ADVANCED THEORY AND SIMULATIONS
www.advtheorysimul.com

Benchmarking the Computational Costs and Quality of Vibrational Spectra from Ab Initio Simulations

Shima Taherivardanjani, Roman Elfgen, Werner Reckien, Estela Suarez, Eva Perlt, and Barbara Kirchner**

The scaling behavior of ab initio molecular dynamics simulations for the different size bulk systems of liquid methanol is presented and thereby the characteristics of every system performing on either a local compute cluster or a supercomputer are analyzed. Additionally, the influence of different parameters on the quality of the infrared and Raman spectra is investigated using different simulation frameworks, including time step, convergence criteria, density functional approximation, and basis set. Both the maximally localized Wannier functions and the radical Voronoi tessellation approaches are employed to evaluate vibrational spectra from the trajectories. The results of infrared and Raman spectra are classified in two frequency regions, 500 to 1600 cm^{-1} and 2500 to 4000 cm^{-1}, in order to compare and discuss the experimental spectra and the results derived from ab initio molecular dynamics simulations comprehensively. The outcome of this study guides future experimental and theoretical researchers in order to acquire a profound perception into vibrational spectra, which evolves the way of elucidating molecular structure.

1. Introduction

Vibrational spectra represent an important and experimentally feasible way to determine material compositions and structural information of molecules by exploring characteristic frequencies of atomic vibrations.[1–3] The theoretical prediction of vibrational spectra does not only support the interpretation of experimental measurements, but based on atomistic simulations, they also provide detailed molecular level insights and help to develop a fundamental understanding of molecular vibrations and interplay.[1,4,5] The relevance of reliable vibrational data for the calculation of thermodynamic properties from theory has been demonstrated for different systems including methanol–water mixtures.[6–8]

Especially for the treatment of bulk systems and flexible molecules, ab initio molecular dynamics (AIMD) simulations represent a powerful technique that has already been applied successfully with respect to spectroscopic properties.[9–13] AIMD simulations solve the electronic structure problem on-the-fly at each time step of the simulation, mostly using state-of-the-art density functional theory (DFT) methods. For the prediction of vibrational spectra, however, additional information is required: infrared (IR) spectra rely on the temporal changes of the molecular dipole moments,[14] while for Raman spectra, the molecular polarizabilities have to be determined.[15–17]

Among a number of techniques to calculate these molecular properties, many examples in literature are based on the well-established scheme using maximally localized Wannier functions.[10,18] In this approach, a set of localized Wannier function centers[10] is determined, representing the orbital locations in a system that are assigned to the individual molecules via a minimum distance criterion.[19] The Wannier function centers are subsequently used to calculate dipole moments[18,20,21] and polarizabilities.[22] The great advantage of this method is the low additional amount of data produced, and thus, memory space required, which guarantees a much faster processing of the data during the simulation and afterward. However doubtless, this method contains several drawbacks. From a technical point of view, the convergence of the iterative localization procedure is not guaranteed. Furthermore, the result of the localization is not necessarily smooth along a molecular dynamics trajectory as could be demonstrated for Raman spectra of aromatic compounds.[23] Both effects—not converged localizations and discontinuous data points—can lead to unphysical results for the obtained spectra.

S. Taherivardanjani, R. Elfgen, W. Reckien, B. Kirchner
Mulliken Center for Theoretical Chemistry
Institute for Physical and Theoretical Chemistry
Beringstr. 4, Bonn D-53115, Germany
E-mail: kirchner@thch.uni-bonn.de

E. Suarez
Institute for Advanced Simulation
Jülich Supercomputing Centre, Forschungszentrum Jülich GmbH
Wilhelm-Johnen-Straße, Jülich D-52425, Germany

E. Perlt
Otto Schott Institute of Materials Research
Faculty of Physics and Astronomy
Friedrich-Schiller-Universität Jena
Löbdergraben 32, Jena D-07743, Germany
E-mail: eva.von.domaros@uni-jena.de

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adts.202100293

DOI: 10.1002/adts.202100293

Using the Wannier-based approach, the trans and cis isomers of *N*-methylacetamide in aqueous solution were studied from AIMD simulations.[9] Next to the Wannier approach, a second method based on the molecular current correlation function computed from the Born charges and atomic velocities was applied.[9] Clusters consisting of a chloride anion Cl^- and one or two methanol (hereafter abbreviated MeOH) molecules have been explored combining infrared predissociation experiments and DFT-based molecular dynamics simulations at 100 K.[24]

Recently, Gaigeot and Rijs[1,25] provided a review of neutral peptide gas phase IR spectra. The authors compared theoretical approaches such as dynamical (DFT) techniques and their recently developed graph theory with experimental data in order to underline how the synergy between far-IR spectroscopy and theory can provide an unprecedented picture of the structure of neutral biomolecules in the gas phase. The former author, Gaigeot, has in a recent review considered the molecular dynamics (MD) simulations for anharmonic gas-phase IR and Raman spectra, by which the spectra of large (bio-) molecular systems over long time-scale trajectories can be accurately and efficiently achieved.[26]

Luber proposed an alternative approach based on a subsystem-density functional theory based embedding scheme.[27] An overview over the historical development of spectra calculations is provided by Shaoqing Wang.[28] Wang introduces the Eckart frame algorithm and the multiorder derivative algorithm for vibrational frequency calculations directly based on the raw data of atomic trajectory from AIMD. Mandal et al. have recently proposed a novel strategy, named s–MTACE, which provides the opportunity of performing around tenfold faster hybrid DFT-based MD simulations of large condensed matter systems by analyzing trajectories.[29]

The application of the Voronoi tessellation[23] for dissecting the electron density is another alternative to the use of the Wannier center approach. Starting from the atomic positions, the Euclidean space is divided into a set of Voronoi cells, assigning each point in space to exactly one atomic cell. To account for different atom types, their respective van der Waals radii can be taken into account when partitioning the electron density to define the Voronoi cells. This scheme is referred to as the radical Voronoi tessellation.[30] A major drawback of the Voronoi approach lies in the necessity to print out the electron density as volumetric data on a grid (usually in cube files), which can produce tremendously large files depending on the system size and length of the simulation.

Even if the process of printing out the electron density is performed only every eighth time step, which was shown to be reasonable for the calculation of IR and Raman spectra,[23] the resulting cube-files easily exceed a size of several hundred gigabytes (GBs). Recently, Brehm and Thomas introduced a novel compression algorithm (bqb), which significantly enhances the handling of volumetric data.[31] A comparison of the Wannier versus Voronoi approach was recently published by Apostolidou at the example of the OH radical in water.[32] The first simulation of resonance Raman spectra based on the radical Voronoi tessellation in combination with real-time time-dependent density functional theory was introduced by Brehm and Thomas.[33] The authors computed the spectrum of uracil in aqueous solution in good agreement with experiment.

It can be seen that depending on the system to be investigated the choice of the method for the dipole and polarizability calculations, and hence, the calculation of vibrational spectra can become challenging with respect to computational costs. In this regard, a great variety of algorithms termed linear-scaling methods have been proposed to enhance the accomplishment of large scale electronic structure calculations. As an instance, the properties of ethanol using the linear scaling method together with finite-element basis functions were inspected by Tsuchida.[34] Tsuchida substantiated the claim that linear-scaling methods lessen the computational cost of large-scale AIMD simulations with maintaining accuracy. Wilhelm et al. initiated another algorithm to calculate the correlation energy in the random phase approximation in a Gaussian basis demanding operations and memory.[35] Wilhelm and co-workers also introduced an optimized low-scaling GW (where G is the Green's function and W stands for the screened Coulomb interaction[36]) algorithm in a Gaussian basis with colossally parallel execution on supercomputers.[37] This low-scaling GW algorithm is an accurate method for computing electron addition and removal energies of molecules and solids. The application of large-scale GW electronic-structure calculations for complex materials on large timescales was investigated by Del Ben and co-workers.[38]

Hutter et al. showed that the Car–Parrinello molecular dynamics code paired with multiscale modeling provides an optimal solution for systems up to thousand atoms on future generations of supercomputers.[39] In line with the study by Hutter and co-workers of the machine-dependent scaling behavior for the AIMD simulation of liquid water,[40] we herein focus on studying the scaling behavior of AIMD simulations including the electron density treatment via the Wannier versus the Voronoi approaches. For this purpose, we consider methanol bulk systems of different sizes. Being the simplest alcohol, methanol is a close analog to the water molecule.[41] However, the presence of a methyl group in the methanol structure leads to a different hydrogen bonding network in this molecule, forming only two hydrogen bonds, as opposed to water with its complicated hydrogen-bonding networks.[42,43] Liquid methanol has been considered not only as an interesting subject for spectroscopical and theoretical studies,[44–49] but also as the all-purpose solvent in industrial process and laboratory chemistry.[43]

We also test the influence of certain relevant simulation parameters, such as the system size, the size of the time step, the criterion for the SCF convergence, the functional, and the basis set on the quality of the resulting IR and Raman spectra. Based on the results, it will be discussed whether a modular approach, in which the single processes are parallelized, would be appropriate and worthwhile for the calculation of vibrational spectra from AIMD simulations. Finally, we aim to provide guidance in finding the best computational setup for the efficient calculation of high-quality vibrational spectra.

2. Theoretical Background

2.1. Vibrational Spectra from AIMD Simulations

For the calculation of different spectroscopic properties in liquid media and for flexible molecules, the seminal work by Green and Kubo constitutes the starting point.[50,51] They showed that

phenomenological coefficients describing many transport processes and time (t) dependent phenomena can be written as integrals over auto-correlation functions $C(t)$:

$$C(t) = \langle A(0)A(t)\rangle_0 \tag{1}$$

with A being any dynamical function and the brackets indicating the average over the ensemble of the system at reference time 0.

In order to further decompose the auto-correlation functions into their particular constituent frequencies, so to finally receive a spectrum resolving the molecular vibrational frequencies, a Fourier transformation has to be applied. Therefore, it is crucial to efficiently calculate auto-correlation functions and their corresponding Fourier transforms, which can be achieved using the Wiener–Khintchine theorem:[52,53]

$$\langle C(\tau)C(t+\tau)\rangle_\tau = \frac{1}{2\pi}\int\left|\int f(t)e^{-i\omega t}dt\right|^2 e^{i\omega t}d\omega \tag{2}$$

With τ being the time lag, the Fourier transform of $C(t)$ is generated, its square is calculated within the frequency range, and the inverse transformation is applied. By using the fast Fourier transformation formalism, the Wiener–Khintchine scheme scales with $\mathcal{O}(n\log(n))$, while a direct application of the definition would lead to an algorithm scaling with $\mathcal{O}(n^2)$.

The calculation of IR and Raman spectra is usually based on appropriate quantum mechanical expressions for the absorption and the scattering of light, respectively, which can be derived from time-dependent perturbation theory.[54–60] Here, the time-correlation formalism[61] elucidates that the IR absorption coefficient and the Raman scattering cross section are proportional to the auto-correlation of the molecular dipole moment and polarizability, respectively. Nevertheless, the resulting equations depend upon ensemble averages of the corresponding quantum mechanical operators and cannot be applied to the AIMD simulations directly, in which the nuclei move according to the classical equations of motion.[10]

The intensity of the resulting spectrum corresponds to the amplitude of the oscillation in the correlation function and in this way notable changes in the dipole moment or the polarizability lead to peaks in the spectrum and vice versa.

The corresponding wavenumber-dependent IR spectra $A(\tilde{\nu})$ are given by

$$A(\tilde{\nu}) = \frac{N_\mathrm{A}}{12\varepsilon_0 c k_\mathrm{B} T}\int\langle\dot{\mu}(\tau)\cdot\dot{\mu}(t+\tau)\rangle_\tau e^{-2\pi i c\tilde{\nu}t}dt \tag{3}$$

Here, N_A is the Avogadro constant, ε_0 the electric constant, c the speed of light in vacuum, k_B the Boltzmann constant, T the temperature, $\tilde{\nu}$ the wavenumber, and μ the dipole moment.

Similar to this, the wavenumber-dependent differential Raman scattering cross section $I(\tilde{\nu})$ is calculated according to

$$I(\tilde{\nu}) = \frac{h}{8\varepsilon_0 c k_\mathrm{B} T}\frac{(\tilde{\nu}_\mathrm{in}-\tilde{\nu})^4}{\tilde{\nu}\left(1-\exp\left(-\frac{hc\tilde{\nu}}{k_\mathrm{B}T}\right)\right)}\int\left(\langle\dot{\alpha}_{xx}(\tau)\cdot\dot{\alpha}_{xx}(t+\tau)\rangle_\tau + \langle\dot{\alpha}_{xy}(\tau)\cdot\dot{\alpha}_{xy}(t+\tau)\rangle_\tau\right)e^{-2\pi i c\tilde{\nu}t}dt \tag{4}$$

where in addition, the Planck constant h is involved and the dipole moment auto-correlation is replaced by the cross-correlation of the polarizabilities α. The calculation of vibrational spectra is carried out using TRAVIS.[62,63]

2.2. Dipole Moments and Polarizabilities

2.2.1. Maximally Localized Wannier Functions

Maximally localized Wannier functions are obtained by a unitary transformation of the Kohn–Sham orbitals, which are inherently available in each time step of the trajectory, and thus a set of n localized Wannier orbitals is calculated.[64]

The position expectation values $\mathbf{r}_i$ of the orbitals are considered as the locations of the electron pairs, which can, in the bulk phase, be assigned to the individual molecules via a minimum distance criterion. The molecular dipole moment $\boldsymbol{\mu}_\mathrm{W}$ is now calculated by the sum over the point charges of the nuclei and the Wannier function centers representing n electron pairs:

$$\boldsymbol{\mu}_\mathrm{W} = -2e\sum_{i=1}^{n}\mathbf{r}_i + e\sum_{j=1}^{N}Z_j\mathbf{R}_j \tag{5}$$

Here, e is the elementary charge, the first sum is over the positions of all n Wannier function centers $\mathbf{r}_i$ assigned to the molecule, and the second sum is over all N nuclei of the molecule with the positions $\mathbf{R}_j$ and the charges Z_j.

A possible calculation scheme for polarizabilities emanates from the linear approximation of the dipole moments induced by an external electric field $\mathbf{E}$:

$$\boldsymbol{\mu}_\mathrm{ind} = \boldsymbol{\alpha}\mathbf{E} \tag{6}$$

where $\boldsymbol{\alpha}$ is the second-order polarizability tensor.

This suggests that the electronic structure calculation has to be repeated three times applying an electric field of the same strength in each direction in space, x, y, and z, in order to record the dipole moment changes with respect to the calculation without the electric field. Since the polarizability is the dipole moment change divided by the electric field strength, the polarizabilities are obtained by finite differences of the dipole moments.

As mentioned above, the significant advantage of this method is the low additional amount of data produced, and consequently, less memory space is required, since the Wannier function centers are characterized solely by Cartesian coordinates of each electron pair. This guarantees a much faster data transfer during the simulation, which is in almost every case parallelized over a number of processors and makes the subsequent handling and storage of the data much easier. One disadvantage of the maximally localized Wannier functions approach is the requirement of additional computational effort for the localization itself. Especially for big, complex systems composed of several hundreds of atoms, the convergence of the Wannier localization can be slow and, consequently, can significantly extend the simulation time, making the calculation of vibrational spectra more expensive.

Although it is also possible to use the total dipole moment and polarizability of the simulation box to calculate IR and Raman

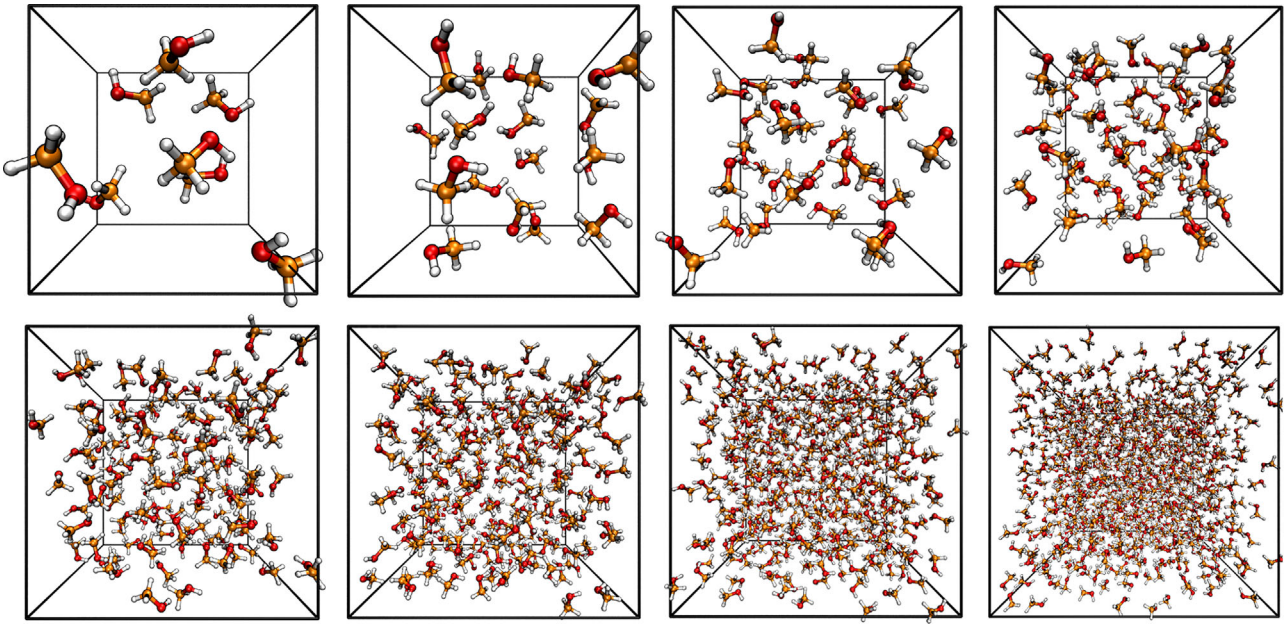

Figure 1. Simulation boxes after pre-equilibration via classical MD for systems composed of 8, 16, 32, 64, 128, 256, 512, and 1024 MeOH molecules, respectively. Please note, for the spectroscopic study in Section 3.2.2, we also simulated a 4-molecule system.

spectra, the assignment of individual molecular quantities improves the sampling and allows for the separation of vibrational contributions of systems that originate from different components like for example in mixtures or solute–solvent systems.

2.2.2. Radical Voronoi Tessellation

The radical Voronoi tessellation[65] allows for the partitioning of the electron density into atomic or molecular contributions, only based on distance criteria, to further calculate dipole moments of individual molecules and, therefrom, vibrational spectra. In the case of an atomistic simulation with N atoms, the atom positions are considered as sites $\mathbf{s}_i$ to uniquely divide the Euclidean space into a set of Voronoi cells, assigning each point in space to exactly one cell.[23,66] Each cell, therefore, consists of all points that are closer to a particular site than to any other one. Additionally, a radius $\mathbf{r}_i$ is assigned to each site $\mathbf{s}_i$ and the distance to the sites is replaced by the power distance to spheres around the sites with the corresponding radii. For the correct treatment of intra- and intermolecular distances and interactions within liquid systems, the choice of the van der Waals radii has been shown to be the best option.[67] The resulting Voronoi cells C_i are, therefore, defined by

$$C_i = \{\mathbf{x} \in \mathbb{R}^3 \mid (\mathbf{x}-\mathbf{s}_i)^2 - \mathbf{r}_i^2 \leq (\mathbf{x}-\mathbf{s}_j)^2 - \mathbf{r}_j^2 \; \forall \, j \neq i\}; \; i,j = 1, \ldots, N \tag{7}$$

Voronoi cells of all N_k atoms belonging to the same molecule k are united in order to obtain molecular cells M_k. The recognition of the molecules is available in TRAVIS.[62,63] Molecular dipole moments can, hence, be obtained by integrating the electron density $\rho(\mathbf{r})$, which is accessible for each time step in an AIMD simulation, as a function of the spacial coordinate $\mathbf{r}$ over the afore defined molecular cells:

$$\boldsymbol{\mu}_V = \int_{M_k} \mathbf{r} \cdot \rho(\mathbf{r}) \, d\mathbf{r}; \; k = 1, \ldots, N_k \tag{8}$$

Analogous to the maximally localized Wannier functions, the polarizabilities are obtained by finite differences of the dipole moments. However, for the electronic structure calculations without and with external electric field in each Cartesian direction, in total four cube-files containing the information of the electronic density have to be produced.

3. Computational Methodology

3.1. Test Systems

For the systematic investigation of the scaling behavior and the simulation parameter-dependent accuracy of AIMD simulations including the treatment of the electron density for the subsequent calculation of vibrational spectra, model systems composed of either 8, 16, 32, 64, 128, 256, 512, or 1024 MeOH molecules were chosen (see **Figure 1**). All systems were considered to study the scaling behavior on different hardware (see Part I below and Section 5) while only systems with up to 64 molecules were considered to study the dependence of spectra on different computational parameters (see Part II below and Section 6).

To generate the simulation boxes for the pre-equilibration of the systems using classical MD simulations, first, one single MeOH molecule was built using the visualization package MOLDEN (version 5.4).[68] In order to derive the initial starting geometries of the simulation boxes, the program PACKMOL (version 16.228) was used.[69,70] Considering the experimental room temperature density of MeOH,[71] $\rho^{exp} = 0.79 \text{g cm}^{-3}$, leads to

Table 1. Changes in densities and box length before (ρ^{exp}, a^{exp}) and after (ρ^{theor}, a^{theor}) the classical MD pre-equilibration runs.

# MeOH	ρ^{exp} [g cm^{-3}]	ρ^{theor} [g cm^{-3}]	change [%]	a^{exp} [pm]	a^{theor} [pm]	change [%]
8	0.79	0.83833	+6.12	8.12	7.98	−1.72
16	0.79	0.75424	−4.53	10.25	10.41	+1.56
32	0.79	0.75764	−4.10	12.92	13.10	+1.39
64	0.79	0.77592	−1.78	16.27	16.37	+0.61
128	0.79	0.75712	−4.16	20.50	20.80	+1.46
256	0.79	0.76359	−3.35	25.83	26.13	+1.16
512	0.79	0.76300	−3.42	32.55	32.93	+1.17
1024	0.79	0.76325	−3.39	41.01	41.48	+1.15

the resulting box length a^{exp} in pm, which are summarized in **Table 1**. Table 1 provides an overview of changes in the systems' densities and box length a before and after the pre-equilibration performed by classical MD simulations. As one can notice, the maximum deviation from experimental density is about 6% for eight molecules.

In order to pre-equilibrate each system, classical MD simulations were carried out using the molecular dynamics simulator LAMMPS (version from August 17, 2017).[72] The simulation details for these pre-equilibration runs are given in the Supporting Information.

3.2. AIMD Simulations and Electron Density Treatment

The AIMD simulations and electron density calculations were carried out using DFT methods, since these are computationally efficient electronic structure methods for dynamical spectroscopy of large (bio-)molecular systems with over 100 ps timescale trajectories,[26] with the CP2K[73,74] program package applying the QUICKSTEP module.[40] Here, the hybrid Gaussian and plane waves (GPW) approach was used to calculate the energies and forces on the atoms. As a standard setup, the molecularly optimized short range double-ζ basis set (MOLOPT-DZVP-SR-GTH)[75] was applied to all atoms together with the BLYP functional and the corresponding BLYP Goedecker–Teter–Hutter pseudopotentials for core electrons.[76–78] Throughout this work, molecularly optimized basis sets have been applied (MOLOPT-XZV(P)-SR-GTH), which will henceforth be referred to as (SR)-XZV(P). A 350 Ry density CUTOFF criterion (again standard setup) with the finest grid level was employed, together with multigrids number 4 (NGRID 4 and REL_CUTOFF 50) using the smoothing for the electron density (NN10_SMOOTH) and its derivative (NN10).[40] The dispersion corrections in all simulations were accounted for by using the DFT-D3 type of a pair potential van der Waals density functional.[79,80] A value of 10^{-6} was used as target accuracy for the SCF convergence (EPS_SCF 1.0E-6). The DIIS minimizer[73] was used to reach a faster orbital transformation via direct inversion in the iterative subspace. The maximum number of SCF iterations to be performed for one iteration was set to 45 while for the outer SCF loops it was 30. Apart from some special cases, which will be defined and amplified in the following, a time step of 0.5 fs and a simulation temperature of 300 K were chosen.

3.2.1. Part I: Scaling Behavior

For the systematic investigation of the scaling behavior, each of the pre-equilibrated systems was subject to an AIMD simulation using either 4, 8, 16, 32, 64, or 128 cores. To compare different treatments of the electron density the following three protocols were applied in each case: production of the trajectory only (T), production of the trajectory and a Wannier localization (TW), and production of the trajectory together with the calculation and printing of the electron density (TD). In each case, the trajectories were produced over 500 time steps at 0.5 fs and the Wannier localizations and electron density calculations were performed in every time step.

3.2.2. Part II: Simulation Parameter Dependency of Vibrational Spectra

To analyze the influence of certain relevant simulation parameters on the quality of the resulting IR and Raman spectra, longer AIMD equilibration and production runs had to be performed.

In a first equilibration, four systems containing 8, 16, 32, and 64 MeOH molecules were simulated over 25 ps. The temperature was increased to 340 K (slightly above the experimental boiling point of MeOH ≈ 338 K)[81] and massive thermostats were applied to all atoms. Meanwhile, the density CUTOFF criterion was decreased to 280 Ry and the target accuracy for the SCF convergence released to a value of 10^{-5} in order to reach a fast convergence in a first equilibration. A second equilibration run over another 25 ps was performed to adjust the system to the desired simulation conditions. Therefore, massive thermostats were kept, but the temperature was decreased to 300 K, the density CUTOFF criterion increased to 350 Ry, and the SCF convergence criterion tightened to a value of 10^{-6}.

To investigate the influence of different parameters, the simulation setup with 16 MeOH molecules, a timestep of 0.5 fs, an SCF convergence threshold of 10^{-6}, the BLYP functional, and a (SR)-DZVP basis set have been chosen as reference setup. From here on, 13 copies from the last snapshot of the 16-MeOH system were generated and separate simulations were run—each differing in one simulation parameter as compared to the reference run. To investigate the influence of the system size, the pre-equilibrated 8-, 32-, and 64-MeOH systems were additionally considered. For all systems (17 in total), first, another short

Table 2. Different simulation parameters for testing their influence on the quality of the resulting vibrational spectra.

# MeOH	Time step [fs]	SCF threshold	Functional	Basis set
16	**0.5**	$\mathbf{10^{-6}}$	**BLYP**	**(SR)-DZVP**
4	0.5	10^{-6}	BLYP	(SR)-DZVP
8	0.5	10^{-6}	BLYP	(SR)-DZVP
32	0.5	10^{-6}	BLYP	(SR)-DZVP
64	0.5	10^{-6}	BLYP	(SR)-DZVP
16	**0.8**	10^{-6}	BLYP	(SR)-DZVP
16	**0.3**	10^{-6}	BLYP	(SR)-DZVP
16	0.5	$\mathbf{10^{-4}}$	BLYP	(SR)-DZVP
16	0.5	$\mathbf{10^{-5}}$	BLYP	(SR)-DZVP
16	0.5	$\mathbf{10^{-7}}$	BLYP	(SR)-DZVP
16	0.5	10^{-6}	**PBE**	(SR)-DZVP
16	0.5	10^{-6}	**revPBE**	(SR)-DZVP
16	0.5	10^{-6}	**B97-3c**[a)]	(SR)-DZVP
16	0.5	10^{-6}	**B97-3c**[a)]	**(SR)-TZVP**
16	0.5	10^{-6}	**B3LYP**	(SR)-DZVP
16	0.5	10^{-6}	**PBE0**	(SR)-DZVP
16	0.5	10^{-6}	BLYP	**SZV**
16	0.5	10^{-6}	BLYP	**(SR)-TZVP**

The bold face parameters highlight the changes compared to the standard setup given in the first row. [a)] B97-3c parametrization of the B97D functional.

equilibration was run for 10 ps and subsequently, a production run over 50 ps was carried out. Furthermore, a system with only four MeOH molecules was considered. Since the calculation of the density of such a limited system and hence, a classical pre-equilibration are hardly possible, this step was foregone and instead, four molecules out of the 8-MeOH system were selected and simulated within the AIMD computational setup. During the production run, the Wannier localization was performed at every fifth time step using the method CRAZY[73] and, in case of no convergence, the JACOBI_FALLBACK method[73] with a maximum number of 2000 iterations (MAX_ITER 2000) was applied. The electron density was printed out at every eighth time step, which saves a large amount of computational time and disk space and has proven to be sufficient for the calculation of dipole moments, polarizabilities, and therefrom IR and Raman spectra, respectively (see ref. [23]). The modified simulation parameters for the third equilibration and the production runs are summarized in **Table 2**.

4. Hardware

4.1. UniBonn Local Compute Cluster

The simulations of the systems up to 256 MeOH molecules have been performed with the local cluster of the Kirchner group at UniBonn. The local cluster consists of 36 servers and one head node which is used for login and management of the jobs. Each local node is equipped with 128 GB of DDR3 memory and four AMD-Opteron 6376 CPUs, clocked at 2.3 GHz. The local nodes are interconnected by a QSFP 40 GB s^{-1} QDR InfiniBand network. The configuration of the local cluster is given in **Table 3**.

Table 3. Hardware configuration of the two platforms used: UniBonn Local Cluster and JURECA Cluster.

Characteristics	UniBonn local cluster	JURECA cluster
Processor	AMD Opteron 6376 (Abu Dhabi)	Intel Xeon E5-2680v3 (Haswell)
CPUs per node	4	2
Cores (threads)	4×16 (128)	2×12 (48)
Frequency [GHz]	2.3	2.5
Memory per node [GB]	128 RAM	128 RAM
Node count	36	1872
Network technology	QSFP 40GB/S QDR InfiniBand	EDR InfiniBand (full fat-tree)

4.2. JURECA Cluster

The simulations for the 512- and 1024-MeOH systems have been performed with the JURECA Cluster,[82] which has been in operation at the Jülich Supercomputing Centre (JSC) from Forschungszentrum Jülich GmbH (FZJ) between 2015 and 2020. JURECA consists of 1872 compute nodes, accompanied by additional 24 nodes used for login, visualization, etc. Each node is equipped with two 12-core Intel Xeon E5-2680 v3 processors, clocked at 2.5 GHz. The compute nodes used for this study present 128 GB DDR4 memory and a maximum memory bandwidth of 136 GB s^{-1}. The nodes are interconnected via a full fat-tree Mellanox InfiniBand network in the EDR generation, which delivers 100 GB s^{-1} link bandwidth and 1 µs MPI latency. The JURECA hardware configuration is summarized briefly in Table 3.

It is worth noting that the JURECA Cluster is interconnected with another, accelerator-based system: the JURECA Booster. Both together realize the so-called Modular Supercomputing architecture,[83,84] a heterogeneous system design that aims at addressing the requirements of diverse applications. Although the Booster has not been used for this work, there are future plans to test the architecture by running different parts of the simulation. This motivates the preliminary analysis presented in **Figure 2**. This figure shows the computational resources that are necessary for the production of the trajectory (black lines)—including the energy calculation, gradient calculation, and integration—and those that are necessary if Wannier localization (TW, left panels) or printing out of the density (TD, right panels) are carried out. Although the simulation itself requires the major part of the resources, both wavefunction postprocessing methods increase the walltime significantly. However, the interface between both operations is only the wavefunction, so that in principle, the trajectory production could proceed with the next MD step while the current wavefunction is analyzed on separate CPUs with access to shared memory. As shown in Figure 2, such a modular approach is currently not feasible since the wavefunction is not stored and the DFT calculation for the energy evaluation is performed in each step. Consequently, if only the wavefunction analysis is conducted as indicated by W and D in the figure, the same amount of resources as for TW or TD, respectively, is required.

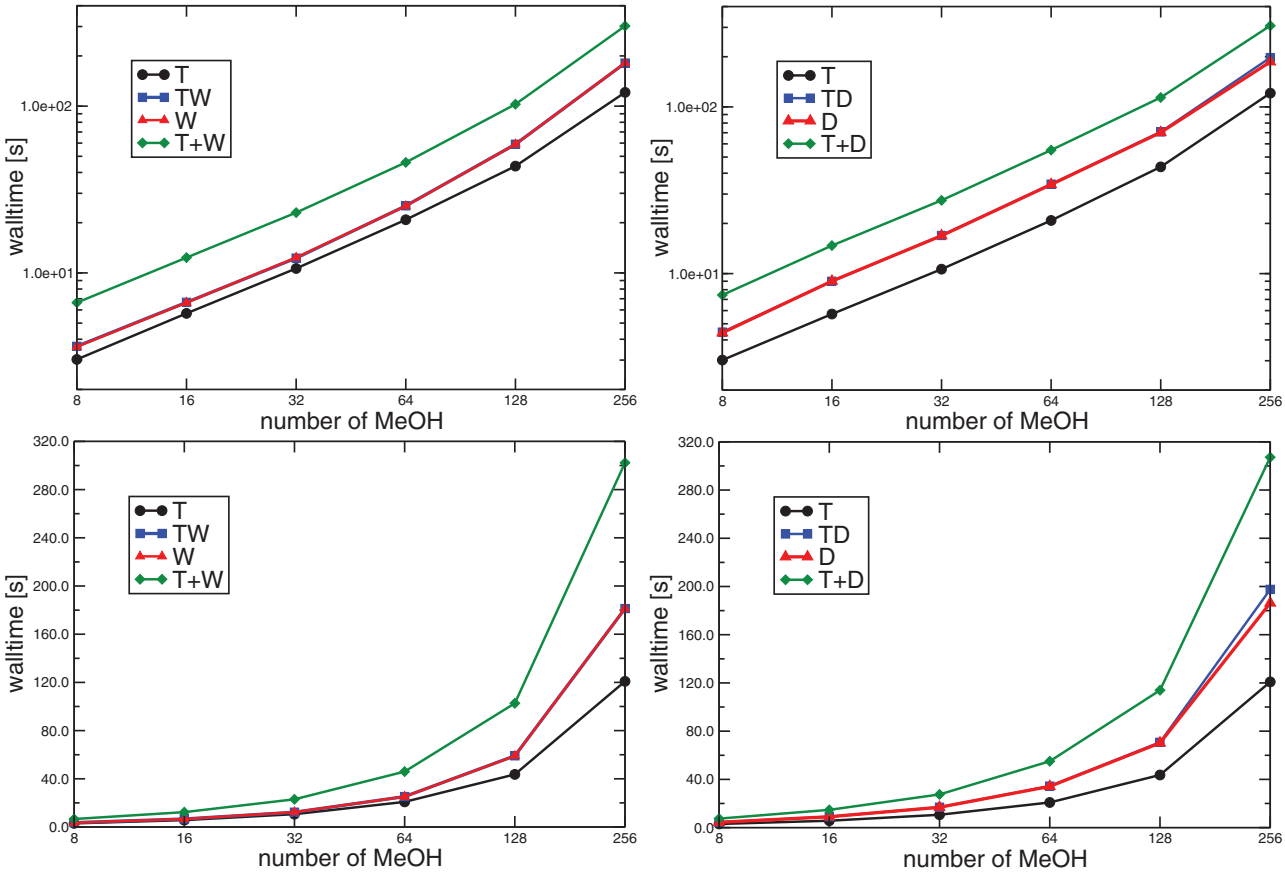

Figure 2. Decomposition of the total walltime for different computational modules (modularity): trajectory production and electron density treatment via Wannier (W) and printing out the electron density (D). Left panels: Modularity for W, right panels: modularity for D; upper panels: logarithmic scale, lower panels: linear scale.

5. Scaling Behavior

VandeVondele et al. illustrated the stability of QUICKSTEP on parallel computers.[40] This was achieved by performing AIMD simulations for liquid water benchmark systems of different sizes (32 to 1024 H_2O molecules) under ambient conditions using the same set of input parameters. In addition, the scaling tests were performed using two different methods for the minimization of the total ground state energy of a system by an iterative SCF procedure: the traditional diagonalization (TDia) scheme and the efficient orbital transformation (OT) method. For an extended description of both methods see refs. [40, 85]. When plotting the average time needed per MD step as a function of the number of CPUs on a double logarithmic scale, it can be manifested that using the TDia scheme, an efficient simulation of small systems (32 to 64 H_2O molecules) is possible on a relatively small number of CPUs. On the other hand, medium-sized systems (128 to 256 H_2O molecules) can be simulated with fair efficiency on 32-64 CPUs, but large systems cannot be applied on the given memory constrains. The OT method yields a significantly improved scaling behavior for all systems, up to six times faster than the TDia scheme, and reduced memory requirements for the simulation of up to 1024 water molecules. Thus, the efficient OT method is the method of choice for the scaling tests presented in this work.

Based on the study described above, we extend the scaling tests by adding the Wannier localization and the process of printing out the electron density of the system for the subsequent Voronoi tessellation in order to derive dipole moments and polarizabilities. See Section 3.2.1 for computational details and definition of different subsets.

For a comparison of the scaling behavior between the different subsets T, TW, and TD, each for the systems containing 8, 16, 32, 64, 128, and 256 MeOH molecules, calculations were performed on either 4, 8, 16, 32, or 64 cores. One has access to the overall walltime from the starting point of the CP2K program until the very end. This, however, includes moving and copying processes between single calculation steps within one run. In consequence, to reach a better comparability of the subsets (e.g., to compare TW with T and W), in the following, we consider the averaged duration of only one time step. To do so, the mean duration of one time step is calculated for each of the given subsets by averaging over the last 400 time steps (from overall 500 time steps). The reason behind this is clearly recognizable in **Figure 3**. The presence of large oscillations in the first 100 steps (around up to 10 s more than the entire ones) indicates occasional slow conver-

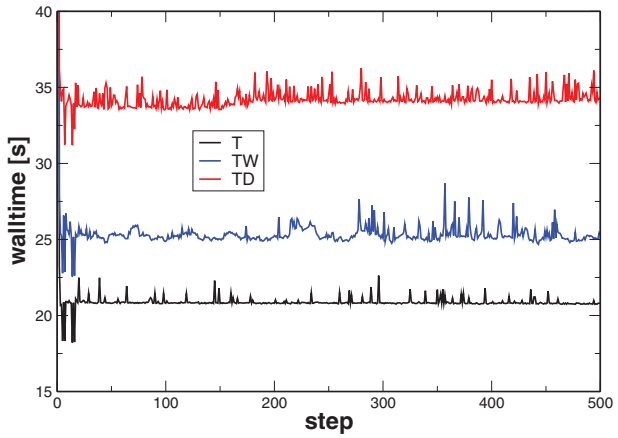

Figure 3. Step duration development: 64 MeOH, 64 cores.

Table 4. Average duration of one time step for each subset, performed on the local cluster in Bonn.

# MeOH	Subset	Walltime/MD step [s]				
		4 cores	8 cores	16 cores	32 cores	64 cores
	T	17.92	9.73	5.76	3.93	3.06
8	TW	18.35	10.08	6.04	4.31	3.69
	TD	18.79	10.68	6.70	4.94	4.51
	T	34.26	18.46	9.56	5.95	5.75
16	TW	35.18	19.08	10.11	6.55	6.65
	TD	36.37	20.56	11.69	8.19	9.03
	T	67.78	34.13	18.71	11.39	10.63
32	TW	70.74	36.08	20.12	12.82	12.35
	TD	71.89	38.35	22.95	15.75	16.86
	T	158.10	80.67	43.17	25.02	20.87
64	TW	169.39	88.03	48.09	29.42	25.33
	TD	167.77	89.77	52.98	34.24	34.27
	T	321.58	165.29	88.04	53.97	43.66
128	TW	380.94	198.69	109.26	70.95	58.92
	TD	347.15	184.94	107.12	77.47	70.26
	T	938.94	475.82	256.74	148.85	120.79
256	TW	1287.57	659.53	364.37	221.86	182.43
	TD	1118.81	562.52	336.54	242.42	200.37

gence in early time steps which can affect the analysis, making it reasonable to cut them off.

As one can see from **Table 4**, the trajectory production for the 64-MeOH system takes about 20.87 s in each step using 64 cores. Post-processing of the electron density in the TW and TD subsets extends the time for each MD step to 25.33 and 34.27 s, respectively. Despite the fluctuations in the graphs in Figure 3, the average slopes are zero, proving the average time step for all three subsets. The resulting mean durations of one time step for the different MeOH systems, submitted on the different numbers of cores of the local and JURECA clusters are presented in Tables 4 and **5** for the subsets T, TW, and TD.

Table 6 lists the characteristics of the employed benchmark MeOH systems containing 4 to 1024 MeOH molecules. The number of Gaussian type orbitals expands linearly from 184 to 47 104, and 53.8×10^5 plane waves are required for the auxiliary basis of the 512–MeOH system.

Summarizing the **Figure 4**, it shows how the scaling behavior of the TW subset on the local cores in comparison to the TD calculation is significantly smoother (except for 256-MeOH system) considering the deviations from the T subset. However, this deviation grows considerably when the number of cores increases from 16 to 64. The TW scheme seems to be more appropriate to simulate systems up to 64 MeOH on a relatively small number of cores, rather than the TD scheme.

Another feature of this figure depicts that for a small number of cores, in case of TW, the walltime for systems ranging from 8 to 32 MeOH is comparable with the individual T subset for the three smaller systems. Concerning the larger systems (512 and 1024 MeOH molecules) performed on JURECA, the TW scheme deviates from T for almost 100 s and 200 s, respectively, implying a longer computation time for these larger systems on the JURECA cluster. To recapitulate, the calculation of dipole moments, polarizabilities, and from that IR and Raman spectra seems reasonable and less expensive if the Wannier localization scheme is used along with a larger number of cores.

The information about the computational effort and memory space required for all investigated systems is shown in **Table 7**. It can be observed that the time per MD step and the size of cube files increase drastically in accordance with an increasing number of atoms, implying very high I/O demands, and in consequence, potential limitations on the scalability of the code on supercomputers. One solution to this is to compress the data. For this purpose the highly effective bqb algorithm was developed (for more details see ref. [31]), which allows to avoid the production of large cube files in CP2K entirely. For example, the required disk space for largest system diminishes from 684 to 21.4 GBs, a reduction by almost 97 %. Details about how to invoke the usage of the bqb algorithm in CP2K are given in the Supporting Information.

6. Parameter Dependence of the Calculated Spectra

In this section, the influence of different simulation parameters on the quality of IR and Raman MeOH bulk spectra is presented and discussed. Therefore, different system sizes (4, 8, 16, 32, and 64 MeOH molecules), simulation time steps (0.3, 0.5, and 0.8 fs), thresholds for the SCF convergence (1.0E-5, 1.0E-6, and 1.0E-7), GGA density functionals (BLYP,[86,87] PBE,[88] revPBE,[89] and B97-3c[90]), and hybrid density functionals (PBE0[91] and B3LYP[92]), which reveal relatively fast and reliable simulation results for covalently bound systems,[93] and basis set sizes[94] (SZV, (SR)-DZVP, and (SR)-TZVP) are compared, as shown in Table 2. Please note that in this study, the B97-3c parametrization of the B97 functional as implemented in CP2K has been used with a (SR)-DZVP and a (SR)-TZVP basis set instead of the proposed modified mTZVP basis.[90,95]

In addition, the IR and Raman spectra are calculated based on dipole moments and polarizabilities, respectively, once emanating from the Wannier localization scheme and once from radical

Table 5. Average duration of one time step for each subset, performed on the JURECA cluster at Jülich.

# MeOH	Subset	Walltime/MD step [s]						
		24 cores	48 cores	96 cores	192 cores	288 cores	384 cores	576 cores
	T	331.86	180.30	93.67	56.02	40.92	32.63	28.78
512	TW	557.76	294.15	163.72	113.81	95.81	94.10	–
	T	–	–	411.12	229.20	160.71	126.19	108.31
1024	TW	–	–	902.14	504.29	394.88	352.70	–

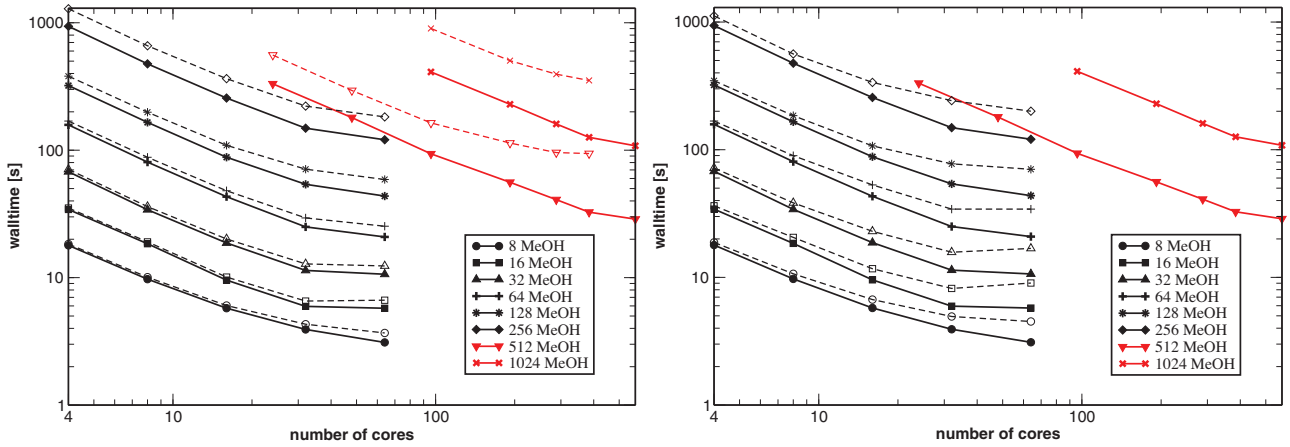

Figure 4. Solid lines: Production of trajectory (T); dashed lines, left panel: production of trajectory plus Wannier localization (TW); dashed lines, right panel: production of trajectory plus printing of electron density (TD). Black: Performed on local compute cluster (using 4, 8, 16, 32, or 64 cores); red: performed on the high-performance computer JURECA Cluster at Jülich (using 24, 28, 96, 192, 288, 384, or 576 cores).

Table 6. Characteristics for employed benchmark systems composed of liquid MeOH at ambient conditions (300 K, 100 kPa).

# MeOH	*a* [pm]	M	N	Grid points ($\times 10^5$)
4	798	184	28	1.5
8	798	368	56	2.2
16	1041	736	112	2.9
32	1310	1472	224	4.4
64	1637	2944	448	7.5
128	2080	5888	896	14.7
256	2613	11776	1792	26.8
512	3293	23552	3584	53.6
1024	4148	47104	7168	—

The edge length of the cubic simulation boxes *a*, the numbers of Gaussian type orbitals (*M*), the number of occupied orbitals (*N*), and the grid points for the plane waves are shown.

Table 7. Required computational resources using 64 (384) cores, 500 time steps. bqb is developed in ref. [31].

# MeOH	# Atoms	# Electrons	Walltime/MD step [s]	Filesize (Cube) [GB]	Filesize (bqb) [GB]
8	48	112	3.1	4.5	0.1
16	96	224	5.8	11	0.3
32	192	448	10.6	21	0.7
64	384	896	20.9	42	1.3
128	768	1792	43.7	85	2.7
256	1536	3584	120.8	173	5.4
512	3072	7168	421.3	339	10.6
1024	6144	14336	215.4	684	21.4

Voronoi tessellation of the electron density (c.f. Section 3.2.2 for computational details). A detailed discussion of the system size, the functional choice, and the basis set will follow—the time step and SCF convergence will be shown in the Supporting Information. Please note, absolute intensities are not discussed in this study. The experimental IR and Raman spectra in all cases are taken from ref. [96].

For the peak assignment and for comparison with simulated spectra, static calculations of the MeOH monomer and dimer using the hybrid functional B3LYP-D3/def2-SVP, the GGA BLYP-D3/def2-SVP, as well as the composite method B97-3c were performed with Turbomole 7.4.1[97–99] and standard settings. **Table 8** lists the calculated frequencies and their assignments for the major experimental peaks shown in the low- and high-frequency regions, see top panels in **Figures 5** and **7**. We observe that the composite B97-3c predicts frequencies at larger values as compared to B3LYP and BLYP. In the methanol monomer, lower frequencies are predicted by BLYP as compared to B3LYP, while in the dimer both functionals yield comparable values.

Table 8. Identification of molecular vibrations with regards to the experimental IR spectrum in both low- and high-frequency regions applying static calculations, B3LYP-D3/def2-SVP, BLYP-D3/def2-SVP, and B97-3c/def2-mTZVP levels of theory. Please note that the columns are labeled only by the name of functionals.

$\tilde{\nu}$ [cm^{-1}]/ B3LYP	BLYP	B97-3c	Assignment
	MeOH monomer		
1070	1027	1191	C-O stretching, O-H rotation
1105	1061	1123	C-O stretching
1170	1133	1207	H-C-H bending (rocking), H-O-C-H torsion
1365	1332	1417	H-C-H angle bending, O-H rotation, H-O-C-H torsion
1470	1424	1527	H-C-H angle bending, H-C-H bending (rocking)
1495	1453	1561	H-C-H angle bending, O-H rotation
2960	2869	3049	C-H stretching
3010	2910	3096	H-C-H asymmetric stretching
3100	3018	3166	C-H stretching
3820	3661	4001	O-H stretching
	MeOH dimer		
1070	1054	1094	C-O stretching, O-H rotation, H-C-H bending (twisting)
1085	1091	1112	C-O stretching
1110	1136	1137	H-C-H bending (rocking), C-O stretching, H-O-C-H torsion
1350	1311	1415	H-C-H bending (rocking), H-O-C-H torsion, O-H rotation
1425	1424	1454	H-C-H angle bending, O-H rotation
1460	1459	1562	H-C-H bending (wagging), H-C-H angle bending, C-O stretching
2950	2967	3034	C-H stretching, O-H stretching, H-C-H bending (wagging)
3085	3056	3177	C-H stretching, C-H rotation, O-H stretching
3690	3513	3931	C-H stretching, O-H stretching
3835	3682	4002	O-H stretching

Overall, the obtained Raman spectra are somewhat disappointing, especially regarding relative peak positions and intensities, none of which showed improvement by changing any of the investigated parameters. The reasons for that have yet to be identified and might be related to the necessity to obtain the polarizability tensor by numerical differentiation of the dipole moments making it prone to numerical errors. Also, the cross-correlation terms present in Equation (4) converge very slowly so that longer simulation times might improve the Raman spectra.

Beginning with the time step (see Figures S1–S4, Supporting Information), in both low- and high-wavenumber regions there are shifts in IR and Raman spectra signals for the Voronoi integration (henceforth termed Voronoi) and Wannier localization (henceforth termed Wannier) in comparison to the experimental spectrum.[96] The first sharpest peaks in Figures S2 and S4, Supporting Information, for both Voronoi and Wannier are slightly shifted to lower wavenumbers. The positions of the second and third peaks in the experimental spectrum at around 1120 and 1450 cm^{-1}, respectively, are reproduced well by both methods. Moreover, the Voronoi results of the 0.8 fs time step in high-frequency region for both IR and Raman spectra is not observable, indicating that this time step does not perform satisfactorily. Thus, the overall best agreement with experiment is achieved using a time step of 0.5 fs.

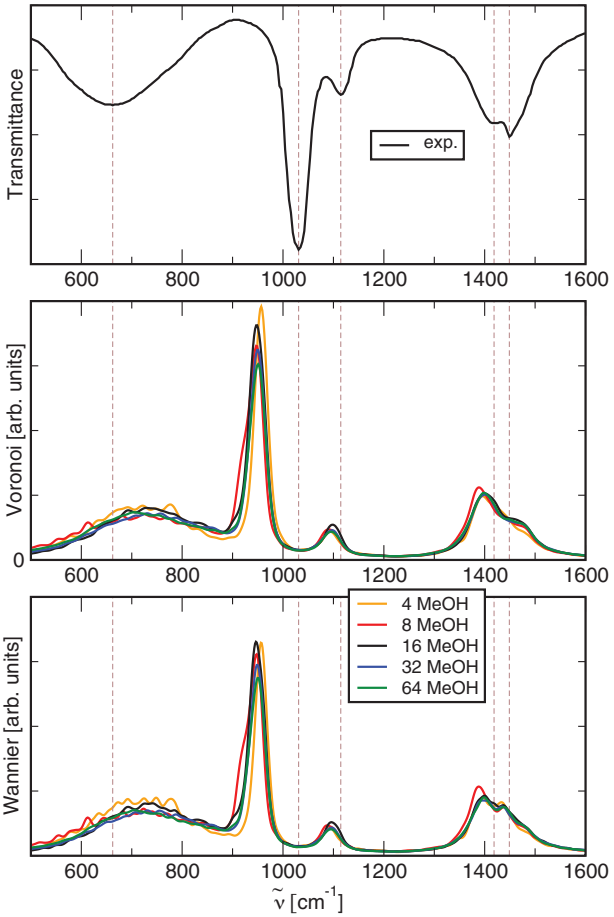

Figure 5. Influence of the system size on the spectra. IR spectra in the low-frequency range using the Voronoi (middle) and Wannier (bottom) approach. The top panel shows the experimental spectrum.[96]

A similar behavior for the peak locations can be detected in case of the different SCF convergence criteria (see Figures S5–S8, Supporting Information). All SCF convergence parameters fulfill the experimental peaks relatively well, with the fluctuations and deviations from experiment being slightly stronger for the more loose criterion of 10^{-5}.

6.1. System Size

In order to investigate the influence of the system size, we simulated different systems containing 4, 8, 16, 32, and 64 molecules using a time step of 0.5 fs, an SCF convergence threshold of 10^{-6},

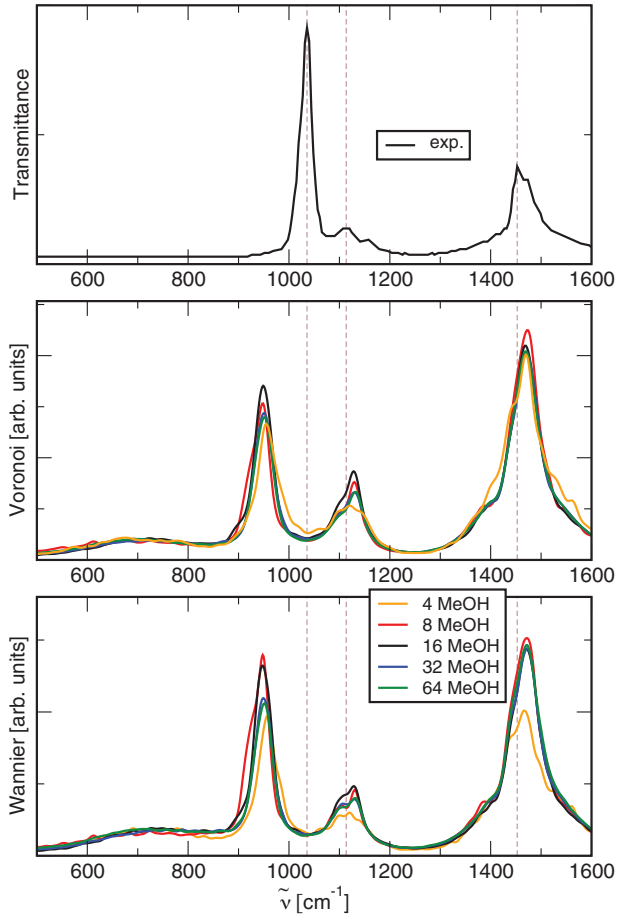

Figure 6. Same as Figure 5, showing Raman spectra in the low-frequency range.

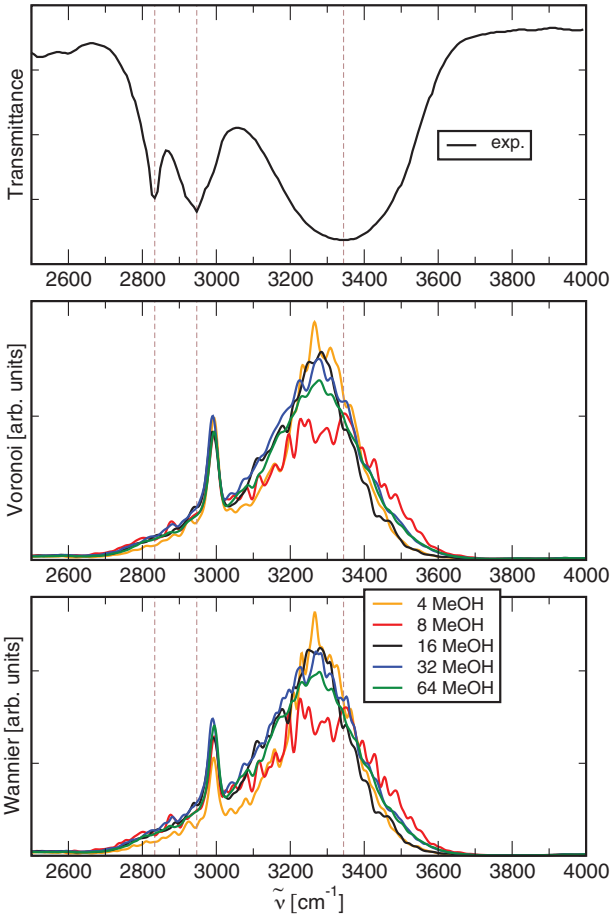

Figure 7. Same as Figure 5, showing IR spectra in the high-frequency range.

and the BLYP functional with the (SR)-DZVP basis set. Figures 5 and 7 show IR spectra for different system sizes in the low and high-frequency region, respectively, while **Figures 6** and **8** display the corresponding Raman spectra, each in comparison to the experimental spectra. As is visible from Table 1, the error in the density can be as large as 6 %. If both the system size and the performed equilibration time are sufficient (larger systems require longer equilibration), the error in density converges to about 3 %.

The low-frequency region (500 to 1600 cm^{-1}) of the IR spectrum is shown in Figure 5. Comparing to the experimental spectrum (with the sharpest peak at 1030 cm^{-1}), it can be observed that for all system sizes this sharp peak (C–O stretching, O–H rotation, and H–C–H bending, see Table 8) is shifted to the lower wavenumbers, located at around 955 cm^{-1} for 4-MeOH system, at 945 cm^{-1} for the 8- and 16-MeOH systems, and at 950 cm^{-1} for the 32- and 64-MeOH systems.

The relative intensity of the sharpest peak decreases in height following the order starting from the 4-MeOH system, then 16-, 8-, 32-, and lastly the 64-MeOH system, with the highest peak for the 4-MeOH system and the lowest one for the 64-MeOH system. The Wannier spectrum does not noticeably deviate from the Voronoi spectrum, except that 4- and 16-MeOH systems possess a comparable relative intensity for the peak at around 955 and 945 cm^{-1}, respectively. Also the small experimental peaks at approximately 1115 and 1420 cm^{-1} involving mostly H–C–H bending are well reproduced by all system sizes.

For the Raman spectrum in the low-frequency region (Figure 6), the peak locations do not converge properly to the experimental reference. While the experimental spectrum possesses the highest intensity peak at around 1040 cm^{-1}, both Voronoi and Wannier show their first peaks with an approximate 90 cm^{-1} shift to the lower wavenumber region. The most intense peak is obtained for the 8-MeOH system in Wannier and 16-MeOH system in Voronoi, both at around 950 cm^{-1}, and the least intense one for the 4-MeOH system at about 955 cm^{-1}. Similarly, the peak at 1475 cm^{-1} is more intense in both calculated spectra than in the experimental spectrum. In this case, a shift to higher wavenumbers is observable which renders the scaling[100] of the peak locations difficult.

In the high-frequency region of the IR spectrum (**Figure 7**), the first experimental peak at around 2835 cm^{-1} of C–H stretching regarding Table 8 is almost not notable in both Voronoi and Wannier calculations. The second peak of the calculated spectra is shifted by approximately 50 cm^{-1} to higher wavenumbers for both calculated spectra. The broad band of O–H and C–H stretching (Table 8) is shifted by about 110 cm^{-1} to lower wavenumbers

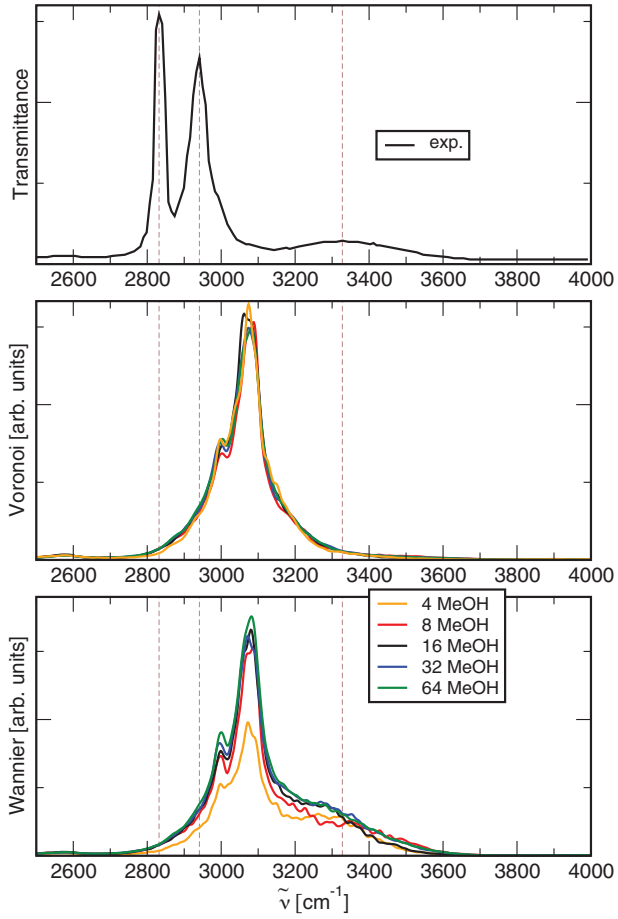

Figure 8. Same as Figure 5, showing Raman spectra in the high-frequency range.

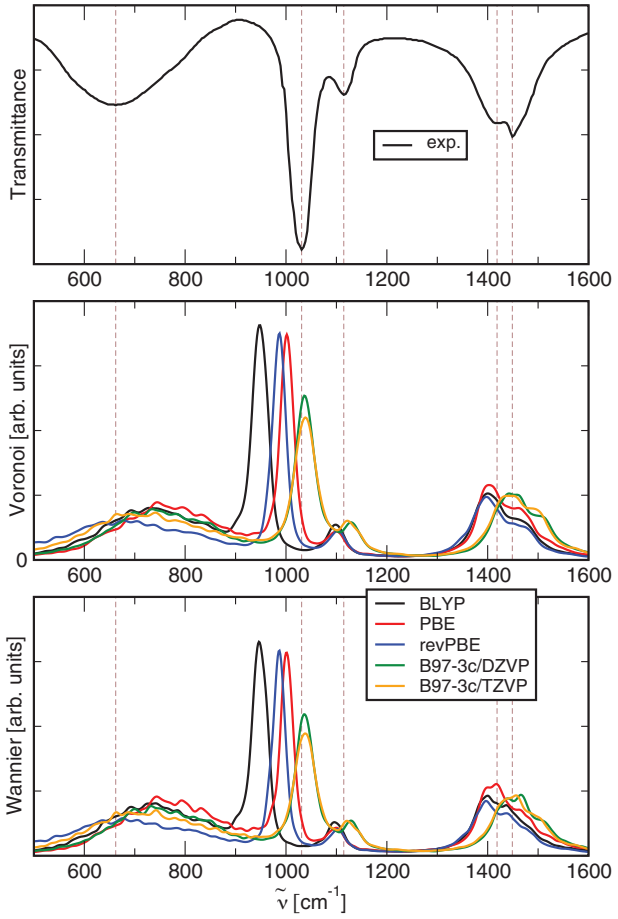

Figure 9. Influence of different GGA functionals on the spectra. IR spectra in the low-frequency range using the Voronoi (middle) and Wannier (bottom) approach. The top panel shows the experimental spectrum.

as compared to the experiment (at 3345 cm^{-1}), with the highest intensity for the 4- and the lowest for the 8-MeOH systems.

Having a glance at the Raman spectra in the high-frequency regime (Figure 8), the calculated spectra from both Wannier and Voronoi methods show a larger deviation from the experimental reference than the IR spectra (c.f. Figure 7). The first peak of the experimental spectrum is located at around 2830 cm^{-1} and the second one at 2940 cm^{-1}, both almost similar in intensity. Contrary, we observe for both the Voronoi and the Wannier spectra only a small shoulder at around 3000 cm^{-1}, followed by a pronounced peak at approximately 3075 cm^{-1}. The highest intensity is found for the 4-MeOH system in case of Voronoi and for the 64-MeOH system in case of Wannier approach.

6.2. Functional

We now compare the influence of the density functional approximation, which is shown in **Figures 9–12** for the low- and high-frequency regions of IR and Raman spectra, respectively, for GGA functionals and in **Figures 13–16** for two hybrid functionals. All simulations were performed on the 16 MeOH-System with a timestep of 0.5 fs, the SCF convergence threshold being 10^{-6} and with the (SR)-DZVP basis set. It is obvious that different functional sets perform differently and that the choice of the functional has much more influence on the IR and Raman spectra rather than the choice of the system size, time step, and SCF convergence criterion in our chosen parameter range. The positions of the first sharp and second smaller peaks originating from the B97-3c/DZVP and B97-3c/TZVP at around 1030 and 1130 cm^{-1} in the IR spectrum (Figure 9),[101] respectively, are in a very good agreement with the experimental spectrum for both Voronoi and Wannier, together with the two small peaks at around 1450 cm^{-1}. The intensities, however, are smaller from B97-3c/TZVP as compared to B97-3c/DZVP and further increase in the following order: PBE, revPBE, and BLYP, with the latter one possessing the most intense peak. It should be noted that in the 600 to 800 cm^{-1} region, the revPBE functional shows the strongest concordance with the experimental spectrum.

A similar behavior can be observed for the Raman spectra in the low-frequency region (**Figure 10**). However, this is not true for the last two small peaks at around 1115 and 1450 cm^{-1}, which represent a satisfactory match of PBE functional to the experimental spectrum.

Regarding the performance of GGA functionals in the high-frequency region given in Figures 11 and 12, alike the low-frequency region, the broadest, most conspicuous peak at around

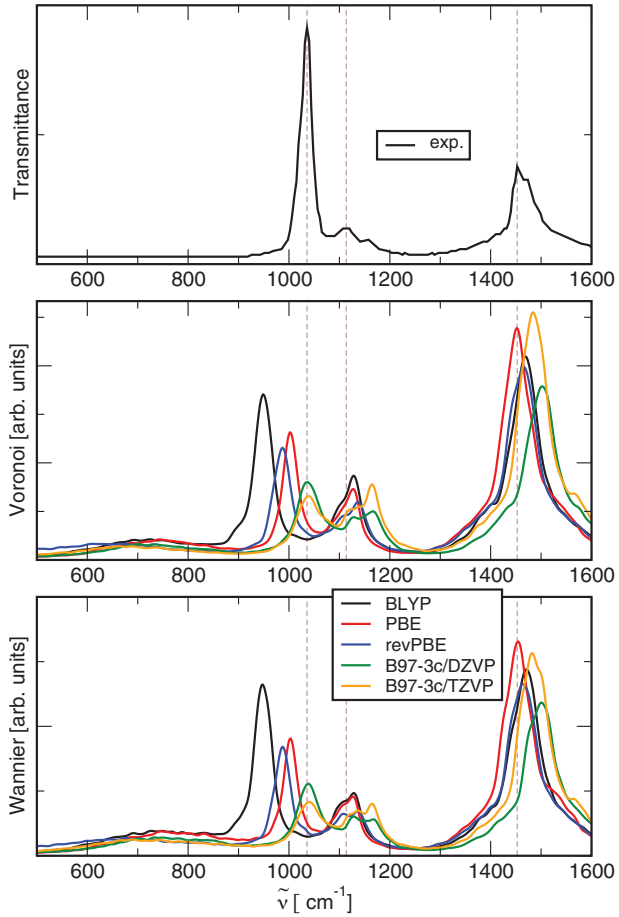

Figure 10. Same as Figure 9, showing Raman spectra in the low-frequency range.

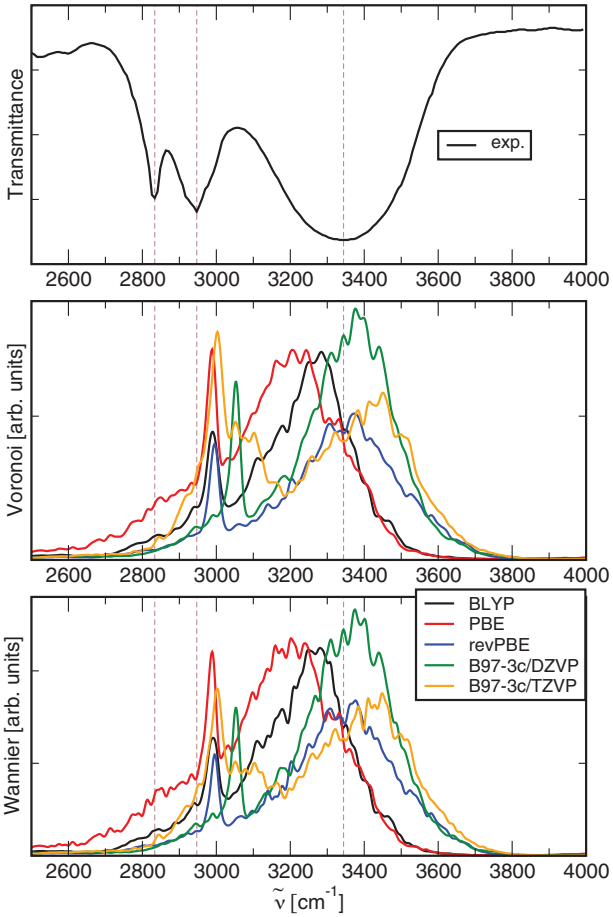

Figure 11. Same as Figure 9, showing IR spectra in the high-frequency range.

3345 cm^{-1} in the experimental IR spectrum is reproduced best by the B97-3c/DZVP, B97-3c/TZVP, and revPBE functionals for both Voronoi and Wannier. However, the B97-3c/DZVP functional represents the experimental peak at about 2830 and 2940 cm^{-1} with an approximately 220 cm^{-1} shift to higher wavenumbers for both peaks at around 3050 and 3160 cm^{-1}, respectively, in both Voronoi and Wannier panels of Figure 12. Given the shifted location of the broad peak for the other functional sets, these apparently meet the experimental 2940 cm^{-1} peak better.

In the high-frequency Raman spectrum (Figure 12), the broad but flat peak at 3330 cm^{-1} is reproduced best by B97-3c/DZVP followed by revPBE in the Wannier spectrum. The experimental double peak at around 2830 and 2940 cm^{-1} is reproduced by all calculated spectra for both Voronoi and Wannier in that a tiny peak appears next to a much larger peak, but the double peak of B97-3c/DZVP and B97-3c/TZVP exhibits a shoulder in both Voronoi and Wannier Raman spectrum.

Focusing now on the hybrid functional results of the low-frequency region, shown in Figures 13 and 14, the B3LYP functional provides peak intensities and locations in a very good agreement with the experiment, while the PBE0 functional shows larger shifts of the peaks. Please note, the GGA functionals show here a shift to lower wavenumbers.

The high-wavenumber region in the IR spectrum displayed in **Figure 15** is improved upon the usage of the hybrid functionals. The aforementioned double peak is now reproduced by both methods: Voronoi and Wannier. Interestingly, the Wannier flat peak at around 3500 cm^{-1} in the corresponding Raman spectrum (Figure 16), which is located at around 3320 cm^{-1} in the experimental spectrum, evidently outperforms the Voronoi one.

To conclude the comparison of different functionals, the trend observed from the frequency analysis following a static calculation in Table 8 is also observed in the spectra obtained from AIMD simulations. In the simulations, all functionals—PBE0 being an exception—tend to underestimate the experimentally observed vibrational frequencies, which is in agreement with scaling factors for static harmonic frequencies being larger than one for these functional/basis set combinations.[102]

6.3. Basis Set

Finally, the influence of the basis set on the quality of vibrational spectra is studied and results are shown in **Figures 17–20**. All simulations were performed on the 16 MeOH-System with a timestep of 0.5 fs, the SCF convergence threshold being 10^{-6} and

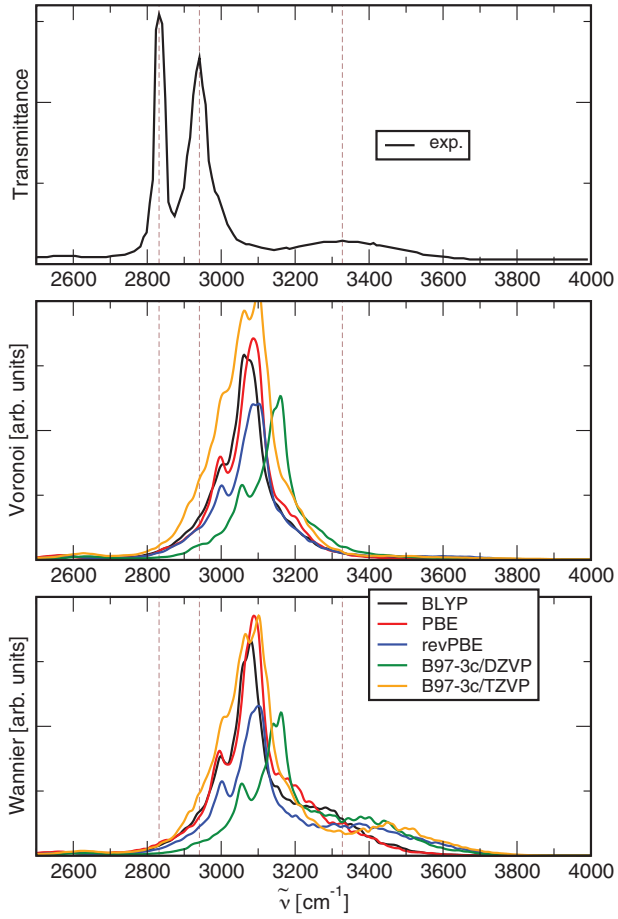

Figure 12. Same as Figure 9, showing Raman spectra in the high-frequency range.

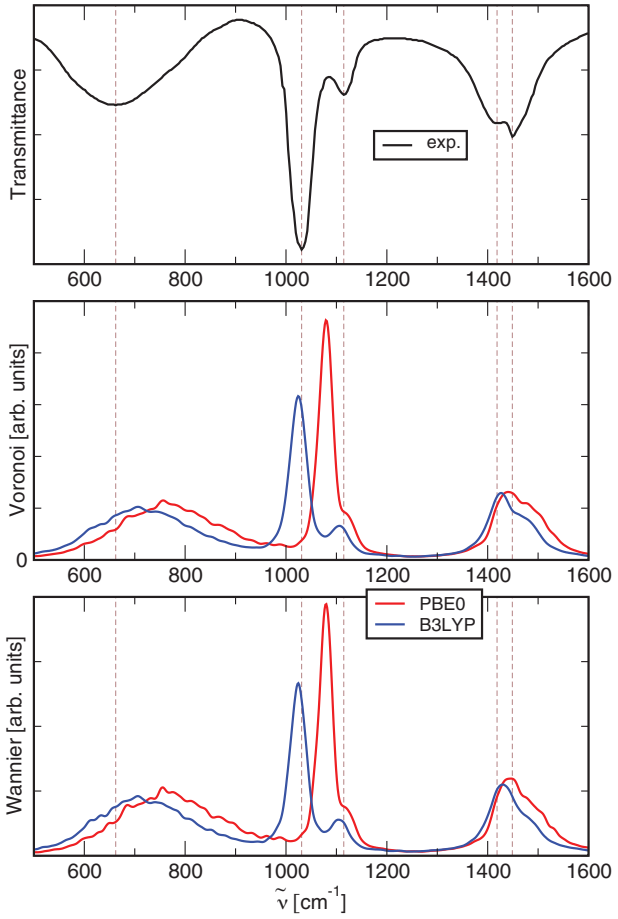

Figure 13. Influence of two different hybrid functionals on the spectra. IR spectra in the low-frequency range using the Voronoi (middle) and Wannier (bottom) approach. The top panel shows the experimental spectrum.

with the BLYP functional. Figure 17, showing the low-frequency IR spectrum, emphasizes that the DZVP and TZVP basis sets both reproduce the first experimental peak at 1030 cm^{-1} for Wannier and Voronoi methods with a shift to the lower frequencies at 950 cm^{-1}. The entire experimental peaks at around 1115, 1415, and 1450 cm^{-1} are located at about 1100, 1405, and 1465 cm^{-1} for DZVP and, respectively, at around 1095, 1395, and 1460 cm^{-1} for TZVP. This supports the claim that the DZVP and TZVP basis sets represent the experimental spectrum adequately when it comes to the peak location. The SZV basis set on the other hand exhibits an unsatisfying trend in case of the Voronoi and provides no spectrum at all in case of Wannier. The inferior performance of the SZV basis set which causes the scarcity of Raman spectrum for Wannier can be conceivably clarified by SZV basis set lacking polarization functions and flexibility to exhibit polarization and hence, dipole moments.

While in the low-frequency Raman spectrum in **Figure 18** the second and the last peaks of experimental spectrum are placed at around the same wavenumber for the DZVP and TZVP basis sets, SZV behaves improperly in case of the Voronoi, especially for the last peak. However, both the spectra resulting from the usage of the DZVP and TZVP basis set do not meet the relative intensity of peaks appropriately.

In the high-frequency region, the usage of the TZVP basis set for calculating the spectrum surpasses that of the DZVP basis set for the IR spectrum in the case of peak locations. The locations of first two experimental IR peaks in **Figure 19** are represented adequately by the spectrum calculated with the TZVP basis for both Voronoi and Wannier.

Considering the Raman spectrum in the high-frequency region in **Figure 20**, the broad experimental peak is absent in Voronoi spectra for all basis sets. In addition, there are differences in peak locations as well as the relative intensities for the first two experimental peaks comparing the Voronoi and Wannier spectra. Nevertheless, the spectrum obtained from the usage of the TZVP basis set displays a better representation of the experimental spectrum than the one obtained from the DZVP basis set calculation. Resembling to the low-frequency region, SZV does not manifest the IR and Raman experimental spectra for Wannier. In addition, the disparity between the experimental IR spectrum and SZV IR spectrum for the Voronoi can be discerned. It shows that the experimental double peak has been shifted extensively to the very end of the frequency region in the Voronoi, at 3590 and 3720 cm^{-1} for the first and the second peaks, respectively.

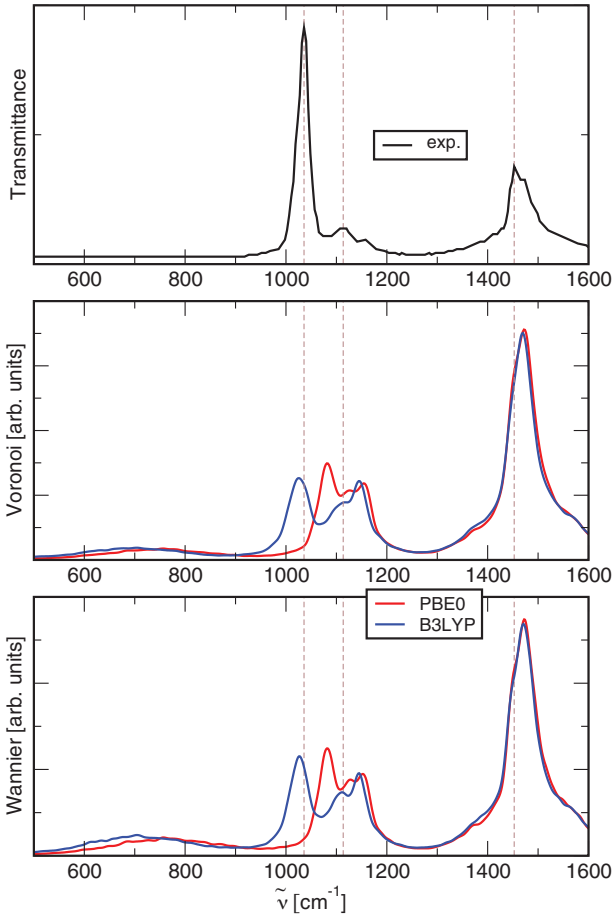

Figure 14. Same as Figure 13, showing Raman spectra in the low-frequency range.

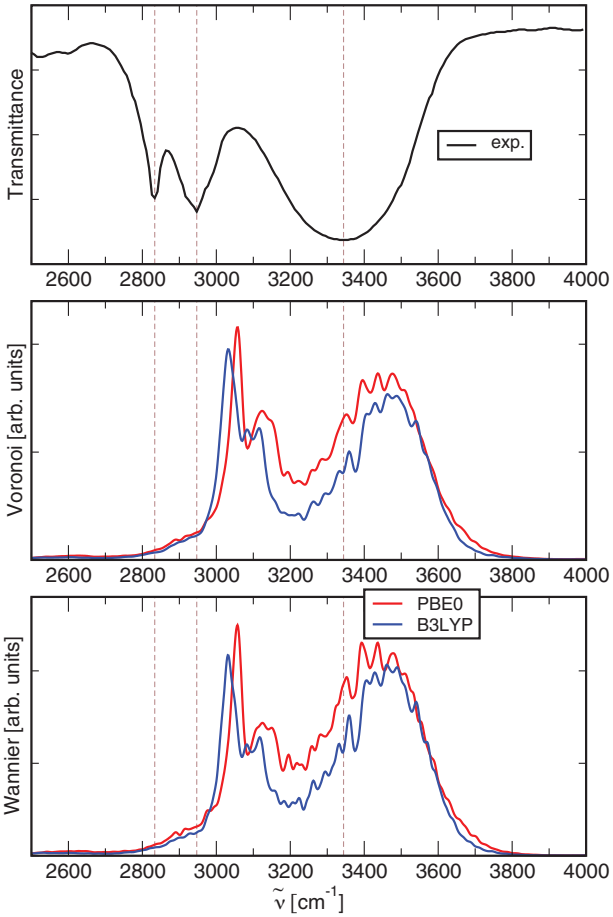

Figure 15. Same as Figure 13, showing IR spectra in the high-frequency range.

7. Conclusion

The aim of this article was to arrive at a better understanding of what parameters determine both the quality and the computational cost of vibrational spectra from AIMD simulations. Moreover, given the relevance of vibrational spectroscopy combined with the large computational effort required for theoretical spectra, a well-grounded recommendation for the computational setup shall be given. Therefore, the scaling behavior of AIMD simulations followed by either maximally localized Wannier functions or radical Voronoi tessellation was examined at the example of the liquid MeOH and the resulting IR and Raman spectra were investigated while parameters such as time step, convergence criteria, system size, functional, and basis set were varied.

Calculations on different computer systems and architectures using up to 576 cores show a very efficient parallelization of the trajectory production and both the Wannier localization and Voronoi tessellation. Although the Wannier scheme is slightly superior in terms of parallelization, it has to be kept in mind that this approach can lead to unphysical fluctuations in the electron density and give rise to additional peaks in the calculated spectrum. A major drawback of the Voronoi tessellation—the need to store large amounts of volumetric data on a grid—has been addressed by the recent introduction of the bqb format,[31] which could shrink the storage requirements by up to 97% in this study.

Regarding the technical parameters, it could be confirmed that for the SCF convergence criterion and the MD time step widely used values of 10^{-6} and 0.5 fs embody good choices and yield reliable spectra.

A key factor that determines the required computational resources is the system size. It has been observed that if only peak positions are sought, relatively small systems including eight units of methanol already yield satisfying results. The relative peak intensities, however, require at least 32 methanol molecules with the high-frequency IR region being most affected.

The assessment of the performance of different density functional approximations for vibrational spectra is very complex and a general recommendation can hardly be given. In particular, the experimental Raman spectrum in the high-frequency range could not be reproduced with satisfying accuracy by any of the studied functionals. Overall, the best agreement was observed for the B3LYP hybrid functional. However, a number of computationally less demanding GGA functionals also yielded good results. Within the investigated GGA functionals, the deviations of the BLYP functional were less systematic as compared to PBE, revPBE, and B97-3c. For the systems investigated in this study,

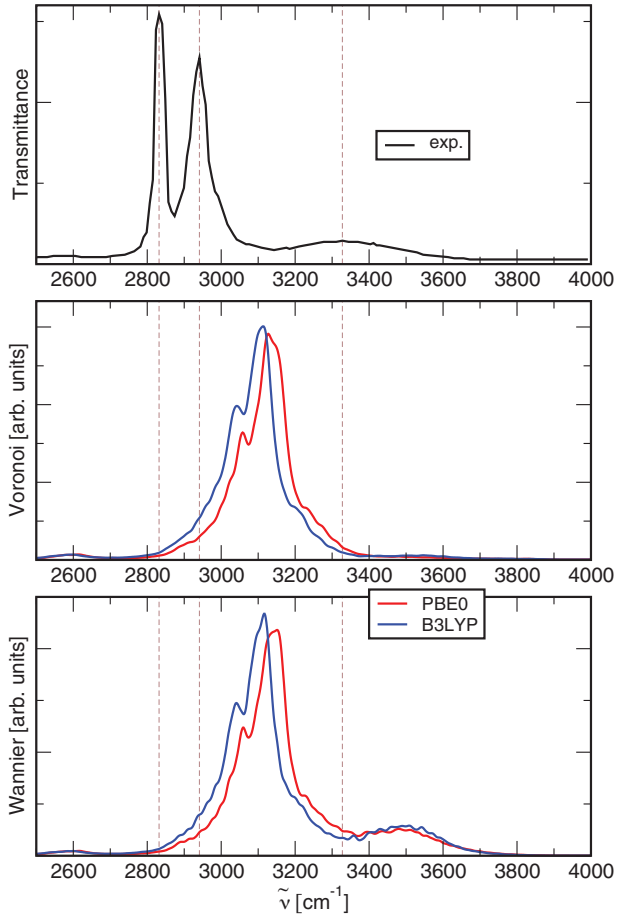

Figure 16. Same as Figure 13, showing Raman spectra in the high-frequency range.

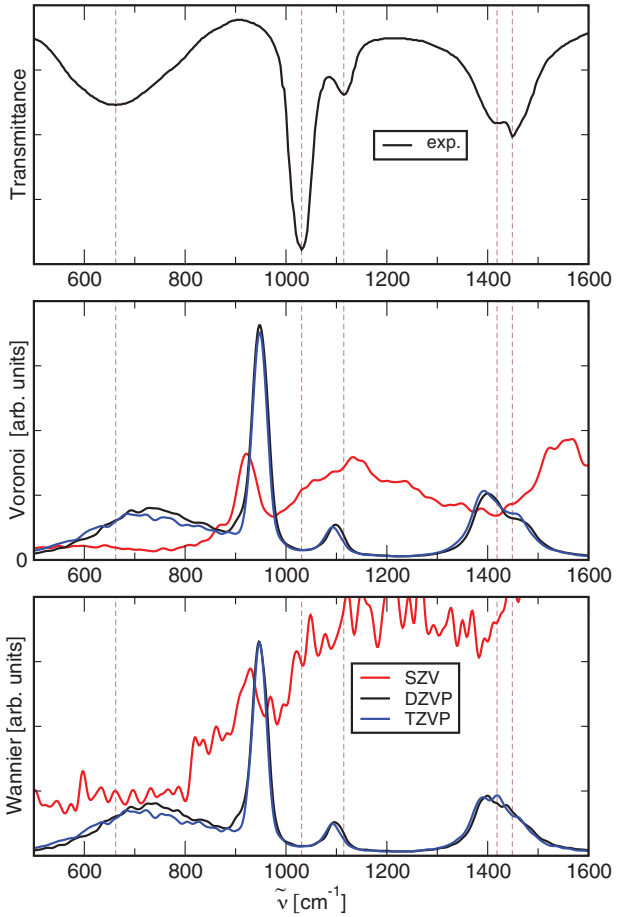

Figure 17. Influence of different basis sets on the spectra. IR spectra in the low-frequency range using the Voronoi (middle) and Wannier (bottom) approach. The top panel shows the experimental spectrum.

the best performing GGA functional is B97-3c/DZVP with an overall good agreement and a systematic deviation that can be treated by applying scale factors and/or a shift, which is common practice for the evaluation of vibrational spectra from DFT.[102] Moreover, observations from our study indicate that the determination of such scaling factors for vibrational data obtained from AIMD simulations is possible. Since effects like conformational sampling or the approximation of the energy landscape by harmonic potentials are no issues in simulations, these scaling factors would be more targeted to the approximations of the functionals themselves and would thus be of high interest. However, such an analysis would require more data points and more functionals and different liquids would need to be investigated, which would justify a separate study.

Concerning the basis set, it could be shown that the spectra produced by the SZV basis set were not even close to the experimental reference, which is not surprising given the lack of polarization functions and the very small flexibility of this small basis set. Based on the findings in this study the recommended choice would be a TZVP basis set. In the case of limited computational resources, the DZVP basis set also yields reasonable spectra except for a notable deviation in the high-frequency range of the Raman spectrum. A more detailed inspection of the cardinal number of the basis as compared to added polarization functions and the resulting spectra as opposed to the required resources would go beyond the scope of this study.

Overall, a considerable number of parameters have been investigated, but the list by far cannot be considered complete. Important parameters to be investigated in further studies include other liquids, the cutoff value for the plane wave basis, the total simulation time, and parameters of the numerical differentiation for the calculation of polarizability tensors.

In general, the evaluation of vibrational spectra from AIMD simulations—while providing highly reliable results including both an accurate description of the electronic structure and a phase space sampling—is computationally very demanding, because the electronic structure has to be determined and subsequently processed. This need for time-dependent electronic structure data makes it necessary to either store electronic densities, which produces large amounts of data, or to perform the analysis on-the-fly while running the simulation. With efficient compression algorithms for electron densities such as the bqb format, it would be possible to separate both processes and map the different elements of the workflow onto different hardware

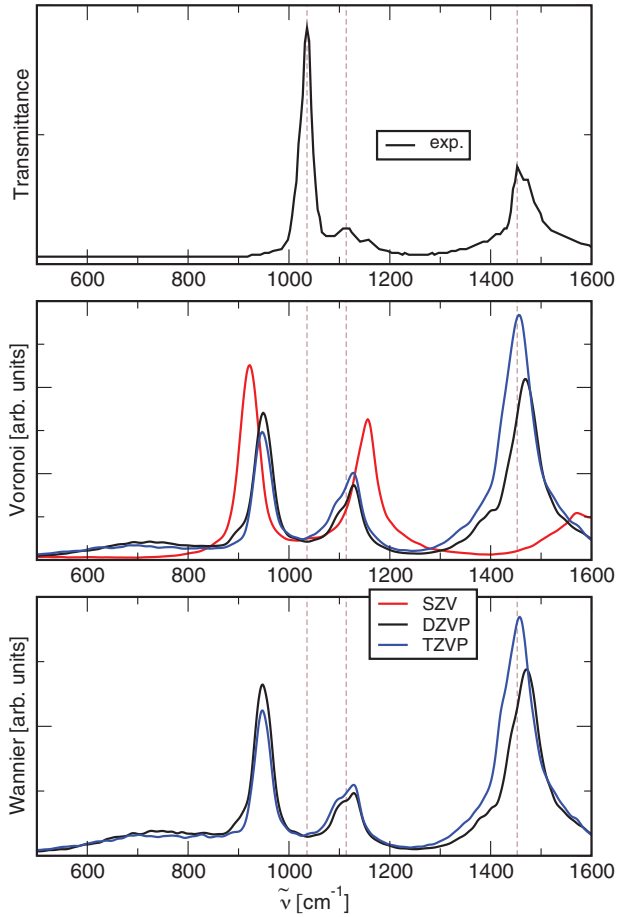

Figure 18. Same as Figure 17, showing Raman spectra in the low-frequency range.

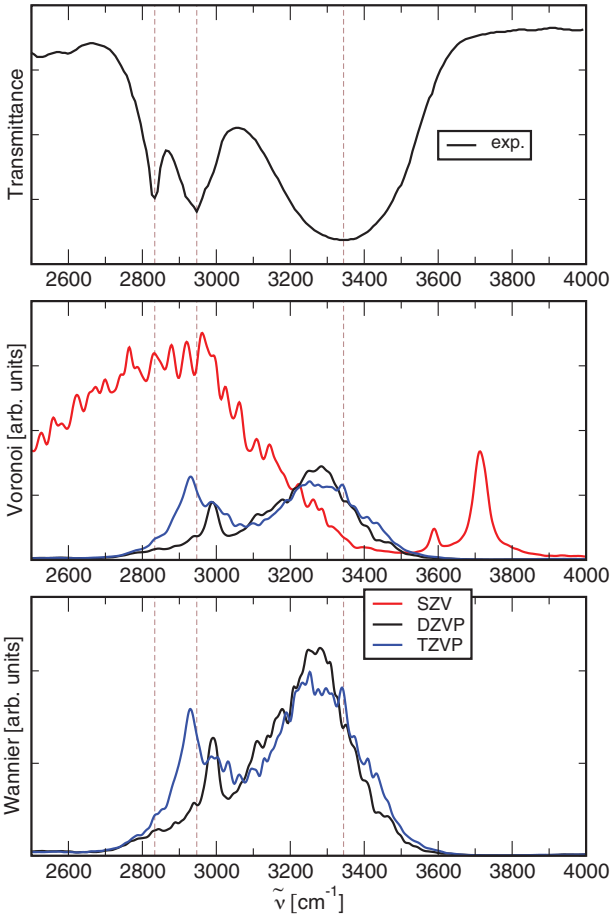

Figure 19. Same as Figure 17, showing IR spectra in the high-frequency range.

platforms, chosen so that they best fulfill the respective code requirements. In future work, we plan to use modular supercomputers to investigate the behavior of the various parts of CP2K[73] and TRAVIS.[62,63] Code analysis and benchmarking campaigns are envisioned to identify the computer modules best fitting each part of the application workflow. We foresee that those parts that are less parallelizable and require a higher single-thread performance will run better on a general purpose cluster. Other code components could, however, profit from acceleration devices, such as graphic cards or many-core processors.

This study has shown that the sole evaluation of spectra based on an existing trajectory is as expensive as generating the trajectory from scratch and calculating spectra on-the-fly, since the electronic structure problem has to be solved in any case. Therefore, the efficient storage of the time evolution of the wavefunction along the trajectory would be necessary in order to efficiently separate the simulation from the spectra evaluation steps.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors would like to thank the reciprocal support and the possibility to use the JURECA Cluster at the Jülich Supercomputing Centre (JSC) from Forschungszentrum Jülich GmbH (F.Z.J.). F.Z.J. thanks the funding received by the Helmholtz Gemeinschaft for the CASCADE-EXN project.

Moreover, the authors thank the International Max Planck Research School on Reactive Structure Analysis for Chemical Reactions (IMPRS RECHARGE) for the support of R.E.

Homepage: https://imprs.cec.mpg.de

E.P. acknowledges funding by the Carl Zeiss Foundation within the Breakthroughs 2019 program.

Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Research data are not shared.

Keywords

ab initio molecular dynamics simulations, benchmark, maximally localized Wannier functions, radical Voronoi tessellation, vibrational spectra

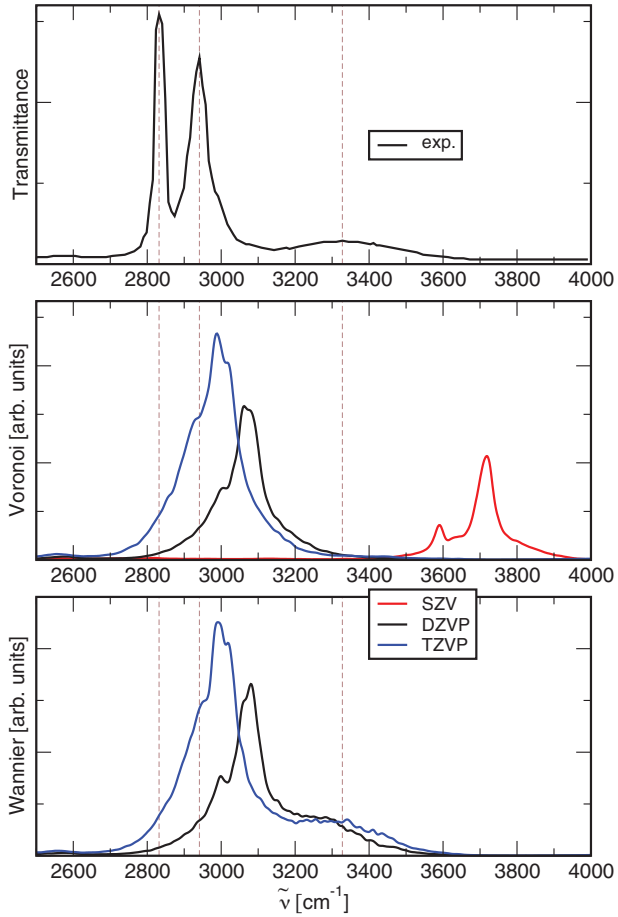

Figure 20. Same as Figure 17, showing Raman spectra in the high-frequency range.

Received: August 5, 2021
Revised: October 19, 2021
Published online:

[1] S. Bakels, M.-P. Gaigeot, A. M. Rijs, *J. Chem. Rev.* **2020**, *120*, 3233.
[2] P. T. Panek, C. R. Jacob, *J. Phys. Chem. Lett* **2016**, *7*, 3084.
[3] M. K. Ghosh, J. Lee, C. H. Choi, M. Cho, *J. Phys. Chem. A* **2012**, *116*, 8965.
[4] V. Barone, M. Biczysko, J. Bloino, M. Borkowska-Panek, I. Carnimeo, P. Panek, *Int. J. Quantum Chem.* **2012**, *112*, 2185.
[5] E. Kraka, W. Zou, Y. Tao, *Wiley Interdiscip. Rev.: Comput. Mol. Sci* **2020**, *10*, e1480.
[6] M. Brüssel, E. Perlt, S. B. C. Lehmann, M. von Domaros, B. Kirchner, *J. Chem. Phys.* **2011**, *135*, 194113.
[7] G. Matisz, A.-M. Kelterer, W. M. F. Fabian, S. Kunsági-Máté, *Phys. Chem. Chem. Phys.* **2015**, *17*, 8467.
[8] M. von Domaros, E. Perlt, *J. Chem. Phys.* **2017**, *146*, 124114.
[9] M.-P. Gaigeot, R. Vuilleumier, M. Sprik, D. Borgis, *J. Chem. Theory Comput.* **2005**, *1*, 772.
[10] M. Thomas, M. Brehm, R. Fligg, P. Vöhringer, B. Kirchner, *Phys. Chem. Chem. Phys.* **2013**, *15*, 6608.
[11] B. Kirchner, J. Vrabec, *Multiscale molecular methods in applied chemistry*, Vol. *307*, Springer Science & Business Media, Berlin **2012**.
[12] K. Wendler, M. Brehm, F. Malberg, B. Kirchner, L. Delle Site, *J. Chem. Theory Comput.* **2012**, *8*, 1570.
[13] A. Bussy, J. Hutter, *Phys. Chem. Chem. Phys.* **2021**, *23*, 4736.
[14] K. Wendler, J. Thar, S. Zahn, B. Kirchner, *J. Phys. Chem. A* **2010**, *114*, 9529.
[15] M. Salanne, B. Rotenberg, S. Jahn, R. Vuilleumier, C. Simon, P. A. Madden, *Theor. Chem. Acc.* **2012**, *131*, 1.
[16] W. J. Glover, P. A. Madden, *J. Chem. Phys.* **2004**, *121*, 7293.
[17] D. M. Wilkins, A. Grisafi, Y. Yang, K. U. Lao, R. A. DiStasio, M. Ceriotti, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 3401.
[18] B. Kirchner, J. Hutter, *J. Chem. Phys.* **2004**, *121*, 5133.
[19] N. Marzari, D. Vanderbilt, *Phys. Rev. B* **1997**, *56*, 12847.
[20] P. L. Silvestrelli, M. Parrinello, *Phys. Rev. Lett.* **1999**, *82*, 3308.
[21] P. L. Silvestrelli, M. Parrinello, *J. Chem. Phys.* **1999**, *111*, 3572.
[22] J. J. Molina, S. Lectez, S. Tazi, M. Salanne, J.-F. Dufrêche, J. Roques, E. Simoni, P. A. Madden, P. Turq, *J. Chem. Phys.* **2011**, *134*, 014511.
[23] M. Thomas, M. Brehm, B. Kirchner, *Phys. Chem. Chem. Phys.* **2015**, *17*, 3207.
[24] J. P. Beck, A. Cimas, J. M. Lisy, M.-P. Gaigeot, *Spectrochim. Acta, Part A* **2014**, *119*, 12.
[25] D. R. Galimberti, S. Bougueroua, J. Mahé, M. Tommasini, A. M. Rijs, M.-P. Gaigeot, *Faraday Discuss.* **2019**, *217*, 67.
[26] M.-P. Gaigeot, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2021**, 119864.
[27] S. Luber, *J. Chem. Phys.* **2014**, *141*, 234110.
[28] S. Wang, *ACS Omega* **2019**, *4*, 9271.
[29] S. Mandal, V. Thakkur, N. N. Nair, *J. Chem. Theory Comput.* **2021**, *17*, 2244.
[30] N. Medvedev, V. Voloshin, V. Luchnikov, M. Gavrilova, *J. Comput. Chem.* **2006**, *27*, 1676.
[31] M. Brehm, M. Thomas, *J. Chem. Inf. Model.* **2018**, *58*, 2092.
[32] C. Apostolidou, *Adv. Theory Simul.* **2020**, *3*, 2000174.
[33] M. Brehm, M. Thomas, *J. Chem. Theory Comput.* **2019**, *15*, 3901.
[34] E. Tsuchida, *J. Phys.: Condens. Matter* **2008**, *20*, 294212.
[35] J. Wilhelm, P. Seewald, M. Del Ben, J. Hutter, *J. Chem. Theory Comput.* **2016**, *12*, 5851.
[36] M. Shishkin, G. Kresse, *Phys. Rev. B* **2006**, *74*, 035101.
[37] J. Wilhelm, D. Golze, L. Talirz, J. Hutter, C. A. Pignedoli, *J. Phys. Chem. Lett.* **2018**, *9*, 306.
[38] M. Del Ben, C. Yang, Z. Li, H. Felipe, S. Louie, J. Deslippe, in *2020 SC20: Int. Conf. for High Performance Computing, Networking, Storage and Analysis (SC)*, IEEE Computer Society, Washington, DC **2020**, pp. 36–46.
[39] J. Hutter, A. Curioni, *ChemPhysChem* **2005**, *6*, 1788.
[40] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, J. Hutter, *Comput. Phys. Commun.* **2005**, *167*, 103.
[41] H. Yu, D. P. Geerke, H. Liu, W. F. van Gunsteren, *J. Comput. Chem.* **2006**, *27*, 1494.
[42] U. Buck, F. Huisken, *Chem. Rev.* **2000**, *100*, 3863.
[43] E. H. Anwander, M. M. Probst, B. M. Rode, *Chem. Phys.* **1992**, *166*, 341.
[44] N. Asahi, Y. Nakamura, *J. Chem. Phys.* **1998**, *109*, 9879.
[45] J.-W. Handgraaf, E. J. Meijer, M.-P. Gaigeot, *J. Chem. Phys.* **2004**, *121*, 10111.
[46] J. A. Morrone, M. E. Tuckerman, *J. Chem. Phys.* **2002**, *117*, 4403.
[47] C. Wang, J. Tan, L. Liu, *AIP Adv.* **2017**, *7*, 035115.
[48] U. Buck, B. Schmidt, *J. Chem. Phys.* **1993**, *98*, 9410.
[49] K. Lin, X. Zhou, Y. Luo, S. Liu, *J. Phys. Chem. B* **2010**, *114*, 3567.
[50] M. S. Green, *J. Chem. Phys.* **1954**, *22*, 398.
[51] R. Kubo, *J. Phys. Soc. Japan* **1957**, *12*, 570.
[52] N. Wiener, *Acta Math.* **1930**, *55*, 117.
[53] A. Khintchine, *Math. Ann.* **1934**, *109*, 604.
[54] S. Baroni, P. Giannozzi, A. Testa, *Phys. Rev. Lett.* **1987**, *58*, 1861.
[55] X. Gonze, *Phys. Rev. A* **1995**, *52*, 1096.

[56] X. Gonze, *Phys. Rev. B* **1997**, *55*, 10337.
[57] X. Gonze, C. Lee, *Phys. Rev. B* **1997**, *55*, 10355.
[58] A. Putrino, D. Sebastiani, M. Parrinello, *J. Chem. Phys.* **2000**, *113*, 7102.
[59] S. Baroni, S. De Gironcoli, A. Dal Corso, P. Giannozzi, *Rev. Mod. Phys.* **2001**, *73*, 515.
[60] P. Umari, A. Pasquarello, *Phys. Rev. B* **2003**, *68*, 085114.
[61] D. McQuarrie, *Sausalito, CA* **2000**, 222–223.
[62] M. Brehm, B. Kirchner, *J. Chem. Inf. Model.* **2011**, *51*, 2007.
[63] M. Brehm, M. Thomas, S. Gehrke, B. Kirchner, *J. Chem. Phys.* **2020**, *152*, 164105.
[64] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, D. Vanderbilt, *Rev. Mod. Phys.* **2012**, *84*, 1419.
[65] B. Gellatly, J. L. Finney, *J. Mol. Biol.* **1982**, *161*, 305.
[66] A. Goede, R. Preissner, C. Frömmel, *J. Comput. Chem.* **1997**, *18*, 1113.
[67] Y. V. Zefirov, P. Zorkii, *Russ. Chem. Rev.* **1989**, *58*, 421.
[68] G. Schaftenaar, J. H. Noordik, *J. Comput.-Aided Mol. Design* **2000**, *14*, 123.
[69] J. M. Martinez, L. Martinez, *J. Comp. Chem.* **2003**, *24*, 819.
[70] L. Martinez, R. Andrade, E. G. Birgin, J. M. Martinez, *J. Comp. Chem.* **2009**, *30*, 2157.
[71] U. Domanska, M. Laskowska, *J. Chem. Eng. Data* **2009**, *54*, 2113.
[72] S. Plimpton, *J. Comp. Phys.* **1995**, *117*, 1.
[73] CP2k, A General program to perform molecular dynamics simulations, CP2k developers group under the terms of the GNU General Public License.
[74] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, et al., *J. Chem. Phys.* **2020**, *152*, 194103.
[75] J. VandeVondele, J. Hutter, *J. Chem. Phys.* **2007**, *127*, 114105.
[76] S. Goedecker, M. Teter, J. Hutter, *Phys. Rev. B* **1996**, *54*, 1703.
[77] C. Hartwigsen, S. Goedecker, J. Hutter, *Phys. Rev. B* **1998**, *58*, 3641.
[78] M. Krack, *Theor. Chem. Acc.* **2005**, *114*, 145.
[79] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[80] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comp. Chem.* **2011**, *32*, 1456.
[81] G. Hodson, E. Wilkes, S. Azevedo, T. Battaglene, In *BIO Web of Conferences*, volume *9*. EDP Sciences, **2017**, 02028.
[82] D. Krause, P. Thörnig, *J. Large-Scale Res. Facil.* **2016**, *2*, A62.
[83] E. Suarez, N. Eicker, T. Lippert, *Supercomputing Evolution at JSC*, Vol. *49*, John von Neumann Institute for Computing, Germany **2018**.
[84] E. Suarez, N. Eicker, T. Lippert, *Modular Supercomputing Architecture: From Idea to Production*, Vol. *3*, CRC Press, Boca Raton, FL **2019**.
[85] J. VandeVondele, J. Hutter, *J. Chem. Phys.* **2003**, *118*, 4365.
[86] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
[87] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[88] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.
[89] Y. Zhang, W. Yang, *Phys. Rev. Lett.* **1998**, *80*, 890.
[90] J. G. Brandenburg, C. Bannwarth, A. Hansen, S. Grimme, *J. Chem. Phys.* **2018**, *148*, 064104.
[91] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158.
[92] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.
[93] F. Stein, J. Hutter, V. V. Rybkin, *Molecules* **2020**, *25*, 5174.
[94] J. VandeVondele, J. Hutter, *J. Chem. Phys.* **2007**, *127*, 114105.
[95] E. Perlt, P. Ray, A. Hansen, F. Malberg, S. Grimme, B. Kirchner, *J. Chem. Phys.* **2018**, *148*, 193835.
[96] SDBSWeb, National institute of advanced industrial science and technology, http://sdbs.riodb.aist.go.jp (accessed: August 26, 2014).
[97] S. G. Balasubramani, G. P. Chen, S. Coriani, M. Diedenhofen, M. S. Frank, Y. J. Franzke, F. Furche, R. Grotjahn, M. E. Harding, C. Hättig, A. Hellweg, B. Helmich-Paris, C. Holzer, U. Huniar, M. Kaupp, A. Marefat Khah, S. Karbalaei Khani, T. Müller, F. Mack, B. D. Nguyen, S. M. Parker, E. Perlt, D. Rappoport, K. Reiter, S. Roy, M. Rückert, G. Schmitz, M. Sierka, E. Tapavicza, D. P. Tew, et al., *J. Chem. Phys.* **2020**, *152*, 184107.
[98] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165.
[99] K. Reiter, M. Kühn, F. Weigend, *J. Chem. Phys.* **2017**, *146*, 054102.
[100] B. Kirchner, J. Blasius, L. Esser, W. Reckien, *Adv. Theory Simul.* **2021**, *4*, 2000223.
[101] The recommended basis set for this parametrization of the B97-D functional in CP2K is TZVP. In this study, we additionally applied the DZVP basis set. For this reason, the B97-3c/TZVP results are discussed in comparison to other GGA functionals rather than in the basis set section.
[102] M. K. Kesharwani, B. Brauer, J. M. Martin, *J. Phys. Chem. A* **2015**, *119*, 1701.