JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Generalized Normal Coordinates for the Vibrational Analysis of Molecular Dynamics Simulations

Gerald Mathias*[,†,§] and Marcel D. Baer[‡,§]

[†]Lehrstuhl für BioMolekulare Optik, Ludwig−Maximilians Universität München, Oettingenstrasse 67, 80538 München, Germany

[‡]Chemical and Materials Science Division, Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352, United States

ABSTRACT: The computation of vibrational spectra via molecular dynamics (MD) simulations has made lively progress in recent years. In particular, infrared spectra are accessible employing ab initio MD, for which only the total dipole moment has to be computed "on the fly" from the electronic structure along the trajectory. The analysis of such spectra in terms of the normal modes of intramolecular motion, however, still poses a challenge to theory. Here, we present an algorithm to extract such normal modes from MD trajectories by combining several ideas available in the literature. The algorithm allows one to compute both the normal modes and their vibrational bands without having to rely on an equipartition assumption, which hampered previous methods. Our analysis is based on a tensorial definition of the vibrational density of states, which spans both the frequency resolved cross- and auto-correlations of the molecular degrees of freedom. Generalized normal coordinates are introduced as orthonormal transforms of mass-weighted coordinates, which minimize their mutual cross-correlations. The generalized normal coordinates and their associated normal modes are iteratively constructed by a minimization scheme based on the Jacobi diagonalization. Furthermore, the analysis furnishes mode local temperatures, which provide not only a measure for the convergence of the computed intensities but also permits one to correct these intensities a posteriori toward the ensemble limit. As a first non-trivial test application we analyze the infrared spectrum of isoprene based on ab initio MD, which is an important building block of various dye molecules in molecular biology.

1. INTRODUCTION

Vibrational spectra of molecules, such as infrared (IR), Raman, or resonance Raman spectra, encode rich information about chemical bonds, intramolecular forces, and molecular structures. Moreover, in the condensed-phase vibrational spectra can be strongly modified by interactions with polar solvents or more complex environments, such as a protein matrix. Here, vibrational spectra also serve as a probe for intermolecular interactions and, thus, the solvent environment.

The decoding of vibrational spectra, i.e., the assignment of vibrational bands to molecular structures and motions, is often tedious for polyatomic molecules. On the experimental side, it involves, for example, site specific isotope substitutions or mutations of amino acid side chains, when it comes to protein spectra. Therefore, theory has become more and more important to help decipher information encoded in vibrational spectra.[1−11] With the development of density functional theory (DFT) and reliable gradient corrected functionals, one is nowadays able to compute intramolecular forces to high accuracies at the computational expense of a Hartree−Fock calculation. On the analysis side, the routine tools are the normal-mode analysis (NMA) techniques based on an expansion of the potential energy surface around the equilibrium structure, i.e., the minimum structure at zero temperature. For medium-sized quasi-rigid molecules in the gas phase, such NMA methods based on the harmonic approximation can reach excellent accuracy if method specific scaling factors are employed.[12] In addition to the frequency spectrum, these methods deliver the normal modes of intramolecular motion and the approximate intensities for IR and Raman absorption.

However, finite temperature effects, such as shifting and broadening of bands, due to the fluctuating structure of the molecule or its solvent environment in condensed phase are beyond the scope of traditional NMA methods. Here, proper sampling of these fluctuations is indispensable, and, thus, molecular dynamics (MD) simulations are the method of choice whenever the system is too large to allow a quantum mechanical treatment of the nuclei.

Within MD simulations the calculation of linear absorption cross-sections is straightforward, as it relates to the Fourier transform of the dipole autocorrelation function, if a proper quantum correction is employed.[3,5,9−11,13−16] Also peaks in the power spectra of the atomic velocities give a reasonable overview of frequencies that are important for the intramolecular motion.[16,17]

Assigning these bands to vibrational motifs is by no means straightforward. A simple but often sufficient approach for band assignment is to project the trajectory on some internal coordinates of the molecule and to take the Fourier transforms of their correlation functions.[5,7,18] This approach is of course neither universal nor complete. Much more involved are instantaneous normal-mode techniques[19−21] and the instantaneous normal-mode analysis (INMA),[13,22−25] for which one computes full or partial Hessians of the system along the trajectory and then

Received: February 22, 2011
Published: May 10, 2011

averages over the corresponding frequencies and intensities. To avoid negative frequencies, the INMA technique employs local structural relaxations prior to the calculation of the Hessian. These methods get quite involved, when no analytic second derivatives are at hand, and thus, the number of sampling points is limited for larger molecules.

It is much more appealing to use the information encoded in the MD trajectory directly to construct the normal modes, as such an approach only involves an a posteriori effort for the analysis. Here, advances like the principal mode analysis (PMA)[22,26−30] and related methods[16,31] use cross-correlations of Cartesian coordinates or momenta to derive eigenvalue equations for the normal modes. Particularly Martinez et al. provide a general framework of Fourier transformed time correlation functions of coordinates and momenta and relate these to both NMA and PMA.[16,31]

These methods rely to a lesser or greater extent on the equipartition assumption for the construction of the normal modes. As a result, if equipartition is violated, then the normal coordinates given by Martinez et al. are in general not orthonormal transforms of mass-weighted coordinates, even for a purely harmonic system.[16,31] As a remedy the authors suggest to orthonormalize the normal coordinate vectors a posteriori. Furthermore, Schmitz and Tavan show that for methods relying on the diagonalization of correlation matrices, one can get artificial mixing of normal coordinates.[22,28] For PMA this is the case if $T_k/\omega_k^2 \approx T_l/\omega_l^2$, i.e., if the ratios of the mode local temperatures T_k and squared eigenfrequencies ω_k of two modes become comparable.

In this article we combine ideas of these cited works and describe a method to determine generalized normal coordinates directly from (ab initio) MD simulations. The method does not rely on an equipartition assumption and is, therefore, also applicable if even approximate equipartition cannot be reached. We define these generalized normal coordinates as orthogonal linear combinations of mass-weighted Cartesians, which minimize their mutual correlation. To measure this correlation, we define a functional based on a tensorial formulation of the vibrational density of states (VDOS) and provide a corresponding minimization strategy based on the Jacobi diagonalization. Employing the convolution theorem we show how these correlation functions can be efficiently calculated in the Fourier domain.

For the analysis in a molecular frame of reference, the Eckart frame,[32] we compare the transformation of coordinates and velocities via a mass-weighted root-mean-square deviation (rmsd) fit and, alternatively, the transform via internal coordinates. In particular, the performance of both methods for large amplitude motions, such as the rotation of a methyl group, are examined.

Furthermore, we derive mode specific IR intensities, which are efficiently and accurately calculated within the framework of correlation functions in Fourier space. These mode local intensities decompose the global IR spectrum obtained form the dipole auto-correlation functions and allow to assign peaks in the IR spectrum to normal-mode vibrations. Furthermore, these local intensities permit us to correct the overall IR spectrum a posteriori toward the ensemble limit.

As a non−trivial application we examine the vibrational modes of isoprene (*trans*-2-methyl-1,3-butadiene) in the gas phase. To mimic the gas phase, we consider the isolated but rotating molecule, whose angular momentum is sampled by 41 initial conditions from a canonical distribution. Thereby we include the rovibrational couplings, albeit classical, in our model and can observe rotationally induced frequency shifts and line broadenings. These effects are not accessible by standard frequency calculations that ignore molecular rotation.

Isoprene has been chosen because it is an important building block of biologically important dye molecules, such as carotenoids and retinal. In particular, the latter is an important probe for vibrational spectroscopists, as it plays a central role in bacterial photosynthesis and mammalian vision. Still, the decoding of changes in the corresponding spectra due to changes in the environment and in the molecule remains a challenge to theory.[2,33] For isoprene we are, therefore, particularly interested in how well the C−C and C═C stretching modes, as well as the hydrogen out of plane (HOOP) modes, are described by the ab initio MD simulations, as these modes serve as marker modes for this class of molecules. Furthermore, we demonstrate how our analysis aids to check the convergence of the computed IR spectra with respect to sampling and examine the quality of the suggested temperature correction.

Beyond the comparatively well-behaved isoprene molecule, an extended version of the algorithm, which can deal with multiple reference structures and conformations, has been already employed for the vibrational analysis of floppy molecules, like CH_5^+ and its isotopologues, as well as for microsolvated hydronium and Zundel ions,[9,11,34] which exhibit large amplitude motion during the dynamics. Here, the algorithm has demonstrated its ability to deal with such challenging problems. The necessary extension to multiple reference structures is, however, beyond the scope of this article and will be described in a subsequent publication.

2. THEORY

We start our derivation of generalized normal coordinates with the basic relations of atomic velocities, kinetic energy, and temperature, which ultimately lead us to a tensorial definition of the VDOS. Close examination of the latter object in the case of purely harmonic motion will result in a principle of minimal cross-correlation between generalized normal coordinates, together with a corresponding minimization procedure. After discussing the transform to the Eckart frame employing an rmsd fit or internal coordinates, we give a computationally convenient recipe to assign individual IR intensities to the vibrational modes.

2.1. Vibrational Density of States (VDOS). MD simulations render the time evolution of a system from time t_i to t_f by a finite trajectory of Cartesian coordinates $\mathbf{x}(t)$: $[t_i,t_f] \rightarrow \mathbb{R}^n$ and velocities $\mathbf{v}(t) = \dot{\mathbf{x}}(t)$ of n DOFs, where the time derivative of a property a is denoted by $\dot{a}$. A molecular system of N atoms consists of $n = 3N$ Cartesian coordinates $\mathbf{x} = (\mathbf{r}_1^T, ..., \mathbf{r}_N^T)^T$, where the superscript 'T' denotes a matrix transpose. In the course of our vibrational analysis we will make extensive use of the Fourier transform (FT):

$$\hat{a}(\omega) = \int_{t_i}^{t_f} a(t) e^{-i\omega t} dt \tag{1}$$

of dynamical properties, such as velocities and dipole moments, where we mark the FT of a property by the caret accent. In order to write the finite FT integral as an infinite one we introduce a window function $W(t)$ and define the FT integral as

$$\hat{a}(\omega) = \int W(t) a(t) e^{-i\omega t} dt \tag{2}$$

In a rather general fashion, we demand that W is non-negative, bounded and that

$$\mathrm{supp}(W) \subseteq [t_\mathrm{i}, t_\mathrm{f}] \tag{3}$$

i.e., $W(t) = 0$ outside the interval $[t_\mathrm{i},t_\mathrm{f}]$. Within these limits, W can additionally serve to remedy artifacts of the FT, which arise from the finite simulation length and the discrete sampling.[35,36] The simplest form of W is a rectangular window, which can be trivially realized[35] with the Heaviside step function θ as

$$\mathrm{rect}(t; t_\mathrm{i}, t_\mathrm{f}) = \theta(t - t_\mathrm{i})\theta(t_\mathrm{f} - t) \tag{4}$$

Having established the basic notation we start the derivation of our vibrational analysis by looking at the kinetic energy of the system along the trajectory, which is given by

$$E_\mathrm{kin}(t) = \frac{1}{2}\,\mathbf{v}^\mathrm{T}(t)\mathbf{M}\mathbf{v}(t) \tag{5}$$

where the mass matrix $\mathbf{M}$ is diagonal with elements $M_{kl} = \delta_{kl}m_k$ containing masses m_k associated with each DOF k, δ being the Kronecker symbol. From the theory of small vibrations we introduce mass-weighted coordinates:

$$\mathbf{c}(t) = \mathbf{M}^{1/2}\mathbf{x}(t) \tag{6}$$

which simplifies the kinetic energy:

$$E_\mathrm{kin}(t) = \frac{1}{2}\,\dot{\mathbf{c}}(t)\cdot\dot{\mathbf{c}}(t) \tag{7}$$

to comprise only a scalar product of the mass-weighted velocities. The average temperature T of the system is computed from the time average of the kinetic energy by

$$T = \frac{2}{k_\mathrm{B}(n - n_\mathrm{c})}\frac{\int W^2(t)\dot{\mathbf{c}}(t)\cdot\dot{\mathbf{c}}(t)\mathrm{d}t}{\int W^2(t)\mathrm{d}t} \tag{8}$$

where n_c denotes the number of constrained DOFs, e.g., the global translation or rotation. Here we used the rather unusual weighting function $W^2(t)$ for which the choice $W(t) = \mathrm{rect}(t)$ leads, however, straight to the common temperature definition, as $\mathrm{rect}^2 = \mathrm{rect}$. Due to the weighting function $W^2(t)$ we can now employ Plancherel's theorem,[37] which states that the FT transform preserves scalar products of functions and write the temperature:

$$T = \frac{2}{k_\mathrm{B}(n - n_\mathrm{c})}\frac{\int \hat{\dot{\mathbf{c}}}^*(\omega)\cdot\hat{\dot{\mathbf{c}}}(\omega)\mathrm{d}\omega}{\int \hat{W}^*(\omega)\hat{W}(\omega)\mathrm{d}\omega} \tag{9}$$

in the Fourier domain (see eq 2), where the asterisk marks the complex conjugate. From the integral kernel of the numerator, which is an even function in ω, we readily derive the scalar VDOS as

$$\kappa(\omega) = \frac{4}{k_\mathrm{B}T}\frac{\hat{\dot{\mathbf{c}}}^*(\omega)\cdot\hat{\dot{\mathbf{c}}}(\omega)}{\int \hat{W}^*(\omega)\hat{W}(\omega)\mathrm{d}\omega} \tag{10}$$

which is normalized such that

$$\int_0^\infty \kappa(\omega)\mathrm{d}\omega = n - n_\mathrm{c} \tag{11}$$

Thus, in the ensemble limit, where each and every DOF contributes equally to $\kappa(\omega)$ due to equipartition, the integral:

$$n[\omega_s, \omega_e] = \int_{\omega_s}^{\omega_e} \kappa(\omega)\mathrm{d}\omega \tag{12}$$

"counts" the number of vibrational modes contributing with their kinetic energy to the interval $[\omega_s,\omega_e]$. For this property, it is essential to include the mass dependence of the kinetic energy in the definition of the VDOS in eq 10. Note that $\kappa(\omega)$ is the Fourier transform of the time correlation function of the mass-weighted velocities, due to the cross-correlation theorem:

$$\hat{a}^*(\omega)\hat{b}(\omega) = \iint a^*(\tau)b(\tau + t)\mathrm{d}\tau\mathrm{e}^{-i\omega t}\mathrm{d}t \tag{13}$$

which is closely related to the convolution theorem.

For an assignment of vibrational modes to bands in the VDOS we have to analyze the individual contributions of the atomic DOFs. For this purpose, we define the tensorial VDOS $\boldsymbol{\Theta}$: $\rightarrow \mathbb{R}^n \times \mathbb{R}^n$ as

$$\boldsymbol{\Theta}(\omega) = \frac{2}{k_\mathrm{B}T}\frac{\hat{\dot{\mathbf{c}}}^*(\omega)\otimes\hat{\dot{\mathbf{c}}}(\omega) + \hat{\dot{\mathbf{c}}}(\omega)\otimes\hat{\dot{\mathbf{c}}}^*(\omega)}{\int \hat{W}^*(\omega)\hat{W}(\omega)\mathrm{d}\omega}$$

$$= \frac{4}{k_\mathrm{B}T}\frac{\mathscr{R}[\hat{\dot{\mathbf{c}}}^*(\omega)\otimes\hat{\dot{\mathbf{c}}}(\omega)]}{\int \hat{W}^*(\omega)\hat{W}(\omega)\mathrm{d}\omega}$$

Here, we have replaced the scalar product in the definition of $\kappa(\omega)$, given in eq 10, by the outer or tensorial product $\hat{\dot{\mathbf{c}}}\otimes\hat{\dot{\mathbf{c}}} = \hat{\dot{\mathbf{c}}}\hat{\dot{\mathbf{c}}}^T$ and inserted the definition of the temperature from eq 8. Furthermore, we have included the symmetrization $\mathscr{R}(\hat{\dot{\mathbf{c}}}^*\otimes\hat{\dot{\mathbf{c}}}) = 1/2(\hat{\dot{\mathbf{c}}}^*\otimes\hat{\dot{\mathbf{c}}} + \hat{\dot{\mathbf{c}}}\otimes\hat{\dot{\mathbf{c}}}^*)$ to ensure that the off−diagonal elements of $\boldsymbol{\Theta}(\omega)$ are real valued. This symmetrization corresponds to averaging over the time forward and the time reversed trajectories, $\mathbf{x}(t)$ and $\mathbf{x}(-t)$, which are both valid trajectories in time-reversible dynamics.

The diagonal elements $\Theta_{kk}(\omega)$ are related to $\kappa(\omega)$ by the trace:

$$\kappa(\omega) = \sum_k \Theta_{kk}(\omega) \tag{14}$$

and, therefore, $\Theta_{kk}(\omega)$ yields the individual contribution to $\kappa(\omega)$ by DOF k. Since κ is derived from the global temperature eq 8 the integral:

$$\int_0^\infty \Theta_{kk}(\omega)\mathrm{d}\omega = \frac{T_k}{T} \tag{15}$$

furnishes the temperature T_k of DOF k.

The off−diagonal elements $\Theta_{k\neq l}(\omega)$ contain the cross-correlations between the k-th and l-th DOF. Note that $\boldsymbol{\Theta}(\omega)$ is an extension of the Fourier transformed time correlation function of the velocities $\mathbf{P}^{\dot{\mathbf{x}}}(\omega)$ defined by Martinez et al.,[31] for which we have included the mass matrix $\mathbf{M}$ as a metric tensor in the definition of the correlation function.

2.2. VDOS in the Harmonic Case. In order to establish a procedure of assigning the vibrational bands in $\kappa(\omega)$ to atomic motion we first focus on the properties of $\boldsymbol{\Theta}(\omega)$ in the case of nonergodic sampling of a purely harmonic system with $n_\mathrm{c} = 0$. From the harmonic analysis we introduce the orthogonal transform $\mathbf{D}^\mathrm{h} \in O_n$ of the mass-weighted coordinates, with O_n

denoting the orthogonal group on $\mathbb{R}^n$, to the so-called normal coordinates:

$$\mathbf{q} = \mathbf{D}^{\mathrm{h}}\mathbf{c} \tag{16}$$

which factorize the Hamiltonian of the system as

$$H = \frac{1}{2}\sum_{k=1}^{n}(\dot{q}_k^2 + \omega_k^2 q_k^2) \tag{17}$$

with the eigenfrequencies $\omega_k > 0$.[38] Note that $\mathbf{D}^{\mathrm{h}}$ is not unique, because a mere permutation of the coordinate numbering or a linear combination of modes that are degenerate in frequency does not change the physics of the system but leads formally to a different $\mathbf{D}^{\mathrm{h}}$. The resulting classical equations of motion are solved by

$$q_k(t) = a_k \sin(\omega_k t + \varphi_k) \tag{18}$$

in which the set of amplitudes a_k and phases φ_k depends on the initial conditions.

The finite FT of the normal coordinate velocities are then given as

$$\hat{\dot{q}}_k = \frac{a_k\omega_k}{2}\left[e^{i\phi_k}\hat{W}(\omega-\omega_k) + e^{-i\phi_k}\hat{W}(\omega+\omega_k)\right] \tag{19}$$

where $\hat{W}$ is the FT of the window function defined in eq 2. For example, a rectangular window function has the FT:[35]

$$\int \mathrm{rect}(t;t_i,t_f)e^{-i\omega t}\mathrm{d}t = e^{-i\omega\frac{t_i+t_f}{2}}\frac{\sin[(t_f-t_i)\omega/2]}{\omega/2} \tag{20}$$

The elements of $\boldsymbol{\Theta}(\omega)$ evaluate to

$$\Theta_{kl}(\omega) = \frac{2na_ka_l\omega_k\omega_l}{\sum_j a_j^2\omega_j^2} \times \frac{\cos(\phi_{kl})\mathscr{R}[\hat{Y}_{kl}(\omega)] - \sin(\phi_{kl})\mathscr{I}[\hat{Y}_{kl}(\omega)]}{\int \hat{W}^*(\omega)\hat{W}(\omega)\mathrm{d}\omega}$$

with the abbreviations $\varphi_{kl} = \varphi_k - \varphi_l$ and

$$\hat{Y}_{kl}(\omega) = \hat{W}(\omega-\omega_k)\hat{W}^*(\omega-\omega_l) + \hat{W}^*(\omega+\omega_k)\hat{W}(\omega+\omega_l) \tag{21}$$

Here, we have assumed that $\hat{W}(\omega+\omega_k)\hat{W}(\omega-\omega_l) \approx 0$, i.e., that bands belonging to positive and negative frequencies do not overlap. The mode temperatures eq 15:

$$T_k = T\frac{na_k^2\omega_k^2}{\sum_l a_l^2\omega_l^2} \tag{22}$$

are solely determined by the initial conditions and are independent of the sampling length in the harmonic case.

The corresponding integrals of the off-diagonal elements $\Theta_{k\neq l}$ vanish, if the overlap between $\hat{W}(\omega-\omega_k)$ and $\hat{W}(\omega-\omega_l)$ is negligible. Moreover, also the two-norm:

$$\| \Theta_{kl} \| = \left(\int [\Theta_{kl}(\omega)]^2\mathrm{d}\omega\right)^{1/2} \tag{23}$$

of $\Theta_{k\neq l}$ vanishes in this case, which simply reflects that the q_k and q_l are uncorrelated.

If we employ, instead of $\mathbf{D}^{\mathrm{h}}$, an arbitrary transform $\mathbf{D} \in O_n$ on the mass-weighted coordinates, the $\|\Theta_{k\neq l}\|$ will be in general nonzero.

In the case of finite overlap between $\hat{W}(\omega-\omega_k)$ and $\hat{W}(\omega-\omega_l)$, e.g., when modes k and l are degenerate, the $\|\Theta_{k\neq l}\|$ of the normal coordinates will, in general, not vanish for a single trajectory. However, for the ensemble average over many initial conditions we find $\|\Theta_{k\neq l}\| = 0$, due to the phase factors $\cos(\varphi_{kl})$ and $\sin(\varphi_{kl})$ in eq 21.

2.3. Generalized Normal Coordinates. Our findings for the purely harmonic case in the previous section now lead us to a generalization of normal coordinates for anharmonic Hamiltonians. We define the measure of the overall intermode coupling as the functional:

$$\mathrm{off}[\boldsymbol{\Theta}(\omega)] = \left[\sum_{k\neq l}\int [\boldsymbol{\Theta}_{kl}(\omega)]^2\mathrm{d}\omega\right]^{1/2} \tag{24}$$

which vanishes for the ensemble average for the normal coordinates of a strictly harmonic system. For the general anharmonic case we call

$$\mathbf{q}^{\mathrm{g}} = \{\mathbf{Dc}\,\big|\,\min_{\mathbf{D}\in O_n}\mathrm{off}[\mathbf{D}\boldsymbol{\Theta}(\omega)\mathbf{D}^{-1}]\} \tag{25}$$

generalized normal coordinates, which are equivalent to the standard normal coordinates in the harmonic limit. Note that $\boldsymbol{\Theta}$ transforms as

$$\boldsymbol{\Theta}'(\omega) = \mathbf{D}\boldsymbol{\Theta}(\omega)\mathbf{D}^{-1} \tag{26}$$

for any $\mathbf{D}\in O_n$, and since the trace is invariant under orthonormal transforms, $\kappa(\omega)$ is preserved, see eq 14.

If $\boldsymbol{\Theta}$ would be a regular matrix, then the task of minimizing off$[\boldsymbol{\Theta}]$ would correspond to diagonalizing $\boldsymbol{\Theta}$. Due to the frequency dependence of $\boldsymbol{\Theta}$ this is not achievable by a single transform for all values of ω simultaneously. Hence, we have to find the transform which brings $\boldsymbol{\Theta}(\omega)$ close to diagonal form for all ω in a balanced fashion. For this purpose, we modify the well-known and robust Jacobi diagonalization algorithm.[39,40] The Jacobi diagonalization relies on the idea that any $\mathbf{D} \in O_n$ can be written as a product:

$$\mathbf{D} = \prod_{k=1}^{j}\mathbf{J}(u_k, v_k, \vartheta_k) \tag{27}$$

of j Jacobi (or Givens) rotations:

$$\mathbf{J}(u,v,\vartheta) = \begin{pmatrix} & & & u & & v & & \\ & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ & \vdots & \ddots & \vdots & & \vdots & & \vdots \\ u & 0 & \cdots & c & \cdots & s & \cdots & 0 \\ & \vdots & & \vdots & \ddots & \vdots & & \vdots \\ v & 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ & \vdots & & \vdots & & \vdots & \ddots & \vdots \\ & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \tag{28}$$

with $s = \sin(\vartheta)$ and $c = \cos(\vartheta)$. Thus, multiplying $\mathbf{x}$ by $\mathbf{J}(u,v,\vartheta)$ applies a rotation in the u,v plane by the angle ϑ to $\mathbf{x}$. For the Jacobi diagonalization of a matrix $\mathbf{A}$ one chooses ϑ for a given pair u,v such that A'_{uv} vanishes after the transformation $\mathbf{A}' = \mathbf{JAJ}^T$ and

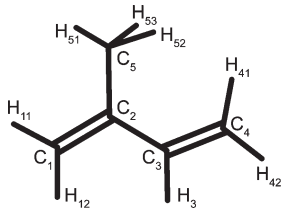

Figure 1. Structure and atomic labels of isoprene.

iterates the procedure over u and v. Considering eq 24 we see that the analogous approach is to minimize $\|\Theta_{u\neq v}\|$, since it is in general not possible to bring it to zero. Thus, the optimal rotation angle ϑ obeys

$$\frac{\partial}{\partial\vartheta'}\int [\mathbf{J}(u,v,\vartheta')\boldsymbol{\Theta}(\omega)\mathbf{J}(u,v,\vartheta')^T]_{uv}^2 \mathrm{d}\omega\Big|_{\vartheta'=\vartheta} = 0 \qquad (29)$$

This leads to an algebraic equation of fourth order, which is given in the Appendix together with its analytic roots ϑ_1, ..., ϑ_4. For each iteration complex solutions and values $|\vartheta_j| > \pi/4$ are discarded.[41]

For the matrix Jacobi algorithm one often picks the largest off-diagonal element to determine u,v for the next step.[41] Since, for our purpose, we cannot easily judge if such an element can be further reduced, we loop over all possible pairs $u < v$, i.e., perform a Jacobi sweep. When convergence is reached after multiple sweeps, $\mathbf{D}^g$ is reconstructed via eq 27.

2.4. Molecular Frame of Reference. A molecular frame of reference is crucial for any approach that constructs normal modes from atomic motion. Within the frame of reference one, furthermore, assumes that the molecule is quasi-rigid, i.e., that the intramolecular motions describe small oscillations around a single reference structure $\mathbf{x}^0$, which corresponds to a local minimum and does not undergo any conformational changes.

During a simulation molecules rotate and translate in general. Therefore, for a vibrational analysis one transforms the molecular coordinates to the Eckart frame of reference connected to an average molecular structure.[28,31,32] This is achieved by applying a mass-weighted root-mean-square (rms) fit to the Cartesian coordinates of a reference structure $\mathbf{x}^0$. Here, one determines new atomic coordinates $\mathbf{r}'_i = \mathbf{R}\mathbf{r}_i + \mathbf{l}$ by a translation vector $\mathbf{l}$ and a rotation matrix $\mathbf{R} \in O_3$ for each time step, which minimize the mass-weighted distance:

$$d_{\mathrm{rms}} = \sqrt{\frac{\sum_{i=1}^{N} m_i(\mathbf{r}'_i - \mathbf{r}_i^0)^2}{\sum_{i=1}^{N} m_i}} \qquad (30)$$

and transforms the velocities accordingly.

However, if the assumption of quasi-rigidity does not apply globally, then problems with the rms approach can arise, which we discuss on our test molecule isoprene, which is displayed in Figure 1. Here, the methyl group at the C_5 position rotates nearly freely around C_2—C_5 bond at ambient conditions due to the shallow rotational barrier. For the hydrogens this corresponds to a large amplitude motion. In particular, rotations of 120° around this bond preserve the conformation of the molecule but lead to a large local rmsd because of the underlying classical dynamics, for which the methyl hydrogens are discriminable, in contrast to a quantum mechanical treatment. Correspondingly, averaging over the motion of the hydrogens will render poor results.

An alternative to a rmsd fit which may help to mend these problems is to describe the intramolecular DOF by a complete nonredundant set of internal coordinates.[16] These can comprise primitive coordinates, such as bond distances, bond and dihedral angles, or linear combinations of such primitives.[42] For example the rotation of a methyl group can be described by a sum of dihedral angles, which is invariant under rotations of 120°. For a molecule we need $n - n_c$ linearly independent internal coordinates $s_i(\mathbf{x})$. Thus $\mathbf{s}(\mathbf{x}) = [s_1(\mathbf{x}), ..., s_{n-n_c}(\mathbf{x})]^T$ describes a transform to curvilinear coordinates. This transform eliminates the translational and rotational DOFs.

Due to the restriction of our vibrational analysis to orthonormal transforms of mass-weighted Cartesians, we need to map this motion of internal coordinates back to Cartesian displacements from the reference structure, i.e., to the Eckart frame. However, a closed reverse transform to $\mathbf{x}$ is not possible, since $\mathbf{s}$ and $\mathbf{x}$ are of different dimensions.

Because we actually need only a transform to the mass-weighted velocities $\dot{\mathbf{c}}$, we start from the kinetic energy:

$$E_{\mathrm{kin}} = \frac{1}{2}\dot{\mathbf{s}}^T\mathbf{G}^{-1}\dot{\mathbf{s}} \qquad (31)$$

of the internal coordinate velocities $\dot{\mathbf{s}}$, where the matrix $\mathbf{G} = \mathbf{B}^T\mathbf{M}^{-1}\mathbf{B}$ is constructed from the derivative of the internal coordinates:

$$B_{ij} = \left.\frac{\partial s_j}{\partial x_i}\right|_{x^0} \qquad (32)$$

at the reference structure $\mathbf{x}^0$ and the mass matrix $\mathbf{M}$.[38]

We can factorize $\mathbf{G} = \mathbf{K}\mathbf{K}^T$ by

$$\mathbf{K} = (\mathbf{M}^{-1/2}\mathbf{B})^T \qquad (33)$$

and correspondingly the inverse $\mathbf{G}^{-1} = (\mathbf{K}^+)^T\mathbf{K}^+$ by the pseudoinverse:

$$\mathbf{K}^+ = \mathbf{K}^T(\mathbf{K}\mathbf{K}^T)^{-1} \qquad (34)$$

of $\mathbf{K}$, which satisfies $\mathbf{K}\mathbf{K}^+ = \mathbf{E}_{n-n_c}$. Defining the transformed mass-weighted velocities:

$$\dot{\mathbf{c}}' = \mathbf{K}^+\dot{\mathbf{s}} \qquad (35)$$

yields the correct kinetic energy, which can be checked by inserting these definitions into eqs 7 and 31.

This procedure introduces an error in second order due to the expansion of the internal coordinates around $\mathbf{x}^0$. The magnitude of this error mainly depends on the choice of $\mathbf{x}^0$ and the internal coordinates $\mathbf{s}$.[42,43] In the Results Section we will assess this error introduced by this forward and backward transform by comparing the resulting $\kappa(\omega)$ for both the rms fit and the internal coordinate transform.

This representation of intramolecular motion allows a characterization of the normal coordinates with respect to the internal coordinates by $\dot{\mathbf{q}} = \mathbf{D}^g\mathbf{K}^+\dot{\mathbf{s}}$. Correspondingly, the displacement of internal coordinate j associated with normal coordinate k is

$$\left.\frac{\partial q_k}{\partial s_j}\right|_{\mathbf{x}^0} = (\mathbf{D}^g\mathbf{K}^+)_{kj} \qquad (36)$$

Normalization of the contributions of s_j to all q_i to account for the different amplitudes of the s_j yields the relative contribution:

$$\lambda_{kj} = \frac{\partial q_k}{\partial s_j}\left[\sum_i \left(\frac{\partial q_i}{\partial s_j}\right)^2\right]^{-1/2} \tag{37}$$

of s_j to q_k.

2.5. Assignment of IR Intensities. What remains after the construction of the generalized normal coordinates is their assignment to bands in experimental spectra.[16] For the case of IR spectra we now develop a convenient procedure to compute the individual absorption of these modes. From a classical dynamics trajectory we approximate the tensorial IR absorption cross-section of our finite sample as[14]

$$\alpha^{\mathrm{QC}}(\omega) = \left[\frac{2\pi\omega}{3V\hbar c_0 \hat{n}(\omega)}\right]\left(1-\exp\left(-\frac{\hbar\omega}{k_\mathrm{B}T}\right)\right) \times Q(\omega)\frac{\mathscr{R}[\hat{\boldsymbol{\mu}}^*(\omega)\otimes\hat{\boldsymbol{\mu}}(\omega)]}{\int \hat{W}^*(\omega)\hat{W}(\omega)\mathrm{d}\omega} \tag{38}$$

where V denotes the sample volume, c_0 the speed of light, and $\hat{n}(\omega)$ is the defraction coefficient. Once again, we have used the cross-correlation theorem to express the FT of the time correlation function of the finite dipole trajectory[14] $\boldsymbol{\mu}(t)$ as the (tensorial) product of its FT $\hat{\boldsymbol{\mu}}(\omega)$ and normalized it similar to the definitions of $\kappa(\omega)$ and $\boldsymbol{\Theta}(\omega)$ in eqs 10 and 14. The quantum correction factor $Q(\omega)$ is usually introduced as an ad hoc correction that essentially imposes the detailed balance condition, which proper quantum time correlation functions satisfy, onto the FT of the classical correlation function.[14] For systems without orientational order, such as molecules in the gas phase, the scalar absorption is given as $\alpha^{\mathrm{QC}} = \sum_{k=1}^{3}\boldsymbol{\alpha}_{kk}^{\mathrm{QC}}$. For molecular systems the so-called harmonic approximation for the quantum correction factor:

$$Q^{\mathrm{HC}}(\omega) = \frac{1}{k_\mathrm{B}T}\frac{\hbar\omega}{1-\exp\left(-\frac{\hbar\omega}{k_\mathrm{B}T}\right)} \tag{39}$$

has proven to render reliable results for the mid-IR range.[5,13,14,22] Applying the harmonic quantum correction we rewrite eq 38 as

$$\alpha^{\mathrm{HC}}(\omega) = \frac{\gamma}{k_\mathrm{B}T\hat{n}(\omega)}\frac{\mathscr{R}[\hat{\mathbf{j}}^*(\omega)\otimes\hat{\mathbf{j}}(\omega)]}{\int \hat{W}^*(\omega)\hat{W}(\omega)\mathrm{d}\omega} \tag{40}$$

where the parameter $\gamma = 2\pi/(3Vc_0)$ collects the constants and where we have introduced the current[15,16] $\mathbf{j}(t) = \dot{\boldsymbol{\mu}}(t)$. For this equality we recall the identity $\hat{\dot{a}} = i\omega\hat{a}$ for the time derivative, and thus, $\hat{\mathbf{j}} = i\omega\hat{\boldsymbol{\mu}}$.

In order to derive IR intensities for our normal modes, we first have to transform $\boldsymbol{\mu}$ to the frame of reference given by $\mathbf{x}^0$. Here, the transformation matrix $\mathbf{R}$ determined for the coordinates along the trajectory is equally applied to $\boldsymbol{\mu}$. Then $\boldsymbol{\mu}$ is expanded to linear order[28,31] as

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu}_0 + \sum_k \mathbf{a}_k q_k(t) + \mathscr{O}(2) \tag{41}$$

where the yet undetermined expansion coefficients $\mathbf{a}_k$ correspond to the derivatives of the dipole moment along the normal coordinates q_k in the harmonic case. Defining the coefficient matrix $\mathbf{A} = (\mathbf{a}_1, ..., \mathbf{a}_n) \in \mathbb{R}^3 \times \mathbb{R}^n$ we write eq 41 as $\boldsymbol{\mu}(t) = \boldsymbol{\mu}_0 + \mathbf{A}\mathbf{q}(t)$ and get, due to linearity,

$$\hat{\mathbf{j}}(\omega) = \mathbf{A}\hat{\dot{\mathbf{q}}}(\omega) \tag{42}$$

For a numerically stable and tractable way to determine $\mathbf{A}$ within our framework, both sides of eq 42 are multiplied with $\otimes\hat{\dot{\mathbf{q}}}^*$ to get

$$\hat{\mathbf{j}}(\omega)\otimes\hat{\dot{\mathbf{q}}}^*(\omega) = \mathbf{A}[\hat{\dot{\mathbf{q}}}(\omega)\otimes\hat{\dot{\mathbf{q}}}^*(\omega)] \tag{43}$$

To relate $\mathbf{A}$ and $\boldsymbol{\Theta}$ we define the correlation function:

$$\mathbf{C}(\omega) = (n-n_\mathrm{c})\frac{\mathscr{R}[\hat{\mathbf{j}}(\omega)\otimes\hat{\mathbf{c}}^*(\omega)]}{\int \hat{\mathbf{c}}^*(\omega)\cdot\hat{\mathbf{c}}(\omega)\mathrm{d}\omega} \tag{44}$$

Analogous to the definition of $\boldsymbol{\Theta}$ in eq 14, taking the real value here corresponds to a symmetrization in time, which assumes that coordinates and dipoles are in phase. Note that $\mathbf{C}$ can be conveniently calculated along with the initial tensorial VDOS $\boldsymbol{\Theta}$. If we now transform $\mathbf{C}$ to the coordinate system defined by $\mathbf{D}^\mathrm{h}$, then we get with eqs 14 and 43:

$$\mathbf{C}(\omega)(\mathbf{D}^\mathrm{h})^{-1} = \mathbf{A}\boldsymbol{\Theta}(\omega) \tag{45}$$

The effective dipole gradient matrix $\mathbf{A}$ now is determined numerically by a least-squares fit as

$$\frac{\partial}{\partial\mathbf{A}}\int \mathrm{d}\omega\left\|\mathbf{A}\boldsymbol{\Theta}(\omega) - \mathbf{C}(\omega)(\mathbf{D}^\mathrm{h})^{-1}\right\|_2 = 0 \tag{46}$$

Note that we can safely restrict this problem to values of ω, where the IR absorption is finite, e.g., $\omega < 4000\ \mathrm{cm}^{-1}$.

The reconstructed IR absorption resulting from $\mathbf{A}$, eqs 38,40, and 42, is given by

$$\alpha'(\omega) = \frac{\gamma}{\hat{n}(\omega)}\mathbf{A}\boldsymbol{\Theta}(\omega)\mathbf{A}^T \tag{47}$$

where the normalizing factors in the respective definitions of $\mathbf{C}$ and $\boldsymbol{\alpha}$ cancel with $k_\mathrm{B}T$ according to eq 8. We can associate the contributions of the diagonal elements of $\boldsymbol{\Theta}$ to $\boldsymbol{\alpha}'$ given by

$$\alpha'_{kk}(\omega) = \frac{\gamma}{\hat{n}(\omega)}(\mathbf{a}_k\otimes\mathbf{a}_k)\Theta_{kk}(\omega) \tag{48}$$

with individual mode absorptions,[16] which describe both the band shapes and the total absorptions of the modes q_k. Note that the reconstructed IR absorption $\boldsymbol{\alpha}'(\omega)$ additionally includes cross-terms $\boldsymbol{\alpha}'_{kl}(\omega)$, which account for the remaining correlations between modes q_k and q_l.

Having decomposed the total IR absorption into individual contributions of the generalized normal coordinates and their intermode couplings, we can approximate the absorption of an ergodic ensemble. If we assume predominantly harmonic modes, then all cross-correlations $\Theta_{k\neq l}(\omega)$ should vanish. Furthermore, in the ergodic limit all modes should contribute according to the mean temperature of the system. Therefore, we estimate the temperature-corrected IR absorption in the ergodic limit as

$$\alpha^{\mathrm{TC}}(\omega) = \sum_{k:T_k>0}\frac{T}{T_k}\alpha'_{kk}(\omega) \tag{49}$$

Note that we have $T_k = 0$ for the constrained DOFs.

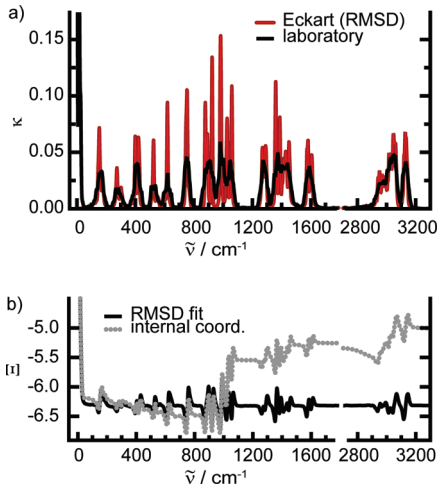

Figure 2. (a) VDOSs κ computed in the laboratory frame of reference (black) and in the Eckart (red) frame of reference obtained by an rmsd fit. (b) Integral difference Ξ (eq 50) computed for the transform to the Eckart frame via either rmsd fit (black line) or internal coordinates (gray dotted line).

3. METHODS

With the primary aim to demonstrate the usefulness of our novel method we have chosen the isoprene molecule, which is not only a basic building block of carotenoids and related dye molecules, like retinal in biological systems, but also serves as a realistic application of our approach.

Ab initio molecular dynamics (MD) simulations[44] within the framework of Kohn−Sham density functional theory (DFT) have been performed with the CP2k simulation package (see http://cp2k.berlios.de). Here, energies, forces, and dipole moments are computed "on the fly", and the system is propagated on the Born−Oppenheimer surface.

We have employed the BLYP functional[45,46] together with Goedecker−Teter−Hutter pseudopotentials[47,48] for core electrons and a TZV2P basis set[49] for the valence electrons. Within the quickstep algorithm[50,51] electrostatic interactions have been treated with a density cutoff of 280 Ry. The Poisson equation for the isolated isoprene molecule in a box of 12 Å side length has been solved by the Martyna−Tuckerman solver.[52] Geometry optimization yielded a reference structure $\mathbf{x}^0$, for which harmonic frequencies were calculated for comparison.

For the dynamics a set of 41 initial conditions has been sampled by an NVT simulation employing massive Nosé−Hoover chains at T = 300 K.[53] From these initial coordinates and velocities subsequent NVE runs of 25 ps each were run. The integration time step was 0.5 fs, and coordinates, velocities, and dipole moments were sampled each step and used for the vibrational analysis.

A set of nonredundant internal coordinates for the reference structure $\mathbf{x}^0$ has been obtained by the `TURBOMOLE v5.9` quantum chemistry package.[54]

The generalized normal coordinate algorithm has been implemented in the C code `normcor`. With standard linear algebra and fast FT libraries the fully converged minimization and normal coordinate analysis for isoprene took less than 10 min on an 3.0 GHz desktop. The optimization problem eq 25 converged to numerical accuracy after 13 Jacobi sweeps.

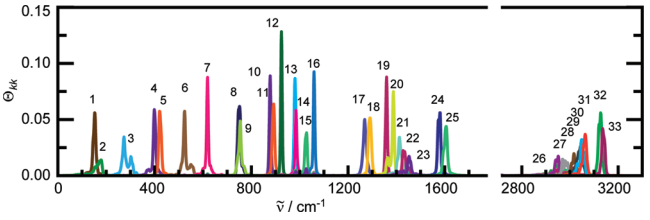

Figure 3. Power spectra $\Theta_{kk}(\omega)$ (colored lines) of the generalized normal coordinates denoted with labels 1−33.

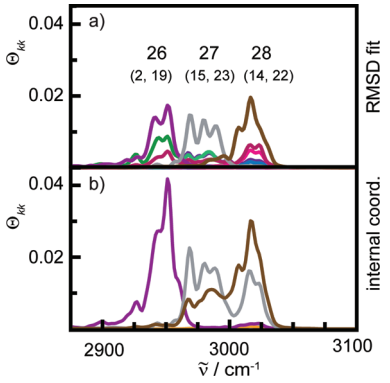

Figure 4. Power spectra $\Theta_{kk}(\omega)$ in the methyl stretching region for: (a) the rmsd fit and (b) the internal coordinate transform.

In order to remedy artifacts from the discrete and finite FT we employed a truncated Gaussian-type windowing function $W(t)$ (see eq 2), which corresponded to a convolution of the FTs by a Gaussian of width 5 cm^{-1} in the frequency domain.

4. RESULTS AND DISCUSSION

To set the stage for the normal coordinate analysis of isoprene, we first compare the VDOSs obtained in the laboratory $\kappa^{\mathrm{l}}(\omega)$ and in the Eckart frame of reference $\kappa^{\mathrm{E}}(\omega)$, which are shown in Figure 2a. After transforming the velocities to the Eckart frame, the intense peak below 50 cm^{-1} vanishes in $\kappa^{\mathrm{E}}(\omega)$, which is present in $\kappa^{\mathrm{l}}(\omega)$. These low-frequency contributions are associated with the rotational and translational DOFs of the molecule. Most other bands sharpen and more distinct features appear.

To check possible errors introduced by the transform to the Eckart frame the integral difference:

$$\Xi(\omega) = \int_0^{\omega} [\kappa^{\mathrm{E}}(\omega') - \kappa^{\mathrm{l}}(\omega')]\mathrm{d}\omega' \qquad (50)$$

for which $\kappa^{\mathrm{E}}(\omega)$ was obtained either by an rmsd fit or by the internal coordinate transform, is displayed in Figure 2b. Both integral differences drop to about −6.2 within the first 50 cm^{-1}, accounting for the elimination of the six translational and rotational DOFs. For higher frequencies, however, they behave differently: $\Xi(\omega)$ obtained for the rmsd-fitted VDOS displays oscillations around the level reached after the initial drop. These oscillations correspond to the sharpening of bands already observed in Figure 2a. In contrast, $\Xi(\omega)$ obtained by the internal coordinate transform displays a general drift atop of the oscillations and reaches a limit of −5.0 at 3200 cm^{-1}. Thus, with respect to the rmsd fit the VDOS based on the internal

Table 1. Peak Positions ω_p and Corresponding Harmonic Frequencies ω_h of the Vibrational Bands (in cm^{-1}) as well as the Corresponding Internal Coordinate Coefficients (cf. eq 37) Obtained from the Analysis Based on Internal Coordinates[a]

	ω_p	ω_h	internal coordinate contributions
1	152	145	0.99 C_2—C_3 tors. + 0.42 H_3 wag − 0.36 C_4H wag
2	161	171	0.96 Me tors.
3	274	277	0.85 C_2—C_3=C_4 bend + 0.80 C_5—C_2—C_3 def. + 0.43 C_1 rock + 0.38 C_2—C_3
4	399	398	0.86 C_2 outp. − 0.47 C_3=C_4 tors.
5	420	423	0.86 C_1 rock − 0.57 C_1H rock − 0.41 C_2—C_3
6	524	526	0.47 C_3=C_4 + 0.46 C_1=C_2 + 0.46 C_2—C_3=C_4 bend − 0.46 C_4 H rock
7	619	616	0.72 C_1=C_2 tors. + 0.47 C_3=C_4 tors.
8	751	746	0.65 C_1=C_2 tors. − 0.46 C_3=C_4 tors.
9	754	754	0.69 C_2—C_5 + 0.58 C_2—C_3 + 0.45 C_1=C_2
10	878	889	0.92 C_1H wag
11	893	899	0.91 C_4H wag
12	925	911	0.57 C_1 H rock + 0.45 C_2—C_5 − 0.36 C_4H rock
13	981	984	0.81 H_3 wag − 0.50 C_3=C_4 tors.
14	985	988	0.62 Me asym. def.′ + 0.59 Me rock′ − 0.48 Me asym. def. + 0.42 Me rock
15	1029	1024	0.71 Me asym. def. − 0.65 Me rock + 0.51 Me asym. def.′ + 0.47 Me rock′
16	1059	1056	0.61 C_4H rock − 0.38 H_3 rock
17	1269	1272	0.39 C_1H rock
18	1291	1287	0.79 H_3 rock − 0.39 C_3=C_4
19	1360	1364	0.90 Me sym. def.
20	1388	1378	0.70 C_1H def. + 0.49 C_4H def.
21	1414	1414	0.65 C_4H def. − 0.47 C_1H def.
22	1428	1423	0.54 Me rock + 0.39 Me asym. def.
23	1452	1450	0.53 Me rock′ − 0.41 Me asym. def.′
24	1580	1574	0.42 C_1=C_2 − 0.37 C_3=C_4
25	1605	1604	0.40 C_3=C_4 + 0.32 C_1=C_2
26	2951	2908	0.61 H_{53} + 0.60 H_{51} + 0.52 H_{52}
27	2968	2956	0.75 H_{53} − 0.59 H_{51}
28	3017	2999	0.82 H_{52} − 0.52 H_{51}
29	3045	3016	0.81 H_3 − 0.41 H_{12}
30	3049	3023	0.61 H_{12} + 0.55 H_{11} + 0.51 H_3
31	3063	3036	0.72 H_{42} + 0.66 H_{41}
32	3127	3103	0.72 H_{11} − 0.64 H_{12}
33	3135	3118	0.71 H_{41} − 0.64 H_{42}

[a] Hydrogen stretches H_i are marked by their atom labels i. The nomenclature, according to Pulay et al.,[42] for the internal coordinates of the methylenes is: def., $2\alpha - \beta_1 - \beta_2$; rock, $\beta_1 - \beta_2$; and wag, $H_1 + H_2$ out of plane and of the methyl group (Me) as: Me sym. def., $\alpha_1 + \alpha_2 + \alpha_3 - \beta_1 - \beta_2 - \beta_3$; Me asym. def., $2\alpha_1 - \alpha_2 - \alpha_3$; Me asym. def.′, $\alpha_2 - \alpha_3$; Me rock, $2\beta_1 - \beta_2 - \beta_3$; and Me rock′, $\beta_2 - \beta_3$, where the H−C−H angles are denoted as α_i and the H−C−C angles as β_i.

coordinate transform gains about one DOF over the full frequency range. Correspondingly, the error introduced by this transfrom is about 3% if we relate this to the number of internal DOFs of the molecule, 33. This error is not uniformly distributed but has major contributions around 1000, 1300, and 3000 cm^{-1}.

Having inspected the transforms to the Eckart frame, we turn to the contributions of the generalized normal coordinates to the respective VDOS, here based on the rmsd fit. The resulting power spectra $\Theta_{kk}(\omega)$ after minimizing off[$\boldsymbol{\Theta}$] are well localized in frequency space, as can be seen in Figure 3. Most band shapes are characterized by single sharp peaks, except for a few, which have small auxiliary bands like modes 3 or 6 or which cover larger frequency intervals, like the overlapping bands of modes 22 and 23 or modes 28−31. Some large peaks appearing in $\kappa(\omega)$ decompose into nearly degenerate and overlapping bands, e.g., modes 8 and 9 at 750 cm^{-1} or 32 and 33 at 3130 cm^{-1}.

However, not all parts of the VDOS spectrum are fully disentangled, which is shown in Figure 4a for modes 26−28. Although these modes dominate $\kappa(\omega)$ in this frequency region, many other modes (given in parentheses), whose main peaks are far away in the spectrum, yield sizable contributions. In contrast, the mode spectra obtained after the internal coordinate transform shown in Figure 4b do not exhibit such admixtures of low-lying modes. Here, however, modes 27 and 28 are still entangled, i.e., they cover the same spectral range and are bimodal.

For an assignment of these modes to atomic motion Table 1 provides the contributions of the internal coordinates s to the generalized normal coordinates q according to eq 37 based on the data obtained after the internal coordinate transform. Here, modes 26−28 are identified as the symmetric and the two asymmetric stretching modes of the methyl group, respectively. Seemingly the two asymmetric modes cannot be disentangled, whereas the symmetric stretch is clearly identified. The admixtures observed

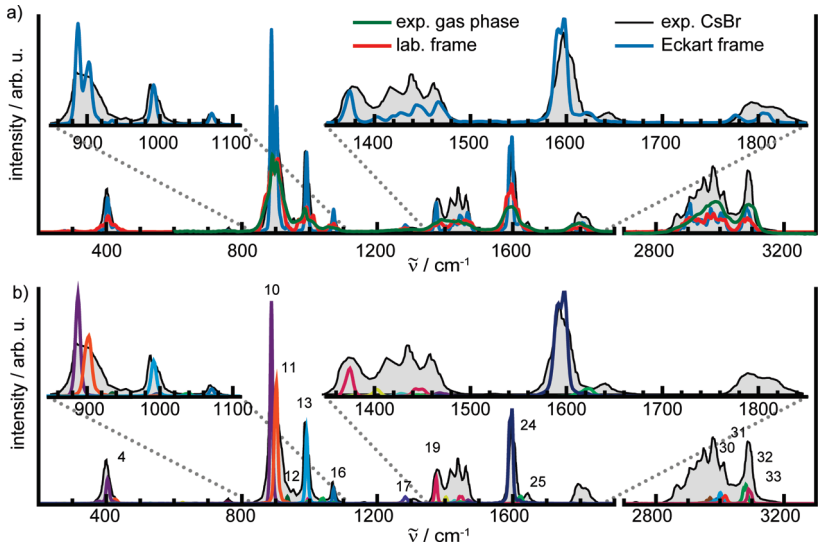

Figure 5. Comparison of the computed IR absorptions to experiment. Experimental data of isoprene in a Cs Br matrix (shaded area) was taken from ref 56, baseline corrected, and intensity scaled to match the dominant methylene wagging band at 900 cm^{-1}. Gas-phase data was obtained from the National Institute of Standards and Technology (NIST) database (green curve).[55] The frequencies of the computed spectra have been scaled by 1.015 in the lower part of the spectrum and by 0.985 in the C−H stretch region. Panel (a) compares the overall IR absorption cross-section in the laboratory frame (red curve) and in the Eckart frame (blue curve); insets zoom in on interesting spectral regions. Panel (b) compares the individual mode absorptions $\alpha_{kk}(\omega)$ (colored curves) to peaks of the experimental spectrum in CsBr.

for the modes based on the rmsd fitted trajectories stem from other modes of the methyl group: its rotation (mode 2), rocking and deformation (modes 14, 15, 19, 22, and 23). Although the mode classification seems to work better based on the internal coordinate transform, it is still not fully satisfying because the asymmetric modes are not disentangled. The reason is that rotations of the methyl group, which leave the conformation invariant but change the numbering of the hydrogens, are not properly accounted for in both methods. Only modes invariant under such rotations, i.e., the symmetric methyl stretch (mode 26) and the symmetric deformation (mode 19) are correctly identified. A drawback of the internal coordinate transform is that it notably modifies the VDOS in the Eckart frame, in particular near the vibrational contributions of the methyl group at 1000, 1400, and 3000 cm^{-1}, see Figure 2. An extension of our algorithm that can handle such permutations and resolves these problems will be presented in a subsequent publication.

The other modes listed in Table 1 besides those located at the methyl group are nicely characterized by the internal coordinate contributions. For example the modes 24 and 25 represent the asymmetric and symmetric C═C stretches. Also for many other modes in phase and out of phase combination of internal coordinates are found, e.g., modes 7 and 8, 20 and 21, or 31 and 33.

Table 1 furthermore compares the peak positions of the modes ω_p to the frequencies calculated by the harmonic approximation ω_h. The differences are only a few wavenumbers for the first 25 modes and do not exceed 10 cm^{-1}. Thus, anharmonic effects are small for this molecule and induce only subtle shifts. In contrast, for the hydrogen stretches modes 26 to 33 a systematic blue-shift on the order of 25 cm^{-1} is observed. This is somehow counterintuitive if we think of the C−H bond as a Morse-type binding potential, for which we would expect a red-shift due to anharmonicities. However, this blue-shift was no longer observed when we had set the angular momentum to zero within the initial conditions and rerun the MD simulations (data not shown). Thus, these blue-shifts are caused by couplings to rotation.

For comparison to experiment, the interactions of the molecule and its vibrational modes with radiation are required. Figure 5a shows the total IR absorption spectrum $\boldsymbol{\alpha}$ eq 38 computed in the laboratory frame (red curve) and in the Eckart frame of reference (blue curve), for which the dipole moment had been transformed along with the rmsd fit of the coordinates. The frequencies have been slightly scaled by 1.015 in the region below 1900 cm^{-1} and by 0.985 in the region above 2700 cm^{-1} to match experiment. The experimental data are a gas-phase spectrum of isoprene[55] and a solid-state spectrum obtained in a CsBr matrix.[56]

Comparing the spectra, we find a strikingly close agreement of the computed IR absorption in the laboratory frame (red curve) and the experimental gas-phase data (green curve): band positions, shapes, and relative intensities match almost perfectly. Thus, including molecular rotation within the classical dynamics seems to be sufficient to mimic the rotational broadening of the vibrational bands. Differences are only observed for the C═C stretching modes near 1600 cm^{-1}, whose computed band is slightly sharper and the hydrogen stretch region, where the overall intensity is smaller than in experiment. Note in particular that in the spectral regions of the methyl deformation and stretching modes near 1400 and 2900 cm^{-1}, the computed and the gas-phase spectrum nicely match.

After the dipole moments are transformed to the Eckart frame the computed IR absorption bands sharpen (blue curve), similar to the observations for the VDOS. Here, a comparison to an experimental spectrum that is not rotationally broadened is more appropriate, the spectrum in Cs Br (shaded area). For most peaks in the computed spectrum, we find a close correspondence in the experimental one, see also insets. However, the solid state clearly modifies the experimental spectrum of isoprene, in particular the intensities, e.g., of the hydrogen stretches or in the region between 1400 and 1500 cm^{-1}.

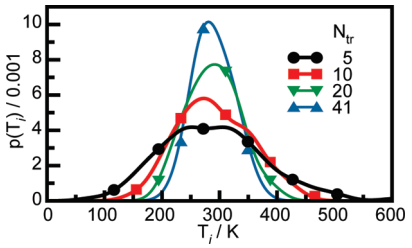

Figure 6. Mode temperature distributions depending on the number of trajectories *n*.

A mode assignment for these experimental peaks is given in Figure 5b by the IR absorption bands of the individual modes. Most of the peaks in the experimental spectrum are soundly assigned, e.g., the prominent peak of the methylene wagging modes at 900 cm^{-1}, see inset. However, the small band near 1800 cm^{-1} is missing in the reconstructed mode spectrum but present in the computed IR absorption cross-section in Figure 5a. Closer inspection reveals that this band is associated to the overtones of the methylene wagging modes. Because such overtones are not considered in the expansion of the dipole moment eq 41, they are not present in the reconstructed spectrum.

For the methyl modes we find similar deficiencies in the assignment of their IR intensities, as we have observed in the disentangling of these modes. Only the symmetric methyl deformation (mode 19) yields the proper IR absorption, whereas for the other methyl modes, it is grossly underestimated. Here, the sum of mode IR absorptions computed for the Eckart frame (blue line in Figure 5a) does not reach the total IR absorption in the laboratory frame. Again these problems are caused by not properly accounting for the rotation of the methyl group and can be only resolved by considering multiple reference structures.

A major advance of the algorithm presented here is that it provides the mode temperatures along with their generalized normal coordinates. Furthermore, these mode temperatures allow to assess the convergence of the computed IR spectrum. Figure 6 shows the distribution $p(T)$ of these mode temperatures for the full set of 41 trajectories (blue curve) obtained by a kernel density estimate. Here, we have represented each of the 33 mode temperatures by Gaussians of width σ = 24 K and normalized their sum. This resulting distribution has a standard deviation of about 39 K, which is about 13% of the mean temperature 300 K. This seems quite sizable if we recall that this distribution is based on more than 1 ns of ab initio trajectory.

We examined the convergence of this distribution by selecting blocks of N_{tr} = 5, 10, and 20 trajectories from the whole set and recalculated the mode temperatures and their distribution for each block size. These distributions are compared to the full set in Figure 6. The corresponding standard deviations amount to 91 K for N_{tr} = 5, 71 K for N_{tr} = 10, and 51 K for N_{tr} = 20. Their ratios approximately follow the rule that doubling the number of trajectories scales the standard deviation of the resulting mode temperature distribution by about $1/\sqrt{2}$, as is expected for the standard deviation. Note in particular, that if the number of trajectories is small one can have quite a few modes exhibiting a temperature twice the mean temperature or only half of it. According to eqs 15 and 48, however, this implies that the IR absorption of such modes is grossly over or underestimated.

Finally, we check how well the suggested temperature correction can remedy such deficiencies due to insufficient sampling.

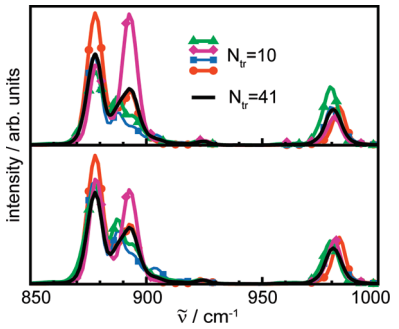

Figure 7. Uncorrected (upper panel) and temperature-corrected (lower panel) IR absorption for sets of 10 trajectories (colored curves) and all 41 trajectories (black curve).

Figure 7 compares the uncorrected and corrected spectrum for sets of 10 trajectories and for all 41 trajectories for the prominent double band of the methylene wagging modes, 11 and 12, and the HOOP mode, 13, at 981 cm^{-1}. For the latter the temperature correction brings the spectra of the small and the unconverged sets close to the one of the full set. In contrast, for the two overlapping methylene wagging bands, the discrepancies of intensities remain larger. However, the temperature corrections are qualitatively correct and shift the intensities toward the relative heights of the full set. Seemingly, to resolve such details extended sampling is inevitable. Nevertheless, the temperature correction can help to identify unconverged parts of the spectrum, in particular, if they are connected to statistical outliers of the mode temperatures.

5. CONCLUSIONS AND OUTLOOK

We have developed a consistent and automated procedure to deduce generalized normal coordinates from finite MD trajectories. It is based on the definition of the VDOS in a tensorial form and does not rely on any equipartition assumption. The generalized normal coordinates are defined as orthonormal transforms of mass-weighted coordinates, which exhibit minimal correlation as measured by the off-diagonal norm of the tensorial VDOS. A minimization scheme based on the Jacobi diagonalization algorithm robustly converges and renders the generalized normal coordinates, their band positions and shapes, their mode temperatures, and their remaining cross-correlations, which result from the finiteness of the sampling and anharmonic intermode couplings. Within this framework we have embedded the assignment of the vibrational modes to internal coordinate motion. Furthermore, mode-specific IR intensities are derived which also take cross-correlations into account.

As a non-trivial application of the algorithm, we presented the analysis of the IR spectrum of isoprene in the gas phase, i.e., an isolated but rotating molecule, which has been sampled by 41 ab initio MD trajectories. The algorithm renders rich and detailed information on the vibrational modes, especially their compositions in terms of internal coordinates and their respective absorption strength. Computed band positions and widths, which are rotationally broadened, nicely agree with an experimental gas-phase spectrum and assign the experimental peaks of a solid-state spectrum.

The distribution of mode temperatures still shows a sizable width if the full set of trajectories is used for the analysis and dramatically broadens for smaller sets. The corresponding temperature correction for the absorption qualitatively rectifies the

relative intensities for overlapping modes and even quantitatively for isolated peaks.

We have extended the algorithm presented here to manage multiple conformations of molecules and to resolve permutational symmetries, which will be described in a subsequent publication. This extension fully resolves the problems observed for the methyl group of isoprene in this article by employing three instead of one reference structure. Furthermore, it allows to treat large amplitude motion more complex than a methyl group rotation,[9,11,34] for which also multiple conformations of a molecule have to be considered.

APPENDIX A: OPTIMAL JACOBI ANGLE

To determine the optimal rotation angle ϑ as defined by eq 29, we first note that since the overall Frobenius norm of $\boldsymbol{\Theta}$ is invariant with respect to orthonormal transforms,[41] it is sufficient to consider just the term $N_{uv} = \int [\Theta_{uv}(\omega)]^2 \mathrm{d}\omega$. Thus, according to eqs 24 and 29 we have to solve

$$\left.\frac{\partial}{\partial \vartheta'} N'_{uv}\right|_{\vartheta' = \vartheta} = 0 \tag{51}$$

for the transformed term:

$$N'_{uv} = \int [(c^2 - s^2)\Theta_{uv}(\omega) + sc(\Theta_{uu}(\omega) - \Theta_{vv}(\omega))]^2 \mathrm{d}\omega \tag{52}$$

again with $c = \cos\vartheta$ and $s = \sin\vartheta$. With $t = \tan(\vartheta)$, the identities $c = 1/(t^2 + 1)^{1/2}$ and $s = tc$, and the integrals:

$$\begin{aligned} \chi_1 &= \int [\Theta_{uv}(\omega)]^2 \mathrm{d}\omega = N_{uv}, \\ \chi_2 &= \int \Theta_{uv}(\omega)[\Theta_{uu}(\omega) - \Theta_{vv}(\omega)]\mathrm{d}\omega, \\ \chi_3 &= \int [\Theta_{uu}(\omega) - \Theta_{vv}(\omega)]^2 \mathrm{d}\omega \end{aligned}$$

Equation 52 transforms to

$$N'_{uv} = \frac{\chi_1(t^2-1)^2 - 2\chi_2 t(t^2-1) + \chi_3 t^2}{(1+t^2)^2} \tag{53}$$

and, with $\eta = (4\chi_1 - \chi_3)/\chi_2$ its derivative is given by

$$\frac{\partial}{\partial t} N'_{uv} = 2\chi_2 \frac{(t^4 + \eta t^3 - 6t^2 - \eta t + 1)}{(1+t^2)^3} \tag{54}$$

which has the four roots

$$\begin{aligned} t_{kl} = \frac{1}{4}(&-\eta + (-1)^k \sqrt{16 + \eta^2} + (-1)^l \sqrt{2} \\ &\times \sqrt{16 + \eta^2 - (-1)^k \eta \sqrt{16 + \eta^2}}) \end{aligned} \tag{55}$$

$k,l \in \{1,2\}$. For our transform we choose the root t_{kl} with $|t_{kl}| \leq 1$ and

$$\begin{aligned} \left.\frac{\partial^2 N'_{uv}}{\partial t^2}\right|_{t = t_{kl}} = &-\frac{2\chi_2}{(t_{kl}^2+1)^4}[2t_{kl}(t_{kl}^4 - 14t_{kl}^2 + 9) \\ &+ \eta(3t_{kl}^4 - 8t_{kl}^2 + 1)] > 0 \end{aligned}$$

to obtain a minimum of N'_{uv}.

AUTHOR INFORMATION

Corresponding Author

*E-mail: gerald.mathias@physik.uni-muenchen.de.

Former Addresses

[§]Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany.

ACKNOWLEDGMENT

We thank Sergei Ivanov and Harald Forbert for helpful discussions and Dominik Marx for encouragement and support. Theodoros Zelleke is acknowledged for implementing the internal coordinate transform. Funding was provided by the Deutsche Forschungsgemeinschaft through grant MA 1547/3 to D.M. within FOR 436.

REFERENCES

(1) Nonella, M.; Mathias, G.; Eichinger, M.; Tavan, P. *J. Phys. Chem. B* **2003**, *107*, 316–322.
(2) Vogel, R.; Siebert, F.; Mathias, G.; Tavan, P.; Fan, G.; Sheves, M. *Biochemistry* **2003**, *42*, 9863–9874.
(3) Rousseau, R.; Kleinschmidt, V.; Schmitt, U. W.; Marx, D. *Angew. Chem., Int. Ed.* **2004**, *43*, 4804–4807.
(4) Asvany, O.; Kumar, P. P.; Redlich, B.; Hegemann, I.; Schlemmer, S.; Marx, D. *Science* **2005**, *309*, 1219–1222.
(5) Mathias, G.; Marx, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6980–6985.
(6) Masia, M.; Forbert, H.; Marx, D. *J. Phys. Chem. A* **2007**, *111*, 12181–12191.
(7) Baer, M.; Mathias, G.; Kuo, I.-F. W.; Tobias, D. J.; Mundy, C. J.; Marx, D. *ChemPhysChem* **2008**, *9*, 2703–2707.
(8) Cimas, A.; Maitre, P.; Ohanessian, G.; Gaigeot, M.-P. *J. Chem. Theory Comput.* **2009**, *5*, 2388–2400.
(9) Ivanov, S. D.; Asvany, O.; Witt, A.; Hugo, E.; Mathias, G.; Redlich, B.; Marx, D.; Schlemmer, S. *Nat. Chem.* **2010**, *2*, 298–302.
(10) Heyden, M.; Sun, J.; Funkner, S.; Mathias, G.; Forbert, H.; Havenith, M.; Marx, D. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 12068–12073.
(11) Baer, M.; Marx, D.; Mathias, G. *Angew. Chem., Int. Ed.* **2010**, *49*, 7346–7349.
(12) Neugebauer, J.; Hess, B. A. *J. Chem. Phys.* **2003**, *118*, 7215–7225.
(13) Nonella, M.; Mathias, G.; Tavan, P. *J. Phys. Chem. A* **2003**, *107*, 8638–8647.
(14) Ramírez, R.; López-Ciudad, T.; Kumar, P. P.; Marx, D. *J. Chem. Phys.* **2004**, *121*, 3973–3981.
(15) Gaigeot, M.-P.; Vuilleumier, R.; Sprik, M.; Borgis, D. *J. Chem. Theory Comput.* **2005**, *1*, 772–789.
(16) Gaigeot, M.-P.; Martinez, M.; Vuilleumier, R. *Mol. Phys.* **2007**, *105*, 2857–2878.
(17) Pauwels, E.; Verstraelen, T.; De Cooman, H.; Van Speybroeck, V.; Waroquier, M. *J. Phys. Chem. B* **2008**, *112*, 7618–7630.
(18) Kumar, P. P.; Marx, D. *Phys. Chem. Chem. Phys.* **2006**, *8*, 573–586.
(19) Stratt, R. M. *Acc. Chem. Res.* **1995**, *28*, 201–207.
(20) Keyes, T. *J. Phys. Chem. A* **1997**, *101*, 2921–2930.
(21) Kalbfleisch, T.; Keyes, T. *J. Chem. Phys.* **1998**, *108*, 7375–7383.
(22) Schmitz, M.; Tavan, P. *J. Chem. Phys.* **2004**, *121*, 12247–12258.
(23) Babitzki, G.; Mathias, G.; Tavan, P. *J. Phys. Chem. B* **2009**, *113*, 10496–10508.
(24) Miani, A.; Helfand, M. S.; Raugei, S. *J. Chem. Theory Comput.* **2009**, *5*, 2158–2172.
(25) Rieff, B.; Mathias, G.; Bauer, S.; Tavan, P. *Photochem. Photobiol.* **2011**, *87*, 511–523.

(26) Wheeler, R. A.; Dong, H.; Boesch, S. E. *ChemPhysChem* **2003**, *3*, 382–384.
(27) Wheeler, R. A.; Dong, H. *ChemPhysChem* **2003**, *4*, 1227–1230.
(28) Schmitz, M.; Tavan, P. *J. Chem. Phys.* **2004**, *121*, 12233–12246.
(29) Carbonniere, P.; Dargelos, A.; Ciofini, I.; Adamo, C.; Pouchan, C. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4375–4384.
(30) Calvo, F.; Parneix, P.; Van-Oanh, N. T. *J. Chem. Phys.* **2010**, *133*, 074303.
(31) Martinez, M.; Gaigeot, M.-P.; Borgis, D.; Vuilleumier, R. *J. Chem. Phys.* **2006**, *125*, 144106/14.
(32) Eckart, C. *Phys. Rev.* **1935**, *47*, 552–558.
(33) Babitzki, G.; Mathias, G.; Tavan, P. *J. Phys. Chem. B* **2009**, *113*, 10496–10508.
(34) Baer, M.; Marx, D.; Mathias, G. *ChemPhysChem* **2011**, DOI: 10.1002/cphc.201000955.
(35) Roos, B. W. *Analytic Functions and Distributions in Physics and Engineering*; John Wiley & Sons: New York, 1969.
(36) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in C*; Cambridge University Press: Cambridge, U.K., 1988, pp 553−558.
(37) Reed, M.; Simon, B. *Methods of Modern Mathematical Physics: II Fourier Analysis, Self-Adjointness*; Academic Press: New York, 1975, p 10.
(38) Wilson, E.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw Hill: New York, 1955, pp 19−27, 54−63.
(39) Bunse-Gerstner, A.; Byers, R.; Mehrmann, V. *SIAM. J. Matrix Anal. Appl.* **1993**, *14*, 927–949.
(40) Cardoso, J.-F.; Souloumiac, A. *SIAM J. Mat. Anal. Appl.* **1996**, *17*, 161–164.
(41) Golub, G.; van Loan, C. *Matrix Computations*, 3rd ed.; Johns Hopkins University Press: Baltimore, MD, 1996, pp 426−438.
(42) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. *J. Am. Chem. Soc.* **1979**, *101*, 2550–2560.
(43) Fogarasi, G.; Zhou, X.; Taylor, P. W.; Pulay, P. *J. Am. Chem. Soc.* **1992**, *114*, 8191–8201.
(44) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*; Cambridge University Press: Cambridge, U.K., 2009, pp 11−75.
(45) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
(46) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
(47) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.
(48) Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys. Rev. B* **1998**, *58*, 3641–3662.
(49) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.
(50) Lippert, G.; Hutter, J.; Parrinello, M. *Mol. Phys.* **1997**, *92*, 477–487.
(51) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103–128.
(52) Martyna, G.; Tuckerman, M. *J. Chem. Phys.* **1999**, *110*, 2810–2821.
(53) Martyna, G. J.; Klein, M. L.; Tuckerman, M. E. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
(54) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
(55) Stein, S. E. *NIST/EPA Gas-Phase Infrared Database*; National Institute of Standards and Technology: Gaithersburg, MD, 2004.
(56) Schrader, B. *Raman/Infrared Atlas of Organic Compounds*; VCH: Weinheim, Germany, 1989, p C2−01.