

Gaussian-4 theory using reduced order perturbation theory

Larry A. Curtiss

Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA; Chemistry Division, Argonne National Laboratory, Argonne, Illinois 60439, USA; and Center for Nanoscale Materials, Argonne National Laboratory, Argonne, Illinois 60439, USA

Paul C. Redfern

Chemistry Division, Argonne National Laboratory, Argonne, Illinois 60439, USA

Krishnan Raghavachari

Department of Chemistry, Indiana University, Bloomington, Indiana 47401, USA

(Received 18 April 2007; accepted 18 July 2007; published online 26 September 2007)

Two modifications of Gaussian-4 (G4) theory [L. A. Curtiss *et al.*, *J. Chem. Phys.* **126**, 084108 (2007)] are presented in which second- and third-order perturbation theories are used in place of fourth-order perturbation theory. These two new methods are referred to as G4(MP2) and G4(MP3), respectively. Both methods have been assessed on the G3/05 test set of accurate experimental data. The average absolute deviation from experiment for the 454 energies in this test set is 1.04 kcal/mol for G4(MP2) theory and 1.03 kcal/mol for G4(MP3) theory compared to 0.83 kcal/mol for G4 theory. G4(MP2) is slightly more accurate for enthalpies of formation than G4(MP3) (0.99 versus 1.04 kcal/mol), while G4(MP3) is more accurate for ionization potentials and electron affinities. Overall, the G4(MP2) method provides an accurate and economical method for thermochemical predictions. It has an overall accuracy for the G3/05 test set that is much better than G3(MP2) theory (1.04 versus 1.39 kcal/mol) and even better than G3 theory (1.04 versus 1.13 kcal/mol). In addition, G4(MP2) does better for challenging hypervalent systems such as H₂SO₄ and for nonhydrogen species than G3(MP2) theory. © 2007 American Institute of Physics.
[DOI: 10.1063/1.2770701]

I. INTRODUCTION

The Gaussian-*n* series of model chemistries¹⁻⁵ was developed with the goal of calculating molecular energies within chemical accuracy. Recently, we reported details of our fourth generation model, labeled Gaussian-4 (G4) theory,⁶ that has significantly improved accuracy compared to its predecessor Gaussian-3 (G3) theory.⁴ G4 theory contains several new features not included in G3 theory that lead to improved performance. These features include an extrapolation to the HF limit, use of density functional geometries and zero-point energies, use of coupled cluster energies, an improved set of polarization functions for the G3Large basis set, and modification of the higher level correction (HLC) terms. G4 theory has an average absolute deviation from experiment of 0.83 kcal/mol from an assessment on the 454 energies in the G3/05 test set⁷ composed of enthalpies of formation, ionization energies, electron affinities, proton affinities, and hydrogen bond energies. This is a significant improvement over G3 theory that has an average absolute deviation of 1.13 kcal/mol for the same test set. In addition to our Gaussian-*n* methods, a number of successful approaches for calculating thermochemical data are available from other groups, including ones using very large basis sets with high levels of correlation energy⁸⁻¹⁴ such as Martin's *W_n* methods that can achieve very high accuracy but at considerable computational expense as well as those using more approximate composite techniques¹⁵⁻²¹ such as Petersson's

Complete Basis Set (CBS) methods that combine energies from lower level calculations, enabling calculations on larger systems at some loss in accuracy.

Modifications of the Gaussian-*n* methods, such as G2(MP2), G3(MP2), and G2(MP2, SVP), that require less computer time have been proposed by eliminating fourth- and third-order perturbation theory (MP4, MP3) calculations using large basis sets.²²⁻²⁵ These computationally less intensive methods have found wide use in a variety of applications. For example, G3(MP2) theory has recently been used in the calculation of gas phase acidities of a series of Brønsted acids.²⁶ Thus, it is of interest to consider analogous simplified models by elimination of some of the higher order perturbation theory energies in G4 theory and to assess the accuracy of the resulting methods.

In this paper, two modifications of G4 theory with reduced orders of perturbation theory are presented. The first method, referred to as G4(MP2) theory, eliminates the MP3 and MP4 large basis set calculations, while the second, G4(MP3) theory, eliminates only the MP4 large basis set energies. The accuracies of the two methods are assessed on the 454 energies in the G3/05 test set. It is found that both methods give significantly improved results compared to their G3 counterparts, G3(MP2) and G3(MP3). In fact, both new methods are more accurate than full G3 theory at a less cost. In Sec. II, the two new theoretical methods are described. In Sec. III, the G4(MP2) and G4(MP3) results for the G3/05 test set are presented. In Sec. IV, several aspects of

the methods are discussed in more detail and results for some challenging hypervalent systems not included in the test set are presented. Finally, conclusions are drawn in Sec. V.

II. DESCRIPTION OF G4(MP2) AND G4(MP3) METHODS

A. G4(MP2) theory

G4(MP2) theory is based on G4 theory.⁶ It uses B3LYP/6-31G(2df,p) optimized geometries for a series of single point energy calculations at higher levels of theory. In a previous study,⁵ this basis set was found to perform well for most geometries with some exceptions (hydrogen bonded complexes and Jahn-Teller distorted systems). The zero-point energy, $E(\text{ZPE})$, is based on B3LYP/6-31G(2df,p) frequencies scaled by 0.9854, the same as in G4 theory. We have previously analyzed the use of scaled vibrational frequencies in the G_n methods and found that they work well over a range of molecules.^{27,28} The first energy calculation is at the triples-augmented coupled cluster level of theory, CCSD(T), with the 6-31G(d) basis set, i.e., CCSD(T)/6-31G(d). This energy is then modified by a series of energy corrections to obtain a total energy E_0 ,

$$E_0[\text{G4(MP2)}] = \text{CCSD(FC,T)/6-31G(d)} + \Delta E_{\text{MP2}} + \Delta E_{\text{HF}} + \Delta E(\text{SO}) + E(\text{HLC}) + E(\text{ZPE}), \quad (1)$$

where FC denotes that the calculation is done using frozen core, i.e., only the valence electrons are correlated. The correction at the second-order Moller-Plesset level (MP2) is given by

$$\Delta E_{\text{MP2}} = [E(\text{MP2(FC)/G3MP2Large XP}) - [E(\text{MP2(FC)/6-31G(d)})]. \quad (2)$$

The G3MP2LargeXP basis set in Eq. (2) is the same as the G3LargeXP basis set used in G4 theory except that the core polarization functions of G3LargeXP are deleted because the MP2 calculation is done using frozen core orbitals. The G3LargeXP basis set has extra d functions added to the G3Large basis set of G3 theory as detailed in Ref. 6. The G3MP2LargeXP basis set is given in the supplementary information²⁹ and is also available on the web.³⁰

The other corrections in Eq. (1) are similar to those in G4 theory. The $\Delta E(\text{HF})$ is calculated as the difference between $E(\text{HF/G3MP2largeXP})$ and $E(\text{HF/limit})$ [i.e., $\Delta E(\text{HF}) = E(\text{HF/limit}) - E(\text{HF/G3MP2LargeXP})$]. The $E(\text{HF/limit})$ is calculated via a linear two-point extrapolation procedure^{31,32} using two large correlation consistent basis sets.³³⁻³⁵ There are several differences from the HF extrapolation procedure used in G4 theory. First, the extrapolation is based on triple- and quadruple-zeta basis sets, aug-cc-pVTZ and aug-cc-pVQZ, instead of quadruple- and quintuple-zeta basis sets used in G4. Second, these two basis sets have a tight d function added to the d polarization set of Al-Ar, denoted as aug-cc-pV(T+d)Z and aug-cc-pV(Q+d)Z.³⁶ Otherwise, they are similar to those used in G4 theory (see the Appendix for details). The justification for the change in the extrapolation procedure is discussed in the next section.

A spin-orbit term is included for atoms (from the experiment) and for molecules with first-order corrections (from accurate theoretical calculations), as detailed in Ref. 6.

The HLC is added to account for remaining deficiencies in the energies and has the same form as in G4 theory, with six parameters. The HLC includes $-An_\beta$ for closed shell molecules, $-A'n_\beta - B(n_\alpha - n_\beta)$ for open shell systems, and $-Cn_\beta - D(n_\alpha - n_\beta)$ for atoms and atomic ions. Here, n_β and n_α are the number of β and α valence electrons, respectively, with $n_\alpha \geq n_\beta$. There is also a correction parameter, E , for the energy of pairs of electrons in molecular and atomic nonhydrogen species having only one pair of valence electrons. These *single electron pair* species are special as the basis set requirements are considerably less stringent for these species than for those with more valence electrons. The HLC parameters are derived with a procedure similar to that used in G4 theory.⁶ The A , A' , B , C , D values are derived to give the smallest average absolute deviation from the experiment for the G3/05 test set and E is derived for the subset of 13 species in the G3/05 test that involve single pairs of electrons (see Ref. 6 for a list of these species). The parameters that yield the smallest average absolute deviation for the G3/05 test set are given in Table I.

The CCSD(T) and MP2 calculations are done with a frozen core approximation. The frozen core is the same as for G3 theory, i.e., $1s$ for Li-Mg, $1s, 2s, 2p$ for B-Ca, and $1s, 2s, 2p, 3s, 3p$ for Ga-Kr. Therefore, the $3d$ electrons on Ga-Kr, $3s, 3p$ on K, Ca, and $2s, 2p$ on Na, Mg are included in the correlation space. This has been defined as the "small core" treatment of correlation by Rassolov *et al.*³⁷ and is the same as in G4 theory. In rare cases, the standard method for choosing the frozen core (based on orbital energy criteria) fails and leads to large errors. In these cases, we use a population criteria³⁷ for determining the orbitals to be included in the correlation treatment. The only molecule in the G3/05 test set for which this changes the orbitals to be frozen is LiNa, where the difference with the experimental binding energy is reduced from -13.2 to -0.5 kcal/mol. The standard frozen orbital choices also fail in a few other molecules such as GaF₃ that are not included in the G3/05 test set.³⁸

Note that in the case of the third row non-transition-metal species, the HLC is based only on the valence electrons, i.e., $4s, 4p$, although more orbitals ($3d$ for Ga-Kr and $3s, 3p$ for K, Ca) are included in the correlation calculations as discussed above. This is also the case for Na and Mg where the $2s, 2p$ are included in the valence space. Finally, the zero-point correction $E(\text{ZPE})$ is obtained from scaled (0.9854) B3LYP/6-31G(2df,p) frequencies.⁵

In addition to the two HF calculations for extrapolation to the HF/limit, the calculation of the G4(MP2) energy requires only two single point energy calculations at the correlation level, CCSD(FC,T)/6-31G(d) and MP2(FC)/G3MP2largeXP. The absence of the MP4/6-31G(2df,p) and MP4/6-31+G(d,p) calculations in G4(MP2) theory compared to G4 theory provides significant savings in computational time.

TABLE I. Comparison of G4(MP3) and G4(MP2) methods with previous methods.

	G4	G4(MP3)	G4(MP2)	G3	G3(MP3)	G3(MP2)
Average absolute deviation (kcal/mol)						
Enthalpies of formation (270)	0.80	1.04	0.99	1.19	1.46	1.40
Nonhydrogens (79)	1.13	1.61	1.44	2.10	2.73	2.57
Hydrocarbons (38)	0.48	0.69	0.63	0.69	0.77	0.71
Subst. hydrocarbons (100)	0.68	0.78	0.83	0.82	0.92	0.88
Inorganic hydrides (19)	0.92	1.06	0.94	0.95	1.20	1.07
Radicals (34)	0.66	0.89	0.86	0.83	1.01	1.19
Ionization energies (105)	0.91	1.01	1.07	1.09	1.20	1.39
Atomic (26)	0.65	0.83	1.13	1.03	1.22	1.64
Molecular (79)	0.99	1.07	1.05	1.12	1.20	1.31
Electron affinities (63)	0.83	0.97	1.23	0.97	1.21	1.48
Atomic (14)	0.91	1.37	1.84	1.32	2.06	2.79
Molecular (49)	0.81	0.85	1.06	0.87	0.97	1.11
Proton affinities (10)	0.84	0.91	0.67	1.14	1.18	0.88
Hydrogen bonded complexes (6)	1.12	1.31	1.28	0.60	0.61	0.64
All (454)	0.83	1.03	1.04	1.13	1.35	1.39
G3/99 (376)	0.80	0.99	1.01	1.06	1.27	1.31
Root mean square deviation (kcal/mol)						
All (454)	1.19	1.52	1.49	1.67	2.08	2.09
Higher level correction (mhartree)						
A	6.947	8.620	9.472	6.386	7.902	9.279
B	2.441	3.179	3.102	2.977	3.684	4.471
C	7.116	8.376	9.741	6.219	7.368	9.345
D	1.414	2.323	2.115	1.185	1.983	2.021
A'	7.128	8.839	9.769			
E	2.745	3.420	2.379			

B. G4(MP3) theory

G4(MP3) theory is similar to G4(MP2) theory except that it includes third-order perturbation theory effects at the larger basis set levels as in G4 theory. As in G4(MP2) theory, the first energy calculation is at the CCSD(T)/6-31G(*d*) level. This energy is then modified by a series of energy corrections to obtain a total energy E_0 ,

$$\begin{aligned}
 E_0[\text{G4(MP3)}] = & \text{CCSD(FC, T)/6-31G}(d) + \Delta E_{\text{MP3}} \\
 & + \Delta E_{\text{MP2}} + \Delta E_{\text{HF}} + \Delta E(\text{SO}) + E(\text{HLC}) \\
 & + E(\text{ZPE}).
 \end{aligned} \quad (3)$$

The correction at the MP3 level is given by

$$\begin{aligned}
 \Delta E_{\text{MP3}} = & [E(\text{MP3(FC)/6-31}(2df,p))] \\
 & - [E(\text{MP3(FC)/6-31G}(d))].
 \end{aligned} \quad (4)$$

The correction at the MP2 level is given by

$$\begin{aligned}
 \Delta E_{\text{MP2}} = & [E(\text{MP2(Full)/G3 Large XP})] \\
 & - [E(\text{MP2(FC)/6-31G}(2df,p))].
 \end{aligned} \quad (5)$$

The basis sets in Eqs. (4) and (5) are the same as in G4 theory. The ΔE_{HF} correction in Eq. (3) is obtained in the same manner as described for G4(MP2) theory in the previous section. The $\Delta E(\text{SO})$, $E(\text{HLC})$, and $E(\text{ZPE})$ corrections in Eq. (3) are the same as those in G4 theory and are discussed for G4(MP2) theory. The HLC parameters for G4(MP3) are derived from the G3/05 test set and are also listed in Table I.

The G4(MP3) energy requires three single point energy calculations in addition to the two HF calculations for extrapolation to the HF/limit. The correlation level calculations are CCSD(FC, T)/6-31G(*d*), MP3(FC)/6-31G(2*df*, *p*), and MP2(Full)/G3large. The first calculation also provides the MP3(FC)/6-31G(*d*) energy required for Eq. (4) and the second calculation provides the MP2(FC)/6-31G(2*df*, *p*) energy required for Eq. (5). We have included all electrons in the MP2(Full)/G3LargeXP basis set calculation in Eq. (5), whereas the CCSD(T) and MP3 calculations in Eqs. (3) and (4) are performed with a frozen core treatment as described for G4(MP2) theory.

All calculations in this paper were done with the GAUSSIAN03 computer program.³⁹ The basis sets, total energies, and deviations from the experiment are available on the web³⁰ and in the supplementary information.²⁹

III. ASSESSMENT OF G4(MP2) AND G4(MP3) THEORIES

The G3/05 test set⁷ contains 454 experimental energies selected based on their small quoted uncertainties. This is the latest in a series of test sets of accurate experimental data that have been developed for assessing the accuracy of quantum chemical methods for energy calculations.^{40–42} The G3/05 test set contains enthalpies of formation of neutrals, atomization energies, ionization potentials, electron affinities, proton affinities, and hydrogen bond energies. The species in the test set have first row, second row, and third row

main group (K, Ca, Ga–Kr) elements. The enthalpies of formation at 298 K are calculated as described in Ref. 40, while the ionization potentials, electron affinities, atomization energies, and proton affinities are calculated at 0 K.

The total G4(MP2) and G4(MP3) energies of the atoms, molecules, and ions in the G3/05 test set were calculated as described in the previous section. These were then used to calculate the 454 energies in the G3/05 test set for comparison to the experiment. Table I contains a summary of the average absolute deviations and root mean square deviations of G4(MP2) and G4(MP3) theories from the experiment for several subsets of energies included in the G3/05 test set. The B3LYP/6-31G(2*df*,*p*) geometries of the species in the G3/05 test set are available on the internet.³⁰

A. G4(MP2) theory

The results in Table I indicate that for the 454 energies, the average absolute deviation at the G4(MP2) level is 1.04 kcal/mol. This result is not as good as G4 theory, which has an average absolute deviation of 0.83 kcal/mol, due to the neglect of higher orders of perturbation theory. However, G4(MP2) theory is a dramatic improvement over G3(MP2) theory²⁴ on this test set (1.39 kcal/mol) and is, in fact, more accurate than full G3 theory (1.13 kcal/mol). About 86% of the G4(MP2) deviations fall within the range -2.0 – $+2.0$ kcal/mol compared to 92% for G4, 85% for G3, and 77% for G3(MP2).

The average absolute deviation for G4(MP2) for enthalpies of formation (including atomization energies) is 0.99 kcal/mol for the 270 species in the G3/05 test set. This is significantly better than the performance of G3(MP2) theory (1.40 kcal/mol). The largest improvements are for nonhydrogens (2.57 versus 1.44 kcal/mol), inorganic hydrides (1.07 versus 0.94 kcal/mol), and radicals (1.19 versus 0.86 kcal/mol). G4(MP2) theory is also more accurate than G3 theory for enthalpies of formation (0.99 versus 1.19 kcal/mol) with the major improvement for nonhydrogens (1.44 versus 2.10 kcal/mol), while the other types of enthalpies of formation are about the same.

The average absolute deviations of G4(MP2) theory for ionization potentials, electron affinities, proton affinities, and hydrogen bonded complexes in the G3/05 test set are 1.07, 1.23, 0.67, and 1.28 kcal/mol, respectively, compared to 1.39, 1.48, 0.88, and 0.64 for G3(MP2). Except for the hydrogen bonded complexes, all of the other types of energies are improved in G4(MP2) relative to G3(MP2). The use of B3LYP/6-31(2*df*,*p*) for geometries causes the hydrogen bonded complexes to have larger errors in the G4 methodology.⁶ In the case of the electron affinities, G4(MP2) does much poorer for atomic species (1.84 kcal/mol) than for molecular species (1.06 kcal/mol), although in both cases they are better than G3(MP2). The results in Table I indicate that G4(MP2) is similar in accuracy to G3 for ionization potentials (1.07 versus 1.09 kcal/mol), better for proton affinities (0.67 versus 1.14 kcal/mol), and worse for hydrogen bonded complexes (1.28 versus 0.60 kcal/mol) and electron affinities (1.23 versus 0.97 kcal/mol). We have

TABLE II. Relative CPU times used in G3, G3(MP2), G4, and G4(MP2) single point energy calculations (from single processor times).

Method	SiCl ₄	Benzene	Hexane	Heptane
G3	1.0	1.0	1.0	1.0
G4	1.9	3.0	3.0	2.5
G3(MP2)	0.2	0.2	0.2	0.2
G4(MP2)	0.3	0.5	0.5	0.3

shown that the results for the hydrogen bonded complexes improve dramatically when diffuse functions are included in the basis set for geometry optimization.⁶

In Table I, we also include overall average absolute deviations of the various methods for the G3/99 test set⁷ of 376 energies for comparison. The conclusions concerning the relative performance of the methods are the same for the G3/99 test set as it is for G3/05. The relative times for G4(MP2) versus other *Gn* methods for some representative molecules are given in Table II. G4(MP2) is approximately six to eight times faster than G4, two to three times faster than G3, and two times slower than G3(MP2). The time difference with respect to G3(MP2) will become smaller for larger molecules due to the reduced scaling of the large basis set HF calculations (relative to the scaling of the correlation steps). Thus, G4(MP2) theory represents a significant improvement in accuracy over G3(MP2) theory for approximately the same cost in computer time. In addition, it is more accurate than G3 theory for significantly less cost in computer time.

B. G4(MP3) theory

The results in Table I indicate that for the 454 energies, the average absolute deviation at the G4(MP3) level is 1.03 kcal/mol. This is nearly the same as G4(MP2) theory (1.04 kcal/mol). The results in Table I indicate that G4(MP3) performs better than the G3 and G3(MP3) (Ref. 25) methods, which have average absolute deviations from the experiment of 1.13 and 1.35 kcal/mol, respectively. The average absolute deviation for G4(MP3) for enthalpies of formation (including atomization energies) is 1.04 kcal/mol for the 270 species in the G3/05 test set. This is slightly larger than for G4(MP2) theory (0.99 kcal/mol) with the main difference being for the nonhydrogen subset (1.61 versus 1.44 kcal/mol). The poorer performance of the MP3 version of G4 theory is consistent with a recent study indicating slow basis set convergence for third-order perturbation.⁴³ The average absolute deviations of G4(MP3) for ionization potentials, electron affinities, proton affinities, and hydrogen bonded complexes in the G3/05 test set are 1.01, 0.97, 0.91, and 1.31 kcal/mol, respectively. G4(MP3) does significantly better for electron affinities than G4(MP2).

Overall, the performance of G4(MP3) is not that much better than G4(MP2) and it also requires significantly more computer time due to the MP3 single point calculations. Therefore, for most purposes, G4(MP2) theory would be the best choice for a faster method than G4 theory.

TABLE III. Comparison of extrapolation of HF energies for the G3/05 test set.

	G4			G4(MP2)		
	(i) 4,5	(ii) 3,4	(iii) 3,4(td)	(i) 4,5	(ii) 3,4	(iii) 3,4(td)
	Average absolute deviation with expt. (kcal/mol)					
Enthalpies of formation (270)	0.80	1.03	0.83	0.99	1.05	0.99
Ionization energies (105)	0.91	0.92	0.91	1.07	1.08	1.07
Electron affinities (63)	0.83	0.87	0.85	1.23	1.25	1.23
Proton affinities (10)	0.84	0.92	0.82	0.65	0.71	0.67
Hydrogen bonded complexes (6)	1.12	1.14	1.14	1.30	1.28	1.28
All (454)	0.83	0.98	0.86	1.04	1.08	1.04
	Higher level correction (mhartree)					
<i>A</i>	6.947	6.867	6.855	9.469	9.363	9.472
<i>B</i>	2.441	2.305	2.291	3.109	3.075	3.102
<i>C</i>	7.116	6.941	6.989	9.765	9.591	9.741
<i>D</i>	1.414	1.430	1.408	2.076	2.073	2.115
<i>A'</i>	7.128	7.059	7.054	9.764	9.662	9.769
<i>E</i>	2.745	2.801	2.737	2.296	2.330	2.379

IV. NEW ASPECTS AND APPLICATIONS OF G4(MP2) AND G4(MP3)

A. Choice of basis sets for use in the Hartree-Fock energy extrapolation

As mentioned in Sec. II the basis sets used for extrapolation to obtain the HF limit in G4(MP2) theory are different from those in G4 theory. Since one of the primary motivations for this work is to reduce requirements for computer resources, we investigated two smaller basis set pairs for the HF extrapolation (for use in G4(MP2) theory) and compared them to the one used in G4 theory. The results are listed in Tables III and IV. The basis set pairs being compared are (i) aug-cc-pVQZ and aug-cc-pV5Z basis sets³³⁻³⁵ as modified for the 4,5 extrapolation in G4 theory, (ii) aug-cc-pVTZ and aug-cc-pVQZ basis sets³³⁻³⁵ for a 3,4 extrapolation, and (iii) the same basis sets as in (ii) except aug-cc-pV(T+d)Z and aug-cc-pV(Q+d)Z basis sets³⁶ are used for Al–Ar. In both (ii) and (iii), the basis sets have been modified as in G4 theory, i.e., using reduced number of diffuse functions and a smaller H basis set. The difference between (ii) and (iii) is the addition of a tight *d* function to the *d*-polarization set of Al–Ar with reoptimization of the exponents.³⁶ Martin has shown that additional tight *d* functions make important contribution to the total atomization calculated at the Hartree-

TABLE IV. Extrapolation of energies to HF limit: comparison with the 4,5 extrapolation for the G3/05 test set (in kcal/mol).

	(ii) 3,4	(iii) 3,4(td)
All species		
Average absolute deviation from 4,5	0.52	0.26
Max deviation from 4,5	-11.2, +0.51	-3.1, +0.72
All species except those with F		
Average absolute deviation from 4,5	0.38	0.15
Max deviation from 4,5	-7.3, +0.51	-1.3, +0.72

Fock level for some second row species.^{44,45} Such functions have also been considered by other authors at correlated levels of theory.⁴⁶⁻⁴⁹

Table III presents results for the different extrapolations for the G3/05 test set. They indicate that the 3,4 extrapolation including tight *d* functions gives an average absolute deviation from the experiment of 1.04 kcal/mol for the G3/05 test set, while the 3,4 extrapolation without tight *d* functions yields 1.08 kcal/mol. Interestingly, the former is the same as the overall accuracy based on using the 4,5 extrapolation in G4(MP2). In the case of G4 theory, the 3,4 extrapolation with tight *d* functions gives an average absolute deviation of 0.86 kcal/mol compared to 0.83 for the 4,5 extrapolation. The 3,4 extrapolation without tight *d* functions is significantly worse in this case (0.98 kcal/mol).

We investigated the extrapolations in more detail by examining differences in the projected HF energies since the results in Table III may be masked by the effects of the higher level correction (lower part of Table III). Table IV contains the average absolute deviations of the two 3,4 extrapolations from the 4,5 extrapolated HF limit. This assumes that the 4,5 extrapolation is close to convergence. We have done some tests using a 5,6 extrapolation that show little change in the extrapolated limit. Also, adding a tight *d* function to the basis sets of (i) has little effect on the extrapolated limit. The results in Table IV indicate that, for the atoms and molecules in the G3/05 test, the addition of the tight *d* function significantly improves the Hartree-Fock convergence for smaller basis sets. The 3,4 extrapolation with no tight *d*'s added, i.e., (ii), has an average absolute deviation of 0.52 kcal/mol from the 4,5 extrapolation and a maximum deviation of 11.2 kcal/mol. Addition of tight *d* functions to Al–Ar, i.e., (iii), reduces the average absolute deviation to 0.26 kcal/mol and the maximum deviation to 3.1 kcal/mol. For the 3,4 extrapolation without tight *d* functions, the largest deviations from 4,5 are primarily for molecules containing P–Cl atoms where the additional *d* function is especially important for convergence. In the case of the 3,4 extrapola-

TABLE V. Comparison of G4, G4(MP3), and G4(MP2) average absolute deviations with and without core correlation (CC). (Results are for the G3/05 test set and in each case the HLC was reoptimized.)

Type	Average absolute deviation (kcal/mol)					
	G4		G4(MP3)		G4(MP2)	
	CC	NoCC	CC	NoCC	CC	NoCC
Enthalpies of formation (270)	0.80	1.13	1.04	1.25	1.47	0.99
Ionization energies (105)	0.91	0.94	1.01	0.99	1.09	1.07
Electron affinities (63)	0.83	0.91	0.97	1.01	1.17	1.23
Proton affinities (10)	0.84	0.75	0.91	0.84	0.76	0.67
Hydrogen bonded complexes	1.12	1.13	1.31	1.39	1.21	1.28
All (454)	0.83	1.05	1.03	1.15	1.32	1.04

tion including the tight d 's, there are still deviations though they are much smaller, except for some fluorine containing compounds such as C_2F_6 . This indicates that further small improvement for the 3,4 extrapolation might be obtained by adding tight d functions to fluorine in the 3,4 basis sets.

Overall, we conclude that the 3,4 tight d -function extrapolation, though it has some residual differences from the 4,5 extrapolation, is adequate for the most part, and we have included it as part of G4(MP2) theory. In the case of full G4 theory, the more rigorous 4,5 extrapolation procedure is used.

B. Effect of core correlation on G4, G4(MP2), and G4(MP3) energies

Table V contains results for G4 and G4(MP3) when core-related correlation is not included. This is referred to as "NoCC" and is obtained by using the MP2(FC)/G3LargeXP energy instead of the MP2(Full)/G3LargeXP energy and reoptimizing the HLC parameters. The results in Table V indicate that the average absolute deviation of G4 theory increases from 0.83 to 1.05 kcal/mol, while that of G3(MP3) theory increases from 1.03 to 1.15 kcal/mol. The largest effect is on enthalpies of formation [the average absolute deviation of the G4 and G4(MP3) enthalpies increases from 0.80 to 1.13 kcal/mol and from 1.04 to 1.25 kcal/mol, respectively]; there is little change in the accuracy of the ionization potentials and electron affinities from core correlation. Hence, the inclusion of core-related correlation increases the accuracy of both the G4 and G4(MP3) methods on the G3/05 test set.

The situation is different for G4(MP2). In this case, we note that the method, as defined earlier, does not include core correlation effects. In contrast to G4 and G4(MP3), the average absolute deviation of G4(MP2) increases from

1.04 to 1.32 kcal/mol when the core correlation effects are included. In particular, there is a significant deterioration in the results for enthalpies of formation with the inclusion of core correlation effects. The reason for the opposite effect of core correlation in G4(MP2) is not clear, but there is obviously some fortuitous cancellation that gives rise to the accuracy that is achieved by G4(MP2) theory. This has also been found for G3(MP2) theory,^{24,25} and this is the main reason for not including core correlation effects in the basic definition of the G4(MP2) energy, in contrast to G4 or G4(MP3) that includes such effects.

C. Hexavalent sulfur systems

The G4 methods show a substantial improvement for the energies of nonhydrogen systems in the G3/05 test set compared to the corresponding G3 methods. We have also found significant improvement for some species that are not included in the test set. In a recent paper, Gutowski and Dixon²⁶ have pointed out that G3 and G3(MP2) theories perform poorly for enthalpies of molecules with hexavalent sulfur and suggested that this is due to problems inherent in the methods in handling large changes in the valence state of elements in molecules compared to their atomic states. They found errors of 8–10 kcal/mol in species such as H_2SO_4 . We have calculated G4 and G4(MP2) energies for three problem molecules (H_2SO_4 , FSO_3H , CH_3SO_3H) reported in the study of Gutowski and Dixon where the G3 methods are deficient.²⁶ The results are summarized in Table VI. In all cases, the error in the enthalpies of formation decreases substantially from those for the G3 methods. The G4 and G4(MP2) results for H_2SO_4 and FSO_3H are now within 2–4 kcal/mol of the experiment, which have relatively large uncertainties. They are within 2–3 kcal/mol of the best theoretical values of Gutowski and Dixon. The G4 and

TABLE VI. Results for hexavalent sulfur species.

	G4	G4(MP2)	G3	G3(MP2)	Expt. ^a ΔH_f^0 (298 K)	Theory ^b ΔH_f^0 (298 K)
SO_3	-92.6	-93	-89.4	-88.2	-94.59	
H_2SO_4	-171.8	-171.4	-168.3	-165.8	-175.7±2	-174.3
FSO_3H	-177.4	-177.3	-173	-170.7	-180.0±2	-179.5
CH_3SO_3H	-132.3	-132.0	-129.5	-127.3		-131.9

^aReference 52.

^bBest theoretical value from Ref. 26.

TABLE VII. Use of B3LYP/6-31G* geometries and zero-point energies in G4(MP2) theory.

G305 test set	Average absolute deviation from expt. (kcal/mol)	
	G4(MP2)	G4(MP2)//B3 ^a
Enthalpies of formation (270)	0.99	1.10
Nonhydrogens (79)	1.44	1.62
Hydrocarbons (38)	0.63	0.75
Subst. hydrocarbons (100)	0.83	0.95
Inorganic hydrides (19)	0.94	0.94
Radicals (34)	0.86	0.85
Ionization energies (105)	1.07	1.11
Electron affinities (63)	1.23	1.23
Proton affinities (10)	0.67	0.70
Hydrogen bonded complexes (6)	1.28	1.43
All (454)	1.04	1.12

^aBased on B3LYP/6-31G* geometries and zero-point energies (scale factor of 0.98) and the same HLC parameters as G4(MP2).

G4(MP2) results for CH₃SO₃H are within 0.3 kcal/mol of the theoretical values of Gutowski and Dixon. The table also includes results for SO₃, which was part of the G3/05 test set. The results in Table VI show significant improvement at the G4 levels of theory for this molecule with an error of less than 2 kcal/mol compared to the error of 5–6 kcal/mol at the corresponding G3 levels. The improvements in these systems are largely due to the inclusion of the Hartree-Fock extrapolation in the G4 and G4(MP2) methods and should also improve the performance on other high valency systems. The importance of the Hartree-Fock contribution to these types of systems has been pointed out by Martin.⁴⁴

D. Use of B3LYP/6-31G* geometries and zero-point energies

The G4(MP2) and G4(MP3) methods presented here are based on B3LYP/6-31G(2*df*,*p*) optimized geometries and zero-point energies. For the size of the molecules in the G3/05 test set, the increase in basis set size from 6-31G* in the G3 methods to 6-31G(2*df*,*p*) causes little problem in utilization of additional computer resources. However, in the application of G4(MP2) theory to very large systems, the use of a smaller basis set may be desirable, especially for the zero-point energy calculation step. This basis set has been used in the G3//B3LYP and G3(MP2)//B3LYP methods.⁵⁰ We have assessed the use of B3LYP/6-31G* geometries and zero-point energies in place of the B3LYP/6-31G(2*df*,*p*) geometries and zero-point energies in G4(MP2). We refer to this method as G4(MP2)//B3. The scale factor for the zero-point energies is 0.98 and the HLC is left unchanged. The results are summarized in Table VII. The average absolute deviation increases from 1.04 to 1.12 kcal/mol with the largest increase occurring for the nonhydrogen subset due to the sensitivity of some of these structures to the basis set. Little improvement is obtained with optimization of the HLC. Overall, the results are still slightly better than G3 theory. The G4(MP2)//B3 method should be a useful alternative if the use of B3LYP/6-31G* geometries and zero-

point energies is desirable. The deviations with experiment for the G3/05 test set are given in the supplementary information²⁹ and on the web.³⁰

V. CONCLUSIONS

The following conclusions can be drawn concerning the two modifications of G4 theory based on reduced orders of perturbation theory.

- (1) The neglect of third- and fourth-order perturbation theories in large basis set calculations in G4 theory leads to a method, G4(MP2) theory, that has an average absolute deviation of 1.04 kcal/mol for the G3/05 test set. This is substantially better than its counterpart in G3 theory based on reduced perturbation orders, G3(MP2), which has an average absolute deviation of 1.39 kcal/mol. It also has an overall accuracy that is better than G3 theory, especially for enthalpies of formation. Thus, the G4(MP2) method provides an economical method for accurate energy calculations.
- (2) The neglect of just the fourth-order perturbation theory in large basis set calculations in G4 theory, G4(MP3), is also significantly improved over its G3 counterpart. However, there is only a small difference between G4(MP2) and G4(MP3) so the former is recommended as a more economical method.
- (3) In addition, the G4 methods show significant improvement for hypervalent systems such as H₂SO₄ that have large errors in the G3 methods, consistent with the large improvement for these methods for the subset of nonhydrogens, including hypervalent species, in the G3/05 test set.

ACKNOWLEDGMENTS

The authors acknowledge grants of computer time at the Laboratory Computing Resources Center at Argonne National Laboratory. This work was supported by the U.S. Department of Energy, Division of Materials Sciences, under Contract No. DE-AC-02-06CH11357 and by a NSF grant, CHE-0616737, at Indiana University.

APPENDIX: HF BASIS SETS

The basis sets for extrapolation to the HF limit are based on aug-cc-pVTZ or aug-cc-pVQZ basis sets^{33–35} available from the EMSL database (<http://www.emsl.pnl.gov/forms/basisform.html>) with the exception of potassium.⁵¹ The basis sets were modified to save CPU time.

The modified triple-zeta basis sets for H and He use the *s* part of aug-cc-pVTZ combined with *p* polarization functions taken from cc-pVDZ. They have no diffuse functions. The modified triple-zeta basis set for Li–Mg uses the aug-cc-pVTZ basis set but includes only *s* and *p* diffuse functions (no *d*, *f* diffuse functions). The only exceptions to this are Na, Mg, K, and Ca, which have no diffuse functions added. For Al–Ar, the aug-cc-pV(T+*d*)Z basis set³⁶ is used with the same modifications to the diffuse functions.

TABLE VIII. Basis sets used in single point HF energy calculations G4(MP2) and G4(MP3) theories.

Atoms	aug-cc-pVTZ, aug-cc-pV(T+d)Z ^a		aug-cc-pVQZ, aug-cc-pV(Q+d)Z ^b	
	Literature ^c	Modified ^d	Literature ^c	Modified ^d
H, He	3s2p1d+diffuse spd	3s1p	4s3p2d1f+diffuse spd	4s2pd
Li–Ne	4s3p2d1f+diffuse spd	4s3p2d1f+diffuse sp	5s4p3d2f1g+diffuse spd	5s4p3d2f1g+diffuse sp
Na, Mg	5s4p2d1f	5s4p2d1f	6s5p3d2f1g	6s5p3d2f1g
Al–Ar	5s4p3d1f+diffuse spd	5s4p3d1f+diffuse sp	6s5p4d2f1g+diffuse spd	6s5p4d2f1g+diffuse sp
K, Ca	6s5p3d1f	6s5p3d1f	7s6p4d2f1g	7s6p4d2f1g
Ga–Kr	6s5p3d1f+diffuse spd	6s5p3d1f+diffuse sp	7s6p4d2f1g+diffuse spd	7s6p4d2f1g+diffuse sp

^aaug-cc-pV(T+d)Z is for Al–Ar; aug-cc-pVTZ is for all the other elements.

^baug-cc-pV(Q+d)Z is for Al–Ar; aug-cc-pVQZ is for all the other elements.

^cSee the Appendix for references.

^dModified basis set used in G4(MP2) and G4(MP3) theories (see the Appendix).

The modified quadruple-zeta basis sets for H and He use the *s* part of aug-cc-pVQZ combined with the *2pd* polarization functions from the smaller cc-pVTZ basis set. They have no diffuse functions. The modified quadruple-zeta basis sets for Li–Mg use the standard aug-cc-pVQZ basis set but include only *s* and *p* diffuse functions (no *d*, *f*, or *g* diffuse functions). The only exceptions to this are Na, Mg, K, and Ca, which have no diffuse functions added. For Al–Ar, the aug-cc-pV(Q+d)Z basis set³⁶ is used with the same modifications to the diffuse functions. All of the basis sets are summarized in Table VIII and are included in the supplementary information and are also available on the web.

¹J. A. Pople, M. Headgordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, *J. Chem. Phys.* **90**, 5622 (1989).

²L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.* **93**, 2537 (1990).

³L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, *J. Chem. Phys.* **94**, 7221 (1991).

⁴L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, *J. Chem. Phys.* **109**, 7764 (1998).

⁵L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.* **114**, 108 (2001).

⁶L. A. Curtiss, P. C. Redfern, and K. Raghavachari, *J. Chem. Phys.* **126**, 084108 (2007).

⁷L. A. Curtiss, P. C. Redfern, and K. Raghavachari, *J. Chem. Phys.* **123**, 124107 (2005).

⁸D. Feller and D. A. Dixon, *J. Chem. Phys.* **115**, 3484 (2001).

⁹T. H. Dunning, *J. Phys. Chem. A* **104**, 9062 (2000).

¹⁰A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kallay, and J. Gauss, *J. Chem. Phys.* **120**, 4129 (2004).

¹¹A. Tajti, P. G. Szalay, A. G. Csaszar, M. Kallay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vazquez, and J. F. Stanton, *J. Chem. Phys.* **121**, 11599 (2004).

¹²D. Feller and K. A. Peterson, *J. Chem. Phys.* **110**, 8384 (1999).

¹³J. M. L. Martin and G. de Oliveira, *J. Chem. Phys.* **111**, 1843 (1999).

¹⁴A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, *J. Chem. Phys.* **125**, 144108 (2006).

¹⁵N. J. DeYonker, T. Grimes, S. Yockel, A. Dinescu, B. Mintz, T. R. Cundari, and A. K. Wilson, *J. Chem. Phys.* **125**, 104011 (2006).

¹⁶G. P. F. Wood, L. Random, G. A. Petersson, E. C. Barnes, M. J. Frisch, and J. A. Montgomery, *J. Chem. Phys.* **125**, 094106 (2006).

¹⁷J. A. Montgomery, M. J. Frisch, J. W. Ochterski, and G. A. Petersson, *J. Chem. Phys.* **112**, 6532 (2000).

¹⁸J. A. Montgomery, M. J. Frisch, J. W. Ochterski, and G. A. Petersson, *J. Chem. Phys.* **110**, 2822 (1999).

¹⁹P. L. Fast, M. L. Sanchez, and D. G. Truhlar, *Chem. Phys. Lett.* **306**, 407 (1999).

²⁰P. L. Fast, J. C. Corchado, M. L. Sanchez, and D. G. Truhlar, *J. Phys. Chem. A* **103**, 5129 (1999).

²¹B. J. Lynch and D. G. Truhlar, *J. Phys. Chem. A* **107**, 3898 (2003).

²²L. A. Curtiss, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.* **98**, 1293 (1993).

²³L. A. Curtiss, P. C. Redfern, B. J. Smith, and L. Radom, *J. Chem. Phys.* **104**, 5148 (1996).

²⁴L. A. Curtiss, P. C. Redfern, K. Raghavachari, V. Rassolov, and J. A. Pople, *J. Chem. Phys.* **110**, 4703 (1999).

²⁵L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, *Chem. Phys. Lett.* **313**, 600 (1999).

²⁶K. E. Gutowski and D. A. Dixon, *J. Phys. Chem. A* **110**, 12044 (2006).

²⁷L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, *Chem. Phys. Lett.* **270**, 419 (1997).

²⁸P. C. Redfern, P. Zapol, L. A. Curtiss, and K. Raghavachari, *J. Phys. Chem. A* **104**, 5850 (2000).

²⁹See EPAPS Document No. E-JCPSA6-127-304733 for basis sets, energies, and deviations from experiments. This document can be reached through a direct link in the online article's HTML reference section or via the EPAPS homepage (<http://www.aip.org/pubservs/epaps.html>).

³⁰chemistry.anl.gov/compmat/compterm.htm

³¹A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper, and J. Olsen, *Chem. Phys. Lett.* **302**, 437 (1999).

³²D. Feller, *J. Chem. Phys.* **98**, 7059 (1993).

³³A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning, *J. Chem. Phys.* **110**, 7667 (1999).

³⁴D. E. Woon and T. H. Dunning, *J. Chem. Phys.* **98**, 1358 (1993).

³⁵T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).

³⁶T. H. Dunning, K. A. Peterson, and A. K. Wilson, *J. Chem. Phys.* **114**, 9244 (2001).

³⁷V. A. Rassolov, J. A. Pople, P. C. Redfern, and L. A. Curtiss, *Chem. Phys. Lett.* **350**, 573 (2001).

³⁸L. A. Curtiss, P. C. Redfern, V. Rassolov, G. Kedziora, and J. A. Pople, *J. Chem. Phys.* **114**, 9287 (2001).

³⁹M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 03, Gaussian, Inc., Wallingford, CT, 2004.

⁴⁰L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, *J. Chem. Phys.* **106**, 1063 (1997).

⁴¹L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.* **109**, 42 (1998).

⁴²L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, *J. Chem. Phys.* **112**, 7374 (2000).

⁴³K. Kahn, A. A. Granovsky, and J. Noga, *J. Comput. Chem.* **28**, 547 (2007).

⁴⁴J. M. L. Martin, *J. Mol. Struct.: THEOCHEM* **771**, 19 (2006).

⁴⁵J. M. L. Martin, *Chem. Phys. Lett.* **310**, 271 (1999).

⁴⁶A. K. Wilson and T. H. Dunning, *J. Phys. Chem. A* **108**, 3129 (2004).

⁴⁷C. W. Bauschlicher and A. Ricca, *J. Phys. Chem. A* **102**, 8044 (1998).

⁴⁸J. M. L. Martin, *J. Chem. Phys.* **108**, 2791 (1998).

⁴⁹A. K. Wilson and T. H. Dunning, *J. Chem. Phys.* **119**, 11712 (2003).

⁵⁰A. G. Baboul, L. A. Curtiss, P. C. Redfern, and K. Raghavachari, *J. Chem. Phys.* **110**, 7650 (1999).

⁵¹K. A. Peterson (private communication).

⁵²M. W. Chase, Jr., C. A. Davies, J. R. Downey, Jr., D. J. Frurip, R. A. McDonald, and A. N. Syverud, *J. Phys. Chem. Ref. Data* **14**, suppl. 1 (1985).