Eur. Phys. J. D **10**, 353–362 (2000)

THE EUROPEAN PHYSICAL JOURNAL D

EDP Sciences

Molecular structure calculations via path integral simulations: Estimating finite-discretization errors

L. Knoll[1] and D. Marx[2,a]

[1] Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, 69029 Heidelberg, Germany
[2] Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany

Received 13 September 1999 and Received in final form 18 November 1999

Abstract. Path integral simulations are now recognized as a useful tool to determine theoretically the structure of complex molecules at finite temperatures including quantum effects. In addition to statistical errors due to incomplete sampling, also systematic errors are inherent in this procedure because of the finite discretization of the path integral. Here, useful "back of the envelope" estimates to assess the systematic errors of bond-length distribution functions are introduced. These analytical estimates are tested for two small molecules, HD^+ and H_3^+, where quasi-exact benchmark data are available. The accuracy of the formulae is shown to be sufficient in order to allow for a reliable assessment of the quality of the discretization in a given simulation. The estimates will also be applicable in condensed phase path integral simulations, and the basic idea can be generalized to other observables than those presented.

PACS. 31.15.Kb Path-integral methods – 02.70.Lq Monte Carlo and statistical methods

1 Introduction and motivation

The notion of *"the structure"* of a molecule is ambiguous, even if one refers exclusively to isolated non-interacting molecules, *e.g.* in the gas phase, and this problem becomes more intricate in condensed phases. A host of different structural concepts, as determined from either experiment or calculations, can be defined. Here, we refer the reader to reference [1] (in particular to Tab. 1.3.3) or to reference [2] for overviews and precise definitions. Probably the most widespread theoretical structure concept is that of the so-called "equilibrium structure" r_e as determined from the global minimum of the electronic ground state potential energy hyper-surface. It is this structure that is typically reported as a result of quantum-chemical geometry optimization calculations. However, this is an inherently classical concept in the sense that it refers to a structure that *classical* nuclei would assume at $T = 0$ K, *i.e.* in the absence of any thermal excitations. There is, of course, no direct experimental access to this quantity since the nuclei perform quantum-mechanical zero-point motion even at $T = 0$ K. More realistic structural indicators are the "average structure" $\langle r \rangle$ (obtained from suitably averaging distance or position operators using the many-body nuclear wave function, most often in the quantum-mechanical rotational-vibrational ground state) or the "most probable structure" r_P (defined similarly, but based on the absolute maximum of the nuclear density) [2]. Further complications arise if excitations have to be taken into account in thermal equilibrium at $T > 0$ K, or even more so in situations where different modes have different effective "temperatures" depending on their coupling and thermalization rate.

Experimentally it is, among others, the so-called "r_0 structure" which is traditionally obtained from converting rotational constants (spectroscopically obtained in the ro-vibrational ground state) into distances between nuclei, which are again assumed to behave like point-particles. More recently, dramatic progress in determining more directly the real-space structure of molecules was made in the framework of Coulomb Explosion Imaging (CEI) [3]. Such CEI experiments can yield information about the diagonal part of the fully correlated many-body nuclear density matrix of small molecules, and thus they provide among others access to the average $\langle r \rangle$ or the most probable r_P structures and even to the width of the distribution functions along certain coordinates.

The traditional route towards determining ro-vibrationally averaged structures (such as average or most probable structures) is to calculate globally the electronic ground state Born-Oppenheimer potential-energy hyper-surface and to solve for the nuclear wave functions in terms of eigenvectors and eigenvalues using this external potential. Although very successful and accurate in the limit of small molecules (see for instance the series of benchmark calculations [4,5] for H_3^+) this philosophy becomes soon intractable for systems with many degrees of freedom. An interesting alternative for large systems is the approximate vibrational self-consistent field approach where the

[a] e-mail: dominik.marx@theochem.ruhr-uni-bochum.de

interactions are obtained "on the fly" from MP2 electronic structure calculations [6]. Similarly, the diffusion Monte Carlo technique [7,8] was combined with "on the fly" MP2 calculations [9]. These approaches avoid the bottleneck to construct *a priori* a high-dimensional global potential energy surface.

Another philosophy, the *ab initio* path integral method [10], was advocated lately as a technique to tackle complex many-body quantum problems at finite temperatures. Traditionally, path integral simulations (see Refs. [11–15] for introductions and reviews) are applied to systems in the condensed phase or to clusters using fairly simple interaction potentials. This is no more necessary in the case of the *ab initio* path integral technique [10,16–19] since the interactions are calculated from "on the fly" electronic structure calculations in analogy to conventional Car-Parrinello simulations [20]. This renders structural investigations of molecules feasible, *e.g.* CH_5^+ and $C_2H_3^+$ molecules [21,22], intra-molecular proton transfer in malonaldehyde [23], protonated water and hydrogen clusters [24,25], as well as lithium clusters [26] were analyzed in detail.

A caveat of such *ab initio* path integral simulations [10,16–19] where the interactions are determined from "on the fly" electronic structure calculations is, of course, that it is numerically more expensive than analogous traditional simulations using model potentials [11–15]. Therefore, it would be highly desirable to estimate in advance the expected systematic error for a certain discretized Trotter approximation to the continuous path integral without necessarily taking recourse to straightforward but expensive convergence studies. This would open the way to predetermine the discretization based on a given error criterion without actually doing several converged calculations with varying Trotter number. Alternatively, one could estimate the accuracy of a given maximum discretization used in a particular calculation.

In this paper we present an analytical method to estimate the effect of a finite Trotter discretization on structural properties of small molecules based on the harmonic approximation of the mode(s) under consideration. In order to check the method and to be able to compare with known (quasi-exact) benchmark results we use published potential energy surfaces for diatomic and triatomic molecules, HD^+ and H_3^+. However, the same estimates can be applied in more conventional condensed-phase path integral simulations. For simplicity, we perform the calculations with the Monte Carlo variant of canonical path integral simulations rather than with Nosé-Hoover-chain [27] thermostated molecular dynamics which underlies our *ab initio* path integral technique [10,16–19].

2 Molecular structure and path integral simulations

Within the Born-Oppenheimer approximation the Hamiltonian for a set of N nuclei with masses M_i that move on the electronic ground state potential energy surface V reads

$$\hat{H} = \sum_{i=1}^{N} \frac{\hat{\mathbf{p}}_i^2}{2M_i} + V(\{\hat{\mathbf{r}}_i\}), \tag{2.1}$$

where $\mathbf{r}_i$ and $\mathbf{p}_i$ are the positions and momenta of the nuclei; atomic units (a.u.) are used throughout this paper. Given all eigenvectors $\{\Phi_k\}$ and the corresponding eigenvalues $\{E_k\}$ (where k denotes the multi-index required to completely specify the quantum state) in the respective electronic potential $V(\{\mathbf{r}_i\})$, the ground state nuclear structure of the molecule consisting of N nuclei is completely determined by the N-body ground state density

$$P(\{\mathbf{r}_i\}) = \langle \{\mathbf{r}_i\} | \Phi_0 \rangle \langle \Phi_0 | \{\mathbf{r}_i\} \rangle = |\Phi_0(\{\mathbf{r}_i\})|^2, \tag{2.2}$$

which is the joint-probability to find in real space atom 1 at $\mathbf{r}_1$ and atom 2 at $\mathbf{r}_2$... and atom N at $\mathbf{r}_N$. Excitations at $T > 0$ K can be included in a canonical sense and lead to the N-body thermal density

$$P(\{\mathbf{r}_i\}, \beta) = \frac{1}{\sum_k \exp[-\beta E_k]} \times \sum_k \langle \{\mathbf{r}_i\} | \Phi_k \rangle \langle \Phi_k | \{\mathbf{r}_i\} \rangle \exp[-\beta E_k] \tag{2.3}$$

by summing over all states, weighting the different states with the Boltzmann factor, and normalizing with the partition function; $\beta = 1/k_B T$. The conventional approach consists in calculating the wave functions $\{\Phi_k(\{\mathbf{r}_i\})\}$, energy levels $\{E_k\}$ and using (2.3) to compute the thermal density and thus ro-vibrationally averaged structures.

In order to relate this to quantities that we obtain in the path integral approach we start with the partition function

$$Z = \mathrm{Tr}\ \exp\left[-\beta \hat{H}\right] \tag{2.4}$$

for the canonical equilibrium ensemble. In its well-known path integral representation this transforms to [28,29]

$$Z = \oint \mathcal{D}\{\mathbf{r}_i\}\, \mathrm{e}^{-\mathcal{A}([\mathbf{r}_i])}, \tag{2.5}$$

where the path integration is carried out over all closed paths weighted with the classical Euclidean action

$$\mathcal{A}([\mathbf{r}_i]) = \int_0^\beta \mathrm{d}\tau \left\{ \left(\sum_{i=1}^{N} \frac{1}{2} M_i \left(\dot{\mathbf{r}}_i(\tau)\right)^2 \right) + V(\{\mathbf{r}_i(\tau)\}) \right\} \tag{2.6}$$

of the particular path $[\mathbf{r}_i]$. Here, and in the following, we assume that the nuclei are distinguishable and thus we do not include their bosonic or fermionic permutation statistics. After discretization of the path integral in position space one obtains the following discrete representation

of the partition function as a multidimensional integral

$$Z_n = \mathcal{N}_n \int \mathrm{d}\{\mathbf{r}_i^{(1)}\} \cdots \int \mathrm{d}\{\mathbf{r}_i^{(n)}\} \times \exp\left[-\beta \sum_{j=1}^{n} \left\{ \left(\sum_{i=1}^{N} \frac{nM_i}{2\beta^2} \left(\mathbf{r}_i^{(j)} - \mathbf{r}_i^{(j+1)}\right)^2 \right) + \frac{1}{n} V\left(\{\mathbf{r}_i^{(j)}\}\right) \right\}\right] \quad (2.7)$$

with a known normalization factor $\mathcal{N}_n$ that depends on temperature and masses. In addition, all paths have to be closed, *i.e.*, periodic boundary conditions in imaginary time $\mathbf{r}_i^{(n+1)} = \mathbf{r}_i^{(1)}$ are required to satisfy the trace property (2.5). The exact quantum partition function (2.5) is obtained in the limit $Z = \lim_{n\to\infty} Z_n$. The corresponding discretized many-body density takes the form

$$P_n(\{\mathbf{r}_i\}, \beta) = \frac{1}{Z_n} \int \mathrm{d}\{\mathbf{r}_i^{(2)}\} \cdots \int \mathrm{d}\{\mathbf{r}_i^{(n)}\} \times \exp\left[-\beta \sum_{j=1}^{n} \left\{ \left(\sum_{i=1}^{N} \frac{nM_i}{2\beta^2} \left(\mathbf{r}_i^{(j)} - \mathbf{r}_i^{(j+1)}\right)^2 \right) + \frac{1}{n} V\left(\{\mathbf{r}_i^{(j)}\}\right) \right\}\right], \quad (2.8)$$

and similarly the exact path integral representation of (2.3) is obtained as $P(\{\mathbf{r}_i\}, \beta) = \lim_{n\to\infty} P_n(\{\mathbf{r}_i\}, \beta)$. Expectation values in the canonical ensemble of operators $\hat{A}$ for a given Trotter number are denoted as $\langle \hat{A} \rangle_n$, and $\langle \hat{A} \rangle_\infty$ or $\langle \hat{A} \rangle$ in short-hand notation corresponds to the exact value at infinite discretization.

In the discretized representation (2.7) or (2.8) the well-known polymer isomorphism [30] becomes evident: every particle i is represented by a closed necklace of n beads $(\mathbf{r}_i^{(1)}, \mathbf{r}_i^{(2)}, ..., \mathbf{r}_i^{(n)})$ in real-space, which is a discretized approximation to a full path $[\mathbf{r}_i]$ from $\mathbf{r}_i(0)$ to $\mathbf{r}_i(\beta)$. The potential V only interacts between beads with the same Trotter index j, and the kinetic part of the Hamiltonian (2.1) is transformed into a harmonic interaction between neighboring beads $\mathbf{r}_i^{(j-1)}$, $\mathbf{r}_i^{(j)}$, and $\mathbf{r}_i^{(j+1)}$ of the same particle i.

In the remainder of this section a few technical details of the present simulations are given. The path integral, while difficult to solve numerically, can be easily evaluated using Monte Carlo techniques [13]. In this approach the integral is sampled statistically, by moving the beads $\{\mathbf{r}_i^{(j)}\}$ according to a Metropolis algorithm [31]. Although this approach works well with the so-called primitive discretization (2.7) for small discretizations n, it leads to inefficient simulations for large n. The reason is a slowdown of the path movements through configuration space because of the strong harmonic forces acting between neighboring beads which leads to long correlation times. To overcome these problems, a host of different methods have been provided [13]. One route is to apply collective moves in order to update the paths by propagating them directly in Fourier space. Instead, we have chosen a different approach to non-local moves as provided by the transformation to the so-called staging coordinates [17,27].

As a measure of the improvements achieved by this staging transformation, the efficiency of different computational schemes can be computed as follows (see *e.g.* Chap. V.B. in Ref. [13]). Given an operator $\hat{A}$ with mean $\langle \hat{A} \rangle$ variance $\nu_A = \langle (\hat{A} - \langle \hat{A} \rangle)^2 \rangle$, one can define a correlation time

$$\kappa_A = 1 + 2 \sum_{k=1}^{N_{\mathrm{mc}}} \frac{\langle (\hat{A}_1 - \langle \hat{A} \rangle)(\hat{A}_k - \langle \hat{A} \rangle) \rangle}{\nu_A}, \quad (2.9)$$

where $\hat{A}_k$ is the operator evaluated at the kth step of the Markov chain; note that $\kappa_A \geq 1$. Taking into account the correlation in the generated Markov chain, the statistical error Σ_A associated with the expectation value of the operator $\hat{A}$ is then given by

$$\Sigma_A = \sqrt{\frac{\kappa_A \nu_A}{N_{\mathrm{mc}}}}, \quad (2.10)$$

where N_{mc} is the number of Monte Carlo steps performed in the simulation. The efficiency ξ_A of a simulation with respect to the expectation value of this particular operator is defined as

$$\xi_A = \frac{1}{\Sigma_A^2 N_{\mathrm{mc}} T_{\mathrm{CPU}}} = \frac{1}{\kappa_A \nu_A T_{\mathrm{CPU}}}, \quad (2.11)$$

where T_{CPU} is the computational time for each Monte Carlo step. Thus, ξ_A measures how quickly the *statistical* error of a given quantity decreases as a function of increasing length of the Markov chain or total computer time. By using the staging transformation, the efficiency of the computational scheme could be improved by two orders of magnitude in the limit of large Trotter numbers n, see Figure 1.

3 Estimation of systematic discretization errors

The Trotter discretized version (2.7) of the quantum partition function (2.5) is of course only exact in the limit of an infinite discretization $n \to \infty$. Thus, any finite representation which is necessary to carry out numerical simulations introduces *systematic* errors in addition to the statistical errors due to finite and thus incomplete numerical sampling. It is know that these finite-n errors have a leading order of $(\beta/n)^2$ for almost all physical observables [32]. The traditional approach to extrapolate expectation values $\langle \hat{A} \rangle_n$, obtained at a given finite discretization n, to the correct $n \to \infty$ limit is the so-called Trotter scaling

$$\langle \hat{A} \rangle_n = \langle \hat{A} \rangle_\infty + \frac{c_2}{n^2} + \frac{c_4}{n^4} + \ldots, \quad (3.1)$$

where $\langle \hat{A} \rangle_\infty$ is the extrapolated expectation value and c_2 and c_4 are constants for a given temperature. Thus, one

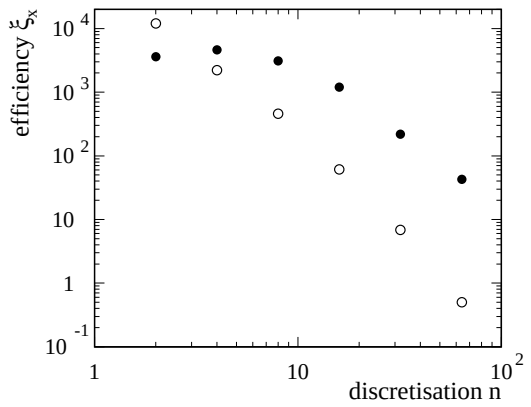

Fig. 1. Efficiency ξ_x for the expectation value of the position operator $\hat{x}$ using the primitive (○) and the staging (●) algorithm for test calculations of a one-dimensional harmonic oscillator (see (3.2) with $\beta = 10$, $\omega = 1$ and $M^\star = 1$) as a function of Trotter discretization n.

has to perform separate simulations using several different discretizations n in order to exploit (3.1).

As already motivated in the introduction, it is highly desirable to be able to estimate the accuracy of a single calculation with finite n for a given position-dependent observable $\hat{A}$ without actually knowing the converged answer. This is particularly important in cases where it is not possible to perform calculations with several discretizations n because of the rapidly increasing computational cost of the individual calculations. In such a case, one approach is to perform a calculation with the largest n that is affordable with the available computer resources and to estimate the accuracy based on an approximate expression. In the following we will introduce such error estimates in the framework of molecular structure determination *via* path integral simulations. Such estimates are particularly relevant if structures determined by path integral simulations are compared to measured real-space structures as provided by the CEI technique [3].

The systematic error in determining the width of inter-atomic distance distributions in molecules, which is introduced by using the discretized path integral expression (2.7), can be estimated easily if one assumes the potential to be harmonic near the minimum. This ansatz, although clearly not sufficient for actual structure determinations of very floppy molecules, might nevertheless be sufficient for quick and rough *error* assessments, see below. Based on this approximation one can trivially separate the different vibrational eigenmodes of the molecule and obtain $3N-6$ non-interacting harmonic oscillators and thus normal modes. For each one of them the Hamiltonian is of course of the form

$$\hat{H}_{\mathrm{HO}} = \frac{\hat{p}^2}{2M^\star} + \frac{1}{2}M^\star\omega^2\hat{x}^2, \tag{3.2}$$

where $M^\star$ is the reduced mass of the mode and x denotes the relative displacement from the minimum. The analytical expression of the density matrix of the harmonic oscillator Hamiltonian is known at a given discretization n (see *e.g.* Ref. [33]). Its diagonal part (which is related to inter-atomic distance distributions in molecules) is of the form

$$P_n(x,\beta) = \mathcal{N}_n \exp\left\{-M^\star\omega\left(1+\left(\frac{\omega\beta}{2n}\right)^2\right)^{1/2} \times \tanh\left(n\,\mathrm{arsinh}\left(\frac{\omega\beta}{2n}\right)\right)x^2\right\}, \tag{3.3}$$

where

$$\mathcal{N}_n = \left(\frac{M^\star\omega}{2\pi}\right)^{1/2}\left(\frac{1+\left(\frac{\omega\beta}{2n}\right)^2}{\sinh\left(2n\,\mathrm{arsinh}\left(\frac{\omega\beta}{2n}\right)\right)}\right)^{1/2} \tag{3.4}$$

is the factor required to normalize the distribution. For $n\to\infty$ one recovers

$$P_\infty(x,\beta) = \left(\frac{M^\star\omega}{2\pi\sinh(\omega\beta)}\right)^{1/2} \times \exp\left(-M^\star\omega\tanh\left(\frac{\omega\beta}{2}\right)x^2\right) \tag{3.5}$$

the exact quantum mechanical result.

The width $\sigma_n = \langle(\hat{x}-\langle\hat{x}\rangle_n)^2\rangle_n^{1/2}$ of the finite-n distribution can easily be obtained from (3.3) and leads to the following relative deviation from the exact width σ_∞

$$\frac{\sigma_n^2}{\sigma_\infty^2} = \frac{1}{\left(1+\left(\frac{\omega\beta}{2n}\right)^2\right)^{1/2}}\frac{\tanh\left(\frac{\omega\beta}{2}\right)}{\tanh\left(n\,\mathrm{arsinh}\left(\frac{\omega\beta}{2n}\right)\right)}; \tag{3.6}$$

note that the mass-dependency drops out at this stage. One can show that for $n \geq 1$ the second factor in (3.6) is bound between unity and 1.089, the maximum being reached for $n=1$, $\omega\beta/2 = 1.432$. For $n > 10$ the function is bound between unity and 1.0012 and approaches unity even closer for larger n. For discretizations $n > 10$ one can thus approximate the relation (3.6) further

$$\frac{\sigma_n}{\sigma_\infty} \approx \left(1+\left(\frac{\omega\beta}{2n}\right)^2\right)^{-1/4} \tag{3.7}$$

and obtains an estimate which does only depend on the vibrational frequency ω of the particular normal mode and the ratio β/n. This equation can be simplified even more in the limit $\omega\beta/2n \ll 1$ and one finds with the abbreviation $\Delta\sigma_n = \sigma_\infty - \sigma_n$ a very handy estimate

$$\frac{\Delta\sigma_n}{\sigma_\infty} \approx \left(\frac{\omega\beta}{4n}\right)^2 \tag{3.8}$$

of the relative Trotter discretization error concerning the width of a specific eigenmode of a molecule. Using the last two equations, this error can easily be estimated for every eigenmode of the molecule if the respective vibrational frequency is (roughly) known. Similarly, one can determine the Trotter number n which is necessary to perform a simulation with a desired accuracy.

There are cases where the ground state energy is known without having much information about the individual vibrational frequencies. Also for this case simple but useful estimates can be obtained. If the potential surface is again approximated by a multidimensional oscillator potential with f degrees of freedom, the energy E_0 in the quantum-mechanical ground state is simply given by

$$E_0 \approx \frac{1}{2}\sum_{i=1}^{f}\omega_i, \tag{3.9}$$

where ω_i is the vibrational frequency of the ith mode. Assuming that all frequencies are of a similar magnitude, *i.e.* $\omega \approx 2E_0/f$, one can combine the last two equations

$$\frac{\Delta\sigma_n}{\sigma_\infty} \approx \left(\frac{E_0\beta}{2fn}\right)^2, \tag{3.10}$$

whereas one obtains

$$\frac{\Delta\sigma_n}{\sigma_\infty} \approx \left(\frac{E_0\beta}{2n}\right)^2 \tag{3.11}$$

in the limit that a single mode of frequency ω dominates the vibrational spectrum, *i.e.* $\omega \approx 2E_0$. Combining these two limits we arrive at the following inequality

$$\left(\frac{E_0\beta}{2fn}\right)^2 \lesssim \frac{\Delta\sigma_n}{\sigma_\infty} \lesssim \left(\frac{E_0\beta}{2n}\right)^2, \tag{3.12}$$

where the right-hand-side strictly holds for for all modes, and the left-hand-side holds for at least the largest eigenmode of the molecule. This allows to estimate the Trotter number n to establish a predefined accuracy $\Delta\sigma_n/\sigma_\infty$ if only the ground state energy E_0 is known. Since for many systems either the ground state vibrational energy or even the vibrational frequencies are known from measurements or other calculations, one can now easily estimate "on the back of an envelope" the discretization n that is needed to push the systematic errors below a certain limit.

It is not clear *a priori* how well the underlying harmonic approximation for the ground state potential energy surface will do in practice. Our error estimates cannot be expected to be very accurate, but we will demonstrate in the next section that these crude relations agree nicely with the results of numerically exact path integral simulations, even for quite anharmonic potentials at moderate temperatures. Of course the estimates are expected to work the better the lower the temperature for a given potential energy surface. Thus, we believe that our harmonic estimation scheme will be useful in future large-scale path integral calculations of molecular systems.

4 Test applications

In order to test the reliability of the harmonic estimates, calculations for two small molecules, HD^+ and H_3^+, were carried out. These systems were selected because (i) very precise *ab initio* potential energy surfaces are known, (ii) quasi-exact reference data (energies, bond lengths and widths) are available, and (iii) these molecules can be measured *via* CEI at the Heidelberg Storage Ring TSR [34] so that a direct comparison with experiment is possible.

4.1 A diatomic molecule: HD^+

The calculations for HD^+ were done based on the numerical potential of reference [35] provided on a grid with a spacing of 0.25 a.u. which was interpolated with cubic splines for easy use; the masses $M_H = 1.007825$ a.m.u. and $M_D = 2.0141$ a.m.u. were used for all calculations reported in this paper. In order to be able to compare our simulation with published ground state results we have chosen a low simulation temperature of $T = 50$ K, where one does not expect any vibrational and almost no rotational excitations; the lowest rotational excitation energy amounts to about 65 K [36]. Because in our path integral simulations the molecules are allowed to move freely in three-dimensional space, an energy of $3k_BT/2$ accounting for the center of mass motion of the molecule must be subtracted from the total energy calculated in order to compare with calculations that do not include this degree of freedom. The simulations were carried out for several Trotter discretizations in the range $n = 2, ..., 256$.

The calculated energy and structure parameters are compared to results from benchmark calculations [36,37] of the HD^+ ground state in Table 1. Inspecting the averages obtained from simulations with increasing n one observes qualitatively a convergence to some specific values. More quantitatively, one finds that the averages $\langle\hat{A}\rangle_n$ determined at a certain Trotter discretization n show a linear behavior as a function of $1/n^2$ in the range $n = 64, ..., 256$. Thus one can use relation (3.1) truncated at the leading term in order to extract $\langle\hat{A}\rangle_\infty$, which in turn can be compared to the exact benchmark data. These extrapolated values are identical to the quasi-exact result within the statistical error.

In particular the structure of this diatomic molecule as defined by the average bond length $\langle r\rangle$ and the width σ of the associated distribution function agrees with the quasi-exact benchmark data. This test calculation demonstrates that path integral simulations are able to produce ground state data of similar quality as traditional approaches. At this stage, however, path integral simulations can go beyond what is easily accessible by more traditional methods, namely enter the regime of large excitations. The complete distribution function of the bond length $P(r, \beta)$ is shown in Figure 2 for various temperatures ranging from 50 K to 2500 K. As expected, the 300 K distribution does not deviate much from the low temperature result at 50 K, *i.e.*, the 50 K simulation is indeed close to the ground state of the molecule. Increasing the ro-vibrational excitations

Table 1. Calculated expectation values for the ground state of HD^+ from path integral Monte Carlo simulations at $T = 50$ K as a function of the Trotter discretization n compared to quasi-exact benchmark data [36,37] with $r_e = 2.016$ a.u.; $\langle r\rangle$: average bond length, $\sigma = \langle(r - \langle r\rangle)^2\rangle^{1/2}$: width of the bond length distribution, $\Delta\sigma_n$: estimated root-mean-square error of σ according to (3.7) using $\sigma_\infty = 0.2150$ a.u. and $\omega = 1913$ cm^{-1}, $\langle E_{\rm tot}\rangle$: total energy including translations and rotations at 50 K, $\langle E_0\rangle$: total energy excluding center of mass translations (*i.e.* $\langle E_{\rm tot}\rangle - 3/2k_{\rm B}T$ in the path integral simulation at $T = 50$ K); the Trotter extrapolation $n \to \infty$ was done using (3.1) to lowest order. The statistical error bars in parenthesis give ± one standard deviation on the reported digits (*e.g.* $2.055(2) = 2.055 \pm 0.002$) and were obtained according to (2.10). All data are reported in atomic units.

n	$\langle r\rangle$	σ	$\Delta\sigma_n$	$\langle E_{\rm tot}\rangle$	$\langle E_0\rangle$
4	2.007(1)	0.0804(14)	0.1334	−0.60165(3)	−0.60181(3)
16	2.026(2)	0.1513(11)	0.0625	−0.60012(4)	−0.60028(4)
64	2.050(1)	0.2065(8)	0.0089	−0.59804(3)	−0.59820(3)
128	2.053(1)	0.2125(3)	0.0024	−0.59779(2)	−0.59795(2)
256	2.054(2)	0.2145(2)	0.0006	−0.59778(2)	−0.59794(2)
$n \to \infty$	2.055(2)	0.2150(2)		−0.59774(1)	−0.59790(1)
exact	2.0544	0.2148			−0.597898

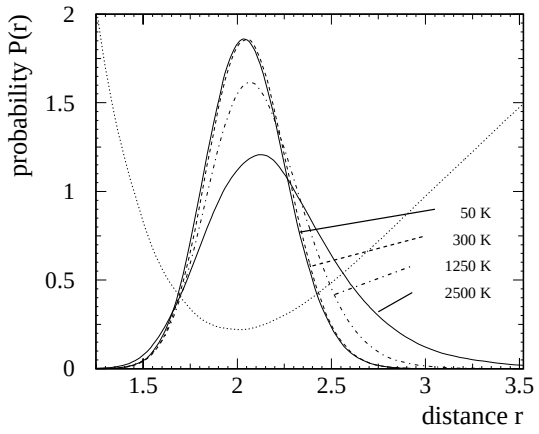

Fig. 2. Bond-length distribution function $P(r, \beta)$ of HD^+ obtained with $\beta/n < 25$ at different temperatures as defined in the figure. The underlying potential is schematically included by a dotted line for comparison.

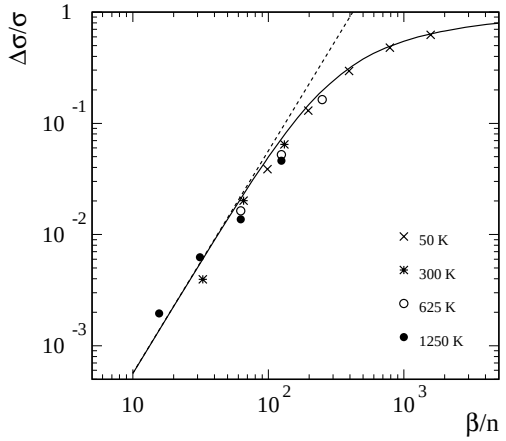

Fig. 3. Relative discretization error $\Delta\sigma_n/\sigma_\infty = (\sigma_\infty - \sigma_n)/\sigma_\infty$ of the width of the bond-length distribution $P_n(r, \beta)$ of HD^+ (partly shown in Fig. 2) as a function of β/n at different temperatures. Symbols denote the simulation data for various temperatures as defined in the figure. The solid line shows the estimated error (3.7) and the dashed line is the more approximate estimated error (3.8), which is valid for small β/n. A value of $\omega = 1913$ cm^{-1} was used for the analytical estimates [36].

leads to a substantial broadening of the distribution and shifts the average bond length to larger distances. Path integral simulations are able to quantify these effects.

Having shown that the path integral simulations do indeed converge to the exact answer if the systematic errors are accounted for by extrapolation, we now focus on estimating these errors for a *specific* value of the discretization *without* the need to perform simulations for various discretizations n. Thus, (3.7) or the more approximate relation (3.8) are now tested as estimates of the discretization error for distance distribution functions $P_n(r, \beta)$. A detailed comparison between these estimates and the actual discretization error computed from the simulations for different temperatures is carried out in Figure 3. For the analytical error estimates (lines) the vibrational frequency of 1913 cm^{-1} reported in reference [36] was used. For the simulation points (symbols) the "exact" width σ_∞ at a given temperature was computed by one very accurate calculation with a four-times larger discretization than the ones shown in the plot, and at 50 K the data for the ground state (see Tab. 1 and Refs. [36,37]) were used.

The comparison of the numerically exact reference points from path integral simulations to the full line obtained from (3.7) demonstrates that the harmonic estimate describes the magnitude of the systematic errors on the width of the distance distribution function very satisfactorily. This is even true at quite high temperatures, where it is known from Figure 2 that the anharmonicity of the potential is sampled. The more approximate estimate (3.8) converges to the true answer in the limit of small β/n as expected.

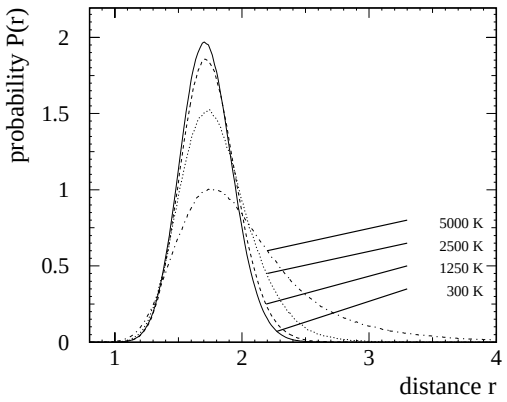

Fig. 4. Bond-length distribution function $P(r,\beta)$ of H_3^+ (averaged over all three H–H distances) obtained with $\beta/n < 25$ at different temperatures as defined in the figure.

4.2 A triatomic molecule: H_3^+

In the following we focus on the triatomic case H_3^+ where a potential energy surface calculated to micro-Hartree accuracy [4] was used; note that now an even more refined surface is available [5]. The path integral sampling was performed in the nine-dimensional set of staging coordinates $\{\mathbf{u}_i\}$, whereas the potential was evaluated by transforming (*via* the Cartesian coordinate set $\{\mathbf{r}_i\}$) to the so-called MBB representation [38] of the three-dimensional intra-molecular potential energy surface (using the very accurate parameters reported in Tab. 7 of Ref. [4]). Furthermore it is assumed that all nuclei are distinguishable particles. Thus, all effects due to spin statistics are neglected, *i.e.*, there is no symmetry-induced coupling of nuclear and rotational wave functions which restricts the accessible rotational quantum states. As a particular consequence the lowest energy ro-vibrational state in the present path integral simulation is the $J = 0$ rotational state in the vibrational ground state. Upon thermal excitation, the rotational states are sequentially populated without restrictions due to their symmetry.

In order to test our analytic estimates of the systematic errors, the distribution of the average bond length of the H_3^+ molecule was computed at various temperatures using different discretizations n. In Figure 4 the density distribution functions $P_n(r,\beta)$ of the bond lengths averaged over the three occurring H–H distances are depicted in the temperature range from 300 K to 5000 K. A significant broadening effect and thus shift of $\langle r\rangle$ to larger average bond lengths is observed as the temperature is raised. The results obtained for the lowest temperature 300 K and the largest discretization $n = 256$ are very close to the quasi-exact results for the A_1 ($J = 0$) ro-vibrational ground state obtained by Jaquet [39] using the same potential surface [4], see Table 2. There are, however, deviations which are small but nevertheless statistically significant. A first source for the discrepancy, the systematic error due to the finite Trotter number $n = 256$, is estimated to be much smaller than the statistical error so that it can be neglected. A second reason could be due to the fact that the simulations are performed at 300 K, whereas the benchmark data are obtained at 0 K. The first vibrationally excited state (with $J = 0$) is approximately 2500 cm^{-1} above the ro-vibrational ground state (with $J = 0$) [39]. Furthermore, there are many low-lying rotational states in the vibrational ground state, in particular three excitations are in the range 60–170 cm^{-1} [39]. The simulation temperature of 300 K corresponds to about 200 cm^{-1}, so that low-lying rotational excitations in the vibrational ground state are expected, whereas contributions due to populating the first vibrationally excited state are insignificant. Rotational excitations in general result in an increase of the average bond length $\langle r\rangle$ and a broadening of its width σ, which is indeed confirmed by the data in Table 2. Concerning the estimate of the energy one has to correct the total energy $\langle E_{\rm tot}\rangle$ for translations and rotational excitations at 300 K in order to compare to the benchmark data. The energy of the center of mass translations is easily taken into account exactly which leads to $\langle E_{\rm tot-trans}\rangle = \langle E_{\rm tot}\rangle - 3/2k_{\rm B}T$. In the absence of data for the rotational energy levels the rotational energy can only be estimated roughly by using the classical estimate of $1/2k_{\rm B}T$ per degree of freedom. This procedure will of course overestimate the correction due to rotations, and thus underestimate $\langle E_0\rangle$. Thus, $\langle E_{\rm tot-trans}\rangle$ is expected to severely overshoot the benchmark ground state energy, whereas $\langle E_0\rangle$ has to be smaller. Comparing the actual data in Table 2, it seems that the classical correction for rotations leads to an acceptable underestimation of the quasi-exact ground state energy, which is even within the statistical error bar. Whence, these rotational excitation effects are consistently observed and are the source for the remaining small discrepancy between the quasi-exact data at 0 K and their path integral estimates obtained at 300 K.

Having established that the exact ground state data are reproduced, we now test for the triatomic case the systematic error estimates derived in Section 3. The relative error $\Delta\sigma_n/\sigma_\infty$ of the width of the distance distribution functions should only depend on the ratio β/n in the limit of large discretizations n according to (3.7). The results of the actual calculations are plotted in Figure 5; the limiting width σ_∞ was determined at every temperature using an essentially converged discretization with $n_{\rm conv} = 4n_{\rm max}$, $n_{\rm max}$ being the largest discretization shown in the figure. The error estimate (3.7) is shown as a solid line and the more approximate relation (3.8) is included for comparison as a dashed line. For this purpose $\omega = 2521$ cm^{-1} from reference [4] is used as the reference frequency; note that here even a rough frequency estimate is sufficient. The triatomic case confirms the conclusion obtained for the diatomic molecule. The exact harmonic estimate (3.7) provides a very good estimation of the systematic errors due to the finite discretization of the path integral over the entire range of relevant temperatures and discretizations. The prediction of the approximate harmonic estimate (3.8) is obeyed only in the limit of sufficiently large Trotter numbers for a given temperature, *i.e.* a linear behavior with a slope of two

Table 2. Calculated expectation values for the ground state of H_3^+ from path integral Monte Carlo simulations at $T = 300$ K compared to quasi-exact benchmark data [4,39] with $r_e = 1.65000$ a.u.; $\langle r\rangle$: average bond length, $\sigma = \langle (r - \langle r\rangle)^2\rangle^{1/2}$: width of the bond length distribution, $\langle E_{tot}\rangle$: total energy including translations and rotations at 300 K, $\langle E_{tot-trans}\rangle$: total energy excluding center of mass translations (*i.e.* $\langle E_{tot}\rangle - 3/2 k_B T$ in the path integral simulation at $T = 300$ K), $\langle E_0\rangle$: total energy excluding center of mass translations and classical rotations (*i.e.* $\langle E_{tot}\rangle - 3k_B T$). The statistical error bars in parenthesis give $\pm$ one standard deviation on the reported digits (*e.g.* $1.726(2) = 1.726 \pm 0.002$) and were obtained according to (2.10). The systematic error $\Delta\sigma_n$ of σ due to the finite discretization is estimated using (3.7) with a vibrational frequency $\omega = 2521$ cm^{-1} and found to be of the order of $\pm\ 10^{-5}$ a.u., which is small compared to the statistical error in this calculation. Lengths are reported in atomic units, energies in cm^{-1}.

n	$\langle r\rangle$	σ	$\langle E_{tot}\rangle$	$\langle E_{tot-trans}\rangle$	$\langle E_0\rangle$
256	1.726(2)	0.2036(1)	4920(75)	4620(75)	4320(75)
exact	1.72170	0.203389			4361.816

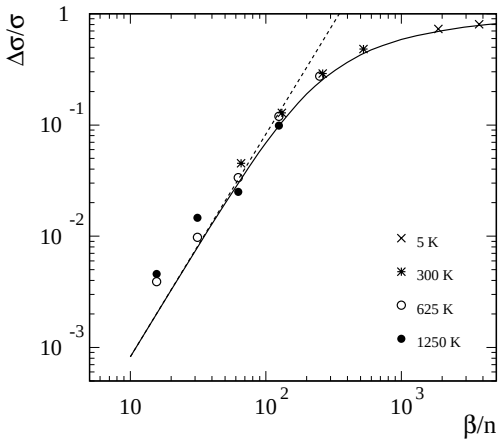

Fig. 5. Relative discretization error $\Delta\sigma_n/\sigma_\infty = (\sigma_\infty - \sigma_n)/\sigma_\infty$ of the width of the bond-length distribution $P_n(r,\beta)$ of H_3^+ (partly shown in Fig. 2) as a function of β/n at different temperatures. Symbols denote the simulation data for various temperatures as defined in the figure. The solid line shows the estimated error (3.7) and the dashed line is the more approximate estimated error (3.8), which is valid for small β/n. A value of $\omega = 2521$ cm^{-1} was used for the analytical estimates [4].

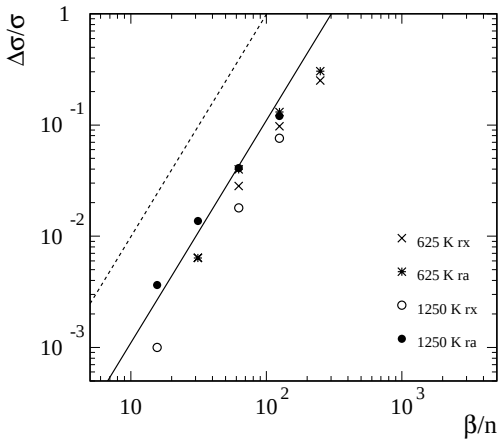

Fig. 6. Relative discretization error $\Delta\sigma_n/\sigma_\infty = (\sigma_\infty - \sigma_n)/\sigma_\infty$ of the width of the bond-length distribution $P_n(r,\beta)$ for the r_x and r_a normal modes (4.1) of H_3^+ as a function of β/n at different temperatures. Symbols denote the simulation data for two temperatures as defined in the figure. The dashed line is the estimated upper limit (3.11) and the solid line is relation (3.10), see also (3.12). A value of $E_0 = 4362$ cm^{-1} was used for the analytical estimates [4].

is obtained in the double-logarithmic representation chosen for Figure 5. Thus, in spite of the fact that the potential energy surface of H_3^+ is already quite anharmonic (note that $r_e = 1.65000$ a.u. [4] whereas $\langle r\rangle = 1.72170$ a.u. in the A_1 ($J = 0$) ground state and $\langle r\rangle = 1.76707$ a.u. in the E ($J = 0$) first vibrationally excited state [39], see also Tab. 2 and the broadening effects in Fig. 4) we find that these harmonic estimates are useful to assess the systematic discretization error in path integral simulations.

Finally, the usefulness of the relation (3.12), which assumes that only the ground state energy of the molecule is known in order to estimate the systematic finite-discretization error of bond length distributions, is assessed. According to (3.12) the discretization error is expected to approximately lie in a certain band for discretizations that satisfy $\beta/n \gg E_0$. Compared to the diatomic case, there is not only one bond length present in this molecule. For H_3^+, the following three symmetry-adapted coordinates

$$r_a = \frac{1}{\sqrt{3}}(r_{12} + r_{23} + r_{31})$$
$$r_x = \frac{1}{\sqrt{6}}(2r_{12} - r_{23} - r_{31})$$
$$r_y = \frac{1}{\sqrt{2}}(r_{23} - r_{31}), \qquad (4.1)$$

can be used, where r_{ij} denotes the distance between the ith and jth hydrogen atom. In Figure 6 the prediction is tested against numerical calculations separately for the distances r_a and r_x at two different temperatures. It is found that all deviations are smaller than the predicted upper limit (dashed line, according to (3.11)), and in addition they are rather close to the lower solid line given by (3.10). This behavior is easily understood because the three harmonic frequencies in H_3^+ are (for the present purpose) of similar magnitude so that (3.10) is the better approximation. In summary, also the estimate (3.12) that

is solely based on the ground state energy turns out to describe quite well the actual data and can be used to assess discretization errors in path integral simulations.

5 Conclusions

Path integral simulations provide a powerful way to determine the structure of molecules in their ground state and also in thermal equilibrium with the surrounding black body radiation. They automatically include the coupling between rotations and vibrations, taking into account the full nuclear Hamiltonian. For the path integral calculations performed in this study, the only non-trivial input needed is a global potential energy surface for the molecule, which can be obtained by standard quantum chemical calculations. Note that in the case of *ab initio* path integral simulations [10,16–19] not even the potential energy surface is required *a priori*. A particularly appealing feature of such path integral simulations is that the diagonal part of the many-body nuclear density matrix, which contains exhaustive information about the real-space structure of molecules, is obtained as an immediate result of these calculations. The many-body nuclear density can be compared directly to data obtained from Coulomb Explosion Imaging experiments [3].

The main focus of this paper is to study the systematic errors of the density matrix and in particular of bond length distribution functions which are introduced into path integral calculations due to the necessarily finite discretization of the continuous path integral. Estimating these errors in advance is highly desirable because convergence studies of the path integral and proper Trotter scaling can be prohibitively expensive. Two approaches to address this problem were introduced, both relying on the harmonic approximation. Simple and easy to use analytical formulae to estimate these errors in path integral structure calculations were derived, the only necessary input being approximate frequencies and/or energies. These relations were tested and validated by explicit calculations for diatomic (HD^+) and triatomic (H_3^+) molecules, for which quasi-exact benchmark data are available and where detailed convergence studies are also easily feasible. With these data in hand it was also demonstrated that path integral simulations lead in the low-temperature limit to ground state structures that agree quantitatively with those obtained by more traditional methods. It was found that harmonic estimates for the systematic discretization errors work astonishingly well even in the case of quite anharmonic potentials. Thus, the analytical formulae presented in Section 3 allow for quick *a priori* "back of the envelope" estimates of the discretization needed in order to push the systematic error of a path integral simulation below a certain limit of accuracy without performing convergence tests. Finally, we stress that these or similar error estimates will not only be useful for path integral simulations of molecules, but also of condensed matter systems such as solids or liquids.

It is a great pleasure to thank Ralph Jaquet for sharing with us his insights and results concerning H_3^+, Zeev Vager for useful and stimulating discussions, Roland Wester for the spline-interpolated HD^+ potential surface, and last but not least Dirk Schwalm for his interest and support of this study.

References

1. J.H. Callomon, E. Hirota, T. Iijima, K. Kuchitsu, W.J. Lafferty, in Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology, New Series Group II, Atomic and Molecular Physics **15**, Suppl. II/7, *Structure Data of Free Polyatomic Molecules*, edited by K.-H. Hellwege, A.M. Hellwege (Springer, Berlin, 1987), p. 1.
2. R.D. Brown, in *Structure and Conformations of Non-Rigid Molecules*, edited by J. Laane, M. Dakkouri, B. van der Veken, H. Oberhammer (Kluwer, Dordrecht, 1993), p. 99.
3. Z. Vager, R. Naaman, E.P. Kanter, Science **244**, 426 (1989).
4. R. Röhse, W. Kutzelnigg, R. Jaquet, W. Klopper, J. Chem. Phys. **101**, 2231 (1994).
5. W. Cencek, J. Rychlewski, R. Jaquet, W. Kutzelnigg, J. Chem. Phys. **108**, 2831 (1998); R. Jaquet, W. Cencek, W. Kutzelnigg, J. Rychlewski, J. Chem. Phys. **108**, 2837 (1998).
6. G.M. Chaban, J.O. Jung, R.B. Gerber, J. Chem. Phys. **111**, 1823 (1999).
7. M.A. Suhm, R.O. Watts, Phys. Rep. **204**, 293 (1991).
8. K.B. Whaley, Int. Rev. Phys. Chem. **13**, 41 (1994).
9. J.K. Gregory, D.C. Clary, Chem. Phys. Lett. **237**, 39 (1995).
10. D. Marx, M. Parrinello, Z. Phys. B **95**, 143 (1994).
11. M.J. Gillan, in *Computer Modelling of Fluids, Polymers, and Solids*, edited by C.R.A. Catlow, S.C. Parker, M.P. Allen (Kluwer, Dordrecht 1990).
12. K.E. Schmidt, D.M. Ceperley, in *The Monte Carlo Method in Condensed Matter Physics*, edited by K. Binder (Springer Verlag, Berlin, 1992), pp. 205.
13. D.M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995); for errata and updates see: `http://www.ncsa.uiuc.edu/Apps/CMP/papers/cep95a/cep95a.html`.
14. Ch. Chakravarty, Int. Rev. Phys. Chem. **16**, 421 (1997).
15. D. Marx, M.H. Müser, J. Phys. Cond. Matt. **11**, R117 (1999).
16. D. Marx, M. Parrinello, J. Chem. Phys. **104**, 4077 (1996).
17. M.E. Tuckerman, D. Marx, M.L. Klein, M. Parrinello, J. Chem. Phys. **104**, 5579 (1996).
18. D. Marx, M.E. Tuckerman, G.J. Martyna, Comput. Phys. Commun. **118**, 166 (1999).
19. D. Marx, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B.J. Berne, G. Ciccotti, D.F. Coker (World Scientific, Singapore, 1998), Chap. 15.
20. R. Car, M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
21. D. Marx, M. Parrinello, Nature **375**, 216 (1995); Z. Phys. D **41**, 253 (1997); D. Marx, A. Savin, Angew. Chem. Int. Ed. Engl. **36**, 2077 (1997); D. Marx, M. Parrinello, Science **286**, 1051 (1999), see `http://www.sciencemag.org/cgi/content/full/286/5442/1051a`.
22. D. Marx, M. Parrinello, Science **271**, 179 (1996).

23. M.E. Tuckerman, D. Marx (in preparation).
24. M.E. Tuckerman, D. Marx, M.L. Klein, M. Parrinello, Science **275**, 817 (1997).
25. I. Štich, D. Marx, M. Parrinello, K. Terakura, Phys. Rev. Lett. **78**, 3669 (1997); J. Chem. Phys. **107**, 9482 (1997).
26. R. Rousseau, D. Marx, Phys. Rev. Lett. **80**, 2574 (1998); J. Chem. Phys. **111**, 5091 (1999).
27. M.E. Tuckerman, B.J. Berne, G.J. Martyna, M.L. Klein, J. Chem. Phys. **99**, 2796 (1993).
28. R.P. Feynman, A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
29. R.P. Feynman, *Statistical Mechanics* (Addison-Wesley, Redwood City, 1972).
30. D. Chandler, P.G. Wolynes, J. Chem. Phys. **74**, 4078 (1981); D. Chandler, in *Liquids, Freezing, and Glass Transition*, edited by J.P. Hansen, D. Levesque, J. Zinn-Justin, (Elsevier, Amsterdam, 1991), p. 193.
31. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. **21**, 1087 (1953).
32. R.M. Fye, Phys. Rev. B **33**, 6271 (1986); R.M. Fye, R.T. Scalettar, Phys. Rev. B **36**, 3833 (1987); M. Suzuki, J. Stat. Phys. **43**, 883 (1986); see also *Quantum Monte Carlo Methods in Equilibrium and Nonequilibrium Systems*, edited by M. Suzuki (Springer, Berlin, 1987).
33. H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics and Polymer Physics* (World Scientific, Singapore, 1990).
34. R. Wester, F. Albrecht, M. Grieser, L. Knoll, R. Repnow, D. Schwalm, A. Wolf, A. Baer, J. Levin, Z. Vager, D. Zajfman, Nucl. Instrum. Meth. A **413**, 379 (1998).
35. T.E. Sharp, Atomic Data **2**, 119 (1971).
36. D.M. Bishop, L.M. Cheung, Phys. Rev. A **16**, 640 (1977).
37. D.M. Bishop, L.M. Cheung, Mol. Phys. **36**, 501 (1978).
38. W. Meyer, P. Botschwina, P.G. Burton, J. Chem. Phys. **84**, 891 (1986).
39. R. Jaquet, private communication (1998); the calculation is based on a C_{2v} representation of H_3^+ (where A_1, E, A_2 denote the states with increasing vibrational excitation), the potential energy surface of reference [4], and M_H = 1.007825 a.m.u.