

SANDIA REPORT

SAND99-8239

Unlimited Release

Printed April 1999

RECEIVED

FEB 11 2000

OSTI

Bond Additivity Corrections for Quantum Chemistry Methods

C.F. Melius, M. D. Allendorf

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of
Energy under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.



Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Prices available from (615) 576-8401, FTS 626-8401

Available to the public from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A03
Microfiche copy: A01



DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Bond Additivity Corrections for Quantum Chemistry Methods*

Carl F. Melius and Mark D. Allendorf
Sandia National Laboratories
Livermore, CA 94551-0969

Abstract

New Bond Additivity Correction (BAC) methods have been developed for the G2 method, BAC-G2, as well as for a hybrid DFT/MP2 method, BAC-Hybrid. These BAC methods use a new form of BAC corrections, involving atomic, molecular, and bond-wise additive terms. These terms enable one to treat positive and negative ions as well as neutrals. The BAC-G2 method reduces errors in the G2 method due to nearest-neighbor bonds. The parameters within the BAC-G2 method only depend on atom types. Thus the BAC-G2 method can be used to determine the parameters needed by BAC methods involving lower levels of theory, such as BAC-Hybrid and BAC-MP4. The BAC-Hybrid method should scale well for large molecules. The BAC-Hybrid method uses the differences between the DFT and MP2 as an indicator of the method's accuracy, while the BAC-G2 method uses its internal methods (G1 and G2MP2) to provide an indicator of its accuracy. Indications of the average error as well as worst cases are provided for each of the BAC methods.

*Work supported by an LDRD from within the Department of Energy.

1. Introduction

The thermochemical properties for many molecular species are not known. This is particularly true for transient intermediates occurring in chemical reactions. These species can be highly reactive and short lived. Also, the environment in which they are produced can involve high temperatures and pressures that make experimental diagnostic techniques difficult to apply. It has therefore been necessary to develop theoretical quantum chemical techniques that can provide accurate thermochemical data¹. It is important that these computational procedures have predictive capability. That is, it is not sufficient for the calculated results to be accurate on the average, but one must also know when the results may be wrong (i.e., what is the uncertainty in a particular calculated value).

In the 1980's, we developed a bond-additivity correction procedure for quantum chemical calculations called BAC-MP4.^{2,3} The BAC-MP4 method has proven reliable in calculating the thermochemical properties of molecular species, including radicals as well as stable closed-shell species. However, this procedure is currently limited to neutral species and cannot adequately treat positive and negative ions with the same level of accuracy. In addition, this procedure is limited to small molecules. When developed over fourteen years ago, three- and four-heavy-atom systems could be handled routinely. Today, even with faster computers, we are still limited to routinely doing six- and seven- and occasionally ten-heavy-atom systems, due to the seventh power in the scaling of the computational time. While a sub-procedure of BAC-MP4, BAC-MP2, has been used for larger systems⁴, it does not provide the needed accuracy. Several other

variations on the BAC-MP4 method have been developed. A somewhat larger basis set has been used⁵, i.e., BAC-MP4/6-311++G**, which reduced the magnitude of the uncertainty in the calculation. In addition, the MP2-optimized geometry⁶ was used instead of the Hartree-Fock geometry, i.e., BAC-MP4/MP2 as opposed to BAC-MP4/HF. These methods, however, all depend on the same functional form for the BAC corrections and thus contain many of the same limitations inherent in the original BAC-MP4 method..

We therefore have developed new BAC procedures, based on the improved computational chemistry-algorithms and computer capabilities that exist today. In particular, limitations in the quality of geometry optimization based on HF-optimized geometry are no longer necessary due to the progress made in density functional methods. These newer methods provide geometries and frequencies that are comparable to those using the MP2 method.

However, no one procedure can yield the desired accuracy and yet scale to larger molecules. We have therefore developed a new BAC procedure that can be applied to a variety of quantum chemistry methods. Some of these methods, such as BAC-G2, can be applied to small molecules to establish accurate reference heats of formation. These methods can incorporate high levels of electron correlation, such as quadratic CI (QCI)⁷ or coupled cluster (CCST(T))⁸ theory, which can accurately determine the energies of chemical bonds. Other methods, such as BAC-Hybrid, can be used for large molecules. The extension to larger molecules involves more than just the scaling of the electronic energy expression. It also must address the number of conformers that may exist as well as the time required to search for stationary points (i.e., the reaction surface for a chemical species becomes combinatorially more complex as the number of atoms increases).

In the following sections, we describe the BAC procedure in general. We then define the new forms for the bond additivity corrections, including the atomic, molecular, and bond-wise additive terms. In particular, we introduce the BAC-G2 and BAC-Hybrid methods. We then present results on the accuracy of these new methods and discuss the implications of these results.

2. The BAC Method for Thermochemical Properties

In this section, we describe the BAC method. To maximize its usefulness, the BAC method must be able to treat unstable or radical species as well as closed-shell stable molecules. It should also be able to treat activation barriers along reaction pathways, although the error uncertainties may be somewhat larger. To determine the thermochemical properties for a given molecular species (e.g., its heat of formation, entropy, heat capacity, and free energy), we need to determine its structure, vibrational frequencies, and electronic energy. The overall BAC procedure is shown schematically in Fig. 1.

In the following subsections we present details of the BAC procedure, with particular emphasis on the bond additivity corrections. We present two new BAC procedures: BAC-G2 and BAC-Hybrid.

2.1. Electronic Structure Calculations for Structure, Vibrational Frequencies, and Electronic Energy

The BAC method first involves *ab initio* electronic structure calculations to determine the structure, vibrational frequencies, and electronic energy. The equilibrium geometry of the molecule is determined using a lower-level electronic structure method with a small basis set. For the BAC-MP4 and BAC-G2 methods, we use the Hartree-Fock (HF) method (restricted Hartree-Fock⁹, RHF, for closed shell molecules and unrestricted Hartree-Fock¹⁰, UHF, for open shell molecules). The basis set used to describe the electronic wave function is the split-valence

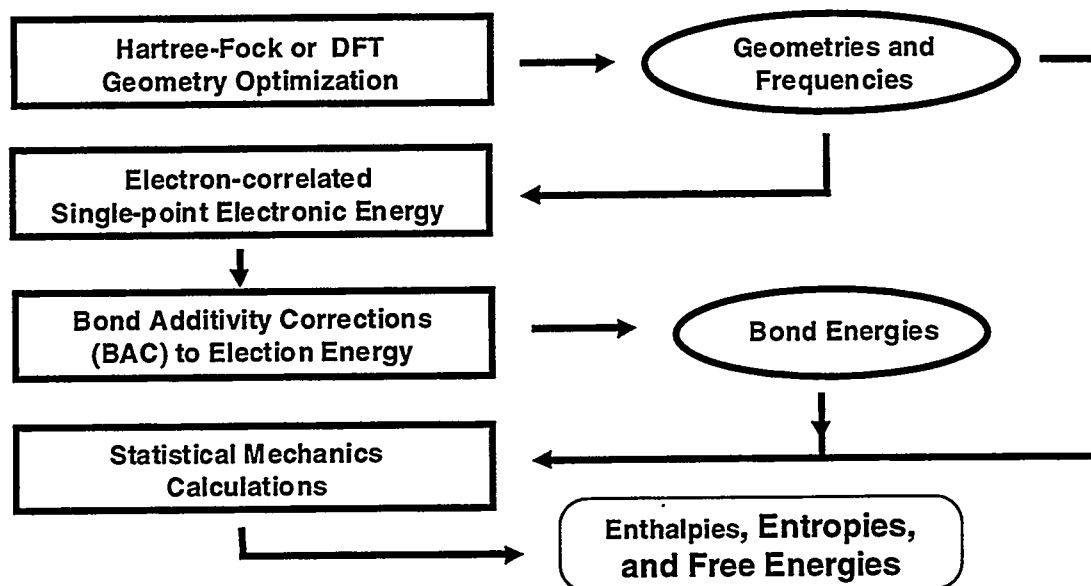


Figure 1. Hierarchical diagram of the BAC method for determining thermochemical properties of molecules. First, a Hartree-Fock or DFT calculation provides the molecular geometry and frequencies. Next, a single-point calculation is performed at a higher-level of theory with a larger basis set to determine the *ab initio* electronic energy, to which bond-additivity corrections (BAC) are added to obtain a total electronic energy. Finally, statistical mechanics calculations are used to determine the enthalpy, entropy, and free energy of the molecule.

with polarization functions on the heavy atoms (denoted 6-31G(D) or 6-31g*¹¹). For the BAC-Hybrid method, one uses the analogous B3LYP (hybrid density functional method mixed with the Hartree-Fock) method¹²⁻¹⁴. The analogous restricted RB3LYP or open-shell (UB3LYP) methods apply¹⁴. In the rest of this paper, we will denote the B3LYP as DFT due to its heavy reliance on the density functional theory contribution to the energy. The basis set used for the BAC-Hybrid method also uses the 6-31G(D) basis.

Having determined the optimized molecular geometry, harmonic vibrational frequencies are calculated¹⁵ at the same level of theory (HF or DFT), using the same basis set (6-31G(D)). Actually, the first and second derivatives of the potential energy surface at the stationary point are calculated. Thus, the harmonic vibrational frequencies along with their corresponding normal modes can be determined for any isotopic form of the molecule. This provides the ability to determine the thermochemistry of isotopically labeled molecules as well as isotopic effects of rate constants using transition state theory (TST). For the HF level of theory, the vibrational frequencies tend to be too large compared to experimental values. We therefore scale the HF harmonic frequencies downward by 12 per cent, as suggested by Pople¹⁶. For the DFT method, no scaling is used. The resulting vibrational frequencies are used subsequently to determine the zero-point energy of the molecule and other statistical mechanical properties.

Having determined the structure and vibrational frequencies of the molecule, the next step is to determine the electronic energy. To obtain useful electronic energies, higher levels of

electronic-structure theory must be used which incorporate electronic correlation. For the BAC-MP4 method, we use Møller-Plesset perturbation theory through fourth order, MP4¹⁷ (also denoted MP4SDTQ). The basis set used is a split-valence basis set with polarization functions on all atoms (6-31G(d,p) or 6-31G**) ¹¹. This method is denoted MP4/6-31G(d,p). For the BAC-G2 method, we use the G2 method itself⁷, with its series of QCI, MP4, and MP2 methods with varying basis sets. The geometry for these single-point calculations is re-optimized with the MP2 method, as defined in the G2 method.⁷

For the BAC-Hybrid method, two separate single-point calculations are performed, one at the DFT level and the other at the MP2 level of theory. First, a B3LYP calculation, as in the geometry optimization, is performed but with a larger basis set, 6-311++G(2DF,2PD). This corresponds to a triple-zeta basis set with balanced polarization functions (i.e., double D's, single F's) on the heavy atoms and analogous polarization functions on the hydrogen atoms. In addition, diffuse functions are added to treat ionic effects. This method is denoted B3LYP/6-311++G(2DF,2PD). Along with the B3LYP calculation, an MP2 calculation is performed with the same basis set. This method is denoted MP2/6-311++G(2DF,2PD).

2.2. Bond Additivity Corrections

While the *ab initio* electronic structure methods involving electron correlation, such as the Møller-Plesset perturbation methods (MP2, MP4) and coupled-cluster approaches, are very sophisticated, the energy provided by most practical methods and basis sets do not provide the level of accuracy required for useful heats of formation. However, we find that when using these methods and basis sets, the resulting errors in the electronic energy are quite systematic. These systematic errors can be separated, to a large extent, into three parts: (1) those due to the individual atoms making up the molecule; (2) those due to the molecule as a whole; and (3) those due to individual bonds. These three sources of errors can be corrected by using empirically derived expressions, denoted collectively as the Bond Additivity Correction (BAC). The corrections are added to the calculated electronic energy. The resulting energy is combined with statistical mechanics calculations to provide thermochemical properties of the molecule. In the following subsections, we present the empirical expressions contributing to the BAC method.

2.2.1. Atomic Corrections

The atomic corrections arise from errors due to intra-atomic electron correlation. These errors result from the treatment of the core electrons, the change in valence electron coupling, and the interaction between the valence and the core. The old procedure (BAC-MP4) adjusted the calculated electronic energy so that it reproduced the experimental heat of formation of the atom.² This is equivalent to choosing the reference state for each element to be a given electronic state of the gas-phase atomic species, rather than some arbitrary physical state, such as graphite for carbon or the diatomic $^3\Sigma_g^-$ state for oxygen. The atomic reference state is useful in quantum chemistry calculations, since the calculated energy of a molecule is referenced to the atomization energy of the molecule. Experimentally, the reference states were chosen to be some common physical state of the element at STP, either a solid, liquid, or gas, because the experimental heat of formation of the gas-phase atom was not known. However, heats of atomization are closer to the objective of determining the heats of formation of small molecules. Furthermore, the experimental heats of formation of atomic species, at least for the lighter elements, are typically quite well established (there are a few exceptions, such as boron). In addition, changing the reference state of an element, requires the heats of formation of all the compounds containing that element to be changed, as was recently the case for phosphorous compounds, illustrating the interdependence of heats of formation.

By choosing the reference state to be that of the gas-phase atom, we can avoid the problems associated with these arbitrary reference states and concentrate on determining the heat of atomization of the molecule, i.e., the negative of the heat of formation of a molecule from its constituent atoms. However, there is still an issue over which atomic state of the atom to use. Typically, one chooses the lowest electronic energy state of the neutral atom. For most compounds containing carbon, for instance, the atom first promotes its electronic configuration from an s^2p^2 (3P) to an s^1p^3 , which itself is a mixture of pure electronic spin states (3S , 3D , etc.). Each of these introduces different electron correlation energy errors and basis-set dependencies. Likewise, contributions from the ionic states of the atom also contribute to the molecular heats of formation. Finally, spin-orbit coupling and relativistic effects contribute to the error in the calculated heats of formation of the atomic states as well as of the overall molecule.

In the BAC-MP4 method, we took the atomic parameter to be the heat of formation of the ground electronic state of the gas phase atom. In the new procedure, we have decided to let the atomic parameters be adjustable, to account for calculation errors due to the changes in electronic configuration, charge, and spin-orbit coupling, as well as errors do to core interactions, core-valance interactions, and other relativistic effects. For many of the elements, these corrections are small, but can become significant as the molecule grows in size.

We therefore define an atom-based BAC correction,

$$E_{\text{BAC-atom}} = \sum_k E_{\text{BAC-atom}}(A_k), \quad (1)$$

where the sum is over all the atoms in the molecule, and $E_{\text{BAC-atom}}(A_k)$ depends on the atom type.

It should be noted that in a chemical equation, the number of atoms of each type occurring on each side of the reaction are identical. Thus, these atomic corrections exactly cancel out. That is, the calculated heat of atomization is not affected by the choice of atomic parameters.

2.2.2. Molecular Corrections

The second type of BAC correction arises from errors in the overall electronic structure of the molecule. The BAC correction for this molecular term is given by

$$E_{\text{BAC-molecule}} = E_{\text{BAC-elec pair}} + E_{\text{BAC-S}^2}, \quad (2)$$

where $E_{\text{BAC-elec pair}}$ depends on the net spin of the molecule and $E_{\text{BAC-S}^2}$ depends on any spin contamination in the electronic wavefunction. The first term arises from the basic nature of the interactions between electrons and occurs in atoms as well as molecules. These errors are systematic, resulting from the way electronic-structure codes treat electron correlation. The second term depends on the choice of wavefunction used in the calculation. In general, it is difficult to determine the functional form for the systematic error due to the interaction of electrons. The original BAC method represented this majority of this error within the bond-wise additive term. This term exponentially increased in value as two atoms approached each other to form a bond. However, the electron interaction error can also arise just by bringing an electron up to a molecule (forming a negative ion), without forming a bond. Likewise, removing a proton from a molecule removes a bond but does not remove the electron, which still interacts with the remaining molecule. Furthermore, when two electrons interact, the systematic electron correlation error is much larger when they are paired up (an electron bonding pair, in the case of a chemical bond) as opposed to when they are triplet-coupled. When they are triplet coupled,

their orbitals must be orthogonal by the Pauli exclusion principle, so their short-ranged interaction is reduced.

The systematic correction for the interactions between electrons can be represented by the total number of electrons that are interacting and by the number of these electrons that are interacting as electron pairs. Since the total number of electrons in a molecule is already treated by the atomic parameter, the molecular parameter is given by the net change in the spin of the molecule relative to that of the individual atoms,

$$E_{\text{BAC-elec pair}} = K_{\text{Elec pair}} (\text{Spin}_{\text{Molecule}} - \sum_{\text{Atom}} \text{Spin}_{\text{Atom}}). \quad (3)$$

where $K_{\text{Elec pair}}$ is an empirically adjusted parameter for a given BAC method. This term is very similar to the empirical correction term found within the G2 method.

Note that this term destroys size consistency. That is, the sum of the energies for two hydrogen atoms at infinity (two doublets) is not the same as the single energy representing a hydrogen molecule pulled apart to infinity (a singlet; note there is also a triplet surface). For chemical reactions, such as an abstraction or an elimination reaction, that maintain the same spin state, this is not a problem. However, care should be taken in relating combinations of chemical species to points on a potential energy surface.

Another error due to electron correlation is related to the spatial extent of the electrons in a molecule. In general, the amount of error follows the trend cations < neutrals < anions. While we investigated the possibility of using parameters to correct for systematic "net charge" dependence, the electron-pair-change correction (Eqn. 3 above) addresses the majority of the error, and no additional systematic correction was found to substantially improve the resultant discrepancies with experiment data for positive and negative ions.

An additional systematic error arises for open-shell molecules that use methods derived from the unrestricted Hartree-Fock approach (UHF). This error, $E_{\text{BAC-S2}}$, arises from spin contamination in the wavefunction (e.g., a doublet state may have quartet-state spin contamination). The resulting correlated energy (from an unrestricted UMP4 calculation, for instance) still contains contributions from the contaminated spin states. We therefore have included an energy correction² using the energy of the projected method. Unfortunately, this energy is not always available in computational methods. For methods such as G2, the energy prior to the application of any corrections is obtained from a QCI calculation. Energies obtained from Møller-Plesset calculations are used in the form of differences between pairs of calculations, each set using the same method but differing in the basis-sets. Thus, the systematic errors from the Møller-Plesset terms tend to cancel. For the unrestricted DFT methods, the spin contamination is smaller, so its effect is less serious. Calculations using restricted open-shell methods, which do not have spin contamination, indicate good results can be obtained without the need for this correction.

A separate "spin-dependent" correction was added to the BAC-MP4 method for closed shell systems^{2,3}. This correction used the spin contamination predicted by an unrestricted calculation of the singlet state to indicate the need for a multi-reference wavefunction (i.e., one expects a larger energy error using a single-reference configuration). However, this extra calculation is itself time-consuming and not guaranteed to find the instability. As the basis-set size increases, this systematic error decreases. Also, higher levels of theory, such as coupled cluster and QCI, more effectively treat the multi-reference correlation energy. In contrast with *ab initio* electron-correlation methods, methods such as DFT tend not to have an unrestricted wavefunction instability. For these reasons, we have eliminated the unrestricted wavefunction spin correction

in the newer BAC methods, since the corrections are not always available. The spin-pairing correction (Eqn. 3) partly compensates for it. The larger basis sets also reduce the size of the correction. However, the spin correction served another useful role as an indicator of greater uncertainty in the accuracy of the calculated value. The contribution to the "error bar" estimate, indicating greater multi-reference contributions to the energy, will be lost.

2.2.3. Bond Additivity Corrections

The third type of BAC correction depends on the formation of chemical bonds. One must distinguish between bonds and pair-wise interactions. A bond is taken to mean the formation of an electron pair between the atoms. The correction is meant to address the type of systematic error involved in the electron-pairing correction given in Eqn. 3. The bond correction is dependent on the identity of the neighboring bonds, which contribute basis-set superposition errors that become increasingly important as the basis set increases in size. Most efforts to develop corrections for quantum chemistry procedures have not adequately addressed this problem. This is due to the small size of the molecules being tested (typically less than seven heavy, i.e., non-hydrogen, atoms) and to the limited accuracy of experimental data for larger non-hydrocarbon, non-saturated gas-phase species.

In the BAC-MP4 method^{2,3}, the correction for each bond A-B in the molecule having neighbors C and D (e.g., C-A-B-D) was of the form

$$E_{\text{BAC-bond}}(\text{AB}) = - \{ A_{\text{AB}} e^{-\alpha_{\text{AB}} R_{\text{AB}}} \} \prod f_{\text{CAB}} \prod f_{\text{DBA}}, \quad (4)$$

where the first term in brackets is the correction for the bond by itself. The correction for nearest neighbors is treated as the product of corrections for each neighbor of the form

$$f_{\text{CAB}} = (1 - g_{\text{AC}} g_{\text{AB}}), \quad (5)$$

where

$$g_{\text{AC}} = k_{\text{C}} e^{-\alpha_{\text{AC}} (R_{\text{AB}} - 1.4)}. \quad (6)$$

The neighboring-bond corrections have the effect of reducing the size of the bond correction due to basis-set superposition errors.

For the new BAC methods, the correction for each bond A-B in the molecule having neighbors C and D (e.g., C-A-B-D) is given by

$$E_{\text{BAC-bond}}(\text{AB}) = -A_{\text{AB}} e^{-\alpha_{\text{AB}} R_{\text{AB}}} + \sum_{\text{C}} B_{\text{CA}} + \sum_{\text{D}} B_{\text{DB}}, \quad (7)$$

where the first term is the correction for the bond alone, while the correction for its nearest neighbors is treated as a sum of corrections for each neighbor of the form

$$B_{\text{CA}} = B_{\text{C}} + B_{\text{A}}. \quad (8)$$

The B_{A} 's are constants dependent only the type of atom. The bond distance dependence in the new form exists only in the first term for the bond itself. Furthermore, α no longer depends on the type of bond.

This new functional form results from several factors. First, the spin-pairing correction (Eqn. 3) is responsible for the most of the error in the bond. In the BAC-MP4 functional form (Eqn.4), a major factor for the increase in the bond correction as the distance decreased is that multiple bonds (spin pairings) are being formed. For instance, the triplet state of ethylene has a bond distance closer to that of ethane. This correction is now treated by Eqn. 3. In addition, the new method assumes larger basis sets are being used, including diffuse functions that have been added to treat ionic character. Thus, the corrections themselves are smaller. Indeed, for the more ionic bonds, the correction might even be negative. Thus, the multiplicative form of Eqn. 4 (which reduces the size of the correction) is no longer meaningful. In general, bond corrections in Eqn. 7 have the opposite effect of the spin-pairing correction (Eqn. 3). The electron-correlation error due to the pairing of electrons (given by Eqn.3) is reduced by having a bond, due to the ability to correlate the electrons of the pair on different atoms.

The definition of a bond is itself subjective. A hydrogen bond would not count as a BAC correction. On the other hand, a dative bond, such as in N_2O , would count, since it involves rearrangement of electron pairing. Hydrogen-boron bonds are particularly subjective. In diborane, B_2H_6 , the bridged hydrogens are considered as having two bonds each, while the boron-boron interaction is considered not to be a bond. For transition-state structures, both the bonds being formed and the bonds being broken are included in the BAC corrections. In the BAC-MP4 method, the stretching of the bond distance for the breaking and forming of bonds in the transition state structure decreases the size of the correction for each, which compensates for having only one bond defined for the reactants or products.

For the new BAC methods, the bond corrections do not go to zero at infinity, due to the terms $\sum B_{CA} + \sum B_{DB}$ in Eqn. 7. In addition, for the BAC-DFT method, the exponential term does not go to zero at infinity since α is 0. However, it should be noted that the spin pairing term (Eqn. 3) for the atoms at infinity does not correspond to the individual atoms.

2.2.4 Total BAC Correction

The resulting total BAC energy correction is given by the three contributions to the error: atomic (Eqn. 1), molecular (Eqn. 2), and bond-wise (Eqn. 7).

$$E_{BAC}(Total) = \sum_{ij} E_{BAC-bond}(A_i A_j) + \sum_k E_{BAC-atom}(A_k) + E_{BAC-clc\ pair} + E_{BAC-S2}, \quad (9)$$

where ij is summed over chemical bonds in the molecule while k is summed over the atoms in the molecule. The last term, E_{BAC-S2} , is only used in the BAC-MP4 and BAC-MP2 methods.

We have developed BAC corrections for two new procedures, BAC-G2 and BAC-Hybrid. These two BAC methods themselves involve several methods. For instance, the formulations in the BAC-G2 method are also used for the BAC-G2MP2 and BAC-G1 methods. The BAC-Hybrid entails both a new BAC-MP2 method as well as a BAC-DFT method. The methods all use Eqn. 9, with the bond-wise corrections $E_{BAC-bond}$ given by Eqn. 7. The differences are found in the specific form of the parameters in the formulae.

For the BAC-G2 method, the α exponent in Eqn. 7 is taken to be 3.0 \AA^{-1} , while the pre-exponential coefficient A_{AB} is taken to be the geometric mean of the individual atom types, i.e.,

$$A_{AB} = (A_{AA} A_{BB})^{1/2}. \quad (10)$$

For the BAC-DFT method, the α exponent in Eqn. 7 is taken to be 0.0, i.e., the bond-dependent error does not depend on the bond distance. For the BAC-MP2 method, the α exponent in Eqn. 7 is taken to be 3.0 \AA^{-1} , as for the BAC-G2 method. The resulting BAC-G2 parameters for selected elements are given in Table I.

Table I. BAC-G2 Parameters (Energies in kcal-mol⁻¹)

$$K_{\text{Elec pair}} = 0.430$$

Atom	K_{Atom}	B_{Atom}	A_{ii}
H	0.485	-0.146	1.462
C	1.081	0.051	0.0
N	1.498	-0.010	2.281
O	-0.501	-0.010	114.3
F	-1.942	0.215	373.1
Si	0.097	0.008	297.4
P	0.277	0.042	1121.8
S	0.155	0.003	921.4
Cl	-0.776	0.087	1433.7
Br	-1.703	-0.106	-0.106

For the new BAC-MP2 method, the open-shell molecules contain the spin contamination term

$$E_{\text{BAC-S2}} = K_{\text{spin cont}} (\text{PUMP2} - \text{UMP2}), \quad (11)$$

which is just the energy difference between the projected and unprojected energies. This is equivalent to saying that the PUMP2 energy is used for the MP2 energy for open-shell systems.

2.2.5 BAC Error Estimates

We had previously developed an error estimate² (or confidence level) for the BAC-MP4 energy. The approach was based on the convergence of the molecular energy at lower levels of perturbation theory, i.e., BAC-MP2, BAC-MP3, and BAC-MP4SDQ. In addition, both the BAC correction due to spin contamination, $E_{\text{BAC}} (\text{Spin } s_2)$, and the UHF instability, $E_{\text{BAC}} (\text{Spin UHF-I})$, have uncertainties. We defined the BAC-MP4 error estimate to be

$$\begin{aligned} \text{Error (BAC-MP4)} = & \text{Sqrt} \{ 1.0 \text{ kcal-mol}^{-1} + (\Delta H_{\text{BAC-MP4}} - \Delta H_{\text{BAC-MP3}})^2 \\ & + (\Delta H_{\text{BAC-MP4}} - \Delta H_{\text{BAC-MP4SDQ}})^2 \\ & + 0.25 (E_{\text{BAC}} (\text{Spin } s_2) \text{ or } E_{\text{BAC}} (\text{Spin UHF-I}))^2 \}. \end{aligned} \quad (10)$$

For the BAC-G2 method, we define a similar estimate based on the similarities between the G1 and G2-MP2 methods and the G2 method itself⁷, i.e.,

$$\begin{aligned} \text{Error (BAC-G2)} = & \text{Sqrt} \{ 1.0 \text{ kcal-mol}^{-1} + (\Delta H_{\text{BAC-G2}} - \Delta H_{\text{BAC-G2MP2}})^2 \\ & + (\Delta H_{\text{BAC-G2}} - \Delta H_{\text{BAC-G1}})^2 \}. \end{aligned} \quad (11)$$

For the BAC-Hybrid method, there are no perturbation expansion terms to use. Instead, we use the difference between the BAC-MP2 and BAC-DFT methods as an indication of consistency, i.e.,

$$\begin{aligned} \text{Error (BAC-Hybrid)} = & \text{Sqrt} \{ 1.0 \text{ kcal-mol}^{-1} + (\Delta H_{\text{BAC-MP2}} - \Delta H_{\text{BAC-DFT}})^2 \}. \end{aligned} \quad (12)$$

2.3. Heats of Formation, Entropies, and Free Energies

Having obtained the structure, vibrational energies, *ab initio* electronic energy, and the BAC energy corrections, we can then calculate the enthalpies, entropies, and free energies of a molecule using statistical mechanics (see Fig. 1). The equations used in these calculations are an extension of the subroutines in the Gaussian codes¹⁸. These subroutines use standard statistical mechanical expressions for an ideal gas in the canonical ensemble to compute the entropy, heat capacity, and internal energy.

The heat of formation at 0 K (ΔH_{f0}^0) can be obtained from the calculated electronic energy. First, the electronic energy is added to the zero-point energy. Unscaled frequencies are used, both for the old BAC-MP4 procedure, which used the Hartree-Fock method for geometry optimization and frequencies, as well as for the new BAC method, which uses the DFT method for geometry optimization and frequencies. Next, the resulting energy is subtracted from the electronic energies of the atoms to give an electronic heat of atomization. Referencing this energy against the experimental ΔH_{f0}^0 of the atoms in the gas phase yields the uncorrected ΔH_{f0}^0 . Adding the BAC corrections to this energy finally yields ΔH_{f0}^0 at 0 K. Statistical-mechanical calculations then provide the thermochemical properties of the molecule, e.g., partition functions, enthalpy, heat capacity, and entropy for given temperatures. The values of these thermochemical properties as a function of temperature are then fit to an analytic polynomial for the modeling of thermochemical processes.

Note that for finite temperatures, the G2 energies without BAC corrections do not correspond to those from the output of a Gaussian G2 calculation, since hindered rotors are included in the BAC procedure. Also, the heats of formation are not absolute values, but rather, depend on the changes in the heats of formation of the atoms relative to the reference state of the element.

3. Results

The accuracy of the various BAC methods has been determined by comparing their predictions with the experimental heats of formation for approximately one hundred to two hundred molecular species. For the BAC-MP4 method, the average difference is 1.3 kcal-mol⁻¹ for 93 stable (closed shell) species (see Table II).

Table II. Comparison of $\Delta H_{f, 298}^\circ$'s for BAC-MP4 vs. Experiment (Energies in kcal-mol⁻¹).

Compound Class	Number of Compounds	Average Energy Difference
C _x H _y	27	1.3
C _x O _y H _z	23	1.4
C _x N _y H _z	10	0.9
C _x N _y O _z H _w	8	1.3
N _x O _y H _z	12	1.2
Fluorinated species	13	1.2
Total	93	1.25

Table III shows the results for chlorinated methanes obtained from the MP2 method and various DFT hybrid methods, using the 6-311++G(2df,2pd) basis set. The B3LYP_{OPT} method refers to the B3LYP method at the smaller basis set, 6-31G*, indicating the dependence of the error on basis-set size. CH₄ and CCl₄ were used to determine the parameters. One can see that while the raw predictions of the MP2 method are worse than the DFT method, the BAC-MP2 method gives a slightly better result than does the BAC-DFT method.

Table III. BAC $\Delta H_{f, 298}^\circ$'s using various methods and basis sets. For each molecule, the raw values are in the first row and the total BAC corrections are in the second row (Energies in kcal-mol⁻¹).

Molecule	MP2	B3LYP	B3LYP _{OPT}	HF6/DFT4
CH ₄	10.4 0.0	-0.3 0.0	-0.3 0.0	-9.1 0.0
CH ₃ Cl	7.0 0.2	2.0 -1.7	2.7 -2.5	-11.2 -1.9
CH ₂ Cl ₂	3.9 0.8	5.9 -2.1	8.0 -2.9	-11.6 -2.0
CHCl ₃	0.1 0.8	10.7 -1.6	14.6 -2.0	-11.1 -1.2
CCl ₄	-4.4 0.0	16.4 0.0	22.1 0.0	-10.0 0.0

In table IV, we present the resulting accuracy of the BAC-Hybrid and BAC-G2 methods for selected compounds. The BAC-Hybrid method used a smaller set of the test compounds, due to known deficiencies in treating biradical species. The accuracy of the hybrid method HF6/DFT4, with 60% HF and 40% DFT, is worse than that obtained from the B3LYP, particularly for positive and negative ions. Therefore, it was not pursued further.

The BAC-MP2 method gives somewhat better results than the BAC-DFT method. Thus, the BAC-MP2 method was used to define the BAC-Hybrid method, with the BAC-DFT method being used for determining the error estimate.

Table IV. Accuracy of various BAC procedures (Energies in kcal-mol⁻¹).

Method	Avg. Error	RMS Error	Max Error
BAC-Hybrid (122 neutral compounds)			
MP2/6-311++G(2df,2pd)	0.78	1.13	3.69
B3LYP/6-311++G(2df,2pd)	0.93	1.28	4.04
B3LYP/6-31G(d)	1.10	1.54	5.43
HF6/DFT4/6-311++G(2df,2pd)	1.28	1.90	7.26
BAC-G2 methods (143 neutral compounds)			
G2	0.69	0.90	2.68
G2MP2	0.74	0.95	3.02
G1	0.75	0.98	2.88
QC1/6-311G(d)	0.82	1.11	3.92
MP4/6-311+G(d,p)	0.95	1.32	4.38

In table V we show those molecules for which the BAC-G2, BAC-MP2 and BAC-DFT methods, respectively, yield the least agreement with experiment. For comparison, we present the old BAC-MP4 method. These data help define the limitations of the method and indicate classes of compounds for which the present procedure may be inadequate (see Discussion section).

Table V. Outliers for the BAC-G2, BAC-MP2, and BAC-DFT methods. Differences between theory and experiment (energies in kcal-mol⁻¹).

Molecule	BAC-G2 ^a	BAC-MP2	BAC-DFT	BAC-MP4
Carbon-containing				
C	-1.5 (0.4)	-6.0	13.0	-8.6
CCl ₄	0.1 (-2.9)	-0.8	1.8	2.7
CF ₄	0.4 (-5.6)	0.8	1.7	-0.3
² CH	-1.3 (-0.6)	1.6	8.2	4.8
⁴ CH	2.3 (3.9)	-0.3	8.2	0.0

HNCO	-3.9 (-4.7)	-1.9	-3.8	-3.7
CH ₂ S	4.0 (3.4)	5.3	5.5	1.6
CH ₃ O	1.8 (0.5)	4.3	-2.4	2.4
(NH ₂) ₂ CO	2.8 (1.8)	5.2	5.0	1.1
CN	2.4 (3.3)	6.0	5.2	4.2
C ₂ F ₄	-1.3 (-7.8)	-0.7	-0.5	1.7
C ₂ H ₅ NO ₂	-2.4 (-4.1)	-1.6	-0.2	-0.4
Cyclopropene	1.3 (2.7)	4.5	0.1	1.8
Cyclopropane	-0.5 (0.6)	1.4	-4.0	-1.0
Methylene- cyclopropane	-2.3 (-0.6)	1.6	-5.3	-1.1
Cyclobutene	3.6 (5.3)	4.9	3.5	3.2
Benzene	0.6 (3.6)	-0.4	1.7	-2.8
Non-carbon-containing				
ClO	1.0 (2.1)	6.5	-2.6	4.8
ClOO	5.2 (6.4)	-19.8	-0.5	12.3
OCIO	-0.5 (4.8)	0.1	0.8	13.6
FNO	-5.2 (-7.7)	-5.2	-2.5	-1.2
HNO	1.5 (0.7)	3.0	4.0	-0.5
³ HNO	-0.5 (-0.4)	5.9	-3.4	0.0
O ₂ NNO ₂	1.0 (-2.6)	-8.6	-1.7	-5.5
HNNH	-3.1 (-3.5)	-1.0	-1.1	-3.9
NO ₃	3.9 (1.5)	-8.2	-1.5	5.7
O ₂	1.6 (2.4)	-8.4	2.0	0.0
O ₃	-1.6 (-1.1)	-16.5	7.9	0.4
SO ₂	0.7 (5.0)	-1.6	0.7	-1.6
SO ₃	-0.6 (6.8)	1.1	0.3	2.8
HOSO ₂	1.6 (8.1)	5.0	-1.9	-24.3 (-3.2) ^b

^aG2 raw energies in parentheses.

^bBAC-MP4/MP2 method in parentheses (geometry optimized using the MP2 method).

4. Discussions

The results show that, as expected, the BAC-G2 method gives the best results (see table IV). In fact, for the majority of species, the G2 method itself is sufficiently accurate without needing the BAC method. Furthermore, the BAC-G2 method requires many fewer parameters than the other methods. While this method is the most accurate of those presented in this paper, it is also computationally the most expensive. The BAC-MP2 and BAC-DFT are the cheapest computationally and will scale better for larger molecules.

A significant advantage of the BAC-G2 method is that its parameters depend only on the type of atomic elements contained in the molecule.. The BAC-MP4 and BAC-Hybrid methods, on the other hand, use parameters that depend on pairs of element types. [Note, this can lead to inappropriate BAC parameterization because agreement may be forced with experiment when the experiment itself is wrong.] Thus, by using the BAC-G2 method, only a few reference

compounds containing a given element are needed to determine the parameters for all compounds containing that element. The BAC-G2 method can then be used to determine the BAC parameters for each bond type of interest for the BAC-MP4 and BAC-Hybrid methods, for which experimental data are either not available or are not trusted. Of course, care should be taken since the G2 method uses similar methodologies in its procedure, i.e., the Møller-Plesset perturbation theory expansions, so errors inherent in G2 could propagate.

There are certain species for which the G2 method is not sufficiently accurate (see table V). Species involving the halogens and oxides are particularly problematic. A major part of this error is possibly due to issues of molecular size. The errors in the G2 method tend to grow as the number of non-hydrogen atoms in the molecule increases. Unfortunately, the G2 method is rather limited in the number of heavy atoms it can handle, so it is difficult to determine the exact dependence of the error on the size. The BAC method attempts to correct for this size dependence with nearest-neighbor bond corrections (see Eqn. 7), though the atomic (see Eqn. 1) and molecular (see Eqn. 3) BAC corrections also depend on the size. As one can see from table V, the BAC-G2 results represent a significant improvement over the uncorrected values for the larger molecules. Equally important, one wants predictive capability. Thus, reducing the maximum error (the last column of table IV) as much as possible is very important. A small number here means not only that the BAC-G2 method yields accurate results, but also that it does so for all compounds, not just on the average. The 143 species used in the accuracy evaluation are not just stable species, but were chosen to include different types of chemical bonds and resonant structures. Ions were not included, due to uncertainty in the accuracy of this data. The heats of formation of the 143 neutral species are quite well established experimentally. This strongly suggests that for the outliers in the BAC-G2 method (table V), those BAC-G2 heats of formation that differ from experiment by more than 3 kcal-mol⁻¹ are more likely to be correct than the corresponding experimental value. This is particularly likely if the other BAC methods give similar results.

For the BAC-Hybrid method, only 122 species were used for the determination of parameters. The larger list of 143 compounds used for the BAC-G2 method includes species known to require multireference configurations; these were excluded in the determination of the parameters, but the results of these outliers are shown in table V. It is encouraging that the BAC-MP2 and BAC-DFT tend to not give the same incorrect results. This may be due to the different approaches used by these methods to treat electron-pair correlation. Thus, by combining BAC-MP2 and BAC-DFT together within the BAC-Hybrid method, one can use the difference between the two methods as an indicator of the accuracy (see Eqn. 12). It should also be noted that many of these outliers represent unique electron correlation issues that will not extend to the larger molecules for which the BAC-Hybrid method would be used. When in doubt about a chemical moiety, one should find a smaller molecular prototype and perform the calculation at a higher level of theory. One can then use the differences in the BAC methods for the smaller molecule as an additional BAC correction factor, i.e., a chemical moiety correction factor for a specific chemical moiety.

$$E_{\text{BAC}}(\text{Total}) = \sum_{ij} E_{\text{BAC-bond}}(A_i A_j) + \sum_k E_{\text{BAC-atom}}(A_k) + E_{\text{BAC-elec pair}} + E_{\text{BAC-S2}} + E_{\text{BAC-chemical-moiety}} \quad (9')$$

This correction has been used to correct the results of the old BAC-MP2 method for large aromatic hydrocarbons using the BAC-MP4 results on a small aromatic species.

5. Conclusions

New BAC procedures have been developed that extend the capabilities of the BAC-MP4 method. The BAC procedures are designed to balance the computational time between optimization searching, frequency evaluation, and single-point energy calculations. The new BAC parameterization can treat positive and negative ions as well as neutrals. Diffuse basis functions permit better treatment of ionic character. The BAC-G2 method provides high accuracy for small species. The BAC-Hybrid method uses the DFT procedure for improved geometry optimization and frequency evaluation. It uses both the MP2 and DFT procedures for the single-point calculations. The BAC-DFT and BAC-MP2 methods provide complimentary treatments of electron correlation, providing an error estimate at these low levels of theory. The DFT and MP2 methods are well suited for massively parallel and distributed computational platforms and should scale well for large molecules.

The new BAC parameterization should be readily extendible to other quantum chemistry procedures being developed. It should be readily extendible to hybrid methods, such as mixing density functional geometry optimizations with coupled cluster methods. It should therefore be possible to derive parameters for a BAC-CCSD method, in which the geometry is determined by the DFT method, while the single-point calculation is performed using the coupled-cluster with triples method, CCSD(T)⁸. It should also be possible to apply this BAC method to CBS procedures, analogous to the implementation of the BAC-G2 method, or to a combined CCSD(T)/CBS²⁰ methods. In addition, the BAC method should be applicable to new algorithms such as quadratically-scaling local-correlation methods. Errors introduced in the local-correlation approximation are likely to be systematic, so the BAC corrections should be applicable. These new BAC methods can be used to determine the parameters needed for BAC methods using the lower levels of theory. Further work needs to be done looking at nonlinear errors that accumulate as a function of size of the molecule.

References

- ¹Computational Thermochemistry, K. K. Irikura and D. J. Frurip, Ed., ACS Symp. Series Vol. 677 (1998).
- ²C. F. Melius, in Chemistry and Physics of Energetic Materials, S. N. Bulusu, Ed. (Kluwer Academic Publishers, Dordrecht, 1990), Vol. 309, pp. 21.
- ³P. Ho and C. F. Melius, J. Phys. Chem. 94, 5120-5127 (1990).
- ⁴C. F. Melius, M. E. Colvin, N. M. Marinov, W. J. Pitz, and S. M. Senkan, *Twenty-Sixth Symposium (International) on Combustion* (The Combustion Institute, Pittsburgh, PA, 1996), p. 685-692.
- ⁵M. R. Zachariah and C. F. Melius, J. of Phys. Chem. 101, 913-918 (1997).
- ⁶B. C. Garrett, M. L. Koszykowski, C. F. Melius, and M. Page, J. Phys. Chem. 94, 7096 (1990).
- ⁷L. A. Curtiss, K. Raghavachari, G. W. Trucks, J. A. Pople, J. Chem. Phys., 94 (1991) 7221.
- ⁸J. D. Watts, J. Gauss, R. J. J. Bartlett, Chem. Phys. 1993, 98, 8718.
- ⁹C. C. J. Roothaan, Rev. Mod. Phys. 23, 69 (1951).
- ¹⁰J. A. Pople and R. K. Nesbet, J. Chem. Phys. 22, 571 (1954).
- ¹¹P. C. Hariharan and J. A. Pople, Theor. Chim. Acta 28, 213 (1973); M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, J. Chem. Phys. 77, 3654 (1982).
- ¹²A. D. Becke, J. Chem. Phys., 98 (1993) 5648.
- ¹³C. Lee, W. Yang, R. G. Parr, Phys. Rev. B, 37 (1988) 785.
- ¹⁴P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, J. Phys. Chem., 98 (1994) 11623.
- ¹⁵J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, Int. J. Quant. Chem. S13, 225 (1979).

- ¹⁶J. A. Pople, H. B. Schlegel, R. Krishnan, D. J. DeFrees, J. S. Binkley, M. J. Frisch, and R. A. Whiteside, *Int. J. Quant. Chem.* S15, 269 (1981).
- ¹⁷R. Krishnan and J. A. Pople, *Int. J. Quant. Chem.* 14, 91 (1978); R. Krishnan, M. J. Frisch and J. A. Pople, *J. Chem. Phys.* 72, 4244 (1980).
- ¹⁸M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith; G. A. Petersson, J. A. Montgomery, K. Raghavachari; M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe; C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, J. A. Pople, *Gaussian 94, Revision D.3*, Gaussian, Inc., Pittsburgh, , 1995.
- ¹⁹G. A. Petersson, in *Computational Thermochemistry*, K. K. Irikura and D. J. Frurip, Ed., ACS Symp. Series Vol. 677 (1998), p. 237.
- ²⁰J. M. L. Martin, *Chem. Phys. Lett.* 259, 669 (1996).

DISTRIBUTION:

1	MS 0601	P. Ho, 1126
1	MS 9001	M. E. John, 8000
1	MS 9003	D. L. Crawford, 5200
1	MS 9405	T. M. Dyer, 8700
1	MS 9214	Ida Nielsen, 8130
1	MS 9214	Carl Melius, 8130
1	MS 9214	Carmen Piqueras, 8130
1	MS 9214	Curtis Janssen, 8950
1	MS 9056	W. J. McLean, 8300
1	MS 9052	D. R. Hardesty, 8361
1	MS 9052	M. D. Allendorf, 8361
1	MS 9052	R. Behrens, 8361
1	MS 9052	Anthony McDaniel, 8361
1	MS 9041	J. S. Binkley, 8345
1	MS 9051	J. F. Grcar, 8345
1	MS 9055	F. P. Tully, 8353
1	MS 9055	J. A. Miller, 8353
3	MS 9018	Central Technical Files, 8940-2
1	MS 0899	Technical Library, 4916
1	MS 9021	Technical Communications Department, 8815/ Technical Library, MS 0899, 4916
1	MS 9021	Technical Communications Department, 8815 For DOE/OSTI
1	MS 0188	D. Chavez, LDRD Office, 4001

This page intentionally left blank