Gaussian-4 theory

Larry A. Curtiss
Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439; Chemistry Division, Argonne National Laboratory, Argonne, Illinois 60439; and Center for Nanoscale Materials, Argonne National Laboratory, Argonne, Illinois 60439

Paul C. Redfern
Chemistry Division, Argonne National Laboratory, Argonne, Illinois 60439

Krishnan Raghavachari
Department of Chemistry, Indiana University, Bloomington, Indiana 47401

(Received 16 October 2006; accepted 3 January 2007; published online 28 February 2007)

The Gaussian-4 theory (G4 theory) for the calculation of energies of compounds containing first- (Li–F), second- (Na–Cl), and third-row main group (K, Ca, and Ga–Kr) atoms is presented. This theoretical procedure is the fourth in the Gaussian-*n* series of quantum chemical methods based on a sequence of single point energy calculations. The G4 theory modifies the Gaussian-3 (G3) theory in five ways. First, an extrapolation procedure is used to obtain the Hartree-Fock limit for inclusion in the total energy calculation. Second, the *d*-polarization sets are increased to 3*d* on the first-row atoms and to 4*d* on the second-row atoms, with reoptimization of the exponents for the latter. Third, the QCISD(T) method is replaced by the CCSD(T) method for the highest level of correlation treatment. Fourth, optimized geometries and zero-point energies are obtained with the B3LYP density functional. Fifth, two new higher level corrections are added to account for deficiencies in the energy calculations. The new method is assessed on the 454 experimental energies in the G3/05 test set [L. A. Curtiss, P. C. Redfern, and K. Raghavachari, J. Chem. Phys. **123**, 124107 (2005)], and the average absolute deviation from experiment shows significant improvement from 1.13 kcal/mol (G3 theory) to 0.83 kcal/mol (G4 theory). The largest improvement is found for 79 nonhydrogen systems (2.10 kcal/mol for G3 versus 1.13 kcal/mol for G4). The contributions of the new features to this improvement are analyzed and the performance on different types of energies is discussed. [DOI: 10.1063/1.2436888]

I. INTRODUCTION

A number of approaches, based on quantum chemical methods, have been developed over the past decade and a half to make accurate predictions of thermochemical data. The Gaussian-*n* (G*n*) theories ($n=1,2,3$),[1–3] which we have developed, employ a set of calculations with different levels of accuracy and basis sets with the goal of approaching the exact energy. In the G*n* approach, a high level correlation calculation [e.g., QCISD(T) and CCSD(T)] with a moderate sized basis set is combined with energies from lower level calculations (e.g., MP4 and MP2) with larger basis sets to approximate the energies of more expensive calculations. In addition, several molecule-independent empirical parameters [higher level correction (HLC) terms] are included to estimate remaining deficiencies, assuming that they are systematic. An alternate and more accurate approach for the calculation of thermochemical data is based on CCSD(T) calculations using very large correlation consistent basis sets extrapolated to the complete basis set limit with addition of corrections for some smaller effects not included in the calculations such as core-valence effects, relativistic effects, and atomic spin-orbit effects.[4–11] This type of approach is limited to smaller molecules because of the use of very large basis sets. An intermediate approach referred to as correlation consistent composite approach (ccCA) that uses correlation consistent basis sets with no parametrization has recently been introduced.[12,13] Other composite techniques related to the Gaussian-*n* methods have also been introduced. These include the complete basis set methods of Montgomery, Jr. *et al.*[14,15] and the multicoefficient methods of Truhlar and co-workers.[16,17]

Concurrent with the development of the Gaussian-*n* theories, we have compiled a series of data sets of accurate experimental data, which have been used in the assessment of theoretical methods for energy calculations. The first in this series was the G2 test set of 125 energies.[2] This was followed by the G2/97 (301 energies),[18] G3/99 (376 energies),[19] and G3/05 (454 energies)[20] test sets. Each succeeding test set included energies from the preceding test sets and additional species of larger sizes and of different types. The test sets contain thermochemical data such as enthalpies of formation, ionization potentials, electron affinities, and proton affinities chosen based on a listed accuracy of ±1 kcal/mol or better in critical compilations. The latest test set, G3/05,[20] contains 270 enthalpies of formation, 105 ionization energies, 63 electron affinities, 10 proton affinities, and 6 hydrogen-bonded complexes. The expansion from G3/99 to G3/05 was done by including 14 new enthalpies of formation of nonhydrogen species, 58 energies of molecules containing third-row elements, and 6 hydrogen bonded com-

plexes. The nonhydrogens were added because these have generally been the most difficult type of molecule for the G*n* methods to handle. The energies of the third-row elements and hydrogen bonded complexes were added to the test set for the first time in G3/05. This new test set provides a more rigorous database with which to evaluate quantum chemical methods than the previous ones.

When the G3 theory was originally published, it was assessed on the G2/97 test set and was found to have an average absolute deviation of 1.02 kcal/mol from experiment.[3] The two succeeding test sets, G3/99 and G3/05, gave average absolute deviations of 1.07 and 1.13 kcal/mol, respectively, for the G3 theory.[19,20] The increase in the error is partially due to the poor performance of the method on the subset of first- and second-row nonhydrogen species with the deviation of that subset increasing from 1.66 kcal/mol (G2/97) to 2.10 kcal/mol (G3/99) to 2.34 kcal/mol (G3/05). Each succeeding nonhydrogen subset contains, on average, larger species, which is also responsible for the increase in the average error. (The average absolute deviation given above for the nonhydrogen subset of G3/05 does not include 18 species from the third-row entries which are smaller and, when included in the statistics, reduce the error to 2.10 kcal/mol.) The other contributor to the increase in deviation from the G2/97 test set is an increase in the error in the substituted hydrocarbon and inorganic hydride subsets, although these are still under 1 kcal/mol. Thus, molecules with no hydrogens are the most challenging for the G3 theory.

In this paper, we describe modifications to the G3 theory that help to improve its overall performance as well as its performance on the nonhydrogen species. We refer to the resulting method as the G4 theory. This methodology is consistent with the G*n* approach in past publications: i.e., it is a composite technique aimed at getting accurate energies without requiring extensive computer resources. This approach depends on a cancellation of errors as well as well-defined parameters to achieve this. The modifications included in the G4 theory include (1) an extrapolation procedure to obtain the Hartree-Fock (HF) limit for inclusion in the total energy calculation, (2) increase of the *d*-polarization sets to 3*d* on the first-row atoms and to 4*d* on the second-row atoms, with reoptimization of the exponents for the 4*d* set, (3) the replacement of the QCISD(T) method by CCSD(T), (4) geometries and zero-point energies obtained at the B3LYP/6-31G(2*df*,*p*) level, and (5) two new higher level correction parameters to account for deficiencies in radicals and in species having only one electron pair in the valence space. The HF energy extrapolation used here eliminates any error that may be present due to the incompleteness of the basis set in the HF energy, thus reducing the sources of error in the calculation. This was partially implemented in a version of the G3 theory published in 2001, referred to as the G3X theory,[21] in which a *g* function was added to the second-row atoms. The use of density functional geometries and zero-point energies was also included in the G3X method. The details of these five modifications are described in Sec. II. It is shown in Sec. III that the G4 theory gives a significant overall improvement on the G3/05 test set, particularly for nonhydrogens. In Sec. III, we also discuss reasons for the remaining outliers, i.e., molecules with errors larger than 2 kcal/mol. We present a "complete" version of the G4 theory that is based on a single calculation using the full basis set. This method, G4 (complete), corrects a problem with the G3 theory, where calculations without additivity approximation have a significantly larger average absolute deviation than the version with additivity. Conclusions are given in Sec. IV.

II. DESCRIPTION OF THE GAUSSIAN-4 THEORY

The Gaussian-4 theory is a composite technique in which a sequence of well-defined *ab initio* molecular orbital calculations is performed to arrive at a total energy for a given molecular species. The steps in the G4 theory and the differences with the G3 theory[3] are as follows.

(1) The equilibrium structure is obtained at the B3LYP/6-31G(2*df*,*p*) level. Spin-restricted theory is used for singlet states and spin-unrestricted theory for others. This differs from the G3 theory in which the geometries are calculated at the MP2(full)/6-31G* level. As shown in our paper on the G3X theory,[21] the use of the B3LYP density functional theory[22] for geometries leads to an improvement in overall results compared to using the MP2 theory, which was used in the formulation of G1, G2, and G3 theories.

(2) The B3LYP/6-31G(2*df*,*p*) equilibrium structure is used to calculate harmonic frequencies, which are then scaled by a factor of 0.9854 (Ref. 21) to take account of known deficiencies at this level. These frequencies give the zero-point energy E(ZPE) used to obtain E_0. This is a change from the G3 theory, in which HF/6-31G* was used for zero-point energies and the scale factor was appropriate for vibrational frequencies.[23] In most cases this modification should be more reliable.

(3) The Hartree-Fock energy limit E(HF/limit) is calculated. This is a new step that was not included in previous methods. The Hartree-Fock basis set limit is determined using a linear two-point extrapolation scheme[24,25] and Dunning's aug-cc-pV*n*Z basis sets,[26–28]

$$E_{\mathrm{HF/aug\text{-}cc\text{-}pV}n\mathrm{Z}} = E_{\mathrm{HF/limit}} + B\exp(-\alpha n), \tag{1}$$

where n is the number of contractions in the valence shell of the basis set and α is an adjustable parameter. The above formula yields a set of two linear equations with two unknowns from which the Hartree-Fock limit can be determined analytically, $E_{\mathrm{HF/limit}}(n,n+1) = (E_{\mathrm{HF}/n+1} - E_{\mathrm{HF}/n}\exp(-\alpha))/(1-\exp(-\alpha))$. We investigated various pairs of $n,n+1$ values and α values. We found that calculating the Hartree-Fock limit using $n=4$, $n+1=5$ (aug-cc-pVQZ and aug-cc-pV5Z basis sets), and $\alpha=1.63$ gave nearly converged values for a set of large molecules from the G3/05 test set. In order to reduce the computational time required, we modified the standard aug-cc-pVQZ and aug-cc-pV5Z basis sets by reducing the number of diffuse functions on heavy atoms and by reducing the hydrogen basis set as described in Appendix A. These basis set modifications

saved significant computer time, without reducing the accuracy.

(4) A series of single point correlation energy calculations is then carried out. The first is based on the complete fourth-order Moller-Plesset perturbation theory with the 6-31G(d) basis set, MP4/6-31G(d). This is modified by corrections from additional calculations.

(a) a correction for diffuse functions,

$$\Delta E(+) = E[\text{MP4/6-31+G}(d)] - E[\text{MP4/6-31G}(d)], \tag{2}$$

(b) a correction for higher polarization functions,

$$\Delta E(2df,p) = E[\text{MP4/6-31G}(2df,p)] - E[\text{MP4/6-31G}(d)], \tag{3}$$

(c) a correction for correlation effects beyond a fourth-order perturbation theory using a coupled cluster theory,

$$\Delta E(\text{CC}) = E[\text{CCSD(T)/6-31G}(d)] - E[\text{MP4/6-31G}(d)], \tag{4}$$

(d) a correction for larger basis set effects and for the nonadditivity caused by the assumption of separate basis set extensions for diffuse functions and higher polarization functions,

$$\Delta E(\text{G3LargeXP}) = E[\text{MP2(full)/G3LargeXP}] - E[\text{MP2/6-31G}(2df,p)] - E[\text{MP2/6-31+G}(d)] + E[\text{MP2/6-31G}(d)]. \tag{5}$$

The basis sets in (a)–(c) are the same as those in the G3 theory.[3] In step (d) the G3LargeXP basis is used instead of the G3Large basis set[3] to correct for some additivity problems discussed in Sec. III. The difference between the two basis sets is added d-polarization functions in G3LargeXP (with XP standing for extra polarization functions). The new exponents and a description of how they were obtained are given in Appendix A. The $2df$ polarization set in G3Large on the first row is replaced by a $3df$ set in G3LargeXP, the $3d2f$ polarization functions on the second row (Al–Cl) are replaced by $4d2f$, and no changes are made for Na, Mg, K, Ca, and Ga–Kr. The other difference in step 4 is that the QCISD(T)/6-31G* calculation is replaced by CCSD(T)/6-31G*. This is done because in some cases the QCISD(T) method has rather dramatic failures, which does not occur for the CCSD(T) method.[29,30] The MP4 and CCSD(T) calculations are done in the frozen core approximation, while the MP2 calculation with the large basis set is done with all electrons correlated. The electrons included in the frozen core for the G4 theory are the same as those for the G3 theory; i.e., the $3d$ on Ga–Kr and $3s$ and $3p$ on K and Ca are included in the correlation space and the $2s,2p$ on Na and Mg are included in the valence space. This has been defined elsewhere[31] as the "small core" treatment of correlation.

(5) The MP4/6-31G(d) energy and the four correlation corrections from step 4 are combined in an additive manner along with a correction for the HF limit (step 3) and a spin-orbit correction, $\Delta E(\text{SO})$

$$E(\text{combined}) = E[\text{MP4/6-31G}(d)] + \Delta E(+) + \Delta E(2df,p) + \Delta E(\text{CC}) + \Delta E(\text{G3LargeXP}) + \Delta E(\text{HF}) + \Delta E(\text{SO}). \tag{6}$$

The $\Delta E(\text{HF})$ is calculated as the difference between $E(\text{HF/G3LargeXP})$ calculated in step 4 and $E(\text{HF/limit})$ calculated in step 3 [i.e., $\Delta E(\text{HF}) = E(\text{HF/limit}) - E(\text{HF/G3LargeXP})$]. A spin-orbit term is included for atoms, as described previously for the G3 theory for the first- and second-row[3] and for the third-row main group elements.[32–34] This is taken from experiment[35] where available and accurate calculations[3,36] elsewhere. The atomic spin-orbit corrections are listed in Table I. A spin-orbit correction is also included for all molecules with first order corrections ($^2\Pi$ diatomics). In this case the spin-orbit correction is taken from accurate theoretical calculations and are given in Ref. 32.

(6) A HLC is added to take into account remaining deficiencies in the energy calculations,

$$E_e(\text{G4}) = E(\text{combined}) + E(\text{HLC}). \tag{7}$$

The form of the HLC is the same as for the G3 theory[3] except that two additional parameters (A' and E) have been added. The HLC parametrization used in the G4 theory is $-An_\beta$ 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 (including atomic ions). The n_β and n_α are the number of β and α valence electrons, respectively, with $n_\alpha \geqslant n_\beta$. The number of valence electron pairs corresponds to n_β. The A' parameter has been added to the G4 theory to account for deficiencies in pairs of electrons in radical molecular species including ions. In addition, we have added a parameter, E, that corrects for the energy of pairs of electrons in molecular and atomic species whose valence electrons consist only of one pair of s electrons (not including systems having one or more $1s$ electrons). These *single electron pair* species represent a special case for which the basis set requirements are not as great. The A, A', B, C, and D values are chosen to give the smallest average absolute deviation from experiment for the whole G3/05 test set. The value of E is determined by the minimization of the root mean square deviation of the energies involving the subset of 13 species from the G5/03 test set that involve single pairs of electrons (see Table IV for this subset). For the G4 theory, $A=6.947$ mhartree, $B=2.441$ mhartree, $C=7.116$ mhartree, $D=1.414$ mhartree, $A'=7.128$ mhartree, and

TABLE I. Total G4 energies (in hartrees) of atomic species and spin-orbit corrections (in mhartrees).

Atomic species	Atomic species	E_0(G4)	ΔE(SO)[a]	Atomic species	E_0(G4)	ΔE(SO)[a]
H (^2S)	H (^2S)	−0.501 42	0.0	F^+ (^3P)	−99.066 11	−0.67
He (^1S)	He (^1S)	−2.904 91	0.0	Ne^+ (^2P)	−128.108 67	−1.19
Li (^2S)	Li (^2S)	−7.466 36	0.0	Na^+ (^1S)	−161.928 92	0.0
Be (^1S)	Be (^1S)	−14.657 65	0.0	Mg^+ (^2S)	−199.630 07	0.0
B (^2P)	B (^2P)	−24.646 65	−0.05	Al^+ (^1S)	−242.001 35	0.0
C (^3P)	C (^3P)	−37.834 17	−0.14	Si^+ (^2P)	−288.937 90	−0.93
N (^4S)	N (^4S)	−54.573 67	0.0	P^+ (^3P)	−340.749 63	−1.43
O (^3P)	O (^3P)	−75.045 50	−0.36	S^+ (^4S)	−397.601 63	0.0
F (^2P)	F (^2P)	−99.704 98	−0.61	Cl^+ (^3P)	−459.540 26	−1.68
Ne (^1S)	Ne (^1S)	−128.900 99	0.0	Ar^+ (^2P)	−526.822 78	−2.18
Na (^2S)	Na (^2S)	−162.117 89	0.0	K^+ (^1S)	−599.553 54	0.0
Mg (^1S)	Mg (^1S)	−199.912 04	0.0	Ca^+ (^2S)	−677.139 71	0.0
Al (^2P)	Al (^2P)	−242.221 07	−0.34	Ga^+ (^1S)	−1924.129 16	0.0
Si (^3P)	Si (^3P)	−289.237 04	−0.68	Ge^+ (^2P)	−2076.150 05	−5.37
P (^4S)	P (^4S)	−341.134 63	0.0	As^+ (^3P)	−2234.951 93	−8.04
S (^3P)	S (^3P)	−397.980 18	−0.89	Se^+ (^4S)	−2400.620 74	0.0
Cl (^2P)	Cl (^2P)	−460.015 05	−1.34	Br^+ (^3P)	−2573.151 78	−6.71
Ar (^1S)	Ar (^1S)	−527.400 45	0.0	Kr^+ (^2P)	−2752.713 29	−8.16
K (^2S)	K (^2S)	−599.712 19	0.0	Li^- (^1S)	−7.490 42	0.0
Ca (^1S)	Ca (^1S)	−677.362 61	0.0	B^- (^3P)	−24.655 71	−0.03[b]
Ga (^2P)	Ga (^2P)	−1924.350 57	−2.51	C^- (^4S)	−37.879 08	0.0
Ge (^3P)	Ge (^3P)	−2076.440 70	−4.41	O^- (^2P)	−75.098 47	−0.26[b]
As (^4S)	As (^4S)	−2235.312 07	0.0	F^- (^1S)	−99.833 64	0.0
Se (^3P)	Se (^3P)	−2400.977 97	−4.3	Na^- (^1S)	−162.139 76	0.0
Br (^2P)	Br (^2P)	−2573.585 37	−5.6	Al^- (^3P)	−242.235 93	−0.28[b]
Kr (^1S)	Kr (^1S)	−2753.225 82	0.0	Si^- (^4S)	−289.286 57	0.0
He^+ (^2S)	He^+ (^2S)	−2.001 39	0.0	P^- (^3P)	−341.159 86	−0.45[b]
Li^+ (^1S)	Li^+ (^1S)	−7.267 61	0.0	S^- (^2P)	−398.055 13	−0.88[b]
Be^+ (^2S)	Be^+ (^2S)	−14.313 78	0.0	Cl^- (^1S)	−460.146 71	0.0
B^+ (^1S)	B^+ (^1S)	−24.343 23	0.0	K^- (^1S)	−599.730 00	0.0
C^+ (^2P)	C^+ (^2P)	−37.421 83	−0.2	Ge^- (^4S)	−2076.487 18	0.0
N^+ (^3P)	N^+ (^3P)	−54.040 65	−0.43	Br^- (^1S)	−2573.710 12	0.0
O^+ (^4S)	O^+ (^4S)	−74.547 31	0.0			

[a]Spin-orbit corrections are from Ref. 33, except where noted.
[b]Calculated value, Ref. 34.

E=2.745 mhartree. The effect of adding the two new parameters is discussed in more detail in the next section. Note that in the case of the third-row species (K, Ca, and Ga–Kr) the HLC is based only on the valence electrons, i.e., 4s and 4p, although more orbitals are included in the correlation calculation (see step 4 above). This is slightly different from what was done in the G3 theory for the third row[32] where the HLC for K and Ca included 3s and 3p as valence electrons for the calculation of the HLC.

(7) Finally, the total energy at 0 K is obtained by adding the zero-point energy, obtained from the frequencies of step 2, to the total energy,

$$E_0(\mathrm{G4}) = E_e(\mathrm{G4}) + E(\mathrm{ZPE}). \quad (8)$$

The energy E_0 is referred to as the "G4 energy."

The single point entry calculations in the G4 theory are summarized in Fig. 1. The final total energy is effectively at the CCSD(T,full)/G3LargeXP+HFlimit level if the different additivity approximations work well. The validity of such approximations is discussed in the next section. All calculations in this paper were done with the GAUSSIAN 03 computer program.[37] All of the basis sets in the G4 theory, with the exception of G3LargeXP and the third-row basis sets, are standard in GAUSSIAN 03. The latter basis sets are available on the web[38] and in the supplementary information.[39]

	HF	MP2	MP4	**CCSD(T)**
6-31G(d)	X	X	X	**X**
6-31+G(d)	X	X	X	
6-31+G(2df,p)	X	X	X	
G3LargeXP	**X**	X		
Limit	**X**			

FIG. 1. Single point energies used in the G4 theory (bold entries are new or modified from the G3 theory).

TABLE II. Comparison of average absolute deviations and root mean square deviations for G3 and G4 theories for the G3/05 test.

	Average absolute deviation (kcal/mol)		Root mean square deviation (kcal/mol)	
Type of energy	G3	G4	G3	G4
Enthalpies of formation (270)	1.19	0.80	1.78	1.10
Nonhydrogens (79)	2.10	1.13	2.85	1.49
Hydrocarbons (38)	0.69	0.48	0.93	0.70
Substitutes hydrocarbons (100)	0.82	0.68	1.12	0.94
Inorganic hydrides (19)	0.95	0.92	1.19	1.12
Radicals (34)	0.83	0.66	0.97	0.79
Ionization energies (105)	1.09	0.91	1.61	1.45
Electron affinities (63)	0.97	0.83	1.35	1.03
Proton affinities (10)	1.14	0.84	1.29	1.04
Hydrogen bonded complexes (6)	0.60	1.12	0.77	1.53
All (454)	1.13	0.83	1.67	1.19

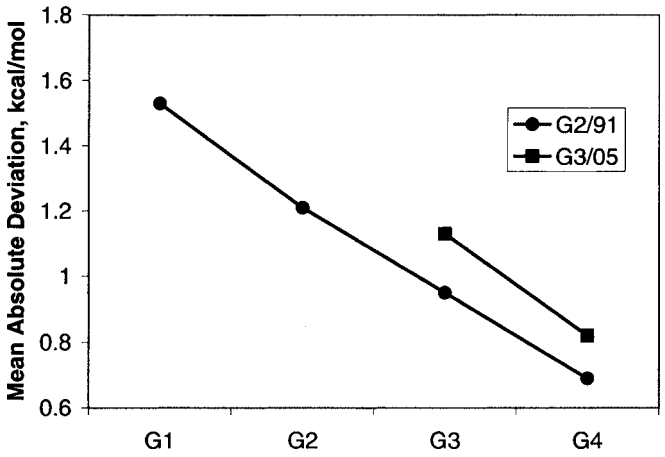

FIG. 2. Comparison of accuracies of G*n* methods for the G2/91 (also referred to as the G2 test set) and G3/05 test sets.

III. ASSESSMENT OF THE G4 THEORY ON THE G3/05 TEST SET

The G3/05 test set[20] contains 454 energies including enthalpies of formation of neutrals, atomization energies, ionization potentials, electron affinities, proton affinities, and hydrogen bond energies. This test set includes energies for species containing first-, second-, and third- and (K, Ca, and Ga–Kr) row elements. The enthalpies of formation at 298 K were calculated as in Ref. 18. The ionization potentials, electron affinities, atomization energies, and proton affinities were calculated at 0 K.[18]

The G4 theory was used to calculate the energies of atoms, molecules, and ions in the G3/05 test set. Table I contains the G4 total energies of the atomic species and the spin-orbit corrections $\Delta E(\mathrm{SO})$, which are included in the total energies. The G4 total energies for the molecules and their geometries are available elsewhere,[38,39] as are the G4 deviations from experiment of all 454 energies. Table II contains a summary of the average absolute deviations and root mean square deviations of the G4 theory from experiment, with results for G3 included for comparison.

A. Comparison of G3 and G4 theories

The results in Table II indicate that for the 454 energies, the average absolute deviation from experiment at the G4 level is 0.83 kcal/mol, which is a significant improvement over 1.13 kcal/mol for the G3 theory. The root mean square deviation of the G4 theory (1.19 kcal/mol) also significantly improves compared to that of the G3 theory (1.67 kcal/mol). The G4 theory especially improves for enthalpies of formation (1.19 kcal/mol for G3 versus 0.80 kcal/mol for G4), with the largest improvement occurring for nonhydrogens, although all of the other enthalpy types (hydrocarbons, substituted hydrocarbons, inorganic hydrides, and radicals) also show improvement due to the changes made in the G4 theory. The average absolute deviation for enthalpies of formation of 79 nonhydrogens decreases from 2.10 kcal/mol (G3) to 1.13 kcal/mol (G4). The improvement for the remaining species is smaller since the G3 theory is already well within the target accuracy of 1 kcal/mol. The next largest improvement occurs for the 38 hydrocarbons for which the average absolute deviation decreases from 0.69 kcal/mol (G3) to 0.48 kcal/mol (G4). The average absolute deviation for the 100 substituted hydrocarbons decreases from 0.82 to 0.68 kcal/mol, while that for the 34 radicals decreases from 0.83 to 0.66 kcal/mol. Finally, the enthalpies for 19 inorganic hydrides improve slightly from 0.95 to 0.92 kcal/mol.

The G4 theory also has significant improvements over the G3 theory for ionization energies electron affinities, and proton affinities. These three quantities have average absolute deviations of 0.91, 0.83, and 0.84 kcal/mol, respectively, at the G4 level compared to 1.09, 0.97, and 1.14 kcal/mol at the G3 level. The only type of energy for which the accuracy decreases is hydrogen bond energies, which have an average absolute deviation of 1.12 kcal/mol at the G4 level compared to 0.60 kcal/mol at the G3 level. This increase is due to poor results for the water and hydrogen fluoride dimers where the B3LYP/6-31G(2*df*,*p*) geometries fail. This will be discussed in more detail later.

The G3X theory,[21] which we previously published, is an extension of the G3 theory that has two of the new features that are included in the G4 theory. The average absolute deviation of G3X from experiment is 1.01 kcal/mol for the G3/05 test set. The G4 theory has an additional three new features that make a substantial improvement over G3X to reduce the average absolute deviation to 0.83 kcal/mol. First, the HF limit step in G4 is more rigorous than the inclusion of a single *g* function on second-row atoms for the G3X theory. While the inclusion of a *g* function is an improvement, it can still miss significant HF energy. For example, for SF_6 the HF basis set in G3X still misses 1.8 kcal/mol of the HF binding energy compared to the extrapolated limit. Second, the expanded *d*-polarization set in the G4 theory corrects a problem in the complete (i.e., nonadditive) version of G3 and G3X theories, as discussed in Sec. III C. Third, the higher level correction parameter set is increased from four to six. A comparison of the accuracy of the Gn methods, $n=1-4$, is shown in Fig. 2.

TABLE III. Contributions of different modifications of G3 theory to the performance of G4 theory on the G3/05 test set. [G3//B2*df* uses B3LYP/6-31G(2*df*,*p*) optimized geometries and scaled zero-point energies instead of MP2/6-31G* geometries and HF/6-31G* scaled zero-point energies; G3(CC)//B2*df* replaces QCISD(T)/6-31G* with CCSD(T)/6-31G*; G3(CC,HF)//B2*df* adds the estimated HF limit to the energy where ΔE(HF) $=E$(HF/limit)$-E$(HF/G3Large); G3(CC,HF,XP)//B2*df* includes extra polarization functions on the G3Large basis set, i.e, G3LargeXP. See text for more details of the modifications. In each case the HLC as defined for the G4 theory was reoptimized using the G3/99 test set.]

	Average absolute deviation (kcal/mol)					
	All (454)	Neutral enthalpies[a]	IPs	EAs	PAs	H-bond complexes
G3	1.13	1.19(2.10,0.69,0.82,0.95,0.83)	1.09	0.97	1.14	0.60
G3//B2*df*	1.06	1.09(1.88,0.55,0.82,0.87,0.75)	1.04	0.99	1.09	1.06
G3(CC)//B2*df*	1.06	1.08(1.85,0.55,0.81,0.83,0.79)	1.06	1.01	1.05	1.06
G3(CC,HF)//B2*df*	0.92	0.87(1.26,0.61,0.68,0.88,0.78)	1.07	0.93	0.83	1.13
G3(CC,HF,XP)//B2*df*	0.90	0.84(1.27,0.47,0.69,0.89,0.68)	1.01	0.95	0.84	1.12
G3(CC,HF,XP,HLC5)	0.87	0.83(1.27,0.48,0.68,0.87,0.65)	0.92	0.93	0.83	1.12
G3(CC,HF,XP,HLC6)=G4	0.83	0.80(1.13,0.48,0.68,0.92,0.66)	0.91	0.83	0.84	1.12

[a]Values in parentheses are for the molecules in the subsets in the following order: nonhydrogens, hydrocarbons, substituted hydrocarbons, inorganic hydrides, and radicals. The neutral enthalpy subset here includes results for atomization energies for third-row species that were added to the test set in Ref. 18.

B. Assessment of the new features of the G4 theory

We calculated the contributions of the five new features in the G4 theory to the average absolute deviation from experiment relative to the G3 theory in a stepwise manner, including reoptimization of the HLC at each step. The results for the G3/05 test set are summarized in Table III and are discussed in this section.

1. Geometries and zero-point energies (G3//2df)

The use of B3LYP/6-31G(2*df*,*p*) geometries and zero-point energies (steps 1 and 2) reduces the overall average absolute deviation from 1.13 to 1.06 kcal/mol with the improvement largely coming from the enthalpies for formation. The use of density functional geometries is most important for the nonhydrogens where the average absolute deviation of the subset decreases from 2.10 to 1.88 kcal/mol. This has been noted in our paper on the G3X theory,[21] with examples such as PF_5, SO_2, SO_3, and SF_6, where the density functional theory gives improved geometries. On the other hand, as noted above, the B3LYP/6-31G(2*df*,*p*) method does poorly on geometries of several hydrogen bonded dimers (hydrogen fluoride and water dimers), which increases the average absolute deviation for this set.

2. CCSD(T) [G3(CC)//2df]

The use of the CCSD(T) method [step (4)(c)] results in no change in the average absolute deviation for the G3/05 test set, as expected based on previous studies.[29,30] The change from QCISD(T) to CCSD(T) is expected to help in specific cases where the former fails. Such cases are not included in the test set, but may be encountered in applications of the methods.

3. HF limit [G3(CC,HF)//2df]

Of the five modifications, the inclusion of the HF limit energy in the formulation of the G4 theory (steps 3 and 5) has the largest effect, as it reduces the overall average absolute deviation from 1.06 to 0.92 kcal/mol. The improvement largely comes from the enthalpies of formation where it decreases the average absolute deviation from 1.08 to 0.87 kcal/mol. In particular, the inclusion of the HF limit is most important for the nonhydrogens where the HF/G3LargeXP energy is farthest from the HF limit. The average absolute deviation of the 79 molecules in the nonhydrogen set decreases from 1.85 to 1.26 kcal/mol. Most of the other subsets show some improvement from this new feature, but not as large.

The use of smaller basis sets than those described in step 3 to obtain the extrapolation to the HF limit was investigated. We tried using the aug-cc-pVTZ and aug-cc-pVQZ basis sets for the extrapolation in Eq. (1), but this pair of basis sets did not converge adequately for larger nonhydrogen species. We also investigated extrapolations based on the G3Large basis set with variation of the polarization sets and found that this approach was also not adequate for obtaining the HF limit energy. Finally, Martin[40] reported that tight polarization functions had a significant effect on the atomization energy of SO_2 at the HF level. We have investigated the inclusion of tight polarization functions in the basis sets used for the SO_2 HF extrapolation and found that the proposed extrapolation gives essentially the same result whether or not tight polarization functions are added.

4. Basis set change [G3(CC,HF,XP)//2df]

The addition of more *d* functions to the G3Large basis set [step (4)(d)], i.e., G3LargeXP, for use in the MP2(full) calculation leads to a slight improvement in the average absolute deviation from 0.92 to 0.90 kcal/mol. The largest improvement is for hydrocarbon enthalpies and ionization potentials (see Table III). This change also corrects some deficiencies in the complete calculation, i.e., without the use of any additivity approximations, as discussed in Sec. III C.

5. Higher level corrections

The addition of the two new HLC parameters reduces the average absolute deviation from 0.90 to 0.83 for the whole test set, with both parameters contributing to this improvement. The addition of the added A' parameter for mol-

TABLE IV. Comparison of deviations for energies involving species having a single valence electron pair in the G3/05 test set. (Not including those with one or more 1*s* electrons.)

	Deviation (kcal/mol)	
	G3(CC,HF,XP,HLC5)[a]	G4
ΔH_f(BeH)	−1.01	1.81
ΔH_f(Li_2)	2.33	−0.30
ΔH_f(Na_2)	3.74	1.11
ΔH_f(LiNa)	3.21	0.58
ΔH_f(BeF_2)	−3.45	−0.82
ΔH_f($MgCl_2$)	−0.72	1.91
D_0(K_2)	−2.27	0.36
IP(Be)	−3.54	−0.86
IP(Mg)	−0.30	−0.62
IP(Ca)	−1.65	1.03
EA(Li)	−3.52	−0.84
EA(Na)	−3.77	−1.09
EA(K)	−2.33	0.35

[a]Results without inclusion of the HLC parameter (*E*) for the single electron pair species.

ecules provides for a different parameter for paired electrons in the closed shell species (*A*) compared to open shell species (*A′*), including radical ions and neutrals. This is important since spin polarization in the latter case can lead to differences in the correction needed compared to closed shell systems. Note that a similar parameter could be added for atoms, but since there are few closed shell atomic systems in the test set (He, Ne, Ar, Kr, F^-, Cl^-, B^+, and Al^+), this is not as important. The closed shell atoms Be, Mg, and Ca are covered by the new *E* parameter. The most significant improvement from the addition of the *A′* parameter occurs for ionization potentials (IPs) whose average absolute deviation decreases from 1.01 to 0.92 kcal/mol. The neutral radical enthalpies also improve from 0.68 to 0.65, and the electron affinities improve slightly from 0.95 to 0.93 kcal/mol. The overall average absolute deviation for the G3/05 test set decreases from 0.90 to 0.87 kcal/mol with the addition of this parameter. The addition of the HLC parameter for single pairs (*E*) further reduces the average absolute deviation to 0.83 kcal/mol, with the largest effect on molecules and atoms having a single valence electron pair. A summary of the energies involving these species is given in Table IV, showing the overall improvement when this HLC parameter is added. The *E* parameter is smaller by about 4 mhartree than the value for *A*, indicating that the reason for doing this was valid, i.e., that the basis set requirement and hence the correction required for these systems are much smaller.

Finally, we note that when the HLC is not included in G4 theory, the average absolute deviation increases to 8.6 kcal/mol, which is slightly lower than the value of 9.1 kcal/mol for the G3 theory. A detailed table of the deviations without the HLC included in the G4 theory is given in the supplementary tables.[39]

6. Timings

The cost of a G4 calculation is increased compared to a G3 theory calculation due largely to the two HF calculations used to extrapolate to the HF limit in step 3. The ratios of computer times for a G4 calculation compared to a G3 calculation for four representative molecules, benzene, hexane, octane, and silicon tetrachloride, are given in Table V. Also given are the relative times for a G2 calculation for benzene and silicon tetrachloride. For these molecules the G4 calculation takes two to three times more CPU time than G3, so that the savings that was gained in G3 compared to G2 is eliminated. However, since this increase is a result of HF calculations, which scale only as $\sim n^{3-4}$, the increase is not a serious problem for these calculations. The CPU increase is less for nonhydrogens (e.g., $SiCl_4$) compared to hydrocarbons due to the basis sets on the hydrogens in the HF calculations.

TABLE V. Relative CPU times used in G2, G3, and G4 single point energy calculations.

Method	$SiCl_4$	Benzene	Hexane	Heptane
G2	2.4	1.9	⋯	⋯
G3	1.0	1.0	1.0	1.0
G4	1.9	3.0	3.0	2.5

C. Assessment of additivity approximation in G4 theory

The additivity approximations in the Gaussian-*n* approach to the computation of molecular energies are essential to reducing computer resource requirements. These approximations can be assessed by calculating energies at the highest correlation level [QCISD(T) or CCSD(T)] with the largest basis set of the specific method and reoptimization of the HLC. Ideally, the geometries and zero-point energies from this level of theory would also be used, but due to the cost we have not done this. The additivity approximations in the G2 theory were previously investigated by calculating QCISD(T)/6-311+G(3*df*,2*p*)//MP2(full)/6-31G* energies, scaled HF/6-31G* zero-point energies, and a reoptimized HLC.[41] This type of calculation has been referred to as G2(complete). A comparison of the G2(complete) results for the G2 test set of 125 energies found the approximations to work well, with only two energies differing by more than 1 kcal/mol (the electron affinities of NH_2 and OH). The average absolute deviation of the full basis set calculation was very close to that using the additive approximation (1.17 kcal/mol vs 1.21 kcal/mol).

In the current work we have investigated the additivity approximations of the G4 theory along with those of the G3 theory, which were not previously investigated. Results using analogous G3(complete) and G4(complete) methods have been obtained for the G2 test set of 125 energies. The G4(complete) method uses CCSD(T,full)/G3LargeXP//B3LYP/6-31G(2*df*,*p*) energies along with scaled B3LYP/6-31G(2*df*,*p*) zero-point energies, Hartree-Fock extrapolation, atomic spin-orbit corrections, and reoptimized HLC parameters. The G3(complete) method uses QCISD(T,full)/G3Large//B3LYP/6-31G(2*df*,*p*) energies along with scaled HF/6-31G* zero-point energies, atomic spin-orbit corrections, and reoptimized HLC parameters. A

TABLE VI. Comparison of G3 and G4 methods with and without additivity approximation on the G2 test set of 125 energies.[a,b]

			Average absolute deviation (kcal/mol)			
Breakdown	G2	G2(complete)	G3	G3(complete)	G4	G4(complete)
All (125)	1.21	1.17	0.95	1.07	0.72	0.79
ΔH_f			0.87	0.99	0.65	0.78
IP			0.95	0.90	0.70	0.84
EA			1.01	1.45	0.82	0.70
PA			1.32	1.34	1.01	1.00

[a]G3(complete) is a QCISD(T,full)/G3Large//MP2(FU)/6-31G* energy calculation with scaled HF/6-31G* zero-point energies. G4(complete) is a CCSD(T,full)/G3LargeXP//B3LYP/6-31G(2*df*,*p*) plus HF/limit extrapolation calculation with scaled B3LYP/6-31G(2*df*,*p*) zero-point energies.
[b]All methods have HLC parameters individually optimized for the G2/91 test set. HLC parameters for the G4 theory: $A=8.669$ mhartree, $B=3.126$ mhartree, $C=8.723$ mhartree, $D=2.337$ mhartree, $A'=8.949$ mhartree, and $E=3.216$ mhartree. HLC parameters for G4(complete): $A=9.815$ mhartree, $B=3.727$ mhartree, $C=9.534$ mhartree, $D=2.872$ mhartree, $A'=9.871$ mhartree, and $E=4.904$ mhartree.

summary of the results is given in Table VI. Previous results for the G2 theory are also included. The results in Table VI are surprising because G3 and G4 theories both perform better than the corresponding versions without the additivity approximations. The average absolute deviation of G3 is 0.12 kcal/mol smaller than G3(complete) and that of G4 is 0.07 kcal/mol smaller than G4(complete). In each case about 11 species differ by more than 1 kcal/mol. These species fall into two categories. The first is enthalpies involving silicon and phosphorus (specifically, SiH_2, SiH_3, SiH_4, PH_2, PH_3, SiO, and Si_2H_6) and the second is the electron affinities of O, F, OH, and NH_2.

The problem with the enthalpies of formation of the Si and P species is largely due to the correlation treatment of core electrons. Martin *et al.*[42] have pointed out that the MP2(full)/G3Large level of theory gives a reasonable account of core correlation due to a cancellation of errors from using MP2 instead of CCSD(T) (that typically leads to an overestimation) and a relatively small basis set (that leads to an underestimation). From our calculations we find that the evaluation of core correlation at the MP2 level versus the QCISD(T) [or CCSD(T)] level is responsible for most of the difference between G*n* and G*n*(complete), $n=3,4$, for the enthalpies of the seven Si and P species listed above. The use of MP2(full)/G3Large for core correlation in the additive methods gives better agreement with experiment due to the cancellation of errors.

There are two sources for the nonadditivity of the G3 and G4 methods in the calculation of electron affinities. The first source, which has been corrected in the G4 theory, is a deficiency in the *d*-polarization functions and the second, which is not corrected in the G4 theory, is due to nonadditivity of the effects of *d*-polarization and diffuse functions at the higher correlation levels. The *d*-polarization function deficiency is caused by the use of two types of basis sets, i.e., 6-31G(*d*) in QCISD(T) and MP4 calculations and 6-311G(*d*) basis in the large basis set (G3Large) for MP2 calculations. The 6-311G(*d*) basis has a much tighter *d* function (e.g., 1.75 for F) than 6-31G(*d*) (e.g., 0.8 for F). This causes a problem in G3(complete) because the QCISD(T)/G3Large calculation uses a 2*df* polarization set[3] having a small *d* exponent (e.g., 1.75/2=0.875 for F) that is not diffuse enough for electron affinities (EAs) of species such as F and OH. This is not a problem in the G3 theory because it includes a MP4/6-31G(2*df*,*p*) calculation that has a more diffuse *d*-function exponent (e.g., 0.8/2=0.4 for F). The first nonadditivity problem is corrected in the G4 theory by the G3LargeXP basis set, which has an expanded polarization set of 3*df* on the first row and of 4*d*2*f* on the second row. Use of the expanded polarization set results in a dramatic improvement for electron affinities with the average absolute deviation of G4(complete) being 0.70 kcal/mol compared to 1.45 kcal/mol for G3(complete) for the G2 test set. The second source of error, i.e., nonadditivity due to the separation of *d*-polarization and diffuse functions in the higher correlation calculations, is still present in the G4 theory. As a result, the G4(complete) electron affinities of F, O, OH, and NH_2 differ by more than 1 kcal/mol with the G4 theory, but G4(complete) is now in better agreement with experiment than the G4 theory in these cases.

The average absolute deviation of the G4 theory (0.72 kcal/mol) is still smaller than that of the G4(complete) theory (0.79 kcal/mol) for the G2 test set. The poorer agreement with experiment for the complete method is largely due to the problem with the core correlation calculation for the seven Si and P species. If these seven species are not included in the assessment, the average absolute deviations of the two methods are about the same.

D. Analysis of problem energies

While the G4 theory is significantly improved relative to the G3 theory on the G3/05 test set as discussed above, there are still some problem energies. Of the 454 energies only 35 have errors greater than 2 kcal/mol. These are listed in Table VII. In this section we discuss reasons for the larger errors in cases where there exists an apparent explanation.

1. Enthalpies of formation of neutrals

Eighteen of the 270 enthalpies of formation (or atomization energies in the case of third-row species) in the G3/05 test deviate from experiment by more than 2 kcal/mol. Three

TABLE VII. Outliers in the G3/05 test set that are greater than 2 kcal/mol. (Units are kcal/mol. Experimental values are from Refs. 16–18).

Property	Species	Expt.	Expt.-G4 (kcal/mol)	Property	Species	Expt.	Expt.-G4 (kcal/mol)
ΔH_f(298 K)	BF_3	−271.4	−2.8	IP	CH_4	291.0	−2.4
	COS	−33.1	2.5		BF_3	358.8	−2.4
	CS_2	28.0	3.0		BCl_3	267.5	−2.6
	C_2F_4	−154.4	3.2		B_2F_4	278.3	8.9
	C_2Cl_4	−3.0	3.1		see-C_3H_7	170.0	−2.9
	CH_2=CHCl (vinyl chloride)	8.9	3.6		CN	313.6	−3.7
	$C_4H_4N_2$ (pyrimidine)	46.8	2.5		CH_3F	287.6	−2.6
	CH_3–C(=O)–CCH	15.6	−2.6		B_2H_4	223.7	2.3
	$Si(CH_3)_4$ (tetramethylsilane)	−55.7	−3.4		Si_2H_5	175.3	−2.5
	PF_5	−38.1	−2.3		NaBr	191.6	−4.7
	$POCl_3$	−133.8	−2.4	EA	F	78.4	−2.3
	Cl_2O_2S	−84.8	−2.2		CH_3	1.8	2.2
	PCl_3	−69.0	−2.8		C_2	75.4	2.2
	AlF	−63.5	2.3		CH_2NC	24.4	−2.3
	Al_2Cl_6	−309.7	5.2		HS	54.4	2.3
	$ClFO_3$	−5.1	−4.2	HB	H_2O dimer, ΔH_f	−3.6	−2.3
D_0	KF	117.6	2.3		HF dimer, D_0	−2.97	−2.6
	GeH_4	270.5	−2.5				

species (C_2F_4, C_2Cl_4, and CH_2CHCl) may have problems with the experimental values. This has been discussed elsewhere.[19,43] Of the remaining 15 energies, 12 are nonhydrogens. Only three of these have errors greater than 3 kcal/mol (CS_2, Al_2Cl_6, and $ClFO_3$). The difficulties with these nonhydrogens are probably due to the slow convergence of correlation energy with basis set and errors in core correlation effects that cannot be accounted for adequately by the current higher level correction.[44] In addition, relativistic effects are not included in the G4 theory, and for some nonhydrogen systems these are probably not accounted for by the higher level correction. The remaining three energies in this group are for substituted hydrocarbons [$Si(CH_3)_4$, $C_4H_4N_2$, CH_3COC_2H].

2. Ionization potentials

Ten of the 105 ionization potentials from the G4 theory deviate by more than 2 kcal/mol from experiment. The IPs of CH_4, BF_3, and BCl_3 are too large because the B3LYP method fails for the Jahn-Teller distortions in the ions, as discussed previously.[45] These errors are corrected when MP2(full)/6-31G* geometries are used, as shown in Table VIII. The large error for the IP of B_2F_4 is probably due to errors in the experimental IP, as discussed previously.[19] The error for CN is probably related to the fact that CN^+ is isoelectronic with C_2, which is known to be a challenging molecule. The large error (−4.9 kcal/mol) in the IP of NaBr is similar to the error found for the G3 theory.[32] In that case we investigated the need to include the Na 2*s* and 2*p* electrons in the correlation treatment and found that this did not improve the results. In addition, the calculation of this IP with G4 (complete) gives about the same result. We note that the error is much less at the G2 level of theory (−2.4 kcal/mol)[33] so the large error may be due to a basis set effect. The reasons for the >2 kcal/mol errors in the remaining four IPs in Table VI (C_3H_7, Si_2H_5, CH_3F, and B_2H_4) are unclear.

3. Electron affinities

Five of the 63 electron affinities from the G4 theory deviate from experiment by more than 2 kcal/mol. The F electron affinity is too large due to the additivity problem of the basis sets in the G4 theory, as discussed in Sec. III C. This is corrected when done without additivity approximations. The error for C_2 is likely due to the highly correlated nature of the neutral C_2 molecule that is known to be a challenge to describe accurately. The CH_3 radical weakly binds an electron (EA=1.8 kcal/mol), so it is not surprising that it is underbound by 2.2 kcal/mol due to the lack of enough

TABLE VIII. Effect of use of improved geometry on H-bonded dimers and ionization potentials involving cations with Jahn-Teller distortions.

	Expt. Theor. (kcal/mol)	
	G4	G4 with improved geometry[a]
H_2O dimer, D_0	−2.28	−0.48
CH_3OH dimer, D_0	0.02	0.46
$CH_3(CO)CH_3$ dimer, D_0	−1.36	−1.33
HCl dimer, D_0	−0.25	−0.24
CH_3COOH dimer, D_0	0.19	0.15
HF dimer, D_0	−2.61	−0.15
CH_4, IP	−2.40	−0.93
BF_3, IP	−2.38	0.28
BCl_3, IP	−2.57	−0.92
AAD (kcal/mol)	1.56	0.55

[a]For hydrogen bonded complexes the improved optimized geometries are from B3LYP/6-31+G(2*df*,*p*) calculations, and for the Jahn-Teller species they are from MP2(full)/6-31G* calculations.

TABLE IX. Basis sets used in single point HF energy calculations for the G4 theory.

	aug-cc-pVQZ		aug-cc-pV5Z	
Atoms	Literature[a]	G4 modified[b]	Literature[a]	G4 modified[b]
H,He	4*s*3*p*2*d*1*f*+diffuse *spdf*	4*s*2*pd*	5*s*4*p*3*d*2*f*1*g*+diffuse *spdfg*	5*s*3*p*2*d*
Li–Ne	5*s*4*p*3*d*2*f*1*g*+diffuse *spdfg*	5*s*4*p*3*d*2*f*1*g*+diffuse *sp*	6*s*5*p*4*d*3*f*2*g*1*h*+diffuse *spdfgh*	6*s*5*p*4*d*3*f*2*g*1*h*+diffuse *sp*
Na, Mg	6*s*5*p*3*d*2*f*1*g*	6*s*5*p*3*d*2*f*1*g*	7*s*6*p*4*d*3*f*2*g*1*h*	7*s*6*p*4*d*3*f*2*g*1*h*
Al–Ar	6*s*5*p*3*d*2*f*1*g*+diffuse *spdfg*	6*s*5*p*3*d*2*f*1*g*+diffuse *sp*	7*s*6*p*4*d*3*f*2*g*1*h*+diffuse *spdfgh*	7*s*6*p*4*d*3*f*2*g*1*h*+diffuse *sp*
K,Ca	7*s*6*p*4*d*2*f*1*g*	7*s*6*p*4*d*2*f*1*g*	8*s*7*p*5*d*3*f*2*g*1*h*	8*s*7*p*5*d*3*f*2*g*1*h*
Ga–Kr	7*s*6*p*4*d*2*f*1*g*+diffuse *spdfg*	7*s*6*p*4*d*2*f*1*g*+diffuse *sp*	8*s*7*p*5*d*3*f*2*g*1*h*+diffuse *spdfgh*	8*s*7*p*5*d*3*f*2*g*1*h*+diffuse *sp*

[a]See Appendix A for references.
[b]Modified basis set used in the G4 theory (see Appendix A).

diffuse functions. This was also true for the G3 theory. The reason for the errors of greater than 2 kcal/mol for the remaining two systems (HS and CH_2NC) is unclear, but they are only slightly more than 2 kcal/mol.

4. Hydrogen bonded dimers

As discussed earlier, two hydrogen bonded dimers, the water dimer and the hydrogen fluoride dimer, have errors of >2 kcal/mol due to the need to include diffuse functions in the basis set used for geometry optimization,[20] which is not included in the G4 theory. Results for the G4 energies for all hydrogen bonded dimers when a diffuse function is included in the basis set (6-31+G(2*df*,*p*)) are given in Table VIII. The errors in the water and hydrogen fluoride dimers are reduced to under 1 kcal/mol and the average absolute deviation is reduced from 1.12 kcal/mol to 0.47 kcal/mol.

We note that an analysis of the overall absolute deviation for the G3/05 test set using improved geometries in the case of the hydrogen bonded complexes and Jahn-Teller systems gives a value of 0.81 kcal/mol. In addition, if the four energies with probable problems with the experimental data (enthalpies of C_2F_4, C_2Cl_4, and CH_2CHCl and ionization potential of B_2F_4) are excluded, the average absolute deviation is reduced to 0.78 kcal/mol.

IV. CONCLUSIONS

In this paper we have presented Gaussian-4 theory (G4 theory) for the calculation of energies of molecular species containing first-row (Li–F), second-row (Na–Cl), and third-row main group elements. The G4 theory modifies the Gaussian-3 (G3) theory in five ways, including an estimate of the HF energy limit, an expanded polarization set for the large basis set calculation, use of CCSD(T) energies, use of DFT geometries and zero-point energies, and two added higher level correction parameters. The overall average absolute deviation for the 454 energies in this test set is 0.83 kcal/mol, a significant improvement over the G3 theory. The largest improvement occurs for enthalpies of formation of nonhydrogen species, which are reduced to 1.13 kcal/mol. With the exception of hydrogen-bonded complexes the other types of energies in the G3/05 test set are also improved in the G4 theory. The inclusion of the HF limit energy in the G4 method contributes to the largest improvement over the G3 theory. The addition of the two new higher level correction parameters, one for paired electrons in open shell species and the other for species with only one pair of valence electrons, also contributes to a significant improvement. Finally, the expanded *d*-polarization set corrects a problem with nonadditivity that is present in the G3 theory for some anions.

ACKNOWLEDGMENTS

The authors acknowledge grants of computer time at the National Energy Research Supercomputer Center and the Laboratory Computing Resources Center at Argonne National Laboratory. They thank Dr. K. Peterson for providing them with the potassium basis set. 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 A: BASIS SETS FOR HF LIMIT EXTRAPOLATION

These basis sets are based on aug-cc-pVQZ or aug-cc-pV5Z basis sets[26–28] available from the EMSL database (http://www.emsl.pnl.gov/forms/basisform.html) with the exception of potassium.[46] In most cases the basis sets were modified to save CPU time. The basis set compositions are summarized in Table IX. The G4 quadruple zeta basis sets for H and He use the *s* part of cc-pVQZ combined with the 2*pd* polarization functions from the smaller cc-pVTZ basis set. They have no diffuse functions. The G4 quadruple zeta basis sets for the other atoms 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. The 5Z basis set for H and He uses the *s* part of cc-pV5Z combined with 3*p*2*d* polarization functions taken from cc-pVQZ. They have no diffuse functions. The other atoms use the aug-cc-pV5Z basis set, but include only *s* and *p* diffuse functions (no *d*, *f*, *g*, or *h* diffuse functions). The only exceptions to this are Na, Mg, K, and Ca, which have no diffuse functions added. The basis sets are included in the supplementary information[39] and are available on the web.[38]

APPENDIX B: EXPANDED *d*-POLARIZATION BASIS

In the expanded *d*-polarization set in the G3LargeXP basis set, the 2*df* polarization set in G3Large on the first row

TABLE X. Polarization exponents used in the G3LargeXP basis set.

Atom	4*d*	*d*	*d*/4	4th *d*	MP2(full)/G3LargeXP
LI	0.8	0.2	0.05		−7.464 22
Be	1.02	0.255	0.0638		−14.635 86
B	1.604	0.401	0.1003		−24.614 20
C	2.504	0.626	0.1565		−37.798 67
N	3.652	0.913	0.2283		−54.538 46
O	5.168	1.292	0.323		−74.998 55
F	7.0	1.75	0.4375		−99.650 43
Ne	9.126	2.304	0.576		−128.843 11
Na	0.7	0.175	0.0438		−162.102 97
Mg	0.7	0.175	0.0438		−199.890 69
Al	2.16	0.54	0.135	0.39	−242.189 16
Si	3.04	0.76	0.19	0.47	−289.198 96
P	3.68	0.92	0.23	0.61	−341.093 58
S	4.08	1.02	0.26	0.59	−397.926 05
Cl	4.8	1.2	0.3	0.72	−459.951 86
Ar	5.6	1.4	0.35	0.85	−527.332 84

is replaced by a 3*df* with a 4:1:1/4 geometrical progression, and the 3*d*2*f* polarization functions on the second row (Al–Cl) are replaced by 4*d*2*f*. We investigated the need to reoptimize the first-row exponents and found that it was not necessary, but that optimization of the 4*d* exponents was required. This was done by optimizing the 3*d* set and a fourth *d* exponent sequentially until converged. The resulting *d* exponents for both first and second rows are listed in Table X. The species used in the Al through Kr optimizations were Al, AlF_3, Si, SiH_4, SiF_4, P, PH_3, PF_3, S, SO_2, SH_2, Cl, HCl, CCl_4, and Ar. Values for a given atom type were averaged. Note that the fourth exponent falls approximately midway between the two smaller exponents of the 3*d* set. A similar type of exponent spread has been found recently by Dunning and co-workers[47,48] to be necessary for an accurate description of the second row. The 3*d*2*f* polarization set is still used for Na and Mg in the second row. The *d* functions in the 3*d*2*f* polarization set were not changed for the third row (K, Ca, Ga–Kr) from their values in G3Large because of the filled *d* shell. The *f* exponents, diffuse exponents, and tight polarization functions in the G3LargeXP basis set remain the same as in G3Large.

[1] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, J. Chem. Phys. **90**, 5622 (1989); L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople, *ibid.* **93**, 2537 (1990).
[2] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94**, 7221 (1991).
[3] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, J. Chem. Phys. **109**, 7764 (1998).
[4] T. H. Dunning, J. Phys. Chem. A **104**, 9062 (2000).
[5] D. Feller and D. A. Dixon, J. Chem. Phys. **115**, 3484 (2001).
[6] D. Feller and D. A. Dixon, J. Phys. Chem. A **104**, 3048 (2000).
[7] J. M. L. Martin and G. De Oliveira, J. Chem. Phys. **111**, 1843 (1999).
[8] A. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay, and J. Gauss, J. Chem. Phys. **120**, 4129 (2004).
[9] D. Feller and K. A. Peterson, J. Chem. Phys. **108**, 154 (1998).
[10] D. Feller and K. A. Peterson, J. Chem. Phys. **110**, 8384 (1999).
[11] A. 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).
[12] N. J. DeYonker, T. R. Cundari, and A. K. Wilson, J. Chem. Phys. **124**, 114104 (2006).
[13] N. J. DeYonker, T. Grimes, S. Yockel, A. Dinescu, B. Mintz, T. R. Cundari, and A. K. Wilson, J. Chem. Phys. **125**, 104111 (2006).
[14] J. A. Montgomery, Jr.,M. J. Frisch, J. W. Ochterski, and G. A. Petersson, J. Chem. Phys. **110**, 2822 (1999).
[15] J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, and G. A. Petersson, J. Chem. Phys. **112**, 6532 (2000).
[16] P. L. Fast, M. L. Sanchez, and D. G. Truhlar, Chem. Phys. Lett. **306**, 407 (1999).
[17] B. J. Lynch and D. G. Truhlar, J. Phys. Chem. A **107**, 3898 (2003).
[18] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **106**, 1063 (1997); L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, *ibid.* **109**, 42 (1998).
[19] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **112**, 7374 (2000).
[20] L. A. Curtiss, P. C. Redfern, and K. Raghavachari, J. Chem. Phys. **123**, 124107 (2005).
[21] L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, J. Chem. Phys. **114**, 108 (2001).
[22] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[23] L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, Chem. Phys. Lett. **270**, 419 (1999).
[24] A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper, and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).
[25] D. Feller, J. Chem. Phys. **98**, 7059 (1993).
[26] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[27] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **98**, 1358 (1993).
[28] A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning, Jr., J. Chem. Phys. **110**, 7667 (1999).
[29] L. A. Curtiss, K. Raghavachari, P. C. Redfern, A. G. Baboul, and J. A. Pople, Chem. Phys. Lett. **314**, 101 (1999).
[30] L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, Chem. Phys. Lett. **359**, 390 (2002).
[31] V. A. Rassolov, J. A. Pople, P. C. Redfern, and L. A. Curtiss, Chem. Phys. Lett. **350**, 573 (2001).
[32] L. A. Curtiss, P. C. Redfern, V. Rassolov, G. Kedziora, and J. A. Pople, J. Chem. Phys. **114**, 9287 (2001).
[33] L. A. Curtiss, M. P. McGrath, J.-P. Blaudeau, N. E. Davis, and R. Binning, J. Chem. Phys. **103**, 6104 (1995).
[34] J.-P. Blaudeau, M. P. McGrath, L. A. Curtiss, and L. Radom, J. Chem. Phys. **107**, 5016 (1997).
[35] C. Moore, Natl. Bur. Stand. (U.S.) Circ. No. 467 (U.S. GPO, Washington, D.C., 1952).
[36] J.-P. Blaudeau (private communication).
[37] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 03, Gaussian, Inc., Wallingford, CT, 2004.
[38] chemistry.anl.gov/compmat/comptherm.htm
[39] See EPAPS document no. E-JCPSA6-126-310707 for supplementary information including G4 total energies, both E_e and E_o values, for all species in the G3/05 test set, the G4 deviations from experiment, the

G3LargeXP basis set, and the modified aug-cc-pVQZ or aug-cc-pV5Z basis sets. This document can be reached via a direct link in the online article's HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).
[40] J. M. L. Martin, J. Chem. Phys. **108**, 2791 (1998).
[41] L. A. Curtiss, J. E. Carpenter, K. Raghavachari, and J. A. Pople, J. Chem. Phys. **96**, 9030 (1992).
[42] J. M. L. Martin, A. Sunderman, P. L. Fast, and D. G. Truhlar, J. Chem. Phys. **113**, 1348 (2000).
[43] L. Pollack, T. L. Windus, W. A. Windus, and D. A. Dixon, J. Phys. Chem. A **109**, 6934 (2005).
[44] K. E. Gutowski and D. A. Dixon, J. Phys. Chem. A **110**, 12044 (2006).
[45] A. G. Baboul, L. A. Curtiss, P. C. Redfern, and K. Raghavachari, J. Chem. Phys. **110**, 7650 (1999).
[46] K. Peterson (private communication).
[47] T. H. Dunning, Jr., K. A. Peterson, and A. K. Wilson, J. Chem. Phys. **114**, 9244 (2001).
[48] A. K. Wilson and T. H. Dunning, Jr., J. Phys. Chem. A **108**, 3129 (2004).