

# Decomposing total IR spectra of aqueous systems into solute and solvent contributions: A computational approach using maximally localized Wannier orbitals

Radu Iftimie<sup>a),b)</sup>

Department of Chemistry, New York University, New York, New York 10003

Mark E. Tuckerman<sup>c)</sup>

Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, New York, New York 10003

(Received 15 November 2004; accepted 21 March 2005; published online 6 June 2005)

The theoretical principles underpinning the calculation of infrared spectra for condensed-phase systems in the context of *ab initio* molecular dynamics have been recently developed in literature. At present, most *ab initio* molecular dynamics calculations are restricted to relatively small systems and short simulation times. In this paper we devise a method that allows well-converged results for infrared spectra from *ab initio* molecular dynamics simulations using small systems and short trajectories characteristic of simulations typically performed in practice. We demonstrate the utility of our approach by computing the imaginary part of the dielectric constant  $\epsilon''(\omega)$  for H<sub>2</sub>O and D<sub>2</sub>O in solid and liquid phases and show that it compares well with experimental data. We further demonstrate that maximally localized Wannier orbitals can be used to separate the individual contributions of different molecular species to the linear spectrum of complex systems. The new spectral decomposition method is shown to be useful in present-day *ab initio* molecular dynamics calculations to compute the magnitude of the “continuous absorption” generated by excess protons in aqueous solutions with good accuracy even when other species present in the solutions absorb strongly in the same frequency window. © 2005 American Institute of Physics.

[DOI: 10.1063/1.1908950]

## I. INTRODUCTION

Understanding the structure and dynamics of aqueous environments is essential for garnering insight into many processes of chemical and biological relevance. A first step towards understanding the fundamental role played by water in biological systems consists of investigating the nature and dynamics of the hydrogen-bond patterns that water molecules form with their neighbors. Unfortunately, the microscopic details cannot be resolved experimentally primarily because most experimental methods, such as diffusion of light, x ray, NMR, or neutron diffraction, which are so powerful for the structure determination of macromolecules, have difficulties<sup>1</sup> associated with localizing small and floppy H<sub>2</sub>O units.

Infrared spectroscopy represents one of the most sensitive methods for the study of hydrogen bonds.<sup>2,3</sup> Indeed, until a few decades ago, the oversensitivity of the traditional IR transmission spectroscopy methods to the presence of hydrogen bonds often led to catastrophic saturation effects<sup>4</sup> that limited dramatically the utility of the approach for the study of aqueous systems. However, the progress recently achieved in the use of the attenuated total reflection technique in linear spectroscopy<sup>5</sup> provides experimentalists with an extremely

convenient and accurate method to obtain optical and dielectric constants of liquids. Moreover, experimental advances in the field of nonlinear midinfrared spectroscopy<sup>6</sup> have opened the way for ultrafast time-resolved investigations of the dynamics of the hydrogen-bond network in aqueous solutions.

However, several limitations exist regarding the nature of the information that can be obtained via IR experimental methods. The information obtained in a typical spectroscopy experiment represents an average taken over distances that are limited at the lower bound by the wavelength of the external electromagnetic radiation. In the case of mid-IR spectroscopy of water, this means that only the response of the system averaged over approximately 10<sup>10</sup> molecules can be resolved. In particular, information regarding the nature of the *short-range* forces acting between neighboring molecules in aqueous solutions must be disentangled from the total IR signal, an operation that is often performed using controversial techniques.

In contrast with experimental approaches, theoretical calculations of the linear optical constants can be performed using linear response theory without explicitly including an external electromagnetic field. Within linear response theory, the imaginary part of the dielectric constant of a system is proportional to the Fourier transform of the autocorrelation function of the time derivative of the total dipole moment. Consequently, if total dipole moments could be unambiguously decomposed into a sum of local, molecule-based dipole moments, one could define the contribution of a mo-

<sup>a)</sup>Electronic mail: radu.ion.iftimie@umontreal.ca

<sup>b)</sup>Present address: Département de Chimie, Université de Montréal, Montréal H3C 3J7, Canada.

<sup>c)</sup>Electronic mail: mt33@nyu.edu

lecular species to the IR spectrum by computing the time-dependent values of the cross correlation between the local and total dipole moments.

In the present work it will be demonstrated that a *chemically meaningful* decomposition of the IR spectrum of aqueous solutions can be performed using Car–Parrinello<sup>7</sup> (CP) molecular dynamics. Recently, we derived<sup>8,9</sup> new CP equations of motion that intrinsically evolve maximally localized<sup>10–12</sup> Wannier<sup>13</sup> orbitals, thereby preserving the representation or “gauge” in which the orbitals are maximally localized (we have termed this the preservation of the “Wannier gauge”<sup>8,9</sup>). The Wannier gauge provides a natural definition for local dipole moments. We will first demonstrate that a practical procedure for obtaining high-quality IR spectra from *ab initio* molecular dynamics trajectories of only a few picosecond can be formulated by employing the harmonic approximation<sup>14</sup> for the calculation of quantum time correlation functions via classical molecular dynamics in conjunction with windowing techniques<sup>15</sup> for harmonic analysis with discrete Fourier transforms. This assertion will be supported by comparing the absolute values of the frequency-dependent dielectric constants for liquid and solid H<sub>2</sub>O and D<sub>2</sub>O with experimental data. The origin of any noticeable discrepancies between theory and experiment will be quantified and explained, and remedies will be proposed that are likely to improve agreement in future investigations.

The second part of the present work will present the details of the spectral decomposition approach based on maximally localized orbitals and illustrate the method in a study of completely dissociated mineral acids. It will be demonstrated that the absolute spectrum of the excess proton in water obtained by subtracting the spectrum of bulk water from that of an aqueous solution of HCl compares well with experimental data as well as with that of a solution containing H<sub>3</sub>O<sub>aq</sub><sup>+</sup> and FHF<sub>aq</sub><sup>-</sup>, from which the spectra of the bulk water and FHF<sub>aq</sub><sup>-</sup> have been extracted. This “in silico” experiment will serve as a proof of consistency for the method, since both H<sub>3</sub>O<sub>aq</sub><sup>+</sup> and FHF<sub>aq</sub><sup>-</sup> form hydrogen bonds with large proton polarizabilities<sup>16</sup> and generate<sup>3</sup> “continua of absorption” in IR spectra that would be otherwise difficult to separate.

## II. COMPUTING ABSOLUTE FREQUENCY-DEPENDENT OPTICAL AND DIELECTRIC CONSTANTS IN *AB INITIO* MOLECULAR DYNAMICS SIMULATIONS

Several experimental setups can be currently employed to measure the response of a macroscopic system to the action of an external electromagnetic field the spectrum of which extends from the microwave to the visible regions. In a typical optical experiment, for example, one could measure the index of refraction  $n(\nu)$  of the system. In an absorptivity experiment, on the other hand, the quantity being measured would be the Neperian Beer–Lambert absorptivity  $\alpha(\nu)$ , while in a dielectric relaxation experiment it is the dielectric constant  $\varepsilon(\nu)$  that would be reported. These quantities, however, are not independent, all of them being representations of the same physical phenomenon, that is, the *linear* response of matter to the action of an external electromagnetic

field. In the first part (A) of this section, we will first review the relations that connect the different linear optical and dielectric constants. In the second part (B), we will propose a practical scheme for obtaining well-converged values for the product  $\alpha(\nu)n(\nu)$  for the entire IR spectral range. In the final part (C), the power of the method will be illustrated by computing different optical and dielectric constants for liquid water and ice icosahedral (Ih) for various isotopic labeling of the hydrogen atoms.

### A. The Kramers–Krönig relations

The optical and dielectric constants are complex frequency-dependent quantities defined as

$$\hat{n}(\nu) = n(\nu) + ik(\nu) \quad (1)$$

and

$$\hat{\varepsilon}(\nu) = \varepsilon'(\nu) + i\varepsilon''(\nu), \quad (2)$$

where the “hat” denotes complex quantity,  $n$  and  $k$  are the real and imaginary components of the complex refractive index  $\hat{n}$ , and  $\varepsilon'$  and  $\varepsilon''$  are the real and imaginary components of the complex dielectric constant  $\hat{\varepsilon}$ . Both  $\hat{n}$  and  $\hat{\varepsilon}$  are complex functions of the frequency  $\nu$ , and their mathematical definition extends to negative frequencies by employing the following relations:

$$\hat{n}(-\nu) = \hat{n}^*(\nu) \quad (3)$$

and

$$\hat{\varepsilon}(-\nu) = \hat{\varepsilon}^*(\nu), \quad (4)$$

where the star denotes a complex conjugation. The optical and dielectric theories<sup>17</sup> of the interaction between an applied electromagnetic field and matter demonstrate that the index of refraction  $\hat{n}$  and dielectric constant  $\hat{\varepsilon}$  are related by the identity

$$\hat{n}^2(\nu) = \hat{\varepsilon}(\nu), \quad (5)$$

and that the (Neperian) Beer–Lambert absorptivity coefficient  $\alpha(\nu)$  can be related to  $k(\nu)$  via the equation

$$\alpha(\nu) = 4\pi\nu k(\nu)/c, \quad (6)$$

where  $c$  is the speed of light in vacuum. The real and imaginary parts of an analytic function such as  $\hat{n}(\nu)$  that can be written as the Fourier transform of a linear and causal process are not independent but obey the so-called Kramers–Krönig relations<sup>18,19</sup>

$$n(\nu) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\tilde{\nu} k(\tilde{\nu})}{\tilde{\nu}^2 - \nu^2} d\tilde{\nu}, \quad (7)$$

$$k(\nu) = -\frac{\nu}{\pi} P \int_{-\infty}^{\infty} \frac{n(\tilde{\nu})}{\tilde{\nu}^2 - \nu^2} d\tilde{\nu},$$

where  $P$  indicates that the principal value of the integrals is taken. Using the antisymmetry relations from Eq. (3), it can be demonstrated<sup>20,21</sup> that the Kramers–Krönig relations are equivalent to the following Hilbert transforms:

$$n(\nu) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{k(\tilde{\nu})}{\tilde{\nu} - \nu} d\tilde{\nu},$$

$$k(\nu) = \frac{-1}{\pi} P \int_{-\infty}^{\infty} \frac{n(\tilde{\nu})}{\tilde{\nu} - \nu} d\tilde{\nu}.$$
(8)

In practice, the principal value of each Hilbert transform can be calculated<sup>20,21</sup> by taking two successive Fourier transforms, namely,

$$n(\nu) = 4 \int_0^{\infty} \int_0^{\infty} k(\tilde{\nu}) \sin(2\pi\tilde{\nu}\tau) \cos(2\pi\nu\tau) d\tilde{\nu} d\tau,$$

$$k(\nu) = 4 \int_0^{\infty} \int_0^{\infty} n(\tilde{\nu}) \cos(2\pi\tilde{\nu}\tau) \sin(2\pi\nu\tau) d\tilde{\nu} d\tau.$$
(9)

Similar equations can be written for the real and imaginary components of the complex dielectric permittivity,

$$\epsilon'(\nu) = 4 \int_0^{\infty} \int_0^{\infty} \epsilon''(\tilde{\nu}) \sin(2\pi\tilde{\nu}\tau) \cos(2\pi\nu\tau) d\tilde{\nu} d\tau,$$

$$\epsilon''(\nu) = 4 \int_0^{\infty} \int_0^{\infty} \epsilon'(\tilde{\nu}) \cos(2\pi\tilde{\nu}\tau) \sin(2\pi\nu\tau) d\tilde{\nu} d\tau.$$
(10)

Equations (9), (10), and (5) will be used later to compute all the optical and dielectric constants from a knowledge of  $\alpha(\nu)n(\nu)$ .

## B. Linear response theory and the “harmonic approximation”

Using the time correlation formalism<sup>22,23</sup> of linear response theory, the value of the frequency-dependent Beer-Lambert absorptivity coefficient  $\alpha(\nu)$  can be written as

$$\alpha(\nu) = \frac{\pi\nu[1 - \exp(-\beta\hbar 2\pi\nu)]}{3\hbar V c n(\nu)\epsilon_0} \int_{-\infty}^{\infty} \exp(2\pi i\nu\tau) \times \langle \hat{\mathbf{M}}(0) \cdot \hat{\mathbf{M}}(\tau) \rangle_{\text{qm}} d\tau,$$
(11)

where qm indicates a quantum mechanical ensemble average,  $\alpha$  has Neperian units,  $\hbar$  is Planck’s constant,  $V$  is the sample’s volume, and  $\epsilon_0$  is the vacuum permittivity. In Eq. (11)  $\hat{\mathbf{M}}(0)$  and  $\hat{\mathbf{M}}(\tau)$  represent the values of the quantum-mechanical total dipole-moment operator written in the Heisenberg representation at times  $t=0$  and  $t=\tau$ , respectively, and are defined using Système International (SI) charge units.

For condensed-phase systems involving many degrees of freedom, such as those of interest in this paper, explicit eigenstate calculations have a far too high computational overhead to be carried out while direct path integral approaches suffer from such severe phase oscillations that convergence cannot be achieved on any realistic time scale.<sup>24</sup> In contrast, classical molecular dynamics can often be used to obtain the analogous correlation functions for a corresponding classical system. However, there is no general prescription for rigorously relating a quantum correlation function to

a classical one. The precise relationship depends on the underlying Hamiltonian generating the classical and quantum dynamics. Nevertheless, Bader and Berne<sup>14</sup> have demonstrated that when the classical and quantum systems are both characterized by identical *effective harmonic* Hamiltonians, and when the total dipole moment can be written as a linear combination of effective normal modes of the system, the quantum correlation function can be computed from the classical one by means of the identity

$$\int_{-\infty}^{\infty} d\tau \exp(2\pi i\nu\tau) \langle \hat{\mathbf{M}}(0) \cdot \hat{\mathbf{M}}(\tau) \rangle_{\text{qm}}$$

$$= \frac{\beta\hbar 2\pi\nu}{1 - \exp(-\beta\hbar 2\pi\nu)} \int_{-\infty}^{\infty} d\tau \exp(2\pi i\nu\tau)$$

$$\times \langle \mathbf{M}(0) \cdot \mathbf{M}(t) \rangle_{\text{cl}},$$
(12)

which, in the general case, becomes an approximation termed the “harmonic approximation.” Here, cl indicates a classical ensemble average, i.e., phase space integration.

It is important to note that a similar result can be obtained, this time for general Hamiltonians, if one assumes the validity of the centroid molecular dynamics (CMD) theory of Cao and Voth<sup>25</sup> and if the total dipole moment can be written as a linear function of the nuclear cartesian coordinates and/or their momenta.<sup>26,27</sup> The only difference between the CMD result and Eq. (12) consists in replacing the classical ensemble average by a “centroid” average.<sup>26,27</sup> In fact, Ramirez *et al.*<sup>28</sup> have recently demonstrated that explicit CMD calculations can actually improve the harmonic approximation calculations, but at the expense of increasing the computational cost by at least an order of magnitude. Moreover, in the case of *ab initio* electronic structure calculations subject to periodic boundary conditions, the electronic dipole moment is *not* a linear function of the nuclear coordinates, a fact that serves to complicate matters significantly. Although the problem of computing time correlation functions involving nonlinear operators in the CMD theory has recently found a formal solution,<sup>29</sup> it is not clear how one should perform the calculation of the dipole-moment centroid symbols as well as that of the multiple-time centroid correlation functions required by the formalism without rendering the simulation prohibitively expensive.

Our approach in the present work is to adopt the computationally less expensive harmonic approximation first and test its accuracy in a series of calculations of the IR spectra of aqueous solutions. Our tests will include calculations of the absolute frequency-dependent optical and dielectric properties of liquid water and ice Ih, as well as the IR difference spectrum of the excess proton in water (see Sec. III). It is important to note that there is no overwhelming *a priori* evidence that any of these systems should behave as a collection of effective harmonic oscillators. On the contrary, Bakker and Nienhuys,<sup>30</sup> using an ultrafast mid-IR nonlinear spectroscopy approach, have recently concluded that proton vibrations inside the hydrogen-bonded network of liquid water are very anharmonic to the point where “in the second excited state of the O–H stretch vibration, the hydrogen atom becomes delocalized between the oxygen atoms of two

neighboring water molecules.” In the same vein, Tuckerman *et al.*<sup>31</sup> and Marx *et al.*<sup>32</sup> have concluded from their first-principles computational studies of aqueous solutions of dissociated acids that the excess proton moves in a double-well potential if zero-point vibration effects are neglected and in a shallow, almost flat potential when zero-point effects are included. Both pictures suggesting that strong anharmonic effects are present.

### C. Results for H<sub>2</sub>O and D<sub>2</sub>O liquid water and ice I<sub>h</sub>

Using the harmonic approximation in Eq. (12), a computationally more convenient expression for  $\alpha(\nu)n(\nu)$  and consequently for

$$\varepsilon''(\nu) = \frac{c}{2\pi\nu} \alpha(\nu)n(\nu) \quad (13)$$

can be obtained via integration by parts. The final result can be written as

$$\alpha(\nu)n(\nu) = \frac{1}{6cV\epsilon_0 k_B T} \int_{-\infty}^{\infty} d\tau \exp(2\pi i\nu\tau) \langle \dot{\mathbf{M}}(0) \cdot \dot{\mathbf{M}}(\tau) \rangle_{cl}, \quad (14)$$

where  $\dot{\mathbf{M}}(t) = d\mathbf{M}(t)/dt$  is the time-derivative of the (classical) total dipole moment at time  $t$ . The form of the right-hand side of Eq. (14) suggests that there are two main sources of error that affect the accuracy of the simulation: First, systematic as well as random errors caused by a quasiergodic behavior of the dynamics and finite simulation time affect the calculation of the correlation function

$$C(\tau) = \langle \dot{\mathbf{M}}(0) \cdot \dot{\mathbf{M}}(\tau) \rangle_{cl}. \quad (15)$$

Second, the infinite and continuous Fourier transform in Eq. (14) must be estimated utilizing the short and discrete time series of  $C(\tau)$  obtained at the end of the simulation. Several methods to perform this transformation have been proposed in literature, but uncontrolled approximations are always involved, resulting in the appearance of an additional error term that is usually difficult to quantify. We have analyzed the merits of two of the most commonly used approaches that enable one to approximate the values of a continuous Fourier transformation of an infinite signal from a finite, discrete, and uniform sample of the signal. These are the maximum entropy method<sup>33</sup> (MEM) and the windowed discrete Fourier transform<sup>15</sup> (WDFT) approach. A direct comparison with experimental data suggests that short (typically 10 ps) trajectories of the total dipole moment sampled every 0.4 fs can be optimally converted into IR spectra by constructing a relatively short time series  $C(\tau)$  evenly symmetrized about the origin with the right-end point removed, in conjunction with the four-term Blackman WDFT method.<sup>15</sup> In the present approach we typically computed 2 048 values for  $C(\tau)$  or, equivalently, imposed the restriction  $\tau_{\min} \leq \tau < \tau_{\max}$  on all time correlation functions. Typical values of  $\tau_{\max(\min)} = +(-)409.6$  fs.

Figures 1 and 2 show a comparison between the values of  $\alpha(\nu)n(\nu)$  obtained from CP simulations and experimental data for liquid and solid H<sub>2</sub>O and D<sub>2</sub>O, the latter using the

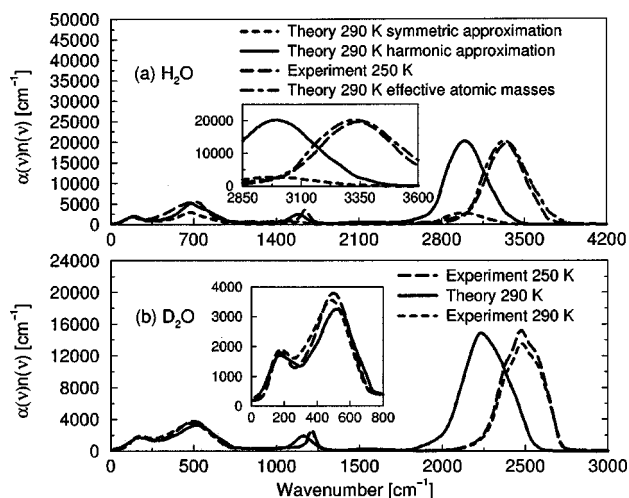


FIG. 1. Comparison between theory and experiment for liquid water: (a) H<sub>2</sub>O and (b) D<sub>2</sub>O. The experiment and the simulation correspond to data obtained at different temperatures. The simulation was performed at  $T = 290$  K, while the experimental spectra were extrapolated (see text) at  $T = 250$  K. A comparison between the results obtained using two different quantum corrections (see text) is included for H<sub>2</sub>O. Scaling the abscissa of the theoretical OH stretch absorption spectrum obtained using the harmonic approximation by 1.115 (Refs. 53 and 60) brings theory and experiment in very good agreement [see inset (a)]. This suggests that the mass of the OH stretch vibration mode is effectively increased by 450 a.u. in our Car-Parrinello simulations (see text). A comparison between theory and experimental data at different temperatures is included for D<sub>2</sub>O. The inset in (b) highlights the remarkable agreement between theory and experiment at small wave numbers.

ice I<sub>h</sub> configuration. The simulations were performed using the density functional theory,<sup>34</sup> the BLYP<sup>35,36</sup> generalized gradient exchange-correlation functional, the plane-wave basis sets, and the Troullier–Martins,<sup>37</sup> pseudopotentials. The plane-wave basis-set cutoff was taken to be 70 Ry, and a fictitious electron mass  $\mu = 400$  a.u., in conjunction with an integration time step  $\delta = 0.1$  fs, was utilized. The simulated systems consist of primitive cubic (liquid phase) and orthorhombic (ice I<sub>h</sub>) boxes containing 32 molecules subject to

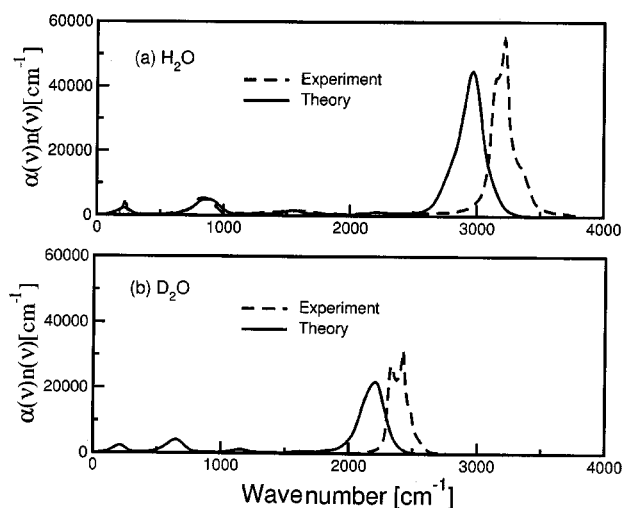


FIG. 2. Comparison between theory (solid line) and experiment (dashed line) for ice I<sub>h</sub>: (a) H<sub>2</sub>O and (b) D<sub>2</sub>O. The experiment and the simulation correspond to data obtained at different temperature. The simulation was performed at  $T = 230$  K, while the experimental spectra were extrapolated (see text) at  $T = 150$  K.

periodic boundary conditions. As noted above,  $\tau_{\min(\max)}$  was taken to be  $-(+)$ 409.6 fs. The dimensions of the primitive boxes were calculated using experimental values for the density. The systems were first equilibrated for a period of 10 ps using a Nose–Hoover chain thermostat<sup>38</sup> on each degree of freedom, after which the thermostats were removed and dynamical trajectories were collected for another 10 ps. The average temperature during the production runs was approximately 290 K for the liquid-phase simulations and 230 K for the ice  $I_h$  systems. All the calculations were carried out using the PINY\_MD<sup>39</sup> package.

The experimental numbers used to construct the plots in Figs. 1 and 2 were taken from the most recent and reliable measurements reported in literature and, in the case of liquid water, were extrapolated to 250 K using the empirical relations suggested by Maréchal.<sup>40–42</sup> In the case of liquid water, the original experimental data consists of tabulated values for  $k(\nu)$  and  $n(\nu)$  obtained by Bertie *et al.*<sup>43,44</sup> and by Maréchal<sup>40,41</sup> using mid-IR attenuated total reflectance spectroscopy, by Zelsmann<sup>45</sup> using far-IR transmission spectroscopy, and by Williams and co-workers<sup>46,47</sup> using near-normal-incidence reflectance spectroscopy. The data for ice  $I_h$  was taken from the works of Bergren *et al.*<sup>48</sup> and Bertie *et al.*<sup>49</sup> who employed a transmission spectroscopy with explicit correction for a thin-film approach. A visual inspection of Figs. 1 and 2 suggests that the use of the harmonic approximation in combination with windowing techniques provides a powerful practical approach for computing accurate IR spectra of condensed-phase systems from short *ab initio* simulations. In particular, the values of the frequencies at which IR absorption occurs are predicted by our computational procedure with a relative error that is always smaller than 10%. A similar conclusion can be drawn for the absolute IR intensity if the comparison is made with experimental data at lower temperatures. In addition, the computed integrated area under the absorption peaks matches almost exactly the experimental value.

The agreement between the computed and measured intensities obtained using the present approach constitutes a marked improvement over previously reported spectra<sup>50</sup> of BLYP water, which employed different quantum corrections<sup>50,51</sup> and a maximum entropy inversion method. For comparison, Fig. 1(a) shows the IR spectrum that would have been obtained using the “symmetrized approximation” on the quantum correlation function, which consists<sup>50,52</sup> of approximating the symmetrized quantum-mechanical dipole-moment correlation function by the corresponding classical correlation function

$$\frac{1}{2}[\langle \hat{\mathbf{M}}(0) \cdot \hat{\mathbf{M}}(t) \rangle_{\text{qm}} + \langle \hat{\mathbf{M}}(t) \cdot \hat{\mathbf{M}}(0) \rangle_{\text{qm}}] \approx \langle \mathbf{M}(0) \cdot \mathbf{M}(t) \rangle_{\text{cl}}. \quad (16)$$

The results displayed in Fig. 1(a) demonstrate the superiority of the harmonic approximation over the entire frequency range. The symmetrized approximation can be improved in the small frequency region by adopting the so-called “Egelstaff correction”<sup>51</sup> as demonstrated by Silvestrelli *et al.*<sup>50</sup> and by Guillot,<sup>52</sup> but its behavior at large wave numbers remains practically unchanged.

It is important to realize that the small differences that exist between experiment and simulation appear to be systematic. In particular, the OH stretch frequency spectrum is consistently redshifted in our simulations with respect to experiment. The observed 10% shift is similar, however, to that observed in a previous *ab initio* MD simulation of the IR spectrum of liquid  $D_2O$ .<sup>50</sup> This redshift artifact could be caused by a slight drag on the nuclear degrees of freedom due to the fictitious electronic mass parameter used in the CP method.<sup>53</sup> The use of an incomplete plane-wave basis constitutes another possible cause for the redshift observed in the computed spectra. According to Sprik *et al.*,<sup>54</sup> the stretching frequency peak could shift to larger wave numbers by as much as  $100 \text{ cm}^{-1}$  if one uses a plane-wave energy cutoff of 150 Ry instead of the 70-Ry value that was used in the present work.

Another important observation consists of the fact that the computed absolute IR intensities compare favorably with experimental data at *lower temperatures* (250 K instead of 290 K for liquid water, and 150 K instead of 230 K for ice  $I_h$ ). This is consistent with recent theoretical studies<sup>55–58</sup> which conclude that the room-temperature structural (i.e., radial distribution functions) and dynamical (i.e., diffusion coefficients) properties of BLYP water are similar to those measured in experiments performed in the supercooled regime.<sup>59</sup> In addition, it should be mentioned that the presence of a nonzero electron mass following the nuclei has the effect of renormalizing the mass of the latter and causes a small decrease in the effective simulation temperature.<sup>53</sup>

The redshift artifact observed in the computational IR spectra can easily be corrected by employing smaller masses for the nuclei. In particular, an examination of the isotope effects suggest that the combined effects of employing a fictitious electron mass  $\mu=400$  a.u. and a basis-set cut-off of 70 Ry has the effect of increasing the effective mass of the H atom by 450 a.u. Indeed, one can see by examining the inset from Fig. 1(a) that the theoretical and the experimental absorption peaks are almost identical if the abscissa of the theoretical spectrum is scaled by  $\sqrt{(1840+450)/1840}=1.115$ . Several empirical methods have been proposed for obtaining such frequency scalings,<sup>53,60</sup> both of which yield scaling factors (or effective masses) close to 1.115 (450 a.u.), although they parse the error differently between finite  $\mu$  and basis-set contributions. Note that changing the definition of the instantaneous temperature to take into account the renormalization of the nuclear masses is also likely<sup>53</sup> to improve the accuracy of the computed absolute IR *intensities*. It should be mentioned at this point, however, that improved agreement between theoretical and experimental observables will likely involve a development of a different exchange-correlation functional as suggested by a series of recent detailed comparative studies of structural and dynamical properties using the Born–Oppenheimer dynamics with the BLYP functional.<sup>56–58</sup>

Using Eqs. (5), (6), (9), and (10) and the computed values for  $\alpha(\nu)n(\nu)$ , we have been able to calculate all the optical and dielectric spectra of the systems under investigation. The infinite double integrals in Eqs. (9) and (10) were approximated by performing discrete Fourier transforms up

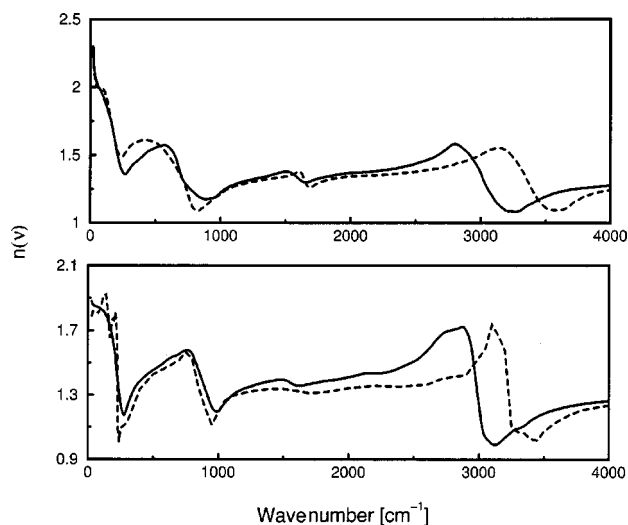


FIG. 3. Comparison between the computed (solid line) and experimental (dashed line) index of refraction for  $\text{H}_2\text{O}$  in liquid (top panel) phase and ice  $\text{I}_h$  (bottom panel).

to a cutoff wave number of  $\tilde{\nu}_{\text{co}} = \nu_{\text{co}}/c = 8000 \text{ cm}^{-1}$  and taking the values of  $k(\tilde{\nu}_{\text{co}})$  and  $n(\tilde{\nu}_{\text{co}})$  from experimental data. To illustrate the power of the computational approach, the theoretical and experimental index of refraction for hydrogenated liquid water and ice  $\text{I}_h$  are depicted in Fig. 3. It is important to emphasize at this point that the simulated IR spectra are uniformly accurate for the entire infrared region down to wave numbers as small as  $40 \text{ cm}^{-1}$ , which is our spectral resolution. The reliability of the simulation results for the entire IR frequency region stands in sharp contrast with that of most of the experimental methods. Transmission experiments for ice  $\text{I}_h$ , for example, require<sup>49</sup> the use of ice films about  $1\text{-}\mu\text{m}$  thick if the sample is not to be black at the peak near  $3300 \text{ cm}^{-1}$ , whereas at  $30 \text{ cm}^{-1}$  films several millimeters thick are required.<sup>49</sup> In the same vein, the reliability of attenuated total reflection spectroscopy experiments decreases dramatically below some cutoff wave number due to an intense absorption from the crystal used in these experiments<sup>40</sup> ( $\approx 660 \text{ cm}^{-1}$  for ZnSe crystals). Computed optical or dielectric infrared constants could therefore help experimentalists extend their experimental data to spectral regions that would otherwise be challenging to investigate, and, therefore, avoid compromising the accuracy of some of the reported optical or dielectric constants obtained via an incomplete Kramers–Krönig transformation.

### III. USING MOLECULAR DIPOLE MOMENTS TO SEPARATE SOLVENT AND SOLUTE CONTRIBUTIONS TO IR SPECTRA OF AQUEOUS SOLUTIONS

#### A. Defining molecular dipole moments

The limited scope of experimental IR investigations of the structure and dynamics of aqueous solutions becomes immediately apparent once the system studied contains solute molecules that absorb radiation in a spectral range where pure water is not transparent. Theory<sup>61</sup> indicates that when two species are at the origin of absorption of an electromagnetic field and the Beer–Lambert linear dependence of the

total absorption on the concentration of the species is fulfilled, the total spectrum of the imaginary dielectric constant  $\varepsilon''$  can be written as a linear superposition of two spectra,  $\varepsilon''_1$  and  $\varepsilon''_2$ , corresponding to the two species. Consequently, the method most often used by experimentalists to separate overlapping spectra of solutes and water consists of subtracting the partial volume-corrected spectra of pure water from that of the aqueous solution. However, the Beer–Lambert law often breaks down at large solute concentration. In addition, the experimental spectra are often incomplete, which introduces uncontrolled errors in the spectra of  $\varepsilon''(\nu)$  that are obtained via Kramers–Krönig transformations. Finally, even for relatively simple solutions such as those of mineral acids in water, it is not clear how many individual species should be considered to participate in the total absorption spectrum observed. For example, given that anomalous proton transport causes a continuous interconversion between the Eigen ( $\text{H}_9\text{O}_4^+$ ) and Zundel ( $\text{H}_5\text{O}_2^+$ ) cation species, should these species be treated separately? This question becomes even more subtle when considering the results of Marx *et al.*,<sup>32</sup> who suggest that these species are only limiting forms of a more complex fluxional defect.

Progress in this area has been recently reported by Max and Chapados<sup>62</sup> who pioneered the use of attenuated total reflection (ATR) spectra, instead of the less reliable spectra of  $\varepsilon''$ , in conjunction with statistical methods such as principal factor analysis<sup>63,64</sup> to overcome some of the aforementioned difficulties. However, using the ATR spectra instead of the  $\varepsilon''$  spectra has no theoretical basis as ATR intensities can be obtained from dielectric spectra via nonlinear,<sup>61</sup> complex transformations. Clearly, a theory of these spectral decomposition approaches is needed to justify and complement the empirical methods.

The mathematical form of Eqs. (13) and (14) suggests that a theoretical approach to the spectral decomposition problem could be formulated, provided that an unambiguous molecular decomposition of the dipole moment can be accomplished. More precisely, consider expressing the total dipole moment  $\mathbf{M}$  of a system as the sum of chemically meaningful molecular dipole moments  $\boldsymbol{\mu}_A$ , i.e.,

$$\mathbf{M} = \sum_A \boldsymbol{\mu}_A, \quad (17)$$

where the sum extends over all molecules present in the solution. Then the dipole-moment correlation function  $C(\tau)$  defined in Eq. (15) can be decomposed into molecular components by means of the equation

$$C(\tau) = \sum_A \langle \dot{\mathbf{M}}(0) \cdot \dot{\boldsymbol{\mu}}_A(\tau) \rangle_{\text{cl}} = \sum_i \langle \dot{\boldsymbol{\mu}}_A(0) \cdot \dot{\mathbf{M}}(\tau) \rangle_{\text{cl}}, \quad (18)$$

and the contribution  $\varepsilon''_A(\nu)$  of the molecule  $A$  to the total dielectric spectrum  $\varepsilon''(\nu)$  becomes

$$\varepsilon''_A(\nu) = \frac{1}{12\pi V \varepsilon_0 k_B T} \int_{-\infty}^{\infty} d\tau \exp(2\pi i \nu \tau) \langle \dot{\mathbf{M}}(0) \cdot \dot{\boldsymbol{\mu}}_A(\tau) \rangle_{\text{cl}}. \quad (19)$$

The development of a first-principles theory that relates empirical but very useful atom, bond and molecule-based

chemical concepts, such as atomic charges, bond orders, or molecular dipole moments, to measurable quantities such as the total charge density remains a largely untapped area of theoretical chemistry. In the case of molecular dipole moments, for instance, the primary challenge of the theory consists in defining a series of possibly overlapping but chemically meaningful “molecular surfaces,” not necessarily defined in the 3D physical space, that would enable one to partition the electron density or the electron orbitals between the various molecules present in the system. Several solutions to this molecular boundary problem have been suggested, including the use of the original<sup>65</sup> and the extended<sup>66,67</sup> versions of the Mulliken population analysis, of the Voronoi polyhedra,<sup>68</sup> of the atoms-in-molecule (AIM) theory,<sup>69</sup> of the natural bond orbitals,<sup>70</sup> and of the maximally localized orbitals.<sup>71,72</sup> In addition, a different approach has been recently pioneered by Pasquarello and Resta<sup>73</sup> that avoids the partitioning of the static dipole moment altogether and suggests that “dynamic molecular dipole moments” be used to compute molecular IR spectra.

However, in the absence of a *first-principles* coherent theory of atomic and molecular chemical descriptors such as those mentioned earlier, the merits of any of the previously mentioned partition approaches can only be determined by comparing simulation results with experiment and checking for consistency with chemical intuition. In the present work, we employ maximally localized Wannier orbitals. Maximally localized Wannier orbitals (MLWO) can be obtained through an appropriate unitary transformation of the commonly used Kohn–Sham orbitals and represent condensed-phase analogs of localized molecular orbitals known from the quantum chemistry works of Boys,<sup>74</sup> Edmiston and Ruedenberg,<sup>71</sup> and Pipek and Mezey.<sup>72</sup>

A meaningful partitioning of the electronic orbitals (and hence the density) among the molecules can be performed using the centers of these MLWO. According to a basic tenet of the theory of polarization formulated by King-Smith and Vanderbilt<sup>75</sup> and Resta,<sup>76</sup> the change in macroscopic polarization of a condensed-phase system is exactly determined by the displacements of the nuclei and of the Wannier function centers, suggesting that the latter provide a classical correspondence to the location of an electron pair. Using this Wannier picture of electron pairs, molecular dipole moments can be computed by summing over the coordinates of the nuclei and of the electron pairs that belong to a particular molecule, i.e.,

$$\boldsymbol{\mu}_A = \sum_{j \in A} \left( Z_j \mathbf{R}_j - 2 \sum_{\alpha \in j} \mathbf{R}_\alpha \right), \quad (20)$$

where the first sum is taken over all the atoms  $j$  that belong to molecule  $A$  and the second sum comprises all Wannier centers that are closer to the atom  $j$  than to any other atom in the system.

We have recently derived<sup>8,9</sup> new Car–Parrinello equations of motion that intrinsically evolve maximally localized orbitals, allowing one to compute the positions of the Wannier centers “on the fly.” We have exploited the fact that the CP equations of motion in our dynamical localization approach constitute an exact scheme to derive efficient integra-

tion algorithms, including multiple time-step methods.<sup>77</sup> Such algorithms allowed us to reduce the computational overhead of the method (including the calculation of the total dipole moment) to  $\approx 10\%$ , a significant improvement over other methods<sup>12,78–80</sup> that have appeared in recent literature. The maximum-localization principle that we employed in the present work is the condensed-phase extension<sup>10–12</sup> of the original Boys<sup>74</sup> principle.

## B. Dynamical generation of Wannier orbitals

The procedure described in the previous subsection for generating and decomposing IR spectra requires that the maximally localized Wannier orbitals be determined at each step of an *ab initio* molecular dynamics simulation. In the CP approach, the nuclear trajectory is generated with forces determined from electronic structure calculations performed on the fly as the simulation proceeds, employing a density-functional description of the electronic structure. In addition, an initially minimized set of orbitals is “propagated” from one nuclear configuration to the next via a fictitious dynamics as opposed to an explicit minimization or diagonalization. This dynamics is generated by a Lagrangian of the form

$$L = \sum_{i=1}^n \mu \int d\mathbf{r} \dot{\psi}_i^*(\mathbf{r}, t) \dot{\psi}_i(\mathbf{r}, t) + \frac{1}{2} \sum_{I=1}^N M_I \dot{\mathbf{R}}_I^2 - E[\{\psi_j\}, \{\mathbf{R}\}] + \sum_{ij} \left( \int d\mathbf{r} \psi_i^*(\mathbf{r}, t) \psi_j(\mathbf{r}, t) - f_{ij} \delta_{ij} \right), \quad (21)$$

where  $n$  is the number of orbitals,  $N$  is the number of nuclei,  $\psi_1(\mathbf{r}, t), \dots, \psi_n(\mathbf{r}, t)$  are the electronic orbitals,  $\mathbf{R}_1(t), \dots, \mathbf{R}_N(t)$  are the nuclear positions,  $f_1, \dots, f_n$  are the set of orbital occupation numbers, and  $\Lambda_{ij}$  are a set of Lagrangian multipliers used to maintain the orthogonality of the orbitals via a constraint. Note that the time label on the orbitals, although the same as that of the nuclei, designates the fictitious orbital dynamics and should not be interpreted as an actual true dynamical evolution of the electrons. The fictitious dynamical parameter  $\mu$  determines the time scale on which this fictitious evolution occurs. The equations of motion generated by this Lagrangian are

$$\mu \ddot{\psi}_i(\mathbf{r}, t) = - \frac{\delta E}{\delta \psi_i^*(\mathbf{r}, t)} + \sum_j \Lambda_{ij} \psi_j(\mathbf{r}, t), \quad (22)$$

$$M_I \ddot{\mathbf{R}}_I = - \frac{\partial E}{\partial \mathbf{R}_I},$$

known as the *Car–Parrinello equations*. An important property of the Lagrangian is that the orbitals are invariant with respect to an arbitrary unitary transformation among the occupied orbitals

$$\psi_i'(\mathbf{r}, t) = \sum_j U_{ij} \psi_j(\mathbf{r}, t), \quad (23)$$

provided  $\Lambda$  transforms as  $\Lambda' = U^\dagger \Lambda U$ . This condition is known in quantum-field theory as *global gauge invariance*. The invariance constitutes an invariance with respect to the group  $SU(n)$ , i.e., the group of all  $n \times n$  unitary matrices with

unit determinant. The unitary freedom can be used to generate maximally localized orbitals for a fixed electronic configuration via a minimization of the so-called *spread functional*, which, for systems subject to periodic boundary conditions, takes the form<sup>10–12,81</sup>

$$\Omega[\{\psi_j\}] = \frac{1}{(2\pi)^2} \sum_i \sum_\alpha \Omega_{\alpha f}(|z_{ii,\alpha}|^2), \quad (24)$$

$$z_{\alpha,ii} = \int d\mathbf{r} \psi_i^*(\mathbf{r}) e^{i\mathbf{G}_\alpha \cdot \mathbf{r}} \psi_j(\mathbf{r}),$$

with  $f(|z|^2) = 1 - |z|^2$ . The minimization condition can be expressed in a convenient manner by writing the unitary transformation  $U$  as  $U = \exp[-i\boldsymbol{\tau} \cdot \boldsymbol{\theta}]$ , where  $\boldsymbol{\tau}$  is a vector of  $n^2 - 1$  Hermitian  $n \times n$  matrices known as the *generators* of  $SU(n)$ , and  $\boldsymbol{\theta}$  is a vector of  $n^2 - 1$  arbitrary real numbers. The minimization of  $\Omega$  is then expressed as

$$\frac{\partial \Omega[\{\psi_j'\}]}{\partial \theta^a} = 0, \quad (25)$$

which yields  $n^2 - 1$  equations for the particular set of components of  $\boldsymbol{\theta}$  that lead to maximally localized orbitals.

In order to generate maximally localized orbitals at each step of a simulation, we seek a set of CP-type equations that propagate the maximally localized orbitals from one configuration to another. This can only be achieved by requiring that the CP Lagrangian be invariant with respect to a *time-dependent* unitary transformation of the form

$$\psi_i'(\mathbf{r}, t) = \sum_j U_{ij}(t) \psi_j(\mathbf{r}, t). \quad (26)$$

The invariance under such a transformation is known as local (in time) gauge invariance. Unfortunately, the CP Lagrangian, as formulated, does not possess this symmetry because the fictitious kinetic energy originally proposed by Car and Parrinello<sup>7</sup> does not remain invariant under the transformation in Eq. (26).

Our solution to this problem is to modify the fictitious kinetic energy by the introduction of a covariant derivative  $D_0$  defined in

$$D_0 = \frac{\partial}{\partial t} - i\boldsymbol{\tau} \cdot \mathbf{A}(t), \quad (27)$$

where  $\mathbf{A}(t)$  is a vector of  $n^2 - 1$  (spatially uniform) gauge fields that transform according to

$$\boldsymbol{\tau} \cdot \mathbf{A}'(t) = U(t) \boldsymbol{\tau} \cdot \mathbf{A}(t) U^\dagger(t) - i \frac{\partial U}{\partial t} U^\dagger(t) \quad (28)$$

under Eq. (26). This covariant derivative is then used to formulate a new CP Lagrangian according to

$$L = \mu \int d\mathbf{r} (D_0 \psi(\mathbf{r}, t))^\dagger (D_0 \psi(\mathbf{r}, t)) + \frac{1}{2} \sum_{I=1}^N M_I \dot{\mathbf{R}}_I^2 - E[\{\psi_j\}, \{\mathbf{R}\}] + \sum_{ij} \left( \int d\mathbf{r} \psi_i^*(\mathbf{r}, t) \psi_j(\mathbf{r}, t) - f_{ij} \delta_{ij} \right), \quad (29)$$

where  $\psi(\mathbf{r}, t) \equiv (\psi_1(\mathbf{r}, t), \dots, \psi_n(\mathbf{r}, t))^T$ . In Eq. (29), the gauge fields  $\mathbf{A}(t)$  need to be treated as dynamical variables. However, if a set of conjugate momenta for these variables is computed according to  $\partial L / \partial \dot{A}^a$ , it is straightforward to show that these momenta are 0. Hence, although Eq. (29) is now invariant with respect to Eq. (26), it contains  $n^2 - 1$  unspecified variables in  $\mathbf{A}(t)$ , a fact that reflects the unitary freedom in the choice of the orbitals. We now have complete freedom to choose the gauge fields so as to affect a particular choice of the unitary matrix, a procedure known as *gauge fixing*. For dynamical generation of Wannier functions, the gauge-fixing condition required is simply that  $\Omega$  be a minimum on the gauge surface, i.e., the surface on which  $U(t) = I$ . Writing  $U(t) = \exp[-i\boldsymbol{\tau} \cdot \boldsymbol{\theta}(t)]$ , the required condition is expressed as

$$\chi^a[\psi] = \left. \frac{\partial \Omega[\psi']}{\partial \theta^a(t)} \right|_{\boldsymbol{\theta}(t)=0} = 0. \quad (30)$$

The gauge-fixing condition is imposed using a technique developed by Dirac.<sup>82,83</sup> The details of this method are beyond the scope of this paper; however, the result is a new set of CP equations of motion for propagating the maximally localized orbitals from one nuclear configuration to the next. These equations, written in Hamiltonian form, read<sup>8,9</sup>

$$\begin{aligned} \dot{\psi}_i(\mathbf{r}, t) &= \frac{\pi_i(\mathbf{r}, t)}{\mu} - \sum_{a,b} \sum_j [D^{-1}]^{ba} \{\chi^a, H_{\text{CP}}\} \tau_{ij}^b \psi_j(\mathbf{r}, t), \\ \dot{\pi}_i(\mathbf{r}, t) &= - \frac{\delta E}{\delta \psi_i^*(\mathbf{r}, t)} - \sum_{a,b} \sum_j [D^{-1}]^{ba} \{\chi^a, H_{\text{CP}}\} i \tau_{ij}^b \pi_j(\mathbf{r}, t) \\ &\quad + \sum_j \Lambda_{ij} \psi_j(\mathbf{r}, t), \end{aligned} \quad (31)$$

$$\dot{\mathbf{R}}_I = \frac{\mathbf{P}_I}{M_I} \quad \dot{\mathbf{P}}_I = - \frac{\partial E}{\partial \mathbf{R}_I},$$

where the matrix  $D$  is given in

$$D^{ab} = \{\chi^a, \mathcal{F}^b\}, \quad (32)$$

with

$$\mathcal{F}^a = - \sum_{ij} \tau_{ij}^a [\langle \psi_i | \pi_j \rangle - \langle \pi_i | \psi_j \rangle], \quad (33)$$

where  $\{\dots, \dots\}$  is the classical Poisson bracket and  $H_{\text{CP}}$  is the CP Hamiltonian. The momentum  $\pi_i(\mathbf{r}, t)$  conjugate to the orbitals is given by the usual formula  $\pi_i(\mathbf{r}, t) = \delta L / \delta \dot{\psi}_i^*(\mathbf{r}, t)$ .

## C. Results

The applicability of the Wannier-center-based approach to the calculation of molecular infrared spectra was tested in a computer simulation in which the IR difference spectrum

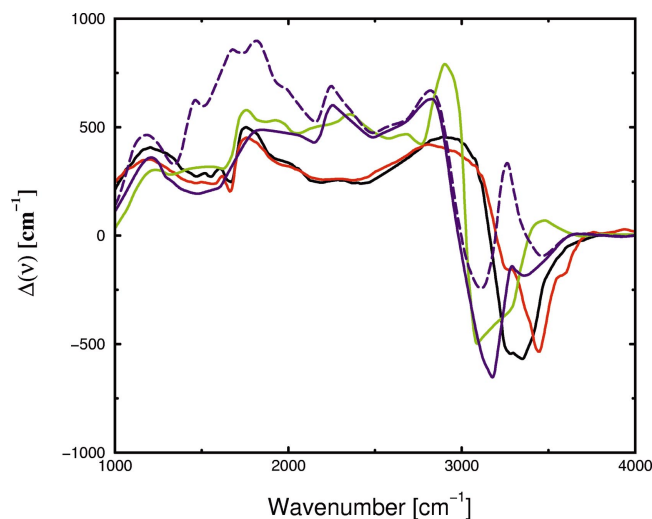


FIG. 4. (Color) Comparison between experimental and theoretical difference spectra  $\Delta(\nu)$  defined in Eq. (34) scaled to a concentration  $c=0.85$  M. The experimental data were obtained in a spectral reflectance study of 3.6 M (black) and 8.8 M (red) aqueous solutions of HBr. The green curve and blue curves represent the computed spectra that were obtained using Eqs. (19) and (20) in a study of HCl (green) and  $\text{H}_2\text{F}_2$  solutions. The dashed curve includes the spectrum of the bifluoride  $\text{FHF}^-$  ion (see text). Note that after subtracting the spectrum of the bifluoride ion from that of the  $\text{H}_2\text{F}_2$  (the result is depicted as the solid blue and green curves), the solution agreement with the spectra of the excess proton from the HCl and HBr systems is recovered.

$\Delta(\nu)$  between the absorption  $[\alpha(\nu)n(\nu)]_c$  of an aqueous solution of a strong acid of concentration  $c$  mol/L and the absorption  $[\alpha(\nu)n(\nu)]_0$  of pure water was computed; i.e.,

$$\Delta(\nu) = [\alpha(\nu)n(\nu)]_c - [\alpha(\nu)n(\nu)]_0, \quad (34)$$

was calculated and compared with experimental data. It should be noted that both terms on the right side of Eq. (34) could be computed in a single simulation of the aqueous solution using the Wannier-orbital approach. Specifically, the second term of the right-hand side of Eq. (34) was taken to be the spectra generated by those “bulk” water molecules that were never hydrogen bonded to the solutes. Simulations were performed on two systems: a 0.85-M aqueous solution of HCl and a 0.85-M aqueous solution of  $\text{H}_2\text{F}_2$  in which the solute was completely dissociated. The simulation details are identical to those reported in the previous section, except that the plane-wave basis-set cutoff employed here is 80 Ry.

Figure 4 shows a comparison between theoretical and experimental values of the difference spectra  $\Delta(\nu)$  defined in Eq. (34) with  $c=0.85$  M. The experimental values plotted in Fig. 4 were obtained by digitizing and scaling the near-normal-incidence spectra of HBr reported by Downing and Williams.<sup>84</sup> The agreement between experiment and theory is good, which demonstrates that maximally localized orbitals can be successfully employed to decompose IR spectra into molecular contributions. In particular, the IR-decomposition approach based on maximally localized orbitals allowed us to separate the spectrum of the excess proton from that of  $\text{FHF}^-$  ion, even though both ions absorb strongly between 1200 and 3000  $\text{cm}^{-1}$ .

#### IV. CONCLUSIONS AND PERSPECTIVES

We have formulated a practical procedure for obtaining high-quality IR spectra from *ab initio* molecular dynamics trajectories of only a few picosecond. The computational investigations undertaken in the present work strongly suggest that IR spectra of aqueous solutions can be accurately calculated within the linear response theory framework by employing the harmonic approximation<sup>14</sup> for the calculation of quantum time correlation functions. Although more investigations are necessary to completely settle the question, the good agreement between theory and experiment that was obtained in the present work for the IR spectra of pure water and aqueous solutions of mineral acids suggests that the validity of the harmonic approximation extends beyond the framework of “effective harmonic systems”<sup>14</sup> for which it was originally conceived. From a practical point of view, the present work suggests that one of the best solutions to the problem of estimating the continuous and infinite Fourier transform in Eq. (11) consists in employing short, but well-converged discrete dipole-moment time correlation functions in conjunction with windowed discrete Fourier transform techniques.

A molecular spectral decomposition method was formulated based on a definition of molecular dipole moments that employs the positions of the centers of maximally localized electron orbitals. We have utilized this method to demonstrate that the absolute spectrum of the excess proton in water can be accurately obtained by subtracting the spectrum of bulk water from that of an aqueous solution of HCl. We have demonstrated that the excess proton spectrum can be accurately disentangled from that of an aqueous solution containing the bifluoride ion  $\text{FHF}_{\text{aq}}^-$  despite the fact that both  $\text{H}_3\text{O}_{\text{aq}}^+$  and  $\text{FHF}_{\text{aq}}^-$  form strong hydrogen bonds and absorb electromagnetic radiation in overlapping frequency intervals.

One of the most exciting aspect of the computational IR spectroscopic approaches presented in this work is their applicability to a host of important problems relevant to chemistry, materials science, and biology that cannot be answered using today’s experimental methodology. In a separate publication we will describe a comparative IR absorption study of the chemical species that are formed in a solution of hydrofluoric acid in water. Experimental evidence<sup>85</sup> suggests that aqueous solutions of HF contain a mixture of undissociated structures, ion pairs, and hydrated protons as well as fluorine ions. However, it is challenging to devise experiments that could isolate each of these chemical species and record their individual properties, including IR spectra. In particular, the IR spectrum generated by ion pairs remains unknown despite the fact that it could prove to be a powerful analytical tool in the investigation of the composition of aqueous solutions of organic acids, including amino acids and proteins.

Turning to potential applications, the theoretical IR spectroscopy could be used to double check the reliability of the reverse Monte Carlo techniques<sup>86</sup> that are often used to suggest typical many-body configurations of amorphous systems from experimental pair-correlation functions. It is well known that the uniqueness theorem of statistical mechanics<sup>87</sup>

establishes a one-to-one correspondence between pair-correlation functions and interaction potentials only if the latter are pairwise additive. However, when polarization as well as other many-body interactions must be taken into account, as is the case for hydrogen-bonded systems, pair-correlation functions alone cannot uniquely determine higher-order correlation functions. The linear IR spectroscopy, on the other hand, can often highlight the nature of such many-body interactions: The “continuous absorption” that can be observed in the spectra of acids in water reflects the highly polarizable nature of the hydrogen bonds in these systems.<sup>16</sup> The spectral decomposition approach outlined in the present work could thus prove instrumental in suggesting what topological artifacts in the hydrogen-bond network model obtained via a reverse Monte Carlo (RMC) approach should be corrected in order to improve the agreement between experimental and theoretical IR spectra. Some important amorphous hydrogen-bonded systems, the structure elucidation of which could benefit from a computational *ab initio* IR decomposition investigation, include the protein-water interface, the hydration shell of nucleic acids, and the low-density and high-density amorphous forms of ice.

Finally, computational spectral decomposition approaches can be utilized to devise mid-IR pump-probe experiments, in which the selective excitation of a particular molecule or chemical bond is necessary, by identifying optimal spectral windows for the pump and probe lasers. As an example, it is worth mentioning the recent mid-IR pump-probe experiments of Omta *et al.*<sup>88</sup> who employed aqueous solutions of perchlorate ions and concluded that solute ions have a negligible effect on the reorientational viscosity of solvent molecules beyond the first solvation shell. A key component in the design and analysis of their experiment was the fact that IR spectra of perchlorate solutions show an absorption peak at 3600 cm<sup>-1</sup>, which the authors attributed to OH groups hydrogen bonded to perchlorate ions. The spectral decomposition approach presented here can be used to confirm their interpretation of the IR spectra and suggest how similar experiments can be performed using different aqueous systems<sup>85</sup> that exhibit spectral features which are challenging to interpret on an experimental basis alone.

## ACKNOWLEDGMENTS

This work was supported by NSF Grant No. CHE-0121375 and NSF Grant No. CHE-0310107. R.I. is the recipient of a postdoctoral fellowship from the Natural Science and Engineering Research Council of Canada.

<sup>1</sup>D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2267 (1998).

<sup>2</sup>G. C. Pimentel and A. L. McClellan, *The Hydrogen Bond* (Freeman, San Francisco, 1960).

<sup>3</sup>D. Hadzi and S. Bratos, in *The Hydrogen Bonds: Recent Developments in Theory and Experiments*, edited by P. Schuster, G. Zundel, and C. Sandorfi (North-Holland, Amsterdam, 1976), Vol. II.

<sup>4</sup>Y. Maréchal and A. Chamel, *J. Phys. Chem.* **100**, 8551 (1996).

<sup>5</sup>E. Bertie and H. H. Eysel, *Appl. Spectrosc.* **39**, 392 (1985).

<sup>6</sup>T. Elsaesser and H. J. Bakker, *Ultrafast Hydrogen Bonding Dynamics and Proton Transfer Processes in the Condensed Phase* (Kluwer, Dordrecht, Boston, 2002).

<sup>7</sup>R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).

<sup>8</sup>R. Iftimie, J. W. Thomas, and M. E. Tuckerman, *J. Chem. Phys.* **120**, 2169 (2004).

<sup>9</sup>J. W. Thomas, R. Iftimie, and M. E. Tuckerman, *Phys. Rev. B* **69**, 125105 (2004).

<sup>10</sup>N. Marzari and D. Vanderbilt, *Phys. Rev. B* **56**, 12847 (1997).

<sup>11</sup>R. Resta, *Phys. Rev. Lett.* **82**, 370 (1999).

<sup>12</sup>G. Berghold, C. J. Mundy, A. Romero, J. Hutter, and M. Parrinello, *Phys. Rev. B* **61**, 10040 (2001).

<sup>13</sup>G. H. Wannier, *Phys. Rev.* **52**, 191 (1937).

<sup>14</sup>J. S. Bader and B. J. Berne, *J. Chem. Phys.* **100**, 8359 (1994).

<sup>15</sup>F. J. Harris, *Proc. IEEE* **66**, 51 (1978).

<sup>16</sup>G. Zundel, *Adv. Chem. Phys.* **111**, 1 (2000).

<sup>17</sup>G. R. Fowles, *Introduction to Modern Optics* (Holt, Rinehart and Winston, New York, 1975).

<sup>18</sup>F. Stern, *Solid State Phys.* **15**, 331 (1963).

<sup>19</sup>S. Maeda and P. N. Schatz, *J. Chem. Phys.* **36**, 571 (1962).

<sup>20</sup>H. Frohlich, *Theory of Dielectrics* (Oxford University Press, Oxford, 1963).

<sup>21</sup>*Theory of Dielectrics*, edited by A. G. Marshall (Plenum, New York, 1982).

<sup>22</sup>R. G. Gordon, *Adv. Magn. Reson.* **3**, 1 (1968).

<sup>23</sup>D. A. McQuarrie, *Statistical Mechanics* (University Science Books, Sausalito, California, 2000).

<sup>24</sup>R. P. Feynman, *Int. J. Theor. Phys.* **21**, 467 (1982).

<sup>25</sup>J. Cao and G. A. Voth, *J. Chem. Phys.* **99**, 10070 (1994).

<sup>26</sup>S. Jang and G. A. Voth, *J. Chem. Phys.* **111**, 2357 (1999).

<sup>27</sup>R. Ramírez and T. López-Ciudad, *Phys. Rev. Lett.* **83**, 4456 (1999).

<sup>28</sup>R. Ramírez, T. López-Ciudad, P. K. P., and D. Marx, *J. Chem. Phys.* **121**, 3973 (2004).

<sup>29</sup>D. R. Reichman, P. N. Roy, S. Jang, and G. A. Voth, *J. Chem. Phys.* **113**, 919 (2000).

<sup>30</sup>H. J. Bakker and H.-K. Nienhuys, *Science* **297**, 587 (2002).

<sup>31</sup>M. E. Tuckerman, D. Marx, M. L. Klein, and M. Parrinello, *Science* **275**, 817 (1997).

<sup>32</sup>D. Marx, M. E. Tuckerman, J. Hutter, and M. Parrinello, *Nature (London)* **397**, 601 (1999).

<sup>33</sup>N. Wu, *The Maximum Entropy Method* (Springer, New York, 1997).

<sup>34</sup>W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory* (Springer, New York, 2000).

<sup>35</sup>A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).

<sup>36</sup>C. Lee, W. Yang, and R. C. Parr, *Phys. Rev. B* **37**, 785 (1988).

<sup>37</sup>N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 1993 (1991).

<sup>38</sup>G. J. Martyna, M. L. Klein, and M. Tuckerman, *J. Chem. Phys.* **97**, 2635 (1992).

<sup>39</sup>M. E. Tuckerman, D. A. Yarne, S. O. Samuelson, A. L. Hughes, and G. J. Martyna, *Comput. Phys. Commun.* **128**, 333 (2000).

<sup>40</sup>Y. Maréchal, *J. Chem. Phys.* **95**, 5565 (1991).

<sup>41</sup>Y. Maréchal, *J. Phys. II* **3**, 557 (1993).

<sup>42</sup>In order to perform the extrapolation, the values of  $\epsilon''(\nu; T)$  obtained at  $T_0=273$  K and  $T_1=348$  K were taken from the experimental literature, and an approximate value for  $\epsilon''(\nu; T=250)$  was obtained using the following empirical relation of Maréchal:<sup>40,41</sup>

$$\epsilon''(\nu; T) = \epsilon''(\nu; T_0) - a(T)[\epsilon''(\nu; T_0) - \epsilon''(\nu; T_1)],$$

where  $a(T) = [(T - T_0)/(T_1 - T_0)]\{1 - 0.2[(T - T_1)/(T_1 - T_0)]\}$

<sup>43</sup>J. E. Bertie and Z. Lan, *Appl. Spectrosc.* **50**, 1047 (1996).

<sup>44</sup>J. E. Bertie, M. K. Ahmed, and H. H. Eysel, *J. Phys. Chem.* **93**, 2210 (1989).

<sup>45</sup>H. R. Zelsmann, *J. Mol. Struct.* **350**, 95 (1995).

<sup>46</sup>H. D. Downing and D. Williams, *J. Geophys. Res.* **80**, 1656 (1975).

<sup>47</sup>L. W. Pinkley, P. P. Sethna, and D. Williams, *J. Opt. Soc. Am.* **67**, 494 (1977).

<sup>48</sup>M. S. Bergren, D. Schuh, M. G. Sceats, and S. A. Rice, *J. Chem. Phys.* **69**, 3477 (1978).

<sup>49</sup>J. E. Bertie, H. J. Labbe, and E. Whalley, *J. Chem. Phys.* **50**, 4501 (1969).

<sup>50</sup>P. L. Silvestrelli, M. Bernasconi, and M. Parrinello, *Chem. Phys. Lett.* **277**, 478 (1997).

<sup>51</sup>P. A. Egelstaff, *Adv. Phys.* **11**, 203 (1962).

<sup>52</sup>B. Guillot, *J. Chem. Phys.* **95**, 1543 (1991).

<sup>53</sup>P. Tangney and S. Scandolo, *J. Chem. Phys.* **116**, 14 (2002).

<sup>54</sup>M. Sprik, J. Hutter, and M. Parrinello, *J. Chem. Phys.* **105**, 1142 (1996).

<sup>55</sup>J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli, *J. Chem. Phys.* **120**, 300 (2004).

<sup>56</sup>I.-F. W. Kuo, C. J. Mundy, M. J. McGrath *et al.*, *J. Phys. Chem. B* **108**,

- 12990 (2004).
- <sup>57</sup>M. V. Fernandez-Serra and E. Artacho, cond-mat/0407237.
- <sup>58</sup>M. V. Frenandez-Serra, G. Ferlat, and E. Artacho, cond-mat/0407724.
- <sup>59</sup>R. Mills, J. Phys. Chem. **77**, 685 (1973).
- <sup>60</sup>I.-F. W. Kuo and D. J. Tobias, J. Phys. Chem. A **106**, 10969 (2002).
- <sup>61</sup>Y. Maréchal, J. Mol. Struct. **648**, 27 (2003).
- <sup>62</sup>J.-J. Max and C. Chapados, Appl. Spectrosc. **53**, 1601 (1999).
- <sup>63</sup>J.-J. Max and C. Chapados, Can. J. Chem. **78**, 490 (2000).
- <sup>64</sup>J.-J. Max and C. Chapados, J. Chem. Phys. **116**, 4626 (2002).
- <sup>65</sup>R. S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).
- <sup>66</sup>A. J. Stone, Chem. Phys. Lett. **83**, 233 (1981).
- <sup>67</sup>A. J. Stone and M. Alderton, Mol. Phys. **56**, 1047 (1985).
- <sup>68</sup>E. R. Batista, S. S. Xantheas, and H. Jónsson, J. Chem. Phys. **111**, 6011 (1999).
- <sup>69</sup>R. Bader, *Atoms in Molecules: A Quantum Theory* (Oxford University Press, Oxford, UK, 1990).
- <sup>70</sup>A. E. Reed, L. A. Curtiss, and F. Weinhold, Chem. Rev. (Washington, D.C.) **88**, 899 (1988).
- <sup>71</sup>C. Edmiston and K. Ruedenberg, Rev. Mod. Phys. **35**, 457 (1963).
- <sup>72</sup>J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916 (1989).
- <sup>73</sup>A. Pasquarello and R. Resta, Phys. Rev. B **68**, 174302 (2003).
- <sup>74</sup>S. F. Boys, Rev. Mod. Phys. **32**, 296 (1960).
- <sup>75</sup>R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993).
- <sup>76</sup>R. Resta, Rev. Mod. Phys. **66**, 899 (1994).
- <sup>77</sup>M. Tuckerman, G. Martyna, and B. Berne, J. Chem. Phys. **97**, 1990 (1992).
- <sup>78</sup>F. Gygi, J. L. Fattebert, and E. Schwegler, Comput. Phys. Commun. **155**, 1 (2003).
- <sup>79</sup>M. Sharma, Y. Wu, and R. Car, Int. J. Quantum Chem. **95**, 821 (2003).
- <sup>80</sup>B. Kirchner and J. Hutter, J. Chem. Phys. **121**, 5133 (2004).
- <sup>81</sup>P. L. Silvestrelli, Phys. Rev. B **59**, 9703 (1999).
- <sup>82</sup>P. A. M. Dirac, Can. J. Math. **2**, 129 (1950).
- <sup>83</sup>P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva University Press, New York, 1964).
- <sup>84</sup>H. D. Downing and D. Williams, J. Phys. Chem. **80**, 1640 (1976).
- <sup>85</sup>P. A. Giguere and S. Turrell, J. Am. Chem. Soc. **102**, 5473 (1980).
- <sup>86</sup>M. A. Howe, J. Phys.: Condens. Matter **2**, 741 (1990).
- <sup>87</sup>R. Evans, Mol. Simul. **4**, 409 (1990).
- <sup>88</sup>S. W. A. W. Omta, M. F. Kropman, and H. J. Bakker, Science **301**, 347 (2003).