Quantum dynamics using path integral coarse-graining (EP)

View Online   Export Citation   CrossMark

Félix Musil,[1] (ID) Iryna Zaporozhets,[1,2,3] (ID) Frank Noé,[1,2,3,4] (ID) Cecilia Clementi,[1,2,3,a)] (ID) and Venkat Kapil[5,b)] (ID)

AFFILIATIONS

[1] Department of Physics, Freie Universität Berlin, Arnimallee 12, 14195 Berlin, Germany
[2] Department of Chemistry, Rice University, Houston, Texas 77005, USA
[3] Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, USA
[4] Microsoft Research, Cambridge, United Kingdom
[5] Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Note: This paper is part of the 2022 JCP Emerging Investigators Special Collection.
[a)] E-mail: cecilia.clementi@fu-berlin.de
[b)] Author to whom correspondence should be addressed: vk380@cam.ac.uk

ABSTRACT

The vibrational spectra of condensed and gas-phase systems are influenced by the quantum-mechanical behavior of light nuclei. Full-dimensional simulations of approximate quantum dynamics are possible thanks to the imaginary time path-integral (PI) formulation of quantum statistical mechanics, albeit at a high computational cost which increases sharply with decreasing temperature. By leveraging advances in machine-learned coarse-graining, we develop a PI method with the reduced computational cost of a classical simulation. We also propose a simple temperature elevation scheme to significantly attenuate the artifacts of standard PI approaches as well as eliminate the unfavorable temperature scaling of the computational cost. We illustrate the approach, by calculating vibrational spectra using standard models of water molecules and bulk water, demonstrating significant computational savings and dramatically improved accuracy compared to more expensive reference approaches. Our simple, efficient, and accurate method has prospects for routine calculations of vibrational spectra for a wide range of molecular systems - with an explicit treatment of the quantum nature of nuclei.

Predictive simulations of thermodynamic and time-dependent properties of condensed and gas-phase systems lay the foundations of computational materials design and discovery.[1] Accurate modeling of many systems, such as those containing light nuclei like H, C, N, and O, must account for their quantum-mechanical behavior to include the zero-point motion of collective modes[2] and the tunneling of the system across classically inaccessible barriers.[3] The phenomena emerging from the quantum dynamics of light nuclei are ubiquitous in chemistry and material science of molecular systems, for instance, the relative diffusion of H_2 in clathrate hydrates[4] and kinetic isotope effects in porous organic crystals,[5] proton-transfer rates in molecular switches,[6] the red shift in the IR spectra of O–H stretch mode in ice,[7] and the characterization of (bio-)molecular systems using vibrational spectroscopy.[8]

Unfortunately, the exact description of quantum dynamics—requiring the solution of Schrödinger's equation—is possible only for the smallest of systems, such as molecules containing a few atoms.[9,10] Extension to larger systems requires "local" approximations or truncation of the interaction potential and/or an approximate solution of the many-body Schrödinger equation,[11–13] similar to the electronic structure problem. Alternatives that render an "approximate but full" quantum-mechanical treatment of all degrees of freedom are based on the (semi-)classical dynamics of the system.[14] In this context, imaginary time path integral (PI) simulations, although originally formulated for incorporating the quantum statistical effects,[15] are becoming increasingly popular for studying the approximate quantum dynamics of distinguishable particles.[14] The state-of-the-art PI approaches[16–19] neglect real-time

quantum coherence but include effects arising from the quantum statistical distribution of the nuclei.[20] These methods give a good description of the dynamical response of condensed phase systems, for instance, vibrational spectra of bulk water[21] and ice,[19] where quantum coherence effects last only over short times. On the other hand, predicting the vibrational response of molecules and clusters is still a challenge, the reasons being twofold. First, the artifacts of PI methods (arising from the neglect of real-time coherence and approximations to non-centroid Matsubara fluctuations[22]), such as spurious broadening, frequency shifts, and an incorrect temperature-dependence of the relative intensities of quantal modes,[23,24] become worse with a reduction in temperature.[25] Second, the computational cost of PI methods increases steeply with inverse temperature,[26] making them expensive to obtain a direct comparison with experiments. In the last few decades, many new approaches have aimed at individually improving the accuracy[27–30] or efficiency[19,31–33] of PI dynamical methods. These studies highlight the urgency for an accurate, efficient, and generally applicable method that can treat the quantum dynamics of molecules, clusters, and bulk systems at the same footing, and aid in modeling their vibrational response.

In this work, we combine PI and coarse-graining methods via machine learning to render the calculation of quantum vibrational spectra accurate and computationally affordable. We demonstrate the capabilities of our method on paradigmatic aqueous systems. Our approach builds on the state-of-the-art centroid molecular dynamics[16,34] (CMD) approach, which time-evolves the system *classically* on the free energy surface (FES) of the centroid of the imaginary time path—a modified PES that includes nuclear quantum effects. Its two key features, a temperature elevation (T_e) *ansatz* and path integral coarse-graining simulations (PIGS), ensure accuracy and computational efficiency. On the one hand, the T_e *ansatz* alleviates the spurious redshift in CMD, leading to an improvement in the accuracy of the vibrational spectra over the state-of-the-art methods. On the other hand, with PIGS, we machine-learn the centroid FES in a general manner, i.e., without making prior assumptions about the functional form of the system's PES, and use it to evolve the system classically. This method, referred to as T_e PIGS, enables the calculation of the IR spectra of aqueous systems, including nuclear quantum effects, in excellent agreement with numerically exact or more expensive reference methods. Furthermore, we demonstrate that our approach is transferable across phases and temperatures, allowing modeling of vibrational spectra at *cryogenic* temperatures at orders of magnitude lower computational cost using classical MD.

We first discuss CMD and introduce T_e PIGS in the context of a simple yet realistic anharmonic system that highlights the deficiencies of PI methods at low temperatures: a 2D radial Morse oscillator mimicking an O–H bond, described by the Hamiltonian,

$$\hat{H} = (2\mu)^{-1}\left[\hat{p}_x^2 + \hat{p}_y^2\right] + D\left[1 - e^{-\alpha\left(\sqrt{\hat{q}_x^2+\hat{q}_y^2}-r_0\right)}\right]^2, \quad (1)$$

where the parameters μ, D, α, and r_0 are defined in Sec. I A of the supplementary material. For simplicity, we ignore the coupling between the rotations and vibrations of the system, making this a simple but physically relevant one-dimensional problem that highlights the limitations of PI-based dynamical methods. The O–H bond has a large zero-point energy $E_0 \approx 1843$ cm^{-1} ≈ 2652 K and

a $0 \to 1$ transition energy $E_1 - E_0 \approx 3568$ cm^{-1} ≈ 5134 K (see Sec. I A of the supplementary material for more details), meaning that even at a temperature of 600 K, the system resides almost exclusively in its ground state, yet probes the anharmonic regions of the PES due to zero-point motion. As shown in Fig. 1(a), this results in a temperature-independent line position of the IR spectrum, at least up to 600 K, as well as a red shift of around 200 cm^{-1} with respect to the classical spectrum due to quantum nuclear motion.

The CMD approach[16] is based on the imaginary time path integral isomorphism[35] between the thermodynamics of a quantum system at inverse temperature β and a classical ring polymer made of P replicas of the system at $\beta_P = \beta/P$,

$$Z = \text{Tr}\left[e^{-\beta\hat{H}}\right] \propto \lim_{P\to\infty} \int d\mathbf{q}\, e^{-\beta_P\left[\sum_j U(\mathbf{q}^{(j)}) + U^{\text{spr}}(\mathbf{q})\right]}, \quad (2)$$

where $\mathbf{q} \equiv \{\mathbf{q}^{(1)}, \ldots, \mathbf{q}^{(P)}\}$ is a shorthand for positions of the P replicas of the system with $\mathbf{q}^{(j+P)} \equiv \mathbf{q}^{(j)}$ implied, $U(\mathbf{q})$ defines the classical PES, and $U^{\text{spr}}(\mathbf{q})$ is a temperature-dependent spring term[35] that connects consecutive replicas of the system. Within CMD, the system is time evolved *classically* on $U^{\text{CMD}}(\mathbf{q}_c; \beta)$, defined as the free energy surface of the centroid of the imaginary time path (modulo a constant) at β,

$$U^{\text{CMD}}(\mathbf{q}_c; \beta) = -\beta^{-1}\log\left\langle \delta\left(\frac{1}{P}\sum_{j=1}^{P}\mathbf{q}^{(j)} - \mathbf{q}_c\right)\right\rangle_\beta, \quad (3)$$

where $\sum_{j=1}^{P}\mathbf{q}^{(j)}/P$ is the centroid of the ring polymer and $\langle\cdot\rangle$ is an average over the path integral Hamiltonian in Eq. (2). The thermodynamic force acting on the centroid can be calculated *on the fly* from a constrained PI simulation at each CMD step,[16]

$$\mathbf{f}^{\text{CMD}}(\mathbf{q}_c; \beta) = -\nabla U^{\text{CMD}}(\mathbf{q}_c; \beta) = \left\langle \mathbf{f}_c\,\delta\left(\frac{1}{P}\sum_{j=1}^{P}\mathbf{q}^{(j)} - \mathbf{q}_c\right)\right\rangle_\beta, \quad (4)$$

where $\mathbf{f}_c = \sum_{j=1}^{P}\mathbf{f}^{(j)}/P$, and $\mathbf{f}^{(j)}$ is the physical force acting on the jth replica. Alternatively, the centroid can be evolved on its FES within a PI simulation in a partially adiabatic manner[36] by decoupling the centroid from the rest of the system. Dynamical properties, such as the IR spectrum, can be easily calculated via classical time correlation functions (TCFs) based on the centroid trajectory, similar to the classical MD.

As shown in Fig. 1(b), the IR spectrum of the O–H bond computed with CMD is in excellent agreement with the numerically exact result at 600 K. Unfortunately, at lower temperatures, the system experiences a spurious red shift that gets worse as the temperature is reduced. This well-known artifact is referred to as the "curvature problem"[25] and arises in cases where the ring polymer has a shape such that its centroid lies outside the ring (see Fig. 2), and thus does not represent the quantum-mechanical probability density of the system (as obtained by the replicas). Interestingly, the curvature problem is purely a structural artifact, and, as shown in the inset of Fig. 1(b), it can be diagnosed from PI trajectories: the misalignment of the distribution of the centroid at 300 and 150 K to the (physical) probability density obtained from the replicas. As mentioned above, another issue with CMD, and more generally with PI
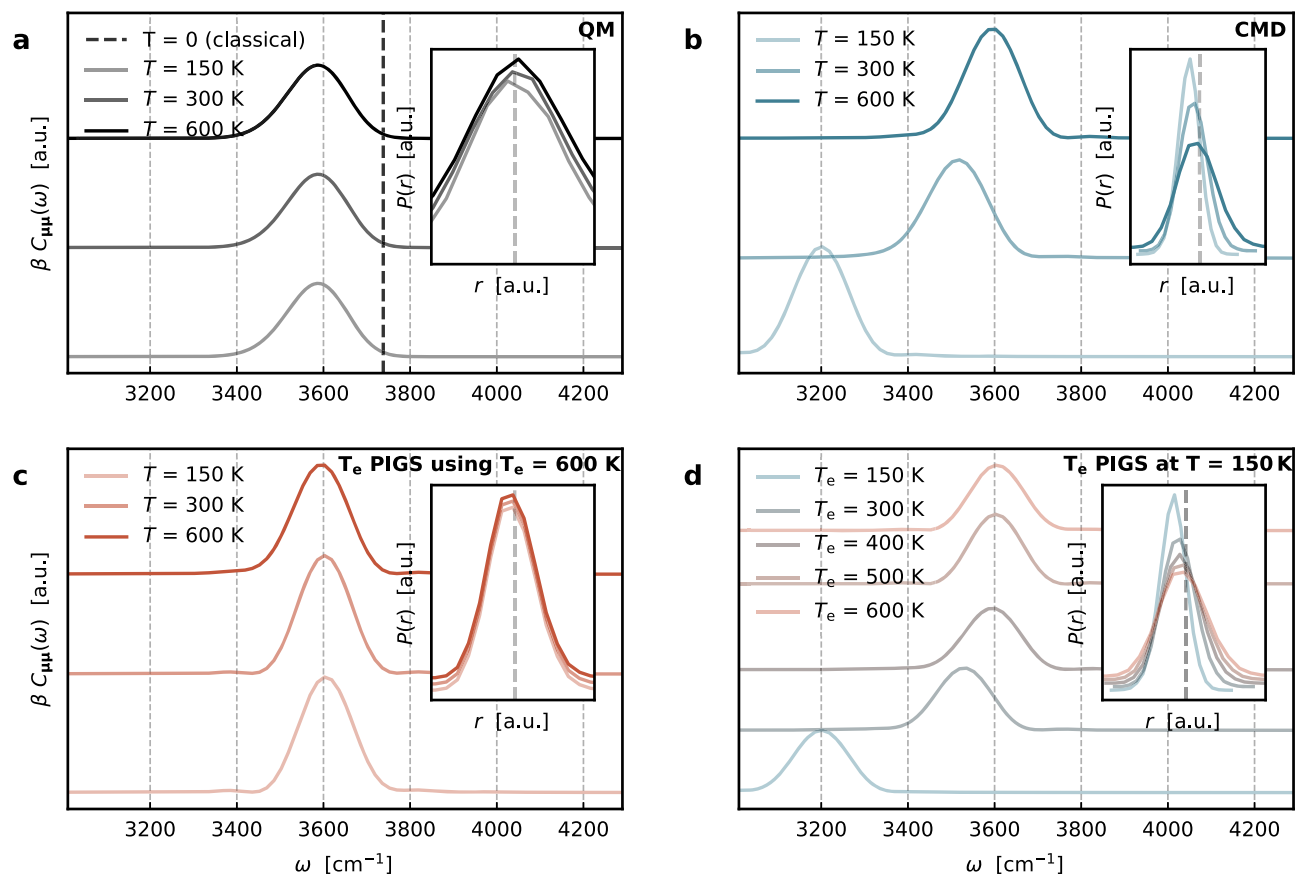
FIG. 1. Temperature-dependent IR spectrum of an O–H bond described by a Morse oscillator. The IR spectrum of a 2D Morse oscillator mimicking an O–H at 600, 300, and 150 K calculated by (a) solving the Schrödinger equation numerically (QM),[30] (b) using centroid molecular dynamics (CMD), and (c) T_e path integral coarse-graining simulations (PIGS) using T_e = 600 K. Panel (d) shows the dependence of the T_e PIGS IR spectrum on the temperature of the centroid free energy surface (T_e). Insets show the temperature-dependent radial probability densities calculated by using the respective methods, and the dashed gray vertical line is the mode of quantum-mechanical distribution. Note that the range of y axes is not kept the same to aid clarity.

methods, is that the required number of replicas scales inversely with the temperature and the maximum physical frequency of the system, $P \sim \beta\hbar\omega_{max}$.[37] Therefore, its computational cost increases steeply as the temperature is reduced. These challenges prevent investigations of systems at cryogenic temperatures where most experimental data are available.

Recently, Trenins et al.[30] have proposed a *quasi*-centroid molecular dynamics (QCMD) scheme that evolves the system on the FES of an *ad hoc* curvilinear function of replica positions—a *quasi*-centroid that does not "fall out" of the hull of the path integral. A careful selection of this function does not hamper the quantum Boltzmann statistics and results in a compact PI ring polymer that alleviates the curvature problem. This results in an excellent agreement of the vibrational spectrum of a molecule of water, liquid water, and proton-disordered hexagonal ice[38] with (numerically exact) reference methods. More recently, Fletcher et al.[33] have proposed an efficient *fast* QCMD scheme that avoids the costly *on-the-fly* quasi-centroid forces by precomputing an analytic FES based on a PI trajectory. This reduces the cost of predicting an accurate vibrational

spectrum to that of the classical MD, resulting in an improvement over standard PI methods for the calculation of vibrational spectra of exemplary molecular systems.[33] Unfortunately, the computational cost of these approaches still grows unfavorably with the temperature due to the need for a low-temperature PI trajectory for fitting the FES. Moreover, an extension to general systems requires a general/universal procedure to fit the FES of the centroid and careful knowledge of appropriate curvilinear coordinates that do not suppress the sampling of the physical regions of the configurational space.[39] Nonetheless, the advancements brought forth by (*fast*) QCMD[30,33] suggest that further improvements in CMD provide a promising route toward the accurate and efficient calculation of vibrational spectra.

Taking inspiration from *fast* implementations of CMD,[40,41] we propose path integral coarse-grained simulations (PIGS) that perform MD on a modified PES—including quantum nuclear effects. In particular, we leverage the recent developments in the definition of coarse-grained machine-learning potentials[42–44] and of high-order correlation functions[45,46] to obtain this modified potential by
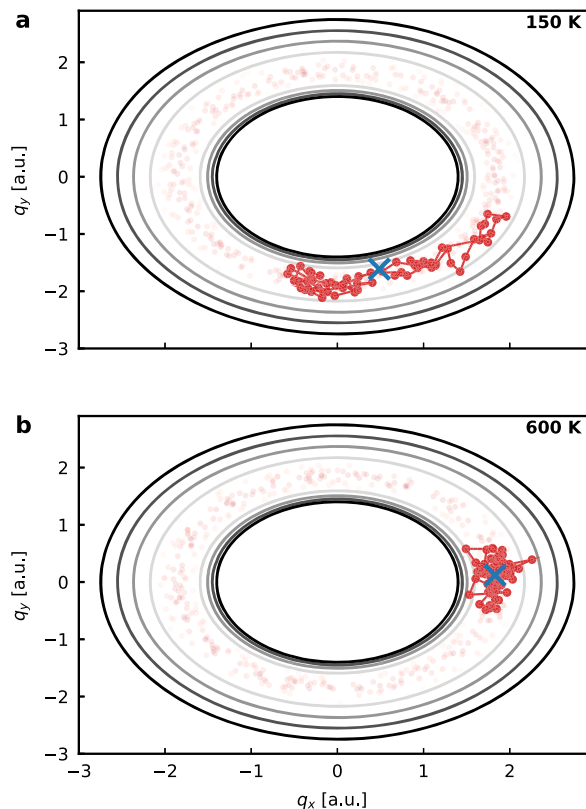
FIG. 2. The curvature problem in centroid molecular dynamics Contour plots of a 2D Morse radial potential up to half the dissociation energy (black to gray solid lines), the probability density of the system (red to white indicating high to low), and a snapshot of the ring polymer (red) and its centroid (blue) at (a) 150 K and (b) 600 K.

coarse-graining the imaginary time path integral to that of a classical system. In this work, we use PIGS to accelerate CMD in a general manner, i.e., it applies to systems exhibiting a wide range of interparticle interactions. To obtain the centroid FES at β_e, we use the force matching method[47,48] that is typically used to build thermodynamically consistent bottom-up coarse-grained models for macromolecular systems, i.e., the coarse-grained model reproduces the thermodynamic properties of the all-atom system projected onto the coarse-grained coordinates. It has been shown[48] that Eq. (3) can be recast as a variational principle—$U^{\mathrm{CMD}}(\mathbf{q}_c; \beta)$ corresponds to the minimum of the force matching functional,

$$U^{\mathrm{CMD}}(\mathbf{q}_c; \beta) = \arg\min_{G \in C(\mathbb{R}^n)} \left\langle \left\| \mathbf{f}_c + \nabla G(\mathbf{q}_c) \right\|^2 \right\rangle_\beta, \quad (5)$$

where the average is performed at inverse temperature β with the PI centroid constrained at $\mathbf{q}_c \in \mathbb{R}^n$, and the minimization is performed over the space $C(\mathbb{R}^n)$ of all real continuous functions $G : \mathbb{R}^n \to \mathbb{R}$. In this work, we optimize a machine-learning model as a surrogate for the potential of mean force to reproduce the thermodynamics of the centroid without having to explicitly simulate P replicas of the

system. In practice, we only learn the difference between the centroid FES and the classical PES, as it is done in Ref. 40. This keeps the amount of training data to a minimum and provides an appropriate prior in the absence of data or for collective modes that do not exhibit quantum nuclear effects.

The centroid potential of mean force associated with an atomic configuration is expressed as a sum of the classical PES associated with U and atom-centered contributions,

$$\tilde{U}^{\mathrm{CMD}}(\mathbf{q}_c; \boldsymbol{\theta}) = U(\mathbf{q}_c) + \sum_i A_{a_i}(\mathbf{q}_c - \mathbf{q}_{ci}; \boldsymbol{\theta}), \quad (6)$$

where $\boldsymbol{\theta}$ is a set of model parameters, $\mathbf{q}_c \equiv \{\mathbf{q}_{c1}, \ldots, \mathbf{q}_{cN}\}$ is a shorthand for the set of atomic positions of a structure, $\mathbf{q}_{ci}$ is the position of the ith atom of species a_i, and the A_{a_i} functions are parameterized using a machine-learning model that captures the multi-body interactions emerging from the coarse-graining procedure.[49] We model the atomic potentials of mean force, A_{a_i}, by representing the atomic environment with the normalized SOAP power spectrum[50–52] and pass these three-body features to a multi-layer perceptron with three hidden layers of width $[400, 200, 200]$ and the hyperbolic tangent activation function. The parameters of the model are obtained by minimizing the force matching loss,

$$\left\| \mathbf{f}_c + \nabla \tilde{U}^{\mathrm{CMD}}(\mathbf{q}_c; \boldsymbol{\theta}) \right\|^2, \quad (7)$$

over a set of centroid sample forces $\mathbf{f}_c$ and configurations $\mathbf{q}_c$ obtained from a PIMD simulation at β performed with the i-PI code[53] (see Sec. I C of the supplementary material for more details). The model and its training have been implemented in pytorch,[54] and the codes are available upon request.

We next propose a simple and physically motivated temperature elevation (T_e) *ansatz* that alleviates the curvature problem of CMD as well as eliminates the unfavorable temperature dependence associated with the computational cost of the vibrational spectra. We note that for a system in its ground state, i.e., $\hbar\omega \gg \beta^{-1}$, the IR spectrum—related by the dipole correlation function—is only *trivially* dependent on the temperature (see Sec. II A of the supplementary material). It is easy to show that after rescaling with the inverse temperature β, the Kubo-transformed time-correlation function is temperature-independent for a system in its ground state,

$$\beta \tilde{C}_{\dot{\mu}\dot{\mu}}(\omega; \beta) = (\hbar\omega)^{-1} \sum_{j,k} \langle j|\hat{\mu}|k\rangle\langle k|\hat{\mu}|j\rangle \delta(\omega - \hbar^{-1}(E_k - E_j)), \quad (8)$$

where E_j and E_k are the vibrational energies of the j-th and k-th eigenstates of the system.

As shown in Fig. 1(b), the CMD does not follow a temperature-independent line position because of the curvature problem at low temperatures. of the curvature problem at low temperatures. To avoid this artifact, we propose a T_e *ansatz*, i.e., we rewrite a time-correlation function in terms of a CMD time-correlation function computed at an elevated temperature T_e,

$$\tilde{C}_{\dot{\mu}\dot{\mu}}^{T_e}(\omega, \beta; \beta_e) = \frac{\beta_e}{\beta} \tilde{C}_{\dot{\mu}\dot{\mu}}^{\mathrm{CMD}}(\omega, \beta_e). \quad (9)$$

This *ansatz* is valid for any quantum system (of distinguishable particles) and can be used to improve PI-based methods that exhibit diminishing performance at low temperatures. To

ease the calculation of quantum dynamical properties for general (high-dimensional) systems, we approximate this relation as

$$\tilde{C}_{\mu\mu}^{T_e}(\omega,\beta;\beta_e) \approx Z^{-1} \mathcal{F}\left[\int d\mathbf{q}' e^{-\beta U^{CMD}(\mathbf{q}';\beta_e)} \boldsymbol{\mu}(0) \cdot \boldsymbol{\mu}(t) \right], \quad (10)$$

where $\mathcal{F}$ is a Fourier transform with respect to time so that it is simply estimated by evolving the centroid at β on a FES calculated at a high temperature β_e. This approximate form of Eq. (9) bears advantages such as it is exact in the harmonic and the classical limit, and does not require any posterior rescaling of the TCF. These limits also suggest that Eq. (10) should give a good description of the stiff modes, weakly coupled with the rest of the system, without perturbing the dynamics of the low-frequency (classical) modes. Notably, an elevated temperature leads to computational efficiency as it reduces the number of replicas needed to discretize the imaginary time path and improves sampling efficiency - similar to the use of a high temperature in the adiabatic free energy dynamics approach to improve sampling of a high-dimensional space. In this work, we perform fully adiabatic CMD by separately computing the FES at a higher temperature; however, we also plan to implement and study a partially adiabatic implementation of the T_e *ansatz* with CMD.

As shown in Fig. 1(b), the CMD IR spectrum remains in excellent agreement with the exact result at 600 K, and the radial distribution function of the centroid of the O–H bond is aligned with the physical distribution. These observations suggest that the system does not exhibit the curvature problem at 600 K, and thus, we test the T_e *ansatz* for $T_e = 600$ K, where $\beta_e = (k_B T_e)^{-1}$. As shown in Figs. 1(c) and 1(d), increasing the "elevated temperature" progressively improves the description of the IR spectrum at 150 K. Moreover, using $T_e = 600$ K leads to an excellent agreement of the temperature-dependent IR spectrum with the exact result. Note that for the Morse potential, there exists a wide window of suitable T_e (see Sec. II B of the supplementary material for more details), which is large enough to alleviate the curvature problem as well as small enough for the harmonic approximation in Eq. (10) to be valid, and we observed similar features with the water molecule and bulk water. Figure 1(c) shows that the line position of the predicted IR spectra and the radial distribution of the O–H bond are largely temperature-independent using $T_e = 600$ K, as expected for a system in its ground state. Finally, the number of replicas needed to calculate $U^{CMD}(\mathbf{q}'; \beta_e)$ is $P \approx \beta_e \hbar \omega_{max} < \beta \hbar \omega_{max}$, which is independent of β. Thus, within the T_e *ansatz*, the cost of simulating the quantum dynamics of a system in its ground state at β does not scale with temperature.

The workflow for computing the quantum vibrational spectrum of a system using PIGS and the T_e *ansatz* can be summarized as follow. First, we perform short PIMD simulations exploring a range of temperatures and select the *lowest* temperature T_e by checking for the alignment between the centroid and the physical radial distributions. In this study, we use radial distribution functions (RDFs) to check for this alignment; however, more general systems might require many-body correlation functions such as SOAP[50] or ACE[58] (of which RDFs form a subset), combined with the state-of-the-art dimensionality reduction schemes to compare the centroid and physical distributions. Then, we use Eqs. (6) and (7) to machine-learn the centroid FES at β_e as a sum of local

atom-centered components. In the final step, we predict the quantum dynamical properties by performing MD at the desired temperatures. These simple steps enable the development of an effective PES that includes quantum nuclear effects for dynamics—a FES trained on a *single* high-temperature PI trajectory—and that is *transferable* across temperatures. The computational details of all the simulations are described in Secs. I B, I D, and I E of the supplementary material.

We demonstrate the capabilities of the T_e PIGS approach by applying it to the IR spectrum of a water molecule—a challenging system that exhibits a strong red shift of the stretching modes due to zero-point motion and a fine temperature-dependent splitting of the vibrational peaks due to the coupling of rotations and vibrations. We study the IR spectrum at 150, 300, and 600 K using the well-known Partridge–Schwenke model,[55] which exhibits spectroscopic accuracy, and compare with experimental and numerically exact results.[60] We also compare with the state-of-the-art approaches such as CMD and thermostatted ring polymer molecular dynamics (TRPMD).[18]

The water molecule resides in its vibrational ground state at least up to 600 K but exhibits a sensitive dependence of the rovibrational splitting of the modes with the temperature, as seen from the numerically exact IR spectra[30] in Fig. 3. The TRPMD approach largely captures the correct line positions of the (envelopes of) the stretching and the bending bands but is artificially broadened.[18] The CMD approach gives a good description of the full IR spectrum at 600 K but is artificially red-shifted at 300 and 150 K due to the onset of the curvature problem. This is confirmed by observing in Fig. 3 that the centroid distributions at 300 and 150 K indicate a lower (unphysical) O–H bond length and misalignment with the physical radial distribution. Nevertheless, at 600 K, we do not observe any "symptoms" of the curvature problem, and thus, we use $T_e = 600$ K for the T_e PIGS approach.

As shown in Fig. 3, the T_e PIGS spectra at 600 K are in excellent agreement with the exact and independent CMD results, underlying the accuracy of the learned FES. More importantly, T_e PIGS describes the fundamental frequencies and the rotational splittings of the stretching and the bending modes at 300 and 150 K, a substantial improvement over state-of-the-art PI methods in terms of accuracy. However, we note that like CMD, T_e PIGS does not capture the subtle temperature dependence of the relative intensities of the bending mode with respect to the stretching mode. This is largely an artifact of PI-based approximate quantum dynamics methods, which lack quantum coherence in their dynamics[23] arising from the momenta not being drawn from their quantum Boltzmann distribution.[24] On the other hand, T_e PIGS is computationally efficient since the FES is estimated from a 100 ps PI trajectory with eight replicas and a timestep of 0.5 fs, suggesting computational gains of at least a factor of $500\times$, $1000\times$, and $2000\times$ at 600, 300, and 150 K, respectively, if one knows an appropriate T_e, compared to a 1 ns long CMD simulations needed to converge the TCFs. The overall cost for identifying T_e, namely, a set of 100 ps PIMD simulations across 100–600 K, was at least two orders of magnitude lower than that of a CMD simulation needed to calculate the spectrum at 150 K. Moreover, we expect this cost to be overestimated, due to the over-conservative nature of our benchmark. These could easily be made an order of magnitude less expensive using out-of-the-box accelerated-sampling methods[15] based on
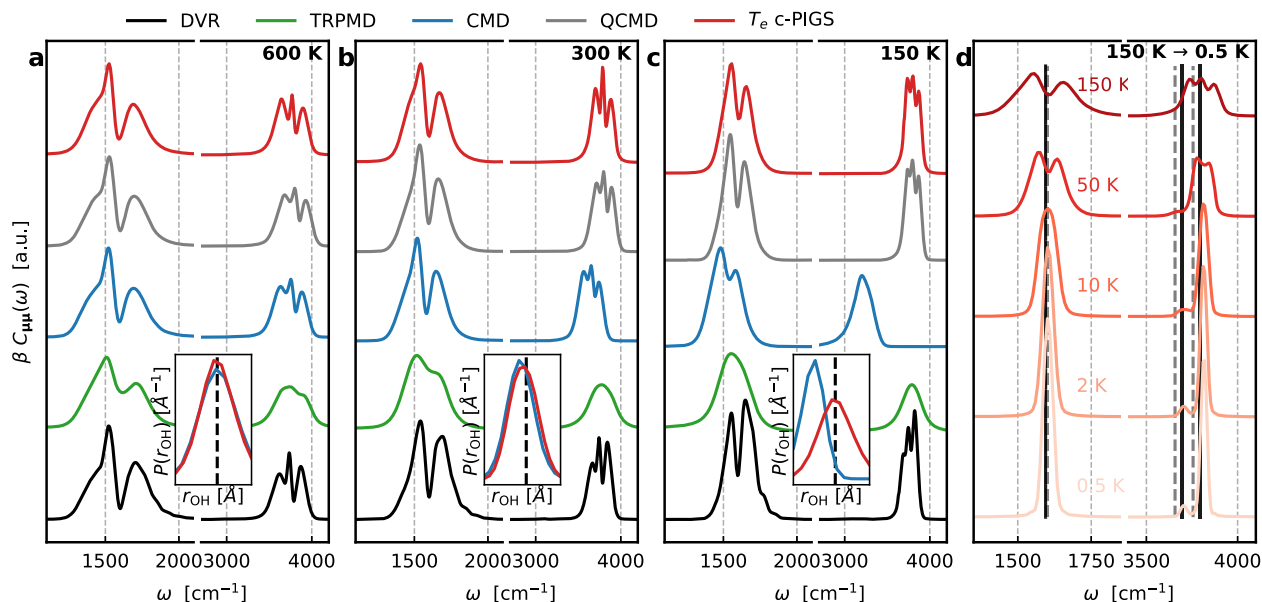
FIG. 3. Quantum dynamics of a water molecule. Comparison of the vibrational spectrum of a water molecule described by the Partridge–Schwenke model[55] at (a) 600, (b) 300, and (c) 150 K using discrete value representation (yielding numerically exact results) in black, the proposed T_e PIGS approach using the centroid free energy surface calculated at 600 K in red, CMD in blue, and TRPMD in green. The insets show the probability distribution of the O–H bond length calculated from the CMD and T_e PIGS, and the black dashed line indicates the mode of the distribution obtained from PIMD—constant across temperatures. Panel (d) displays the temperature-dependent IR spectrum of a water molecule from 600 K down to 0.5 K. The gray dashed line indicates the reference $0 \rightarrow 1$ transition frequency, while the black line is shifted by 40 cm^{-1} to reflect the results obtained with Matsubara dynamics.[22]

generalized Langevin equation thermostats[61–63] and high-order splittings of the Boltzmann operator[32,64–66] combined with replica-exchange[67] across 100–600 K. Other accelerated techniques such as ring polymer contraction and multiple time stepping[37,68] promise a classical computational cost for the calculation of quantum dynamical properties,[21,31] contingent to finding an inexpensive surrogate for the high-frequency modes of the system; however, they bear the inaccuracies of traditional path-integral methods that we address using the T_e *ansatz*.

To showcase the absence of computational scaling with temperature, we calculate the IR spectrum of a water molecule down to 0.5 K. A decrease in the temperature expectedly reduces the rotational splitting of the vibrational modes, and below 10 K, the system falls into its rotational ground state with well-resolved peaks for the three vibrational modes. This temperature is in excellent agreement with the experimentally known rotational constants of a water molecule, i.e., 13–35 K.[69] Furthermore, the frequencies of the ground state vibrational bending mode matches the experimental/exact results[60] up to 8 cm^{-1}—the resolution of calculated spectra—and those of the two stretching modes are expectedly blue-shifted by around 40 cm^{-1} in excellent agreement with the results obtained from Mastubara dynamics[22]—the true reference for PI-based dynamical methods. We emphasize that the calculation of the quantum mechanical spectrum at cryogenic temperatures using classical dynamics at this level of accuracy-to-cost ratio is unprecedented to our knowledge.

As a final test of our approach, we study the IR spectrum of condensed phase aqueous systems: bulk water at 300 K and hexagonal

ice at 150 K. The presence of inter-molecule or crystal modes that couple with high-frequency modes should constitute a challenge for T_e PIGS as has been the case with previous PI approaches.[27,28] An additional challenge is that ice melts below the onset temperature of the curvature problem, which could complicate the calculation of the FES at β_e. To circumvent this issue, we exploit the local nature of the FES [see Eq. (6)] and insights from deep inelastic neutron scattering experiments[7] that probe the quantum nuclear motion of atoms. These experiments suggest that the local potential felt by the nuclei due to quantum delocalization is short-ranged[70] and sensitive/unique to their local environments.[71] Given that Eq. (6) makes the fitted FES size-extensive and local environments of H and O atoms in ice are present in liquid water,[72] we conjecture that the T_e FES of hexagonal ice can be constructed from a high-temperature simulation of liquid water. We estimate the vibrational spectra using the q-TIP4P/f water potential[59] and a linear dipole moment surface. Although this model does not exhibit "experimental" or "spectroscopic" accuracy, it has been extensively used for comparing the performance of various PI-[19,30,38,73] and wavefunction-based[11,12] methods, and it exhibits a good agreement with the experimental spectra when combined with an appropriate dipole moment surface.[74]

As shown in Fig. 4, we obtain well-resolved spectra for both liquid water at 300 K and hexagonal ice at 150 K using a $T_e = 600$ K k_B FES fitted on a 10 ps PI simulation of liquid water. Our results are in good agreement with CMD, TRPMD, and QCMD for room-temperature liquid water, where the curvature problem is small.[75] In the case of hexagonal ice at 150 K, we see a quantitative and
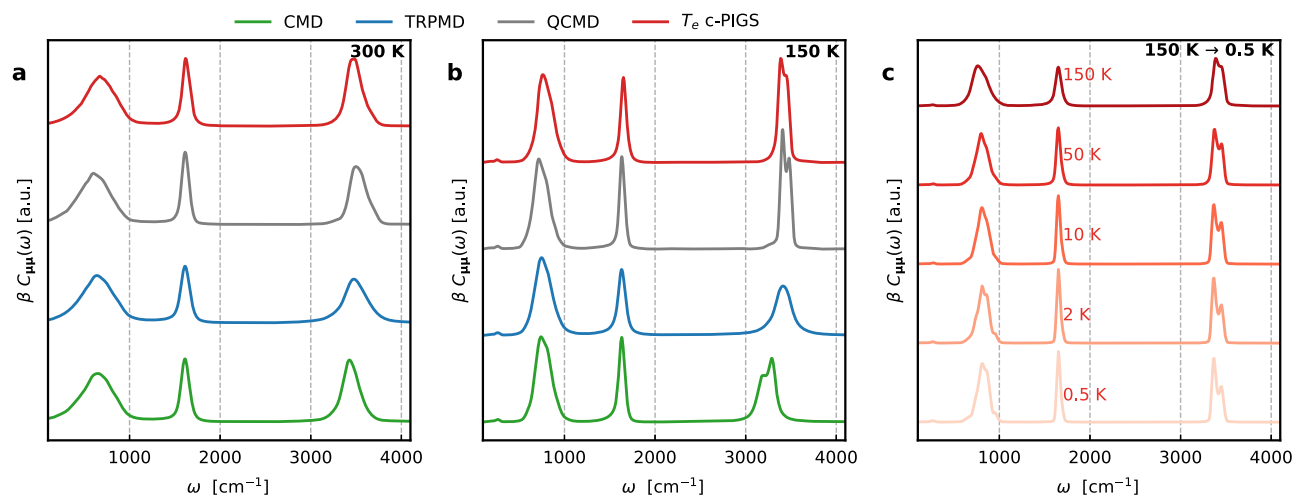
FIG. 4. Quantum dynamics of bulk water. Comparison of the vibrational spectrum of (a) liquid water at 300 K and (b) hexagonal ice at 150 K, described by a q-TIP4P/f[59] water model, calculated using QCMD in black, the proposed T_e PIGS approach using the centroid free energy surface calculated at 600 K in red, CMD in blue, and TRPMD in green. Panel (c) displays the T_e PIGS IR spectrum predictions for hexagonal ice from 150 K down to 0.5 K.

qualitative improvement in the description of the high-frequency band with respect to CMD and TRPMD, and an excellent agreement with QCMD.[30] In addition to the increased accuracy, our approach is over three orders of magnitude less expensive (as measured by the number of force evaluations) than the state-of-the-art approaches at 300 and 150 K. Furthermore, the absence of temperature-dependent scaling behavior of the cost allows us to also compute the IR spectrum of hexagonal ice all the way down to 0.5 K. As expected, we observe that the system resides in its ground state and does not exhibit any line shifts of the high-frequency modes.

In summary, we propose a new approach, T_e PIGS, that accurately simulates the quantum dynamics of light nuclei at the cost of classical MD by combining PI quantum mechanics, machine learning, and bottom-up coarse-graining. This combination is both timely and significant as exemplified by other recent works (which we discovered during the review process) that implement a *fast* version of CMD[76,77] using machine-learning potentials. We use a physically motivated temperature elevation *ansatz* that moves the system on the centroid FES obtained at a high temperature and fitted efficiently by using machine-learning-based coarse-graining approaches.[42–44] The T_e *ansatz* is exact in the high-frequency harmonic limit where the vibrational mode essentially exists in its ground state, as well as for low-frequency modes for which the ring polymer distribution collapses on the centroid. Furthermore, our study of condensed aqueous phase water shows that it also performs well for intermediate frequencies $\hbar\omega \sim \beta^{-1}$, for instance, the librational/rotational modes. These limits suggest that our approach could be useful for studying a wide range of systems with high-frequency modes that exhibit weak coupling to the rest of the system. We show that the T_e PIGS FES is *transferable* across temperatures and phases, as well as size extensive, meaning that it allows for further computational savings by learning the FES on a smaller system. We believe that our approach constitutes a substantial improvement in the accuracy over routinely used state-of-the-art

methods such as CMD and TRPMD. In addition, its low computational cost and the absence of its scaling behavior with temperature allow for accurate IR spectra predictions even at cryogenic temperatures, which are considered prohibitive with state-of-the-art methods.

The routine use of the T_e PIGS approach on materials and the chemical system will require a careful and thorough study of a diverse set of "difficult" systems, which could challenge the accuracy of the T_e *ansatz*. For instance, one could imagine highly fluxional molecules, systems exhibiting a near-continuum of strongly coupled high-frequency modes, or dynamical processes dominated by "rare" quantum tunneling events being testing cases probing the regimes in which the T_e *ansatz* is not exact or expected to be accurate. Similarly, the extension of T_e PIGS to general systems described by first-principles electronic structure methods will require the development of a hierarchical framework leveraging recent advances in active-learning strategies for the development of accurate and reliable machine-learning PES, enhanced sampling of the quantum Boltzmann distribution using accelerated path-integral methods, and, subsequently, the calculation of the centroid FES at an elevated temperature. In addition, these systems would also form an ideal test bed for understanding what range of elevated temperatures could be used to treat a frozen ground-state vibrational mode. Could a "universal" range of $T_e \in [500, 600]$ K. be used for chemical systems, displaying high-frequency modes in the 3000–5000 cm^{-1} regime, to completely avoid the process of selecting an elevated temperature for a general system? Finally, one could envisage the use of the PIGS approach for the calculations of the equilibrium properties. We plan to address all these directions in future studies. Overall, we believe that the simplicity, low cost, and high accuracy of T_e PIGS could open up prospects for routine modeling of quantum-vibrational spectra of general systems and direct comparisons with experiments, often performed at low temperatures.[78]

Section I of the supplementary material contains computational details of the DVR calculations (I A), the PIMD simulations (I B), the ML model (I C), CMD simulations (I D), and MD/TRPMD/PIGS (I E). Section II contains the derivation of the linear scaling of the Kubo-transformed TCF of a ground state system (II A) and a discussion on selecting a suitable T_e.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Félix Musil: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (supporting); Software (lead); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (equal). **Iryna Zaporozhets**: Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Validation (supporting); Visualization (supporting); Writing – review & editing (supporting). **Frank Noé**: Formal analysis (equal); Funding acquisition (equal); Project administration (supporting); Supervision (equal); Writing – review & editing (equal). **Cecilia Clementi**: Conceptualization (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Writing – original draft (supporting); Writing – review & editing (equal). **Venkat Kapil**: Conceptualization (lead); Data curation (equal); Formal analysis (lead); Funding acquisition (equal); Investigation (lead); Methodology (lead); Project administration (equal); Resources (equal); Software (supporting); Supervision (lead); Validation (equal); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in pigs.

REFERENCES

[1] K. N. Houk and F. Liu, Acc. Chem. Res. **50**, 539 (2017).

[2] L. Pereyaslavets, I. Kurnikov, G. Kamath, O. Butin, A. Illarionov, I. Leontyev, M. Olevanov, M. Levitt, R. D. Kornberg, and B. Fain, Proc. Natl. Acad. Sci. U. S. A. **115**, 8878 (2018).

[3] M. Benoit, D. Marx, and M. Parrinello, Nature **392**, 258 (1998).

[4] J. R. Cendagorta, H. Shen, Z. Bačić, and M. E. Tuckerman, Adv. Theory Simul. **4**, 2000258 (2021).

[5] M. Liu, L. Zhang, M. A. Little, V. Kapil, M. Ceriotti, S. Yang, L. Ding, D. L. Holden, R. Balderas-Xicohténcatl, D. He, R. Clowes, S. Y. Chong, G. Schütz, L. Chen, M. Hirscher, and A. I. Cooper, Science **366**, 613 (2019).

[6] Y. Litman, J. O. Richardson, T. Kumagai, and M. Rossi, J. Am. Chem. Soc. **141**, 2526 (2019).

[7] C. J. Burnham, G. F. Reiter, J. Mayers, T. Abdul-Redah, H. Reichert, and H. Dosch, Phys. Chem. Chem. Phys. **8**, 3966 (2006).

[8] M. Rossi, W. Fang, and A. Michaelides, J. Phys. Chem. Lett. **6**, 4233 (2015).

[9] G. Li and H. Guo, J. Mol. Spectrosc. **210**, 90 (2001).

[10] H. R. Larsson, M. Schröder, R. Beckmann, F. Brieuc, C. Schran, D. Marx, and O. Vendrell, "State-resolved infrared spectrum of the protonated water dimer: Revisiting the characteristic proton transfer doublet peak," Chem. Sci. **13**, 11119–11125 (2022).

[11] Y. Wang and J. M. Bowman, J. Chem. Phys. **134**, 154510 (2011).

[12] S. M. Gruenbaum, C. J. Tainter, L. Shi, Y. Ni, and J. L. Skinner, J. Chem. Theory Comput. **9**, 3109 (2013).

[13] C. Qu, P. L. Houston, R. Conte, A. Nandi, and J. M. Bowman, J. Phys. Chem. A **125**, 5346 (2021).

[14] S. C. Althorpe, Eur. Phys. J. B **94**, 155 (2021).

[15] T. E. Markland and M. Ceriotti, Nat. Rev. Chem. **2**, 0109 (2018).

[16] J. Cao and G. A. Voth, J. Chem. Phys. **100**, 5106 (1994).

[17] I. R. Craig and D. E. Manolopoulos, J. Chem. Phys. **121**, 3368 (2004).

[18] M. Rossi, M. Ceriotti, and D. E. Manolopoulos, J. Chem. Phys. **140**, 234116 (2014).

[19] V. Kapil, D. M. Wilkins, J. Lan, and M. Ceriotti, J. Chem. Phys. **152**, 124104 (2020).

[20] T. J. H. Hele, M. J. Willatt, A. Muolo, and S. C. Althorpe, J. Chem. Phys. **142**, 134103 (2015).

[21] O. Marsalek and T. E. Markland, J. Phys. Chem. Lett. **8**, 1545 (2017).

[22] G. Trenins and S. C. Althorpe, J. Chem. Phys. **149**, 014102 (2018).

[23] R. L. Benson and S. C. Althorpe, J. Chem. Phys. **155**, 104107 (2021).

[24] T. Plé, S. Huppert, F. Finocchi, P. Depondt, and S. Bonella, J. Chem. Phys. **155**, 104108 (2021).

[25] A. Witt, S. D. Ivanov, M. Shiga, H. Forbert, and D. Marx, J. Chem. Phys. **130**, 194510 (2009).

[26] F. Uhl, D. Marx, and M. Ceriotti, J. Chem. Phys. **145**, 054101 (2016).

[27] M. Rossi, V. Kapil, and M. Ceriotti, J. Chem. Phys. **148**, 102301 (2017).

[28] M. J. Willatt, M. Ceriotti, and S. C. Althorpe, J. Chem. Phys. **148**, 102336 (2018).

[29] J. Liu, J. Chem. Phys. **140**, 224107 (2014).

[30] G. Trenins, M. J. Willatt, and S. C. Althorpe, J. Chem. Phys. **151**, 054109 (2019).

[31] V. Kapil, J. VandeVondele, and M. Ceriotti, J. Chem. Phys. **144**, 054111 (2016).

[32] S. Shepherd, J. Lan, D. M. Wilkins, and V. Kapil, J. Phys. Chem. Lett. **12**, 9108 (2021).

[33] T. Fletcher, A. Zhu, J. E. Lawrence, and D. E. Manolopoulos, J. Chem. Phys. **155**, 231101 (2021).

[34] J. Cao and G. A. Voth, J. Chem. Phys. **99**, 10070 (1993).

[35] D. Chandler and P. G. Wolynes, J. Chem. Phys. **74**, 4078 (1981).

[36] T. D. Hone, P. J. Rossky, and G. A. Voth, J. Chem. Phys. **124**, 154103 (2006).

[37] T. E. Markland and D. E. Manolopoulos, J. Chem. Phys. **129**, 024105 (2008).

[38] R. L. Benson, G. Trenins, and S. C. Althorpe, Faraday Discuss. **221**, 350 (2019).

[39] C. Haggard, V. G. Sadhasivam, G. Trenins, and S. C. Althorpe, J. Chem. Phys. **155**, 174120 (2021).

[40] T. D. Hone, S. Izvekov, and G. A. Voth, J. Chem. Phys. **122**, 054105 (2005).

[41] F. Paesani, W. Zhang, D. A. Case, T. E. Cheatham, and G. A. Voth, J. Chem. Phys. **125**, 184507 (2006).

[42]J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N. E. Charron, G. de Fabritiis, F. Noé, and C. Clementi, ACS Cent. Sci. **5**, 755 (2019).

[43]B. E. Husic, N. E. Charron, D. Lemm, J. Wang, A. Pérez, M. Majewski, A. Krämer, Y. Chen, S. Olsson, G. de Fabritiis, F. Noé, and C. Clementi, J. Chem. Phys. **153**, 194101 (2020).

[44]J. Wang, S. Chmiela, K.-R. Müller, F. Noé, and C. Clementi, J. Chem. Phys. **152**, 194106 (2020).

[45]F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Chem. Rev. **121**, 9759 (2021); arXiv:2101.04673.

[46]O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K. R. Müller, "Machine learning force fields," Chem. Rev. **121**, 10142–10186 (2021).

[47]S. Izvekov and G. A. Voth, J. Chem. Phys. **123**, 134105 (2005).

[48]W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, J. Chem. Phys. **128**, 244114 (2008).

[49]J. Wang, N. Charron, B. Husic, S. Olsson, F. Noé, and C. Clementi, J. Chem. Phys. **154**, 164113 (2021).

[50]A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013); arXiv:1209.3140v2.

[51]J. P. Darby, J. R. Kermode, and G. Csányi, "Compressing local atomic neighbourhood descriptors," npj Comput. Mater. **8**, 166 (2022).

[52]F. Musil, M. Veit, A. Goscinski, G. Fraux, M. J. Willatt, M. Stricker, T. Junge, and M. Ceriotti, J. Chem. Phys. **154**, 114109 (2021); arXiv:2101.08814.

[53]V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, B. A. Helfrecht, P. Juda, S. P. Bienvenue, W. Fang, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, C. Corminboeuf, T. D. Kühne, D. E. Manolopoulos, T. E. Markland, J. O. Richardson, A. Tkatchenko, G. A. Tribello, V. Van Speybroeck, and M. Ceriotti, Comput. Phys. Commun. **236**, 214 (2019).

[54]A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox and R. Garnett (Curran Associates, Inc., 2019), pp. 8024–8035.

[55]H. Partridge and D. W. Schwenke, J. Chem. Phys. **106**, 4618 (1997).

[56]L. Rosso, P. Mináry, Z. Zhu, and M. E. Tuckerman, J. Chem. Phys. **116**, 4389 (2002).

[57]J. B. Abrams and M. E. Tuckerman, J. Phys. Chem. B **112**, 15742 (2008).

[58]R. Drautz, Phys. Rev. B **99**, 014104 (2019).

[59]S. Habershon, T. E. Markland, and D. E. Manolopoulos, J. Chem. Phys. **131**, 024501 (2009).

[60]Z. Gao, N. Giovambattista, and O. Sahin, Sci. Rep. **8**, 6228 (2018).

[61]M. Ceriotti, D. E. Manolopoulos, and M. Parrinello, J. Chem. Phys. **134**, 084104 (2011).

[62]M. Ceriotti and D. E. Manolopoulos, Phys. Rev. Lett. **109**, 100604 (2012).

[63]N. Mauger, T. Plé, L. Lagardère, S. Bonella, É. Mangaud, J.-P. Piquemal, and S. Huppert, J. Phys. Chem. Lett. **12**, 8285 (2021).

[64]A. Pérez and M. E. Tuckerman, J. Chem. Phys. **135**, 064104 (2011).

[65]V. Kapil, J. Behler, and M. Ceriotti, J. Chem. Phys. **145**, 234103 (2016).

[66]V. Kapil, J. Wieme, S. Vandenbrande, A. Lamaire, V. Van Speybroeck, and M. Ceriotti, J. Chem. Theory Comput. **15**, 3237 (2019).

[67]Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).

[68]M. Tuckerman, B. J. Berne, and G. J. Martyna, J. Chem. Phys. **97**, 1990 (1992).

[69]R. T. Hall and J. M. Dowling, J. Chem. Phys. **47**, 2454 (1967).

[70]L. Lin, J. A. Morrone, R. Car, and M. Parrinello, Phys. Rev. Lett. **105**, 110602 (2010).

[71]C. Andreani, G. Romanelli, and R. Senesi, J. Phys. Chem. Lett. **7**, 2216 (2016).

[72]P. Gasparotto, R. H. Meißner, and M. Ceriotti, J. Chem. Theory Comput. **14**, 486 (2018).

[73]M. Rossi, H. Liu, F. Paesani, J. Bowman, and M. Ceriotti, J. Chem. Phys. **141**, 181101 (2014).

[74]H. Liu, Y. Wang, and J. M. Bowman, J. Phys. Chem. B **120**, 1735 (2016).

[75]F. Paesani and G. A. Voth, J. Chem. Phys. **132**, 014105 (2010).

[76]T. D. Loose, P. G. Sahrmann, and G. A. Voth, "Centroid molecular dynamics can be greatly accelerated through neural network learned centroid forces derived from path integral molecular dynamics," J. Chem. Theory Comput. **18**, 5856–5863 (2022).

[77]C. Wu, R. Li, and K. Yu, Front. Mol. Biosci. **9**, 851311 (2022).

[78]D. Verma, R. M. P. Tanyag, S. M. O. O'Connell, and A. F. Vilesov, Adv. Phys.: X **4**, 1553569 (2019).