] that include additional exponents for describing core-valence effects. The frozen-core approximation, which was utilized in all other calculations in this work, was removed for these optimizations (except that the 1*s* electrons of second row elements were not correlated). The optimization of the coefficient *F* to be used in the Schwenke-type extrapolations was carried out with the general function minimizer in MOLPRO.

III. RESULTS AND DISCUSSION

A. Calculation of reference correlation energies

Previous investigations into the accurate extrapolation of correlation energies by Helgaker *et al.*,[23] Schwenke,[28] and Klopper[25] have centered on a small set of first row systems: Ne, N_2, CH_2 ($^1A_1$), H_2O, CO, HF, and F_2. For all of these systems Klopper and co-workers[25,44] have calculated high accuracy correlation energies with both CCSD-R12 and MP2-R12 methods, where R12 indicates that terms linear in r_{ij} are included explicitly in the wave function, using large uncontracted Gaussian basis sets to estimate the CBS limit. In order to investigate accurate correlation energy extrapolations of molecules containing second row elements, comparable explicitly correlated calculations with the MP2-F12 and CCSD(T)-F12b methods were carried out in this work. In addition to the second row analogs of the molecules listed above (Ar, P_2, SiH_2, H_2S, SiS, HCl, and Cl_2), MP2-F12 reference data was also obtained for systems including group thirteen elements (BH, AlH, BN, and AlP) and molecules containing both first and second row elements: PN, SiO, CS, ClF, and AlN. The lowest closed-shell singlet electronic states of all molecules were used throughout. To ensure the procedure produces correlation energies commensurate to those of Klopper, benchmark correlation energies were also computed for the original seven first row molecules. An alternative route for the production of reference correlation energies could be via the extrapolation of energies from large basis set conventional calculations. However, a conscious decision has been made in this work to avoid extrapolated reference data, as high quality data for molecules including second row elements is very scarce.

The orbital basis sets for the reference energy calculations on the second row atom-containing molecules were based on an AV6Z (22*s*15*p*) set of primitives[45] generally contracted to [10*s*9*p*]. Exponents for the higher angular momentum functions were optimized at the DF-RMP2-F12/3C level of theory in an even tempered manner[46] using the prescription $\zeta_i = \alpha B^{i-1}$, where α is the most diffuse exponent and *B* is the progression between consecutive exponents. In order to saturate the basis, additional functions were optimized and added to each angular momentum until the incremental energy change for adding the next function fell below 0.004 mE_h, resulting in a 12*d*10*f*8*g*6*h* correlating set. The auxiliary basis sets used for the DF of all the integrals except those of the Fock and exchange matrices were also even tempered with a total of 29*s*27*p*23*d*20*f*18*g*17*h*15*i* uncontracted functions. The criteria followed for the construction of these sets was that the DF error, defined as the difference between the conventional MP2 and DF-MP2 correlation energies, should be less than 0.001 mE_h for the atom and less than 0.02 mE_h for homonuclear diatomics; these criteria were met and in the majority of cases the errors fell below these criteria. The reference orbital basis sets for the first row elements were obtained following the same procedure as the second row, with an AV6Z *s* and *p* core,[47] (17*s*11*p*)/[10*s*9*p*], and an optimized set of 9*d*9*f*8*g*7*h* even tempered functions. The non-Fock matrix DF auxiliary sets had 28*s*26*p*22*d*22*f*20*g*18*h*15*i* uncontracted functions. The reference hydrogen orbital basis used was based on the AV6Z basis contracted to [7*s*6*p*5*d*4*f*3*g*2*h*] extended in an even tempered manner with a set of 2*p*1*d*1*f*1*g*1*h* tight functions, along with a 21*s*18*p*14*d*12*f*10*g*8*h*6*i* DF auxiliary set. In all cases the DF of the Fock and exchange matrices used the cc-pV5Z/JKFIT auxiliary basis. In this latter case, the large non-Fock matrix DF auxiliary set was also investigated for this purpose, but these results differed from the cc-pV5Z/

TABLE I. MP2-F12 frozen-core correlation energy contributions from each angular momentum quantum number for Ar, Ne, and P_2. Values in parentheses indicate the incremental change in correlation energy with the addition of basis functions in each angular momenta.

Angular momentum	No. of functions	Total correlation energy (mE_h) *Ansatz* 3C	Total correlation energy (mE_h) *Ansatz* 3C(FIX)
		Ar	
s and p	$(22s15p)/[10s9p]$	−196.636	−170.517
d	12	−248.204(−51.568)	−242.932(−72.415)
f	10	−253.161(−4.957)	−252.024(−9.092)
g	8	−254.299(−1.138)	−254.092(−2.068)
h	6	−254.490(−0.191)	−254.439(−0.347)
CBS (ℓ_{max}^{-7})[a]		−254.54	−254.53
		Ne	
s and p	$(17s11p)/[10s9p]$	−302.041	−298.493
d	9	−317.989(−15.948)	−317.817(−19.325)
f	9	−319.826(−1.837)	−319.803(−1.985)
g	8	−320.121(−0.295)	−320.115(−0.313)
h	7	−320.169(−0.048)	−320.168(−0.052)
CBS (ℓ_{max}^{-7})[a]		−320.18	−320.18
		P_2	
s and p	$(22s15p)/[10s9p]$	−278.391	−240.648
d	12	−311.552(−33.161)	−301.945(−61.297)
f	10	−318.588(−7.036)	−316.468(−14.524)
g	8	−319.972(−1.384)	−319.478(−2.989)
h	6	−320.290(−0.318)	−320.136(−0.678)
CBS (ℓ_{max}^{-7})[a]		−320.37	−320.31

[a]Obtained using correlation energies calculated with the g and h basis sets. See the text.

JKFIT results by only ~10 μE_h. Following the spirit of the OptRI basis sets,[38] the goal for the complementary auxiliary basis sets used in the RI was that after merging with the contracted orbital basis a minimal number of functions should be deleted in the CABS procedure[39] to minimize numerical instabilities while ensuring small RI errors. The CABS basis used in the construction of the RI consisted of even tempered sets of diffuse and tight functions based on values of α and B extracted from the reference orbital basis sets, as well as exponents that intercalated pairs of existing uncontracted functions in the reference orbital basis where the gap was sufficiently large. A full specification of the reference orbital and auxiliary basis sets can be found in the supporting information.[48]

The MP2-F12 correlation energy contributions for each angular momentum quantum number are summarized in Table I for Ne, Ar, and P_2 [for both the 3C and 3C(FIX) *Ansätze*], along with the number of even tempered functions included for each ℓ. The geminal exponent β was fixed to a value of 1.4 in these cases, which was very close to the optimal value for the reference basis set. Table I shows that the amount of correlation energy recovered for each successive angular momentum quantum number drops by around an order of magnitude, suggesting that i functions should contribute less than 0.04 mE_h to the MP2-F12 correlation energy. A comparison of the *Ansätze* indicates that the use of fixed amplitudes produces significantly smaller (in magnitude) correlation energies when only s and p functions are included in the basis set and as the basis set size increases it converges toward the limit more slowly than when the amplitudes are optimized variationally. As the basis sets head toward the CBS limit, however, the correlation energies from the different *Ansätze* do begin to converge.

Since the above reference basis sets are well saturated in each angular momentum, the correlation energies of Table I were also best fit to the expression $E_{corr}=E_{CBS}+A/\ell_{max}^7$. The accuracy of the resulting CBS limits were strongly dependent on the range of ℓ_{max} used: for the Ar atom $E_{CBS}=-254.13$ mE_h (d through h), −254.50 (f through h), and −254.54 (g and h only). The latter value, which is only 0.05 mE_h lower than the reference basis set result obtained with just up through h-type functions, is in good agreement with the finite element result (−254.62 mE_h) of Ref. 49. Similarly, little additional gain (−0.01 mE_h) in correlation energy is obtained via extrapolation of the Ne results, and these are within 0.04 mE_h of the previous finite element value. From the data in Table I, the 3C *Ansatz* was selected to produce reference data in this investigation as in all cases it exhibits the most rapid convergence with basis set as a function of ℓ_{max} and produces the lowest correlation energies. This *Ansatz* is both size extensive and orbital invariant, and additionally GBSSE is not a factor for the total correlation energy of an atom or small molecule.

Table II contains correlation energy contributions for each angular momentum quantum number for Ne, Ar, and P_2 calculated at the CCSD-F12b level of theory using the reference basis sets introduced above (also with $\beta=1.4$). A comparison with Table I shows that while the incremental change in the correlation energy with the addition of d-type functions is very similar between MP2-F12 and CCSD-F12b, the latter curiously converges much slower than MP2-F12 for the higher angular momentum functions. This could possibly be due to the approximate nature of the F12b method compared to CCSD-F12. The result of varying the *Ansätze* is that, just as with MP2-F12, fixed amplitudes result in correlation energies that are smaller in magnitude with low angular momentum basis sets, but in the CCSD-F12b case the 3C and 3C(FIX) *Ansätze* have very similar convergence rates in the case of Ne when high angular momentum functions are introduced. For both Ar and P_2, however, it can be seen that the 3C(FIX) *Ansatz* appears to converge toward the apparent CBS limit more rapidly with basis than 3C, where in the latter case it should be stressed that the amplitudes are optimized at the MP2-F12 level and then fixed in the CCSD-F12b calculation. *Ansatz* 3C(FIX) has been selected to produce CCSD-F12b reference data in this investigation and this is also the default *Ansatz* for this method in MOLPRO. Also shown in Table II are ℓ_{max}^{-7} extrapolations of the CCSD-F12b correlation energies analogous to the MP2-F12 reference results of Table I. Only small improvements over the reference basis set results are observed for the Ne atom, but for both Ar and P_2 the resulting CBS limits do depend on which *Ansatz* is used, with 3C producing too small (in magnitude) CCSD correlation energy limits compared to 3C(FIX) by as much as 0.39 mE_h in the case of P_2.

TABLE II. CCSD-F12b frozen-core correlation energy contributions from each angular momentum quantum number for Ar, Ne, and P_2. Values in parentheses indicate the incremental change in correlation energy with the addition of basis functions in each angular momenta.

Angular momentum	No. of functions	Total correlation energy (mE_h) *Ansatz* 3C	Total correlation energy (mE_h) *Ansatz* 3C(FIX)
		Ar	
s and *p*	$(22s15p)/[10s9p]$	−196.329	−171.212
d	12	−246.736(−50.407)	−245.151(−73.939)
f	10	−259.686(−12.950)	−260.305(−15.155)
g	8	−264.143(−4.456)	−264.630(−4.325)
h	6	−264.635(−0.493)	−264.868(−0.238)
CBS (ℓ_{max}^{-7})[a]		−264.77	−264.93
		Ne	
s and *p*	$(17s11p)/[10s9p]$	−290.308	−286.770
d	9	−308.225(−17.916)	−307.968(−21.197)
f	9	−313.978(−5.753)	−313.919(−5.952)
g	8	−315.439(−1.461)	−315.390(−1.470)
h	7	−315.622(−0.183)	−315.596(−0.207)
CBS (ℓ_{max}^{-7})[a]		−315.67	−315.65
		P_2	
s and *p*	$(22s15p)/[10s9p]$	−262.471	−231.534
d	12	−308.708(−46.237)	−306.644(−75.109)
f	10	−319.479(−10.770)	−320.723(−14.080)
g	8	−321.895(−2.416)	−322.784(−2.060)
h	6	−322.510(−0.615)	−323.005(−0.221)
CBS (ℓ_{max}^{-7})[a]		−322.67	−323.06

[a]Obtained using correlation energies calculated with the *g* and *h* basis sets. See the text.

TABLE III. Reference frozen-core MP2 correlation energies.

Molecule	MP2-F12/3C (mE_h)	MP2-R12[a] (mE_h)	Finite elementMP2[b] (mE_h)
Ne	−320.17	−320.1	−320.22
N_2	−420.94	−421.0	
CH_2 ($^1A_1$)	−155.91	−155.9	
H_2O	−300.53	−300.5	
CO	−403.92	−403.9	
HF	−319.75	−319.7	
F_2	−611.74	−611.7	
Ar	−254.49		−254.62
P_2	−320.29		
SiH_2	−123.55		
H_2S	−221.37		
SiS	−303.23		
HCl	−242.20		
Cl_2	−457.32		
PN	−392.93		
SiO	−385.00		
CS	−346.83		
ClF	−529.94		
BH	−82.04		
AlH	−69.63		
BN ($^1\Sigma^+$)	−369.99		
AlP ($^1\Sigma^+$)	−243.91		
AlN ($^1\Sigma^+$)	−289.23		

[a]Average of MP2-R12/A and MP2-R12/B energies of Klopper (Ref. 25), as presented in Ref. 15.
[b]Finite element MP2 energies of Flores (Ref. 49).

Table III lists the MP2-F12/3C correlation energies for the current set of reference systems with the saturated basis sets described above (β=1.4). It can be seen that the agreement with the averaged MP2-R12/A and MP2-R12/B correlation energies of Klopper[25] for molecules involving first row elements is very good, agreeing to within 0.1 mE_h. The agreement with finite element MP2 correlation energies for Ne and Ar is also excellent. This level of agreement indicates that the new reference data presented in Table III should be accurate to within around 0.1 mE_h of the valence shell MP2 CBS limit. Bischoff *et al.*[50] recently published near basis set limit MP2-F12/2B correlation energies for a set of first row molecules and again the agreement with the results of the current investigation are within 0.1 mE_h. It should be noted, however, that there is a relatively large geometry effect for N_2 and CO, and calculation of these correlation energies at bond lengths corresponding to the all-electron CCSD(T)/cc-pCVTZ level of theory, as used by Bischoff *et al.*,[50] was required to produce comparable correlation energies.

The CCSD-F12b correlation energy contributions calculated with the reference basis sets (β=1.4) are presented in Table IV for the first row systems and their second row analogs. Initial calculations using the CCSD-F12a approximation consistently overestimated the correlation energy by significant amounts, even with the large reference basis set, hence only -F12b results are reported. In the case of the molecules containing just first row elements, the agreement between the CCSD-F12b/3C(FIX) correlation energies shown in Table IV and the previous CCSD-R12/B results of Fliegl *et al.*[44] (Ref. 25 in the case of Ne) is very good, particularly when the same geometries are used in both cases. In all polyatomic cases the energy differences are within 0.1 mE_h, except for F_2 where the present correlation energy is lower by −0.15 mE_h. It is not clear whether this difference is due to the CCSD-F12b treatment or remaining basis set incompleteness in the previous study by Fliegl *et al.*[44] It is perhaps notable, however, that the MP2-R12/B results also reported in the work in Ref. 44 for F_2 lies above the very accurate F12/3C result of Table III by +0.4 mE_h. The difference between the CCSD-F12b correlation energy for neon and the rigorous CBS estimate of Barnes *et al.*[51] is just 0.04 mE_h, while the Klopper result[25] lies above the Barnes *et al.* value by +0.16 mE_h. Definitive CCSD reference data for comparison of the results for second row species were not found in the literature. However, recent conventional CCSD calculations[52] on Ar using basis sets as large as aug-cc-pV8Z together with separate singlet- and triplet-pair CBS extrapolations[25] yield a correlation energy within 0.14 mE_h of the CCSD-F12b result shown in Table IV. Based on these results and the data in Table II, the uncertainty in the present CCSD reference CBS limits is conservatively estimated to be about 0.2 to 0.3 mE_h (with the largest uncertainties assigned to the second row atom-containing molecules).

The CCSD(T)-F12 approximations utilized in this work do not treat the perturbative triples in an explicitly correlated form, and thus the (T) energy converges with respect to basis

TABLE IV. Reference frozen-core CCSD and (T) correlation energies (mE_h).

Molecule	CCSD-F12b/3C(FIX)	(T)[a]	CCSD-R12/B[b]	(T)[c]	CCSD[d]	(T)[d]
Ne	−315.60	−6.43	−315.48		−315.64	−6.50
N_2	−407.50	−21.19		−21.20		
	−407.62[e]		−407.60			
CH_2 ($^1A_1$)	−175.54	−5.63		−5.64		
	−175.57[e]		−175.58			
H_2O	−298.00	−9.80		−9.77		
	−298.05[e]		−298.00			
CO	−395.78	−19.48		−19.49		
	−395.92[e]		−395.88			
HF	−313.99	−8.76		−8.78		
	−314.06[e]		−313.99			
F_2	−601.27	−22.74		−22.67		
	−601.04[e]		−600.89			
Ar	−264.87	−9.83				
P_2	−323.00	−24.99				
SiH_2	−144.96	−4.37				
H_2S	−237.07	−10.11				
SiS	−311.88	−20.99				
HCl	−254.10	−10.59				
Cl_2	−473.41	−23.71				

[a]The (T) contributions were calculated with conventional CCSD(T) using the reference basis with additional *i* functions. See the text.
[b]CCSD-R12/B energies of Fliegl *et al.* (Ref. 44) as reported by Bakowies (Ref. 59). These were calculated at geometries obtained at the frozen-core CCSD(T)-R12 level of theory as reported in Ref. 4. Ne correlation energy as reported in Ref. 25.
[c]Conventional CCSD(T) using *f*-limit basis sets of Schwenke (Ref. 28).
[d]CBS limit data of Barnes *et al.* (Ref. 51) Cited uncertainties: CCSD ±0.018 mE_h and (T) ±0.004 mE_h.
[e]Calculated at the same CCSD(T)-R12 geometry as Fliegl *et al.* (Ref. 44).

set size just as in a conventional CCSD(T) calculation. Including *i*-type functions in the orbital basis set would aid the minimization of the basis set error in the triples, yet this would require *k* functions in the DF auxiliary set if explicit correlation was to be employed. Such high angular momentum functions are currently unsupported in MOLPRO and angular momentum truncation of DF sets has been shown to produce a significant overestimation of the correlation energy in the DF-MP2 case.[53] To obtain reference values for the (T) contributions to the correlation energy, four even tempered *i* functions were optimized at the RMP2 level and added to the reference sets just for these calculations, which were carried out at the conventional CCSD(T) level of theory. As shown in Table IV, the agreement between the (T) energy contributions calculated using the reference basis with additional *i* functions and those calculated by Schwenke[28] using his *f*-limit basis sets is excellent, typically within 0.01 mE_h. In two cases, H_2O and F_2, the (T) correlation energy contributions currently presented are lower in energy than those of Schwenke, by 0.03 and 0.07 mE_h, respectively. In both of these cases, however, the *f*-limit basis sets used by Schwenke had $\ell_{max}=5$, as opposed to his other results where $\ell_{max}=6$, thus a slight underestimation of his (T) energy is to be expected.

B. Selection of DF auxiliary basis sets for MP2-F12 and CCSD(T)-F12b

With the advent of the OptRI auxiliary basis sets[38] the choice of RI basis in explicitly correlated calculations involving first and second row elements has become clearer. However, while the selection of auxiliary basis sets for the DF of both the Fock and exchange matrices and the remaining integrals has been examined briefly for both MP2-F12 and approximate CCSD-F12 methods in previous studies,[34,54] neither of these investigations considered the V*n*Z-F12 orbital basis sets and in the present work a larger range of auxiliary basis sets is considered. In this work a series of calculations were carried out on the first row systems Ne, CO, HF, and their second row analogs Ar, SiS, and HCl for the V*n*Z-F12 (n=D, T, Q) and AV*n*Z (n=D, T, Q, 5) orbital and OptRI basis sets. Every combination of the cc-pV*n*Z/JKFIT (used for DF of the Fock and exchange terms) and AV*n*Z/MP2FIT (used for DF all other integrals) auxiliary basis sets were tested and each resulting correlation energy was compared against the energy for the same orbital/RI basis set pair with the cc-pV5Z/JKFIT and aug-cc-pwCV5Z/MP2FIT (Ref. 42) auxiliary basis sets, working on the assumption that a larger auxiliary basis will produce a more accurate fit. While selecting auxiliary basis sets on the grounds of the errors introduced by their selection in the DF approximation is the goal of this part of the investigation, this also has to be balanced with computational cost, as the larger the size of the auxiliary basis the longer a given calculation will take. The systems under investigation here are too small to produce effective timing data, hence the criteria used is for the fitting sets to be as small as possible while attaining the desired accuracy. The results of this testing with DF-RMP2-F12 are illustrated in

Tables SI and SII (found in the supporting information) as statistics of the error relative to the largest fitting sets: the mean absolute error, the standard deviation, and the maximum error.

For both the V*n*Z-F12 and AV*n*Z orbital basis sets it is immediately obvious that an MP2FIT auxiliary basis with a cardinal number lower than that in the orbital basis cannot be used without introducing significant errors, and in some more extreme cases this even leads to numerical instabilities. Using an MP2FIT basis in MP2-F12 calculations with the same cardinal number as the orbital basis generally produces an average error of only about 0.01–0.04 mE_h when combined with a JKFIT basis also possessing the same cardinal number. The exceptions to this are the double-ζ orbital sets as they have mean absolute errors of 0.06 and 0.07 mE_h for VDZ-F12 and AVDZ, respectively, with the AVDZ/MP2FIT and cc-pVTZ/JKFIT auxiliary sets (cc-pVDZ/JKFIT auxiliary basis sets are not available). While these average errors may seem small relative to the expected accuracy of the correlation energy produced with such small basis sets, the maximum errors in these cases are approximately 0.2 mE_h, which seems overly large when the accuracy of the target reference data is 0.1 mE_h. Simply increasing the MP2FIT basis to AVTZ has little effect on the average error but reduces the maximum error to under 0.1 mE_h. Thus the AVTZ/MP2FIT and cc-pVTZ/JKFIT auxiliary basis sets are recommended in conjunction with both the VDZ-F12 and AVDZ orbital sets and will be used in those cases herein.

For the quadruple- and quintuple-ζ orbital sets it can be seen that increasing the cardinal number of the JKFIT auxiliary basis sets decreases the errors in every case, sometimes by more than an order of magnitude. A noticeable reduction in error, however, is not observed when increasing the MP2FIT basis to AV5Z for the quadruple-ζ orbital sets; in fact some of the error statistics increase. As the errors appear to be well controlled it seems prudent to use the AV*n*Z/MP2FIT and cc-pV*n*Z/JKFIT auxiliary sets that share the same cardinal number as the orbital set for the triple-, quadruple-, and quintuple-ζ cases. This is also currently the default scheme in the MOLPRO program package. To summarize, utilizing the auxiliary basis set combinations described above yield mean absolute errors in the correlation energy of 0.078, 0.019, and 0.006 mE_h for the VDZ-F12, VTZ-F12, and VQZ-F12 orbital sets, respectively, and 0.071, 0.042, 0.008, and 0.000 mE_h for the AVDZ, AVTZ, AVQZ, and AV5Z orbital sets, respectively. Similar results are found when CCSD(T)-F12b is employed for the SiS molecule, hence the same auxiliary basis sets are recommended for both methods.

C. Choice of geminal Slater exponent for the fixed amplitude *Ansatz*

Recommended values of β, the geminal Slater exponent, were previously determined for the 3C *Ansatz* with V*n*Z-F12 and AV*n*Z orbital basis sets by inspection of the basis set error per correlated electron as a function of β.[36] If the diagonal, fixed amplitude *Ansatz*,[16] 3C(FIX), is selected, the amplitudes of the explicitly correlated configurations are held fixed at those determined from the cusp conditions. The error introduced by not optimizing the amplitudes can be minimized by using an optimal value of β, but as shown in an investigation using the CCS(F12) method (a variant of explicitly correlated CC2 theory where conventional double excitations have been removed) the dependence of the correlation energy on the geminal Slater exponent is much greater in the case of fixed amplitudes than when they are optimized variationally, and the optimal value of β differs between the two.[55] A previous investigation into the effect of β on CCSD(T)-F12 energies focused on relative energies in the form of reaction energies, atomization energies, ionization potentials, etc., and suggested a value of 1.0 a_0^{-1} for AVDZ and AVTZ basis sets.[34] For basis set extrapolations, however, it seems logical that a value of β that produces correlation energies as close as possible to the minimum values for a given system should be utilized. In addition this should also remove any reliance on fortuitous cancellation of errors when dealing with relative energies.

TABLE V. Recommended values of the geminal Slater exponent β, using V*n*Z-F12 and AV*n*Z basis sets with the MP2-F12 method and the 3C(FIX) *Ansatz*. Also reported are the mean signed errors per correlated electron (MSE) for using these values of β (relative to the optimal value for the given molecule) and the standard deviation of that error (Std. dev.). See text for further details.

Basis	β (a_0^{-1})	MSE (mE_h)	Std. dev.	MSE with β=1.4
VDZ-F12	0.9	0.136	0.153	0.674
VTZ-F12	1.0	0.021	0.021	0.120
VQZ-F12	1.0	0.004	0.004	0.033
AVDZ	1.0	0.246	0.308	0.674
AVTZ	1.2	0.067	0.078	0.093
AVQZ	1.4	0.016	0.016	0.016
AV5Z	1.5	0.006	0.006	0.006

While the optimal value of β is dependent on both the molecule and orbital basis set, it is impractical in general calculations to optimize β for a given molecule under investigation. In order to reduce this to a specific, compromise value of β for each basis set, optimal values for the first and second row molecular systems, along with the group 13 molecules listed above, were obtained by scanning the MP2-F12/3C(FIX) correlation energy over six different values of β for each system. A polynomial interpolation of these energies was then carried out to produce the minimum energy and the corresponding optimal geminal exponent. The recommended value of β is then finalized by finding the mean over all systems. Table V displays the recommended value of β for each basis set, along with the mean error per correlated electron relative to the minimum energy found by optimizing β individually for each system and the standard deviation of that error. To emphasize the effect of the choice of β on correlation energies with the 3C(FIX) *Ansatz*, Table V also includes mean errors per correlated electron when β is fixed to 1.4 a_0^{-1} for the same test set. In general this somewhat arbitrary (in this case) choice of β, similar values of which have been used in several places in the literature (for example, Refs. 29, 50, and 56), leads to significantly larger

TABLE VI. CBS coefficients for extrapolation of MP2-F12/3C correlation energies using Eqs. (2) and (4) along with statistics of the error in the extrapolated energy relative to reference data (mE_h).

Basis sets	Equation (2) F	Equation (5) pow	MAE	Standard deviation	Max error
First row systems: Ne, N_2, CH_2, H_2O, CO, HF and F_2					
VDZ-F12/VTZ-F12	1.431 442	2.957 854	0.121	0.105	0.267
VTZ-F12/VQZ-F12	1.419 992	4.234 613	0.011	0.006	0.022
AVDZ/AVTZ	1.769 165	2.054 455	0.246	0.175	−0.531
AVTZ/AVQZ	1.636 003	3.283 305	0.070	0.055	−0.148
AVQZ/AV5Z	1.758 345	3.769 156	0.020	0.024	−0.063
Second row systems: Ar, P_2, SiH_2, H_2S, SiS, HCl and Cl_2					
VDZ-F12/VTZ-F12	1.519 031	2.648 547	0.132	0.101	0.296
VTZ-F12/VQZ-F12	1.456 426	4.032 928	0.064	0.065	0.160
AVDZ/AVTZ	1.752 674	2.084 436	0.434	0.373	1.072
AVTZ/AVQZ	1.827 822	2.753 182	0.230	0.118	−0.363
AVQZ/AV5Z	1.636 829	4.230 770	0.042	0.040	0.100
Combined first and second row systems					
VDZ-F12/VTZ-F12	1.489 006	2.746 297	0.185	0.132	0.436
VTZ-F12/VQZ-F12	1.446 336	4.087 183	0.048	0.047	0.173
AVDZ/AVTZ	1.758 592	2.073 695	0.348	0.290	1.037
AVTZ/AVQZ	1.747 998	2.950 594	0.242	0.162	0.561
AVQZ/AV5Z	1.669 114	4.096 907	0.059	0.043	0.160

errors in the correlation energy, by greater than a factor of 5 in some cases.

It can also be seen from Table V that, within a series of basis sets, as the size of the basis set is increased the recommended value of β also increases, particularly for the AVnZ basis sets. This has been noted previously,[15,34,36] as when larger, more flexible one-electron basis sets are used only the areas very near the correlation cusp need to be corrected with explicit correlation, as opposed to both the cusp and tails with smaller basis sets. The increase in β is accompanied by a decrease in the mean error per correlated electron and the standard deviation of the error. The VnZ-F12 basis sets have a smaller sensitivity to the choice of β than the AVnZ basis sets with an equivalent cardinal number.

A similar analysis of optimizing β for the CCSD-F12b method was also carried out. In this case, however, the optimal values turned out to be much larger than expected, e.g., up to values of 2.4 with the AV5Z basis set and about 2.6 for the large reference basis set. These results were at odds with previous results by Tew *et al.*[57] using their CCSD(F12) method, where the optimal values of β were very similar to those at the MP2-F12 level of theory. Subsequent CCSD-F12b test calculations using these relatively large values of β were then carried out on both Ne and F_2 with the reference orbital basis set. In both cases the CCSD correlation energies appeared to be too low by 0.1–0.2 mE_h. Thus the large values of β obtained by minimizing the CCSD-F12b correlation energy seem to be an artifact of the F12b approximation. Hence the recommendation of this work is to simply use the optimal MP2-F12/3C(FIX) values of β given in Table V for all CCSD-F12b calculations, and this scheme has been employed throughout the present work.

D. Extrapolation of correlation energies using series of V*n*Z-F12 and AV*n*Z basis sets

1. MP2 and MP2-F12

Using the combinations of auxiliary and orbital basis sets described in Sec. III B above, a series of MP2-F12/3C calculations were carried out on the molecules listed in Table III. For each pair of orbital basis sets in each series with consecutive cardinal numbers, the coefficient F in Eq. (2) was optimized by minimizing the RMS error with respect to the reference correlation energies also presented in Table III. An exponent (pow) was also optimized for a more general expression of the two-point power-law-based extrapolation of Eq. (1)

$$E_{\ell_{\max}}^{\mathrm{corr}} = E_{\mathrm{CBS}}^{\mathrm{corr}} + A\ell_{\max}^{-\mathrm{pow}} \tag{5}$$

where $\ell_{\max}$ is equal to the cardinal number of the orbital basis set. The optimized coefficients (F and pow) are reported in Table VI, along with the mean absolute error (MAE), standard deviation of the error, and the maximum error, divided into first row systems (Ne, N_2, CH_2, H_2O, CO, HF, and F_2), second row systems (Ar, P_2, SiH_2, H_2S, SiS, HCl, and Cl_2), and all of these combined. The parameters F and pow have been optimized individually for convenience, but the error statistics presented in Table VI are equivalent for both extrapolation schemes [see Eq. (4)]. It is noted that the statistics were produced with the Schwenke-style extrapolation, but in all cases the worst agreement between maximum errors from the two methods was less than a μE_h, which can be attributed to slight numerical differences.

From Table VI it can be seen that, unsurprisingly, using larger basis sets within a series produces a more accurate extrapolation in all cases. A comparison of the data for first and second row systems reveals that the optimized coeffi-

TABLE VII. Optimized CBS extrapolation coefficients for MP2-F12/3C and error statistics for an enlarged test set of systems that includes group 13 elements and systems containing a mixture of first and second row elements (see text for further details). All errors reported in mE_h. These are the recommended values for this method.

Basis sets	Equation (2) F	Equation (5) pow	MAE	Standard deviation	Max error
VDZ-F12/VTZ-F12	1.489 285	2.745 197	0.165	0.140	0.487
VTZ-F12/VQZ-F12	1.449 074	4.071 516	0.041	0.039	0.170
AVDZ/AVTZ	1.754 953	2.080 361	0.372	0.302	1.058
AVTZ/AVQZ	1.736 098	2.982 639	0.199	0.155	0.609
AVQZ/AV5Z	1.665 414	4.111 526	0.050	0.041	0.165

cients for the extrapolations differ considerably between the two rows, suggesting that coefficients optimized for extrapolations of first row elements are unlikely to provide as accurate results for extrapolation of second row systems. The only exception to this is the extrapolation with AVDZ and AVTZ basis sets, where the coefficients remain reasonably similar. It is also interesting to note that in general the errors in the extrapolated CBS limits are larger by almost a factor of two for the second row systems compared to the first row, even with coefficients specifically optimized for each row. When the coefficients are optimized for the combined first and second row systems, the mean errors increase but the errors arising from the VDZ-F12/VTZ-F12 extrapolation are still slightly lower than even that of the AVTZ/AVQZ extrapolation, which is almost certainly due to the former combination's superior performance for second row systems. Indeed, from these relatively inexpensive calculations, the MAE for the extrapolated correlation energy is only 0.19 mE_h with a maximum error of less than half a millihartree. When the basis sets employed in the extrapolation are increased to VTZ-F12/VQZ-F12 the MAE drops to 0.05 mE_h, at which point the accuracy of the correlation energies produced is beyond the estimated precision of the reference data (± 0.1 mE_h). Although the AVQZ/AV5Z extrapolation MAE is slightly worse than that of VTZ-F12/VQZ-F12 at just under 0.06 mE_h, this is still within the limit of the confidence in the reference data. It should also be noted that the optimal exponent of $1/\ell_{\max}$ found in this work for the VTZ-F12/VQZ-F12 basis set pair is nearly identical to the value of 4 used in the original F12 basis set work,[36] which was based on a much smaller set of reference data. As also noted in that earlier work, this is far from the theoretical limit of $\ell_{\max}^{-7}$, but as mentioned above, the orbital basis sets used in this work are not sufficiently saturated in each angular momentum to reach this limit.

Table VII shows the optimized coefficients and error statistics for extrapolations when the systems BH, AlH, BN, and AlP (involving group thirteen elements) and PN, SiO, CS, ClF, and AlN (containing both first and second row elements) are added to the combined set of first and second row systems utilized in Table VI. Analyzing Table VII and the combined first and second row systems section of Table VI together indicates that the inclusion of the additional systems has little effect on the optimal extrapolation coefficients and hence also on the error statistics presented. While using coefficients optimized for the larger set of systems in Table VII is to be recommended, the small nature of the changes suggest that further testing can be carried out on the smaller combined set of fourteen systems used in Table VI.

In order to investigate how the optimized coefficients change between explicitly correlated and conventional MP2, conventional frozen-core MP2 calculations were carried out upon the combined first and second row systems for the AVnZ (where n=D, T, Q, 5, 6) basis sets. The optimized coefficients for both the power-law and Schwenke-type extrapolations on these conventional MP2 correlation energies are presented in Table VIII. Coefficients and error statistics separated into first row and second row systems only are presented in the supporting information as Table SIII. Comparing pow between Tables VI and VIII shows that in general, conventional MP2 has smaller optimal exponents than MP2-F12/3C for the same pair of basis sets, which is to be expected since this simply reflects the explicitly correlated method's greatly increased rate of convergence with respect to basis set size. Of course the extrapolations are also much more accurate in the explicitly correlated case, with a MP2-F12/3C AVDZ/AVTZ extrapolation producing results that are intermediate in accuracy between the extrapolated conventional AVQZ/AV5Z and AV5Z/AV6Z values. The results are even more impressive when the VnZ-F12 basis sets

TABLE VIII. Optimized CBS extrapolation coefficients for conventional MP2 using a combined set of 14 first and second row systems. (The systems are: Ne, N_2, CH_2, H_2O, CO, HF, F_2, Ar, P_2, SiH_2, H_2S, SiS, HCl, and Cl_2.) All error statistics reported in mE_h.

Basis sets	Equation (2) F	Equation (5) pow	MAE	Standard deviation	Max error
AVDZ/AVTZ	1.725 804	2.136 231	2.664	1.535	5.026
AVTZ/AVQZ	1.933 428	2.531 300	1.195	0.679	2.571
AVQZ/AV5Z	2.186 276	2.739 906	0.655	0.586	2.085
AV5Z/AV6Z	2.477 681	2.834 890	0.149	0.169	−0.646

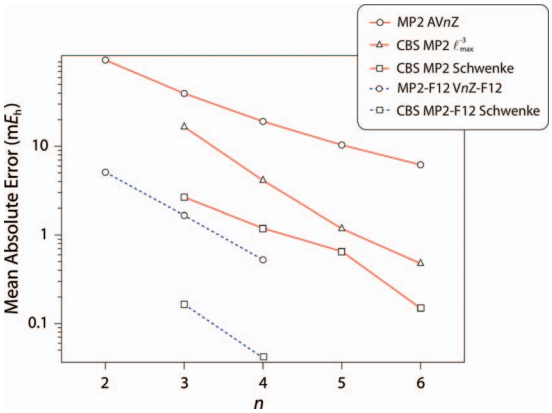

FIG. 1. Mean absolute errors of select CBS extrapolations and mean absolute basis set errors for a set of first and second row systems. The CBS extrapolations encompass the traditional fixed exponent power-law-based scheme (CBS MP2 ℓ^{-3}_{max}) and the Schwenke-style conventional MP2 (CBS MP2 Schwenke) and explicitly correlated MP2 (CBS MP2-F12 Schwenke) schemes. Basis sets are referred to by their cardinal number n and basis set extrapolations by the cardinal number of the largest basis set used in the extrapolation.

are used in MP2-F12/3C, as just a VDZ-F12/VTZ-F12 extrapolation produces results on par with a conventional AV5Z/AV6Z extrapolation. From the data reported in Table SIII it is shown that, with the exception of AVTZ/AVQZ, the extrapolation of conventional MP2 correlation energies for second row systems is much more problematic than for the first row. This is illustrated dramatically for AV5Z/AV6Z, which has an MAE of 0.04 mE_h for first row molecules but 0.20 for the second row. This suggests that while conventional AV5Z/AV6Z extrapolations can provide very good estimates of the CBS limit for first row systems, even larger basis sets or explicitly correlated methods should be employed for second row or mixed systems to obtain similar accuracies. It should be noted that the values of F presented in Table SIII for the first row systems are considerably different from the CCSD extrapolation values of Schwenke.[28] Thus using the same coefficients (either F or pow) for different electron correlation methods is suboptimal (see also Ref. 58).

Figure 1 presents the basis set errors from MP2/AVnZ and MP2-F12/VnZ-F12 in graphical form, along with the errors from selected basis set extrapolations. The errors shown are defined as the mean of the absolute difference between the calculated correlation energy and its reference value over first row systems and their second row analogs. The CBS extrapolations included are the ℓ^{-3}_{max} Helgaker extrapolation of Eq. (1) for conventional MP2 and the Schwenke-style extrapolations for both conventional and explicitly correlated MP2 using the coefficients in Tables VIII and VII, respectively. The improvement in using a Schwenke–style extrapolation over the more traditional power-law-based scheme for conventional MP2 with AVDZ and AVTZ basis sets is particularly striking, with the RMS error dropping by around an order of magnitude and producing results, on average, of almost the same quality as MP2-F12/VTZ-F12. The plot also visually demonstrates the high accuracy of the Schwenke-style extrapolations of the MP2-F12 correlation energies with the VnZ-F12 basis sets; for each pair of basis sets the CBS extrapolation improves the already accurate MP2-F12 correlation energies by about an order of magnitude.

The 3C(D) and 3C(FIX) *Ansätze* are both free of GBSSE and as such can be viewed as more desirable than the 3C method in the majority of cases, especially for energy differences. Optimized coefficients for Schwenke-style CBS extrapolations, along with MAE, standard deviation and max error statistics, from the set of fourteen first and second row systems (i.e., group 13 and molecules comprised of both first and second row elements were not included), using both of these *Ansätze* are listed in Table IX. The values of β chosen for 3C(D) are identical to 3C, while those for the 3C(FIX) calculations correspond to those recommended in Table V. Comparing Tables VI and IX shows that the optimized coefficients and error statistics are reasonably similar for 3C and 3C(D). The change in the value of F is much greater when using 3C(FIX), with the associated error statistics approximately doubling in most cases. In the worst case, i.e., extrapolation of the VTZ-F12 and VQZ-F12 basis sets (or AVTZ/AVQZ), the maximum error is actually larger than

TABLE IX. Optimized CBS extrapolation coefficients for MP2-F12 with *Ansätze* 3C(D) and 3C(FIX) using a combined set of fourteen first and second row systems. All error statistics reported in mE_h.

Basis sets	Equation (2) F	Equation (5) pow	MAE	Standard deviation	Max error
		MP2-F12/3C(D)			
VDZ-F12/VTZ-F12	1.495 083	2.725 445	0.205	0.120	0.453
VTZ-F12/VQZ-F12	1.414 800	4.264 299	0.059	0.026	0.101
AVDZ/AVTZ	1.773 504	2.046 502	0.368	0.282	1.000
AVTZ/AVQZ	1.756 495	2.928 317	0.273	0.175	0.652
AVQZ/AV5Z	1.654 634	4.159 425	0.075	0.050	0.195
		MP2-F12/3C(FIX)			
VDZ-F12/VTZ-F12	1.400 474	3.087 838	0.376	0.300	0.860
VTZ-F12/VQZ-F12	1.400 044	4.354 762	0.128	0.269	1.052
AVDZ/AVTZ	1.649 364	2.298 829	0.701	0.542	1.710
AVTZ/AVQZ	1.645 692	3.252 587	0.401	0.384	1.172
AVQZ/AV5Z	1.615 328	4.325 740	0.069	0.046	0.187

TABLE X. Optimized CBS coefficients for the extrapolation of CCSD-F12b correlation energies using the 3C(FIX) *Ansatz* for the systems presented in Table IV. Error statistics of the extrapolated energies relative to reference data are presented in mE_h.

Basis sets	Equation (2) F	Equation (5) pow	MAE	Standard deviation	Max error
VDZ-F12/VTZ-F12	1.387 834	3.144 518	0.520	0.553	1.658
VTZ-F12/VQZ-F12	1.363 388	4.595 995	0.179	0.113	0.420
AVDZ/AVTZ	1.575 845	2.483 070	0.841	0.447	−1.797
AVTZ/AVQZ	1.416 422	4.255 221	0.204	0.154	0.478
AVQZ/AV5Z	1.502 686	4.910 269	0.042	0.043	0.140

that observed for the VDZ-F12/VTZ-F12 extrapolation even though the MAE decreases. Presumably this merely reflects the greater sensitivity of the 3C(FIX) *Ansatz* to the choice of the geminal exponent β. This suggests that where accuracy of the extrapolated MP2-F12 correlation energy is concerned, use of the 3C(D) *Ansatz* is to be recommended over 3C and 3C(FIX), as it is free from GBSSE and the extrapolation coefficients elucidated from the current investigation provide more accurate results than those for 3C(FIX).

2. CCSD-F12b and CCSD(T)-F12b

As already noted above, a comparison of the optimal values of F for conventional MP2 with those presented by Schwenke for conventional CCSD indicates that extrapolation of different methods of electron correlation benefit from specifically optimized extrapolation coefficients. As stated in Section A, the CCSD(T)-F12b method currently implemented in MOLPRO does not treat the perturbative triples in an explicitly correlated manner. A consequence of this is that the CCSD-F12b and (T) correlation energy contributions will converge at different rates with respect to basis set and thus their extrapolation should be treated separately. CCSD(T)-F12b calculations were carried out on the systems listed in Table IV and values for the coefficient F were optimized by independently minimizing the RMS error with respect to the reference values for both CCSD-F12b and (T) presented in Table IV. The optimized F and pow values, along with the MAE, standard deviation of the error and the maximum error are presented in Tables X and XI, for the CCSD-F12b and (T) extrapolations, respectively.

Comparing the CCSD-F12b error statistics with those from the MP2-F12/3C(FIX) extrapolations in Table IX indicates that with the VnZ-F12 basis sets the CCSD-F12b extrapolations do not produce energies quite as close to their CBS limits as MP2-F12. In contrast, however, with the AVTZ/AVQZ and AVQZ/AV5Z pairs of basis sets the CCSD-F12b extrapolations are statistically more accurate than the MP2-F12 values. In the case of the AVDZ/AVTZ pair, the CCSD-F12b extrapolation yields slightly larger average errors compared to MP2-F12, with a maximum error of 1.80 mE_h. While VTZ-F12/VQZ-F12 extrapolations are statistically worse with the CCSD-F12b method, the MAE is still smaller than the estimated confidence in the reference data ($\pm$0.2 mE_h). In terms of the extrapolation of the (T) correlation contributions shown in Table XI, in all cases the performance is far more accurate than for the CCSD-F12b correlation energy, although the magnitude of the former contribution is also much smaller. Even with the smallest pair of basis sets in each series, the MAE is between about 0.1 and 0.2 mE_h, and the larger basis set combinations will easily surpass the confidence in the CCSD-F12b part of the reference correlation energy and hence the confidence in the overall extrapolated CCSD(T)-F12b correlation energy.

The MAEs in the extrapolated correlation energies and the mean absolute basis set errors for the VnZ-F12 and AVnZ series of basis sets are plotted in Fig. 2 for the CCSD-F12b method with *Ansatz* 3C(FIX). Although the MAE in an extrapolation of the VDZ-F12 and VTZ-F12 basis set energies (0.52 mE_h) seems relatively high compared to the analogous MP2-F12 extrapolation (see Fig. 1), Fig. 2 makes it clear that this is still a considerable improvement over the raw VQZ-F12 or AV5Z CCSD-F12b correlation energies. In fact, the CBS extrapolations still represent an order of magnitude improvement in the raw CCSD-F12b correlation energies, just as in the MP2-F12 case. However, the basis set errors of the CCSD-F12b correlation energies are three to five times larger for a given basis set compared to MP2-F12. Meanwhile, an AVDZ/AVTZ extrapolation produces a CBS

TABLE XI. Optimized CBS coefficients for the extrapolation of the (T) contribution to the CCSD(T)-F12b/3C(FIX) correlation energy. Error statistics for the extrapolated energy are relative to the reference data in Table IV and are reported in mE_h.

Basis sets	Equation (2) F	Equation (5) pow	MAE	Standard deviation	Max error
VDZ-F12/VTZ-F12	1.529 817	2.615 472	0.098	0.082	−0.303
VTZ-F12/VQZ-F12	1.769 474	2.894 993	0.059	0.034	0.097
AVDZ/AVTZ	1.476 233	2.790 300	0.206	0.171	−0.549
AVTZ/AVQZ	1.663 388	3.195 354	0.030	0.025	0.095
AVQZ/AV5Z	1.659 458	4.135 608	0.015	0.015	0.051

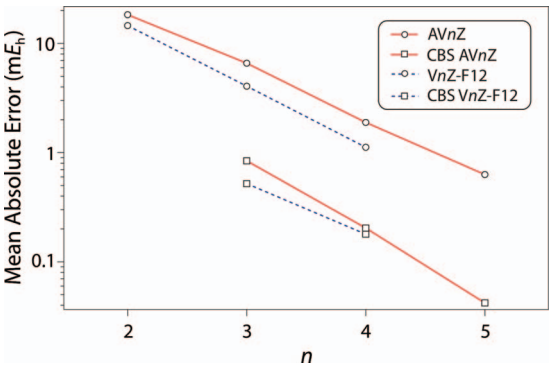

FIG. 2. Mean absolute errors of CBS extrapolations and mean absolute basis set errors of the CCSD-F12b correlation energy with *Ansatz* 3C(FIX) for a set of fourteen systems containing first and second row elements. Basis sets are referred to by their cardinal number, n, and CBS extrapolations by the cardinal number of the largest basis set used in the extrapolation.

correlation energy that is on average very close to the CCSD-F12b/AV5Z value. Figure 3 shows the same data for the (T) contribution to the CCSD(T)-F12b correlation energy and its extrapolation. It can be seen that extrapolations with double- and triple-ζ basis sets produce average correlation energies that are closer to the CBS limit than AV5Z values.

IV. CONCLUSIONS

Optimized CBS extrapolation parameters to be used in both two-point Schwenke-style and the more familiar power-law-based extrapolations of explicitly correlated MP2-F12 and CCSD-F12b correlation energies have been presented for systems containing first and second row atoms (and H). These values were produced by minimizing the error relative to new estimates of reference CBS limits for the frozen-core MP2 and CCSD correlation energies of several small molecules containing both first and second row elements. Separate coefficients were optimized for test sets of systems that included first row elements, second row elements, and a combined set of those first and second row systems. In the MP2 case an expanded set that contained systems having both first and second row elements and systems including group 13 elements was also included. Unsurprisingly, it is shown that the use of larger basis sets leads to more accurate extrapolations, but generally the optimal coefficients for second row systems are quite different to those for first row systems only.

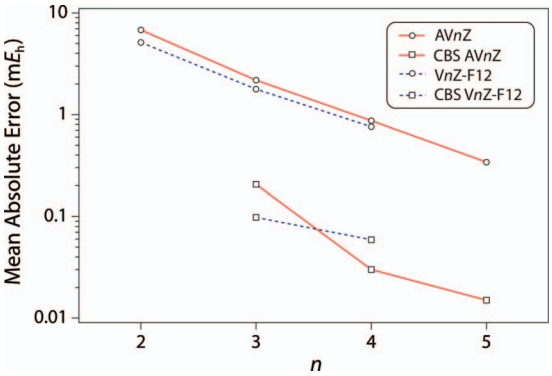

FIG. 3. Mean absolute errors of CBS extrapolations and mean absolute basis set errors of the (T) component of the CCSD(T)-F12b correlation energy with *Ansatz* 3C(FIX) for a set of fourteen systems containing first and second row elements. Basis sets are referred to by their cardinal number, n, and CBS extrapolations by the cardinal number of the largest basis set used in the extrapolation.

Analysis of the error in the extrapolated MP2-F12 correlation energies using *Ansatz* 3C shows that utilizing the relatively inexpensive VDZ-F12 and VTZ-F12 basis sets produces very accurate CBS limit energies, with an MAE of just 0.19 mE_h over the combined set of fourteen first and second row systems. This is similar to the error produced by extrapolation of conventional MP2 with large AV5Z and AV6Z basis sets, although this latter extrapolation is more accurate when only first row systems are considered. Extrapolation of the explicitly correlated energies with either the VTZ-F12 and VQZ-F12 or AVQZ and AV5Z basis sets produces an MAE of less than 0.1 mE_h, which is the level of confidence in the reference data. The choice of *Ansatz* has an effect on both the value of the coefficient used in the extrapolation and the resulting statistics of the error relative to reference data. The difference is small between the 3C and 3C(D) *Ansätze*, yet statistically the extrapolations become worse with 3C(FIX). Unlike *Ansatz* 3C, 3C(D) is not susceptible to GBSSE and thus it becomes the preferred *Ansatz* for the extrapolation of MP2-F12 correlation energies as long as localized orbitals are used. The choice of 3C(D) over 3C is likely to be more important when relative energies are considered, as GBSSE may be more of a factor in such cases. Curiously the differences between 3C and 3C(FIX) were observed to be larger for second row atom-containing systems.

Coefficients have also been optimized for use in Schwenke-style extrapolations of the frozen-core CCSD(T)-F12b total correlation energy, where the CCSD-F12b and (T) contributions are treated separately. Overall these extrapolations do not reach the same accuracy displayed for MP2-F12; for example, using correlation energies from VDZ-F12 and VTZ-F12 calculations results in an MAE of 0.52 mE_h in the CBS limits for the set of systems containing both first and second row systems. However, this MAE is smaller than that produced by the much more expensive AV5Z basis set alone. It should also be noted that the AVTZ/AVQZ and AVQZ/AV5Z extrapolations are statistically more accurate than their MP2-F12 equivalents. When basis sets up to quadruple-ζ or greater quality are utilized in an extrapolation of the correlation energy, the average error is smaller than 0.2 mE_h, which is about the level of confidence in the CCSD reference data. Perhaps reflecting the relatively small magnitude of the (T) contribution, even the smaller basis sets produced relatively accurate extrapolations, with the VDZ-F12 and VTZ-F12 combination producing an MAE of just 0.10 mE_h. In all cases the average error is significantly smaller in magnitude than in the CCSD-F12b extrapolations, implying that the latter is the limiting factor in the accuracy of the combined CCSD(T)-F12 CBS extrapolated correlation energies.

It should be noted at this point that the extra time required for the MP2-F12 and the additional integrals needed for CCSD-F12 is relatively small. For example, for ethanol, C_2H_5OH (C_1 symmetry), with the AVTZ orbital basis set a CCSD calculation takes 24.6 min (user CPU-time on a single Opteron 2380 processor, including two-electron integrals,

Hartree-Fock, integral transformation, and 9 CCSD iterations). The corresponding CCSD-F12a calculation with an AVTZ/OPTRI auxiliary basis set takes only 4 min longer (additional 1.5 min are required for computing the CCSD-F12b energy). Including the (T) energy correction, the CCSD(T) takes 62 min, the CCSD(T)-F12b 68 min. Thus the explicitly correlated calculation requires only 10% additional time; since the MP2-F12 scales as $O(N^5)$ while the CCSD(T) scales as $O(N^7)$, where N is a measure of the molecular size, this fraction quickly decreases with increasing molecular size (timings for the benzene dimer can be found in Ref. 54). The corresponding CCSD(T) calculation for ethanol with the AVQZ basis (516 CGTOs) set takes 730 min (of this 438 min for the triples). With our current computational resources, CCSD or CCSD(T) calculations with the AV5Z basis set were not possible. In the case of just MP2-F12 and conventional MP2, the calculations with the AVTZ basis set took 9.9 and 5.8 min, respectively, while the AVQZ ones required 113 and 95 min for MP2-F12 and MP2, respectively. It should be noted that the savings for a small basis set DF-MP2-F12 calculation versus a large basis set DF-MP2 are much less pronounced, since the DF-MP2 is extremely efficient. For the current example, the DF-MP2-F12/AVTZ calculation takes 4.5 min and a DF-MP2/AV5Z calculation only 19.6 min (total times, including DF-HF).

ACKNOWLEDGMENTS

J.G.H. and K.A.P. acknowledge the support of the National Science Foundation (Grant No. CHE-0723997). G.K. and H.J.W. were supported by the Deutsche Forschungsgemeinschaft (DFG) with the priority program 1145 and by Grant No. DFG WE 984/16-1.

[1] N. J. Deyonker, T. R. Cundari, and A. K. Wilson, J. Chem. Phys. **124**, 114104 (2006); J. Martin and G. de Oliveira, *ibid.* **111**, 1843 (1999); A. Tajti, P. Szalay, A. Császár, M. Kállay, J. Gauss, E. Valeev, B. Flowers, J. Vázquez, and J. Stanton, *ibid.* **121**, 11599 (2004); D. Feller, K. A. Peterson, and D. A. Dixon, *ibid.* **129**, 204105 (2008).
[2] N. B. Balabanov, B. C. Shepler, and K. A. Peterson, J. Phys. Chem. A **109**, 8765 (2005); P. Barletta, S. V. Shirin, N. F. Zobov, O. L. Polyansky, J. Tennyson, E. F. Valeev, and A. G. Császár, J. Chem. Phys. **125**, 204307 (2006); K. A. Peterson and G. C. Mcbane, *ibid.* **123**, 084314 (2005).
[3] Y. Chuang and D. Truhlar, J. Phys. Chem. A **103**, 651 (1999); B. Paizs, P. Salvador, A. G. Császár, M. Duran, and S. Suhai, J. Comput. Chem. **22**, 196 (2001).
[4] M. Heckert, M. Kállay, D. P. Tew, W. Klopper, and J. Gauss, J. Chem. Phys. **125**, 044108 (2006).
[5] J. Almlöf and P. Taylor, J. Chem. Phys. **86**, 4070 (1987).
[6] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[7] K. A. Peterson, in *Annual Reports in Computational Chemistry*, edited by T. D. Crawford (Elsevier, Amsterdam, 2007), Vol. 3, p. 195.
[8] R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[9] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **103**, 4572 (1995).
[10] W. Klopper, F. R. Manby, S. Ten-no, and E. F. Valeev, Int. Rev. Phys. Chem. **25**, 427 (2006).
[11] E. A. Hylleraas, Z. Phys. **54**, 347 (1929); **48**, 469 (1928).
[12] W. Kutzelnigg, Theor. Chim. Acta **68**, 445 (1985).
[13] W. Klopper and C. C. M. Samson, J. Chem. Phys. **116**, 6397 (2002).
[14] S. Ten-no and F. R. Manby, J. Chem. Phys. **119**, 5358 (2003); F. R. Manby, *ibid.* **119**, 4607 (2003).
[15] D. P. Tew and W. Klopper, J. Chem. Phys. **123**, 074101 (2005).
[16] S. Ten-no, Chem. Phys. Lett. **398**, 56 (2004).
[17] A. J. May, E. F. Valeev, R. Polly, and F. R. Manby, Phys. Chem. Chem. Phys. **7**, 2710 (2005).
[18] T. Helgaker, W. Klopper, and D. P. Tew, Mol. Phys. **106**, 2107 (2008).
[19] D. Feller, J. Chem. Phys. **96**, 6104 (1992); **98**, 7059 (1993).
[20] C. Schwartz, Phys. Rev. **126**, 1015 (1962); R. N. Hill, J. Chem. Phys. **83**, 1173 (1985); D. P. Carroll, H. J. Silverstone, and R. M. Metzger, *ibid.* **71**, 4142 (1979).
[21] W. Kutzelnigg and W. Klopper, J. Chem. Phys. **94**, 1985 (1991).
[22] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).
[23] T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).
[24] A. J. C. Varandas, J. Chem. Phys. **126**, 244105 (2007).
[25] W. Klopper, Mol. Phys. **99**, 481 (2001).
[26] K. A. Peterson, D. E. Woon, and T. H. Dunning, Jr., J. Chem. Phys. **100**, 7410 (1994); D. Bakowies, *ibid.* **127**, 164109 (2007).
[27] D. Yamaki, H. Koch, and S. Ten-No, J. Chem. Phys. **127**, 144104 (2007).
[28] D. W. Schwenke, J. Chem. Phys. **122**, 014107 (2005).
[29] E. F. Valeev and T. Daniel Crawford, J. Chem. Phys. **128**, 244113 (2008).
[30] H.-J. Werner, T. B. Adler, and F. R. Manby, J. Chem. Phys. **126**, 164102 (2007).
[31] G. Knizia and H.-J. Werner, J. Chem. Phys. **128**, 154103 (2008).
[32] MOLPRO, version 2009.1, a package of *ab initio* programs, H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz *et al.* Cardiff University, UK, 2009 (http://www.molpro.net).
[33] T. B. Adler, G. Knizia, and H.-J. Werner, J. Chem. Phys. **127**, 221106 (2007).
[34] G. Knizia, T. B. Adler, and H.-J. Werner, J. Chem. Phys. **130**, 054104 (2009).
[35] J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916 (1989).
[36] K. A. Peterson, T. B. Adler, and H.-J. Werner, J. Chem. Phys. **128**, 084102 (2008).
[37] T. H. Dunning, Jr., K. A. Peterson, and A. K. Wilson, J. Chem. Phys. **114**, 9244 (2001).
[38] K. E. Yousaf and K. A. Peterson, J. Chem. Phys. **129**, 184108 (2008); Chem. Phys. Lett. **476**, 303 (2009).
[39] E. F. Valeev, Chem. Phys. Lett. **395**, 190 (2004).
[40] F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).
[41] F. Weigend, A. Köhn, and C. Hättig, J. Chem. Phys. **116**, 3175 (2002).
[42] C. Hättig, Phys. Chem. Chem. Phys. **7**, 59 (2005).
[43] K. A. Peterson and T. H. Dunning, Jr., J. Chem. Phys. **117**, 10548 (2002).
[44] H. Fliegl, W. Klopper, and C. Hättig, J. Chem. Phys. **122**, 084107 (2005).
[45] T. van Mourik and T. H. Dunning, Jr., Int. J. Quantum Chem. **76**, 205 (2000).
[46] D. Feller and K. Ruedenberg, Theor. Chim. Acta **52**, 231 (1979).
[47] T. van Mourik, A. K. Wilson, and T. H. Dunning, Jr., Mol. Phys. **96**, 529 (1999).
[48] See EPAPS supplementary materials at http://dx.doi.org/10.1063/1.3265857 for the reference AO and auxiliary basis sets used in this work as well as the reference geometries. The results of CBS extrapolations on individual molecules using the 3C(D) *Ansatz* for MP2–F12 and 3C(FIX) for both CCSD-F12b and (T) are also given.
[49] J. R. Flores, Int. J. Quantum Chem. **108**, 2172 (2008).
[50] F. A. Bischoff, S. Wolfsegger, D. P. Tew, and W. Klopper, Mol. Phys. **107**, 963 (2009).
[51] E. C. Barnes, G. A. Petersson, D. Feller, and K. A. Peterson, J. Chem. Phys. **129**, 194115 (2008).
[52] D. Feller, personal communication.
[53] J. G. Hill and J. A. Platts, J. Chem. Theory Comput. **5**, 500 (2009).
[54] O. Marchetti and H.-J. Werner, Phys. Chem. Chem. Phys. **10**, 3400 (2008).
[55] S. Höfener, D. P. Tew, W. Klopper, and T. Helgaker, Chem. Phys. **356**, 25 (2009).
[56] T. B. Adler, H.-J. Werner, and F. R. Manby, J. Chem. Phys. **130**, 054106 (2009).
[57] D. P. Tew, W. Klopper, C. Neiss, and C. Hättig, Phys. Chem. Chem. Phys. **9**, 1921 (2007).
[58] J. S. Lee and S. Y. Park, J. Chem. Phys. **112**, 10746 (2000).
[59] D. Bakowies, J. Chem. Phys. **127**, 084105 (2007).