

Basis set convergence of explicitly correlated double-hybrid density functional theory calculations

Amir Karton and Jan M. L. Martin

Citation: *J. Chem. Phys.* **135**, 144119 (2011); doi: 10.1063/1.3647980

View online: <http://dx.doi.org/10.1063/1.3647980>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v135/i14>

Published by the [American Institute of Physics](#).

Related Articles

Density functional theory calculations of dynamic first hyperpolarizabilities for organic molecules in organic solvent: Comparison to experiment
J. Chem. Phys. **135**, 134104 (2011)

Comparing ab initio density-functional and wave function theories: The impact of correlation on the electronic density and the role of the correlation potential
J. Chem. Phys. **135**, 114111 (2011)

Communication: Rationale for a new class of double-hybrid approximations in density-functional theory
J. Chem. Phys. **135**, 101102 (2011)

The performance and relationship among range-separated schemes for density functional theory
J. Chem. Phys. **135**, 074109 (2011)

On the effect of electron correlation on the static second hyperpolarizability of conjugated oligomer chains
J. Chem. Phys. **135**, 014111 (2011)

Additional information on *J. Chem. Phys.*

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

Journal Information: http://jcp.aip.org/about/about_the_journal

Top downloads: http://jcp.aip.org/features/most_downloaded

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

ADVERTISEMENT

**AIP**Advances

Submit Now

**Explore AIP's new
open-access journal**

- **Article-level metrics
now available**
- **Join the conversation!
Rate & comment on articles**

Basis set convergence of explicitly correlated double-hybrid density functional theory calculations

Amir Karton¹ and Jan M. L. Martin^{2,a)}

¹*School of Chemistry, University of Sydney, Sydney, New South Wales 2006, Australia*

²*Department of Chemistry, University of North Texas, Denton, Texas 76203-5017, USA*

(Received 25 August 2011; accepted 20 September 2011; published online 13 October 2011)

The basis set convergence of explicitly correlated double-hybrid density functional theory (DFT) is investigated using the B2GP-PLYP functional. As reference values, we use basis set limit B2GP-PLYP-F12 reaction energies extrapolated from the aug'-cc-pV(Q+d)Z and aug'-cc-pV(5+d)Z basis sets. Explicitly correlated double-hybrid DFT calculations converge significantly faster to the basis set limit than conventional calculations done with basis sets saturated up to the same angular momentum (typically, one "gains" one angular momentum in the explicitly correlated calculations). In explicitly correlated F12 calculations the VnZ-F12 basis sets converge faster than the orbital A'VnZ basis sets. Furthermore, basis set convergence of the MP2-F12 component is apparently faster than that of the underlying Kohn-Sham calculation. Therefore, the most cost-effective approach consists of combining the MP2-F12 correlation energy from a comparatively small basis set such as VDZ-F12 with a DFT energy from a larger basis set such as aug'-cc-pV(T+d)Z. © 2011 American Institute of Physics. [doi:10.1063/1.3647980]

I. INTRODUCTION

In contemporary computational quantum chemistry, the two leading approaches are (correlated) wavefunction *ab initio* methods and density functional theory (DFT). *Ab initio* methods offer a clear road map for convergence to the exact solution but their computational cost becomes prohibitive beyond (what most practicing chemists would consider to be very) small molecules. In particular, basis set convergence of the short-range correlation energy is excruciatingly slow.

Density functional methods do not suffer from this particular problem. While their computational cost formally scales as just $O(N^3)$ with the size of the system (or $O(N^4)$ if involving "exact" Hartree-Fock-like exchange), in practice scalings are even less steep thanks to integral (pre)screening. Thus, DFT methods are routinely applied to fairly large systems. Their basis set convergence behavior is fairly similar to that of Hartree-Fock theory. Unfortunately, DFT intrinsically involves the introduction of an unknown "exchange-correlation" functional. Different DFT functionals effectively represent different (more or less) educated guesses of the exact functional, and while there is a hierarchy of sorts in these approximations (the so-called "Jacob's ladder"¹), no truly systematic path towards the exact solution exists.

In recent years, double-hybrid functionals have emerged as an interesting "third way" or middle road between these two approaches. The idea of carrying out perturbation theory in a basis of Kohn-Sham (KS) orbitals goes back to the work of Görling and Levy;² while Grimme³ was the first to actually turn this into a practical method. The double-hybrid functionals occupy the fifth rung of Jacob's ladder as they mix exact Hartree-Fock exchange as well as MP2 correlation energies

with the exchange-correlation energy of the underlying DFT functional. Where the underlying functional may, in principle, occupy any of the lower rungs (1–4) of Jacob's Ladder.⁴

As has been shown,^{3–7} double-hybrid methods offer a level of agreement with experiment that approaches what is achievable using composite *ab initio* schemes such as G3 and G4 theories.^{8–10} While basis set convergence is faster than for wavefunction *ab initio* methods, they inherit the slow basis set convergence of MP2 to some degree, and hence fully converged atomization energies (or, for that matter, reaction energies of reactions in which several bonds are broken and formed) still require basis sets of *spdfgh* quality, or extrapolation from *spdf* and *spdfg* basis sets. (For purposes such as reaction barrier heights that are less demanding in terms of absolute basis set convergence, basis sets of *spdf* quality are typically sufficient.⁵) Yet basis set sensitivity is mitigated to some degree by the fact that MP2-like correlation in double hybrids typically accounts for only between one-quarter and one-half of the overall correlation energy (the remainder coming from local-density approximation, generalized gradient approximation (GGA), or meta-GGA DFT correlation).

Only one-electron Gaussian basis sets were hitherto considered. Much faster basis set convergence can be reached when terms that depend explicitly on interelectronic distances r_{ij} are included in the wavefunction. Such "explicitly correlated" approaches have been known since the 1970s,¹¹ while the idea itself ultimately goes back, of course, to the 1928 work of Hylleraas on the helium atom.¹² Explicitly correlated calculations display greatly accelerated basis set convergence compared to the conventional calculations. Nevertheless, unless further approximations are invoked, the numerous three- and four-electron integrals that must be evaluated restrict the application of explicitly correlated techniques to very small molecular systems. These approximations are now briefly discussed (for detailed descriptions see, e.g., Refs. 13–15).

^{a)}Electronic mail: gershom@unt.edu.

In the 1980s, Kutzelnigg and Klopper introduced the so-called R12 approach for the MP2 pair function, in which the MP2 pair function (expanded in a conventional orbital basis set) is supplemented with a small number of explicitly correlated basis functions that depend on a linear r_{12} term.¹⁶ Subsequently, in 2002, Klopper and Samson introduced an auxiliary basis set for the resolution of the identity (RI), making it possible to eliminate the costly many-electron integrals in R12 by means of RI.¹⁷ Shortly thereafter, Manby combined MP2-R12 with density fitting, further speeding up integral evaluation.¹⁸ In the MP2-F12 approach, proposed by Ten-no,¹⁹ the linear r_{12} correlation factor is replaced with an exponential function $e^{-\gamma r_{12}}$ which is then expanded in Gaussians. This substantially improves the performance, in particular, when small basis sets are used.

Experience with MP2-F12 has shown that basis set convergence is drastically faster than the conventional orbital calculations:^{20,21} typically, the “gain” amounts to at least one (typically two) angular momenta. While there are some issues with reaching basis set convergence to better than 0.1 kcal/mol due to the various approximations,^{21,22} these are only relevant in high-accuracy coupled cluster and/or multireference configuration interaction calculations, and become irrelevant for double-hybrid DFT calculations where the intrinsic error of the method is an order of magnitude larger.

This suggested us the possibility that the use of an F12 approach for the MP2-like part of a double-hybrid functional might accelerate basis set convergence to the point that an *spdf* (or even *spd*) basis set is adequate for many purposes, which would greatly increase the scope of applicability of these methods. We will show below that this is indeed the case.

II. COMPUTATIONAL METHODS

Most calculations were performed on the Martin group Linux cluster and on the Faculty of Chemistry, HPC facility at the Weizmann Institute of Science. Some very large-scale calculations were carried out on the cluster of the Center for Advanced Scientific Computing and Modeling (CASCAM) at the University of North Texas. The MOLPRO 2010.1 program suite was used throughout.²³ All of the MP2-F12 steps in the double-hybrid B2GP-PLYP (Ref. 5) calculations were carried out with the DF-MP2-F12 approximation^{24,25} (i.e., MP2-F12 with density fitting). In all of these calculations the diagonal, fixed-amplitude 3C(FIX) ansatz¹⁹ was employed. Hereinafter, B2GP-PLYP calculations in conjunction with the DF-MP2-F12 method are denoted by B2GP-PLYP-F12, whereas conventional double-hybrid calculations (i.e., without the F12 technique or density fitting) will simply be referred to as B2GP-PLYP. Unless explicitly indicated otherwise, all conventional B2GP-PLYP calculations were carried out without density fitting approximations. A number of families of basis sets were considered. The first are the correlation consistent basis sets of Dunning and co-workers,^{26–28} which were developed for correlated conventional orbital calculations and are something of a “*de facto* standard” for those (the shorthand notation $A'VnZ$ indicates the combination of regular $cc-pVnZ$ on hydrogen, $aug-cc-pVnZ$ on first-row elements, and

$aug-cc-pV(n+d)Z$ on second-row elements). The second family of basis sets are the $cc-pVnZ-F12$ basis sets of Peterson *et al.*,²⁰ which were developed with explicitly correlated calculations in mind (in the paper, these are denoted by $VnZ-F12$). Optimal values for the geminal Slater exponents (β) used in conjunction with the $VnZ-F12$ and $A'VnZ$ basis sets were taken from Ref. 20. Extrapolation exponents for $VnZ-F12$ basis set were taken from Ref. 21. Finally, in some of the DFT calculations we also consider basis sets which were optimized for Hartree-Fock and DFT; specifically, these are the polarization consistent $A'PC-n$ basis sets of Jensen²⁹ and the $A'nZVPP$ doubly polarized basis sets of Weigend and Ahlrichs.³⁰

The RI approximation was applied using the OptRI auxiliary basis sets of Yousaf and Peterson³¹ within the complementary auxiliary basis approach of Valeev.³² The JKFIT fitting basis sets of Weigend³³ were employed for the density fitting in the Hartree-Fock calculations, while the MP2FIT set of Hättig and co-workers³⁴ was employed for the density fitting of the remaining two-electron integrals in the MP2-F12 calculations.

Finally, we note that the geometries (optimized at the CCSD(T)/ $cc-pV(Q+d)Z$ level of theory) of the small species involved in the test set of 66 isogyric reactions were taken from the supplemental information to Ref. 10. The geometries for the larger species considered in Sec. III D (namely, fulvene, benzene, naphthalene, anthracene, and phenanthrene) were optimized at the B3LYP-D/Def2-TZVP (Ref. 35) level of theory.

III. RESULTS AND DISCUSSION

In the present work, we are interested primarily in investigating the basis set convergence of double-hybrid DFT calculations, in which the MP2 step is carried out with the explicitly correlated F12 method (i.e., the B2GP-PLYP-F12 method). As basis set limit reference values we use our B2GP-PLYP-F12 reaction energies extrapolated from the $A'VQZ$ and $A'V5Z$ basis sets using the $A + B/L^\alpha$ two-point extrapolation formula (where L is the highest angular momentum present in the basis set) with $\alpha = 4.111\,526$ as recommended in Ref. 21. For the purpose of the present work these reaction energies are sufficiently converged to the complete basis set limit (typically within ~ 0.1 kcal/mol) as the intrinsic accuracy of the B2GP-PLYP functional for the type of reactions considered is about one order of magnitude higher (on the order of 1–2 kcal/mol).⁵

Table I lists the 66 reactions in our training set; 46 reactions involve species containing only first-row elements, and 20 reactions involve species containing both first- and second-row elements. Due to the limitations in the MP2-F12 code for open-shell cases and non-Hartree-Fock orbitals, only closed-shell species could be considered. Thus, instead of using a training set of atomization reactions (e.g., the atomization reactions in the W4-11 test set¹⁰), we resort to a set of reactions of similar chemical difficulty but that involve only closed-shell species. In these reactions all (or most) of the bonds in the reactants are being broken, and the nature of these bonds (e.g., bond order, polarity, cyclic vs acyclic, etc.) is quite

TABLE I. Set of 66 reactions involving first- and second-row closed-shell species.

1	$2\text{CH}_4 \rightarrow \text{C}_2 + 4\text{H}_2$	34	$\text{H}_2 + \text{CO} + \text{CF}_4 \rightarrow \text{F}_2 + \text{C}_2\text{F}_2 + \text{H}_2\text{O}$
2	$2\text{NH}_3 \rightarrow \text{N}_2 + 3\text{H}_2$	35	$\text{CF}_4 \rightarrow \text{CF}_2 + \text{F}_2$
3	$2\text{HF} \rightarrow \text{F}_2 + \text{H}_2$	36	$2\text{CF}_2 \rightarrow \text{C}_2\text{F}_2 + \text{F}_2$
4	$\text{H}_2\text{CO} \rightarrow \text{CO} + \text{H}_2$	37	$3\text{H}_2\text{O} \rightarrow \text{O}_3 + 3\text{H}_2$
5	$\text{CH}_4 + \text{H}_2\text{O} \rightarrow \text{CO} + 3\text{H}_2$	38	$\text{O}_3 + 3\text{HF} \rightarrow 3\text{HOF}$
6	$\text{CO} + \text{CH}_4 \rightarrow \text{C}_2\text{H}_2 + \text{H}_2\text{O}$	39	$3\text{CO} + 6\text{H}_2 \rightarrow \text{O}_3 + 3\text{CH}_4$
7	$\text{C}_2\text{H}_4 + \text{H}_2\text{O} \rightarrow \text{CO} + \text{CH}_4 + \text{H}_2$	40	$6\text{CO} + 3\text{H}_2 \rightarrow 2\text{O}_3 + 3\text{C}_2\text{H}_2$
8	$\text{C}_2\text{H}_6 + \text{H}_2\text{O} \rightarrow \text{CO} + \text{CH}_4 + 2\text{H}_2$	41	$6\text{CO} + 6\text{H}_2 \rightarrow 2\text{O}_3 + 3\text{C}_2\text{H}_4$
9	$2\text{CH}_4 + \text{H}_2\text{O} \rightarrow \text{oxirane} + 3\text{H}_2$	42	$6\text{CO} + 9\text{H}_2 \rightarrow 2\text{O}_3 + 3\text{C}_2\text{H}_6$
10	$\text{oxirane} \rightarrow \text{C}_2\text{H}_2 + \text{H}_2\text{O}$	43	$3\text{oxirene} \rightarrow 3\text{C}_2\text{H}_2 + \text{O}_3$
11	$\text{C}_2\text{H}_4 + \text{H}_2\text{O} \rightarrow \text{oxirane} + \text{H}_2$	44	$3\text{oxirane} \rightarrow 3\text{C}_2\text{H}_4 + \text{O}_3$
12	$\text{C}_2\text{H}_6 + \text{H}_2\text{O} \rightarrow \text{oxirane} + 2\text{H}_2$	45	$3\text{CO}_2 \rightarrow 3\text{CO} + \text{O}_3$
13	$2\text{CH}_4 + \text{H}_2\text{O} \rightarrow \text{oxirene} + 4\text{H}_2$	46	$2\text{PH}_3 \rightarrow \text{P}_2 + 3\text{H}_2$
14	$\text{C}_2\text{H}_2 + \text{H}_2\text{O} \rightarrow \text{oxirene} + \text{H}_2$	47	$2\text{HCl} \rightarrow \text{Cl}_2 + \text{H}_2$
15	$\text{C}_2\text{H}_4 + \text{H}_2\text{O} \rightarrow \text{oxirene} + 2\text{H}_2$	48	$\text{CH}_4 + \text{H}_2\text{S} \rightarrow \text{CS} + 3\text{H}_2$
16	$\text{C}_2\text{H}_6 + \text{H}_2\text{O} \rightarrow \text{oxirene} + 3\text{H}_2$	49	$\text{CS} + \text{CH}_4 \rightarrow \text{C}_2\text{H}_2 + \text{H}_2\text{S}$
17	$\text{CH}_4 + \text{NH}_3 \rightarrow \text{H}_2\text{CNH} + 2\text{H}_2$	50	$\text{C}_2\text{H}_4 + \text{H}_2\text{S} \rightarrow \text{CS} + \text{CH}_4 + \text{H}_2$
18	$\text{C}_2\text{H}_2 + 2\text{NH}_3 \rightarrow 2\text{H}_2\text{CNH} + \text{H}_2$	51	$\text{C}_2\text{H}_6 + \text{H}_2\text{S} \rightarrow \text{CS} + \text{CH}_4 + 2\text{H}_2$
19	$\text{C}_2\text{H}_4 + 2\text{NH}_3 \rightarrow 2\text{H}_2\text{CNH} + 2\text{H}_2$	52	$\text{CH}_4 + 2\text{HCl} \rightarrow \text{CCl}_2 + 3\text{H}_2$
20	$\text{C}_2\text{H}_6 + 2\text{NH}_3 \rightarrow 2\text{H}_2\text{CNH} + 3\text{H}_2$	53	$\text{C}_2\text{H}_2 + 4\text{HCl} \rightarrow 2\text{CCl}_2 + 3\text{H}_2$
21	$\text{CH}_4 \rightarrow \text{CH}_2 + \text{H}_2$	54	$\text{C}_2\text{H}_4 + 4\text{HCl} \rightarrow 2\text{CCl}_2 + 4\text{H}_2$
22	$\text{C}_2\text{H}_4 \rightarrow \text{C}_2\text{H}_2 + \text{H}_2$	55	$\text{C}_2\text{H}_6 + 4\text{HCl} \rightarrow 2\text{CCl}_2 + 5\text{H}_2$
23	$\text{C}_2\text{H}_4 \rightarrow 2\text{CH}_2$	56	$\text{CH}_4 + \text{Cl}_2 \rightarrow \text{CCl}_2 + 2\text{H}_2$
24	$\text{C}_2\text{H}_6 \rightarrow \text{C}_2\text{H}_4 + \text{H}_2$	57	$\text{C}_2\text{H}_2 + 2\text{Cl}_2 \rightarrow 2\text{CCl}_2 + \text{H}_2$
25	$\text{CH}_4 + 2\text{HF} \rightarrow \text{CF}_2 + 3\text{H}_2$	58	$\text{C}_2\text{H}_4 + 2\text{Cl}_2 \rightarrow 2\text{CCl}_2 + 2\text{H}_2$
26	$\text{C}_2\text{H}_2 + 4\text{HF} \rightarrow 2\text{CF}_2 + 3\text{H}_2$	59	$\text{C}_2\text{H}_6 + 2\text{Cl}_2 \rightarrow 2\text{CCl}_2 + 3\text{H}_2$
27	$\text{C}_2\text{H}_4 + 4\text{HF} \rightarrow 2\text{CF}_2 + 4\text{H}_2$	60	$3\text{H}_2\text{S} \rightarrow \text{S}_3 + 3\text{H}_2$
28	$\text{C}_2\text{H}_6 + 4\text{HF} \rightarrow 2\text{CF}_2 + 5\text{H}_2$	61	$3\text{HOCl} \rightarrow \text{O}_3 + 3\text{HCl}$
29	$\text{CF}_2 + 2\text{H}_2 \rightarrow \text{CH}_4 + \text{F}_2$	62	$\text{OCS} + \text{CH}_4 \rightarrow \text{CO} + \text{CS} + 2\text{H}_2$
30	$2\text{CF}_2 + \text{H}_2 \rightarrow \text{C}_2\text{H}_2 + 2\text{F}_2$	63	$2\text{OCS} + \text{C}_2\text{H}_2 \rightarrow 2\text{CO} + 2\text{CS} + \text{H}_2$
31	$2\text{CF}_2 + 2\text{H}_2 \rightarrow \text{C}_2\text{H}_4 + 2\text{F}_2$	64	$2\text{OCS} + \text{C}_2\text{H}_4 \rightarrow 2\text{CO} + 2\text{CS} + 2\text{H}_2$
32	$2\text{CF}_2 + 3\text{H}_2 \rightarrow \text{C}_2\text{H}_6 + 2\text{F}_2$	65	$2\text{OCS} + \text{C}_2\text{H}_6 \rightarrow 2\text{CO} + 2\text{CS} + 3\text{H}_2$
33	$\text{CF}_4 + \text{H}_2\text{O} \rightarrow \text{CO} + 2\text{F}_2 + \text{H}_2$	66	$3\text{CS}_2 \rightarrow 3\text{CS} + \text{S}_3$

different from those being formed in the products. Thus, the selected set of reactions should be a challenging test case from the perspective of basis set convergence. In addition, in some cases (most notably, the reactions involving ozone) the reactants and products evidently span the gamut from systems dominated by a single reference configuration to systems dominated by non-dynamical correlation. It should be noted, however, that the test set used here is still less difficult from the point of view of basis set convergence than a test set of atomization reactions.

A. Basis set convergence in “conventional” double-hybrid calculations

Let us first consider basis set convergence with the orbital basis sets in “conventional” double-hybrid calculations, i.e., B2GP-PLYP/A'VnZ. These results are presented in Table II. While convergence is faster than for wavefunction *ab initio* methods as the MP2-like terms only account for part of the correlation energy, it is still quite slow. The A'VDZ basis set results in an unacceptably large root mean square deviation (RMSD) of 4.75 kcal/mol for the entire set of 66 reactions. The A'VTZ basis set leads to a RMSD of 0.85 kcal/mol, implying a 95% confidence interval of about 1.7 kcal/mol which

TABLE II. Statistical analysis of the basis set convergence in conventional B2GP-PLYP calculations for the set of 66 reaction energies (in kcal/mol).^{a,b}

	A'VDZ	A'VTZ	A'VQZ	A'V5Z
All 66 reactions				
RMSD	4.75	0.85	0.30	0.16
MAD	3.84	0.65	0.21	0.11
MSD	-0.16	-0.33	-0.04	-0.01
46 reactions involving only first-row species				
RMSD	4.28	0.80	0.19	0.10
MAD	3.45	0.59	0.13	0.08
MSD	0.29	-0.48	-0.06	-0.02
LD	11.09	2.20	0.50	0.29
LD ^c	42	42	45	45
20 reactions involving both first- and second-row species				
RMSD	5.70	0.94	0.48	0.24
MAD	4.74	0.78	0.39	0.20
MSD	-1.21	-0.01	0.00	0.00
LD	11.42	2.24	1.16	0.60
LD ^c	57	60	60	60

^aThe basis set limit reference reaction energies are extrapolated from the B2GP-PLYP-F12/A'VQZ and B2GP-PLYP-F12/A'V5Z reaction energies, see text.

^bRMSD = root mean square deviation, MSD = mean signed deviation, MAD = mean absolute deviation, LD = largest deviation.

^cReaction number, see Table I. For the errors of all the reactions, see Table SII of the supplemental information (Ref. 38).

is comparable to the intrinsic accuracy of the B2GP-PLYP method (i.e., 1–2 kcal/mol). In order to surpass this level of accuracy the A'VQZ basis set has to be employed. This yields a RMSD of 0.30 kcal/mol for the entire set. For still better convergence the A'V5Z basis set can be considered (RMSD = 0.16 kcal/mol). In most applications, however, basis set convergence beyond 1 kcal/mol is of limited relevance because of the intrinsic error of the functional itself, but for some benchmark purposes (or for parameterizing functionals) one may wish to essentially eliminate the basis set incompleteness error in order to see the intrinsic error in isolation.

B. Basis set convergence in explicitly correlated double-hybrid calculations

Table III presents a statistical analysis for the explicitly correlated B2GP-PLYP-F12 calculations. Using the orbital A'VnZ basis sets in F12 calculations results in faster convergence to the basis set limit compared to the “conventional” B2GP-PLYP calculations. While errors for the B2GP-PLYP-F12/A'VDZ calculations are still unacceptably large (RMSD = 3.53 kcal/mol), B2GP-PLYP-F12/A'VTZ results become tolerable with a RMSD of 0.51 kcal/mol. Using the A'VQZ basis set results in a RMSD of only 0.14 kcal/mol, which is comparable to that of the “conventional” B2GP-PLYP/A'V5Z level of theory (0.16 kcal/mol).

The advantage of F12 calculations over those in pure one-electron basis sets, however, becomes more pronounced when one uses the VnZ-F12 basis sets. The B2GP-PLYP-F12/VDZ-F12 level of theory results in a RMSD of 1.46 kcal/mol for the entire test set (i.e., less than one-third of that obtained at the B2GP-PLYP/A'VDZ level of theory). The RMSD obtained

TABLE III. Statistical analysis of the basis set convergence in explicitly correlated B2GP-PLYP-F12 calculations for the set of 66 reaction energies (kcal/mol).^a

KS MP2	Single basis set approach ^b							Composite basis set approach ^c							
	A'VDZ	A'VTZ	A'VQZ	A'V5Z	VDZ-F12	VTZ-F12	VQZ-F12	A'VTZ	A'VQZ	A'VTZ	A'VQZ	A'TZVPP	A'QZVPP	A'PC-2	A'PC-3
	All 66 reactions														
RMSD	3.53	0.51	0.14	0.05	1.46	0.30	0.08	1.70	0.24	0.63	0.17	0.44	0.14	0.63	0.11
MAD	2.60	0.34	0.10	0.04	1.12	0.25	0.07	1.13	0.17	0.47	0.12	0.33	0.11	0.45	0.09
MSD	-0.03	-0.01	0.01	0.00	-0.23	-0.13	-0.02	-0.40	0.08	-0.13	-0.01	0.04	0.01	0.29	0.06
	46 reactions involving only first-row species														
RMSD	3.61	0.52	0.10	0.04	1.09	0.30	0.08	1.99	0.23	0.73	0.12	0.43	0.12	0.47	0.13
MAD	2.58	0.32	0.07	0.03	0.82	0.26	0.07	1.38	0.16	0.56	0.10	0.32	0.09	0.33	0.11
MSD	0.14	-0.11	0.01	0.01	-0.11	-0.14	-0.01	-0.55	0.12	-0.24	0.03	-0.04	0.03	0.14	0.07
LD	14.06	2.09	0.35	0.14	2.57	0.74	0.22	6.20	0.74	1.93	0.34	1.15	0.28	1.55	0.30
LD ^d	42	42	42	42	33	33	42	28	40	26	42	26	40	45	40
	20 reactions involving both first- and second-row species														
RMSD	3.33	0.49	0.20	0.08	2.07	0.30	0.09	0.69	0.25	0.29	0.25	0.47	0.18	0.90	0.07
MAD	2.66	0.38	0.15	0.06	1.80	0.24	0.07	0.55	0.19	0.25	0.19	0.36	0.15	0.72	0.06
MSD	-0.40	0.21	-0.01	0.00	-0.50	-0.10	-0.05	-0.04	0.00	0.11	-0.09	0.23	-0.04	0.64	0.03
LD	7.43	1.15	0.42	0.17	3.55	0.64	0.16	1.22	0.54	0.59	0.56	0.98	0.35	1.68	0.15
LD ^d	57	60	60	60	60	60	59	60	46	64	57	66	60	66	46

^aSee footnotes a and b of Table II.^bThe same basis set is used in the Kohn-Sham and MP2 steps.^cDifferent basis sets are used in the Kohn-Sham and MP2 steps, see text.^dSee footnote c of Table II.

with the VTZ-F12 basis set (0.30 kcal/mol) may already be acceptable even for benchmark purposes, while the VQZ-F12 results (RMSD = 0.08 kcal/mol) can essentially be regarded as basis set limits.

One should consider the possibility that in F12 calculations carried out with a relatively small basis set the MP2-like correlation energy actually converges faster to the basis set limit than the underlying Kohn-Sham calculation. The last eight columns of Table III consider calculations in which the MP2-F12 part is calculated with a relatively small basis set (of *spd* or *spdf* quality) and the KS energy is calculated with a larger basis set (i.e., a composite basis set approach). Calculating the MP2-like correlation energy with the VDZ-F12 basis set and the KS energy with the A'VTZ basis set results in a RMSD of 0.63 kcal/mol. This has a similar computational cost to the B2GP-PLYP-F12/VDZ-F12 level of theory (*vide infra*), but the RMSD is reduced by more than 50% (!). This seems to be a balanced combination of basis sets, for example, increasing the basis set size either in the KS part (to the A'VQZ basis set) or in the MP2-F12 part (to the VTZ-F12 basis set) results in a slight increase in the overall RMSD (specifically, to 0.73 and 0.66 kcal/mol, respectively). Likewise, calculating the KS energy with the A'VQZ basis set and the MP2-like correlation energy with the VTZ-F12 basis set results in a RMSD of only 0.17 kcal/mol (for comparison, the B2GP-PLYP-F12/VTZ-F12 level of theory gives a RMSD of 0.30 kcal/mol). Increasing the basis set size in the KS part (to the A'V5Z basis set) or in the MP2-F12 part (to the VQZ-F12 basis set) both results in a slightly lower RMSD of 0.13 kcal/mol. Furthermore, using a DFT-optimized basis set for the KS calculation may somewhat lower the RMSD. For example, replacing the A'VTZ basis set in the KS part with

the A'TZVPP and A'PC-2 basis sets results in RMSDs of 0.44 and 0.63 kcal/mol, respectively, in conjunction with the VDZ-F12 basis set for the MP2 part. Likewise, replacing the A'VQZ basis set in the KS part with the A'QZVPP and A'PC-3 basis sets results in RMSDs of 0.14 and 0.11 kcal/mol, respectively, in conjunction with the VTZ-F12 basis set for the MP2 part (see Table III).

It may seem counterintuitive to actually be using the *larger* basis set for the HF-like component and the smaller basis set for the MP2-like component. However, there are precedents in the literature on explicitly correlated calculations, notably a thermochemistry study by Klopper *et al.*³⁶ in which CCSD(F12)/def2-QZVPP calculations were supplemented with many corrections (for higher order correlation effects, inner-shell correlation, relativity, ...)—including a δ HF correction obtained as $E_{\text{HF/cc-pCV5Z}} - E_{\text{HF/cc-cvQZ}}$.

C. Computational cost considerations

It is of interest to compare the relative computational cost of the B2GP-PLYP-F12 procedure in conjunction with the A'VnZ and VnZ-F12 basis sets. Table IV compares the sizes of the orbital (A'VnZ) and explicitly correlated (VnZ-F12) basis sets used in the present work. The VnZ-F12 basis sets are marginally larger than the corresponding A'VnZ basis sets. In particular, the VnZ-F12 basis sets have an extra set of *s* and *p* functions on hydrogen, and an extra set of *s* and *2p* functions on the heavier elements.

Table V summarizes relative timings for the B2GP-PLYP-F12 calculations for benzene, naphthalene, anthracene, and some isomers. Inspection of Tables III and V reveals that the least expensive level of theory that still yields acceptable

TABLE IV. Summary of the basis sets used in the present work.

	Number of basis functions of each angular momentum			Total number of basis functions		
	H	First-row	Second-row	H	First-row	Second-row
A'VDZ	2s1p	4s3p2d	5s4p3d	5	23	32
A'VTZ	3s2p1d	5s4p3d2f	6s5p4d2f	14	46	55
A'VQZ	4s3p2d1f	6s5p4d3f2g	7s6p5d3f2g	30	80	89
A'V5Z	5s4p3d2f1g	7s6p5d4f3g2h	8s7p6d4f3g2h	55	127	136
VDZ-F12	3s2p	5s5p2d	6s6p3d	9	30	39
VTZ-F12	4s3p1d	6s6p3d2f	7s7p4d2f	18	53	62
VQZ-F12	5s4p2d1f	7s7p4d3f2g	8s8p5d3f2g	34	87	96

reaction energies (RMSD = 0.51 kcal/mol) is B2GP-PLYP-F12/A'VTZ. However, the composite basis set approach discussed above (i.e., calculating the Kohn-Sham energy with the A'VTZ and the MP2-like correlation energy with the VDZ-F12) results in a still respectable RMSD of 0.63 kcal/mol, at about half of the computational cost. Alternatively, one can consider the B2GP-PLYP-F12/VTZ-F12 level of theory which has a slightly higher computational cost, but results in a more respectable RMSD of 0.30 kcal/mol. To surpass this level of accuracy one can consider the use of B2GP-PLYP-F12/A'VQZ which attains a RMSD of 0.14 kcal/mol. Yet again, a more successful/cost-effective alternative would be to use the composite basis set approach; calculating the KS and MP2 steps with the A'VQZ and A'VTZ basis sets, respectively, results in an RMSD of 0.24 kcal/mol, and with the A'VQZ and VTZ-F12 basis sets, respectively, in an RMSD of only 0.17 kcal/mol. We also note with regard to the composite basis set approach, that the use of the A'VnZ and VnZ-F12 basis sets in the MP2 step results in similar computational costs, however, the use of the latter seems to be more cost-effective (see Tables III and V).

There are computational cost considerations other than CPU time if benchmark accuracy is required. For instance, the A'V5Z calculation on phenanthrene would entail 2 328 basis functions in C_{2v} symmetry, which in practice would require either laborious integral-direct MP2 calculations or combining a direct KS calculation with an RI-MP2 (Ref. 37) step.

D. Some illustrative examples

In this section, we apply the B2GP-PLYP functional (with and without the explicitly correlated F12 method) to

several hydrocarbons, namely, fulvene, benzene, naphthalene, anthracene, and phenanthrene. Table VI gives reaction energies of a few isogyric and isomerization reactions involving the above aromatic compounds. Let us start with the more challenging isogyric reactions. For the decomposition of benzene into three acetylene molecules, we were able to obtain the basis set limit reaction energy at the B2GP-PLYP-F12/A'V{Q,5}Z level of theory. Conventional B2GP-PLYP calculations in conjunction with the A'VDZ basis set give a reaction energy which is 9 kcal/mol above the basis set limit. Increasing the basis set size to A'VTZ reduces this deviation by almost two orders of magnitude, to 0.14 kcal/mol. We note that the explicitly correlated calculations in conjunction with either the A'VDZ or the A'VTZ basis sets result in similar deviations (see Table VI). However, the B2GP-PLYP-F12/VDZ-F12 level of theory results in a reaction energy which is only 0.37 kcal/mol below the basis set limit value, and increasing the basis set size in the KS step to A'VTZ results in a reaction energy which is 0.30 kcal/mol above the basis set limit value.

For the isogyric decomposition of naphthalene (+H₂) into acetylenes, our best reaction energy is obtained at the B2GP-PLYP-F12/VQZ-F12 level of theory. In conventional B2GP-PLYP calculations, this reaction energy converges rather slowly with respect to the basis set size. For example, the A'VDZ, A'VTZ, and A'VQZ basis sets yield reaction energies which are 15.36, 0.47, and 0.09 kcal/mol above our best value, respectively. B2GP-PLYP-F12 calculations in conjunction with the A'VDZ, A'VTZ, and A'VQZ basis sets give reaction energies which are 11.81, 0.19, and 0.16 kcal/mol above our basis set limit value, respectively. Moving to the VnZ-F12-type basis sets, we note that the VDZ-F12 basis set already results in an acceptable reaction

TABLE V. Relative timings for the B2GP-PLYP-F12 calculations for some hydrocarbons with up to 14 carbon atoms.^{a,b}

KS MP2	A'VDZ	A'VTZ	A'VQZ	A'V5Z	VDZ-F12	VTZ-F12	VQZ-F12	A'VTZ	A'VQZ	A'VTZ	A'VQZ
	A'VDZ	A'VTZ	A'VQZ	A'V5Z	VDZ-F12	VTZ-F12	VQZ-F12	A'VDZ	A'VTZ	VDZ-F12	VTZ-F12
Benzene	1.0	3.4	17.6	107.8	1.4	5.1	26.7	1.4	8.5	1.8	10.2
Fulvene	1.0	3.7	23.5	147.2	1.3	5.4	31.1	1.6	13.8	1.9	15.4
Naphthalene	1.0	3.3	17.1		1.3	5.1	20.6	1.3	8.5	1.6	10.2
Anthracene	1.0	3.5			1.4	6.4		1.4	8.3	1.7	11.1
Phenanthrene	1.0	3.7			1.4	5.1		1.5	6.4	1.9	7.8

^aThe timings are for both the KS and MP2 parts. The numbers given are the ratios between the CPU time for the different levels of theory and that for the B2GP-PLYP-F12/A'VDZ level of theory.

^bAll the calculations ran on 4 cores of otherwise idle dual Intel Nehalem X5550 systems (2.67GHz, 72 GB RAM).

TABLE VI. Basis set convergence of the B2GP-PLYP reaction energies with and without the F12 method for a few isogyric and isomerization reactions involving aromatic compounds (ΔE_e , at the bottom of the well, in kcal/mol).

KS MP2	B2GP-PLYP				B2GP-PLYP-F12									
	A'VDZ	A'VTZ	A'VQZ	A'V5Z	A'VDZ	A'VTZ	A'VQZ	A'V5Z	VDZ-F12	VTZ-F12	VQZ-F12	A'VTZ	A'VTZ	A'VQZ
	Isogyric reactions													
Benzene \rightarrow 3C ₂ H ₂	164.64	155.91	155.85	155.79	162.77	155.94	155.91	155.83	155.41	155.84	155.84	153.88	156.07	156.13
Naphthalene +H ₂ \rightarrow 5C ₂ H ₂	266.89	252.00	251.62		263.34	251.72	251.68		250.68	251.60	251.53	248.40	252.01	252.05
Anthracene +2H ₂ \rightarrow 7C ₂ H ₂	365.24	344.25			359.86	343.64			342.11	343.51		338.92	344.11	344.11
Phenanthrene +2H ₂ \rightarrow 7C ₂ H ₂	371.10	350.12			365.70	349.49			347.96	349.37		344.75	349.96	
	Isomerization reactions													
Benzene \rightarrow fulvene	35.61	35.67	35.79	35.80	35.68	35.76	35.80	35.80	35.61	35.81	35.81	35.70	35.80	35.80
Phenanthrene \rightarrow anthracene	5.86	5.87			5.84	5.85			5.85	5.86		5.83	5.85	

energy of 0.85 kcal/mol below our best estimate. However, increasing the size of the basis set in the KS step to an A'VTZ basis set leads to a reaction energy which is just 0.49 kcal/mol above our estimated basis set limit value.

For the decompositions of anthracene and phenanthrene (+2H₂) into acetylenes, our best estimate is obtained using the A'VQZ basis set for the KS part and the VTZ-F12 basis set for the MP2-like correlation energy. The reaction energies calculated with the A'VDZ basis set (either B2GP-PLYP or B2GP-PLYP-F12) overestimate this basis set limit value by 16–21 kcal/mol, whereas the B2GP-PLYP-F12/VDZ-F12 reaction energy is 2.00 kcal/mol below our best estimate. Using the orbital A'VTZ basis set gives a reaction energy that is 0.13 kcal/mol above our best estimates (with the B2GP-PLYP method), and 0.48 kcal/mol below our best estimates (with the B2GP-PLYP-F12 method).

Turning our attention to the isomerization energies, we note that the phenanthrene \rightarrow anthracene isomerization exhibits a very weak basis set dependence, such that even the B2GP-PLYP/A'VDZ level of theory seems to be fully converged. The benzene \rightarrow fulvene isomerization exhibits a slightly stronger basis set dependence; e.g., a basis set of *spd* quality results in isomerization energies of 0.12–0.19 below our best estimate. However, the composite basis set approach (in which the Kohn-Sham energy is calculated with the A'VTZ basis set and the MP2-like correlation energy with the VDZ-F12 basis set) reproduces the basis set limit value spot-on.

IV. CONCLUSIONS

The basis set convergence of the double-hybrid B2GP-PLYP functional has been investigated for a set of 66 reaction energies in conjunction with the explicitly correlated MP2-F12 approach. Two families of basis sets were considered: the orbital A'VnZ ($n = D, T, Q,$ and 5) basis sets and the VnZ-F12 ($n = D, T,$ and Q) basis sets optimized for use in F12 calculations. With regard to the explicitly correlated B2GP-PLYP-F12 calculations, we draw the following conclusions:

- F12 double-hybrid DFT calculations consistently yield reaction energies that are closer to the basis set limit than the conventional calculations done with basis sets saturated up to the same angular momentum. Typically, one “gains” one angular momentum in the explicitly correlated calculations.
- This improvement, however, is less dramatic than that observed for *ab initio* MP2 (or CCSD) calculations, owing to the fact that in the double-hybrid DFT calculations MP2-type correlation typically accounts for only between one-quarter and one-half of the overall correlation energy, with the remainder coming from the DFT correlation functional.
- In explicitly correlated F12 calculations the VnZ-F12 basis sets converge faster than the orbital A'VnZ basis sets.
- Finally, we find that basis set convergence of the MP2-F12 component is faster than that of the underlying Kohn-Sham calculation. Therefore, the most cost-effective approach consists of combining the MP2-F12 correlation energy from a comparatively small basis set such as VDZ-F12 with a DFT energy from a larger basis set such as A'VTZ.

ACKNOWLEDGMENTS

J.M.L.M. is on the leave of absence as the Thatcher Professor of Chemistry at the Weizmann Institute of Science. This research was supported in part by the Weizmann AERI (Alternative Energy Research Initiative) and by a startup grant from the University of North Texas. A.K. was supported by Australian Research Council Discovery Project Grant No. DP110102336. The authors would like to acknowledge a generous grant of computer time by the Faculty of Chemistry at the Weizmann Institute of Science.

¹J. P. Perdew and K. Schmidt, *AIP Conf. Proc.* **577**, 1 (2001).

²A. Görling and M. Levy, *Phys. Rev. B* **47**, 13105 (1993).

³S. Grimme, *J. Chem. Phys.* **124**, 034108 (2006).

- ⁴S. Kozuch and J. M. L. Martin, "DSD-PBEP86: In search of the best double-hybrid DFT with spin-component scaled MP2 and dispersion corrections," *Phys. Chem. Chem. Phys.* (to be published).
- ⁵A. Karton, A. Tarnopolsky, J.-F. Lamère, G. C. Schatz, and J. M. L. Martin, *J. Phys. Chem. A* **112**, 12868 (2008).
- ⁶S. Kozuch, D. Gruzman, and J. M. L. Martin, *J. Phys. Chem. C* **114**, 20801 (2010).
- ⁷L. Goerigk and S. Grimme, *J. Chem. Theory Comput.* **7**, 291 (2011).
- ⁸L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, *J. Chem. Phys.* **109**, 7764 (1998).
- ⁹L. A. Curtiss, P. C. Redfern, and K. Raghavachari, *J. Chem. Phys.* **126**, 084108 (2007).
- ¹⁰A. Karton, S. Daon, and J. M. L. Martin, *Chem. Phys. Lett.* **510**, 165 (2011).
- ¹¹S. F. Boys and N. C. Handy, *Proc. R. Soc. London, Ser. A* **310**, 43 (1969).
- ¹²E. A. Hylleraas, *Z. Phys.* **54**, 347 (1929); for an English translation, see: H. Hettema, *Quantum Chemistry: Classic Scientific Papers* (World Scientific, Singapore, 2000), Vol. 8, p. 81.
- ¹³T. Helgaker, W. Klopper, and D. P. Tew, *Mol. Phys.* **106**, 2107 (2008).
- ¹⁴E. F. Valeev, *Annu. Rep. Comp. Chem.* **2**, 19 (2007).
- ¹⁵W. Klopper, F. R. Manby, S. Ten-No, and E. F. Valeev, *Int. Rev. Phys. Chem.* **25**, 427 (2006).
- ¹⁶W. Klopper and W. Kutzelnigg, *Chem. Phys. Lett.* **134**, 17 (1987); W. Kutzelnigg and W. Klopper, *J. Chem. Phys.* **94**, 1985 (1991).
- ¹⁷W. Klopper and C. C. M. Samson, *J. Chem. Phys.* **116**, 6397 (2002).
- ¹⁸F. R. Manby, *J. Chem. Phys.* **119**, 4607 (2003); A. J. May, E. F. Valeev, R. Polly, and F. R. Manby, *Phys. Chem. Chem. Phys.* **7**, 2710 (2005).
- ¹⁹S. Ten-no, *Chem. Phys. Lett.* **398**, 56 (2004).
- ²⁰K. A. Peterson, T. B. Adler, and H.-J. Werner, *J. Chem. Phys.* **128**, 084102 (2008).
- ²¹J. G. Hill, K. A. Peterson, G. Knizia, and H.-J. Werner, *J. Chem. Phys.* **131**, 194105 (2009).
- ²²A. Karton and J. M. L. Martin (unpublished).
- ²³H.-J. Werner, P. J. Knowles, F. R. Manby, M. Schütz, *et al.*, MOLPRO, version 2010.1, a package of *ab initio* programs, 2010, see <http://www.molpro.net>.
- ²⁴H.-J. Werner, T. B. Adler, and F. R. Manby, *J. Chem. Phys.* **126**, 164102 (2007).
- ²⁵G. Knizia and H.-J. Werner, *J. Chem. Phys.* **128**, 154103 (2008).
- ²⁶T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
- ²⁷R. A. Kendall, T. H. Dunning, and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).
- ²⁸T. H. Dunning, K. A. Peterson, and A. K. Wilson, *J. Chem. Phys.* **114**, 9244 (2001).
- ²⁹F. Jensen, *J. Chem. Phys.* **115**, 9113 (2001); **116**, 3502 (2002).
- ³⁰F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).
- ³¹K. E. Yousaf and K. A. Peterson, *J. Chem. Phys.* **129**, 184108 (2008); *Chem. Phys. Lett.* **476**, 303 (2009).
- ³²E. F. Valeev, *Chem. Phys. Lett.* **395**, 190 (2004).
- ³³F. Weigend, *Phys. Chem. Chem. Phys.* **4**, 4285 (2002).
- ³⁴F. Weigend, A. Köhn, and C. Hättig, *J. Chem. Phys.* **116**, 3175 (2002); C. Hättig, *Phys. Chem. Chem. Phys.* **7**, 59 (2005).
- ³⁵F. Weigend, *Phys. Chem. Chem. Phys.* **8**, 1057 (2006) and references therein.
- ³⁶W. Klopper, R. Bachorz, C. Hättig, and D. P. Tew, *Theor. Chim. Acta* **126**, 289 (2010).
- ³⁷F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, *Chem. Phys. Lett.* **294**, 143 (1998); F. Weigend and M. Häser, *Theor. Chem. Acc.* **97**, 331 (1997).
- ³⁸See supplementary material at <http://dx.doi.org/10.1063/1.3647980> for basis set limit reaction energies calculated at the B2GP-PLYP-F12/A*V{Q,5}Z level of theory (Table SI). Basis set errors for all the reactions (Table SII). B3LYP-D/Def2-TZVP optimized geometries for fulvene, benzene, naphthalene, anthracene, and phenanthrene.