Efficient methods for including quantum effects in Monte Carlo calculations of large systems: Extension of the displaced points path integral method and other effective potential m...

ARTICLE *in* THE JOURNAL OF CHEMICAL PHYSICS · JANUARY 2013

Impact Factor: 2.95 · DOI: 10.1063/1.4772667 · Source: PubMed

4 AUTHORS, INCLUDING:

Steven L Mielke
University of Minnesota Twin Cities
67 PUBLICATIONS **2,811** CITATIONS
SEE PROFILE

Mohammadhasan Dinpajooh
Arizona State University
12 PUBLICATIONS **29** CITATIONS
SEE PROFILE

Donald Truhlar
University of Minnesota Twin Cities
1,342 PUBLICATIONS **79,450** CITATIONS
SEE PROFILE

Efficient methods for including quantum effects in Monte Carlo calculations of large systems: Extension of the displaced points path integral method and other effective potential methods to calculate properties and distributions

Steven L. Mielke, Mohammadhasan Dinpajooh, J. Ilja Siepmann, and Donald G. Truhlar

Citation: J. Chem. Phys. **138**, 014110 (2013); doi: 10.1063/1.4772667
View online: http://dx.doi.org/10.1063/1.4772667
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v138/i1
Published by the American Institute of Physics.

Efficient methods for including quantum effects in Monte Carlo calculations of large systems: Extension of the displaced points path integral method and other effective potential methods to calculate properties and distributions

Steven L. Mielke,[a] Mohammadhasan Dinpajooh, J. Ilja Siepmann,[a] and Donald G. Truhlar[a]

Department of Chemistry, Chemical Theory Center, and Supercomputing Institute, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455-0431, USA

(Received 9 October 2012; accepted 5 December 2012; published online 4 January 2013)

We present a procedure to calculate ensemble averages, thermodynamic derivatives, and coordinate distributions by effective classical potential methods. In particular, we consider the displaced-points path integral (DPPI) method, which yields exact quantal partition functions and ensemble averages for a harmonic potential and approximate quantal ones for general potentials, and we discuss the implementation of the new procedure in two Monte Carlo simulation codes, one that uses uncorrelated samples to calculate absolute free energies, and another that employs Metropolis sampling to calculate relative free energies. The results of the new DPPI method are compared to those from accurate path integral calculations as well as to results of two other effective classical potential schemes for the case of an isolated water molecule. In addition to the partition function, we consider the heat capacity and expectation values of the energy, the potential energy, the bond angle, and the OH distance. We also consider coordinate distributions. The DPPI scheme performs best among the three effective potential schemes considered and achieves very good accuracy for all of the properties considered. A key advantage of the effective potential schemes is that they display much lower statistical sampling variances than those for accurate path integral calculations. The method presented here shows great promise for including quantum effects in calculations on large systems. © *2013 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4772667]

I. INTRODUCTION

Feynman path integral methods[1–48] provide a robust set of tools for the treatment of quantum effects in molecular simulations at a finite temperature. These methods may be used to obtain essentially exact results, but as a practical matter this is only affordable for small systems. To treat large systems, efficient approximation schemes[31,32,49–57] are needed, and one promising avenue is to use effective classical potential (ECP) schemes.[1,2,31,32,49,53,54,56–61] We have previously presented one such approach, the displaced points path integral (DPPI) method,[31] for calculating absolute partition functions. Here, we consider the extensions of this method and several related methods for calculating ensemble averages and coordinate distributions, and we discuss implementation strategies in the context of two algorithms, one designed to calculate absolute partition functions using uncorrelated sampling methods and the other being a general-purpose Metropolis sampling code.

[a] Authors to whom correspondence should be addressed. Electronic addresses: slmielke@gmail.com, siepmann@umn.edu, and truhlar@umn.edu.

II. BACKGROUND AND THEORY

II.A. Feynman path integrals

The partition function, $Q(T)$, may be calculated as the trace of the density operator ρ via

$$Q(T) = \frac{1}{\sigma^{\mathrm{sym}}} \int d\mathbf{q}\, \rho(\mathbf{q}), \tag{1}$$

where $\mathbf{q}$ is a collection of coordinates that are scaled to a common reduced mass μ, σ^{sym} is a symmetry number,

$$\rho(\mathbf{q}) = \langle \mathbf{q}| \exp(-\beta H) |\mathbf{q}\rangle , \tag{2}$$

H is the Hamiltonian, β is $1/k_{\mathrm{B}}T$, T is temperature, and k_{B} is Boltzmann's constant. If A is an arbitrary operator depending on the coordinates $\mathbf{q}$, its ensemble average may be calculated via

$$\langle A \rangle = \frac{1}{Q} \int d\mathbf{q} A(\mathbf{q}) \rho(\mathbf{q}). \tag{3}$$

A convenient way to calculate $Q(T)$ and ρ is by using Feynman path integrals, which yields[1,2]

$$\rho(\mathbf{q}) = \oint \mathscr{D}\mathbf{q}(s) \exp\left\{ \frac{-1}{\hbar} \int_0^{\hbar\beta} ds\, H[\mathbf{q}(s)] \right\}, \tag{4}$$

where $\oint \mathscr{D}\mathbf{q}(s)$ denotes the sum over all closed paths, $\mathbf{q}(s)$, passing through the point $\mathbf{q}$, and where the paths are parametrized as a function of imaginary time by the real variable s, which is the magnitude of the imaginary time.

Often it is convenient to instead consider the centroid of the paths. For a given path, the path centroid is defined as

$$\bar{\mathbf{q}} = \frac{1}{\hbar\beta}\int_0^{\hbar\beta} ds\mathbf{q}(s). \tag{5}$$

We can then define a centroid density as

$$\rho_c(\mathbf{q}_c) = \oint \mathscr{D}\mathbf{q}(s)\delta\left(\mathbf{q}_c - \frac{1}{\hbar\beta}\int_0^{\hbar\beta} ds\mathbf{q}(s)\right) \times \exp\left\{\frac{-1}{\hbar}\int_0^{\hbar\beta} ds H[\mathbf{q}(s)]\right\} \tag{6}$$

and we can calculate the partition function via

$$Q(T) = \frac{1}{\sigma^{\text{sym}}}\int d\mathbf{q}_c\rho_c(\mathbf{q}_c). \tag{7}$$

The motivation for writing the theory in terms of the centroid density ρ_c rather than the ordinary density ρ, which is a practice introduced by Feynman and Hibbs (FH),[1] is that it should be easier to devise accurate low-order approximations if we work in terms of ρ_c.

When working with the centroid density, ensemble averages must still be calculated at points on the paths—not at the centroid location. When sampling paths based on the centroid location, any point on a path could equally well be considered the initial path position, and thus, we could choose to calculate ensemble averages by

$$\langle A\rangle = \frac{1}{\sigma_{\text{sym}}Q}\int d\mathbf{q}_c\oint \mathscr{D}\mathbf{q}(s)\,\delta\left(\mathbf{q}_c - \frac{1}{\hbar\beta}\int_0^{\hbar\beta} ds\mathbf{q}(s)\right) \times A(\mathbf{q}[s=0])\exp\left\{\frac{-1}{\hbar}\int_0^{\hbar\beta} ds H[\mathbf{q}(s)]\right\}. \tag{8}$$

However, we can get better statistics in Monte Carlo simulations by averaging over the path. Therefore, we could instead use

$$\langle A\rangle = \frac{1}{\sigma_{\text{sym}}Q}\int d\mathbf{q}_c\oint \mathscr{D}\mathbf{q}(s)\bar{A}[\mathbf{q}(s)] \times \delta\left(\mathbf{q}_c - \frac{1}{\hbar\beta}\int_0^{\hbar\beta} ds\mathbf{q}(s)\right)\exp\left\{\frac{-1}{\hbar}\int_0^{\hbar\beta} ds\ H[\mathbf{q}(s)]\right\}, \tag{9}$$

where

$$\bar{A}[\mathbf{q}(s)] = \frac{1}{\hbar\beta}\int_0^{\hbar\beta} ds A[\mathbf{q}(s)]. \tag{10}$$

At high temperatures both Eqs. (1) and (7) reduce to the classical limit (CL) expression

$$Q^{\text{CL}} = \frac{1}{\sigma^{\text{sym}}}\left(\frac{\mu}{2\pi\hbar^2\beta}\right)^{N/2}\int d\mathbf{q}\exp[-\beta V(\mathbf{q})], \tag{11}$$

where $V(\mathbf{q})$ is the potential energy and N is the dimensionality of the coordinate space.

II.B. Effective classical potential methods

Summing over all possible paths or a converged sample of all possible paths involves considerable computational effort. Effective classical potential methods[1,2,31,36,44,49,54,56–61] provide alternative approaches that involve no explicit averaging over paths and might thus be employed with less computational effort. An effective classical potential method is one that replaces the true potential, V, in the CL expression of Eq. (11) by an effective classical potential, W^{eff}, which approximately accounts for quantum effects within a classical calculation. It should be noted here that interaction potentials that are parametrized by fitting[62,63] simulation data with classical sampling to reproduce experimental data, correspond inherently to effective classical potentials, although in practice they have additional approximations.

We will begin by reviewing ECP expressions for the partition function, and in a subsequent section we will consider expectation values. We initially restrict our attention to a particle of mass m moving in one dimension (1D) under a potential $V(q)$ where the domain of V is $-\infty < q < \infty$. For this case, the exact effective classical potential is given as

$$W^{\text{eff}}(q_c) = -\frac{\ln[\lambda\rho_c(q_c)]}{\beta}, \tag{12}$$

where λ is the classical de Broglie wavelength, given by

$$\lambda = \left(\frac{2\pi\hbar^2\beta}{m}\right)^{1/2}. \tag{13}$$

Equation (12) is an exact reformulation of quantum statistical mechanics, but, in general the exact ρ_c is unknown, and hence W^{eff} is unknown. We seek to find useful approximations to the exact W^{eff}. When such an approximation is available, one may calculate an approximate centroid density by

$$\rho_c(q_c) = \frac{\exp[-\beta W^{\text{eff}}(q_c)]}{\lambda} \tag{14}$$

and an approximate partition function via Eq. (7).

Feynman and Hibbs[1] defined a Gaussian smeared potential by

$$V^{\text{eff}}(q_c;\sigma) = \frac{1}{(2\pi\sigma^2)^{1/2}}\int_{-\infty}^{\infty} dr\ V(q_c+q)\exp(-q^2/2\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}}\int_{-\infty}^{\infty} dq\ V(q)\exp[-(q-q_c)^2/2\sigma^2], \tag{15}$$

where σ denotes a general Gaussian smearing parameter. The simplest approximation to an effective classical potential is[1]

$$W^{\text{eff}}(q_c) = V^{\text{eff}}(q_c;\sigma). \tag{16}$$

The Gaussian smeared potential and all quantities calculated using Eq. (16) depend on the value of σ; we will

sometimes, as in Eqs. (15) and (16), display this dependence explicitly to emphasize that a quantity was approximated by Gaussian smearing or to distinguish between different methods.

The FH effective classical potential is obtained by Eq. (16) with σ equal to $\sigma_{\rm FH}$ where

$$\sigma_{\rm FH}^2 = \beta\hbar^2/12m. \tag{17}$$

This ECP yields a lower bound[1] on the exact partition function at all T but only yields useful accuracy in the high-T limit.

Doll and Myers[49] (DM) chose to use Eqs. (15) and (16) with σ given by

$$\sigma_{\rm DM}^2 = \frac{2}{m\beta\omega^2}\ln\left[\frac{\sinh(\beta\hbar\omega/2)}{\beta\hbar\omega/2}\right] \tag{18}$$

leading to

$$W_{\rm DM}^{\rm eff}(q_{\rm c};\sigma_{\rm DM}) = \frac{1}{\left(2\pi\sigma_{\rm DM}^2\right)^{1/2}}\int_{-\infty}^{\infty} dq\, V(q_{\rm c}+q)\exp\left(-q^2/2\sigma_{\rm DM}^2\right), \tag{19}$$

which was selected to yield the exact partition function for a harmonic oscillator (HO) potential, i.e.,

$$V^{\rm HO} = \frac{1}{2}m\omega^2 q^2, \tag{20}$$

where ω is the harmonic frequency. For anharmonic potentials, we will use the instantaneous normal mode[64] (INM) frequencies for ω provided that these are real; the treatment for imaginary INM frequencies is discussed further below. (Note that INM refers to normal mode analysis at a point where the gradient is not zero.) In contrast with the FH ECP, the DM expression is nonvariational. Notice that

$$\sigma_{\rm DM} \xrightarrow{T\to\infty} \sigma_{\rm FH}. \tag{21}$$

Giachetti and Tognetti[58–60] (GT) and Feynman and Kleinert[61] (FK) independently arrived at the same variational effective classical potential scheme, which is one that employs two parameters. We will refer to this as the GTFK scheme, and it is given by

$$W_{\rm GTFK}^{\rm eff}(q_{\rm c};\sigma_{\rm GTFK},\Omega_{\rm GTFK}) = \frac{1}{(2\pi)^{1/2}\sigma_{\rm GTFK}(q_{\rm c})}\int dq\, V(q_{\rm c}+q) \times \exp(-q^2/\{2[\sigma_{\rm GTFK}(q_{\rm c})]^2\}) + \frac{1}{\beta}\ln\left[\frac{\sinh\left[\beta\hbar\Omega_{\rm GTFK}(q_{\rm c})/2\right]}{\beta\hbar\Omega_{\rm GTFK}(q_{\rm c})/2}\right] - \frac{m[\sigma_{\rm GTFK}(q_c)\Omega_{\rm GTFK}(q_c)]^2}{2}, \tag{22}$$

where the smearing parameter and effective frequency are obtained at each position by solution of the simultaneous equations

$$[\sigma_{\rm GTFK}(q_{\rm c})]^2 = \frac{1}{m\beta[\Omega_{\rm GTFK}(q_{\rm c})]^2} \times\left\{\frac{\beta\hbar\Omega_{\rm GTFK}(q_{\rm c})}{2}\coth\left[\frac{\beta\hbar\Omega_{\rm GTFK}(q_{\rm c})}{2}\right]-1\right\} \tag{23}$$

and

$$[\Omega_{\rm GTFK}(q_{\rm c})]^2 = \frac{1}{m(2\pi[\sigma_{\rm GTFK}(q_{\rm c})]^2)^{1/2}}\int dq\frac{d^2V(q)}{dq^2} \times\exp(-(q_{\rm c}-q)^2/\{2[\sigma_{\rm GTFK}(q_{\rm c})]^2\}). \tag{24}$$

For a harmonic oscillator, $\Omega_{\rm GTFK}(q_c)$ becomes independent of q_c and equals the oscillator frequency, but otherwise the GTFK method is computationally very expensive and is impractical for large systems. We will, however, consider a simpler but nonvariational approach that we will refer to as the renormalized effective classical potential scheme or the R scheme, which may be obtained from the GTFK approach for the special case of a harmonic potential. In particular, we have

$$W_{\rm R}^{\rm eff}(q_{\rm c};\sigma_{\rm R},\gamma_{\rm R}) = V^{\rm eff}(q_{\rm c};\sigma_{\rm R}) - \frac{\ln\gamma_{\rm R}}{\beta}, \tag{25}$$

where

$$\sigma_{\rm R}^2 = \frac{\hbar}{2m\omega}\coth(\beta\hbar\omega/2) - \frac{1}{m\beta\omega^2} \tag{26}$$

and

$$\gamma_{\rm R} = \frac{\hbar\beta\omega/2}{\sinh(\hbar\beta\omega/2)}\exp\left(\frac{\hbar\beta\omega}{4}\coth(\hbar\beta\omega/2)-\frac{1}{2}\right) = \frac{\hbar\beta\omega/2}{\sinh(\hbar\beta\omega/2)}\exp\left(m\beta\omega^2\sigma_{\rm R}^2/2\right). \tag{27}$$

Substituting Eq. (25) into Eq. (14) yields

$$\rho_{\rm c}^{\rm R}(q_{\rm c};\sigma_{\rm R},\gamma_{\rm R}) = \frac{\gamma_{\rm R}\exp[-\beta V^{\rm eff}(q_{\rm c};\sigma_{\rm R})]}{\lambda}, \tag{28}$$

which is why we refer to the method as "renormalized." The frequencies obtained in the variational GTFK scheme are guaranteed to be real for well-behaved potentials,[65] but the INM frequencies used in the R scheme and in several other approximation methods can be imaginary; this situation is discussed in Sec. II.D.

The Messina, Schenter, and Garrett[54] (MSG) ECP scheme scales the classical density by a quantum weighting function $\gamma_{\rm MSG}$, i.e.,

$$\rho_{\rm c}^{\rm MSG}(q_{\rm c};\gamma_{\rm MSG}) = \frac{\gamma_{\rm MSG}\exp[-\beta V(q_{\rm c})]}{\lambda}, \tag{29}$$

which is equivalent to choosing

$$W_{\rm MSG}^{\rm eff}(q_{\rm c};\gamma_{\rm MSG}) = V(q_{\rm c}) - \frac{\ln\gamma_{\rm MSG}}{\beta}, \tag{30}$$

where

$$\gamma_{\mathrm{MSG}} = \frac{\hbar\beta\omega/2}{\sinh(\hbar\beta\omega/2)}. \tag{31}$$

Note carefully that Eq. (29) is written using the true potential in contrast with Eq. (28), which is written using a V^{eff}. Thus, in contrast to the R-scheme and the DM, FH, and GTFK schemes, the MSG scheme does not require an expensive numerical quadrature to form the W^{eff}. In Sec. II.C, we will consider the DPPI methods, which also do not require expensive numerical quadratures, but which include an approximate anharmonicity correction for improved accuracy.

II.C. DPPI methods

The simplest DPPI method[31] is designed, for systems of any dimensionality, to include QM effects in classical simulations by replacing the classical potential with an evaluation of the potential at two points displaced from a centroid configuration, where the displacements are chosen so that the method reproduces exactly the partition function for a multidimensional harmonic oscillator. In the DPPI anharmonic sampling (AS) scheme[31] the instantaneous harmonic frequencies for a multidimensional system are evaluated by an INM calculation at each configuration.

In one dimension, the DPPI-AS method[31] may be thought of as being an ECP method obtained by a two-point Gauss–Hermite quadrature approximation of Eq. (19). Substituting this two-point approximation into Eq. (14) gives

$$\rho_{\mathrm{c,DPPI\text{-}AS}}(q_{\mathrm{c}}) = \frac{\exp\{-\beta[V(q_{\mathrm{c}}+c/2)+V(q_{\mathrm{c}}-c/2)]/2\}}{\lambda}, \tag{32}$$

where

$$c = 2\sigma_{\mathrm{DM}}. \tag{33}$$

This two-point quadrature gives the exact partition function for a harmonic potential (and is equivalent to the DM approximation for any potential that is a polynomial of order not greater than 3). For general potentials, this two-point quadrature is approximate; however, numerical tests[31] indicate that it often leads to better results than accurate integration of Eq. (19). For a d-dimensional system, two-point Gauss–Hermite quadrature in each dimension would require 2^d quadrature points. A key feature of the DPPI method is that we require only 2 points independent of d; the multidimensional DPPI scheme is discussed in detail in Sec. II.F.

We can reinterpret the use of Eq. (32) in an illuminating way as follows. According to Eq. (18), σ_{DM} depends on the harmonic frequency. Thus, using Eq. (33) corresponds to using a harmonic approximation to the centroid density to tell us where to sample, but then, when we sample, Eq. (32) corrects this harmonic approximation by using the correct value of V. In this respect, Eqs. (32) and (33) can be viewed as similar to a predictor-corrector method for integrating differential equations. We could consider developing more sophisticated predictor-corrector strategies than the one inherent in DPPI. However, even if we could do this successfully in 1D, it is unlikely that we could do this in the multidimensional implementation of DPPI if we wish to retain the computational efficiency that we only evaluate the potential at two points independent of the dimensionality because there does not appear to be sufficient information available with only two potential evaluations to design a more accurate approximation for multidimensional systems.

TABLE I. Numerical results for $Q(T)$ for a 1D quadratic-quartic potential $V(q) = aq^4 + bq^2$ with $a = 0.01$ a.u. and $b = 0.02$ a.u. for a particle with $\mu = 1224.259\ m_{\mathrm{e}}$.

T	Exact	DPPI-AS	DPPI-RAS
100	7.79×10^{-5}	8.20×10^{-5}	9.70×10^{-5}
150	1.82×10^{-3}	1.93×10^{-3}	2.08×10^{-3}
200	8.82×10^{-3}	9.28×10^{-3}	9.62×10^{-3}
250	2.27×10^{-2}	2.37×10^{-2}	2.42×10^{-2}
300	4.28×10^{-2}	4.42×10^{-2}	4.47×10^{-2}
400	9.46×10^{-2}	9.66×10^{-2}	9.70×10^{-2}
500	1.54×10^{-1}	1.56×10^{-1}	1.56×10^{-1}
600	2.15×10^{-1}	2.17×10^{-1}	2.17×10^{-1}
1000	4.50×10^{-1}	4.51×10^{-1}	4.51×10^{-1}

A two-point Gauss–Hermite quadrature approximation of the effective potential of the R scheme leads to a new method that we will call the displaced points path-integral method with renormalized anharmonic sampling (DPPI-RAS). In this method, we have

$$\rho_{\mathrm{c,DPPI\text{-}RAS}}(q_{\mathrm{c}}) = \frac{\gamma_{\mathrm{R}}\exp\{-\beta[V(q_{\mathrm{c}}+\sigma_{\mathrm{R}})+V(q_{\mathrm{c}}-\sigma_{\mathrm{R}})]/2\}}{\lambda}. \tag{34}$$

Both Eqs. (32) and (34) are exact for a HO, but not for an anharmonic potential. Numerical results for $Q(T)$ for a 1D quadratic-quartic potential studied previously[31,54] are given in Table I. In the supplementary material,[66] we also provide numerical results for a 1D asymmetric double well potential (Table S1) and for a 1D symmetric double well potential (Table S2); both of which have also been studied previously.[31,54] These three comparisons show that at temperatures up to 500 K, the use of Eq. (34), i.e., DPPI-RAS, usually performs slightly less well than the earlier DPPI-AS method that employs Eq. (32); at higher temperatures the results of the two methods are almost the same. Recalling the interpretation of Eqs. (32) and (33) as a predictor-corrector scheme gives one possible way to understand the better performance of the original scheme. Equation (34) may be interpreted as a predictor-corrector scheme where there is an approximate anharmonicity correction for the quantity within the exponential but no correction for the prefactor, γ_{R}. As discussed earlier,[31] the quantum weighting scheme of Messina, Schenter, and Garrett[54] may be thought of as a harmonic approximation without any anharmonic correction, and it consequently performs the poorest of the three methods (results not shown in Tables I, S1, and S2). When considered from this viewpoint it is not surprising that the DPPI-AS scheme is superior to the DPPI-RAS and MSG schemes. The good performance

of DPPI-AS for Q suggests that it is the preferred choice for generating ρ_c.

II.D. Treatment of imaginary frequencies

A special complication occurs in some effective classical potential schemes for regions of the potential where the INM frequencies are imaginary. For small magnitudes of $\beta|\omega|$ analytical continuation provides reasonable results; that is, we can use

$$\omega^2 \to -|\omega|^2$$
$$\frac{\hbar\beta\omega/2}{\sinh(\hbar\beta\omega/2)} \to \frac{\hbar\beta\,|\omega|\,/2}{\sin(\hbar\beta\,|\omega|\,/2)} \tag{35}$$
$$(\hbar\beta\omega/2)\coth(\hbar\beta\omega/2) \to (\hbar\beta\,|\omega|\,/2)\cot(\hbar\beta\,|\omega|\,/2)$$

in the equations for σ_R, γ_R, σ_{GTFK}, etc. Unfortunately, though, the trigonometric functions on the right-hand sides become singular at certain finite values of $\hbar\beta|\omega|/2$. In the full GTFK scheme, the variational optimization of the parameters in the presence of the true anharmonic potential can be shown to produce nonsingular effective parameters under broad conditions.[67] However, in the nonvariational schemes we need some other mechanism to avoid unphysical behavior. Doll and Myers[49] only presented their method in the limited context of a single positive frequency so they never needed to address complications from imaginary INM frequencies. Messina, Schenter, and Garrett[54] noted that if they replaced the infinite parabolic barrier with a truncated barrier they could replace γ_{MSG} with the tunneling correction factor obtained for a truncated barrier by Skodje and Truhlar.[68] This suggestion also provides a way[31] to calculate a well-behaved σ_{DM} for imaginary ω, but it is not clear how to extend this idea to a treatment of σ_R. Also, this approach requires the selection of a cutoff parameter for the truncated barrier, which in past work has been chosen so that the truncated barrier never drops below the global minimum of the potential, but we may not know the global minimum of the potential energy surface (PES) (it is required for calculating absolute partition functions, but for sampling in large systems based on relative energies this quantity may be unavailable). Furthermore, in extended systems, the global minimum may be far away from the current configuration, and it would be unphysical for results at the current configuration to depend on a far away point.

In the present work, we adopt the following scheme. When calculating σ_R, σ_{DM}, γ_R, and γ_{MSG} we use the absolute value of ω in all cases, and when we have an imaginary ω we replace γ_R and γ_{MSG} by their reciprocals. This produces values that are nearly identical to those of analytical continuation for small values of $\beta|\omega|$ and leads to reasonable behavior for large values. A similar approach was adopted in the local generalized Pitzer–Gwinn (LGPG) path integral scheme;[48] if we replace the MSG recommendation for treating imaginary frequencies with this approach, the MSG scheme is identical to the LGPG scheme for the special case of $P = 1$, where P is the number of points used in the path discretization. In the following, when using the MSG scheme with this treatment of imaginary frequencies we will refer to it as the modified MSG (MMSG) scheme.

II.E. Ensemble averages in ECP schemes

As discussed by Voth,[53] it is useful to extend the smearing operation to other variables besides potential energy. In cases where the dynamical variable of interest is a simple function of position, $A(r)$, we therefore define a smeared A by

$$A^{\text{eff}}(q_c;\sigma) = \frac{1}{(2\pi\sigma^2)^{1/2}}\int_{-\infty}^{\infty} dq\, A(q_c+q)\exp(-q^2/2\sigma^2). \tag{36}$$

Voth[53] discussed a scheme in which ensemble averages for a one-dimensional system are carried out by

$$\langle A\rangle = \frac{\int_{-\infty}^{\infty} dq_c\, A^{\text{eff}}(q_c;\sigma)\rho_c(q_c;\sigma,\gamma)}{\int_{-\infty}^{\infty} dq_c\, \rho_c(q_c;\sigma,\gamma)}, \tag{37}$$

where

$$\rho_c(q_c;\sigma,\gamma) = \frac{\gamma\exp[-\beta V^{\text{eff}}(q_c;\sigma)]}{\lambda}. \tag{38}$$

Voth's treatment was restricted to the special case of $\gamma = 1$, and he considered only the FH and GTFK ECPs. Voth[53] suggested that one can obtain the density $\rho(q;\sigma,\gamma)$ corresponding to a particular centroid density, ρ_c, by using

$$\rho(q;\sigma,\gamma) = \int_{-\infty}^{\infty} dq_c\, \delta^{\text{eff}}(q-q_c;\sigma)\rho_c(q_c;\sigma,\gamma), \tag{39}$$

where

$$\delta^{\text{eff}}(q-q_c;\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-(q-q_c)^2/2\sigma^2). \tag{40}$$

Using Eqs. (39) and (40) one can show that

$$\int dq_c\, A^{\text{eff}}(q_c;\sigma)\,\rho_c(q_c;\sigma,\gamma) = \int dq\, A(q)\,\rho(q;\sigma,\gamma). \tag{41}$$

If, for any particular problem, ρ_c yields the exact centroid density, then $\rho(q;\sigma,\gamma)$ calculated by Eq. (39) will be the exact density, and Eq. (41) shows that, in such cases, Eq. (37) will yield exact ensemble averages.

Equation (36) gives

$$[q^{\text{eff}}(q_c;\sigma)]^2 = q_c^2 + \sigma^2 \tag{42}$$

so that

$$V^{\text{eff,HO}}(q_c;\sigma) = \frac{m\omega^2}{2}[q^{\text{eff}}(q_c;\sigma)]^2 = \frac{m\omega^2}{2}\left(q_c^2+\sigma^2\right). \tag{43}$$

The approximate centroid density for a harmonic potential obtained using Eqs. (14) and (16) is

$$\rho_c^{\text{HO}}(q_c;\sigma,\gamma) = \frac{\gamma\exp\left[\dfrac{-\beta m\omega^2}{2}\left(q_c^2+\sigma^2\right)\right]}{\lambda} \tag{44}$$

and using Eqs. (43) and (44) in Eq. (37) gives for a harmonic oscillator

$$\langle q^2\rangle = 1/\beta m\omega^2 + \sigma^2. \tag{45}$$

Thus, we have

$$\langle V^{\rm HO}\rangle = \frac{m\omega^2}{2}\langle q^2\rangle = \frac{1}{2\beta} + \frac{m\omega^2\sigma^2}{2}. \tag{46}$$

For comparison, the exact value of $\langle V\rangle$ for a harmonic oscillator is given by[2]

$$\langle V^{\rm HO}\rangle_{\rm exact} = \frac{\hbar\omega}{4}\coth(\beta\hbar\omega/2). \tag{47}$$

With the centroid density of Eq. (38), for the special case of a harmonic oscillator, Eqs. (39) and (44) yield

$$\rho^{\rm HO}(q;\sigma,\gamma) = \gamma\left(\frac{m}{2\pi\hbar^2\beta}\right)^{1/2}\left(\frac{1}{1+\beta m\omega^2\sigma^2}\right)^{1/2} \times \exp\left(\frac{-\beta m\omega^2\sigma^2}{2} - \frac{\beta m\omega^2 q^2}{2(1+\beta m\omega^2\sigma^2)}\right), \tag{48}$$

which can be compared to the exact HO result[2]

$$\rho^{\rm HO}_{\rm exact}(q) = \left(\frac{m\omega}{2\pi\hbar\sinh(\hbar\beta\omega)}\right)^{1/2} \times \exp\left(\frac{-m\omega q^2\tanh(\hbar\beta\omega/2)}{\hbar}\right). \tag{49}$$

Equating Eqs. (46) and (47), to enforce the requirement that $\langle V\rangle$ be exact for a harmonic oscillator, leads to the requirement that $\sigma = \sigma_{\rm R}$ as given in Eq. (25). Similarly, by equating Eqs. (48) and (49) one can show that choosing $\sigma = \sigma_{\rm R}$ and $\gamma = \gamma_{\rm R}$ leads to the exact density for a HO. Because we obtain the exact density, the R scheme reproduces Q and all ensemble averages exactly for a HO.

If we use $\sigma = \sigma_{\rm DM}$ and $\gamma = 1$ in the treatment above, we get the exact HO density but an inexact $\langle V\rangle$; in order to get accurate expectation values for the other ECP treatments that yield exact HO centroid densities we need to generalize Eq. (37) to

$$\langle A\rangle = \frac{\int_{-\infty}^{\infty} dq_{\rm c}\, A^{\rm eff}(q_{\rm c};\sigma_{\rm R})\rho_{\rm c}^{\rm X}(q_{\rm c};\sigma_{\rm X},\gamma_{\rm X})}{\int_{-\infty}^{\infty} dq_{\rm c}\,\rho_{\rm c}^{\rm X}(q_{\rm c};\sigma_{\rm X},\gamma_{\rm X})}, \tag{50}$$

where "X" can denote the DM, MSG, MMSG, or R schemes as well as the DPPI-AS and DPPI-RAS schemes, and where parameters unused in a particular scheme (i.e., σ in the MSG scheme and γ in the DM scheme) are ignored. Note carefully that we always need to use $\sigma_{\rm R}$ in Eq. (36) when forming $A^{\rm eff}$ even when, as in the DM scheme, we use a different σ for the ECP. (The relative magnitudes of these parameters is worth noting; for $\beta\hbar\omega = 1$, 5, 10, and 100, the ratio $\sigma_{\rm DM}/\sigma_{\rm R} = 1.004$, 1.073, 1.161, and 1.361, respectively.) When we calculate expectation values using Eq. (50) with the DM ECP, we will refer to this as the extended Doll–Myers scheme, and when we use the MSG or MMSG ECP we will refer to this as the extended Messina–Schenter–Garrett (EMSG) scheme or the extended modified Messina–Schenter–Garrett (EMMSG) scheme. As previously mentioned, one of the principal advantages of the DPPI schemes is that we do not need to perform expensive quadratures to calculate the effective potential; however, at first glance it would seem that we need to perform quadratures in Eq. (36) to obtain expectation values, which would greatly diminish this advantage. Fortunately, it is possible, as will be discussed in more detail in Sec. III, to evaluate operator averages using a Monte Carlo strategy that does not require accurate integration of Eq. (36) for each Monte Carlo sample, so we can retain the computational efficiency of the DPPI methods.

Notice that at low temperature we have

$$\sigma_{\rm R}^2 \xrightarrow{T\to 0} \frac{\hbar}{2m\omega}. \tag{51}$$

At $T = 0$ the classical density and the centroid density are both localized as a delta function at the minimum energy structure, and the Gaussian smearing becomes averaging over a Gaussian of the form

$$\exp\left(\frac{-q^2}{2\sigma_{\rm R}^2}\right) = \exp\left(\frac{-m\omega q^2}{\hbar}\right). \tag{52}$$

The right-hand side of Eq. (52) is the square of the ground-state harmonic oscillator wave function or equivalently the exact density of a HO restricted to its ground state.

II.F. Multidimensional DPPI calculations

For a system with N dimensions, Eq. (13) is replaced by

$$\lambda^N = \left(\frac{2\pi\hbar^2\beta}{m}\right)^{N/2}. \tag{53}$$

We find it convenient when calculating absolute partition functions to remove the center of mass motion, and thus the partition functions we tabulate include only contributions from rotation and vibration; in such cases N is $3n^{\rm atoms}-3$, where $n^{\rm atoms}$ is the number of atoms in the system. Sometimes, it is convenient to explicitly sample the translational degrees of freedom, such as when we employ Metropolis sampling, in which case N is $3n^{\rm atoms}$. For some problems, one might want to include quantum effects for only a subset, F, of the degrees of freedom.

After one has selected a centroid configuration to evaluate for sampling, one carries out an INM analysis to get INM eigenvalues ω_j and normalized eigenvectors $\hat{\mathbf{e}}^{(j)}$, $j = 1, 2, \ldots, F$, which when transformed to the same mass scaling as used for $\mathbf{q}_{\rm c}$ are labeled $\mathbf{e}^{(j)}$. Using Eqs. (18) and (33) with the frequency ω replaced by the frequency ω_j of mode j, we calculate a single-mode shift parameter $c^{(j)}$ for each of the F modes. The single-mode shift parameters are then combined into a multi-mode shift vector in $2^{(F-1)}$ possible ways with[31]

$$\mathbf{c}^{[i]} = \sum_{j=1}^{F} s_j^{[i]} c^{(j)}\mathbf{e}^{(j)}, \tag{54}$$

where $s_1^{[i]} = 1$ and the $s_j^{[i]}$ for $j > 1$ are a set of random signs (± 1). Using two-point Gauss–Hermite quadrature in each of

the F degrees of freedom is equivalent to taking

$$\rho_{\mathrm{c}}^{\text{DPPI-MAAS}}(\mathbf{q}_{\mathrm{c}}) = \frac{\exp\left[-\frac{\beta}{2^F}\sum_{i=1}^{2^{F-1}}\{V(\mathbf{q}_{\mathrm{c}}+\mathbf{c}^{[i]}/2)+V(\mathbf{q}_{\mathrm{c}}-\mathbf{c}^{[i]}/2)\}\right]}{\lambda^N}, \tag{55}$$

where MAAS denotes "mean action anharmonic sampling." This scheme scales poorly with the dimensionality of the problem so we instead choose

$$\rho_{\mathrm{c}}^{\text{DPPI-AS}}(\mathbf{q}_{\mathrm{c}}) = \frac{\frac{1}{2^{F-1}}\sum_{i=1}^{2^{F-1}}\exp\left[-\frac{\beta}{2}\left\{V(\mathbf{q}_{\mathrm{c}}+\mathbf{c}^{[i]}/2)+V(\mathbf{q}_{\mathrm{c}}-\mathbf{c}^{[i]}/2)\right\}\right]}{\lambda^N}. \tag{56}$$

We can re-write Eq. (56) as

$$\rho_{\mathrm{c}}^{\text{DPPI-AS}}(\mathbf{q}_{\mathrm{c}}) = \frac{\frac{1}{2^{F-1}}\sum_{i=1}^{2^{F-1}}\exp\left[-\beta W_i^{\text{eff}}(\mathbf{q}_{\mathrm{c}})\right]}{\lambda^N}, \tag{57}$$

where

$$W_i^{\text{eff}}(\mathbf{q}_{\mathrm{c}}) = \frac{1}{2}[V(\mathbf{q}_{\mathrm{c}}-\mathbf{c}^{[i]}/2)+V(\mathbf{q}_{\mathrm{c}}+\mathbf{c}^{[i]}/2)] \tag{58}$$

and we can consider $W_i^{\text{eff}}$to be a multivalued effective classical potential. As discussed previously,[31] one can show that the partition function obtained by DPPI-AS is an upper bound on that of the DPPI-MAAS scheme. The two results are identical for a harmonic potential and they are close for anharmonic potentials. When doing calculations by Monte Carlo we do not need to average over all i in Eq. (56) at each Monte Carlo sample; we merely choose one or more of the 2^{F-1} terms randomly and, in the limit of a large number of samples, we will converge to the correct results (in the calculations presented here, we chose just one).

At non-stationary points there are, in general, $3n^{\text{atoms}}-3$ nonzero frequencies because vibrations and rotations become coupled; thus, there are typically several low-magnitude frequencies associated with rotation-like modes in a full-dimensional INM analysis. Such rotational or nearly rotational modes present numerical complications. A pure rotational mode would have a zero frequency and this leads to a removable singularity in Eq. (18), with the limiting behavior as $\omega \to 0$ being found by L'Hospital's rule to be $\sigma_{\mathrm{DM}} = \sigma_{\mathrm{FH}}$. Similarly, in the MSG scheme the $\omega \to 0$ limit gives $\gamma_{\mathrm{MSG}} = 1$, and in the R scheme this limit gives $\sigma_{\mathrm{R}} = \sigma_{\mathrm{FH}}$. Such zero-frequency modes are properly sampled classically, so the MSG approach, which includes no quantum correction for zero frequency modes, treats these degrees of freedom properly. However, any finite step along an instantaneous rotational mode does not lead to a pure rotation, so without modification the DPPI-type methods will lead to an inaccurate quantum correction for such modes. In the initial presentation of the DPPI methods[31] we used internal coordinates to perform the INM analysis in the $(3n^{\text{atoms}}-6)$ dimensional vibrational subspace,[69] and we formed the displacement vectors of Eq. (54) from the resulting internal-coordinate INM eigenvectors so the question of how to treat rotational motions did not arise; however, for larger systems we wish to work with Cartesian coordinates.

We have considered two possible ways to address complications from finite displacements along rotational modes. The first approach is to project out rotational motions from the Hessian before calculating INMs (see the supplementary material for details), to set $F = 3n^{\text{atoms}}-6$, and to not include rotational-mode displacements in Eq. (54); this precludes a full treatment of vibrational-rotational coupling but avoids the numerical difficulties with low-frequency, and typically highly anharmonic, modes for which effective potential schemes are not especially reliable. This approach will be called the rotational projection method. In the second approach, we do not project rotational motions, we number the unprojected INM modes in order of decreasing absolute frequency and we set $F = 3n^{\text{atoms}}-6$; this corresponds to not taking displacements in the 3 vibrational-rotational modes for which the calculated INM frequencies have the lowest absolute magnitude and is motivated by the fact that these are the most classical modes. In this respect, we note that omitting certain modes from the DPPI displacements leaves these motions treated at a classical level because the integration in Eq. (7) still includes all non-translational degrees of freedom. This unprojected approach is similar in spirit to a method[70] for adding quantal vibrational effects to potentials of mean force. We could go further by including quantum effects on only $F = N'$ modes, where N' is a number smaller than $3n^{\text{atoms}}-6$; the motivation would be that for many purposes it is necessary to include quantum effects only in the high-frequency modes. In the rest of the main text, we only consider the projection scheme, but we provide numerical results of the unprojected scheme with $F = 3n^{\text{atoms}}-6$ and results when including all modes ($F = 3n^{\text{atoms}}-3$) in the supplementary material.

III. IMPLEMENTATION OVERVIEW

An ECP scheme is designed to provide a convenient recipe for the approximate inclusion of QM effects into codes designed for classical mechanical sampling of configuration space. Next, we explain the Monte Carlo procedure we use to treat the QM effects by the DPPI-AS method.

We select a candidate centroid configuration, $\mathbf{q}_{\mathrm{c}}$, by a Monte Carlo sampling procedure, just as for classical sampling. We perform an INM calculation at this $\mathbf{q}_{\mathrm{c}}$. Note that the effective classical potential $W_i^{\text{eff}}(\mathbf{q}_{\mathrm{c}})$ at each centroid configuration $\mathbf{q}_{\mathrm{c}}$ can take on $2^{(F-1)}$ values [i.e., any of the terms generated by the random signs in Eq. (54)], so the value will vary for different evaluations of the same configuration, but we decided above to select only a random term i from the right-hand side of Eq. (56) rather than the average of all possible terms, which is computationally unaffordable for large systems. Monte Carlo sampling ensures that we will eventually converge to the desired mean value provided that we select the retained term randomly. Therefore, we calculate $W_i^{\text{eff}}(\mathbf{q}_{\mathrm{c}})$ only for this random value of i. This is the energy to be used for any Monte Carlo acceptance/rejection decisions if one is using Metropolis sampling.

For each degree of freedom, j, we evaluate a $\sigma_{\mathrm{R},j}$ using Eq. (26) (using $|\omega_j|$ for cases where ω_j is imaginary). Then, for each j, we select a random number, δ_j, from a Gaussian distribution, $\exp(-\delta_j^2/2\sigma_{\mathrm{R},j}^2)/\sqrt{2\pi\sigma_{\mathrm{R},j}^2}$. Then we obtain a configuration $\mathbf{q}_\mathrm{d}$, by

$$\mathbf{q}_\mathrm{d} = \mathbf{q}_\mathrm{c} + \sum_{j=1}^{F} \delta_j \mathbf{e}_j, \tag{59}$$

where the sum runs over the same F modes used in Eq. (54). If any expectation values are desired for operators that are functions of position, we evaluate them at $\mathbf{q}_\mathrm{d}$, and we average these over accepted configurations in the same way that one carries out averages over accepted classical configurations in a classical simulation. (We could, optionally, choose to sample at several different values of $\mathbf{q}_\mathrm{d}$ for each centroid configuration if some expectation values converge slowly. In the following, and for all numerical results presented in this paper or in the supplementary material, we will only use a single value of $\mathbf{q}_\mathrm{d}$ for each centroid configuration.) In addition to using $\mathbf{q}_\mathrm{d}$ for the calculation of ensemble averages, this is also the correct variable to use when accumulating coordinate distributions.

Expectation values of operators that are not just functions of position must be calculated in a different way. For example, to calculate the ensemble average of the energy, Doll and Myers[49] used

$$\langle E_\mathrm{rovib}\rangle = k_B T^2 \left(\frac{\partial \ln Q}{\partial T}\right) = -\left(\frac{\partial \ln Q}{\partial \beta}\right) = -\frac{1}{Q}\left(\frac{\partial Q}{\partial \beta}\right), \tag{60}$$

where Q is a vibrational rotational partition function, and the contribution to $\langle E\rangle$ from translational motion is easily added by

$$\langle E\rangle = \langle E_\mathrm{rovib}\rangle + \frac{3}{2\beta}. \tag{61}$$

We can evaluate $\langle E_\mathrm{rovib}\rangle$ by using Eq. (7) for Q that is obtained by using Eq. (56). Substituting this into Eq. (60) yields

$$\begin{aligned}
\langle E_\mathrm{rovib}\rangle &= -\frac{1}{Q}\frac{\partial}{\partial\beta}\frac{1}{\sigma_\mathrm{sym}}\int d\mathbf{q}_\mathrm{c}\frac{1}{\lambda^N}\frac{1}{2^{F-1}}\sum_i e^{-\beta W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})} \\
&= -\frac{1}{\sigma_\mathrm{sym}}\frac{\beta^{N/2}}{Q}\frac{\partial}{\partial\beta}\int d\mathbf{q}_\mathrm{c}\frac{1}{2^{F-1}}\sum_i \beta^{-N/2} e^{-\beta W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})} \\
&= \frac{\frac{1}{\sigma_\mathrm{sym}}\int d\mathbf{q}_\mathrm{c}\frac{1}{2^{F-1}}\sum_{i=1}^{2^{F-1}}\left(\frac{N}{2\beta}+W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})+\beta\frac{\partial W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})}{\partial\beta}\right)e^{-\beta W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})}}{Q}.
\end{aligned} \tag{62}$$

Note that the weighting factors $e^{-\beta W_i(\mathbf{q}_\mathrm{c})}$ in Eq. (62) involve the $W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})$, which depend (in the DPPI-AS scheme) on the set $\{\sigma_{\mathrm{DM},j}\}$, not the set $\{\sigma_{\mathrm{R},j}\}$ that is used for ensemble averages of functions of $\mathbf{q}_\mathrm{c}$; Eq. (62) also differs from Eq. (36) in that it involves terms evaluated at the centroid location rather than at a $\mathbf{q}_\mathrm{d}$. We again need only sample a single random term (a random choice of i) from Eq. (62) for our results to converge to the desired $\langle E\rangle$.

One could evaluate the derivative with respect to β in Eq. (62) by a finite difference approximation. In this approximation, one would calculate values of $\{\sigma_{\mathrm{DM},j}\}$ for a value of β that is slightly different from that for the simulation temperature and—at the cost of additional potential evaluations—would generate displaced-temperature values of $W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})$, using the same set of random phases as in the first evaluation, that would be used solely for a finite difference calculation of $\partial W_i^\mathrm{eff}/\partial\beta$.

Note that an average energy estimator based on Eq. (62) involves partially averaged quantities and is not appropriate for evaluating fluctuations in the energy, as for instance, are sometimes used to compute heat capacities. Because we define Q so that it does not include contributions from overall translation, one way to calculate the heat capacity is to use

$$C_V = C_\mathrm{rovib}(T) + \frac{3}{2}k_\mathrm{B} \tag{63}$$

and

$$\begin{aligned}
C_\mathrm{rovib}(T) &= k_\mathrm{B}\beta^2\left[\frac{1}{Q}\frac{\partial^2 Q}{\partial\beta^2} - \left(\frac{1}{Q}\frac{\partial Q}{\partial\beta}\right)^2\right] \\
&= k_\mathrm{B}\beta^2\left[\frac{1}{Q}\frac{\partial^2 Q}{\partial\beta^2} - \langle E_\mathrm{rovib}\rangle^2\right].
\end{aligned} \tag{64}$$

An estimator for the first term in Eq. (64) can be constructed in a straightforward fashion by analogy to Eq. (62). This yields

$$\begin{aligned}
&\left\langle\frac{1}{Q}\frac{\partial^2 Q}{\partial\beta^2}\right\rangle \\
&= \frac{1}{Q}\frac{1}{\sigma_\mathrm{sym}}\int d\mathbf{q}_\mathrm{c}\frac{1}{2^{F-1}}\sum_{i=1}^{2^{F-1}}\left\{\left[\left(\frac{N}{2\beta}+W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})\right.\right.\right. \\
&\quad\left.\left.+\beta\frac{\partial W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})}{\partial\beta}\right)^2+\left(\frac{N}{2\beta^2}-2\frac{\partial W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})}{\partial\beta}-\beta\frac{\partial^2 W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})}{\partial\beta^2}\right)\Bigg] \\
&\quad\left.\times\exp\left[-\beta W_i^\mathrm{eff}(\mathbf{q}_\mathrm{c})\right]\right\}.
\end{aligned} \tag{65}$$

As in the case of Eq. (62), Monte Carlo sampling may be done by selecting a single random value of i at each configuration instead of forming the average over all i. When evaluating Eqs. (62) and (65) in an importance sampling scheme, such as Metropolis, where the samples are drawn from a distribution determined from the density, we omit the $1/Q$ factor. When sampling only a single value of i in Metropolis sampling the value of i is only changed when a new move is accepted (i.e., when a configuration is resampled the value of i should not be changed).

IV. CALCULATIONS

In order to evaluate the effectiveness of the DPPI-AS, DPPI-RAS, and MMSG schemes for evaluating operator averages, we have performed calculations for an isolated

water molecule using the Partridge–Schwenke[71] (PS) PES. In addition to the partition function, we considered the average energy $\langle E\rangle$, the average potential energy $\langle V\rangle$, the heat capacity, C_V, and expectation values for the O–H distance $\langle R_{\mathrm{OH}}\rangle$ and the H–O–H angle $\langle\theta_{\mathrm{HOH}}\rangle$. Benchmark values for these quantities were also obtained by accurate path integral calculations, as outlined below. The DPPI-AS calculations were performed using two different codes: a specialized code designed to calculate absolute partition functions for small systems using Jacobi coordinates and specialized importance sampling and stratification techniques, and a code that uses Metropolis sampling techniques and that is designed for treating large systems. The MMSG and DPPI-RAS schemes and partition functions were only calculated using the first code.

In keeping with most published work for calculations on this PES, we use nuclear masses, i.e., $m_{\mathrm{O}} = 29\,148.946\ m_{\mathrm{e}}$, and $m_{\mathrm{H}} = 1836.1527\ m_{\mathrm{e}}$, rather than atomic masses. Uncertainty estimates were obtained by performing ten independent calculations and represent 95% confidence intervals. The zero of energy for the reported partition functions, average energies, and average potential energies is the energy of the global minimum of the potential energy surface. As distributed, the PS PES has a zero of energy at the minimum energy configuration of about $-1.91\times10^{-6}\ E_{\mathrm{h}}$; it is common to see calculations performed on this PES without correcting for this small offset, but in the present calculations we use the energy of the minimum-energy configuration for the zero of energy.

IV.A. Accurate path integral and classical calculations

The computational methods used to obtain the accurate path integral benchmarks have been discussed in detail in earlier publications[26,28,30,37,42,47,48,72,73] and will only be briefly discussed here. Herein, we report barycentric partition functions, i.e., we do not include the contribution from overall translation, which depends on the system volume and can be obtained from an ideal gas treatment; we do, however, include the contribution from translation when reporting expectation values such as C_V and $\langle E\rangle$. Because the system is bounded, we restrict the integration volume of Eq. (7) to a finite domain, Γ, having a volume V_Γ. We integrate the paths with a $(P+1)$-point trapezoidal-rule quadrature with the quadrature nodes denoted $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{P+1}$ where $\mathbf{q}_{P+1} \equiv \mathbf{q}_1$; partition functions obtained with this integration are denoted $Q^{[P]}$. This leads to the expression

$$Q^{[P]} = \frac{Q_{\mathrm{fp}}(T)}{\sigma_{\mathrm{sym}}} \frac{\int_\Gamma d\mathbf{q}_{\mathrm{c}} \oint \mathscr{D}[\mathbf{q}(s)]\delta(\mathbf{q}_{\mathrm{c}} - \bar{\mathbf{q}})\,\Phi_{\mathrm{fp}}[\mathbf{q}(s)] \exp\left\{-\frac{\beta}{P}\sum_{i=1}^{P} V(\mathbf{q}_i)\right\}}{\int_\Gamma d\mathbf{q}_{\mathrm{c}} \oint \mathscr{D}[\mathbf{q}(s)]\delta(\mathbf{q}_{\mathrm{c}} - \bar{\mathbf{q}})\,\Phi_{\mathrm{fp}}[\mathbf{q}(s)]}, \tag{66}$$

where $\Phi_{\mathrm{fp}}[\mathbf{q}(s)]$ is the contribution of a particular path for a free particle (i.e., where the potential is zero)

$$\Phi_{\mathrm{fp}} = \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} ds \frac{\mu}{2}\left\{\frac{d\mathbf{q}(s)}{ds}\right\}^2\right] \tag{67}$$

and Q_{fp} is the contribution of the volume Γ to a free-particle partition function, i.e.,

$$Q_{\mathrm{fp}} = V_\Gamma\left(\frac{\mu}{2\pi\hbar^2\beta}\right)^{N/2}. \tag{68}$$

Note that because we importance sample paths based on a free particle distribution, which we construct using a sequential sectioning scheme,[47] we do not explicitly evaluate the integral in Eq. (67).

The benchmark calculations use uncorrelated sampling strategies; in particular, we use importance sampling involving Gaussian distributions of Jacobi coordinates sampled using a ziggurat method.[42,74] We also employ adaptively optimized stratified sampling using the value of the potential at the path centroid.[47] The harmonically guided variance reduction (HGVR) scheme[48] is employed to decrease the statistical uncertainties of the $Q^{[P]}$ values. The partition functions are extrapolated to the $P\to\infty$ limit using the enhanced same path scheme[37] wherein three simultaneously calculated values of $Q^{[P]}$ having different P values are fitted to the functional form

$$Q^{[P]} = Q^{[P=\infty]} + \frac{C_1}{P^2} + \frac{C_2}{P^3}, \tag{69}$$

where $Q^{[P=\infty]}$, C_1, and C_2 are fitting parameters. Calculations of the heat capacity and $\langle E\rangle$ are obtained by finite difference calculations using three simultaneous calculations of $Q^{[P]}$ performed at different β values. This is accomplished[41,73,75] by scaling the relative paths (i.e., the path locations relative to the centroid position) for the primary calculation by a factor of $(\beta'/\beta)^{1/2}$; partition functions calculated in this way are highly correlated allowing for finite difference calculations that are not degraded by statistical uncertainties. Finite difference calculations of the heat capacity are much less sensitive to the step size if calculated using

$$C_{\mathrm{rovib}} = k_{\mathrm{B}}\beta^2\frac{\partial^2 \ln Q}{\partial\beta^2} \tag{70}$$

rather than Eq. (64).[73]

When calculating expectations values such as $\langle V\rangle$ and coordinate averages we also accumulate an estimate of $Q^{[P]}$ without the HGVR correction for use with Eqs. (9) and (10). Expectation values of $\langle V\rangle$, $\langle E\rangle$, and C_{rovib} are extrapolated to $P\to\infty$ using an analog of Eq. (69), which, in the case of the last two properties, is equivalent to finite difference calculations with extrapolated values of Q. Expectation values

TABLE II. Parameters for the H_2O calculations for various T ranges.

Parameter	Note	$T = 200$ K	$T = 296$–500 K	$T = 750$–1500 K	$T = 2000$–4000 K
$N_{\text{centroids}}$	[a]	1×10^9	1×10^9	1×10^8	1×10^8
H–H domain (a_0)	[b]	1–5	1–5	1–5	1–5
O–HH domain (a_0)	[c]	0–4	0–4	0–4	0–4
H–H center	[d]	2.860858757 a_0	2.860858757 a_0	2.860858757 a_0	2.860858757 a_0
O–HH center	[e]	1.107157485 a_0	1.107157485 a_0	1.107157485 a_0	1.107157485 a_0
θ center	[f]	90°	90°	90°	90°
$\Delta_{\text{H–H}}$	[g]	0.5 a_0	0.5 a_0	0.5 a_0	0.5 a_0
$\Delta_{\text{O–HH}}$	[h]	0.3 a_0	0.3 a_0	0.3 a_0	0.3 a_0
Δ_θ	[i]	20°	20°	20°	20°
N^{Strata}	[j]	40	40	40	40
ΔE^{Strata} (kcal)	[j]	0.5	0.5	0.5	1.0
P^{max}	[k]	960	240	120	120

[a]The number of centroids sampled for the path integral calculations; for each centroid one path was sampled. For the ECP calculations 1×10^7 samples were used for all T values.
[b]The range of H–H distances that defines the sampling domain in unscaled bohr.
[c]The range of O–HH distances that defines the sampling domain in unscaled bohr.
[d]The center parameter in the Gaussian importance function for the H–H distance.
[e]The center parameter in the Gaussian importance function for the O–HH distance.
[f]The center parameter in the Gaussian importance function for the angle between the H–H and O–HH Jacobi vectors.
[g]The width parameter in the Gaussian importance function for the H–H distance.
[h]The width parameter in the Gaussian importance function for the O–HH distance.
[i]The width parameter in the Gaussian importance function for the angle between the H–H and O–HH Jacobi vectors.
[j]The stratified sampling bins samples into N^{strata} different strata based on the energy of the centroid configuration with an energy range for each stratum, except that last one, specified by ΔE^{Strata}. The final stratum includes all samples that exceed the energy cutoff of the other strata.
[k]The largest value of P, the number of path discretizations, included in the calculations.

for coordinate averages are reported for the largest value of P used in the calculations; these averages converge much more rapidly with P than properties such as the heat capacity, and the calculations use large values of P to achieve high accuracy, so the data reported are essentially at the $P \to \infty$ limit and therefore suitable for use as benchmarks. Numerical parameters used in the accurate path integral calculations are listed in Table II; a few parameters, most notably the number of Monte Carlo samples, vary with temperature. In each case, statistical uncertainties were obtained from results of ten independent calculations with these parameters. A much larger number of samples (a factor of 10–100 more depending on the temperature) were employed for the accurate path integral calculations than for the ECP calculations.

Classical results were also calculated by setting $P = 1$ and the number of samples to 10^7 but otherwise using the parameters listed in Table II.

IV.B. ECP calculations

DPPI-AS calculations were performed with two different codes. The first code is identical to the one discussed above for the calculation of the accurate path integral calculations, and identical parameters were used except for the number of Monte Carlo samples—the ECP calculations used 10^7 samples for all T. As we will see below, a key benefit of ECP schemes is that they have significantly lower sampling variances than accurate path integral methods; the use of the same code and the same importance sampling and stratification strategies for both the accurate path integral and ECP calculations facilitates a comparison of these sampling variances. Because this code calculates $Q(T)$ in addition to expectation values, we have the choice of calculating $\langle E \rangle$ using Eq. (64) or by finite difference calculations working directly with Q; similarly, we could calculate the heat capacity either by Eq. (64) or by finite differences of $\ln(Q)$. The results reported below are those obtained using Eqs. (61)–(65), although essentially equivalent results are obtained by working directly with Q. Unless stated otherwise, the tabulated DPPI-AS results were obtained with this code. The MMSG and DPPI-RAS methods were only implemented for testing and only in this first code.

The second code (MCCCS—MN: Monte Carlo for Complex Chemical Systems—Minnesota, version 12.1, http://www.chem.umn.edu/groups/siepmann/software.html) in which we have implemented the DPPI-AS scheme includes various biased[76–78] and Metropolis sampling algorithms, but only regular atom displacements are used for this validation study. This code only calculates operator averages but is designed to treat large systems, whereas the first code was only designed to treat small molecules. The two codes yield consistent results within the statistical uncertainties and both codes achieve a targeted uncertainty with far fewer samples than required in the accurate path integral calculations, although, for the case of an isolated water molecule, the Metropolis code requires a greater number of samples than the first code, which is highly optimized for small molecules. One complication with treating high temperature systems is the possibility of dissociation. At the highest temperature considered here, $T = 4000$ K, contributions to the partition function from energies above the dissociation limit are negligible, but if a molecule does dissociate when one is sampling via a Markov chain it will tend to remain dissociated, leading to highly skewed statistics. To avoid this problem here, we merely discard any move that leads to a

TABLE III. Partition functions, $Q(T)$, for H_2O calculated by various methods, with uncertainties given as 95% confidence intervals.

T	Eigensummation[a]	Accurate path integral	Classical
200	7.85240×10^{-14}	$(7.844 \pm 0.062) \times 10^{-14}$	$(2.55673 \pm 0.00060) \times 10^{-3}$
296	7.05241×10^{-9}	$(7.0551 \pm 0.0025) \times 10^{-9}$	$(1.49847 \pm 0.00033) \times 10^{-2}$
300	N/A	$(9.7199 \pm 0.0032) \times 10^{-9}$	$(1.59200 \pm 0.00032) \times 10^{-2}$
500	1.54265×10^{-4}	$(1.54274 \pm 0.00010) \times 10^{-4}$	$(1.59951 \pm 0.00028) \times 10^{-1}$
750	2.52469×10^{-2}	$(2.52490 \pm 0.00019) \times 10^{-2}$	$(1.00266 \pm 0.00016) \times 10^{0}$
1000	3.84914×10^{-1}	$(3.84926 \pm 0.00023) \times 10^{-1}$	$(3.69891 \pm 0.00044) \times 10^{0}$
1500	7.92834×10^{0}	$(7.92861 \pm 0.00063) \times 10^{0}$	$(2.34533 \pm 0.00069) \times 10^{+1}$
2000	$4.68926 \times 10^{+1}$	$(4.68918 \pm 0.00032) \times 10^{+1}$	$(8.7572 \pm 0.0021) \times 10^{+1}$
2500	$1.63786 \times 10^{+2}$	$(1.63809 \pm 0.00013) \times 10^{+2}$	$(2.4472 \pm 0.0076) \times 10^{+2}$
3000	$4.31097 \times 10^{+2}$	$(4.31401 \pm 0.00033) \times 10^{+2}$	$(5.6929 \pm 0.0014) \times 10^{+2}$
3500	$9.51044 \times 10^{+2}$	$(9.54203 \pm 0.00092) \times 10^{+2}$	$(1.16740 \pm 0.00021) \times 10^{+3}$
4000	$1.85586 \times 10^{+3}$	$(1.87471 \pm 0.00015) \times 10^{+3}$	$(2.18311 \pm 0.00037) \times 10^{+3}$
T	DPPI-AS	MMSG	DPPI-RAS
200	$(7.8546 \pm 0.0025) \times 10^{-14}$	$(9.2213 \pm 0.0030) \times 10^{-14}$	$(8.8981 \pm 0.0028) \times 10^{-14}$
296	$(7.1988 \pm 0.0030) \times 10^{-9}$	$(7.9819 \pm 0.0015) \times 10^{-9}$	$(7.7826 \pm 0.0016) \times 10^{-9}$
300	$(9.9247 \pm 0.0030) \times 10^{-9}$	$(1.0985 \pm 0.0030) \times 10^{-8}$	$(1.0712 \pm 0.0023) \times 10^{-8}$
500	$(1.59536 \pm 0.00047) \times 10^{-4}$	$(1.67656 \pm 0.00039) \times 10^{-4}$	$(1.64656 \pm 0.00046) \times 10^{-4}$
750	$(2.60282 \pm 0.00065) \times 10^{-2}$	$(2.66540 \pm 0.00090) \times 10^{-2}$	$(2.63334 \pm 0.00045) \times 10^{-2}$
1000	$(3.93907 \pm 0.00052) \times 10^{-1}$	$(3.99054 \pm 0.00077) \times 10^{-1}$	$(3.9583 \pm 0.0011) \times 10^{-1}$
1500	$(8.0268 \pm 0.0022) \times 10^{0}$	$(8.0667 \pm 0.0020) \times 10^{0}$	$(8.0357 \pm 0.0016) \times 10^{0}$
2000	$(4.7229 \pm 0.0011) \times 10^{+1}$	$(4.7343 \pm 0.0010) \times 10^{+1}$	$(4.7207 \pm 0.0014) \times 10^{+1}$
2500	$(1.64579 \pm 0.00043) \times 10^{+2}$	$(1.64751 \pm 0.0039) \times 10^{+2}$	$(1.64590 \pm 0.0038) \times 10^{+2}$
3000	$(4.32323 \pm 0.00090) \times 10^{+2}$	$(4.33106 \pm 0.00060) \times 10^{+2}$	$(4.32798 \pm 0.00079) \times 10^{+2}$
3500	$(9.5633 \pm 0.0023) \times 10^{+2}$	$(9.5693 \pm 0.0023) \times 10^{+2}$	$(9.5643 \pm 0.0033) \times 10^{+2}$
4000	$(1.87817 \pm 0.00037) \times 10^{+3}$	$(1.87859 \pm 0.00054) \times 10^{+3}$	$(1.87806 \pm 0.00027) \times 10^{+3}$

[a]Reference 71.

centroid configuration having a potential energy that exceeds the dissociation limit.

V. RESULTS

We have calculated results at 11 temperatures between 200 and 4000 K for which partition functions were previously calculated by Partridge and Schwenke[71] by summation of Boltzmann factors calculated with accurate eigenvalues; we also report results at one additional temperature, $T = 300$ K. Table III compares accurate quantum, classical, and approximate partition functions. The path integral results use a symmetry number of 2 (corresponding to statistical weighting of $\frac{1}{2}$ for both para and ortho states) whereas the eigensummation results use statistical weightings of $\frac{1}{4}$ and $\frac{3}{4}$ for the para and ortho states, respectively; however, at the temperatures considered the treatment of nuclear spin statistics is not expected to lead to significant differences. [The $Q(T)$ data of Partridge and Schwenke[71] were originally tabulated with reference to the ground state energy and with para and ortho weightings of 1 and 3 but have been multiplied here by a factor of $\frac{1}{4}\exp(-\beta E_0)$, where the ground state energy E_0 is 4638.39 cm^{-1}, for comparison to our partition functions that have a reference energy of the global minimum structure.] Comparisons for heat capacity and expectation values

TABLE IV. Calculations of C_V/k_B for H_2O, with uncertainties given as 95% confidence intervals.

T	Accurate path integral	Classical	DPPI-AS	MMSG	DPPI-RAS
200	N/A	6.01688 ± 0.00027	3.02960 ± 0.00051	2.98374 ± 0.00040	2.95870 ± 0.00043
296	3.060 ± 0.015	6.02548 ± 0.00025	3.02328 ± 0.00037	2.99559 ± 0.00049	2.98548 ± 0.00037
300	3.048 ± 0.017	6.02579 ± 0.00025	3.02427 ± 0.00053	2.99655 ± 0.00031	2.98738 ± 0.00022
500	3.23725 ± 0.00078	6.04369 ± 0.00021	3.16836 ± 0.00038	3.17060 ± 0.00032	3.18726 ± 0.00037
750	3.58317 ± 0.00024	6.06656 ± 0.00019	3.51511 ± 0.00044	3.53920 ± 0.00031	3.55419 ± 0.00036
1000	3.96435 ± 0.00009	6.09012 ± 0.00028	3.92982 ± 0.00044	3.95610 ± 0.00036	3.95993 ± 0.00043
1500	4.67024 ± 0.00011	6.14004 ± 0.00028	4.67467 ± 0.00057	4.69159 ± 0.00017	4.68590 ± 0.00054
2000	5.17941 ± 0.00023	6.19323 ± 0.00046	5.19041 ± 0.00057	5.19993 ± 0.00040	5.19070 ± 0.00045
2500	5.52723 ± 0.00031	6.24943 ± 0.00067	5.53669 ± 0.00075	5.54269 ± 0.00069	5.53864 ± 0.00076
3000	5.77702 ± 0.00035	6.31089 ± 0.00098	5.78748 ± 0.00068	5.78815 ± 0.00092	5.7853 ± 0.0013
3500	5.97254 ± 0.00086	6.37915 ± 0.00091	5.97707 ± 0.00078	5.9806 ± 0.0017	5.9773 ± 0.012
4000	6.14194 ± 0.00096	6.4602 ± 0.0011	6.1453 ± 0.0018	6.1505 ± 0.0055	6.1474 ± 0.0040

TABLE V. Calculations of $\langle E \rangle$ in kcal/mol for H_2O, with uncertainties given as 95% confidence intervals.

T	Accurate path integral	Classical	DPPI-AS	MMSG	DPPI-RAS
200	14.445 ± 0.032	2.387982 ± 0.000062	14.47212 ± 0.00019	14.40498 ± 0.00014	14.41926 ± 0.00013
296	15.0220 ± 0.0014	3.536656 ± 0.000095	15.04806 ± 0.00027	14.97421 ± 0.00017	14.98490 ± 0.00012
300	15.0474 ± 0.0013	3.584554 ± 0.000080	15.07212 ± 0.00017	14.99812 ± 0.00014	15.00876 ± 0.00020
500	16.29016 ± 0.00019	5.98300 ± 0.00016	16.29676 ± 0.00036	16.21746 ± 0.00023	16.23073 ± 0.00026
750	17.98197 ± 0.00009	8.99112 ± 0.00031	17.95191 ± 0.00025	17.88031 ± 0.00042	17.90219 ± 0.00033
1000	19.85598 ± 0.00017	12.01088 ± 0.00041	19.80097 ± 0.00046	19.74183 ± 0.00062	19.76851 ± 0.00062
1500	24.15964 ± 0.00032	18.0864 ± 0.0011	24.09403 ± 0.00092	24.05588 ± 0.00052	24.08089 ± 0.00052
2000	29.06938 ± 0.00027	24.2135 ± 0.0014	29.0119 ± 0.0011	28.98727 ± 0.00046	29.0183 ± 0.00012
2500	34.39847 ± 0.00056	30.3944 ± 0.0020	34.3522 ± 0.0014	34.33515 ± 0.00091	34.3495 ± 0.0021
3000	40.02116 ± 0.00057	36.6338 ± 0.0026	39.9816 ± 0.0016	39.9713 ± 0.0019	39.9830 ± 0.0019
3500	45.8609 ± 0.0011	42.9373 ± 0.0030	45.8271 ± 0.0021	45.8181 ± 0.0022	45.827 ± 0.0027
4000	51.8802 ± 0.0013	49.3157 ± 0.0028	51.8530 ± 0.0032	51.8489 ± 0.0055	51.853 ± 0.0035

are given in subsequent tables: C_V in Table IV, $\langle E \rangle$ in Table V, $\langle V \rangle$ in Table VI, $\langle R_{OH} \rangle$ in Table VII, and $\langle \theta_{HOH} \rangle$ in Table VIII. Figure 1 compares the ratio of the DPPI results to the accurate results for the two different code implementations to the ratio of classical and accurate QM calculations. Figure 2 compares HOH angle distributions calculated by accurate path integrals and the DPPI approximation.

VI. DISCUSSION

Table III shows that there is excellent agreement between the accurate path integral calculations of $Q(T)$ and the eigensummation results of Partridge and Schwenke and consequently indicates that our calculations, including the $P \rightarrow \infty$ extrapolations, are capable of very high accuracy. Between 500 and 2500 K the two methods agree to four significant figures; at the higher temperatures the path integral results are slightly larger than the eigensummation ones, which is consistent with truncation error in the latter (converging large numbers of accurate eigenvalues is an extremely computationally demanding task). The DPPI-AS partition functions agree quite well with the accurate results with the largest error being 3.4% at 500 K. We note that these results are, at low T, significantly better than our earlier[31] DPPI-AS calculations, which used an INM treatment based on internal coordinates rather than the Cartesian treatment adopted here. As expected, based on the earlier discussion and 1D benchmarks, the MMSG and DPPI-RAS schemes perform less well than the DPPI-AS calculations with the largest errors being 18% and 13%, respectively. The classical partition functions are more than ten orders of magnitude larger than the accurate ones at 200 K and are still 16% higher than the accurate ones at 4000 K.

It is interesting to note that Wong and Gao[56] report a 17% error for the partition function of H_2O at 200 K, although for a different PES, when they employ their method that uses GTFK to treat uncoupled INMs. Two factors may account for the superior performance of the DPPI-AS method compared to their approach. First, the use of the GTFK method for uncoupled instantaneous normal modes treats intramode anharmonicity very well, but completely neglects intermode anharmonicity, whereas the DPPI approach provides a balanced, although fairly approximate, treatment of both types of anharmonicity. Second, Wong and Gao[56] include all INMs in their quantum treatment, whereas, as discussed previously, we project out rotational motions and treat them classically to avoid errors associated with finite displacements along instantaneous rotational modes. As shown in the supplementary material, using all modes ($F = 3n^{\text{atoms}} - 3$) is less accurate than using $F = 3n^{\text{atoms}} - 6$.

The heat capacity calculations display the most sluggish convergence of all the full-path-integral quantities calculated

TABLE VI. Calculations of $\langle V \rangle$ in kcal/mol for H_2O, with uncertainties given as 95% confidence intervals.

T	Accurate path integral	Classical	DPPI-AS	MMSG	DPPI-RAS
200	6.738 ± 0.027	0.5995 ± 0.0001	6.9214 ± 0.0014	6.9047 ± 0.0018	6.9098 ± 0.0017
296	6.7737 ± 0.0041	0.8897 ± 0.0001	6.9153 ± 0.0013	6.9029 ± 0.0014	6.9065 ± 0.0015
300	6.7739 ± 0.0020	0.9018 ± 0.0001	6.9156 ± 0.0011	6.9024 ± 0.0014	6.9067 ± 0.0018
500	6.78219 ± 0.00039	1.5118 ± 0.0002	6.9200 ± 0.0016	6.9106 ± 0.0014	6.9144 ± 0.0014
750	6.89452 ± 0.00044	2.2843 ± 0.0003	6.9882 ± 0.0017	6.9827 ± 0.0015	6.9852 ± 0.0015
1000	7.09861 ± 0.00039	3.0685 ± 0.0004	7.1619 ± 0.0021	7.1579 ± 0.0012	7.1597 ± 0.0016
1500	7.80978 ± 0.00034	4.6728 ± 0.0011	7.8352 ± 0.0008	7.8334 ± 0.0008	7.8342 ± 0.0008
2000	8.85334 ± 0.00028	6.3286 ± 0.0015	8.8629 ± 0.0014	8.8625 ± 0.0012	8.8650 ± 0.0021
2500	10.13542 ± 0.00057	8.0384 ± 0.0020	10.1399 ± 0.0018	10.1392 ± 0.0012	10.1415 ± 0.0020
3000	11.59515 ± 0.00060	9.8065 ± 0.0026	11.5957 ± 0.0023	11.5973 ± 0.0012	11.5979 ± 0.0022
3500	13.1961 ± 0.0011	11.6389 ± 0.0030	13.1947 ± 0.0024	13.1947 ± 0.0022	13.1952 ± 0.0028
4000	14.9246 ± 0.0012	13.5460 ± 0.0028	14.9256 ± 0.0034	14.9278 ± 0.0046	14.9256 ± 0.0040

TABLE VII. Calculations of $\langle R_{OH} \rangle$, in Å, for H_2O, with uncertainties given as 95% confidence intervals.

T	Accurate path integral	Classical	DPPI-AS	MMSG	DPPI-RAS
200	0.97633 ± 0.00012	0.959784 ± 0.000003	0.976008 ± 0.000013	0.976410 ± 0.000019	0.976192 ± 0.000013
296	0.976528 ± 0.000006	0.960726 ± 0.000003	0.976436 ± 0.000014	0.976690 ± 0.000015	0.976502 ± 0.000013
300	0.976547 ± 0.000007	0.960766 ± 0.000002	0.976453 ± 0.000012	0.976694 ± 0.000014	0.976512 ± 0.000014
500	0.977177 ± 0.000002	0.962751 ± 0.000003	0.977193 ± 0.000019	0.977269 ± 0.000008	0.977154 ± 0.000007
750	0.978056 ± 0.000003	0.965269 ± 0.000006	0.978064 ± 0.000013	0.978070 ± 0.000016	0.978015 ± 0.000017
1000	0.979124 ± 0.000003	0.967844 ± 0.000007	0.979109 ± 0.000011	0.979100 ± 0.000019	0.979077 ± 0.000010
1500	0.981993 ± 0.000003	0.973113 ± 0.000006	0.981942 ± 0.000006	0.981920 ± 0.000009	0.981927 ± 0.000011
2000	0.985788 ± 0.000003	0.978594 ± 0.000011	0.985734 ± 0.000013	0.985725 ± 0.000015	0.985725 ± 0.000013
2500	0.990314 ± 0.000003	0.984292 ± 0.000015	0.990262 ± 0.000014	0.990268 ± 0.000014	0.990271 ± 0.000013
3000	0.995434 ± 0.000004	0.990270 ± 0.000013	0.995391 ± 0.000013	0.995393 ± 0.000020	0.995393 ± 0.000018
3500	1.001071 ± 0.000007	0.996540 ± 0.000014	1.001022 ± 0.000012	1.001026 ± 0.000021	1.001042 ± 0.000019
4000	1.007221 ± 0.000007	1.003184 ± 0.000018	1.007184 ± 0.000021	1.007191 ± 0.000026	1.007191 ± 0.000026

here; for example, at $T = 296$ K, the relative statistical uncertainty in the full-path-integral C_V is a factor of 13 times larger than that for Q; at $T = 200$ K the full-path-integral heat capacity was not even qualitatively converged, and calculations sufficient to converge this to high accuracy were deemed unaffordable. As seen in Table IV, all three ECP schemes yield high accuracy for C_V, with the largest deviations from the accurate results being 2.1%, 2.1%, and 2.4% for DPPI-AS, MMSG, and DPPI-RAS, respectively. Moreover, when calculated in an ECP calculation, the uncertainty in C_V is actually lower than that for Q, in marked contrast to the behavior observed in the full-path-integral calculations. The data in Figure 1 show that the deviations between the DPPI-AS and the accurate results are not a monotonic function of T with the maximum deviation found at 500 K, whereas the deviation is less than 1% at 300 K.

As seen in Table V, the performance of the ECP schemes in estimating $\langle E \rangle$ is even more impressive than it is for the heat capacity. The largest errors are 0.3%, 0.6%, and 0.4% for the DPPI-AS, MMSG, and DPPI-RAS schemes, respectively, and the sampling variances are markedly lower than those for Q or C_V. At $T = 296$ K, for example, the relative uncertainty in the DPPI-AS calculation of $\langle E \rangle$ was more than a factor of 20 lower than that for Q. As seen in Table VI, the accuracy of the ECP schemes for $\langle V \rangle$ is quite good, although not comparable to that for $\langle E \rangle$. The largest errors are 2.7%, 2.4%, and 2.5% for the DPPI-AS, MMSG, and DPPI-RAS methods, respectively. As can be seen from Figure 1, the deviations in $\langle V \rangle$ decrease monotonically with increasing T. The relative uncertainties for $\langle V \rangle$ are significantly higher than those observed for $\langle E \rangle$; in the DPPI-AS method the ratio of the uncertainties in $\langle V \rangle$ to the uncertainties in $\langle E \rangle$ range from a high of a factor of 17 at 750 K to a low of a factor of 3.7 at 4000 K. The higher statistical uncertainties in $\langle V \rangle$ as compared to $\langle E \rangle$ are likely due to the additional uncertainty related to the Monte Carlo evaluation of Eq. (36), by selection of $\mathbf{q}_d$ sample points, simultaneously with the sampling of the information at the centroid; the evaluation of $\langle E \rangle$ via Eqs. (61) and (62) only requires information from the centroid configuration. This situation can be ameliorated by sampling several values of $\mathbf{q}_d$ for each centroid sample.

As seen in Tables VII and VIII, the ECP schemes also perform well for coordinate averages. For example, the DPPI-AS data for $\langle R_{OH} \rangle$ for $T \geq 296$ K agree with the accurate path integral results to better than 10^{-4} Å and the data for $\langle \theta_{HOH} \rangle$ all agree to within 0.07°. The relative statistical uncertainties for the coordinate averages are significantly smaller than those observed for $\langle V \rangle$ and are roughly comparable to those observed for $\langle E \rangle$ and this indicates that the optimal number of times that values for $\mathbf{q}_d$ are sampled per centroid configuration depends on the operator considered. As seen in Fig. 2, the accurate path integral and DPPI-AS HOH angle distributions

TABLE VIII. Calculations of $\langle \theta_{HOH} \rangle$, in deg, for H_2O, with uncertainties given as 95% confidence intervals.

T	Accurate path integral	Classical	DPPI-AS	MMSG	DPPI-RAS
200	104.327 ± 0.024	104.4805 ± 0.0007	104.2614 ± 0.0037	104.3025 ± 0.0026	104.2973 ± 0.0019
296	104.3059 ± 0.0031	104.4667 ± 0.0012	104.2423 ± 0.0020	104.2720 ± 0.0027	104.2688 ± 0.0021
300	104.3016 ± 0.0022	104.4666 ± 0.0016	104.2395 ± 0.0029	104.2719 ± 0.0034	104.2669 ± 0.0024
500	104.2308 ± 0.0005	104.4377 ± 0.0009	104.1945 ± 0.0022	104.2140 ± 0.0026	104.2095 ± 0.0023
750	104.1715 ± 0.0006	104.3988 ± 0.0021	104.1541 ± 0.0034	104.1650 ± 0.0026	104.1614 ± 0.0016
1000	104.1355 ± 0.0009	104.3598 ± 0.0019	104.1276 ± 0.0035	104.1327 ± 0.0019	104.1315 ± 0.0020
1500	104.0830 ± 0.0006	104.2781 ± 0.0023	104.0814 ± 0.0028	104.0827 ± 0.0027	104.0835 ± 0.0032
2000	104.0245 ± 0.0009	104.1874 ± 0.0031	104.0256 ± 0.0027	104.0265 ± 0.0051	104.0272 ± 0.0033
2500	103.9477 ± 0.0013	104.0895 ± 0.0029	103.9495 ± 0.0025	103.9484 ± 0.0040	103.9485 ± 0.0029
3000	103.8472 ± 0.0011	103.9678 ± 0.0049	103.8477 ± 0.0056	103.8457 ± 0.0029	103.8513 ± 0.0048
3500	103.7187 ± 0.0025	103.8236 ± 0.0060	103.7211 ± 0.0066	103.7184 ± 0.0029	103.7182 ± 0.0047
4000	103.5571 ± 0.0019	103.6513 ± 0.0048	103.5613 ± 0.0071	103.5587 ± 0.0080	103.5582 ± 0.0067

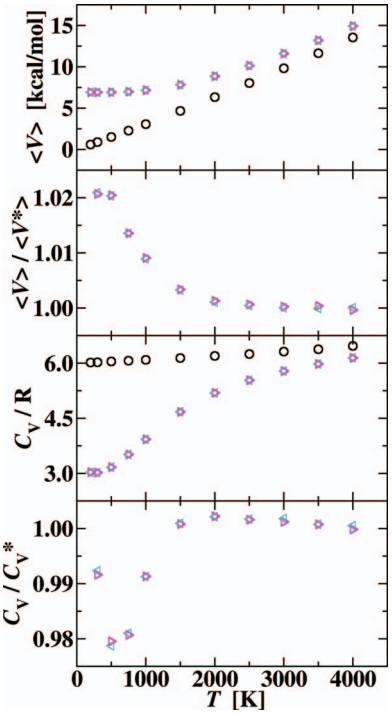

FIG. 1. Comparison of DPPI-AS calculations as implemented in two different Monte Carlo codes (cyan and magenta triangles) to the results of classical calculations (black circles); variables with asterisks are results of accurate path integral calculations.

at 300 and 1000 K are indistinguishable to plotting accuracy, which demonstrates that coordinate distributions may also be obtained with high accuracy.

Numerical results from classical calculations are included for comparison in Tables III–VIII. As expected, these results are qualitatively inaccurate at low T, but quantum effects are still significant even at the highest temperature considered (4000 K). It is interesting to note that the sampling variances for the ECP calculations are very similar to those of the classical calculations (both of which are much lower than would be the case for path integral calculations with an identical number of samples). Thus, the principle cost difference between the ECP and classical calculations is the overhead associated with the INM analysis.

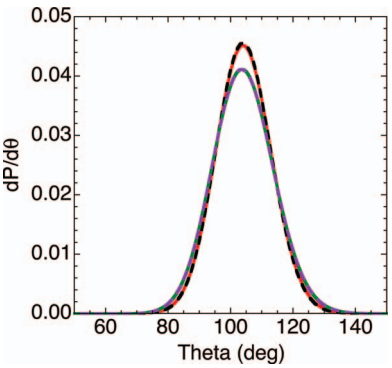

FIG. 2. HOH angle distributions: Accurate path integral at 300 K (solid red line), DPPI-AS at 300 K (dashed black line), accurate path integral at 1000 K (solid green line), DPPI-AS at 1000 K (dashed purple line).

On balance, the DPPI-AS scheme performs better than the other ECP schemes for expectation values, but its advantage over the other two ECP schemes is not as great as was observed for partition functions, at least for the case of an isolated water molecule.

VII. CONCLUDING REMARKS

We have discussed the extensions needed to the DPPI-AS method and to two related effective classical potential methods in order to calculate operator averages. Each of these schemes yields exact results for harmonic potentials and two of them include affordable approximate anharmonicity corrections. A key feature that these methods have in common and which distinguishes them from earlier effective classical potential methods is that the effective classical potentials can be calculated without the need to accurately evaluate high-dimensional numerical quadratures. This feature is critical for making the methods suitable for studying large chemical systems.

These three schemes were tested for several properties: the heat capacity, $\langle E \rangle$, $\langle V \rangle$, and coordinate averages. All three schemes perform well in comparison to accurate path integral calculations with the DPPI-AS scheme being, on balance, the best choice. This method was implemented in two different codes, a specialized code designed to compute absolute partition functions and properties for small systems and a Metropolis sampling code that is capable of treating large systems, and both implementations were shown to yield identical results within tight statistical uncertainties. The ECP schemes display significantly lower sampling variances than accurate path integral calculations and thus are much less computationally demanding. The high accuracy and ease of implementation of the ECP schemes suggests that they hold significant promise for approximately accounting for quantum effects in large systems.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation (NSF) through Grant Nos. CHE-09 56776 and CHE-1051396.

[1]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[2]R. P. Feynman, *Statistical Mechanics* (Benjamin, Reading, 1972).
[3]J. D. Doll and D. L. Freeman, J. Chem. Phys. **80**, 2239 (1980).
[4]D. Chandler and P. G. Wolynes, J. Chem. Phys. **74**, 4078 (1981).
[5]K. S. Schweizer, R. M. Stratt, D. Chandler, and P. G. Wolynes, J. Chem. Phys. **75**, 1347 (1981).
[6]J. D. Doll, J. Chem. Phys. **81**, 3536 (1984).
[7]B. J. Berne and D. Thirumalai, Annu. Rev. Phys. Chem. **37**, 401 (1986).
[8]R. D. Coalson, J. Chem. Phys. **85**, 926 (1986).
[9]R. D. Coalson, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **85**, 4567 (1986).
[10]J. D. Doll, R. D. Coalson, and D. L. Freeman, Phys. Rev. Lett. **55**, 1 (1986).
[11]H. Kono, A. Takasaka, and S. H. Lin, J. Chem. Phys. **88**, 6390 (1988).
[12]N. Makri and W. H. Miller, Chem. Phys. Lett. **151**, 1 (1988).
[13]T. L. Beck, J. D. Doll, and D. L. Freeman, J. Chem. Phys. **90**, 5651 (1989).
[14]J. Cao and B. J. Berne, J. Chem. Phys. **91**, 6359 (1989).
[15]R. D. Coalson, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **91**, 4242 (1989).
[16]N. Makri and W. H. Miller, J. Chem. Phys. **90**, 904 (1989).

[17]J. Lobaugh and G. A. Voth, J. Chem. Phys. **97**, 4205 (1992).
[18]R. Q. Topper and D. G. Truhlar, J. Chem. Phys. **97**, 3647 (1992).
[19]C. Chakravarty, J. Chem. Phys. **99**, 8038 (1993).
[20]W. Janke and T. Sauer, Chem. Phys. Lett. **201**, 499 (1993).
[21]R. Q. Topper, Q. Zhang, Y.-P. Liu, and D. G. Truhlar, J. Chem. Phys. **98**, 4991 (1993).
[22]D. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).
[23]J.-K. Hwang, Theor. Chem. Acc. **98**, 202 (1997).
[24]C. Chakravarty, M. C. Gordillo, and D. M. Ceperley, J. Chem. Phys. **109**, 2123 (1998).
[25]J.-K. Hwang, Theor. Chem. Acc. **101**, 359 (1999).
[26]S. L. Mielke, J. Srinivasan, and D. G. Truhlar, J. Chem. Phys. **112**, 8758 (2000).
[27]J. P. Neirotti, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **112**, 3990 (2000).
[28]J. Srinivasan, Y. L. Volobuev, S. L. Mielke, and D. G. Truhlar, Comput. Phys. Commun. **128**, 446 (2000).
[29]S. Jang, S. Jang, and G. A. Voth, J. Chem. Phys. **115**, 7832 (2001).
[30]S. L. Mielke and D. G. Truhlar, J. Chem. Phys. **114**, 621 (2001).
[31]S. L. Mielke and D. G. Truhlar, J. Chem. Phys. **115**, 652 (2001).
[32]F. V. Prudente, A. Riganelli, and A. J. C. Varandas, J. Phys. Chem. A **105**, 5272 (2001).
[33]C. Predescu and J. D. Doll, J. Chem. Phys. **117**, 7448 (2002).
[34]S. D. Bond, B. B. Laird, and B. J. Leimkuhler, J. Comput. Phys. **185**, 472 (2003).
[35]L. Capriotti, A. Cuccoli, A. Fubini, V. Tognetti, and R. Vaia, Phys. Status Solidi B **237**, 23 (2003).
[36]K. R. Glaesemann and L. E. Fried, J. Chem. Phys. **118**, 1596 (2003).
[37]S. L. Mielke and D. G. Truhlar, Chem. Phys. Lett. **378**, 317 (2003).
[38]C. Predescu, J. Math. Phys. **44**, 1226 (2003).
[39]C. Predescu and J. D. Doll, Phys. Rev. E **67**, 026124 (2003).
[40]C. Predescu, D. Sabo, and J. D. Doll, J. Chem. Phys. **119**, 4641 (2003).
[41]C. Predescu, D. Sabo, J. D. Doll, and D. L. Freeman, J. Chem. Phys. **119**, 12119 (2003).
[42]V. A. Lynch, S. L. Mielke, and D. G. Truhlar, J. Chem. Phys. **121**, 5148 (2004).
[43]M. W. Aviles and E. Curotto, J. Chem. Phys. **122**, 164109 (2005).
[44]K. R. Glaesemann and L. E. Fried, J. Chem. Phys. **123**, 034103 (2005).
[45]C. Predescu, J. Phys. Chem. B **110**, 667 (2005).
[46]S. Chempath, C. Predescu, and A. T. Bell, J. Chem. Phys. **124**, 234101 (2006).
[47]S. L. Mielke and D. G. Truhlar, J. Phys. Chem. A **113**, 4817 (2009).
[48]S. L. Mielke and D. G. Truhlar, J. Chem. Theory Comput. **8**, 1589 (2012).
[49]J. D. Doll and L. E. Myers, J. Chem. Phys. **71**, 2880 (1979).
[50]P. Zhang, R. M. Levy, and R. A. Friesner, Chem. Phys. Lett. **144**, 236 (1988).
[51]J. Cao and B. J. Berne, J. Chem. Phys. **92**, 7531 (1990).
[52]C. H. Mak and H. C. Andersen, J. Chem. Phys. **92**, 2953 (1990).
[53]G. A. Voth, Phys. Rev. A **44**, 5302 (1991).
[54]M. Messina, G. K. Schenter, and B. C. Garrett, J. Chem. Phys. **98**, 4120 (1993).
[55]C. E. Chao and H. C. Andersen, J. Chem. Phys. **107**, 10121 (1997).
[56]K.-Y. Wong and J. Gao, J. Chem. Phys. **127**, 211103 (2007).
[57]K.-Y. Wong and J. Gao, J. Chem. Theory Comput. **4**, 1409 (2008).
[58]R. Giachetti and V. Tognetti, Phys. Rev. Lett. **55**, 912 (1985).
[59]R. Giachetti and V. Tognetti, Phys. Rev. B **33**, 7647 (1986).
[60]R. Giachetti and V. Tognetti, J. Magn. Magn. Mater. **54-57**(2), 861 (1986).
[61]R. P. Feynman and H. Kleinert, Phys. Rev. A **34**, 5080 (1986).
[62]M. G. Martin and J. I. Siepmann, J. Phys. Chem. B **102**, 2569 (1998).
[63]W. L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. **110**, 1657 (1988).
[64]J. E. Adams and R. M. Stratt, J. Chem. Phys. **93**, 1332 (1990).
[65]H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*, 3rd ed. (World Scientific, Singapore, 2004).
[66]See supplementary material at http://dx.doi.org/10.1063/1.4772667 for numerical details, for tables of DPPI numerical results using various schemes for constructing INMs, and for tables of results for two 1D model potentials.
[67]H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics*, 2nd ed. (World Scientific, Singapore, 1995).
[68]R. T. Skodje and D. G. Truhlar, J. Phys. Chem. **85**, 624 (1981).
[69]E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955).
[70]M. Garcia-Viloca, C. Alhambra, D. G. Truhlar, and J. Gao, J. Chem. Phys. **114**, 9953 (2001).
[71]H. Partridge and D. W. Schwenke, J. Chem. Phys. **106**, 4618 (1997).
[72]V. A. Lynch, S. L. Mielke, and D. G. Truhlar, J. Phys. Chem. A **109**, 10092 (2005).
[73]K. E. Anderson, S. L. Mielke, J. I. Siepmann, and D. G. Truhlar, J. Phys. Chem. A **113**, 2053 (2009).
[74]G. Marsaglia and W. W. Tsang, J. Stat. Software **5**, 1 (2000).
[75]T. M. Yamamoto, J. Chem. Phys. **123**, 104101 (2005).
[76]B. Chen and J. I. Siepmann, J. Phys. Chem. B **104**, 8725 (2000).
[77]B. Chen and J. I. Siepmann, Theor. Chem. Acc. **103**, 87 (1999).
[78]J. I. Siepmann and D. Frenkel, Mol. Phys. **75**, 59 (1992).