How Much Can Molecular Dynamics Improve over Harmonic Analysis on Vibrational Spectrum Predictions? Insights from Microcanonical Molecular Simulations of Anharmonic Water Clusters

Yuzhe Zhang,[†] Yiwen Wang,[†] Xi Xu,[‡] Zehua Chen,[†] and Yang Yang[*,†]

†*Theoretical Chemistry Institute and Department of Chemistry, University of Wisconsin-Madison, 1101 University Avenue, Madison, Wisconsin 53706, USA*

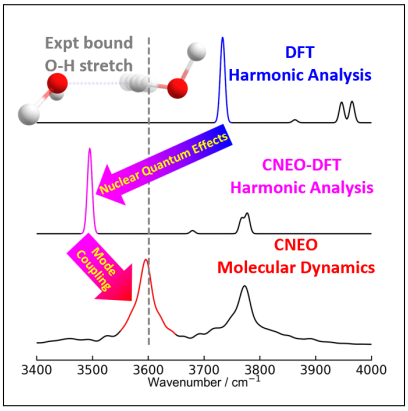
‡*Center for Advanced Materials Research, Beijing Normal University, Zhuhai 519087, China*

E-mail: yyang222@wisc.edu

Abstract

Vibrational spectroscopy is widely used to gain insights into structural and dynamic properties of chemical, biological and material systems. Thus, an efficient and accurate method to simulate vibrational spectra is desired. In this Letter, we propose a micro-canonical molecular simulation scheme for efficient calculations of vibrational spectra. Within the new scheme, we perform constrained nuclear-electronic orbital molecular dynamics simulations and accurately predict vibrational spectra of three challenging water clusters: neutral water dimer (H_4O_2), protonated water trimer ($H_7O_3^+$) and protonated water tetramer ($H_9O_4^+$). We find that in addition to nuclear quantum effects, vibrational mode coupling effects are also crucial for the accurate description of the vibrational motions of these highly anharmonic hydrogen bonded systems, which accounts for the large discrepancy between the vibrational frequencies obtained from molecular simulations and harmonic analyses.

TOC Graphic



Vibrational spectroscopy is a powerful tool for investigating physical properties and chemical processes in a variety of chemical, biological and material systems.[1–4] Combined with experimental measurements, theoretical predictions and simulations of vibrational spectra can provide valuable insights into both static and dynamic properties of a system, including structural information, interactions with the surrounding environment, reaction pathways and dynamic evolutions.

Many theoretical methods have been developed to compute vibrational spectra. Based on pre-calculated high quality potential energy surfaces (PESs), there are highly accurate methods such as vibrational self-consistent field/configuration interaction (VSCF/VCI),[5,6] multiconfiguration time-dependent Hartree (MCTDH),[7] and diffusion Monte Carlo (DMC).[8,9] These methods have been applied to a series of challenging molecular systems and provided accurate predictions of vibrational frequencies and successful identifications of subtle features in vibrational spectra.[10–13] In addition, there are path-integral based methods that can calculate vibrational spectra with reasonable accuracy and less computational cost. Examples often include centroid molecular dynamics (CMD),[14,15] ring polymer molecular dynamics (RPMD),[16] thermostatted RPMD (TRPMD)[17] and quasi-centroid molecular dynamics (QCMD).[18–20] Although CMD may suffer from curvature problems[21,22] and vibrational spectra calculated by RPMD may be contaminated by internal ring-polymer vibrational modes,[23,24] recently developed QCMD and TRPMD can mitigate these problems. Combined with accurate force fields or on-the-fly *ab initio* calculations, these path-integral based methods have been applied to gas phase and condensed phase systems and given vibrational spectra that agree well with experimental results,[17,25,26] although in general the direct combination with *ab initio* calculations remains expensive. A less expensive method is the vibrational second-order perturbation theory (VPT2).[27] VPT2 takes into account anharmonicities and mode coupling effects through a perturbative expansion around the local stationary geometry, and it has been very successful in predicting the vibrational frequencies of many molecular systems. However, VPT2 may face challenges in highly anharmonic

systems such as shared proton systems.[28]

Despite the accuracy of the aforementioned methods, currently in the field, the most widely used methods may still be harmonic analysis and molecular dynamics (MD) simulations based on either force fields or *ab initio* calculations. In harmonic analysis, the PES is assumed harmonic and vibrational frequencies are directly obtained from diagonalizing the mass-weighted Hessian matrix at an optimized geometry. In practical calculations, empirical scaling parameters are often applied to the harmonic frequencies to account for the lack of anharmonicities and nuclear quantum effects and thereby reducing the deviation from experimental results. Despite the popularity of empirical scaling, the choice of scaling parameters is *ad hoc* and even arbitrary, which can vary significantly depending on systems, vibrational modes, and underlying electronic structure methods. Compared to harmonic analysis, MD simulations, especially *ab initio* MD (AIMD) simulations, can often give better vibrational spectra with more enriched information. They can sample the anharmonic region of PESs as well as incorporate coupling effects between different vibrational modes through thermal motions. A drawback of MD simulations is that the performance may have a significant dependence on the simulation temperature. If the simulation temperature is low, the MD trajectory will be limited around the local harmonic region, leading to vibrational frequencies that are close to those from harmonic analysis, whereas higher simulation temperatures often lead to broaden and smeared peaks,[29] and furthermore potentially dissociate weakly bound molecular complexes.

In addition to inaccurate peak positions and lineshapes, another drawback of MD simulations is the requirement of long simulation time. A commonly used procedure to obtain vibrational spectra is first to thermally equilibrate the system with a canonical ensemble (NVT) simulation, subsequently perform many microcanonical (NVE) simulations starting from uncorrelated frames picked from the equilibrated NVT trajectory, and perform Fourier transform on obtained NVE autocorrelation functions before taking an average to obtain the spectrum.[30–32] Unfortunately, this procedure, which we call the NVT-NVE scheme there-

after, requires numerous *NVE* simulation runs to reach sufficient Boltzmann sampling and obtain converged spectrum results. Later, Kirchner and co-workers proposed to directly use equilibrated *NVT* trajectories to calculate vibrational spectra, which we call the direct *NVT* scheme thereafter.[29,33] This direct *NVT* scheme greatly reduced the simulation cost and worked well for bulk systems, although for gas phase small molecules with small couplings between a limited number of degrees of freedom, it often needs an impractically long time for modes to exchange energy and to reach equilibrium.[33] Furthermore, as has been shown in model systems, the direct *NVT* scheme may introduce artifacts caused by thermostats into vibrational spectra.[33]

Recently, our group developed a molecular dynamics framework based on the constrained nuclear-electronic orbital density functional theory (CNEO-DFT) to efficiently incorporate nuclear quantum effects into molecular simulations.[34,35] In CNEO-DFT, both electrons and key nuclei are treated quantum mechanically, but the classical molecular geometry picture is retained using the expectation values of quantum nuclear position operators together with the positions of classical nuclei. This treatment is justified by the physical intuition that quantum nuclei are still relatively localized in space in most chemical systems of interest. Therefore, in CNEO-DFT, constraints on the expectation values of quantum nuclear positions are imposed, which eventually leads to effective PESs that inherently incorporate nuclear quantum effects, especially zero-point effects. On these effective PESs, molecular dynamics, termed CNEO-MD, can be performed, and our group have found that with essentially the same computational cost as conventional DFT-based AIMD (less than 15% more expensive as benchmarked, see Supporting Information of ref 28), CNEO-MD significantly outperforms conventional AIMD in describing the vibrational modes with large hydrogen motion characters.[28,36]

However, like conventional AIMD, CNEO-MD may still suffer from the insufficient Boltzmann sampling problem either from the *NVT-NVE* or the direct *NVT* scheme. Additionally, in the past, CNEO-MD was mostly performed on simple small molecules, and it awaits fur-

ther tests on more complicated systems with much stronger mode coupling effects. In this Letter, we propose a new scheme for efficient calculations of vibrational spectra using only microcanonical molecular simulations. This new scheme, which we call the direct *NVE* scheme, is compared to both *NVT-NVE* and direct *NVT* schemes, and it is found to be able to obtain converged spectra with much less simulation time. We further calculate the vibrational spectra of three highly challenging water clusters: neutral water dimer (H_4O_2), protonated water trimer ($H_7O_3^+$) and protonated water tetramer ($H_9O_4^+$) and demonstrate the excellent performance of CNEO-MD in combination with the direct *NVE* scheme. We find that in these systems, CNEO-MD can significantly improve over CNEO-DFT harmonic analysis in vibrational spectra calculations, thanks to the incorporation of strong mode coupling effects. In contrast, conventional DFT-based AIMD may not necessarily perform better than DFT harmonic analysis because of the missing of nuclear quantum effects.

To introduce the direct *NVE* scheme, we first briefly review infrared spectra calculations with classical simulations. Infrared spectroscopy measures the frequency-dependent absorbance ($A(\omega)$) of infrared waves by a sample. The measured absorbance is directly proportional to the sample's absorption cross-section ($\sigma(\omega)$) and thus the rate of energy loss from the radiation ($\dot{E}_{rad}(\omega)$) through

$$A(\omega) = \sigma(\omega) l/V \tag{1}$$

and

$$\sigma(\omega) = \dot{E}_{rad}(\omega)/S(\omega) \tag{2}$$

where l is the length and V is the volume of the sample under radiation, and $S(\omega)$ is the energy flux, or the time-averaged amplitude of the Poynting vector.

With a continuous and perturbative electromagnetic field $E(t) = E_0 \cos \omega t$, the energy flux is

$$S(\omega) = \frac{n(\omega)\epsilon_0 c}{2\mu_r} E^2 \tag{3}$$

in which $n = \sqrt{\epsilon_r \mu_r}$ is the frequency-dependent refractive index, $c = 1/\sqrt{\mu_0 \epsilon_0}$ is the speed of light in vacuum, and the relative permeability μ_r is often approximated to be unity. The radiation energy loss rate can be computed using Fermi's golden rule:

$$\dot{E}_{rad}(\omega) = \big[R(\omega) - R(-\omega)\big]\hbar\omega = \frac{\omega E_0^2}{12\hbar}(1 - \mathrm{e}^{-\beta\hbar\omega}) \int_{-\infty}^{\infty} \mathrm{d}t \;\; \mathrm{e}^{-i\omega t} \langle \hat{\boldsymbol{\mu}}(0)\hat{\boldsymbol{\mu}}(t)\rangle_{\mathrm{qm}} \tag{4}$$

where $R(\omega)$ and $R(-\omega)$ are the excitation and de-excitation transition rates, $\beta = 1/k_{\mathrm{B}}T$, with k_{B} the Boltzmann constant and T the temperature of the system, and $\langle \hat{\boldsymbol{\mu}}(0)\hat{\boldsymbol{\mu}}(t)\rangle_{\mathrm{qm}}$ is the quantum dipole autocorrelation function.[37–39]

Putting everything together, the infrared spectra can be obtained with the Fourier transform of the quantum dipole autocorrelation function

$$A(\omega) = \frac{\omega \mu_r l}{6\hbar n(\omega)\epsilon_0 c V}(1 - \mathrm{e}^{-\beta\hbar\omega}) \int_{-\infty}^{\infty} \mathrm{d}t \;\; \mathrm{e}^{-i\omega t} \langle \hat{\boldsymbol{\mu}}(0)\hat{\boldsymbol{\mu}}(t)\rangle_{\mathrm{qm}} \tag{5}$$

In practice, because quantum correlation functions are usually difficult to calculate, classical autocorrelation functions obtained from classical simulations are more often used for spectra computation. These two types of autocorrection functions are not rigorously equal, but within the harmonic oscillator model, if the quantum and classical oscillators both reach equilibrium at the temperature T within a canonical ensemble, the Fourier transform of quantum and classical position autocorrelation functions only differ by a frequency-dependent prefactor,

$$\frac{\int_{-\infty}^{\infty} \mathrm{d}t \; \mathrm{e}^{-i\omega t} \langle \hat{x}(0)\hat{x}(t)\rangle_{\mathrm{qm}}}{\int_{-\infty}^{\infty} \mathrm{d}t \; \mathrm{e}^{-i\omega t} \langle x(0)x(t)\rangle_{\mathrm{cl}}} = \frac{\hbar\omega}{kT(1 - \mathrm{e}^{-\beta\hbar\omega})}. \tag{6}$$

Although this prefactor is derived from the harmonic oscillator model, it is often used to approximately account for the differences between quantum and classical position and dipole autocorrelation functions in real systems. Thus, with this approximation, the IR absorbance

$A(\omega)$ can be expressed with classical dipole autocorrelation functions

$$A(\omega) = \frac{B(\omega)\omega^2}{kT} \int_{-\infty}^{\infty} \mathrm{d}t \ \mathrm{e}^{-i\omega t} \langle \boldsymbol{\mu}(0)\boldsymbol{\mu}(t) \rangle_{\mathrm{cl}} \tag{7}$$

where $B(\omega) = \frac{\mu_r l}{6n(\omega)\epsilon_0 cV}$. For further simplification, integration by parts can be performed, and the absorbance can be expressed with the autocorrelation function of the time derivative of the dipole vector

$$A(\omega) \equiv \frac{B(\omega)}{kT} \int_{-\infty}^{\infty} \mathrm{d}t \ \mathrm{e}^{-i\omega t} \langle \dot{\boldsymbol{\mu}}(0)\dot{\boldsymbol{\mu}}(t) \rangle_{\mathrm{cl}} \tag{8}$$

This equation implies that in principle, in order to obtain the IR absorption spectra, one needs to generate enough sampling from the canonical ensemble, which is usually performed by running an NVT trajectory at a given temperature, and then perform NVE simulations starting from each sampled configurations. The final absorption spectra can be obtained by taking the average of the Fourier transformed dipole derivative autocorrelation function, with a prefactor that is essentially frequency independent if the refractive index $n(\omega)$ does not significantly vary with frequency. Equation 8 is the theoretical foundation for the calculations of vibrational spectra with the NVT-NVE scheme. This scheme may seem easy to perform, but to achieve enough sampling from the canonical ensemble and thus converged spectra, the number of configurations needed as well as subsequent NVE trajectories is huge even in model systems.[33]

In order to reduce the computational cost, Kirchner and coworkers introduced the direct NVT scheme for spectrum calculations.[29,33] Instead of an NVT simulation followed by many NVE simulations, the direct NVT scheme only uses one equilibrated NVT trajectory to calculate the ensemble averaged dipole derivative autocorrelation function. In principle, this scheme is not theoretically rigorous because the thermostat will affect the evolution of the system and thus bring artifacts into the calculated spectra, but in practical simulations, especially in condensed phase simulations, if a weak thermostat is used, the artifacts caused by thermostat can be neglected.[33] Despite its success, a scheme that is both computationally

inexpensive and theoretically more rigorous is highly desirable.

In order to achieve the aforementioned goal,, here in this Letter, we propose the direct *NVE* scheme. Different from the previous two schemes, which are built on calculating the dipole derivative autocorrelation functions for a canonical ensemble equilibrated at a given temperature T, we provide the theoretical foundation of the direct *NVE* scheme by revisiting the autocorrelation function of the harmonic model within a microcanonical ensemble.

For a microcanonical ensemble of harmonic oscillator with energy E, the classical position autocorrelation function is

$$\langle x(0)x(t)\rangle_{\mathrm{cl}}^{NVE} = \frac{E}{m\omega^2}\cos\omega t \tag{9}$$

This is highly similar to the classical position autocorrelation function within the canonical ensemble at temperature T, which is

$$\langle x(0)x(t)\rangle_{\mathrm{cl}}^{NVT} = \frac{k_B T}{m\omega^2}\cos\omega t \tag{10}$$

These two equations indicate that within the harmonic oscillator model, the classical position autocorrelation functions of the microcanonical ensemble and the canonical ensemble only differ by a prefactor $k_B T/E$. Previously the classical autocorrelation functions from a canonical ensemble are widely used to approximate the quantum autocorrelation functions, then now a natural question to ask is whether the autocorrelation functions from a microcanonical ensemble can be used for this approximation. If so, we can easily calculate IR spectra with

$$A(\omega) = \frac{B(\omega)}{E}\int_{-\infty}^{\infty} \mathrm{d}t\ \mathrm{e}^{-i\omega t}\langle\dot{\boldsymbol{\mu}}(0)\dot{\boldsymbol{\mu}}(t)\rangle_{\mathrm{cl}}^{NVE}. \tag{11}$$

This equation seems trivially similar to equation 8, but there are two key differences. First, $k_B T$ is in the denominator in equation 8 whereas E is in the denominator here. Second, the autocorrelation function in equation 8 needs to be obtained from a canonical ensemble whereas here it needs to be obtained from a microcanonical ensemble. This second difference is especially beneficial because for a given system, the phase space of a microcanonical

ensemble with a fixed energy E is significantly smaller than that of a canonical ensemble with a fixed temperature T. As such, the sampling need for a microcanonical ensemble simulation is much smaller.

Based on equation 11, the most straightforward way to obtain IR spectra is to run an *NVE* MD trajectory and then perform Fourier transform on the obtained dipole derivative autocorrelation function. In practice, to further accelerate the sampling and to reduce the wall time, we can run multiple *NVE* trajectories in parallel with the same energy but different initial phase space configurations and subsequently average the results from multiple trajectories. This is the direct *NVE* scheme we are introducing in this Letter. Later, we will show that the direct *NVE* scheme performs better than both *NVT*-*NVE* and direct *NVT* schemes in obtaining converged spectra with minimal simulation time.

One should notice that in conventional *NVT* simulations, with the help of thermostats, all modes will equilibrate to the same temperature and there is no need of special treatment on each specific mode. However, in the direct *NVE* scheme, although in principle we only need to fix the total energy of the system and allow the energy to flow naturally between different modes during the simulation, in small gas-phase molecules with relatively weak mode couplings, the energy flow is slow, and it could take a long time to reach ergodicity and thus energy equipartition. This is a main drawback of the direct *NVE* scheme. To alleviate this problem, in practice, we directly start the *NVE* simulations from configurations in which all modes have the same energy. It is admitted that this practice will artificially favor those configurations close to energy equipartition, but we will show later in this Letter that the simulated spectra with this practice are reasonably accurate compared to experimental results in a series of highly challenging water cluster systems.

We also note that in conventional *NVT* simulations, the simulation temperature can have notable effects on the final spectra, including both peak positions and peak intensities. This is in contrast to quantum simulations, in which although peak intensities will change with temperature, peak positions will not. The temperature dependence of peak positions in clas-

sical simulations is a known artifact, but it is a common way to account for anharmonicities and mode coupling effects through temperature elevation, which are missing in harmonic analysis. Similarly, since the direct *NVE* scheme is also based on classical simulations, the final spectra will depend on the simulation energy E. A finite energy E will naturally lead to the incorporation of anharmonic effects and mode coupling effects, which we will show later to be crucial for accurate spectra prediction in the challenging water cluster systems.

Recently, our group developed the CNEO molecular dynamics framework to incorporate nuclear quantum effects, especially zero-point effects, in practical molecular simulations. The key difference between CNEO-MD and conventional DFT-based *ab initio* molecular dynamics is that instead of using DFT, CNEO-DFT is used to evaluate energies and forces on the fly during the CNEO-MD simulations.[36]

CNEO-DFT belongs to the multicomponent quantum theory, which treats more than one type of particles, for example, electrons and nuclei, quantum mechanically .[40–43] The key difference between CNEO and other multicomponent methods is that it assumes quantum nuclei to be distinguishable particles and assigns classical positions to quantum nuclei by imposing a constraint on the expectation value of the position operator for each quantum nucleus,[34,35]

$$\langle \boldsymbol{r}_I \rangle = \boldsymbol{R}_I, \quad \forall I \tag{12}$$

Here $\boldsymbol{R}_I$ denotes the classical nuclear position of the I-th nucleus and $\langle \boldsymbol{r}_I \rangle$ denotes the quantum expectation value of the position operator for the I-th nucleus. This treatment is justified by the physical intuition that in most cases that chemists are interested in, nuclei are still relatively localized in space compared with electrons, and therefore expectation positions of quantum nuclei together with positions of classical nuclei can be used to denote molecular geometries. The CNEO ground state for each geometry can be obtained by minimizing the energy of the multicomponent system subject to the constraints, which lead to additional

terms in the Lagrangian

$$\sum_I \boldsymbol{f}_I \cdot (\langle \boldsymbol{r}_I \rangle - \boldsymbol{R}_I) \tag{13}$$

where $\boldsymbol{f}_I$ is the Lagrange multiplier of the I-th quantum nucleus. Making the Lagrangian stationary with respect to electronic and quantum nuclear orbital variations leads to a set of coupled nuclear and electronic Kohn-Sham equations. These equations can be solved self-consistently, and the converged orbitals can be used to evaluate the CNEO-DFT energy.[34,35] The CNEO-DFT energy is a function of both expectation positions of quantum nuclei and positions of classical nuclei, and the energy gradients with respect to both of them can be calculated.[35] Notably, the Lagrangian multiplier $\boldsymbol{f}_I$ is the classical force acts on the I-th quantum nucleus in the complete basis set limit, but in practice with a finite basis set, the Pulay correction[44] for quantum nuclei is needed and has been implemented.[35]

With CNEO-DFT energies and forces obtained on-the-fly, CNEO-MD simulations can be performed.[36] The theoretical foundation of CNEO-MD is the constrained minimized energy surface molecular dynamics (CMES-MD) framework.[45,46] In CMES-MD, the equations of motion highly resembles classical Newton's equations:

$$m\frac{\mathrm{d}\langle \boldsymbol{x} \rangle}{\mathrm{d}t} = \langle \boldsymbol{p} \rangle \tag{14}$$

and

$$\frac{\mathrm{d}\langle \boldsymbol{p} \rangle}{\mathrm{d}t} \approx -\nabla_{\langle \boldsymbol{x} \rangle} V^{\mathrm{CMES}}(\langle \boldsymbol{x} \rangle) \tag{15}$$

where $\langle \boldsymbol{x} \rangle$ and $\langle \boldsymbol{p} \rangle$ are expectation position and expectation momentum, respectively, and $V^{\mathrm{CMES}}(\langle \boldsymbol{x} \rangle)$ is the CMES obtained from constrained energy minimization with nuclear expectation positions fixed. In practical molecular systems, the CNEO-DFT energy surface serves as the CMES and its energy gradient is used for dynamics simulations. Note that CNEO-MD is essentially an AIMD method with energies and gradients of CNEO-DFT computed on-the-fly at each MD step. Compared with conventional DFT-based AIMD, CNEO-MD only leads to a tiny increase in computational cost but much more accurate vibrational spectra

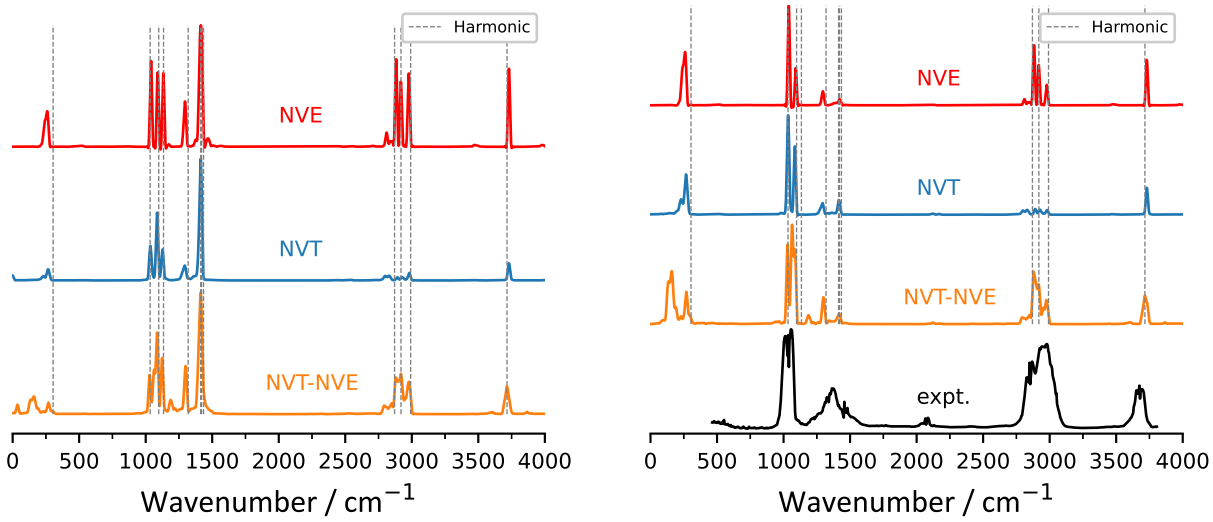
with nuclear quantum effects incorporated.[28,36,47]

In this Letter, we use gas phase methanol molecule as an example to compare the performance of three different MD sampling schemes and demonstrate the strength of the direct *NVE* scheme. To further show the power of CNEO-MD as well as the importance of vibrational mode coupling effects, we choose three highly challenging water clusters, including neutral water dimer(H_4O_2), protonated water trimer ($H_7O_3^+$) and protonated water tetramer ($H_9O_4^+$) for study. It is known that the choice of electronic density functional could significantly influence the calculation of vibrational spectra,[36,48] and here we adopt the PBE0[49,50] functional for both CNEO-DFT and DFT calculations as it was previously found that PBE0 is the best functional for free X-H stretches as benchmarked against the gold standard CCSD(T) method[51,52] for a series of small organic molecules,[36] where X is a heavy atom such as carbon, nitrogen and oxygen. However, because hydrogen bonds are not present in the small organic molecules that were benchmarked before, here for water cluster systems, we additionally benchmarked several commonly used functionals against CCSD(T) on the O-H stretch modes in H_4O_2. We found that the meta-GGA functional ωB97MV[53] with VV10 dispersion correction[54] has the best performance and PBE0 is the runnerup with less than $25\,cm^{-1}$ deviations from the CCSD(T) reference (Table S1). Because of the relatively low computational cost and relatively high accuracy, we keep using PBE0 in this Letter. We use the def2-TZVP[55] electronic basis set for methanol calculations and def2-TZVPPD[55,56] for water clusters calculations. As to nuclei, we only treat protons quantum mechanically because they feature the most significant nuclear quantum effects, but we note that a full quantum treatment for all nuclei is possible.[35] The PB4D basis set[57] is used for protons. Because here protons are relatively localized with little overlap between each other, they are treated as distinguishable particles and no proton-proton correlation is included. We also do not include electron-proton correlation (epc) because in the past we found that without an epc functional,[58–60] CNEO-DFT and CNEO-MD are already sufficiently accurate for vibrational spectra calculations.[28,36] We leave the investigation of epc effects for future studies.

Both DFT and CNEO-DFT harmonic analyses are performed using an in-house version[61] of PySCF,[62,63] and MD simulations are performed with Atomic Simulation Environment (ASE).[64] For the direct *NVE* scheme, unless otherwise specified, 50 2 ps simulations are performed for spectra computation. To initialize each trajectory, the starting geometry is the optimized geometry and a kinetic energy $E = k_B T$ is added to each mode with random positive or negative velocity directions. For the direct *NVT* scheme, the Nosé-Hoover chain thermostat[65–67] that has been implemented in an in-house version of ASE is used. We closely follow the treatment in ref 29: First a 10 ps simulation is performed for the system to reach equilibrium, then another 10 ps simulation is performed for production run. As to the *NVT-NVE* scheme, we use the Andersen thermostat[68] for *NVT* sampling because it does not suffer from the ergodicity issue as the Nosé-Hoover chain thermostat.[69] A 20 ps *NVT* simulation is first performed, and then uncorrelated configurations from the later 15 ps are picked as initial configurations for subsequent 2 ps *NVE* simulations.

We first use the gas phase methanol molecule to compare the performance of direct *NVE*, direct *NVT*, and *NVT-NVE* schemes. In a previous study by our group, it has been found that CNEO-MD can give vibrational spectra that agree excellently with experimental spectra for a series of small organic molecules including methanol, indicating its significantly higher accuracy than conventional DFT-based AIMD.[36] Therefore, here we only run CNEO-MD to test and compare different simulation schemes. Figure 1 presents the power and IR spectra of a single methanol molecule at 300 K or its equivalent energy from direct *NVE*, direct *NVT*, and *NVT-NVE* schemes.

For the direct *NVE* scheme, we run 5 2 ps simulations with the same energy but random initial cmode velocity directions and obtain the final spectrum as an average of 5 individual spectra. As power spectra reflect the energy distribution among different vibrational modes, we can see from Figure 1a that energy is about equally distributed among different modes. This leads to IR spectrum from the direct *NVE* scheme that is in good agreement with the experimental spectrum. For the direct *NVT* scheme, the spectra are calculated from the



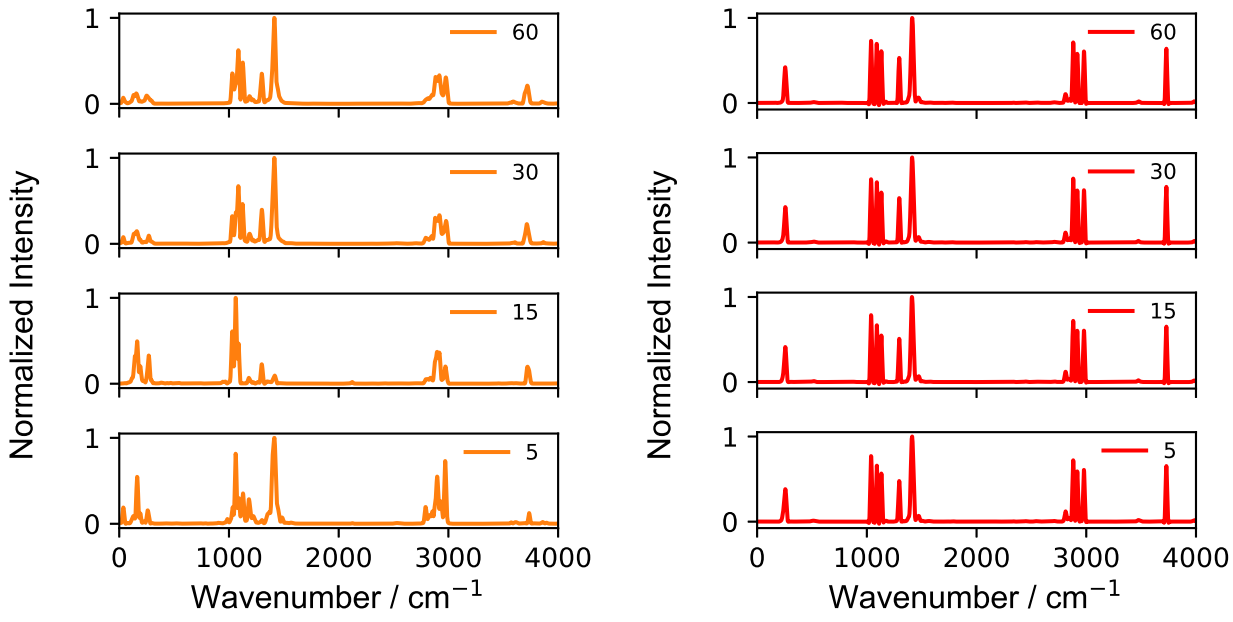
(a) Power spectra

(b) IR spectra

Figure 1: Power and IR spectra of a single methanol molecule by different MD schemes as indicated. Dashed vertical lines are static calculation reference from CNEO-DFT harmonic analysis. The experimental IR spectrum of gas phase methanol is from the National Institute of Standards and Technology (NIST) WebBook.

later 10 ps of a 20 ps simulation. From the power spectrum, we can see that with the same amount of total simulation time as the direct *NVE* scheme, energy equipartition in the direct *NVT* scheme is far from being achieved. Specifically for this trajectory, energies of the O-H and C-H stretch modes above 2500 cm^{-1} are much lower than the remaining modes, which leads to weak IR intensities for the highly IR acitve C-H and O-H stretch modes compared to experiments. This observation agrees well with previous literature that similar unequal energy distributions across different vibrational modes were observed for a single methanol molecule in conventional AIMD simulations.[29] As to the *NVT-NVE* scheme, here the spectra are obtained from an average of 30 2 ps *NVE* trajectories, whose initial configurations are sampled every 500 fs from an *NVT* trajectory. This choice of 500 fs sampling time gap is justified by the fast decay of autocorrelation functions of the *NVT* trajectory (Figure S1). From the power and IR spectra, it can be seen that the results from the *NVT-NVE* scheme are close to the direct *NVE* scheme and in good agreement with the experimental spectrum as well.

Note that here we only run 5 trajectories for the direct *NVE* scheme but 30 trajectories for the *NVT-NVE* scheme. This is because the spectra calculated from the direct *NVE* scheme need fewer trajectories to converge. Figure 2 shows the convergence test results for power spectra as a function of number of random trajectories used in both schemes. We can see that at least 30 trajectories are needed for convergence within the *NVT-NVE* scheme. In contrast, only 5 trajectories are sufficient to achieve convergence within the direct *NVE* scheme for this simple methanol molecule. Although later in more complicated water clusters, more trajectories will be needed for spectrum convergence, the general trend is always true that the direct *NVE* scheme needs significantly fewer trajectories than the *NVT-NVE* scheme to reach convergence. As was discussed before, this difference of convergence behavior is due to the much smaller phase space of a microcanonical ensemble compared to that of the corresponding canonical ensemble, and it makes the direct *NVE* scheme more efficient in vibrational spectrum simulations.



(a) *NVT-NVE* (b) *NVE*

Figure 2: Methanol molecule power spectra as a function of number of trajectories averaged for *NVT-NVE* and direct *NVE* schemes. The direct *NVE* scheme needs significantly fewer trajectories than the *NVT-NVE* scheme to reach spectrum convergence.

Another difference between the direct *NVE* and the *NVT-NVE* schemes is that peaks are generally broader in the *NVT-NVE* scheme. It is understandable that with the *NVT-NVE* scheme, there is a total energy distribution with more fluctuation, whereas the total energy is fixed in the direct *NVE* scheme. However, we emphasize that this does not mean that energy will not flow with the direct *NVE* scheme. Instead, through mode coupling effects, energy can exchange among different modes and also broaden their peaks. Therefore, both schemes are able to capture the mode coupling effects through finite temperature or energy, and it is simply because the fluctuation is larger in the *NVT-NVE* scheme and makes the peaks broader.

To sum up, both the direct *NVE* scheme and the *NVT-NVE* scheme can give reliable spectra but the NVE scheme is significantly more efficient in spectrum convergence with minimal simulation time. The direct *NVT* scheme may be as efficient as the direct *NVE* scheme, but it may suffer from the difficulty of effectively equilibrating different vibrational modes in a single trajectory due to relative weak mode couplings between a limited number of degrees of freedom. However, we note that in condensed phase systems where modes couple more strongly and interact with the surroundings more frequently, there will be little challenge of reaching equipartition and thus the direct *NVT* scheme will be much more successful there.[29,33]

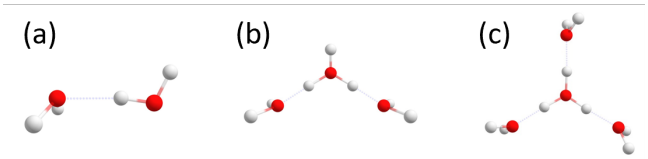


Figure 3: Schematic pictures of (a) H_4O_2, (b) $H_7O_3^+$ and (c) $H_9O_4^+$ water clusters.

To further show the capability of the combination of the direct *NVE* scheme and CNEO-MD, we calculate the IR spectra of three water cluster systems. The structures of three water clusters to be studied are shown in Figure 3 (Structure files are given in the Supporting Information). To start with, water dimer H_4O_2 is the simplest water cluster and its vibrational spectra have been studied experimentally and theoretically for decades.[71–78] In H_4O_2, the

Table 1: Vibrational frequencies of O-H stretch modes in H_4O_2 (in cm^{-1}) from harmonic analysis.

Vibrational mode	expt.[a]	CCSD(T)[a]	DFT	CNEO-DFT
	3745	3941	3964	3778
free O-H	3735	3923	3944	3766
	3660	3836	3861	3679
bound O-H	3601	3753	3731	3494

[a] From ref. 70.

two water molecules are bounded by the intermolecular hydrogen bond, and there are three free O-H stretch modes and one bound O-H stretch mode. The vibrational frequencies of these four O-H stretch modes obtained from CCSD(T), DFT, and CNEO-DFT harmonic analysis, along with experimental references, are presented in Table 1. With the harmonic approximation, the gold standard electronic structure method CCSD(T) overestimates all O-H stretch frequencies by over $150\,cm^{-1}$ and DFT with PBE0 gives similar results. In contrast, although the harmonic approximation is still invoked, because of the incorporation of nuclear quantum effects, CNEO-DFT with PBE0 accurately describes the vibrational frequencies of the free O-H stretches, whose errors are within $30\,cm^{-1}$ of the experimental references. For the bound O-H stretch, interestingly, CNEO-DFT harmonic analysis gives a larger error and underestimates the frequency by 107 cm^{-1}, but it still slightly outperforms DFT, which overestimates the frequency by $130\,cm^{-1}$.

This large error of CNEO-DFT harmonic analysis seems to indicate a failure case for CNEO-DFT, but one should note that a special feature of H_4O_2 is that its bound O-H stretch strongly couples with its low-frequency modes, making the system highly anharmonic.[79,80] Therefore, although CNEO-DFT harmonic analysis incorporates nuclear quantum effects and some single mode anharmonicity, the missing of mode coupling effects beyond the harmonic picture could be the reason for its lower accuracy in describing the bound O-H stretch. Since MD simulations provide an efficient way of incorporating mode coupling effects through finite temperature thermal motions,[29] here we perform CNEO-MD with the new direct *NVE* scheme to investigate the importance of these coupling effects. Note that the water dimer

system is more complex than the methanol molecule and different trajectories tend to give significantly different spectra (Figure S3). Therefore, instead of using 5 trajectories as in the methanol case, here we use 50 trajectories to obtain the converged IR spectrum. The convergence test was performed by randomly choosing 30 trajectories for spectra generation and comparing the spectra to the one generated from all 50 trajectories (Figure S4, the same convergence test is performed on $H_7O_3^+$ and $H_9O_4^+$ as well in Figure S5 and S6). We found that in fact 30 trajectories are already sufficient to reach good spectra convergence.

Figure 4 shows the calculated IR spectra from CNEO-MD and AIMD with the total energy corresponding to 300 K. For comparison, we also show their harmonic analysis results with artificial peak broadening as well as the experimental peak position references. We can see that in spectra obtained from both CNEO-MD and AIMD, the bound O-H stretch blueshifts about $100\,\mathrm{cm}^{-1}$ compared with their corresponding harmonic results. This shift makes CNEO-MD accurately capture the frequency of the bound O-H stretch but makes the AIMD results worse in that the underlying DFT already largely overestimates the vibrational frequency from harmonic analysis without nuclear quantum effects incorporated. Interestingly, for free O-H stretches, because they are barely coupled with other modes, MD simulations hardly change their peak positions compared to harmonic analyses. Therefore, with the incorporation of both nuclear quantum effects and mode coupling effects, CNEO-MD significantly improve over CNEO-DFT harmonic analysis and can excellently describe both free and bound O-H stretches in water dimer. In contrast, due to the missing of nuclear quantum effects, DFT-based AIMD may not necessarily perform better than DFT harmonic analysis. Ironically, for the bound O-H stretch, due to the error cancellation from missing nuclear quantum effects and mode coupling effects, the simple DFT harmonic analysis is much better than the more expensive DFT-based AIMD, which could be the reason for the relatively low popularity of using conventional AIMD for vibrational spectra simulation compared to DFT harmonic analysis.

One should notice that in classical simulations, temperature will affect computed vibra-

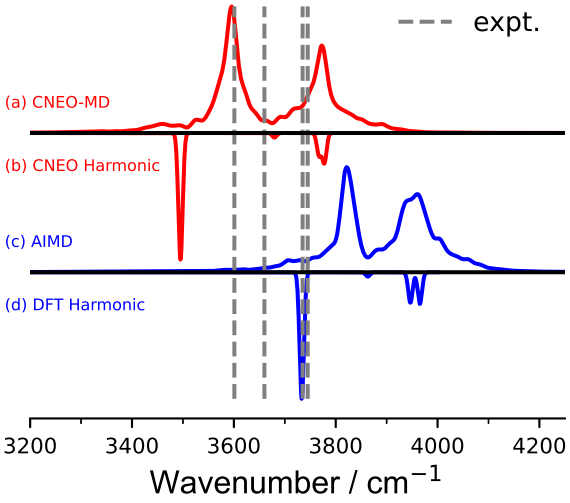


Figure 4: Calculated IR spectra of H_4O_2 by (a) CNEO-MD, (b) CNEO Hessian, (c) AIMD and (d) DFT Hessian. Hessian spectra are obtained by adding 10 cm^{-1} Gaussian width to each transition. Experimental reference is from ref. 70.

tional spectra. Similarly, here in the direct *NVE* scheme, the simulation energy also influences how much the bound O-H stretch shifts. Lower energies, for instance 200 K, make the frequency blueshift less compared with the 300 K simulation presented here, while higher energies, for instance 400 K, could make the frequency shift to even higher frequencies than the experimental reference (Figure S2). Because in principle, peak positions should not change with temperature, this temperature effect is an artifact of classical simulations. Therefore, we emphasize that the simulation temperatures/energies here do not mean that experiments are carried out at this temperature, instead, they should be purely perceived as a parameter that can modulate the mode coupling strength. Although this artifact cannot be avoided in CNEO-MD, it is an effective way of incorporating mode coupling effects, and we find that the 300 K parameter work well for both the water dimer and the protonated water clusters below, and it seems to work well for more hydrogen bonded systems as well.[81]

Now we apply CNEO-MD to two significantly more challenging protonated water cluster systems, $H_7O_3^+$ and $H_9O_4^+$, whose vibrational spectra have been studied extensively with high-level and computationally intensive methods such as VSCF/VCI,[10–12] MCTDH[13] and

DMC.[82] Relatively cheap AIMD and TRMPD have also been applied to these two systems, but the spectra of the former are blueshifted significantly and the latter suffers from artificial broadening.[83]

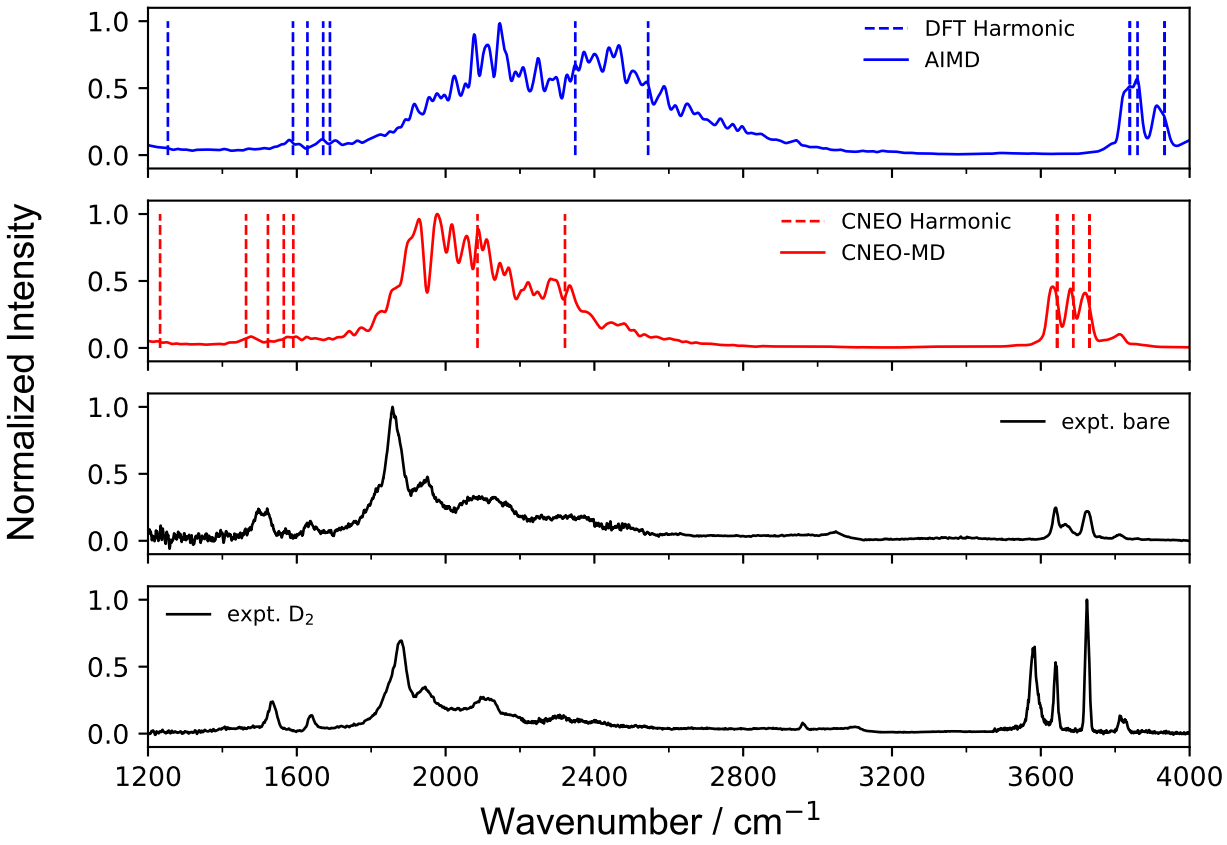


Figure 5: Overview of IR spectra of $H_7O_3^+$ obtained by the indicated methods. Experimental IR spectra are from ref. 10.

Figure 5 shows the calculated IR spectra of $H_7O_3^+$ from AIMD and CNEO-MD along with experimental references. In experiment spectra, the dominant feature is the broad progression of bands from 1800 to 2400 cm^{-1} with the maximum peak located at 1858 cm^{-1} for the bare cluster and 1878 cm^{-1} for the D_2-tagged cluster.[10] VSCF/VCI analysis assigned the most intense peak to the hydronium H_3O^+ asymmetric stretch with strong coupling to some low-frequency modes, and the broad tail extended to 2400 cm^{-1} is attributed to many overtones and combination bands.[10] CNEO harmonic analysis overestimates the asymmetric

stretch of bound O-H of H_3O^+ by 200 cm^{-1} (2086 cm^{-1}), although it is already much better than the DFT harmonic result (2349 cm^{-1}). With the incorporation of mode coupling effects, CNEO-MD further significantly improves the results with an intense peak positioned around 1900 cm^{-1} and shows a progression of bands extended to 2400 cm^{-1}. Although the relative intensity contrast between the main peak and the band progression is not as sharp as that in the experimental spectra, this significant improvement of CNEO-MD again demonstrates that CNEO-MD overall successfully reproduces most features presented in experiment and captures the strong mode coupling effects between bound O-H stretches and other low-frequency modes. In contrast, conventional AIMD predicts an overly broadened progression from 1800 to 2800 cm^{-1} with its most intense peak appearing around 2100 cm^{-1}. For the free O-H stretches around 3600 cm^{-1}, MD barely changes over the harmonic analysis results, and both CNEO harmonic analysis and CNEO-MD can accurately predict their vibrational frequencies, consistent with the water dimer case as well as a series of molecular systems our group tested before.[36]

Figure 6 shows the calculated IR spectra of $H_9O_4^+$ from AIMD and CNEO-MD along with experimental references. For this cluster, there once existed controversies about the existence of linear-chain Zundel isomer in the observed gas phase IR spectrum,[32] but later joint experimental and theoretical efforts concluded that only the Eigen isomer exists in the experimental spectrum.[11] Therefore, here we only perform CNEO-MD and conventional AIMD simulations on the Eigen isomer. The prominent feature of the experimental IR spectra for $H_9O_4^+$ is the band around 2650 cm^{-1}, which is assigned to hydronium H_3O^+ asymmetric stretch coupled to low-frequency modes.[12] CNEO harmonic analysis predicts the frequency of this mode to be 2585 cm^{-1}, which is 65 cm^{-1} lower than experiment. With CNEO-MD, this peak blueshifts to 2600 cm^{-1}, slightly closer to the experimental results. In contrast, DFT harmonic analysis overestimates the frequency by about 200 cm^{-1} and AIMD makes the overestimation slightly worse. Despite again the great success of CNEO-MD in reproducing the main bands of the experimental IR spectra, we should point out a tiny

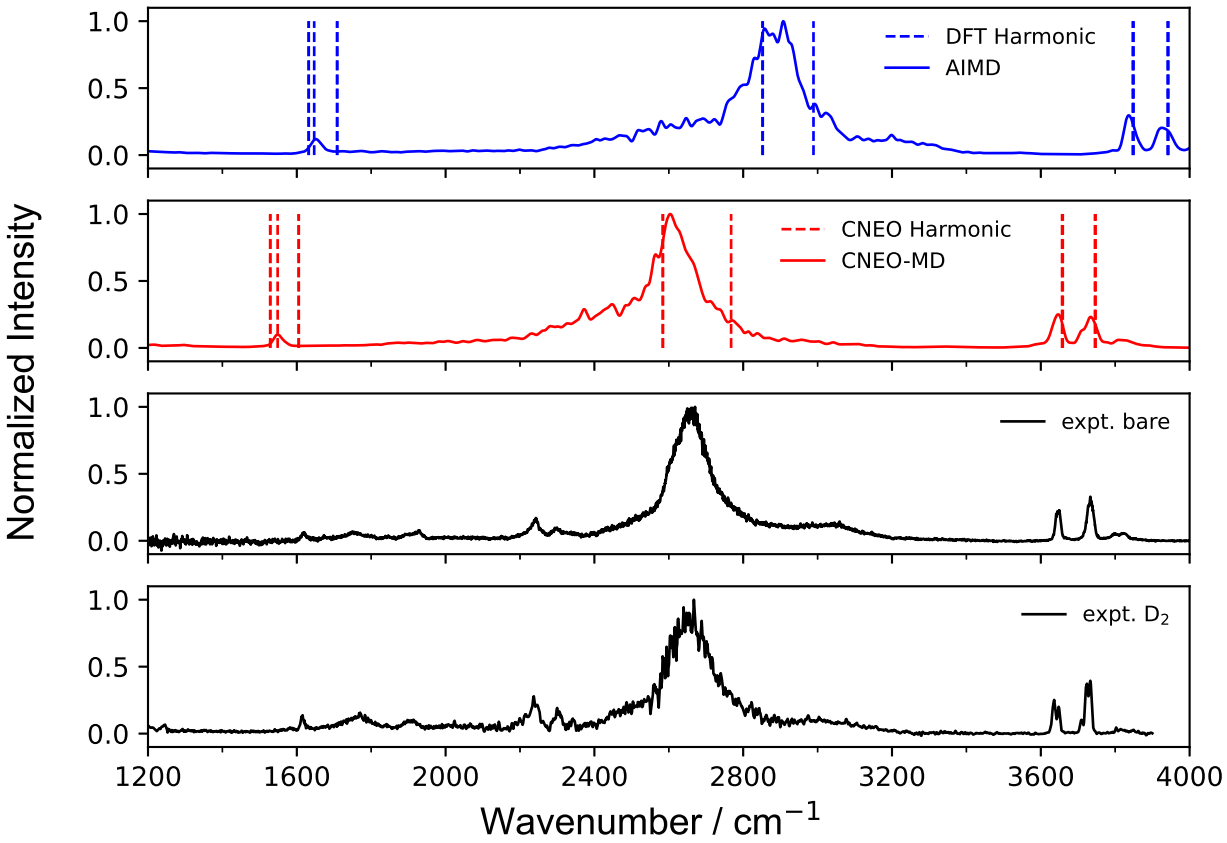


Figure 6: Overview of IR spectra of $H_9O_4^+$ obtained by the indicated methods. Experimental IR spectrum of bare cluster is from ref 12 and experimental IR spectrum of D_2-tagged cluster is from ref 84.

feature that seems absent in CNEO-MD: a diffuse feature that lies about 380 cm^{-1} above the dominant peak in both experimental spectra.[12] With several high level methods, this feature has been attributed to various overtones and combination bands, especially the combination band involving the O-O stretch mode.[12,85] AIMD seems to have captured some diffuse feature at around 3200 cm^{-1}, far from the experimental reference, while CNEO-MD has no such observable feature. It is well known that capturing overtones and combination bands are challenging for classical simulations,[29] and CNEO-MD can only occasionally capture them, depending on the system.[28,36] For the free O-H stretches, same as all cases investigated before, both CNEO harmonic analysis and CNEO-MD can accurately describe them.

Finally, we comment that the two protonated water clusters have very different vibrational features, especially in the dominant bands featuring bound O-H stretches. The $H_7O_3^+$ spectrum is relatively more complex, and CNEO harmonic analysis greatly overestimates its bound O-H stretch frequency, whereas the $H_9O_4^+$ spectrum features an intense band associated with the asymmetric bound O-H stretches, and CNEO harmonic analysis gives a reasonable prediction on the frequency. When it comes to CNEO-MD, we find that it improves much over the poor CNEO harmonic results on $H_7O_3^+$ but only makes a tiny correction on top of the already good CNEO harmonic results on $H_9O_4^+$, and thus we can safely conclude that CNEO-MD can correctly capture the coupling effects between different vibrational modes, whether they are strong or weak. The quantitative agreement with experiment may not be as perfect as those higher-level methods, but as a method based on classical simulations, CNEO-MD is already accurate enough with less than 50 cm^{-1} differences from the experimental references in both cases. Additionally, one should note that a significant percentage of the remaining small error probably comes from the choice of the electronic functional. Even though PBE0 is already one of the best functionals for describing vibrations with significant hydrogen motion character, its deviation from CCSD(T) references can still be around 20 cm^{-1} as benchmarked in Table 1 and in a previous work by our group (see Table S1 and Supporting Information of ref. 36). Therefore, the development of more

accurate functionals for vibrational frequency calculations remains an important task.

In summary, we introduced the direct *NVE* scheme to efficiently compute vibrational spectra by directly running microcanonical ensemble simulations. We proved that under harmonic approximations, the new scheme can obtain essentially the same position and dipole autocorrelation function as conventional *NVT*-*NVE* and direct *NVT* schemes. With the gas phase methanol molecule as an example, we showed that the direct *NVE* scheme can reach spectrum convergence faster with less total simulation time and thus is much more efficient for vibrational spectrum calculations for real molecules.

In combination with the direct *NVE* scheme, we demonstrated again the power of CNEO-MD for accurate and efficient vibrational spectrum calculations. With three highly challenging anharmonic water clusters, H_4O_2, $H_7O_3^+$ and $H_9O_4^+$, as examples, we showed that CNEO-DFT harmonic analysis improves over DFT harmonic analysis with the incorporation of nuclear quantum effects, and CNEO-MD further improves the results with the incorporation of mode coupling effects. Therefore, both nuclear quantum effects and mode coupling effects are crucial for the excellent performance of CNEO-MD. As a cost-effective method based on classical simulations, CNEO-MD in combination with the direct *NVE* scheme holds significant promise as a robust tool for conducting vibrational spectrum simulations in more challenging systems.

Acknowledgement

The authors thank Nan Yang and Mark A. Johnson for providing experimental spectra of protonated water clusters studied in this Letter, and Yunrui Qiu and Zhongyi Wan for helpful discussion on MD schemes. The authors are grateful for the funding support from the National Science Foundation under Grant No. 2238473 and from the University of Wisconsin via the Wisconsin Alumni Research Foundation. The computational resource support from Center for High Throughput Computing at the University of Wisconsin-Madison is

acknowledged.[86]

Supporting Information Available

Structures of water clusters (xyz files); Detailed derivation of the direct *NVE* scheme; benchmark of different electronic functionals on H_4O_2; different temperature H_4O_2 spectra; spectrum convergence test

References

(1) Stuart, B. H. *Infrared spectroscopy: fundamentals and applications*; 2004.

(2) Barth, A. Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **2007**, *1767*, 1073–1101.

(3) Baker, M. J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H. J.; Dorling, K. M.; Fielden, P. R.; Fogarty, S. W.; Fullwood, N. J.; Heys, K. A. et al. Using Fourier transform IR spectroscopy to analyze biological materials. *Nature protocols* **2014**, *9*, 1771–1791.

(4) Baiz, C. R.; Błasiak, B.; Bredenbeck, J.; Cho, M.; Choi, J.-H.; Corcelli, S. A.; Dijkstra, A. G.; Feng, C.-J.; Garrett-Roe, S.; Ge, N.-H. et al. Vibrational spectroscopic map, vibrational spectroscopy, and intermolecular interaction. *Chemical reviews* **2020**, *120*, 7152–7218.

(5) Bowman, J. M. The self-consistent-field approach to polyatomic vibrations. *Accounts of Chemical Research* **1986**, *19*, 202–208.

(6) Bowman, J. M.; Carter, S.; Huang, X. MULTIMODE: a code to calculate rovibrational energies of polyatomic molecules. *International Reviews in Physical Chemistry* **2003**, *22*, 533–549.

(7) Vendrell, O.; Gatti, F.; Meyer, H.-D. Full dimensional (15-dimensional) quantum-dynamical simulation of the protonated water dimer. II. Infrared spectrum and vibrational dynamics. *The Journal of chemical physics* **2007**, *127*.

(8) Reynolds, P. J.; Ceperley, D. M.; Alder, B. J.; Lester Jr, W. A. Fixed-node quantum Monte Carlo for molecules. *The Journal of Chemical Physics* **1982**, *77*, 5593–5603.

(9) McCoy, A. B. Diffusion Monte Carlo approaches for investigating the structure and vibrational spectra of fluxional systems. *International Reviews in Physical Chemistry* **2006**, *25*, 77–107.

(10) Duong, C. H.; Gorlova, O.; Yang, N.; Kelleher, P. J.; Johnson, M. A.; McCoy, A. B.; Yu, Q.; Bowman, J. M. Disentangling the Complex Vibrational Spectrum of the Protonated Water Trimer, H+(H2O)3, with Two-Color IR-IR Photodissociation of the Bare Ion and Anharmonic VSCF/VCI Theory. *The Journal of Physical Chemistry Letters* **2017**, *8*, 3782–3789.

(11) Yu, Q.; Bowman, J. M. High-Level Quantum Calculations of the IR Spectra of the Eigen, Zundel, and Ring Isomers of H+(H2O)4 Find a Single Match to Experiment. *Journal of the American Chemical Society* **2017**, *139*, 10984–10987.

(12) Duong, C. H.; Yang, N.; Kelleher, P. J.; Johnson, M. A.; DiRisio, R. J.; McCoy, A. B.; Yu, Q.; Bowman, J. M.; Henderson, B. V.; Jordan, K. D. Tag-Free and Isotopomer-Selective Vibrational Spectroscopy of the Cryogenically Cooled H9O4+ Cation with Two-Color, IR–IR Double-Resonance Photoexcitation: Isolating the Spectral Signature of a Single OH Group in the Hydronium Ion Core. *The Journal of Physical Chemistry A* **2018**, *122*, 9275–9284.

(13) Schröder, M.; Gatti, F.; Lauvergnat, D.; Meyer, H.-D.; Vendrell, O. The coupling of the hydrated proton to its first solvation shell. *Nature Communications* **2022**, *13*, 6170.

(14) Cao, J.; Voth, G. A. The formulation of quantum statistical mechanics based on the Feynman path centroid density. II. Dynamical properties. *The Journal of chemical physics* **1994**, *100*, 5106–5117.

(15) Jang, S.; Voth, G. A. A derivation of centroid molecular dynamics and other approximate time evolution methods for path integral centroid variables. *The Journal of chemical physics* **1999**, *111*, 2371–2384.

(16) Craig, I. R.; Manolopoulos, D. E. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *The Journal of chemical physics* **2004**, *121*, 3368–3373.

(17) Rossi, M.; Ceriotti, M.; Manolopoulos, D. E. How to remove the spurious resonances from ring polymer molecular dynamics. *The Journal of chemical physics* **2014**, *140*.

(18) Trenins, G.; Willatt, M. J.; Althorpe, S. C. Path-integral dynamics of water using curvilinear centroids. *The Journal of Chemical Physics* **2019**, *151*.

(19) Benson, R. L.; Trenins, G.; Althorpe, S. C. Which quantum statistics–classical dynamics method is best for water? *Faraday Discussions* **2020**, *221*, 350–366.

(20) Fletcher, T.; Zhu, A.; Lawrence, J. E.; Manolopoulos, D. E. Fast quasi-centroid molecular dynamics. *The Journal of Chemical Physics* **2021**, *155*.

(21) Witt, A.; Ivanov, S. D.; Shiga, M.; Forbert, H.; Marx, D. On the applicability of centroid and ring polymer path integral molecular dynamics for vibrational spectroscopy. *The Journal of chemical physics* **2009**, *130*.

(22) Ivanov, S. D.; Witt, A.; Shiga, M.; Marx, D. Communications: On artificial frequency shifts in infrared spectra obtained from centroid molecular dynamics: Quantum liquid water. *The Journal of chemical physics* **2010**, *132*.

(23) Shiga, M.; Nakayama, A. Ab initio path integral ring polymer molecular dynamics: Vibrational spectra of molecules. *Chemical Physics Letters* **2008**, *451*, 175–181.

(24) Habershon, S.; Fanourgakis, G. S.; Manolopoulos, D. E. Comparison of path integral molecular dynamics methods for the infrared absorption spectrum of liquid water. *The Journal of chemical physics* **2008**, *129*.

(25) Marsalek, O.; Markland, T. E. Quantum dynamics and spectroscopy of ab initio liquid water: The interplay of nuclear and electronic quantum effects. *The journal of physical chemistry letters* **2017**, *8*, 1545–1551.

(26) Lieberherr, A. Z.; Furniss, S. T.; Lawrence, J. E.; Manolopoulos, D. E. Vibrational strong coupling in liquid water from cavity molecular dynamics. *The Journal of Chemical Physics* **2023**, *158*.

(27) Barone, V. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *The Journal of chemical physics* **2005**, *122*, 014108.

(28) Zhang, Y.; Xu, X.; Yang, N.; Chen, Z.; Yang, Y. Describing proton transfer modes in shared proton systems with constrained nuclear–electronic orbital methods. *The Journal of Chemical Physics* **2023**, *158*.

(29) Thomas, M.; Brehm, M.; Fligg, R.; Vöhringer, P.; Kirchner, B. Computing vibrational spectra from ab initio molecular dynamics. *Physical Chemistry Chemical Physics* **2013**, *15*, 6608–6622.

(30) Baer, M.; Marx, D.; Mathias, G. Assigning Predissociation Infrared Spectra of Microsolvated Hydronium Cations H3O+·(H2)n (n=0, 1, 2, 3) by Ab Initio Molecular Dynamics. *ChemPhysChem* **2011**, *12*, 1906–1915.

(31) Heine, N.; Fagiani, M. R.; Rossi, M.; Wende, T.; Berden, G.; Blum, V.; Asmis, K. R.

Isomer-selective detection of hydrogen-bond vibrations in the protonated water hexamer. *Journal of the American Chemical Society* **2013**, *135*, 8266–8273.

(32) Wang, H.; Agmon, N. Reinvestigation of the Infrared Spectrum of the Gas-Phase Protonated Water Tetramer. *The Journal of Physical Chemistry A* **2017**, *121*, 3056–3070.

(33) Thomas, M. *Theoretical Modeling of Vibrational Spectra in the Liquid Phase*; Springer, 2016.

(34) Xu, X.; Yang, Y. Constrained nuclear-electronic orbital density functional theory: Energy surfaces with nuclear quantum effects. *The Journal of Chemical Physics* **2020**, *152*.

(35) Xu, X.; Yang, Y. Full-quantum descriptions of molecular systems from constrained nuclear–electronic orbital density functional theory. *The Journal of Chemical Physics* **2020**, *153*.

(36) Xu, X.; Chen, Z.; Yang, Y. Molecular dynamics with constrained nuclear electronic orbital density functional theory: Accurate vibrational spectra from efficient incorporation of nuclear quantum effects. *Journal of the American Chemical Society* **2022**, *144*, 4039–4046.

(37) Gordon, R. Molecular motion in infrared and Raman spectra. *The Journal of Chemical Physics* **1965**, *43*, 1307–1312.

(38) Berne, B. J.; Harp, G. On the calculation of time correlation functions. *Advances in chemical physics* **1970**, 63–227.

(39) Tuckerman, M. *Statistical Mechanics: Theory and Molecular Simulation*; OUP Oxford, 2010.

(40) Thomas, I. Protonic structure of molecules. I. Ammonia molecules. *Physical Review* **1969**, *185*, 90.

(41) Capitani, J. F.; Nalewajski, R. F.; Parr, R. G. Non-Born–Oppenheimer density functional theory of molecular systems. *The Journal of Chemical Physics* **1982**, *76*, 568–573.

(42) Nakai, H. Nuclear orbital plus molecular orbital theory: Simultaneous determination of nuclear and electronic wave functions without Born–Oppenheimer approximation. *International Journal of Quantum Chemistry* **2007**, *107*, 2849–2869.

(43) Pavošević, F.; Culpitt, T.; Hammes-Schiffer, S. Multicomponent Quantum Chemistry: Integrating Electronic and Nuclear Quantum Effects via the Nuclear–Electronic Orbital Method. *Chemical Reviews* **2020**, *120*, 4222–4253.

(44) Pulay, P. Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules: I. Theory. *Molecular Physics* **1969**, *17*, 197–204.

(45) Chen, Z.; Yang, Y. Incorporating nuclear quantum effects in molecular dynamics with a constrained minimized energy surface. *The Journal of Physical Chemistry Letters* **2023**, *14*, 279–286.

(46) Wang, Y.; Chen, Z.; Yang, Y. Calculating vibrational excited state absorptions with excited state constrained minimized energy surfaces. *The Journal of Physical Chemistry A* **2023**,

(47) Xu, X. Constrained Nuclear-Electronic Orbital Density Functional Theory with a Dielectric Continuum Solvent Model. *The Journal of Physical Chemistry A* **2023**,

(48) Taherivardanjani, S.; Elfgen, R.; Reckien, W.; Suarez, E.; Perlt, E.; Kirchner, B. Benchmarking the Computational Costs and Quality of Vibrational Spectra from Ab Initio Simulations. *Advanced Theory and Simulations* **2022**, *5*, 2100293.

(49) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **1996**, *77*, 3865.

(50) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of chemical physics* **1999**, *110*, 6158–6170.

(51) Bartlett, R. J. Many-body perturbation theory and coupled cluster theory for electron correlation in molecules. *Annual review of physical chemistry* **1981**, *32*, 359–401.

(52) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters* **1989**, *157*, 479–483.

(53) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *The Journal of chemical physics* **2016**, *144*.

(54) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional: The simpler the better. *The Journal of chemical physics* **2010**, *133*.

(55) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **2005**, *7*, 3297–3305.

(56) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *The Journal of chemical physics* **2010**, *133*.

(57) Yu, Q.; Pavošević, F.; Hammes-Schiffer, S. Development of nuclear basis sets for multicomponent quantum chemistry methods. *The Journal of Chemical Physics* **2020**, *152*.

(58) Yang, Y.; Brorsen, K. R.; Culpitt, T.; Pak, M. V.; Hammes-Schiffer, S. Development of a practical multicomponent density functional for electron-proton correlation to produce accurate proton densities. *The Journal of Chemical Physics* **2017**, *147*.

(59) Brorsen, K. R.; Yang, Y.; Hammes-Schiffer, S. Multicomponent density functional theory: Impact of nuclear quantum effects on proton affinities and geometries. *The Journal of Physical Chemistry Letters* **2017**, *8*, 3488–3493.

(60) Tao, Z.; Yang, Y.; Hammes-Schiffer, S. Multicomponent density functional theory: Including the density gradient in the electron-proton correlation functional for hydrogen and deuterium. *The Journal of Chemical Physics* **2019**, *151*.

(61) Available at https://github.com/theorychemyang/pyscf.

(62) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S. et al. PySCF: the Python-based simulations of chemistry framework. *WIREs Computational Molecular Science* **2018**, *8*, e1340.

(63) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z.-H. et al. Recent developments in the PySCF program package. *The Journal of Chemical Physics* **2020**, *153*, 024109.

(64) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C. et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.

(65) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **1984**, *81*, 511–519.

(66) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **1985**, *31*, 1695–1697.

(67) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé–Hoover chains: The canonical

ensemble via continuous dynamics. *The Journal of chemical physics* **1992**, *97*, 2635–2643.

(68) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics* **1980**, *72*, 2384–2393.

(69) Cooke, B.; Schmidler, S. C. Preserving the Boltzmann ensemble in replica-exchange molecular dynamics. *The Journal of chemical physics* **2008**, *129*.

(70) Johnson, R. Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database 101. 2002; `http://cccbdb.nist.gov/`.

(71) Huisken, F.; Kaloudis, M.; Kulcke, A. Infrared spectroscopy of small size-selected water clusters. *The Journal of Chemical Physics* **1996**, *104*, 17–25.

(72) Low, G. R.; Kjaergaard, H. G. Calculation of OH-stretching band intensities of the water dimer and trimer. *The Journal of chemical physics* **1999**, *110*, 9104–9115.

(73) Dunn, M. E.; Evans, T. M.; Kirschner, K. N.; Shields, G. C. Prediction of Accurate Anharmonic Experimental Vibrational Frequencies for Water Clusters, (H2O)n, n = 2-5. *The Journal of Physical Chemistry A* **2006**, *110*, 303–309.

(74) Kjaergaard, H. G.; Garden, A. L.; Chaban, G. M.; Gerber, R. B.; Matthews, D. A.; Stanton, J. F. Calculation of vibrational transition frequencies and intensities in water dimer: Comparison of different vibrational approaches. *The Journal of Physical Chemistry A* **2008**, *112*, 4324–4335.

(75) Kuyanov-Prozument, K.; Choi, M. Y.; Vilesov, A. F. Spectrum and infrared intensities of OH-stretching bands of water dimers. *The Journal of chemical physics* **2010**, *132*.

(76) Perlt, E.; Berger, S. A.; Kelterer, A.-M.; Kirchner, B. Anharmonicity of Vibrational Modes in Hydrogen Chloride–Water Mixtures. *Journal of Chemical Theory and Computation* **2019**, *15*, 2535–2547.

(77) Zhang, B.; Yu, Y.; Zhang, Z.; Zhang, Y.-Y.; Jiang, S.; Li, Q.; Yang, S.; Hu, H.-S.; Zhang, W.; Dai, D. et al. Infrared spectroscopy of neutral water dimer based on a tunable vacuum ultraviolet free electron laser. *The Journal of Physical Chemistry Letters* **2020**, *11*, 851–855.

(78) Vogt, E.; Kjaergaard, H. G. Vibrational spectroscopy of the water dimer at jet-cooled and atmospheric temperatures. *Annual Review of Physical Chemistry* **2022**, *73*, 209–231.

(79) Mackeprang, K.; Kjaergaard, H. G.; Salmi, T.; Hänninen, V.; Halonen, L. The effect of large amplitude motions on the transition frequency redshift in hydrogen bonded complexes: A physical picture. *The Journal of Chemical Physics* **2014**, *140*.

(80) Mackeprang, K.; Hänninen, V.; Halonen, L.; Kjaergaard, H. G. The effect of large amplitude motions on the vibrational intensities in hydrogen bonded complexes. *The Journal of Chemical Physics* **2015**, *142*.

(81) Langford, J.; Zhang, Y.; Chen, Z.; Yang, Y. Where is the Hidden Intramolecular H-bonding Vibrational Signal in Proline? *In preparation*

(82) DiRisio, R. J.; Finney, J. M.; Dzugan, L. C.; Madison, L. R.; McCoy, A. B. Using Diffusion Monte Carlo Wave Functions to Analyze the Vibrational Spectra of H7O3+ and H9O4+. *The Journal of Physical Chemistry A* **2021**, *125*, 7185–7197.

(83) Yu, Q.; Bowman, J. M. Classical, Thermostated Ring Polymer, and Quantum VSCF/VCI Calculations of IR Spectra of H7O3+ and H9O4+ (Eigen) and Comparison with Experiment. *The Journal of Physical Chemistry A* **2019**, *123*, 1399–1409.

(84) Wolke, C. T.; Fournier, J. A.; Dzugan, L. C.; Fagiani, M. R.; Odbadrakh, T. T.; Knorke, H.; Jordan, K. D.; McCoy, A. B.; Asmis, K. R.; Johnson, M. A. Spectroscopic snapshots of the proton-transfer mechanism in water. *Science* **2016**, *354*, 1131–1135.

(85) Yagi, K.; Thomsen, B. Infrared spectra of protonated water clusters, H+ (H2O) 4, in Eigen and Zundel forms studied by vibrational quasi-degenerate perturbation theory. *The Journal of Physical Chemistry A* **2017**, *121*, 2386–2398.

(86) Center for High Throughput Computing Center for High Throughput Computing. 2006; `https://chtc.cs.wisc.edu/`.