A path integral approach to molecular thermochemistry

Kurt R. Glaesemann and Laurence E. Fried

Articles you may be interested in

Accurate ab initio potential energy surface, thermochemistry, and dynamics of the Cl(2P, 2P3/2) + CH4 → HCl + CH3 and H + CH3Cl reactions
J. Chem. Phys. **136**, 044307 (2012); 10.1063/1.3679014

Thermochemistry of new molecular species: SBr and HSBr
J. Chem. Phys. **126**, 204314 (2007); 10.1063/1.2737776

Quantitative molecular thermochemistry based on path integrals
J. Chem. Phys. **123**, 034103 (2005); 10.1063/1.1954771

Theoretical study on isotope and temperature effect in hydronium ion using ab initio path integral simulation
J. Chem. Phys. **121**, 5985 (2004); 10.1063/1.1785779

Vacuum ultraviolet pulsed field ionization study of ND 3 : Accurate thermochemistry for the ND 2 – ND 2 + and ND 3 – ND 3 + system
J. Chem. Phys. **120**, 8476 (2004); 10.1063/1.1691406

ARTICLES

A path integral approach to molecular thermochemistry

Kurt R. Glaesemann and Laurence E. Fried
Lawrence Livermore National Laboratory, University of California, Livermore, California 94551

The calculation of thermochemical data requires accurate molecular energies. When such high accuracy is needed, often a method such as G1, G2, or G3 is used to calculate the energy. These methods rely upon the standard harmonic normal mode analysis to calculate the vibrational and rotational contributions to the energy. We present a method for going beyond the harmonic analysis, which uses path integral Monte Carlo to calculate the vibrational and rotational contributions. Anharmonic effects are found to be as large as 2.5 kcal/mol for the molecules studied. Analytical methods for determining an optimal path discretization are presented. A novel potential energy caching scheme, which greatly improves computational efficiency, is also presented. © *2003 American Institute of Physics.* [DOI: 10.1063/1.1529682]

I. INTRODUCTION

It is well known that accurate molecular energies are needed for a variety of reasons, including the calculation of thermochemical rates and equations of state (EOS). Traditional approaches must include a quantum thermal energy (QTE) term that includes the quantum mechanical zero-point energy (ZPE) and the thermal energy. This QTE term is based upon the harmonic approximation of normal modes generated from the second-derivative of the energy at the ground state equilibrium geometry.[1] Rotations are treated through a rigid rotor approximation. This quantum correction is essential, because vibrations are inherently quantum mechanical and the ZPE is larger than the desired accuracy of 1–2 kcal/mol. The QTE calculation done in the traditional electronic structure calculations ignores vibrational anharmonicity and vibration/rotation coupling. This assumption is never completely correct, even at 0 K. The importance of anharmonic effects in molecules is well documented.[2,3] The overall anharmonic contribution to the molecular energy may either increase or decrease with temperature due to canceling repulsive and attractive terms.

One method commonly used for calculating highly accurate energies is the family of Gaussian-Y (G-Y) methods starting with the Gaussian-1 theory (commonly called G1). These methods have enjoyed a great deal of success in the literature. The G-Y methods were created to extrapolate the energy to the large basis set, high level theory limit by doing a series of smaller calculations. There are a variety of methods including the standard G1,[4] G2,[5] and G3.[6,7] Methods that replace the costly MP4 steps with a lower order perturbation theory have been proposed: G2(MP2),[8] G3(MP2),[8] and G3(MP3).[9] Attempts at using smaller basis sets in G2 have been made: G2(SZP).[10] The quadratic CI (QCI) step has been replaced with a coupled cluster CCSD(T) calculation in one variation of G3.[11] Methods incorporating density functional theory (DFT) have been proposed: G3X, G3X(MP3), and G3X(MP2).[12] Gordon *et al.* have proposed several multireference analogs to G2 and G3 for molecular species in

which a single Hartree–Fock determinant is not a good description of the wave function.[13] More accurate, but computationally more expensive methods, such as the Weizmann-2 theory (W2) have also been developed.[14,15] The W2 method includes explicit corrections for anharmonicity that are based upon spectroscopic data. This list of G-Y methods is certainly not complete. Thermochemical accuracy of 1–2 kcal/mol is generally achieved, although there are always some molecules that are statistical outliers. The G-Y methods are designed to have three independent factors contributing to the energy: basis set, electronic structure theory level, and treatment of nuclear rotation and vibration. In some G-Y methods, relatively small spin–orbit and higher-level corrections are added. The nuclear rovibrational effects are treated with a single normal mode analysis at the equilibrium geometry. Since the three primary facets of G-Y are largely independent, a deficiency in any one will limit the overall accuracy.

II. PATH INTEGRAL THERMOCHEMISTRY

We present a treatment of the quantum nature of the nuclei using path integral Monte Carlo (PIMC) (Refs. 16–18) instead of the harmonic normal mode approximation. The PIMC method has found wide application in quantum physics and chemistry,[19–23] and the reader is directed to these references for a full discussion of the method and the underlying physics. PIMC has been applied to the calculation of thermochemical properties using spectroscopic potentials and approximations to the propagator.[24] PIMC has been applied to the calculation of free energies for molecules using empirical potentials.[25–27] For clarity, we present the basic equations for a single particle of mass m, temperature T, and Hamiltonian $H = p^2/2m + V$ in one spatial dimension. We define the inverse temperature $\beta \equiv 1/k_B T$, where k_B is Boltzmann's constant. The PIMC method is based in a statistical mechanics approach to the molecular system. The initial step is defining the canonical ensemble's partition function Q for the molecular system based upon the systems atomic density

matrix: $Q \equiv \int dr \rho(r,r,\beta)$. By inserting $N-1$ resolutions of the identity into the density matrix, the single molecular system of temperature T is transformed to N replicas of the system, each at temperature NT. This yields a partition function that is an integral over the N replicas. This is known as the Trotter[28] path integral canonical partition function Q,

$$Q = \left(\frac{N}{\beta} \frac{m}{2\pi\hbar^2} \right)^{N/2} \int dr_1 \cdots dr_N e^{-\beta V_C}, \qquad (1)$$

where $\hbar$ is Planck's constant divided by 2π and V_C is defined by

$$V_C \equiv \frac{1}{N} \sum_{j=1}^{N} \left[\frac{N^2}{\beta^2} \frac{m}{2\hbar^2} (r_j - r_{j+1})^2 + V(r_j) \right], \qquad (2)$$

with $r_{N+1} \equiv r_1$ and the integral over all degrees of freedom. Equation (2) leads to the well-known interpretation that the quantum path integral is the same as a classical expectation value of the chain potential V_C, which consists of N beads bound together by harmonic interactions and acting with an external potential V. V_C is called the propagator potential because it includes both the molecular potential of each bead and the harmonic interbead potential. The harmonic interbead term comes from the kinetic portion of H, while the V term comes from potential portion of H. Equation (1) is quantum mechanically exact in the large N limit and classically exact in the $N=1$ limit.

The internal energy (often called just the energy) is evaluated as the expectation value,

$$E_s \equiv \frac{\int dr_1 \cdots dr_N \epsilon_s e^{-\beta V_C}}{\int dr_1 \cdots dr_N e^{-\beta V_C}}, \qquad (3)$$

with $r_1 \cdots r_N$ sampled according to V_C. The debate as to which of the common energy estimators ϵ_s is best is not fully resolved (thermodynamic ϵ_T, virial ϵ_V, and centroid virial ϵ_{CV}).[20–23,29–42] We choose to use our recently developed centroid thermodynamic estimator ϵ_{CT} which avoids the costly higher derivatives of the virial estimators, but projects out much of the numerical noise commonly reported with ϵ_T.[22,23] Despite the debate over the relative merits of the various estimators, all the energy estimators give the same answer for the integral provided the integration is carried out with an adequate sampling of dr. This centroid estimator is not to be confused with the powerful method of centroid molecular dynamics (CMD),[43,44] which yields quantum rate constants[45] and nonlinear operator expectation values.[46]

In this work, the integral in Eq. (3) is evaluated via Metropolis Monte Carlo.[47] The Metropolis Monte Carlo method uses a random walk in the dr phase space. Moves that result in an increased contribution to Q are always accepted, while those that result in a lower contribution are accepted at a rate of $\exp(-\beta(\Delta V_C))$. The Metropolis algorithm satisfies the detailed balance condition, which means that a Boltzmann probability distribution is obtained. The Metropolis algorithm also is ergodic, meaning that it is possible to reach any state of the system from any other state of the system if the run is long enough. The other common approach is to use

molecular dynamics (PIMD) to sample the dr phase space.[48] Monte Carlo requires a robust pseudorandom number generator, therefore we utilized an improved Mersenne Twister to generate the necessary random numbers.[49,50]

The enthalpy $\mathcal{H}$ is defined as $\mathcal{H} \equiv E + P\Omega$, where P is the pressure and Ω is the volume. The change in $\mathcal{H}$ during a reaction determines whether the reaction releases heat (exothermic, $\Delta\mathcal{H} < 0$) or absorbs heat (endothermic, $\Delta\mathcal{H} > 0$). The application of the ideal gas equation $P\Omega = nk_BT$ (n is the number of molecules), leads to the easily evaluated expression, $\mathcal{H} = E + k_BT$. $\mathcal{H}$ and the entropy of the system $\mathcal{S}$ defines the Gibbs free energy $\mathcal{G} \equiv \mathcal{H} - T\mathcal{S}$, which determines whether a reaction is thermodynamically favorable. The normal mode analysis method uses simple analytical forms for the contribution of translations, rotations, and vibrations to $\mathcal{S}$. Within path integral thermochemistry, the most straightforward method is to first calculate the constant volume heat capacity as $C_V \equiv (\partial E/\partial T)_\Omega$, and integrate to get the entropy,

$$\mathcal{S} = k_B \ln Q + k_B T \left(\frac{\partial \ln Q}{\partial T} \right)_\Omega = \int_0^T \frac{C_V(T')}{T'} dT'. \qquad (4)$$

As with E, there is a variety of PIMC methods for calculating C_V and the reader is directed to Ref. 23 for a review.

The path integral thermochemistry method intrinsically includes all rotation–vibration interactions and anharmonicities. Other methods that partially take these effects into account have been applied with varying degrees of success and computational intensity to thermochemistry, spectroscopic prediction, and force field generation. Methods utilized include: application of second-order rotovibrational perturbation theory,[51] vibrational self-consistent field with and without additional higher order corrections,[52,53] using theoretically calculated quadratic correction terms,[54] and using spectroscopically obtained anharmonicity constants.[55] The path integral thermochemistry approach should scale well into the high T regime unlike the more traditional sum over state approaches that can become cumbersome as the number of states increase. Path integral thermochemistry suffers from no inherent approximations: the approximations of basis set, level of electronic structure theory, and number of beads are all well defined and systematically improvable. Due to the exploratory nature of this paper, only internal energies are studied and the calculation of enthalpies and free energies will be presented in a later work.

III. EFFICIENT MONTE CARLO SAMPLING

There are certain practical considerations that must be taken into account, since the normal mode analysis is computationally fast compared to PIMC. Due to the necessity of long Monte Carlo runs to adequately sample the available phase space and the large values of N needed to converge Eq. (1), an efficient method must be used. The calculation of the classical potential V for each bead is the computational bottleneck, therefore minimizing the evaluations of V is very useful. The brute force approach requires NM evaluations of V, where M is the number of Monte Carlo steps. We have chosen to use a novel data caching scheme, in which a sparse evenly spaced grid is used to store values of V. This method
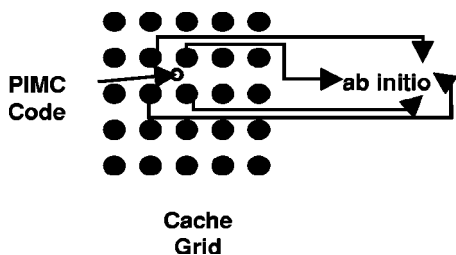
FIG. 1. Values of V are cached and interpolated with a grid.

is summarized pictorially in Fig. 1. Values of V are interpolated as needed from the grid using multilinear interpolation. Multilinear interpolation involves the product of linear functions. The values of the grid points are calculated as needed with the appropriate *ab initio* code. The grid acts as a transparent layer between the PIMC code and the *ab initio* code. The first time a point is needed from a region of phase space, the actual V at the needed point is calculated to verify that the interpolation errors are below a small tolerance. Internal coordinates of the molecules are used instead of Cartesian coordinates to increase the efficiency of the cache. We currently store only V in the cache files, although V' and V'' would be possible. The use of internal coordinates adds additional complexity to the caching of V' and V'', since they must be interconverted between internals and Cartesians. Since the grid only stores V and possibly its derivatives, the grid is independent of T, N, the choice of ϵ_s, M, and m (i.e., isotope). Although the cache could be replaced with a standard potential table, the cache used here differs from a traditional potential table in several ways. The cache is both sparse and dynamically generated, while a potential table is pregenerated for a predefined region of phase space. In this way the cache is more convenient, because one does not have to know in advance what regions of phase space are relevant and no time is wasted calculating unneeded points. The cache code supports operation in a ''lossy'' mode, in which points that have not been utilized recently can be removed if memory limits further growing of the cache. The lossy mode may be useful when studying large molecules. Additional work is currently being undertaken to improve the interpolation method with quadratic terms and to implement a parallel cache filling code. The underlying method may be changed to utilize another interpolation scheme such as inverse local internal coordinates in the future.[56-59]

As N increases, the harmonic terms of V_C become very stiff and very small Monte Carlo steps must be used. To overcome this problem, the beads are moved using normal mode sampling.[60] This method involves moving multiple beads simultaneously according to the interbead term in Eq. (2). This application of normal modes is not an approximation, because the $(r_j - r_{j+1})^2$ term in Eq. (2) is truly harmonic. These normal modes will not be discussed again to avoid confusion with the molecular normal modes.

In all simulations, dynamic adjustment of the Monte Carlo step size is *not* used to increase sampling efficiency. Dynamic adjustment uses different step sizes for different regions of phase space, so that acceptance ratios do not approach either 0% or 100%. Dynamic adjustment can add an

uncontrolled systematic error and result in a corresponding loss of both balance and detailed balance.[61]

IV. ERROR ESTIMATION IN PATH INTEGRAL THERMOCHEMISTRY

It would be very useful to know *a priori* how large N has to be, and consequently how quantum mechanical the system is, without resorting to a discussion of the thermal de Broglie wavelength relative to an arbitrary standard. For the harmonic oscillator such insight is possible. The exact quantum harmonic oscillator equations are simply,

$$E(\nu,\beta) = \frac{h\nu}{2} \frac{1 + e^{-\beta h\nu}}{1 - e^{-\beta h\nu}}, \tag{5}$$

$$C_V(\nu,\beta) = k_B(\beta h\nu)^2 \frac{e^{-\beta h\nu}}{(1 - e^{-\beta h\nu})^2}, \tag{6}$$

with ν being the frequency of the harmonic oscillator and h being Planck's constant. In the $M = \infty$ limit, E and C_V are readily calculated for harmonic oscillators within the PIMC finite bead approximation. The PIMC equations are simply the classical values ($k_B T$, k_B) multiplied by a quantum correction term,[23,30,62]

$$E(\nu,\beta,N) = k_B T \left(\frac{Nr}{\sqrt{4+r^2}} \right) \left(\frac{f^N+1}{f^N-1} \right), \tag{7}$$

$$C_V(\nu,\beta,N) = k_B \left[N \left(\frac{r}{\sqrt{4+r^2}} \right)^3 \left(\frac{f^N+1}{f^N-1} \right) \right.$$
$$+ 2N^2 \left(\frac{r^3}{\sqrt{4+r^2}} + \frac{2r^2+r^4}{4+r^2} \right)$$
$$\left. \times \left(\frac{f^{N-1}}{(f^N-1)^2} \right) \right], \tag{8}$$

where

$$r \equiv \frac{\beta h\nu}{N}, \tag{9}$$

$$f \equiv 1 + \frac{r^2 + r\sqrt{4+r^2}}{2}. \tag{10}$$

These equations are independent of the estimator used for calculating E and C_V.

Within the harmonic normal mode approximation, the translational and rotational contributions to the internal energy are calculated as

$$E^{\text{trans,rot}} = \frac{3}{2}k_B T + \frac{\vartheta}{2}k_B T, \tag{11}$$

where ϑ is zero for an atom, 2 for a linear molecule, and 3 for all other molecules. The resulting contribution to the heat capacity is trivial,

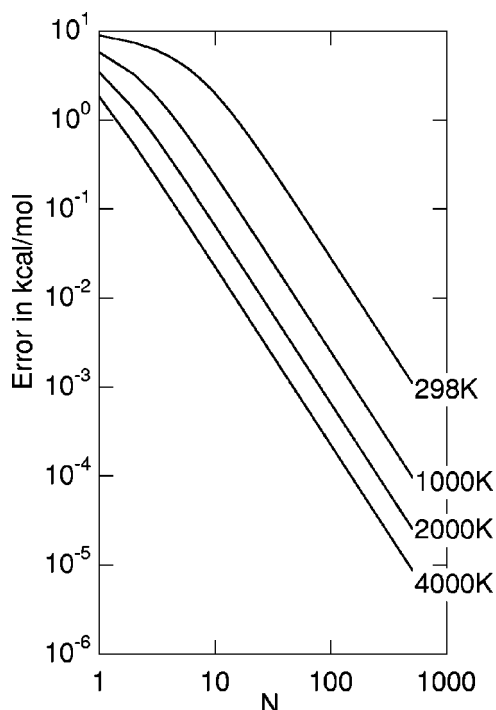$$C_V^{\text{trans,rot}} = \frac{3}{2}k_B + \frac{\vartheta}{2}k_B. \tag{12}$$

FIG. 2. Expected error ΔE in Eq. (13) with frequencies of UMBPT2/6-311G($2df,p$) $\cdot$ CH$_2^+$.

FIG. 3. Expected error ΔC_V in Eq. (14) with frequencies of UMBPT2/6-311G($2df,p$) $\cdot$ CH$_2^+$.

Oscillators with a high $\beta\nu$ (high frequency and low T) require a larger value of N than oscillators with a low $\beta\nu$. For molecules, we simply add up contributions of all frequencies (not the frequencies themselves) and compare the exact results of Eq. (5) with the finite N results of Eq. (7),

$$\Delta E = \sum_{i=1}^{freq} |E(\nu^i,\beta) - E(\nu^i,\beta,N)|, \tag{13}$$

$$\Delta C_V = \sum_{i=1}^{freq} |C_V(\nu^i,\beta) - C_V(\nu^i,\beta,N)|. \tag{14}$$

This provides an *a priori* estimate of the number of beads needed to achieve a given accuracy. Although the highest frequency mode is primarily responsible for limiting accuracy, correctly calculating E requires calculating all normal modes accurately, since they all contribute at all T. At low T the vibrational contribution to C_V becomes insignificant compared the rotational and translational contribution.

As an example, the expected error for MBPT2/6-311G($2df,p$) $\cdot$ CH$_2^+$ at a given T is presented in Figs. 2 and 3. For low T, C_V requires a larger value of N than QTE. For high T, E still must capture the residual ZPE, but C_V only has to capture the change in QTE, therefore E becomes more difficult than C_V. For low T the accuracy of C_V initially decreases with N and then improves. This behavior is a result of the bead degrees of freedom initially increasing C_V, followed by a stiffening of the interbead potential as N increases. It should be noted that this analysis provides only a guide for determining the size of N needed. Steep repulsive walls will most likely require a larger value of N than the harmonic system. For this reason, in this work we conducted a convergence study with respect to $N=1,2,4,8,...$ (not in-
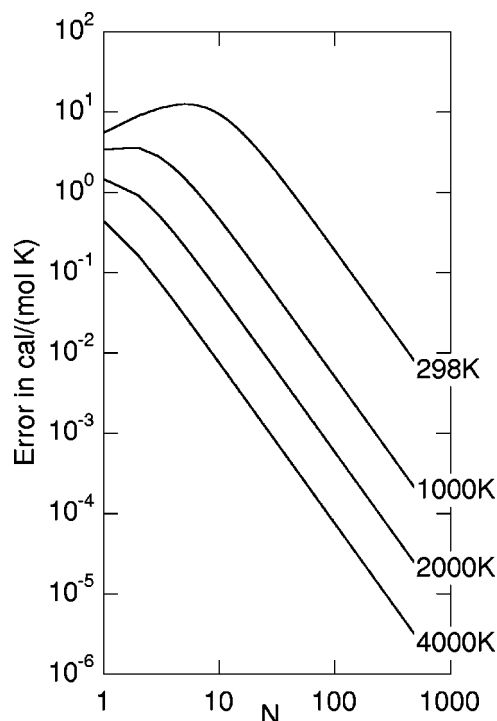
cluded for brevity) and then verified the reasonableness of our choice of N with the harmonic system. The harmonic oscillator estimate of the required N was found to be accurate for the molecules studied here

For the harmonic oscillator, one can gain *a priori* insight into which energy estimator will perform better based upon a similar analysis. The numerical noise associated with each estimator depends upon ν, β, and Monte Carlo walker used (normal modes, Fourier, primitive, molecular dynamics).[30,32,33,36,42] Because with the normal mode sampling method the different estimators have similar autocorrelation times, we choose to ignore autocorrelation within this investigation. For the harmonic oscillator the deviations of the thermodynamic and virial estimators for the internal energy are[30,33]

$$\sigma_{V,i}^2 = \frac{C_V(\nu^i,\beta,N)}{\beta^2} + \frac{3}{4\beta}E(\nu^i,\beta,N), \tag{15}$$

$$\sigma_{T,i}^2 = \sigma_V^2 + \frac{N}{2\beta^2} - \frac{7}{4\beta}E(\nu^i,\beta,N), \tag{16}$$

where the thermodynamic and virial estimators for the kinetic energy are

$$K_T = \frac{N}{2\beta} - \frac{Nm}{2\hbar^2\beta^2}\sum_{j=1}^{N}(r_j - r_{j+1})^2, \tag{17}$$

$$K_V = \frac{1}{2N}\sum_{j=1}^{N} r_j V'(r_j). \tag{18}$$

The more quantum mechanical the system is (low T, high β, large ν), the better behaved ϵ_T is compared to ϵ_V, but the

more quantum systems require large values of N, which favors ϵ_V. It is important to include the zero frequency modes (rotations and translations) for which each of these degrees of freedom has a variance of

$$\sigma_V^2 = 0, \tag{19}$$

$$\sigma_T^2 = \frac{N}{2\beta^2}. \tag{20}$$

The total variance would then be the sum of all variances. This analysis ignores the additional cost associated with calculating V' vs calculating V, thus this merely provides an initial guide. We should note that the centroid estimators would have a lower variance than the noncentroid analogs.[22,23]

V. RESULTS

We apply PIMC to a variety of molecules and compare to the harmonic approximation. The number of Monte Carlo cycles and beads were chosen to converge the energy to within at least 0.01 kcal/mol. Convergence was verified by using the standard blocking technique. This method involves comparing the average of the entire calculation to the average of blocks of data. All calculations were run for at least 10^8 Monte Carlo cycles. The error introduced by the use of the cache is found to be less than 0.01 kcal/mol based upon the internal error checking within the cache code. The cache spacing was set to 0.04 Å for bond distances and 0.08 rad for angles. The fraction of occupied grid points was constantly <50% and decreased with increased T. Approximately 7000 points were needed for a calculation at 300 K.

In the present work, we limit our study to ABA triatomics which can be surprisingly anharmonic.[63] The anharmonicity would be greater in ABC triatomics due to the loss of symmetry and tetra-atomics due to the torsional degree of freedom. With additional improvements to the cache, the method should be extendable to four and five atom systems. Although the goal of path integral thermochemistry is the calculation of E, S, H, G, and C_V, in the current work we concentrate on E and H. Although multiplying calculated frequencies by an empirical constant is common (and used in G-Y), we use unscaled frequencies for comparison, unless otherwise noted. Scaling factors are generally used to correct for the anharmonicity, for using an imperfect basis set, and for imperfect theory. The scaling factors vary greatly depending upon the level of theory, the basis set used, the temperature, and what property is being fit.[2,64] For Hartree–Fock (HF) theory and 3-21G basis set (abbreviated HF/3-21G) the scaling factors range from 0.9085 to 1.0075, depending on whether one is fitting ω, $1/\omega$, QTE(T), $\Delta H_{vib}(T)$, or $S_{vib}(T)$.[2] All molecules studied, except for CH_2^+, are members of the G3 test set. The G3 test set is the set of molecules that G3 theory was calibrated against, because reliable experimental data was available.

Several molecules were studied with V calculated with unrestricted second order many body perturbation theory (UMBPT2) (Refs. 65 and 66) and the 6-311G(d,p) basis set which is part of the family of standard traditional style basis

TABLE I. Harmonic and PIMC calculated UMBPT2/6-311G(d,p) QTE (kcal/mol). The ACE is the difference between the PIMC and harmonic energies.

	T	Harmonic	PIMC	ACE
triplet CH_2	298	12.9	11.8	1.1
	1000	18.0	17.2	0.8
	2000	27.5	27.0	0.5
doublet NH_2	298	14.0	12.3	1.7
	1000	18.8	17.1	1.7
	2000	28.1	26.8	1.3
singlet OH_2	298	15.5	12.9	2.5
	1000	20.2	17.7	2.5
	2000	29.1	26.8	2.3

sets.[67–72] Our PIMC code[73] was interfaced to the GAMESS code[74,75] for the calculation of all necessary *ab initio* V. An unrestricted reference wavefunction was used to correspond to the G-Y methods, to avoid the ambiguity that exists among the multitude of RO-MBPT2 methods,[66,76,77] and to minimize the possibility of a divergence in the perturbation expansion.[78] Basis set superposition error (BSSE) (Ref. 79) was not taken into account via the counterpoise correction because of its unreliability in calculating molecular hypersurfaces.[80] All comparisons were made to single point frequency calculations[81–83] using either GAUSSIAN 98 (Ref. 84) or GAMESS. We have chosen to use a basis set and level of theory that is higher than that used in the G-Y methods, to reduce ambiguity about convergence of level of theory. The anharmonic contribution to the energy (ACE=E_{harm} −E_{PIMC}) for triplet CH_2, doublet NH_2, and singlet OH_2 are presented in Table I. This can be as large as 2.5 kcal/mol. The harmonic and PIMC results differ by more than 1 kcal/mol at 298 K for all three molecules.

The doublet ion radical $\cdot CH_2^+$ was also studied. *Ab initio* PIMC methods have not been applied previously to radicals or ions.[16–21] Therefore, we studied it with several basis sets and levels of theory. In the density functional theory (DFT) (Ref. 85) calculations we used the BLYP (Refs. 86–88) and B3LYP (VWN5 version) (Ref. 89) density functionals. BLYP is used to maximize DFT dominance of the calculation and to minimize empirical parameters.[90] We used FullCI(FC) which is a full configuration interaction (CI) with all orbitals active except for the frozen cores(FC).[91,92] This was done to optimally treat electron correlation. Results are presented in Table II. The harmonic frequencies are converged with the small 6-31G(d) basis set, varying little with basis set. Conversely, the anharmonicity grows with basis set size. As the basis set increases, errors due to basis set incompleteness are reduced. For path integral calculations we require a basis set that is flexible enough to properly represent not just the equilibrium geometry, but also the thermally accessible portions of the PES.

The FullCI(FC) harmonic and PIMC energies agree. This does not necessarily imply that the FullCI(FC) PES is more harmonic than the MBPT2, since the steeper inner wall can partially cancel the effects of the shallow outer potential wall. For B3LYP, the ACE changes sign with T. For FullCI(FC), the absolute value of the ACE increases with

J. Chem. Phys., Vol. 118, No. 4, 22 January 2003

Molecular thermochemistry   1601

TABLE II. Harmonic and PIMC QTE (kcal/mol) for $\cdot CH_2^+$. The ACE is the difference between the PIMC and harmonic energies.

V	T	Harmonic	PIMC	ACE
UMBPT2/6-31G(d)	298	12.6	12.6	−0.03
UMBPT2/6-311G(d,p)	298	12.4	11.5	0.9
UMBPT2/6-31G($2df,p$)	298	12.4	10.4	2.0
BLYP/6-311G(d,p)	298	11.7	11.3	0.4
B3LYP/6-311G(d,p)	298	12.0	11.8	0.2
FullCI(FC)/6-31G(d)	298	12.4	12.3	0.04
UMBPT2/6-31G(d)	1000	17.8	17.8	−0.03
UMBPT2/6-311G(d,p)	1000	17.6	16.4	1.2
UMBPT2/6-31G($2df,p$)	1000	17.7	14.8	1.9
BLYP/6-311G(d,p)	1000	17.1	16.6	0.5
B3LYP/6-311G(d,p)	1000	17.3	17.0	0.3
FullCI(FC)/6-31G(d)	1000	17.6	17.6	−0.03
UMBPT2/6-31G(d)	2000	27.4	27.6	−0.2
UMBPT2/6-311G(d,p)	2000	27.3	26.5	0.8
UMBPT2/6-31G($2df,p$)	2000	27.3	25.4	1.9
BLYP/6-311G(d,p)	2000	26.9	26.6	0.3
B3LYP/6-311G(d,p)	2000	27.0	27.2	−0.1
FullCI(FC)/6-31G(d)	2000	27.2	27.5	−0.2

TABLE III. Harmonic and PIMC QTE (kcal/mol) for UMBPT2 triplet CH_2 at 298 K. The ACE is the difference between the PIMC and harmonic energies.

	Harmonic	PIMC	ACE
STO-3G	14.0	14.2	−0.2
3-21G	12.9	13.1	−0.2
6-31G	13.0	13.2	−0.3
6-31G(d)	13.1	13.2	−0.1
6-31G(d,p)	13.2	13.3	−0.1
6-311G(d,p)	12.9	11.8	1.1
6-311G($2df,p$)	12.9	12.2	0.7
6-311++G($2df,p$)	12.9	12.2	0.7
cc-pVDZ	12.9	13.0	−0.1
cc-pVTZ	12.9	12.9	0.05
cc-pVQZ	12.9	12.9	−0.01
aug-cc-pVDZ	12.9	12.9	−0.1
aug-cc-pVTZ	12.9	13.1	−0.2
aug-cc-pVQZ	12.9	13.0	−0.1

increased T. The FullCI(FC) results are within 0.2 kcal/mol of the UMBPT2/6-31G(d) calculations.

We have also explored the effect of using different energy estimators in these calculations. Results were found to be independent of whether the thermodynamic or centroid thermodynamic estimator was used.[22] For the FullCI(FC) calculations, the computational speedup from using the ϵ_{CT} estimation instead of ϵ_T is a factor of 0 for 298 K, 1.8 for 1000 K, and 17 for 2000 K. The use of the cache provides a more dramatic speedup. For the 2000 K, $N=64$ calculation, the PIMC code takes 10 days with the cache and over 3600 years without. This speedup of over 6 orders of magnitude does not take into account the additional speedup obtained because the lower temperatures are effectively free with the cache code, once the 2000 K run is done.

The convergence with respect to basis set was investigated using the uncharged triplet CH_2 at the UMBPT2 level of theory at 298 K. In addition to traditional style basis sets, we also used the correlation consistent basis sets.[93–99] For computational efficiency, the rotated versions are used, which give the same answers as the unrotated versions.[100] These basis sets were designed to converge consistently to the correct answer as the basis set size is increased. They are designed with the intent of properly treating correlated methods, like MBPT and CI. They provide a straightforward, systematic way to improve the basis set, but they do suffer from a very rapid growth in size as the ζ value increases. Unlike the traditional basis sets which use Cartesian Gaussians, the correlation consistent basis sets use spherical harmonics for $d,f,g,\ldots$ functions. Results are presented in Table III. The harmonic approximation converges quickly to an apparent large basis set limit of 12.9 kcal/mol. The PIMC results are more basis set dependent, indicating that the latent anharmonicity is difficult to converge with respect to basis set. The correlation consistent basis sets appear to offer better convergence in PIMC calculations than the traditional basis sets. The PIMC energies are found to be converged within 0.2 kcal/mol at the triple ζ level. The use of augmented basis

changes the energy by <0.2 kcal/mol. The cache provided a speedup of approximately 8 orders of magnitude for the aug-cc-pVQZ basis set calculations.

We studied singlet CH_2, triplet CH_2, doublet NH_2, singlet O_3, singlet OH_2, and singlet SO_2 at 298 K with UHF/6-31G(d), because this is the level of theory and basis set used in G2 and G3 for calculation of the QTE. Both scaled frequencies (by traditional 0.8929) and unscaled frequency results are presented. The results are presented in Table IV. The SO_2 results show that even for species with heavier (less quantum mechanical) atoms there is a significant difference between PIMC and the harmonic approximation. In all cases the harmonic approximation underestimates the thermal/quantum energy and the frequency scaling only makes the agreement worse. Clearly, the scaling is attempting to correct for an inadequate level of theory (no electron correlation), and a small basis set—not the lack of anharmonicity.

For H_2O PIMC results are compared with atomization energies calculated based on JANAF tables.[101] The MBPT2 method and a 6-311G(d,p) basis set are used to calculate the thermal contribution to the energy. The equilibrium E for H_2O, H, and O is calculated at the CCSD(T)/aug-cc-pVQZ level. Harmonic approximation results for the thermal contribution are also presented for comparison. Results are presented in Fig. 4. The PIMC results match experiment better than the harmonic approximation. For $T>300$ K the harmonic results more nearly match the JANAF results as a result of the curve being the wrong shape.

TABLE IV. Scaled harmonic approximation, harmonic approximation, and PIMC calculated QTE (kcal/mol) for UHF/6-31G(d) at 298 K.

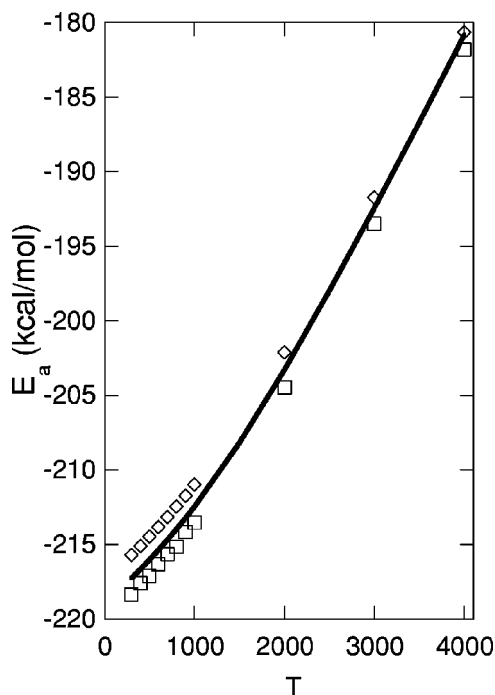|	Harmonic	Scaled	PIMC
singlet CH_2	13.1	11.8	13.4
triplet CH_2	13.3	12.1	13.6
NH_2	14.7	13.3	14.9
O_3	7.3	6.7	8.1
OH_2	16.1	14.7	16.6
SO_2	6.9	6.4	7.6

FIG. 4. Atomization energies for H_2O with MBPT2/6-311G(d,p). $\diamond$ is harmonic, $\square$ is PIMC, and the line is JANAF.

Calculations presented in this paper are computationally affordable: all calculations in this paper were run serially and all but the aug-cc-pVTZ basis set (and larger) calculations took less than a week per energy. Further work is currently being undertaken to determine the level of theory and basis set needed to fully converge the path integral thermochemistry calculation, although previous research into accurately treating the PES is a valuable guide.[102]

VI. CONCLUSIONS

We have calculated quantum anharmonic corrections to the energy of *ab initio* electronic structure calculations using PIMC. The difference between PIMC and the traditional harmonic normal mode approximation is not always the same sign or magnitude, therefore there is clearly no simple scaling relationship that can be consistently exploited. Anharmonicity is found to be more than 2 kcal/mol for some molecules. The effects of basis set size and electron correlation must be carefully investigated. The exploratory calculations presented here suggest that the cc-pVTZ basis set performs well in PIMC thermochemical calculations. More investigation into the level of electron correlation required in a PIMC calculation is needed. Because of the analytical nature of the harmonic oscillator, one can know *a priori* how large the number of beads N must be based upon the molecules normal modes. It is also possible to gain insight into which estimator is best for a particular molecule and temperature by comparing results for the harmonic oscillator. A novel cache code is utilized to overcome the numerical difficulties of repeated potential energy evaluations. Finally, the reader is cautioned that the unaccounted for anharmonicity provides a limit on the obtainable accuracy of methods (such as Gaussian-3) that ignore anharmonicity.

[1] W. D. Gwinn, J. Chem. Phys. **55**, 477 (1971).
[2] A. P. Scott and L. Radom, J. Phys. Chem. **100**, 16502 (1996).
[3] W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory* (Wiley, New York, 1986).
[4] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, J. Chem. Phys. **90**, 5622 (1989).
[5] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94**, 7221 (1991).
[6] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, J. Chem. Phys. **109**, 7764 (1998).
[7] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **112**, 7374 (2000).
[8] L. A. Curtiss, K. Raghavachari, and J. A. Pople, J. Chem. Phys. **98**, 1293 (1993).
[9] L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A Pople, Chem. Phys. Lett. **313**, 600 (1999).
[10] L. A. Curtiss, P. C. Redfern, B. J. Smith, and L. Radom, J. Chem. Phys. **104**, 5148 (1996).
[11] L. A. Curtiss, K. Raghavachari, P. C. Redfern, A. G. Baboul, and J. A. Pople, Chem. Phys. Lett. **314**, 101 (1999).
[12] L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, J. Chem. Phys. **114**, 108 (2001).
[13] T. I. Solling, D. M. Smith, L. Radom, M. A. Freitag, and M. S. Gordon, J. Chem. Phys. **115**, 8758 (2001).
[14] S. Parthiban and J. M. L. Martin, J. Chem. Phys. **114**, 6014 (2001).
[15] J. M. L. Martin and G. de Oliveira, J. Chem. Phys. **111**, 1843 (1999).
[16] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[17] L. S. Schulman, *Techniques and Applications of Path Integrals* (Wiley, New York, 1986).
[18] R. P. Feynman, *Statistical Mechanics* (Benjamin, New York, 1972).
[19] M. Shiga, M. Tachikawa, and S. Miuri, J. Chem. Phys. **115**, 9149 (2001).
[20] D. M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).
[21] C. Chakravarty, Int. Rev. Phys. Chem. **16**, 421 (1997).
[22] K. R. Glaesemann and L. E. Fried, J. Chem. Phys. **116**, 5951 (2002).
[23] K. R. Glaesemann and L. E. Fried, J. Chem. Phys. **117**, 3020 (2002).
[24] S. L. Mielke and D. G. Truhlar, J. Chem. Phys. **115**, 652 (2001).
[25] R. Q. Topper and D. G. Truhlar, J. Chem. Phys. **97**, 3647 (1992).
[26] R. Q. Topper, G. J. Tawa, and D. G. Truhlar, J. Chem. Phys. **97**, 3668 (1992).
[27] R. Q. Topper, G. J. Tawa, and D. G. Truhlar, J. Chem. Phys. **113**, 3930 (2000).
[28] H. F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959).
[29] D. Marx, P. Nielaba, and K. Binder, Int. J. Mod. Phys. C **3**, 337 (1992).
[30] M. F. Herman, E. J. Bruskin, and B. J. Berne, J. Chem. Phys. **76**, 5150 (1982).
[31] M. Creutz and B. Freedman, Ann. Phys. (N.Y.) **132**, 427 (1981).
[32] P. A. Fernandes, A. P. Carvalho, and J. P. P. Ramalho, J. Chem. Phys. **103**, 5720 (1995).
[33] W. Janke and T. Sauer, J. Chem. Phys. **107**, 5821 (1997).
[34] M. Kolář and S. F. O'Shea, J. Phys. A **29**, 3471 (1996).
[35] E. L. Pollock and D. M. Ceperley, Phys. Rev. B **30**, 2555 (1984).
[36] J. Cao and B. J. Berne, J. Chem. Phys. **91**, 6359 (1989).
[37] C. Chakravarty, M. C. Gordillo, and D. M. Ceperley, J. Chem. Phys. **109**, 2123 (1998).
[38] J. D. Doll and D. L. Freeman, J. Chem. Phys. **111**, 7685 (1999).
[39] C. Chakravarty, M. C. Gordillo, and D. M. Ceperley, J. Chem. Phys. **111**, 7687 (1999).
[40] D. L. Freeman and J. D. Doll, J. Chem. Phys. **80**, 5709 (1984).
[41] J. A. Barker, J. Chem. Phys. **70**, 2914 (1979).
[42] A. Giansanti and G. Jacucci, J. Chem. Phys. **89**, 7454 (1988).
[43] M. Pavese, D. R. Berard, and G. A. Voth, Chem. Phys. Lett. **3000**, 93 (1999).
[44] S. Jang and G. A. Voth, J. Chem. Phys. **114**, 1944 (2001).

[45] E. Geva, Q. Shi, and G. A. Voth, J. Chem. Phys. **115**, 9209 (2001).
[46] D. R. Reichman, P. N. Roy, S. Jang, and G. A. Voth, J. Chem. Phys. **113**, 919 (2000).
[47] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[48] M. E. Tuckerman, B. J. Berne, G. J. Martyna, and M. L. Klein, J. Chem. Phys. **99**, 2796 (1993).
[49] M. Matsumoto and T. Nishimura, ACM Trans. Model. Comput. Simul. **8**, 3 (1998).
[50] M. Matsumoto and T. Nishimura, 1/26/2002 version.
[51] J. M. L. Martin and P. R. Taylor, Mol. Phys. **96**, 681 (1999).
[52] G. M. Chaban, J. O. Jung, and R. B. Gerber, J. Chem. Phys. **111**, 1823 (1999).
[53] N. J. Wright, R. B. Gerber, and D. J. Tozer, Chem. Phys. Lett. **324**, 206 (2000).
[54] S. Parthiban and J. M. L. Martin, J. Chem. Phys. **115**, 2051 (2001).
[55] J. M. L. Martin, Chem. Phys. Lett. **310**, 271 (1999).
[56] K. C. Thompson, M. J. T. Jordan, and M. A. Collins, J. Chem. Phys. **108**, 8302 (1998).
[57] B. J. Sung and M. S. Kim, J. Chem. Phys. **113**, 3098 (2000).
[58] M. H. Yang, D. H. Zhang, M. A. Collins, and S.-Y. Lee, J. Chem. Phys. **115**, 174 (2001).
[59] T. Ishida and G. C. Schatz, Chem. Phys. Lett. **314**, 369 (1999).
[60] K. J. Runge and G. V. Chester, Phys. Rev. B **38**, 135 (1988).
[61] M. A. Miller, L. M. Amon, and W. P. Reinhardt, Chem. Phys. Lett. **331**, 278 (2000).
[62] K. S. Schweizer, R. M. Stratt, D. Chandler, and P. G. Wolynes, J. Chem. Phys. **75**, 1347 (1981).
[63] M. Iwai and R. D. Levine, Phys. Rev. A **42**, 3991 (1990).
[64] R. S. Grev, C. L. Janssen, and H. F. Schaefer, III, J. Chem. Phys. **95**, 5128 (1991).
[65] J. A. Pople and R. K. Nesbet, J. Chem. Phys. **22**, 571 (1954).
[66] T. D. Crawford, H. F. Schaefer, and T. J. Lee, J. Chem. Phys. **105**, 1060 (1996).
[67] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).
[68] W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).
[69] P. C. Hariharan and J. A. Pople, Theor. Chim. Acta **28**, 213 (1973).
[70] M. M. Francl *et al.*, J. Chem. Phys. **77**, 3654 (1982).
[71] M. J. Frisch, J. A. Pople, and J. S. Binkley, J. Chem. Phys. **80**, 3265 (1984).
[72] W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **51**, 2657 (1969).
[73] K. R. Glaesemann and L. E. Fried (unpublished).
[74] M. W. Schmidt, K. K. Baldridge, J. A. Boatz *et al.*, J. Comput. Chem. **14**, 1347 (1993).

[75] GAMESS *ab inito* code, 2001 Version, written by M. W. Schmidt, M. S. Gordon, K. R. Glaesemann *et al.*
[76] G. D. Fletcher, M. S. Gordon, and R. S. Bell, Theor. Chem. Acc. **107**, 57 (2002).
[77] C. W. Murray and N. C. Handy, J. Chem. Phys. **97**, 6509 (1992).
[78] M. S. Gordon, M. W. Schmidt, G. M. Chaban, K. R. Glaesemann, W. J. Stevens, and C. Gonzalez, J. Chem. Phys. **110**, 4199 (1999).
[79] E. R. Davidson and S. J. Chakravorty, Chem. Phys. Lett. **217**, 48 (1994).
[80] K. R. Liedl, J. Chem. Phys. **108**, 3199 (1998).
[81] Y. Yamaguchi, Y. Osamura, J. D. Goddard, and H. F. Schaefer, III, *A New Dimension to Quantum Chemistry: Analytical Derivative Methods in Ab Initio* Molecular Electronic Structure Theory (Oxford University Press, New York, 1994).
[82] A. Komornicki and G. Fitzgerald, J. Chem. Phys. **98**, 1398 (1993).
[83] B. G. Johnson and M. J. Fisch, J. Chem. Phys. **100**, 7429 (1994).
[84] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 98, Revision A.4, Gaussian, Inc., Pittsburgh, PA, 1998.
[85] B. J. Lynch and D. G. Truhlar, J. Phys. Chem. A **105**, 2936 (2001).
[86] A. D. Becke, Phys. Rev. A **38**, 3098 (1998).
[87] B. Miehlich, A. Savin, H. Stoll, and H. Preuss, Chem. Phys. Lett. **157**, 200 (1989).
[88] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[89] R. H. Hertwig and W. Koch, Chem. Phys. Lett. **268**, 345 (1997).
[90] P. M. W. Gill, Aust. J. Chem. **54**, 661 (2001).
[91] B. R. Brooks and H. F. Schaefer, J. Chem. Phys. **70**, 5092 (1979).
[92] P. J. Knowles and N. C. Handy, Chem. Phys. Lett. **111**, 315 (1984).
[93] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[94] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **98**, 1358 (1993).
[95] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **100**, 2975 (1994).
[96] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **103**, 4572 (1995).
[97] A. K. Wilson, T. v. Mourik, and T. H. Dunning, Jr., J. Mol. Struct.: THEOCHEM **388**, 339 (1997).
[98] A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning, Jr., J. Chem. Phys. **110**, 7667 (1999).
[99] Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 6/28/02, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory which is part of the Pacific Northwest Laboratory, P.O. Box 999, Richland, Washington 99352, USA, and funded by the U.S. Department of Energy. The Pacific Northwest Laboratory is a multi-program laboratory operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DE-AC06-76RLO 1830.
[100] E. R. Davidson, Chem. Phys. Lett. **260**, 514 (1996).
[101] *NIST-JANAF Thermochemical Tables,* 4th ed., edited by M. W. Chase, Jr. (American Institute of Physics, New York, 1998).
[102] T. H. Dunning, Jr., J. Phys. Chem. A **104**, 9062 (2000).