Communication: Efficient counterpoise corrections by a perturbative approach

Jia Deng, Andrew T. B. Gilbert, and Peter M. W. Gill[a)]

Research School of Chemistry, The Australian National University, ACT 0200, Australia

We investigate the use of Hartree-Fock and density functional perturbative corrections for estimating the counterpoise correction (CPC) for interaction energies at the self-consistent field level. We test our approach using several popular basis sets on the S22 set of weakly bound systems, which can exhibit large basis set superposition errors. Our results show that the perturbative approaches typically recover over 95% of the CPC and can be up to twelve times faster to compute than the conventional methods and therefore provide an attractive alternative to calculating CPCs in the conventional way. © *2011 American Institute of Physics*. [doi:10.1063/1.3632054]

I. INTRODUCTION

The interaction energy of a weakly bonded system is often found by the so-called supermolecular approach, which uses the energy difference

$$E^{\text{int}} = E_{\text{AB}}(\mathbf{A} \cup \mathbf{B}) - E_{\text{A}}(\mathbf{A}) - E_{\text{B}}(\mathbf{B}), \quad (1)$$

between the entire system and its subsystems, where subscripts denote the (sub)system, parentheses denote the basis set, and the fragment structures are taken from the complex.[1] However, this straightforward approach is complicated by the basis set imbalance between the system and its subsystems; subsystem A is modelled by the basis functions $\mathbf{A} \cup \mathbf{B}$ in the supermolecule, but only those in $\mathbf{A}$ in the isolated subsystem. This inconsistency systematically overestimates the interaction energy and gives rise to the so-called "basis set superposition error" (BSSE).[2,3] The BSSE is largest when small basis sets are used, and can become large enough to prevent a qualitative characterization of the interaction energy.

Because BSSE is a consequence of basis set incompleteness, it vanishes in the complete basis set limit. Therefore, a straightforward solution to reducing this error is to perform the calculations in sufficiently large bases. However, such a brute force approach is impractical for large systems and alternative methods for removing the BSSE must be used. The most important of these are the *a priori* Chemical Hamiltonian Approach (CHA),[4] Symmetry-Adapted Perturbation Theory (SAPT),[5] and the *a posteriori* counterpoise correction (CPC) of Boys and Bernardi.[6] Because of its conceptual simplicity, and ease of implementation, the CPC is the most widely used method and is also the focus of this present work.

For a system, AB, consisting of two subsystems A and B, the CPC is defined as

$$E^{\text{CPC}} = E_{\text{A}}(\mathbf{A}) - E_{\text{A}}(\mathbf{A} \cup \mathbf{B}) + E_{\text{B}}(\mathbf{B}) - E_{\text{B}}(\mathbf{A} \cup \mathbf{B}), \quad (2)$$

where, for example, $E_{\text{A}}(\mathbf{A} \cup \mathbf{B})$ is the energy of A calculated using the basis functions for both A and B. Because the basis functions associated with B have no nuclei in that calculation, they are referred to as ghost functions.

The cost of calculating the CPC is typically greater than that of calculating the original interaction energy. Given that the CPC provides only an estimate of the BSSE, it seems wise to consider approximate methods which mitigate this additional cost. Recently, we developed perturbative methods that "bootstrap" a small-basis calculation toward a corresponding large-basis calculation[7–10] and these approaches (and the related dual-basis methods[11–16]) have proven useful in approximating large-basis results at greatly reduced cost. Errors due to this formalism have been shown to be small, while reducing the computational cost by factors between 3 and 10. Originally proposed for Hartree-Fock theory, the techniques have been extended to density functional theory,[9] second-order Møller-Plesset perturbation theory (MP2) (Refs. 10 and 14) and, recently, to complete active space self-consistent field (CAS-SCF) theory.[17]

Although the ghost functions are a large fraction of the basis in a CPC calculation, their energetic contribution is small and such calculations are therefore ideal candidates for our perturbative methods or, indeed, for other dual-basis schemes.[11–16] In this Communication, we use our Hartree-Fock perturbative correction (HFPC) (Refs. 7 and 8) and density functional perturbative correction (DFPC) (Ref. 9) schemes to evaluate the effect of the ghost functions in CPC calculations.

Because of their inability to capture long-range dispersion effects, many SCF methods are poorly suited to describing weakly bound systems. However, this deficiency is ameliorated in the empirical DFT-based methods such as DFT-D (Refs. 18–20) and, because of their cost effectiveness, such methods are likely to continue to be popular for the study of systems with several dozen atoms. Accordingly, we focus on HF and DFT levels of theory and apply these to the S22 dataset[21] of weakly bounded complexes.

II. THEORY

The theory and performance of HFPC and DFPC calculations are described in detail in Refs. 7–9. Here we outline the

---

a)Electronic mail: peter.gill@anu.edu.au.

    135, 081105-1

HFPC scheme and refer the interested reader to our earlier papers for details.

An HFPC calculation begins with a converged SCF calculation in a primary basis of n functions, yielding molecular orbitals and an associated density matrix $\mathbf{P}$. $\mathbf{P}$ is then used to build the new Fock matrix

$$F_{ab}^{[1]} = h_{ab} + \sum_{\lambda\sigma}^{n} P_{\lambda\sigma}\left[(ab|\lambda\sigma) - \frac{1}{2}(a\lambda|b\sigma)\right], \qquad (3)$$

where λ, σ denote primary basis functions and a, b denote functions in a larger $(N > n)$ secondary basis. Diagonalization of $\mathbf{F}^{[1]}$ yields improved molecular orbitals and an improved density matrix $\mathbf{P}^{[1]}$. The HFPC energy

$$E_{\text{HFPC}} = \sum_{ab}^{N} P_{ab}^{[1]} h_{ab} + \frac{1}{2}\sum_{abcd}^{N} P_{ab}^{[1]} P_{cd}^{[1]}[(ab|cd) - \frac{1}{2}(ac|bd)], \qquad (4)$$

is more accurate[7,8] than that from other dual-basis methods because of its use of $\mathbf{P}^{[1]}\mathbf{P}^{[1]}$, rather than $\mathbf{P}\mathbf{P}^{[1]}$.

Although HFPC (and DFPC) are able to jump between any two basis sets, the selection of an optimal primary basis for a given secondary basis and accuracy threshold normally requires careful calibration studies. However, in the case of the CPC calculation of $E_A(\mathbf{A} \cup \mathbf{B})$, it is it is natural to adopt $\mathbf{A}$ as the primary basis and $\mathbf{A} \cup \mathbf{B}$ as the secondary basis, because the ghost functions, $\mathbf{B}$, perturb the energy only slightly. Moreover, such a basis pairing leads to a large N/n ratio which yields significant savings in computational time.

III. RESULTS

A. Computational details

We considered three standard basis sets: 6-31G(d), 6-31+G(d), and 6-31+G($2df$, p). For each of these, we calculated the exact CPC using fully converged SCF calculations and used these results as our reference values. The CPCs were also estimated using our perturbative schemes, which required evaluating each of the energies in Eq. (2) using either the HFPC or DFPC method. The accuracy was measured by the percentage of the CPC that was recovered by the perturbative method.

For the DFT calculations we used the B3LYP hybrid functional and the SG-1 quadrature grid.[22] In addition to the BSSE, Kohn-Sham CPC calculations are also affected by an analogous grid superposition error (GSE). Both the GSEs for the conventional and perturbative calculations were corrected in the same way using a counterpoise correction.

CPC values were calculated for the weakly bound systems in the S22 dataset,[21] which features a diverse selection of non-bonded interactions and a range of system sizes. All calculations were performed using a locally modified copy of the Q-Chem program[23] and the raw energies of all the systems are provided in the supplementary material.[24]

B. Accuracy

Columns 2–7 of Table I list the CPCs and the percentage recovered by the HFPC method for all the systems in the S22 set and for the three basis sets considered. BSSE reflects basis set incompleteness and our results confirm that CPCs typically decrease from 6-31G(d) to 6-31+G(d) to 6-31+G($2df$, p). This effect led Antony and Grimme to argue that polarized triple-ζ basis sets are sufficiently large that one can dispense with the CPC altogether[25] but, unfortunately, such basis sets are impractical for large systems.

Increasing the basis set size also systematically improves the relative performance of the HFPC method. Larger basis sets reduce the usefulness of the ghost functions and thereby allow HFPC to capture greater proportions of the BSSE. This improvement in relative performance is also a feature of the original HFPC method.[7] We see that, even for the small 6-31G(d) basis, where the CPCs are the largest, the HFPC method is still able to capture around 95%. For the larger basis sets, the percentage recovery is even better.

In absolute terms, the error introduced by using HFPC to calculate the CPC can be as large as 0.75 kJ/mol for the 6-31G(d) basis. The errors for formamide dimer, indole-benzene and adenine-thymine are all large. For 6-31+G(d) the maximum error reduces to 0.2 kJ/mol and, given that the counterpoise procedure is only an estimate of the BSSE, this discrepancy will be tolerable in many chemical applications.

Columns 8–13 of Table I show analogous results using DFPC with the B3LYP functional. The DFPC and HFPC results show similar trends but there are two significant differences. First, the B3LYP CPCs are significantly larger than those from HF. This observation was also made by Garza *et al.*[26] who attributed it to the tendency of DFT orbitals to be more diffuse than HF orbitals. Second, our perturbative approach recovers a greater percentage of the B3LYP CPCs than of the HF CPCs. Even for 6-31G(d), DFPC recovers at least 98% of the CPC and, for the larger basis sets, the recovery is usually >99.9%. This is consistent with our earlier results[9] and can be attributed to the fact that DFT energies tend to converge slightly faster than HF energies with respect to basis set size.

C. Efficiency

Calculating the exact CPC by Eq. (2) requires two additional calculations that use all of the basis functions in the system. This is expensive, typically taking almost twice as long as the calculation of the uncorrected interaction energy (1). Our perturbative schemes are much cheaper because the primary density matrix is generated as part of the interaction energy calculation, and therefore the computational overhead associated with our perturbative schemes amounts to a single Fock build and energy evaluation in the secondary basis.

The final columns of Table I show timing comparisons between CPC calculations via the conventional and perturbative approaches. The time taken to calculate the interaction energy, T^{int}, is simply the time required to compute the four energy terms in Eq. (1) but, because we can reuse some of these terms, the overhead of calculating Eq. (2) is only $T^{\text{CPC}} = t_{E_A(\mathbf{A}\cup\mathbf{B})} + t_{E_B(\mathbf{A}\cup\mathbf{B})}$. Columns 14–19 show the timing ratios $T^{\text{CPC}}/T^{\text{int}}$ where, for example, a ratio of 1 implies that

TABLE I. For the systems in the S22 dataset: (Columns 2–7) counterpoise corrections (kJ/mol) at the HF level for various basis sets and the percentage recovered by HFPC; (Columns 8–13) counterpoise corrections (kJ/mol) at the B3LYP level for various basis sets and the percentage recovered by DFPC; and (Columns 14–19) timing ratios for conventional and perturbative counterpoise calculations using the 6-31+G($2df, p$) basis.

Complex	HF						B3LYP						T^{CPC}/T^{int}					
	6-31G(d)		6-31+G(d)		+G($2df, p$)		6-31G(d)		6-31+G(d)		+G($2df, p$)		Exact		Perturbative		Pert./exact	
	CPC	%	CPC	%	CPC	%	CPC	%	CPC	%	CPC	%	HF	B3LYP	HF	B3LYP	HF	B3LYP
Ammonia dimer (C_{2h})	4.29	92.1	1.15	95.7	0.58	96.6	6.82	98.4	1.09	99.1	0.65	100.0	2.24	2.88	0.59	0.37	0.26	0.13
Water dimer (C_s)	3.90	93.6	4.48	95.5	1.72	96.5	6.80	98.4	4.57	99.3	1.63	99.4	1.34	1.49	0.38	0.21	0.28	0.14
Formic acid dimer (C_{2h})	11.06	94.4	7.30	97.7	2.92	97.9	17.53	98.5	5.45	99.8	2.07	100.0	2.18	1.97	0.38	0.29	0.17	0.15
Formamide dimer (C_{2h})	10.68	93.0	4.09	97.8	1.86	98.4	17.36	98.0	3.30	100.0	1.83	100.0	2.63	2.24	0.33	0.25	0.13	0.11
Uracil dimer (C_{2h})	9.16	94.1	5.18	98.1	2.40	98.3	14.82	98.3	3.95	100.0	2.39	100.0	1.59	2.19	0.25	0.24	0.16	0.11
2-Pyridoxine 2-aminopyridine (C_1)	9.48	93.5	4.94	98.0	1.99	98.0	15.15	98.2	3.68	100.0	2.34	100.0	1.22	1.08	0.15	0.12	0.13	0.11
Adenine thymine WC (C_1)	11.15	94.0	5.59	98.2	2.11	100.0	17.97	98.4	4.25	100.0	2.63	100.0	1.23	1.03	0.17	0.09	0.14	0.09
Methane dimer (D_{3d})	0.44	95.5	0.14	100.0	0.11	100.0	0.95	98.9	0.12	100.0	0.09	100.0	2.94	2.65	0.64	0.36	0.22	0.14
Ethene dimer (D_{2d})	3.33	93.7	0.68	97.1	0.50	96.0	3.84	99.0	0.70	100.0	0.67	100.0	2.93	2.62	0.72	0.42	0.24	0.16
Benzene methane (C_3)	2.05	95.6	0.57	98.2	0.48	97.9	2.64	99.2	0.68	100.0	0.64	100.0	1.89	0.82	0.34	0.14	0.18	0.17
Benzene dimer (C_{2h})	9.34	93.5	1.67	98.2	1.52	96.7	9.57	99.1	2.49	100.0	2.50	100.0	2.35	3.74	0.56	0.41	0.24	0.11
Pyrazine dimer (C_s)	8.21	94.4	3.24	97.8	1.91	97.9	8.38	99.2	2.25	100.0	1.87	100.0	1.74	1.60	0.25	0.22	0.15	0.14
Uracil dimer stack (C_2)	11.92	95.6	6.81	97.9	4.54	98.0	14.75	99.3	5.52	100.0	4.39	100.0	2.00	2.21	0.37	0.24	0.19	0.11
Indole benzene stack (C_1)	11.82	93.7	3.02	98.0	2.56	98.0	11.52	98.9	2.80	100.0	2.86	100.0	1.13	1.03	0.21	0.12	0.18	0.12
Adenine thymine stack (C_1)	15.36	95.1	8.58	98.0	4.73	98.1	18.06	99.1	5.37	100.0	4.08	100.0	1.43	1.80	0.17	0.14	0.12	0.08
Ethene ethyne (C_{2v})	2.67	92.1	0.52	98.1	0.38	97.4	2.80	98.6	0.81	100.0	0.64	100.0	1.57	1.27	0.36	0.18	0.23	0.14
Benzene water (C_s)	3.62	96.1	2.49	96.8	1.27	96.9	4.61	99.3	2.56	99.6	1.49	99.3	0.81	0.70	0.14	0.11	0.17	0.15
Benzene ammonia (C_s)	2.71	95.9	1.66	97.0	0.85	96.5	3.57	99.4	1.78	99.4	1.16	100.0	0.96	1.09	0.22	0.17	0.23	0.15
Benzene HCN (C_s)	2.93	96.2	1.16	98.3	0.83	97.6	3.29	99.4	1.27	100.0	1.19	100.0	1.17	1.51	0.23	0.22	0.20	0.15
Benzene dimer (C_{2v})	3.68	95.4	1.18	98.3	1.17	98.3	3.69	99.2	1.43	100.0	1.59	100.0	1.72	2.11	0.35	0.32	0.20	0.15
Indole benzene T-shape (C_1)	6.08	95.2	2.33	97.9	1.77	98.3	6.66	99.2	2.55	100.0	2.36	100.0	1.02	1.05	0.17	0.14	0.17	0.13
Phenol dimer (C_1)	7.52	95.3	5.31	97.4	2.52	97.2	10.86	99.0	5.28	99.8	3.02	100.0	0.89	1.19	0.14	0.17	0.16	0.14
Minimum		92.1		95.5		96.0		98.0		99.1		99.3	0.81	0.70	0.14	0.09	0.12	0.08
Mean		94.5		97.7		97.8		98.9		99.9		99.9	1.68	1.74	0.32	0.22	0.19	0.13
Maximum		96.2		100.0		100.0		99.4		100.0		100.0	2.94	3.74	0.72	0.42	0.28	0.17

the CPC is as expensive to obtain as all the energy terms in Eq. (1).

Interestingly, not only does the DFPC approach capture a slightly higher fraction of the CPC, it does so at a lower relative cost compared to HFPC. This is primarily because the DFT calculations employing ghost functions take longer to converge making the SCF calculation of $E_A(\mathbf{A} \cup \mathbf{B})$ and $E_B(\mathbf{A} \cup \mathbf{B})$ more expensive. Our perturbative scheme effectively includes only one of these SCF cycles and is therefore not plagued by this slower convergence.

IV. CONCLUDING REMARKS

Computational studies of weakly bound systems are important, especially in biological contexts and, in order to calculate accurate interaction energies in these systems, the BSSE must be corrected. This is usually done via the CPC and, in the past, this has necessitated calculations that are significantly more expensive than calculating the original interaction energy. In this work, we have applied the HFPC and DFPC perturbative schemes to estimate the CPC at between 1/4 and 1/12 of the cost of a conventional CPC calculation. Tests on the S22 set of weakly bounded complexes indicate that these schemes recover almost all of the CPC and often introduce no significant error at all.

Recently, other forms of CPCs, mainly dealing with intramolecular BSSEs, have been proposed.[27,28] These procedures involve dividing the molecule into many fragments and calculate each using a large fraction of the full basis. Our approach applies immediately to such calculations and offers major computational savings.

We conclude that the HFPC and DFPC schemes are effective tools for evaluating counterpoise corrections for both HF and DFT calculations.

ACKNOWLEDGMENTS

P.M.W.G. thanks the Australian Research Council for funding (Grant Nos. DP0984806 and DP1094170) and APAC for a generous allocation of supercomputer resources. J.D. thanks the ANU/RSC for a PhD scholarship.

[1]We note that this definition of the interaction energy does not include the relaxation energy associated with structural changes to the fragments upon complexation.

[2]H. B. Jansen and P. Ros, Chem. Phys. Lett. **3**, 140 (1969).

[3]B. Liu and A. D. McLean, J. Chem. Phys. **59**, 4557 (1973).

[4]I. Mayer, Int. J. Quantum Chem. **23**, 341 (1983).

[5]B. Jeziorski, B. Moszynski, and K. Szalewicz, Chem. Rev. **94**, 1887 (1994).

[6]S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).

[7]J. Deng, A. T. B. Gilbert, and P. M. W. Gill, J. Chem. Phys. **130**, 231101 (2009).

[8]J. Deng, A. T. B. Gilbert, and P. M. W. Gill, J. Chem. Phys. **133**, 044116 (2010).

[9]J. Deng, A. T. B. Gilbert, and P. M. W. Gill, Phys. Chem. Chem. Phys. **12**, 10759 (2010).

[10]J. Deng and P. M. W. Gill, J. Chem. Phys. **134**, 081103 (2011).

[11]K. Wolinski and P. Pulay, J. Chem. Phys. **118**, 9497 (2003).

[12]W. Liang and M. Head-Gordon, J. Phys. Chem. A **108**, 3206 (2004).

[13]R. P. Steele, R. A. DiStasio, Jr., Y. Shao, J. Kong, and M. Head-Gordon, J. Chem. Phys. **125**, 074108 (2006).

[14]R. P. Steele, Y. Shao, R. A. DiStasio, Jr., and M. Head-Gordon, J. Phys. Chem. A. **110**, 13915 (2006).

[15]R. P. Steele and M. Head-Gordon, J. Chem. Phys. **124**, 184108 (2006).

[16]R. P. Steele, R. A. DiStasio, Jr., and M. Head-Gordon, J. Chem. Theory Comput. **5**, 1560 (2009).

[17]L. Kong and E. F. Valeev, J. Chem. Phys. **133**, 174126 (2010).

[18]S. Grimme, J. Comput. Chem. **25**, 1463 (2004).

[19]S. Grimme, J. Comput. Chem. **27**, 1787 (2006).

[20]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[21]P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).

[22]P. M. W. Gill, B. G. Johnson, and J. A. Pople, Chem. Phys. Lett. **209**, 506 (1993).

[23]Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. V. Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C. P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Z. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M.W. Gill, and M. Head-Gordon, Phys. Chem. Chem. Phys. **8**, 3172 (2006).

[24]See supplementary material at http://dx.doi.org/10.1063/1.3632054 for raw energies.

[25]J. Antony and S. Grimme, Phys. Chem. Chem. Phys. **8**, 5287 (2006).

[26]J. Garza, J. Ramirez, and R. Vargas, J. Phys. Chem. A **109**, 643 (2005).

[27]F. Jensen, J. Chem. Theory Comput. **6**, 100 (2010).

[28]A. Galano and J. R. Alvarez-Idaboy, J. Comput. Chem. **27**, 1203 (2006).