Fully First-Principles Surface Spectroscopy with Machine Learning

Yair Litman,* Jinggang Lan, Yuki Nagata, and David M. Wilkins*

Cite This: J. Phys. Chem. Lett. 2023, 14, 8175−8182

ABSTRACT: Our current understanding of the structure and dynamics of aqueous interfaces at the molecular level has grown substantially due to the continuous development of surface-specific spectrocopies, such as vibrational sum-frequency generation (VSFG). As in other vibrational spectroscopies, we must turn to atomistic simulations to extract all of the information encoded in the VSFG spectra. The high computational cost associated with existing methods means that they have limitations in representing systems with complex electronic structure or in achieving statistical convergence. In this work, we combine high-dimensional neural network interatomic potentials and symmetry-adapted Gaussian process regression to overcome these constraints. We show that it is possible to model VSFG signals with fully ab initio accuracy using machine learning and illustrate the versatility of our approach on the water/air interface. Our strategy allows us to identify the main sources of theoretical inaccuracy and establish a clear pathway toward the modeling of surface-sensitive spectroscopy of complex interfaces.

Soft matter interfaces, including aqueous interfaces, solid/liquid interfaces, and liquid/liquid interfaces, are ubiquitous in nature and play a crucial role in many important processes, such as (electro)catalytic/electrochemical applications,1 atmospheric aerosol−gas exchanges,2 and mineral dissolution.3 These processes are governed by the molecular level interaction between the molecules at the interface and the other (electrified) molecules/materials.

To probe the interfacial response, a technique must isolate the signal of the relatively few surface molecules at the surfaces from the enormous contribution due to the bulk.4 Vibrational sum frequency generation (VSFG) is a technique where IR and visible beams are spatially and temporally overlapped and the signal generated at the sum of the input beam frequencies is measured. VSFG is a second-order nonlinear optical process, and as in any other even-order nonlinear optical techniques, the centro-symmetric bulk contributions vanish due to the symmetry of the second-order susceptibility, χ(2), making their signal surface-specific.5 The VSFG signal further possesses molecular specificity: a VSFG signal is enhanced when the IR frequency is resonant with an interfacial molecular vibration. When combining probes with different polarizations, VSFG can provide information on the orientation of interfacial molecules,6,7 the depth profile of the interfacial molecules,8 and molecular chirality.9,10 This makes VSFG a powerful method to characterize the identity, structure, and interaction of the molecules at interfaces.

Experimental VSFG data alone are normally insufficient to connect spectroscopic observables with molecular structure, and atomistic simulations are required to achieve a microscopic understanding. The theoretical calculation of VSFG spectra is more challenging than that of more traditional spectrocopies such as linear IR and Raman, since relatively long simulation times (on the order of nanoseconds) are required to converge the statistics and guarantee that the signal in the bulk-like (centrosymmetric) regions vanishes.

The vibrational resonant component of the second-order susceptibility, χ(2)\text{IR}, in an electronically nonresonant condition, can be computed as

\[
χ^{(2)}(ω) = i \int_0^\infty dt e^{-iωt}\langle\alpha_{pq}(t)P_r(0)\rangle
\]

where αpq is the pq component of the polarizability tensor, ωIR is the frequency of the IR pulse, and Pr is the r component of the polarization vector. The evaluation of eq 1 requires an accurate representation of three objects, namely, (i) potential energy surface (PES), (ii) polarization surface (P-S), and (iii) polarizability surface (α-S).

Several approximations have been applied to calculate the VSFG spectrum of aqueous systems with varying degrees of success. The first group of studies use empirical polarizable models of the dipole moment and polarizability based on ab initio data.11−13 The drawback of such strategies is that accurate polarizable models are normally burdensome to construct and validate and cannot handle bond breaking or formation. The second group uses the projection of the transition dipole and polarizability moments into atomic

Received: July 18, 2023
Accepted: August 29, 2023
Published: September 6, 2023
velocities, leading to the surface-specific velocity–velocity correlation function formalism (ssVVCF) approach.\textsuperscript{14,15} This approximation cannot predict the spectral differences observed with different combinations of laser polarization, since due to the approximations involved, it (incorrectly) predicts $\chi_{\alpha\beta}^{(2)} = \chi_{\alpha\beta}^{(2)}$ and $\chi_{\alpha\beta}^{(2)} = 0$. Furthermore, it is unable to capture important spectral features originating from vibrational coupling.\textsuperscript{16–18}

In this Letter, to overcome the problems inherent in “modeling the dipole moment and polarizability, the slow convergence of the time correlation function, and the accurate prediction of the VSFG spectra at different polarization combinations, we propose the combination of several machine-learning (ML) methods. The PES is evaluated using either density functional theory (DFT) or Behler–Parrinello high-dimensional neural networks (HDNNPs),\textsuperscript{19} and $P$ and $\alpha$ are evaluated by a symmetry-adapted Gaussian process regression (SA-GPR) scheme that enables the prediction of tensorial quantities of arbitrary order.\textsuperscript{20,21} The proposed scheme reduces the calculation cost of the fully first-principles modeling by more than 3 orders of magnitude, enabling the simulation of VSFG spectra at various interfaces at any polarization combination.

The modeling of aqueous solution/air interfaces is normally performed by simulating the system under study in a slab geometry, in which the system extends infinitely along two dimensions and has finite size along the third dimension where it is sandwiched by regions of vacuum. This geometry gives rise to two interfaces that generate VSFG signals with opposite plane, $z$, and $y$ (see a more detailed description of the data set and the training procedure in the Supporting Information).

We start by describing the training of the $P$-S and $\alpha$-S using SA-GPR. The reference data were obtained using DFT at the PBE and PBE0 level and POLY2VS, a polarizable water force field developed by Tanimura and co-workers.\textsuperscript{22} The PBE and PBE0 data sets are obtained from first-principles calculations and constitute the ultimate target of this work. The POLY2VS data set provides access to a molecular decomposition of the $\alpha$ and $P$ quantities, allowing us to critically assess the performance of the ML models. Moreover, we considered two types of data sets, one made up exclusively of bulk structures and a second one made of water clusters from monomers up to hexamers (see a more detailed description of the data set and the training procedure in the Supporting Information).

In Table 1, we summarize the different models for the dielectric properties considered in this work. To enable comparisons between data sets, the error estimates are computed as the root-mean-square error (RMSE) percentage of the intrinsic deviation of the data set and expressed per water molecule. The SA-GPR models can accurately learn $P$ and $\alpha$ with errors below a few percentages of intrinsic variation in the training set. These values represent an error below $8.0 \times 10^{-3}$ D/atom and $1 \times 10^{-3}$ Å$^3$/atom for $|P|$ and $\text{Tr}[\alpha]$, respectively. In the Supporting Information, we report the correlation plots and learning curves of the ML models, together with a brief analysis of the training errors.

Global accuracy estimators, such as the RMSE and absolute error presented above, are useful indicators of the overall performance of the ML models. However, when the learned quantities are used for further calculations, as we do here for $\chi^{(2)}$, it is normally not possible to translate RMSE values directly into an objective accuracy measure of the final target. This is mainly because of the difficulty in performing a rigorous error propagation of the uncertainties in eq 1 but also due to the lack of a unique set of descriptors to quantify how “good” a predicted spectrum is. In this work, we take what we consider to be the strictest validation test and use VSFG spectra to judge the accuracy of our models.

The calculation of $\chi^{(2)}$ for a slab geometry needs to be done with care since the contributions of opposite interfaces interfere destructively and lead to a vanishing response. Assuming that the slab thickness is large enough to accommodate a bulk region in the middle and that a molecular decomposition of $P$ is available, it standard practice to set to zero or flip the sign of the molecular dipoles below the center of mass of the slab to avoid this cancellation.\textsuperscript{14,15} However, molecular dipoles are not observables, and therefore, they are arbitrary in nature. In \textit{ab initio} calculations the molecular decomposition can be achieved by using maximally localized Wannier functions,\textsuperscript{23,24} which incurs the difficulty of requiring direct \textit{ab initio} simulations.\textsuperscript{25} Here, we instead utilized an unbiased data-driven approach. Since we are using an atom-centered decomposition to represent the atomic environments, the SA-GPR predictions can be expressed as a sum of molecular contributions as

$$y(X) = \sum_{\gamma} y(X_{\gamma})$$

(2)

where $\gamma$ represents the set of indices that corresponds to a given molecule, and $X$ and $X_{\gamma}$ the atomic environment of the trial configuration and the atomic environment of the $\gamma$ water molecule, respectively. In this way, SA-GPR can be applied to the calculation of VSFG using eq 1 for slab geometries by applying the following modification to eq 2 for the $P$ predictions

$$P(X) = \sum_{\gamma} g(z_{\gamma}) P(X_{\gamma})$$

(3)

where the surface plane is assumed to be parallel to the $xy$ plane, $z_{\gamma}$ is the $z$-coordinate of the $\gamma$ molecule, and
\[ g(z; z_1, z_2) = \begin{cases} \text{sgn}(z - z_{\text{cw}}) \frac{|z - z_{\text{cw}}|}{z_1} & z_1 \geq |z - z_{\text{cw}}| \\ 0 & z_2 \geq |z - z_{\text{cw}}| \\ \text{sgn}(z - z_{\text{cw}}) \frac{|z - z_{\text{cw}}|}{z_2} & z_2 \geq |z - z_{\text{cw}}| \\ \cos \left( \frac{z_1 - |z - z_{\text{cw}}|}{2(z_1 - z_2)} \pi \right) & z_1 \geq |z - z_{\text{cw}}| \geq z_2 \end{cases} \]  

(4)

where \( z_{\text{cm}} \) is the \( z \)-coordinate of the slab center of mass and \( z_1 \) and \( z_2 \) are two parameters that define the transition between the interfacial and bulk regions. The first condition in eq 4 flips the sign of the molecular contributions to the polarization that emerge from water molecules that are in the vicinity of the bottom interface, the second condition sets to zero the contributions to the polarization corresponding to water molecules that are in the vicinity of the slab center, and the third condition is a smooth transition between the previous two conditions. In this way, the signal of the opposite interfaces interfere constructively, leading to a nonvanishing VSFG signal.

Irrespective of the polarization, the \( \text{Im} \chi^{(2)} \) spectrum of the water/air interface presents a sharp positive peak centered at 3700 cm\(^{-1}\) and a broad band in the 3200–3500 cm\(^{-1}\) region with a negative amplitude. The peak is due to the dangling (free) \( \text{O}-\text{H} \) group of the topmost interfacial water layer, and the band is associated with \( \text{O}-\text{H} \) groups of the interfacial water which form hydrogen bonds with other water molecules. The sign of the amplitudes in the \( \text{Im} \chi^{(2)} \) spectrum can be directly associated with the orientation of the water molecules. More specifically, positive (negative) amplitudes imply that the \( \text{O}-\text{H} \) group points toward the air (toward the bulk) phase. \( \chi^{(2)} \) has only 3 independent elements: \( \chi_{\text{xxz}}, \chi_{\text{xyy}} \) and \( \chi_{\text{yyx}} = \chi_{\text{xyy}} \). The rest of the elements are exactly zero except for large chiral environments. In the case of the water/air interface, the nonzero \( \text{Im} \chi^{(2)} \) tensor components differ essentially by the total intensity, the relative intensity of the free \( \text{O}-\text{H} \) and hydrogen-bonded \( \text{O}-\text{H} \) regions, and the presence or absence of a shoulder on the lower-frequency side of the free \( \text{O}-\text{H} \) peak. Furthermore, the significantly small \( \text{Im} \chi^{(2)} \) spectrum compared with other polarization combinations is consistent with the experimental data.

We proceed now with a systematic evaluation of the ML predictions of the VSFG spectra. First, we consider a PES given by the POLY2VS force field and focus on the impact of the different models of \( \alpha \) and \( P \). Panels a–c in Figure 1 show \( \chi_{\text{xxz}}, \chi_{\text{xyy}}, \) and \( \chi_{\text{yyx}} \) respectively. The ML-POLY-A and ML-POLY-B models are trained on POLY2VS reference data computed on bulk water and water cluster configurations, respectively (Table 1). As shown in Figure 1, both models have similar predictions in their \( \chi_{\text{xxz}} \) and \( \chi_{\text{xyy}} \) and provide a semiquantitative agreement with the reference spectra. This is a rather surprising result since the ML-POLY-A model lacks information of interfacial water structures, whereas the ML-POLY-B model had only limited information on fully solvated water molecules. The prediction of \( \chi_{\text{yyx}} \) is considerably worse than those of the other two tensor components, with ML-POLY-A performing only slightly better than ML-POLY-B for the free \( \text{O}-\text{H} \) peak. By performing cross-predictions using \( P/\alpha \) from the ML model and \( \alpha/P \) from the reference, we identified the \( \alpha \) predictions as the source of the observed inaccuracy (see Figure S15). This result highlights the limitation of using RMSE values and correlation plots as error estimators, as all components of \( \alpha \) show a comparable accuracy.

Electrostatic LR effects are known to play an important role at interfaces due to unbalanced interactions that can build up in the presence of a broken translation symmetry. The implications of using short-range models on structural properties of the water/air interface, such as orientations of water molecules or average density are well documented.
ML models that include LR effects are also known to be more accurate. However, their possible impact on nonlinear spectroscopic responses, such as $\chi^{(2)}$, is less clear. To explore the magnitude of this effect, we trained a long-distance equivariant (LODE) model based on the local value of a atom-density potentials. In Figure 1d−f, we show the predictions using a combination of short-range (SR) and LR contributions, since the combined model normally delivers a superior accuracy compared to the individual ones. In particular, we considered different amounts of LR contributions in the range of 50%−85%. The best combination across all the tensor components is obtained for a 25/75 SR/LR model, but the performance is very similar to that of SR models. These results unequivocally demonstrate that LR effects on the descriptions of $P$ and $\alpha$ have a marginal impact on the $\chi^{(2)}$ spectra.

Having established the suitability of the training set and the methodology to described $P$ and $\alpha$ surfaces, we now consider the spectral changes induced by different ab initio PES. In Figure 2, we present the computed VSFG spectra using SAGPR response surfaces trained on PBE (ML-PBE-A) and POLY2VS (ML-POLY-A) data and six different ab initio potential energy surfaces, namely, PBE-D3, BLYP-D3, revPBE-D3, HSE06, B3LYP, and revPBE0-D3. The $D_2O$ trajectories used for this analysis were available from a previous work. Due to the limited length of the simulations, it was necessary to neglect intermolecular terms in eq 1 to obtain a reasonable converged spectrum (see more details in the Supporting Information). Thus, the width of the negative band is poorly described due to the absence of intermolecular couplings. In all cases except for PBE-D3, the spectra do not show a positive signal below 2200 cm$^{-1}$, in agreement with the latest measurements and simulations. For all the considered PESs, the ML-PBE-A model consistently overestimates (underestimates) the intensity of the free O−D peak (the hydrogen bonded (HB) O−D band), while the ML-POLY-A model predicts relative intensities in better agreement with the experiments. The better performance of ML-POLY-A is a direct consequence of the fact that the reference POLY2VS dipole and polarizability surfaces were fitted to reproduce CCSD/aug-cc-pVQZ, rather than DFT, reference values. The calculations with GGA exchange correlation (XC) functionals (PBE-D3, BLYP, and revPBE-D3) are artificially red-shifted in comparison to the results obtained with hybrid ones (HSE06, B3LYP, and revPBE0-D3) in agreement with previous approximations based on the ssVVCF methodology and results on bulk water. Inclusion of nuclear quantum effects (NQEs) is known to induce a frequency red-shift when compared to the corresponding classical nuclei counterpart spectra. Thus, the frequency agreement between the revPBE0-D3 PES and the experimental data is a fortuitous error compensation. The intramolecular vibrational coupling of water molecules in which one O−D is a hydrogen-bond donor and the other one is free has a distinct spectral feature associated with a shoulder on the free O−H peak. This shoulder has been assigned to the asymmetric stretching mode of interfacial water molecules, and contributions to this shoulder from water molecules forming two hydrogen bonds have been shown to be minor due to cancellation of inter- and intramolecular contributions. PBE-D3, BLYP-D3, HSE06, and B3LYP spectra present a shoulder so separated from the free O−D peak that it can be regarded as a separate peak. In contrast, this spectral feature is correctly displayed by the revPBE0-D3 and revPBE-D3 spectra. Moreover, the differences in performance for revPBE0-D3 and revPBE-D3 are relatively small and mainly impact the intensity of the free O−D peak and an overall frequency shift. These results show that even hybrid func-
the H-bond strength, unlike in bulk water. Functionals, such as HSE06 and B3LYP, considerably overestimate O−D vibrations of water molecules at the topmost layer. We believe that the curvature problem is responsible for the net orientation of the water molecules in the bulk region. We believe that these artifacts are responsible for the (small) discrepancies with the fully ab initio spectra. In particular, the spurious additional orientation of water molecules, which happens in opposite directions for revPBE-D3 and revPBE0-D3 HDNNPs, as shown in Figure S17, might be responsible for increasing and decreasing the splitting between free and HB O−D vibrations of water molecules at the topmost layer. However, for all tested cases, irrespective of the PES, the theoretical predictions overestimate the intensity ratio between the free O−D and the hydrogen-bonded band. This under-estimation has been reported by Paesani and co-workers when the free O−D peak and HB-band and a red-shift of −14 cm$^{-1}$.

So far, we have analyzed simulations where the nuclei were assumed to behave classically, and the spectra have been rigidly red-shifted to account for the ignored NQEs. In the right panel of Figure 3, we show the results obtained with centroid molecular dynamics (CMD), which is a well-established method to simulate the vibrational spectra of condensed phase systems at room temperature including an approximate description of NQEs. CMD is based on the path integral formulation of quantum mechanics, and in this method, the nuclei are evolved according to classical equations of motion on the so-called centroid potential of mean force. NQEs induce a broadening of the free O−D peak and HB-band and a red-shift of 81 cm$^{-1}$ with respect to the corresponding classical nuclei simulations. However, the CMD spectrum is still blue-shifted by 57 cm$^{-1}$ with respect to the experimental result. We also obtained the VSFG spectra using the thermostated ring polymer molecular dynamics (TRPMD) method which is known to deliver more accurate frequencies than CMD at 300 K since it does not suffer from the curvature problem.

The combination of two different types of ML algorithms, namely, HDNNPs and SA-GPR, has allowed us to describe the water/air interface and all the components (polarizations) of second-order response from first principles at an affordable computational cost. The ML models were trained in a general fashion, without employing specific information on the nature of the system, such as physical constraints designed for water, mapping models for water,

The Journal of Physical Chemistry Letters

Figure 3. Normalized $\chi^{(2)}$ spectra of the D$_2$O/air interface at 300 K. Simulated spectra using revPBE-D3 (left) and revPBE0-D3 (right) XC functionals. Classical MD simulations using HDNNPs are presented by solid (ML-PBE-A), dashed (ML-POLY-A), and dotted (ML-PBE0-A) green lines. CMD simulations are depicted with solid orange lines. Experimental spectra are depicted as black solid lines with gray shading areas.

---

**b)** Experiments (Lorentzian)

So far, we have analyzed simulations where the nuclei were assumed to behave classically, and the spectra have been rigidly red-shifted to account for the ignored NQEs. In the right panel of Figure 3, we show the results obtained with centroid molecular dynamics (CMD), which is a well-established method to simulate the vibrational spectra of condensed phase systems at room temperature including an approximate description of NQEs. CMD is based on the path integral formulation of quantum mechanics, and in this method, the nuclei are evolved according to classical equations of motion on the so-called centroid potential of mean force. NQEs induce a broadening of the free O−D peak and HB-band and a red-shift of 81 cm$^{-1}$ with respect to the corresponding classical nuclei simulations. However, the CMD spectrum is still blue-shifted by 57 cm$^{-1}$ with respect to the experimental result. We also obtained the VSFG spectra using the thermostated ring polymer molecular dynamics (TRPMD) method which is known to deliver more accurate frequencies than CMD at 300 K since it does not suffer from the curvature problem.

The combination of two different types of ML algorithms, namely, HDNNPs and SA-GPR, has allowed us to describe the water/air interface and all the components (polarizations) of second-order response from first principles at an affordable computational cost. The ML models were trained in a general fashion, without employing specific information on the nature of the system, such as physical constraints designed for water, mapping models for water,
based on an available surrogate model.\textsuperscript{57} While at the moment force fields tailored to describe water outperform the presented results,\textsuperscript{12,28,44} we stress that the procedure and strategy presented are directly applicable to any reference data set, to larger systems, and most importantly, to more complex and reactive interfaces.

Several other ML approaches with varying architectures have been reported in the last five years to predict molecular dipole moments\textsuperscript{58–61} and linear vibrational spectroscopy.\textsuperscript{62–64} However, attempts to simulate VSFG spectroscopy have been lagging behind, most likely due to the destructive interference of the signal, which made those approaches inappropriate for this target. While the strategy presented here could be easily applied to other kernel regression models that use local representations, it remains to be seen if similar ideas could be implemented on neural network architectures.

We evaluated the impact of the different components that are involved in the calculation of the VSFG spectra. By comparing the performance of different XC functionals, we found that the PES is best described by the revPBE-D3 and revPBE0-D3 XC functionals. These results add to the existing evidence confirming that XC functionals constrained by exact functional conditions with suitable dispersion corrections, such as revPBE-D3 and revPBE0-D3, deliver excellent performance in the description of water.\textsuperscript{96} Since the discrepancies between the theoretical and the experimental spectra are larger than the training error of the SA-GPR models, we attribute the largest source of error to the reference data used to describe the P and \( \alpha \) surfaces. A further study using electronic structure methods beyond DFT to obtain more accurate P and \( \alpha \) surfaces is urgently needed. We note that the modeling of the VSFG spectra described by eq 1 is derived in the electric dipole approximation.\textsuperscript{97} We believe that applying the current methodology to a more accurate reference data set would finally resolve existing controversies related to, for example, the relevance of quadrupole contributions in the water bending mode.\textsuperscript{65,66}

We showed that the off-diagonal components of the \( \alpha \) tensor for water are more difficult to learn due to their smaller magnitude when compared to the diagonal ones, which leads to a poorer description of certain non-ring matrix elements. Since spherical components mix together diagonal and off-diagonal elements, it is not possible to train the off-diagonal elements exclusively with current SA-GPR implementations. Future efforts will be directed in this direction, in the development of machine learning models to predict Wannier centers,\textsuperscript{67} atomic polar tensors,\textsuperscript{68} and in the inclusion of explicit LR in the description of the PES.

Two different methods were considered to describe the time evolution of the nuclei: MD and CMD. CMD is appropriate to describe the D\textsubscript{2}O/air interface and predicts a peak position that is in good agreement with experiments. However, the curvilinear motion of dangling O–D in D\textsubscript{2}O shows an extremely broad feature,\textsuperscript{47,69} which calls for the use of other approaches which do not suffer from the curvature problem of CMD nor the broadening problem of TRPMD. The recent method proposed by Musil et al. seems a promising approach to tackle this issue.\textsuperscript{70} We note, however, that none of these can describe accurately the Fermi resonance contributions.

In summary, the work presented here sets a new standard for atomistic simulation of nonlinear spectroscopies of condensed phases. While we have focused on the second-order response which is related to VSFG spectroscopy, the approach presented here paves the way to \textit{ab initio} simulations of 2D-VSFG, 2D-IR, 2D THz-Raman, and 2D ThZ-IR-Vis spectroscopies.\textsuperscript{71–73} We expect that applications of the presented strategy would bring important atomistic insights into the properties of aqueous interfaces at metallic and biological surfaces\textsuperscript{76} in solution and under confinement.\textsuperscript{77}

\section*{ASSOCIATED CONTENT}

\subsection*{Supporting Information}

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcl.3c01989.

- Detailed description of the VSFG calculations, learning curves, error analysis, and further validation tests (PDF)
- Transparent Peer Review report available (PDF)

\section*{AUTHOR INFORMATION}

\subsection*{Corresponding Authors}

Yair Litman \textmd{—} Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K.; Max Planck Institute for Polymer Research, SS128 Mainz, Germany; \textmd{orcid.org/0000-0002-6890-4052}; Email: yl899@cam.ac.uk

David M. Wilkins \textmd{—} Centre for Quantum Materials and Technologies School of Mathematics and Physics, Queen’s University Belfast, Belfast BT7 1NN Northern Ireland, United Kingdom; \textmd{orcid.org/0000-0003-3739-5512}; Email: d.wilkins@qub.ac.uk

\subsection*{Authors}

Jinggang Lan \textmd{—} Department of Chemistry, New York University, New York, New York 10003, United States; Simons Center for Computational Physical Chemistry at New York University, New York, New York 10003, United States; \textmd{orcid.org/0000-0001-6353-2539}

Yuki Nagata \textmd{—} Max Planck Institute for Polymer Research, SS128 Mainz, Germany; \textmd{orcid.org/0000-0001-9727-6641}

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcl.3c01989

\subsection*{Notes}

The authors declare no competing financial interest.

\section*{ACKNOWLEDGMENTS}

Y.L. has been partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 467724959. D.M.W. thanks Queen’s University Belfast for startup funding. J.L. thanks Simons Foundation Postdoctoral Fellowship. We thank Dr. Venkat Kapil for the useful discussion. We acknowledge the computing resources from the Swiss National Supercomputing Centre (CSCS) under project ID S1112 and S1113. Y.L. and Y.N. thank Mischa Bonn and Kuo-Yang Chiang for stimulating discussion and sharing the experimental data, and gratefully acknowledge the support from the MaxWater initiative from the Max Planck Society.

\section*{REFERENCES}

(1) Magnussen, O. M.; Groß, A. Toward an atomic-scale understanding of electrochemical interface structure and dynamics. \textit{J. Am. Chem. Soc.} 2019, \textit{141}, 4777–4790.
(2) Jubb, A. M.; Hua, W.; Allen, H. C. Environmental chemistry at vapor/water interfaces: insights from vibrational sum frequency generation spectroscopy. Annu. Rev. Phys. Chem. 2012, 63, 107−130.

(3) Schott, J.; Pokrovsky, O. S.; Oelkers, E. H. The link between mineral dissolution/precipitation kinetics and solution chemistry. Rev. Mineral. Geochem. 2009, 70, 207−258.

(4) Kraack, J. P.; Hamm, P. Surface-sensitive and surface-specific ultrafast two-dimensional vibrational spectroscopy. Chem. Rev. 2017, 117, 10623−10664.

(5) Morita, A. Theory of sum frequency generation spectroscopy; Springer Nature Singapore: Singapore, 2018; pp 151−200.

(6) Hore, D. K.; Beaman, D. K.; Parks, D. H.; Richmond, G. L. Whole-molecule approach for determining orientation at isotropic surfaces by nonlinear vibrational spectroscopy. J. Phys. Chem. B 2005, 109, 16846−16851.

(7) Gan, W.; Wu, D.; Zhang, Z.; Feng, R.-R.; Wang, H.-F. Polarization and experimental configuration analyses of sum frequency generation vibrational spectra, structure, and orientational motion of the air/water interface. J. Chem. Phys. 2006, 124, 114705.

(8) Zhang, X.; Miranda, P. B.; Kim, D.; Shen, Y. R. Mapping molecular orientation and conformation at interfaces by surface nonlinear optics. Phys. Rev. B 1999, 59, 12632−12640.

(9) Yu, C.-C.; Seki, T.; Wang, Y.; Bonn, M.; Nagata, Y. Polarization-dependent sum-frequency generation spectroscopy for ångström-scale depth profiling of molecules at interfaces. Phys. Rev. Lett. 2012, 128, 226001.

(10) Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Velarde, L.; Hammes-Schiffer, S.; Yan, E. C. Y. Detecting the first hydration shell structure around biomolecules at interfaces. ACS Cent. Sci. 2022, 8, 1404−1414.

(11) Morita, A.; Hynes, J. T. A theoretical analysis of the sum frequency generation spectrum of the water surface. ii. time-dependent approach. J. Phys. Chem. B 2002, 106, 673−685.

(12) Moberg, D. R.; Straight, S. C.; Paesani, F. Temperature dependence of the air/water interface revealed by polarization sensitive sum-frequency generation spectroscopy. J. Phys. Chem. B 2018, 122, 4356−4365.

(13) Nagata, Y.; Hsieh, C.-S.; Hasegawa, T.; Voll, J.; Backus, E. H. G.; Bonn, M. Water bending mode at the water−vapor interface probed by sum-frequency generation spectroscopy: a combined molecular dynamics simulation and experimental study. J. Phys. Chem. Lett. 2013, 4, 1872−1877.

(14) Ohto, T.; Usui, K.; Hasegawa, T.; Bonn, M.; Nagata, Y. Toward ab initio molecular modeling for sum-frequency generation spectra; an efficient algorithm based on surface-specific velocity−velocity correlation function. J. Chem. Phys. 2015, 143, 124702.

(15) Khatib, R.; Sulpizi, M. Sum frequency generation spectra from velocity−velocity correlation functions. J. Phys. Chem. Lett. 2017, 8, 1310−1314.

(16) Tian, C.-S.; Shen, Y. R. Isotopic dilution study of the water/vapor interface by phase-sensitive sum-frequency vibrational spectroscopy, J. Am. Chem. Soc. 2009, 131, 2790−2791.

(17) Nihonyanagi, S.; Ishiyama, T.; Lee, T.-k.; Yamaguchi, S.; Bonn, M.; Morita, A.; Tahara, T. Unified molecular view of the air/water interface based on experimental and theoretical χ(2) spectra of an isotopically diluted water surface. J. Am. Chem. Soc. 2011, 133, 16875−16880.

(18) Yamaguchi, S.; Takayama, T.; Goto, Y.; Otosu, T.; Yagasaki, T. Experimental and theoretical heterodyne-Detected sum frequency generation spectroscopy of isotopically pure and diluted water surfaces. J. Phys. Chem. Lett. 2022, 13, 9649−9653.

(19) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys. Rev. Lett. 2007, 98, 146401.

(20) Grisaß, A.; Wilkins, D. M.; Csányi, G.; Ceriotti, M. Symmetry-adapted machine learning for tensorial properties of atomic systems. Phys. Rev. Lett. 2018, 120, 036002.
dynamics simulation methods for vibrational spectra in the condensed phase. J. Chem. Phys. 2020, 153, 104105.

(45) Schran, C.; Brezina, K.; Marsalek, O. Committee neural network potentials control generalization errors and enable active learning. J. Chem. Phys. 2020, 153, 104105.

(46) Littman, Y.; Chiang, K.-Y.; Seki, T.; Nagata, Y.; Bonn, M. The surface of electrolyte solutions is stratified. arXiv 2022, 2210.01527.

(47) Medders, G. R.; Paesani, F. Dissecting the molecular structure of the air/water interface from quantum simulations of the sum-frequency generation spectrum. J. Am. Chem. Soc. 2016, 138, 3912−3919.

(48) Habershon, S.; Fanourgakis, G. S.; Manolopoulos, D. E. Comparison of path integral molecular dynamics methods for the infrared absorption spectrum of liquid water. J. Chem. Phys. 2008, 129, 074501.

(49) Rossi, M.; Liu, H.; Paesani, F.; Bowman, J.; Ceriotti, M. Communication: On the consistency of approximate quantum charge transfer into a point-charge model for water using machine learning. Eur. Phys. J. B 2021, 94, 181101.

(50) Feynman, R. P.; Hibbs, A. R. Quantum mechanics and path integrals; McGraw-Hill: New York, 1965.

(51) Rossi, M.; Ceriotti, M.; Manolopoulos, D. E. How to remove the spurious resonances from ring polymer molecular dynamics. J. Chem. Phys. 2014, 140, 234116.

(52) Ivanov, S. D.; Witt, A.; Shiga, M.; Marx, D. Communications: on artificial frequency shifts in infrared spectra obtained from centroid model dynamics. J. Chem. Phys. 2010, 132, 031101.

(53) Althorpe, S. C. Path-integral approximations to quantum dynamics. Eur. Phys. J. B 2021, 94, 155.

(54) Treins, G.; Willatt, M. J.; Althorpe, S. C. Path-integral dynamics of water using curvilinear centroids. J. Chem. Phys. 2019, 151, 054109.

(55) Rossi, M.; Kapil, V.; Ceriotti, M. Fine tuning classical and quantum molecular dynamics using a generalized Langevin equation. J. Chem. Phys. 2018, 148, 102301.

(56) Ni, Y.; Skinner, J. L. IR and SFG vibrational spectroscopy of the water bend in the bulk liquid and at the liquid-vapor interface, respectively. J. Chem. Phys. 2015, 143, 014502.

(57) Qi, C.; Yu, Q.; Conte, R.; Houston, P. L.; Nandi, A.; Bomwan, J. M. A delta-machine learning approach for force fields, illustrated by a CCSD(T) 4-body correction to the MB-pol water potential. Digital Discovery 2022, 1, 658−664.

(58) Veit, M.; Wilkins, D. M.; Yang, Y.; DiStasio, R. A., Jr.; Ceriotti, M. Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles. J. Chem. Phys. 2020, 153, 024113.

(59) Stach, C. G.; Wengert, S.; Kunkel, C.; Csányi, G.; Reuter, K.; Margraf, J. T. Kernel charge equilibration: efficient and accurate prediction of molecular dipole moments with a machine-learning enhanced electron density model. Mach. Learn.: Sci. Technol. 2022, 3, 015032.

(60) Christensen, A. S.; Faber, F. A.; von Lilienfeld, O. A. Operators in quantum machine learning: Response properties in chemical space. J. Chem. Phys. 2019, 150, 064105.

(61) Gao, A.; Remsing, R. C. Self-consistent determination of long-range electrostatics in neural network potentials. Nat. Commun. 2022, 13, 1572.

(62) Gasteiger, M.; Behler, J.; Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. Chem. Sci. 2017, 8, 6924−6935.

(63) Han, B.; Ishborn, C. M.; Shi, L. Incorporating polarization and charge transfer into a point-charge model for water using machine learning. J. Phys. Chem. Lett. 2023, 14, 3869−3877.

(64) Sommers, G. M.; Calegari Andrade, M. F.; Zhang, L.; Wang, H.; Car, R. Raman spectrum and polarizability of liquid water from deep neural networks. Phys. Chem. Chem. Phys. 2020, 22, 10592−10602.

(65) Kondu, A.; Tanaka, S.; Ishiyama, T.; Ahmed, M.; Inoue, K.-i.; Nihonyanagi, S.; Sawai, H.; Yamaguchi, S.; Morita, A.; Tahara, T. Bend vibration of surface water investigated by heterodyne-detected sum frequency generation and theoretical study: dominant role of quadrupole. J. Phys. Chem. Lett. 2016, 7, 2597−2601.

(66) Seki, T.; Chiang, K.-Y.; Yu, C.-C.; Yu, X.; Okuno, M.; Hunger, J.; Nagata, Y.; Bonn, M. The bending mode of water: a powerful probe for hydrogen bond structure of aqueous systems. J. Phys. Chem. Lett. 2020, 11, 8459−8469.

(67) Zhang, L.; Chen, M.; Wu, X.; Wang, H.; E, W.; Car, R. Deep neural network for the dielectric response of insulators. Phys. Rev. B 2020, 102, 041211.

(68) Schienbein, P. Spectroscopy from machine learning by accurately representing the atomic polar tensor. J. Chem. Theory Comput. 2023, 19, 705−712.

(69) Kaliannan, N. K.; Aristizabal, A. H.; Wieseler, H.; Zysk, F.; Ohto, T.; Nagata, Y.; Kühne, T. D. Impact of intermolecular vibrational coupling effects on the sum-frequency generation spectra of the water/air interface. Mol. Phys. 2020, 118, 1620358.

(70) Musil, F.; Zaporozhets, I.; Noé, F.; Clementi, C.; Kapil, V. Quantum dynamics using path integral coarse-graining. J. Chem. Phys. 2022, 157, 181102.

(71) Benson, R. L.; Althorpe, S. C. On the “Matsubara heating” of overtone intensities and Fermi splittings. J. Chem. Phys. 2021, 155, 104107.

(72) Plé, T.; Huppert, S.; Finocchi, F.; Depondt, P.; Bonella, S. Anharmonic spectral features via trajectory-based quantum dynamics: A perturbative analysis of the interplay between dynamics and sampling. J. Chem. Phys. 2021, 155, 104108.

(73) Perakis, F.; De Marco, L.; Shalit, A.; Tang, F.; Kann, Z. R.; Kühne, T. D.; Torres, R.; Bonn, M.; Nagata, Y. Vibrational spectroscopy and dynamics of water. Chem. Rev. 2016, 116, 7590−7607.

(74) Savolainen, J.; Ahmed, S.; Hamm, P. Two-dimensional Ramanseterhertz spectroscopy of water. Proc. Natl. Acad. Sci. U. S. A. 2013, 110, 20402−20407.

(75) Greeko, M.; Hasegawa, T.; D’Angelo, F.; Ito, H.; Turchinovich, D.; Nagata, Y.; Bonn, M. Coupling between intramolecular and intermolecular motions in liquid water revealed by two-dimensional terahertz-infrared-visible spectroscopy. Nat. Commun. 2018, 9, 885.

(76) Gonella, G.; Backus, E. H. G.; Nagata, Y.; Bonthuis, D. J.; Loche, P.; Schlaich, A.; Netz, R. R.; Kühne, A.; McCrum, I. T.; Koper, M. T. M.; et al. Water at charged interfaces. Nat. Rev. Chem. 2021, 5, 466−485.

(77) Kapil, V.; Schran, C.; Zen, A.; Chen, J.; Pickard, C. J.; Michelides, A. The first-principles phase diagram of monolayer nanoconfined water. Nature 2022, 609, 512−516.