Toward a Simple Molecular Understanding of Sum Frequency Generation at Air−Water Interfaces†

Joyce Noah-Vanhoucke,† Jared D. Smith,§ and Phillip L. Geissler*,†,§

Department of Chemistry, University of California, Berkeley, California 94720, and Chemical Sciences Division Lawrence Berkeley National Laboratory, Berkeley, California 94720

Received: July 5, 2008; Revised Manuscript Received: January 13, 2009

Second-order vibrational spectroscopies successfully isolate signals from interfaces, but they report on intermolecular structure in a complicated and indirect way. Here, we adapt a perspective on vibrational response developed for bulk spectroscopies to explore the microscopic fluctuations to which sum frequency generation (SFG), a popular surface-specific measurement, is most sensitive. We focus exclusively on inhomogeneous broadening of spectral susceptibilities for OH stretching of HOD as a dilute solute in D$_2$O. Exploiting a simple connection between vibrational frequency shifts and an electric field variable, we identify several functions of molecular orientation whose averages govern SFG. The frequency dependence of these quantities is well captured by a pair of averages, involving alignment of OH and OD bonds with the surface normal at corresponding values of the electric field. The approximate form we obtain for SFG susceptibility highlights a dramatic sensitivity to the way a simulated liquid slab is partitioned for calculating second-order response.

I. Introduction

Molecular organization in heterogeneous environments, and in particular near interfaces, can differ substantially from that in homogeneous bulk materials. As an important example, the microscopic structure of many metals undergoes reconstruction at surfaces, with dramatic consequences for catalysis. The nature and extent of intermolecular rearrangement at liquid interfaces, while no less intriguing, are poorly understood. Of particular interest to biological, chemical, and environmental sciences are interfaces involving water and salt solutions.

Understanding molecular structure at solid−vapor interfaces has been greatly facilitated by surface-specific, high-resolution vibrational spectroscopy, which in those systems can resolve distinct microscopic arrangements. While it is unrealistic to expect well-separated spectral signatures of diverse intermolecular arrangements in a dense fluid, the adaptation of such tools to study liquid interfaces promises to provide substantial new insight into their chemistry and physics. Sum frequency generation spectroscopy (SFG), in particular, has emerged as an important tool for the investigation of the aqueous liquid−vapor interface at the molecular level. SFG spectroscopy is a second-order optical process, and therefore forbidden in bulk centrosymmetric media. Thus, the SFG response, which is proportional to the square of the second-order susceptibility $\chi^{(2)}$, results entirely from the interface where symmetry is necessarily broken. Unfortunately, this complex orientationally weighted vibrational spectrum does not admit a straightforward interpretation of the interfacial structure.

Despite the ambiguous link between the SFG spectrum and interfacial structure, a number of groups have made detailed assignments based on the SFG spectrum of pure water. For instance, Raymond et al. and Liu et al. have both decomposed the SFG spectrum into multiple hydrogen bonding configurations. However, recent theoretical and experimental findings have indicated that such a multistate interpretation is inappropriate even for the much simpler spectral response of bulk liquid water. Computer simulations point to a more continuous view of intermolecular arrangements at the interface. Spectroscopic formalism developed by Morita and Hynes has enabled estimates of SFG response from such simulations. However, the complexity of these calculations, together with uncertainties in several important input parameters (such as nonresonant contributions to SFG susceptibility, transition dipoles and polarizabilities, and the strength of couplings between vibrational modes), make it difficult to draw firm conclusions from comparison with experiment.

In this work, we present a new approach for calculating resonant SFG susceptibility. Our purpose is not to generate more accurate predictions of spectroscopic response, though the methods we use have proved quantitatively accurate in the case of bulk Raman spectra. Instead, we aim to clarify how SFG reveals specific aspects of intermolecular structure, with a focus on hydrogen bonding arrangements at the air−water interface. In section II, we introduce a series of physically motivated approximations that yield from the formalism of Morita and Hynes a set of time-independent orientational averages governing SFG response. These simplifications are reminiscent of our earlier work on bulk vibrational spectroscopy, which emphasized a simple connection between shifts in hydroxyl stretching frequency and fluctuations of a specific component of the liquid’s electric field. We further show that the frequency dependence of relevant orientational averages is well-characterized by a small number of readily interpreted quantities.

We explore the implications of our theoretical results using computer simulations of a molecular model for liquid water in coexistence with its vapor phase. Our numerical methods are described in detail in section III, while practical issues of symmetry breaking that cannot be avoided in the computation of SFG signals are discussed in section II. We report simulation results in section V of a dilute solution of HOD in D$_2$O. These calculations illustrate the benefits of simple perspectives described in section II. However, they also highlight the crucial
importance of resolving problematic symmetry-breaking issues, as well as the need for accurate computation of transition dipoles and polarizabilities that determine the strengths and signs of the contributions we have identified. In section VI, we conclude with a discussion of outstanding theoretical problems and the prospects for assigning structural motifs to features of measured SFG frequency dependence.

II. A Simplified View of SFG Susceptibility

Our analysis of SFG focuses on an interface’s susceptibility, \( \chi(\omega) \), which is primarily responsible for the IR frequency dependence of observed signals.\(^{23} \) More specifically, we consider the imaginary part of the resonant contribution, \( \text{Im} \chi(\omega) \), whose connection to absorptive response offers the most straightforward link with intermolecular structure. In a significant advance, Shen and co-workers have recently demonstrated direct observation of this quantity.\(^{24} \) By contrast, conventional experiments measure only a squared SFG intensity that combines \( \text{Im} \chi(\omega) \) with its real counterpart, as well as a nonresonant contribution. These latter components obscure the frequency dependence of \( \text{Im} \chi(\omega) \) without providing additional useful information about variations in intermolecular structure across the spectrum: The real part of the resonant contribution can be obtained from \( \text{Im} \chi(\omega) \) via Kramers–König relations, while the nonresonant contribution is nearly frequency independent.\(^{10,21,25,26} \)

In this paper, we consider the most common choice of light polarizations, ssp, corresponding to a single element of the susceptibility tensor, \( \chi_{xxz}(\omega) \). We further restrict our attention to a dilute isotopic mixture of liquid water. (Generalizing our results for different polarizations should be straightforward.) As exploited in spectroscopic studies of bulk water, great simplification can be achieved by mixing H\(_2\)O with a large excess of D\(_2\)O, yielding HOD as the dominant solute species. In the case of dilute isotopic mixtures, stretching of an OH bond can be reasonably considered a normal mode of vibration, uncoupled to lower frequency OD stretching motions in its environment. Despite this simplification, few studies have been performed for such mixtures at the interface, and none have probed very dilute mixtures in detail. As will become apparent in our analysis, even intramolecular vibrational coupling can seriously threaten a simple microscopic interpretation of SFG.

Morita and Hynes have derived approximate formulas for the resonant susceptibility that may be evaluated by averaging functions of molecular orientation over an ensemble of thermal fluctuations.\(^{16,17} \) These results, based on the simple connection between SFG susceptibility and average molecular hyperpolarizability, \( \chi \propto \beta \), provide a straightforward numerical prescription for estimating SFG signals using molecular simulations. However, they do not reveal in a transparent way how SFG reports for estimating SFG signals using molecular simulations. How-

\[ \rho_{\text{res}}(\omega) = \int_0^\infty \text{d}t \, e^{i\omega t} \phi(t) \]  
\[ \phi(t) = \langle \alpha_t(\mu) \rangle \]  

Angular brackets denote unrestricted averaging over a canonical distribution of microscopic degrees of freedom, including both the OH vibrational coordinate \( R \) and the solvent configuration \( \Gamma \). We regard \( \mu \) and \( \alpha \) as quantum mechanical operators acting on \( R \), rather than as classical variables as in refs 17, 18, 27, and 28.

The Hamiltonian governing dynamics of the vibrational wave function acquires explicit time dependence from interactions with the fluctuating solvent environment. We consider the solvent’s influence as “pure dephasing”, neglecting vibrational relaxation induced by coupling between the vibrational “system” and the solvent “bath”. We instead focus on solvent-induced modulation of the energy gap \( \hbar\omega_{\text{dip}}(t) \) between ground \( |0\rangle \) and first excited \( |1\rangle \) vibrational states. Ignoring the very low probability of excitations to higher-lying vibrational states, and invoking the Condon approximation, we obtain

\[ \phi(t) = \sum_{j,k,l} \alpha_{jk}^* \mu_k^* \langle j'(k')(l') | 1 | 0 \rangle \exp \left[ -i \int_0^t \text{d}\tau \, \omega_{\text{dip}}(\tau) \right] \]  

Here, \( j_\perp \) and \( k_\perp \) are projections of the Cartesian unit vectors \( \mathbf{j} \), \( \mathbf{k} \), and \( \mathbf{l} \) of the molecular reference frame onto the \( x \)- and \( z \)-axes of the laboratory frame. Figure 1 depicts our choice of coordinate system, which includes the unit vector \( \mathbf{u} \) pointing along the OH bond as a principal direction. The sums of eq 3 thus run over vectors \( \mathbf{u} \), \( \mathbf{v} \) (also lying in the plane of the molecule), and \( \mathbf{w} \equiv \mathbf{u} \times \mathbf{v} \). The coefficients \( \alpha_{jk}^* \) and \( \mu_k^* \) are the transition matrix elements of the polarizability and dipole derivatives with respect to the vibrational coordinate.

The correlation function \( \phi(t) \) in eq 3 closely resembles those describing vibrational dephasing in the contexts of bulk Raman and infrared spectroscopies.\(^{29,30} \) Differing only in the appearance of molecular orientations relative to input light polarization, it could be evaluated using the extensive body of methods developed for computer simulations of those more straightforward probes.\(^{11,29,31,32} \) Independent of the work described in this paper, Skinner and co-workers have performed such a detailed calculation.\(^{33} \) Here, we pursue a simplification of eq 3 that is similar in spirit to those of refs 11 and 22. Our primary approximations from this point are twofold: First, we imagine that spectral line shapes are dominated by inhomogeneous broadening and only weakly manifest effects of motional narrowing. Second, we assert a linear relationship between instantaneous OH vibrational frequency and electrostatic forces experienced by the proton.

Neglect of dynamical broadening is strictly justified only when relaxation of a molecule’s environment, as well as its own

\[ \text{infinite slow dephasing, } \gamma \rightarrow 0, \text{ at which point the two approaches become equivalent. Before doing so, we will cast} \]  

the calculation of SFG susceptibility in a semi-classical form reminiscent of Kubo’s line shape theory for simple absorptive processes.

Following the perturbative treatment of radiation–matter interactions due to Morita and Hynes, we begin by writing the hyperpolarizability as the one-sided Fourier transform of a correlation function involving the molecular dipole \( \mu \) at time zero and the molecular polarizability \( \alpha(t) \) at a later time \( t \):

\[ \rho_{\text{res}}(\omega) = \int_0^\infty \text{d}t \, e^{i\omega t} \phi(t) \]  
\[ \phi(t) = \langle \alpha_t(\mu) \rangle \]  

Angular brackets denote unrestricted averaging over a canonical distribution of microscopic degrees of freedom, including both the OH vibrational coordinate \( R \) and the solvent configuration \( \Gamma \). We regard \( \mu \) and \( \alpha \) as quantum mechanical operators acting on \( R \), rather than as classical variables as in refs 17, 18, 27, and 28.

The Hamiltonian governing dynamics of the vibrational wave function acquires explicit time dependence from interactions with the fluctuating solvent environment. We consider the solvent’s influence as “pure dephasing”, neglecting vibrational relaxation induced by coupling between the vibrational “system” and the solvent “bath”. We instead focus on solvent-induced modulation of the energy gap \( \hbar\omega_{\text{dip}}(t) \) between ground \( |0\rangle \) and first excited \( |1\rangle \) vibrational states. Ignoring the very low probability of excitations to higher-lying vibrational states, and invoking the Condon approximation, we obtain

\[ \phi(t) = \sum_{j,k,l} \alpha_{jk}^* \mu_k^* \langle j'(k')(l') | 1 | 0 \rangle \exp \left[ -i \int_0^t \text{d}\tau \, \omega_{\text{dip}}(\tau) \right] \]  

Here, \( j_\perp \), \( k_\perp \), and \( l_\perp \) are projections of the Cartesian unit vectors \( \mathbf{j} \), \( \mathbf{k} \), and \( \mathbf{l} \) of the molecular reference frame onto the \( x \)- and \( z \)-axes of the laboratory frame. Figure 1 depicts our choice of coordinate system, which includes the unit vector \( \mathbf{u} \) pointing along the OH bond as a principal direction. The sums of eq 3 thus run over vectors \( \mathbf{u} \), \( \mathbf{v} \) (also lying in the plane of the molecule), and \( \mathbf{w} \equiv \mathbf{u} \times \mathbf{v} \). The coefficients \( \alpha_{jk}^* \) and \( \mu_k^* \) are the transition matrix elements of the polarizability and dipole derivatives with respect to the vibrational coordinate.

The correlation function \( \phi(t) \) in eq 3 closely resembles those describing vibrational dephasing in the contexts of bulk Raman and infrared spectroscopies.\(^{29,30} \) Differing only in the appearance of molecular orientations relative to input light polarization, it could be evaluated using the extensive body of methods developed for computer simulations of those more straightforward probes.\(^{11,29,31,32} \) Independent of the work described in this paper, Skinner and co-workers have performed such a detailed calculation.\(^{33} \) Here, we pursue a simplification of eq 3 that is similar in spirit to those of refs 11 and 22. Our primary approximations from this point are twofold: First, we imagine that spectral line shapes are dominated by inhomogeneous broadening and only weakly manifest effects of motional narrowing. Second, we assert a linear relationship between instantaneous OH vibrational frequency and electrostatic forces experienced by the proton.

Neglect of dynamical broadening is strictly justified only when relaxation of a molecule’s environment, as well as its own
bath coupling \( Q = 160.514 \text{ cm}^{-1} \text{Å/V} \) that provide excellent correspondence between the probability distribution \( P(\varsigma) = \langle \delta(\varsigma - \varsigma(\Gamma)) \rangle \) of electric field strength and the Raman spectrum of a dilute HOD/D_{2}O solution. (As before, the variable \( \Gamma \) represents the solvent configuration.) The quantity \( Q \) is a property of a single water molecule, which, to first approximation, is independent of its environment. For the interfacial system under consideration, small changes in the magnitude of \( Q \) in the outer layer of the liquid would have little influence on our results. In the context of SFG, eq 5 allows us to isolate the spectral consequences of symmetry breaking at an interface. Combining eqs 2, 4, and 5, we arrive at our central theoretical result

\[
\text{Im} \beta_{\text{exc}}^{(R)}(\omega) = \frac{\pi}{Q} P(\varsigma) \sum_{j,k,l} \alpha_j^\mu \langle j,k,l,\varsigma \rangle
\]

(6)

where \( P(\varsigma) \) is the distribution of electric field strengths for the entire sample, a distribution dominated by the bulk contributions, as discussed below. Aside from multiplicative constants, the hyperpolarizability within our approximations is just the product of \( P(\varsigma) \) and a linear combination of orientational terms

\[
\langle j,k,l,\varsigma \rangle = \frac{\langle j,k,l,\varsigma - \varsigma(\Gamma) \rangle}{\langle \delta(\varsigma - \varsigma(\Gamma)) \rangle}
\]

(7)

These conditional averages quantify the correlation between a molecule’s hydrogen bonding environment, as characterized by \( \varsigma \), and its orientational bias due to the interface. They provide a concrete basis for rationalizing features of spectra computed according to eq 2 in terms of intermolecular structure.

Note that, in the thermodynamic limit, the electric field distribution \( P(\varsigma) \) in eq 6 is not altered by the interface. Generally speaking, the physical influence of an interface is limited to a microscopic depth \( \Delta_m \) below the surface.\(^{52}\) Many studies have concluded that for liquid water near ambient conditions \( \Delta_m \sim 1 \text{ nm} \) is not much larger than a molecular diameter.\(^{7,14}\) The thickness \( D \) of a typical experimental sample, on the other hand, is macroscopic. The fraction of molecules contributing to \( P(\varsigma) \) whose statistics differ from those of the bulk interior is thus extremely small. Mathematically, we could write \( P(\varsigma) \) as the sum of its bulk counterpart and a perturbation of order \( \epsilon = \Delta_m/D \ll 1 \)

\[
P(\varsigma) = P_{\text{bulk}}(\varsigma) + \epsilon g_m(\varsigma)
\]

(8)

The function \( g_m(\varsigma) \), which quantifies changes in electric field statistics near the interface, is comparable in magnitude to \( P_{\text{bulk}}(\varsigma) \). By contrast, the bulk contribution to orientational averages appearing in eq 6 vanish due to symmetry regardless of \( \varsigma \), so that \( \langle j,k,l,\varsigma \rangle \sim O(\epsilon) \). Replacing \( P(\varsigma) \) by \( P_{\text{bulk}}(\varsigma) \) in our result 6 for the hyperpolarizability therefore introduces an error of only \( O(\epsilon) \), which may be neglected entirely for a macroscopic sample. As a matter of practice, interfacial systems represented in computer simulations are typically only a few nanometers thick, so that the ratio \( \epsilon \) is not necessarily negligible. Nonetheless, for simulated aqueous systems, changes in \( P(\varsigma) \) are not dramatic even in the first few monolayers of the liquid (as evidenced by computed surface Raman and IR spectra that differ little from their bulk counterparts).\(^{17}\) One remains safe in considering only bulk contributions to \( P(\varsigma) \) for the purpose of evaluating \( \text{Im} \beta_{\text{exc}}^{(R)}(\omega) \).
There are 18 choices of the set \((j, k, l)\) that yield distinct averages from eq 7. Several of these averages vanish identically due to symmetry: Because the interfacial system is statistically achiral, any microscopic configuration \(\Gamma\) can be reflected across the \(xz\)- or \(yz\)-planes to yield a mirror image microstate \(\tilde{\Gamma}\) with equal Boltzmann weight. In the case of \(yz\)-plane reflection, the \(x\)-component of OH and OD bond vectors changes sign. Linear combinations of these bond orientations determine basis vectors \(\tilde{u}\) and \(\tilde{v}\) of the molecular frame, which thus transform as \(\tilde{u}_i = -u_i, \tilde{v}_i = -v_i;\) all other components of \(\tilde{u}\) and \(\tilde{v}\) are unaffected by \(yz\)-reflection. As the vector product of the first two basis vectors, \(\mathbf{w}\) transforms differently: \(\tilde{w}_i = w_i, \tilde{w}_i = -w_i,\) and \(\tilde{w}_i = -w_i.\) As a result, products such as \(u_i w_j w_k\) and \(w_i u_j u_k\) yield zero upon averaging, regardless of electric field. Table 1 lists the 10 distinct orientational averages that survive this elimination by symmetry.

The entries in Table 1 are ordered by the magnitude of the coefficients that multiply them in eq 6, as determined by electronic structure calculations for an isolated water molecule. Two of the averages that survive symmetry operations on the system as a whole are multiplied in eq 6 by coefficients that vanish due to internal symmetry of the molecule. For any typical configuration of the liquid, this latter symmetry will be broken due to interactions with surrounding molecules. High-level electronic structure calculations are needed to establish whether \(\langle u_i w_j w_k \rangle\) and \(\langle w_i u_j u_k \rangle\) in fact remain negligible.

Only two of a molecule’s three basis vectors \(\mathbf{u}, \mathbf{v},\) and \(\mathbf{w}\) can be considered independent (since one can always be defined as a vector product of the other two). It may therefore not be surprising that the 10 nonzero orientational averages carry a significant amount of redundant structural information.

For example, we will show in section V that significant amount of redundant structural information. For any typical configuration of the liquid, this latter symmetry will be broken due to interactions with surrounding molecules. High-level electronic structure calculations are needed to establish whether \(\langle u_i w_j w_k \rangle\) and \(\langle w_i u_j u_k \rangle\) in fact remain negligible.

Only two of a molecule’s three basis vectors \(\mathbf{u}, \mathbf{v},\) and \(\mathbf{w}\) can be considered independent (since one can always be defined as a vector product of the other two). It may therefore not be surprising that the 10 nonzero orientational averages carry a significant amount of redundant structural information.

For example, we will show in section V that significant amount of redundant structural information. For any typical configuration of the liquid, this latter symmetry will be broken due to interactions with surrounding molecules. High-level electronic structure calculations are needed to establish whether \(\langle u_i w_j w_k \rangle\) and \(\langle w_i u_j u_k \rangle\) in fact remain negligible.

Only two of a molecule’s three basis vectors \(\mathbf{u}, \mathbf{v},\) and \(\mathbf{w}\) can be considered independent (since one can always be defined as a vector product of the other two). It may therefore not be surprising that the 10 nonzero orientational averages carry a significant amount of redundant structural information.

For example, we will show in section V that significant amount of redundant structural information. For any typical configuration of the liquid, this latter symmetry will be broken due to interactions with surrounding molecules. High-level electronic structure calculations are needed to establish whether \(\langle u_i w_j w_k \rangle\) and \(\langle w_i u_j u_k \rangle\) in fact remain negligible.

Only two of a molecule’s three basis vectors \(\mathbf{u}, \mathbf{v},\) and \(\mathbf{w}\) can be considered independent (since one can always be defined as a vector product of the other two). It may therefore not be surprising that the 10 nonzero orientational averages carry a significant amount of redundant structural information.

For example, we will show in section V that significant amount of redundant structural information. For any typical configuration of the liquid, this latter symmetry will be broken due to interactions with surrounding molecules. High-level electronic structure calculations are needed to establish whether \(\langle u_i w_j w_k \rangle\) and \(\langle w_i u_j u_k \rangle\) in fact remain negligible.

Only two of a molecule’s three basis vectors \(\mathbf{u}, \mathbf{v},\) and \(\mathbf{w}\) can be considered independent (since one can always be defined as a vector product of the other two). It may therefore not be surprising that the 10 nonzero orientational averages carry a significant amount of redundant structural information.

For example, we will show in section V that significant amount of redundant structural information. For any typical configuration of the liquid, this latter symmetry will be broken due to interactions with surrounding molecules. High-level electronic structure calculations are needed to establish whether \(\langle u_i w_j w_k \rangle\) and \(\langle w_i u_j u_k \rangle\) in fact remain negligible.

Only two of a molecule’s three basis vectors \(\mathbf{u}, \mathbf{v},\) and \(\mathbf{w}\) can be considered independent (since one can always be defined as a vector product of the other two). It may therefore not be surprising that the 10 nonzero orientational averages carry a significant amount of redundant structural information.

For example, we will show in section V that significant amount of redundant structural information. For any typical configuration of the liquid, this latter symmetry will be broken due to interactions with surrounding molecules. High-level electronic structure calculations are needed to establish whether \(\langle u_i w_j w_k \rangle\) and \(\langle w_i u_j u_k \rangle\) in fact remain negligible.
understanding how SFG measurements reflect on microscopic structure at the air–water interface.

III. Simulation Methods

We explore the statistics of orientation and electric field at liquid–vapor interfaces using computer simulations of a simple molecular model. For this purpose, we have employed standard techniques of molecular dynamics and Monte Carlo (MC) importance sampling. All results reported here were obtained from systems including \( N = 512 \) molecules, interacting via the SPC/E potential \(^{42}\) computed with Ewald sums, at ambient temperature and at fixed densities that enforce liquid–vapor coexistence. Since we will focus exclusively on static, classical properties, masses of the atoms are irrelevant. For spectroscopic purposes, we view the system as one HOD solute among 511 \( \text{D}_2\text{O} \) solvent molecules. For each sampled configuration of the system, we may regard any one of \( 2 \times 512 \) atoms as the deuteron. Computed averages therefore converge about 1000 times more rapidly than they would in corresponding simulations that explicitly include a single HOD solute.

Periodic boundary conditions were imposed in all three principal directions of the laboratory frame, resulting in a liquid slab oriented perpendicular to the \( z \)-axis. With simulation cell dimensions \( L_x = L_y = 20 \text{ Å} \) in the \( x \)- and \( y \)-directions, the slab adopts a thickness of roughly \( 40 \text{ Å} \). The size of the simulation cell in the \( z \)-direction, \( L_z = 100 \text{ Å} \), was thus sufficient to provide nearly \( 60 \text{ Å} \) of vapor (essentially vacuum on these scales at \( 298 \text{ K} \)) between periodically replicated slabs. Equilibration was achieved by allowing a well-equilibrated configuration from the uniform liquid phase to expand in the \( z \)-direction, followed by extensive sampling (at least \( 200 \text{ ps} \) or, in the case of MC simulations, \( 10^6 \) sweeps) before data was collected over trajectories of at least \( 2 \text{ ns} \) (or \( 10^6 \) MC sweeps). We report vertical coordinates \( \Delta z = z - z_{\text{ Gibbs}} \) of H, O, and D atoms relative to the location \( z_{\text{ Gibbs}} \) of the closest Gibbs dividing surface, i.e., the value of \( z \) where solvent density falls to half its bulk value.

Molecular dynamics simulations were performed using the LAMMPS software package.\(^{43}\) In these calculations, rigid molecular geometries were maintained using the SHAKE algorithm.\(^{44}\) Equilibration runs employed a Nose–Hoover thermostat,\(^{45}\) while production runs were propagated by integrating Newton’s equations of motion using the velocity Verlet algorithm.\(^{46}\) In both cases, \( \Delta t = 1 \text{ fs} \) served as the fundamental time step of integration.

If realized in the laboratory, the liquid samples simulated in this and most other theoretical studies of air–water interfaces would in fact exhibit no second-order spectroscopic response at all. The problem is not a lack but an excess of interfaces. Because a liquid slab possesses two statistically identical interfaces with antiparallel surface normal vectors, it does not in fact break symmetry in the way required for SFG. Computing second-order susceptibilities from such a system thus requires an artificial breaking of symmetry. The natural strategy for doing so is to place an imaginary dividing surface through the slab, parallel to the interfaces. An isolated interfacial region then comprises all molecules residing on one side of the dividing surface.\(^{54}\) Considered independently from the second notionally interfacial region, such a subsystem would yield a nonzero SFG signal. This procedure is so reasonable that many computational SFG studies do not report that it has been implemented, much less how it has been implemented.

There is no ambiguity in assigning a point particle to one side of a planar dividing surface. The internal structure of an HOD molecule, however, mandates that many molecules in a typical configuration of the liquid will straddle the dividing surface. To which interface they should be assigned is surprisingly far from an academic question. We will show in section V that the convention one uses to divide an aqueous liquid slab into two separate interfaces can dramatically influence predictions for the frequency-dependent susceptibility. We explain how such sensitivity arises and describe its relationship to experiments in the following section.

IV. Symmetry-Breaking Conventions

Consider the average of an odd function \( f(u_z) \) of molecular orientation (e.g., \( f = u_z \) or \( f = u_z^3 \)), which vanishes in a centrosymmetric bulk medium, but can be nonzero at an interface. Here, we will mimic the attenuation with depth of radiation incident on a sample by multiplying contributions to \( \langle f(u_z) \rangle \) with a decaying weight \( e^{iDz} \). (The liquid is taken to extend from \( z = -\infty \) to \( z \approx 0 \)). This factor could also be viewed as a mathematical device to break symmetry of a simulated liquid slab in a smooth way. In either case, the decay length \( L \) should be large compared to the scale \( \Delta_{\text{in}} \) over which the interface substantially influences orientational statistics but small compared to the sample thickness \( D \). Although the arguments we make here are unrelated to electric field fluctuations, we will condition the average on \( \zeta \) for application to the quantities controlling SFG:

\[
\langle f(u_z) \rangle_{\zeta} = \int_{-\infty}^{d} dz \int_{-1}^{1} du_z e^{i \zeta} p(u_z, z, \zeta) f(u_z) \quad (13)
\]

The upper limit \( d \) for \( z \)-integration refers to an elevation above the interface at which the density is sufficiently low as to strongly suppress any contribution to \( \langle f(u_z) \rangle_{\zeta} \) from the region \( z > d \). For aqueous systems, \( d = O(\Delta_{\text{in}}) \leq L \).

The joint probability \( p(u_z, z, \zeta) \) that an HOD molecule is located at vertical coordinate \( z \) and has orientation \( u_z \), given electric field \( \zeta \), requires precise specification of molecular location. Let us consider two extreme cases: the vertical coordinate \( z \) of an HOD molecule is assigned to be (1) that of its oxygen atom, \( z_{\text{O}} \), or (2) that of its hydrogen atom, \( z_{\text{H}} \). If an OH bond straddles the dividing surface distinguishing top and bottom interfacial regions of a liquid slab, say \( z_{\text{O}} > 0 \) and \( z_{\text{H}} < 0 \), then convention 1 would associate the molecule with the top region, while convention 2 would assign it to the bottom region. In this configuration, \( u_z < 0 \) from the perspective of the top region, while \( u_z > 0 \) from the perspective of the bottom region, whose surface normal points opposite to \( \zeta \). (Recall that \( u_z \) denotes the projection of an OH bond vector onto the outward surface normal.) As a result, averages such as \( \langle u_z \rangle_{\zeta} \) will be systematically biased to lower values for convention 1 and to higher values for convention 2. As we have presented them, these biases would appear to be an unphysical artifact of the sharp and arbitrary manner in which symmetry has been broken. They would also appear to be quite small in magnitude, and therefore unimportant to the qualitative nature of a predicted SFG signal. Both of these appearances are deceiving.

In an SFG measurement, the contribution of a molecule at vertical coordinate \( z \) to the observed signal is of course not decided in a sharp way according to an artificial dividing surface. Rather, it is determined by the smooth attenuation of incident radiation as it is absorbed and scattered by the sample. Because the length scale \( L \) of this attenuation greatly exceeds the OH bond length \( \approx 1 \text{ Å} \), one might reasonably expect that the distinctions described above become unimportant. To evaluate
this expectation, it is useful to resolve the statistics of orientation and electric field by vertical coordinate \( z \). For an HOD molecule, probability distributions involving \( z \) are only well-defined once we specify whether \( z = z^{(O)} \), \( z = z^{(H)} \) or otherwise. For the first convention introduced above, we would define the joint distribution of \( u \) and \( z \) given \( \zeta \) as
\[
p^{(O)}(u, z | \zeta) = \frac{\langle \delta(u - u_{(O)}(\Gamma)) \delta(z - z^{(O)}(\Gamma)) \delta(\zeta - \zeta(\Gamma)) \rangle}{\langle \delta(\zeta - \zeta(\Gamma)) \rangle} \tag{14}
\]
while for the second we would define
\[
p^{(H)}(u, z | \zeta) = \frac{\langle \delta(u - u_{(H)}(\Gamma)) \delta(z - z^{(H)}(\Gamma)) \delta(\zeta - \zeta(\Gamma)) \rangle}{\langle \delta(\zeta - \zeta(\Gamma)) \rangle} \tag{15}
\]
Since \( z^{(H)} = z^{(O)} + /u_{c} \), the two distributions are related simply by
\[
p^{(O)}(u, z | \zeta) = p^{(H)}(u, z + /u_{c} | \zeta) \tag{16}
\]
To the extent that these probabilities vary slowly with vertical coordinate, averages taken over \( p^{(O)}(u, z | \zeta) \) and \( p^{(H)}(u, z | \zeta) \) differ in general by an amount proportional to the bond length \( / \). This difference is a matter of concern only because the range \( D_{int} \) of interfacial influence is not dramatically larger than \( / \).

Assuming such probabilities vary smoothly with \( z \), we can exploit the connection 16 between different conventions to write
\[
p^{(O)}(u, z | \zeta) = p^{(H)}(u, z | \zeta) + /u_{c} \left[ \frac{\partial p^{(H)}(u, z | \zeta)}{\partial z} \right]_{u_{c}} \tag{17}
\]
Averages of \( f(u_{c}) \) calculated using the two conventions are then related by
\[
\langle f(u_{c}) \rangle^{(O)}_{\zeta} = \langle f(u_{c}) \rangle^{(H)}_{\zeta} + \\
\frac{1}{/} \int_{-\infty}^{\infty} dz \int du_{c} e^{-d_{z}^{(O)}(\zeta)} \frac{\partial p^{(H)}(u_{c}, z | \zeta)}{\partial z} u_{c} f(u_{c}) \tag{18}
\]
Integrating by parts, we obtain
\[
\langle f(u_{c}) \rangle^{(O)}_{\zeta} = \langle f(u_{c}) \rangle^{(H)}_{\zeta} + \\
\frac{1}{/} \int_{-\infty}^{\infty} dz \int du_{c} e^{-d_{z}^{(O)}(\zeta)} p^{(H)}(u_{c}, z | \zeta) u_{c} f(u_{c}) \tag{19}
\]
where boundary terms have vanished by construction.

Following the arguments leading to eq 8, we can separate bulk and interfacial contributions to \( p^{(H)} \):
\[
p^{(H)}(u_{c}, z | \zeta) = p^{(H)}_{bulk}(u_{c}, z | \zeta) + \epsilon p^{(H)}_{int}(u_{c}, z | \zeta) \tag{20}
\]
where \( \epsilon = D_{int}/D \ll 1 \). As with \( p(\zeta) \) in eq 6, the bulk contribution will dominate any average that does not strictly vanish away from the interface. Unlike \( f(u_{c}) \), the even function \( f(u_{c})u_{c} \) will not average to zero in the bulk due to symmetry. The interfacial contribution to the second term in eq 19 can therefore be discarded. Recognizing that orientational averages in the bulk are independent of \( z \) and \( \zeta \), and noting that the depth distribution in bulk is uniform, \( p^{(H)}_{int}(z) = D^{-1} \), we obtain
\[
\langle f(u_{c}) \rangle^{(O)}_{\zeta} = \langle f(u_{c}) \rangle^{(H)}_{\zeta} + \frac{1}{/} \int_{0}^{L} dz /D \langle f(u_{c})u_{c} \rangle_{bulk} \tag{21}
\]
The second term of eq 21 is by no means negligible, since \( f(u_{c})u_{c} \sim O(D_{int}/D) \) and the interfacial region of a dense liquid spans only a handful of molecular layers.

When the differences between smoothly attenuated averages of a function \( f(u_{c}) \) over the distributions 14 and 15 are applied to the specific case \( \langle u_{c} \rangle_{\zeta} \), we obtain
\[
\langle u_{c} \rangle^{(O)}_{\zeta} = \langle u_{c} \rangle^{(H)}_{\zeta} - \frac{1}{3D} \tag{22}
\]
As anticipated, the bias of convention 1 is negative relative to that of convention 2. The difference \( /3D \) is small in an absolute sense (recall that \( D \) is the thickness of the entire sample), but so is \( \langle u_{c} \rangle_{\zeta} \). Remarkably, this result does not depend on the distance \( L \) over which averaging is attenuated. The issue we encounter when using a sharp cutoff is recapitulated when using a smooth cutoff due to an accumulation of bias over the entire range of averaging. Similar consequences of associating a molecular property with that of a single constituent atom have been described in the context of computing electrostatic potentials on molecular centers."
distribution. We distinguish between this extreme and the rest of the distribution by the dividing point \( \mathcal{E}^* \), as indicated in Figure 2. We and others have shown for bulk systems a strong correlation between \( \mathcal{E} \) and hydrogen bond geometry, with more positive values of \( \mathcal{E} \) corresponding to weaker hydrogen bonds.\(^{11,22,37,38,48} \) It is therefore natural to associate the growing weight for \( \mathcal{E} > \mathcal{E}^* \) with a population of HOD molecules at the interface whose protons engage in very weak hydrogen bonds.\(^{23} \) This multiplication by showing the accuracy of eqs 9 and 12. Here, we have decreased. Panels b and c further demonstrate these similarities in D\(_2\)O. Each curve reports molecular dynamics results for water molecules in a 1 Å-thick slice of the slab parallel to the interface. Red lines show results for slices more than 2 Å below the Gibbs dividing surface. Black lines show results for the interfacial region, specifically for slices centered at \( \Delta z = -2 \) Å, \( \Delta z = -1 \) Å, \( \Delta z = 0 \) Å, \( \Delta z = 1 \) Å, \( \Delta z = 2 \) Å, and \( \Delta z = 3 \) Å. (A few of these lines are labeled by the corresponding value of \( \Delta z \).) We plot the logarithm of each distribution, scaled by its value at \( \mathcal{E} = \mathcal{E}^* \) distinguishes a range where electric field statistics are unaltered by the interface from a range that shows dramatic change near the Gibbs dividing surface.

Perhaps more strikingly, electric field statistics within the hydrogen bonded population are essentially invariant with depth. Even molecules above the Gibbs dividing surface experience extremely negative fluctuations in \( \mathcal{E} \) (say, \( \mathcal{E} < -3 \) V/Å, corresponding to abnormally strong hydrogen bonds) with a probability relative to \( \mathcal{E} \) corresponding to weak hydrogen bonds.\(^{11,22,37,38,48} \) It is therefore natural to associate the growing weight for \( \mathcal{E} > \mathcal{E}^* \) with a population of HOD molecules at the interface whose protons engage in very weak hydrogen bonds or lack a hydrogen bond acceptor altogether. Indeed, molecules corresponding to this population possess with high probability a “dangling” or “free” hydroxyl group.\(^{23} \) Others have drawn similar conclusions.\(^{6,49,50} \)

Before focusing on the conceptually simple functions \( \langle u_z \rangle_\mathcal{E} \) and \( \langle u_z \rangle_\mathcal{E} \), we present results for the full set of nonvanishing orientational averages listed in Table 1. These quantities, plotted in Figure 3 as functions of \( \mathcal{E} \), fall roughly into two classes: those that peak in the region \( \mathcal{E} > \mathcal{E}^* \) and are very small in magnitude for \( \mathcal{E} < \mathcal{E}^* \) and those that change sign near \( \mathcal{E} \approx \mathcal{E}_0 \) and then grow approximately linearly in magnitude as \( \mathcal{E} \) decreases. Panels b and c further demonstrate these similarities by showing the accuracy of eqs 9 and 12. Here, we have multiplied orientational averages by \( P(\mathcal{E}) \), as they appear in our prediction (6) for SFG susceptibility. This multiplication emphasizes small differences near the peak of \( P(\mathcal{E}) \), which decays rapidly in the region \( \mathcal{E} > \mathcal{E}^* \).

For the remainder of this paper, we will examine \( \langle u_z \rangle_\mathcal{E} \) and \( \langle u_z \rangle_\mathcal{E} \) as prototypes for the frequency (or, equivalently, electric field) dependence of averages in Figure 3. The first of these functions quantifies correlations that are the subject of many discussions on SFG: If OH bonds with vibrational frequency in a particular range are strongly aligned parallel (antiparallel) to the outward surface normal, \( \langle u_z \rangle_\mathcal{E} \) will be large and positive (negative) at those frequencies. The second function carries more subtle structural information: Nonzero \( \langle u_z \rangle_\mathcal{E} \) at a certain frequency signifies correlations between the electric field acting on the OH bond and orientation of the OD bond vector. The geometry of a water molecule ensures that \( \langle u_z \rangle_\mathcal{E} \) and \( \langle u_z \rangle_\mathcal{E} \) will oppose one another to some degree. Strong alignment of one bond vector in any direction implies partial antialignment of the other. Nonetheless, we will see that \( \langle u_z \rangle_\mathcal{E} \) reveals aspects of molecular arrangement that are not apparent from \( \langle u_z \rangle_\mathcal{E} \).

Figure 4 shows \( \langle u_z \rangle_\mathcal{E} \) and \( \langle u_z \rangle_\mathcal{E} \) as functions of electric field, as well as their products with the distribution \( P(\mathcal{E}) \). The very small magnitude of \( \langle u_z \rangle_\mathcal{E} \) for \( \mathcal{E} < \mathcal{E}^* \) indicates that correlations between OH vibrational frequency and OH bond orientation are significant only for free OH groups, which preferentially align parallel to the surface normal. Hydroxyl groups that do engage
in hydrogen bonding exhibit very little orientational bias. Because \( P(\zeta) \) peaks in the corresponding frequency range, however, any residual bias is greatly amplified. Together, these effects conspire to make the convention for breaking symmetry of a liquid slab an important issue. We have shown that even a smooth cutoff can impart a frequency-independent bias depending on which atom of an HOD molecule is used to gauge its orientational bias that determine our approximation to resonant contributions due to \( \langle u_z \rangle \) and to \( \langle u_z' \rangle \). If mixing is strong, it will not even be possible to decompose a computed SFG susceptibility into simply understood contributions. Orientational averages at all frequencies will be sensitive to alignment of both a molecule’s hydroxyl groups. In this case, one cannot unambiguously reason, e.g., that a negative contribution to the computed susceptibility for \( \zeta < \zeta^* \) indicates downward alignment of strong hydrogen bonds.

VI. Conclusions

We have presented a theoretical framework for understanding, and perhaps ultimately predicting, SFG response for the air—water interface. The same analysis can also be used to investigate the SFG response for dilute electrolyte solutions at the air interface, and such studies are currently underway. By design, our treatment is less elaborate in its details than previous calculations. With complications of bath dynamics, vibrational coupling and relaxation, and fluctuating electronic polarization removed, a relatively simple physical picture emerges for the microscopic origins of interfacial hyperpolarizability. Specifically, we can identify a handful of specific measures of orientational bias that determine our approximation to resonant second-order response. Because the orienting effects of air—water interfaces are rather weak, the frequency dependence of all relevant orientational averages are well captured by a pair of functions involving average projections of OH and OD bond vectors onto the outward surface normal.

These simplifications, together with results of computer simulations, provide several insights that would have been difficult to extract from more detailed expressions for hyperpolarizability or from more elaborate molecular models. Our results reveal, for example, an alarming sensitivity to the convention one chooses for breaking symmetry of a simulated liquid slab. Obtaining accurate predictions of SFG from related calculations will require close attention to this accumulated effect of water molecules’ internal structure. It is not clear that a spectroscopic formalism based on the dipole approximation is sufficient for this purpose, especially when normal modes involve motion of multiple protons (as is the case for pure H2O...
and D₂O). Due to the ambiguity facing current methodologies, we recommend that questions such as, in which direction do hydrogen bonded OH groups point on average at the interface, would be better posed as, do hydrogen bonded OH groups at the interface adopt a noteworthy net orientation on average? An affirmative answer would require an insensitivity to symmetry-breaking convention. For our simulations of the SPC/E model in a liquid slab geometry, the answer according to this criterion is decidedly negative. These issues are unfortunately important and not only for achieving quantitative SFG predictions. Weak, frequency-independent net orientation can alter the qualitative character of computed susceptibilities, due to the weighting of orientational averages by the bulk electric field distribution, which decays rapidly outside the frequency range corresponding to hydrogen bonded species.

We have stopped short of presenting full numerical results for SFG signals that can be measured in the laboratory. The primary reason for doing so is that the coefficients determining the linear combination of orientational averages in eq 6 have not been well characterized for water molecules in solution or in heterogeneous interfacial environments. Our principal result for neat solvent is that eq 6 is essentially a linear combination of the conceptually appealing functions \( \langle \hat{u}_z \rangle \) and \( \langle \hat{u}_z \rangle \). The range of susceptibilities that could be obtained from the data plotted in Figure 4 is not difficult to assess by visual inspection, an exercise we leave to the interested reader. Ambiguity in methods of symmetry breaking, significant nonresonant contributions to most experimental data, and the possible importance of non-Condon effects discourage us from following fitting procedures others have followed to infer these parameters.56 We note, however, that the contributions to second-order susceptibility we have calculated could be combined with reasonable coefficients to roughly reproduce results reported from more involved approaches.

The theoretical perspective we have developed argues strongly against interpreting SFG data as one would a simple absorption spectrum. Associating spectral peaks with distinct intermolecular arrangements, and their intensities as populations, can in fact be misguided even for simple absorptive measurements. In the context of vibrational spectroscopy of bulk liquid water, we have shown for simulated systems that a pronounced shoulder in the Raman line shape signifies nothing more than the nonlinear relationship between frequency and hydrogen bond geometry.11 For liquid–vapor interfaces as well, distributions of \( \rho_{OH} \) for hydrogen bonded species do not suggest a meaningful segregation into ionic and liquidlike structures53 as previously suggested.50 By this measure, only free OH species stand out as a discrete variation in intermolecular structure.

Despite a lack of discrete variety in the local geometry of hydrogen bonded molecules, it is clear that rich SFG signals could nonetheless be obtained from our numerical results. For example, a simple sum of the two contributions plotted in Figure 4 would yield a susceptibility that changes sign twice as a function of frequency. A peak at very negative \( \hat{c}_z \), like that observed in recent experiments,24 would result. Its origin can be traced not to a distinct microscopic structure but instead to a melange of free hydroxyl alignment, weak correlation between OH vibrational frequency and OD orientation, and weighting by the bulk electric field distribution. The location and height of this feature, and indeed its very existence, are sensitive to changing the coefficients of linear combination.

Acknowledgment. This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division and the Chemical Sciences, Geosciences, and Biosciences Interfaces under the Department of Energy Contract No. DE-AC02-05CH11231. Gaussian calculations were performed at the U.C. Berkeley College of Chemistry Molecular Graphics and Computation Facility (NSF Grant No. CHE-0233882). J.D.S is supported by the Camille and Henry Dreyfus Foundation postdoctoral program in environmental chemistry.

References and Notes

(33) Skinner, J. L. Personal communication, 2008.
This separation of interfacial and bulk length scales rests on the assumption that systems of interest are well removed from critical points, as is certainly the case for liquid water at ambient conditions.

Although we notionally generated the configuration $\Gamma$ by reflection, we nonetheless define a molecular reference frame for each of its molecules as we would in the case of $\Gamma$. The resulting coordinate system is therefore right-handed (which would not be the case had we transformed $w$ in the same way as $u$ and $v$). As we have described it, this symmetry argument holds only for achiral molecules such as HOD.

Periodic boundary conditions prevent rigorous association of a molecule and all its images with one side of a dividing surface. However, our simulations include ample space, usually devoid of mass, between replicated liquid regions. We can thus unambiguously identify a liquid subsystem whose center of mass $z_{\text{c.o.m.}}$ in the $z$-direction is well-defined.

When we segregate molecules using a dividing surface, we refer to their periodic images whose $z$-displacements from $z_{\text{c.o.m.}}$ are smaller than $L_z/2$.

We judge a hydroxyl group to be dangling according to the distance $r_{\text{OH}}$ between the proton and the nearest oxygen atom within a cone of half-width $\pi/6$ emanating from the proton in the direction of the OH bond vector. Values of $r_{\text{OH}}$ much greater than the range of typical bulk fluctuations (say, two standard deviations) can safely be considered to lack a hydrogen bond acceptor.

Equation 6 permits a straightforward modification to include non-Condon effects through electric field dependence of the transition dipole.