Supplementary Information of Deciphering Second Harmonic Generation Signals

Yann Foucaud,[*a] Bertrand Siboulet,[a] Magali Duvail,[a] Alban Jonchere,[a] Olivier Diat,[a] Rodolphe Vuilleumier,[*b] and Jean-François Dufrêche[*a]

[a]ICSM, Univ Montpellier, CEA, CNRS, ENSCM, Marcoule, France
[b]PASTEUR, Département de chimie, Ecole normale supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France

Contents

1 Dependency of the hyperpolarisability upon the local chemical environment of the molecule

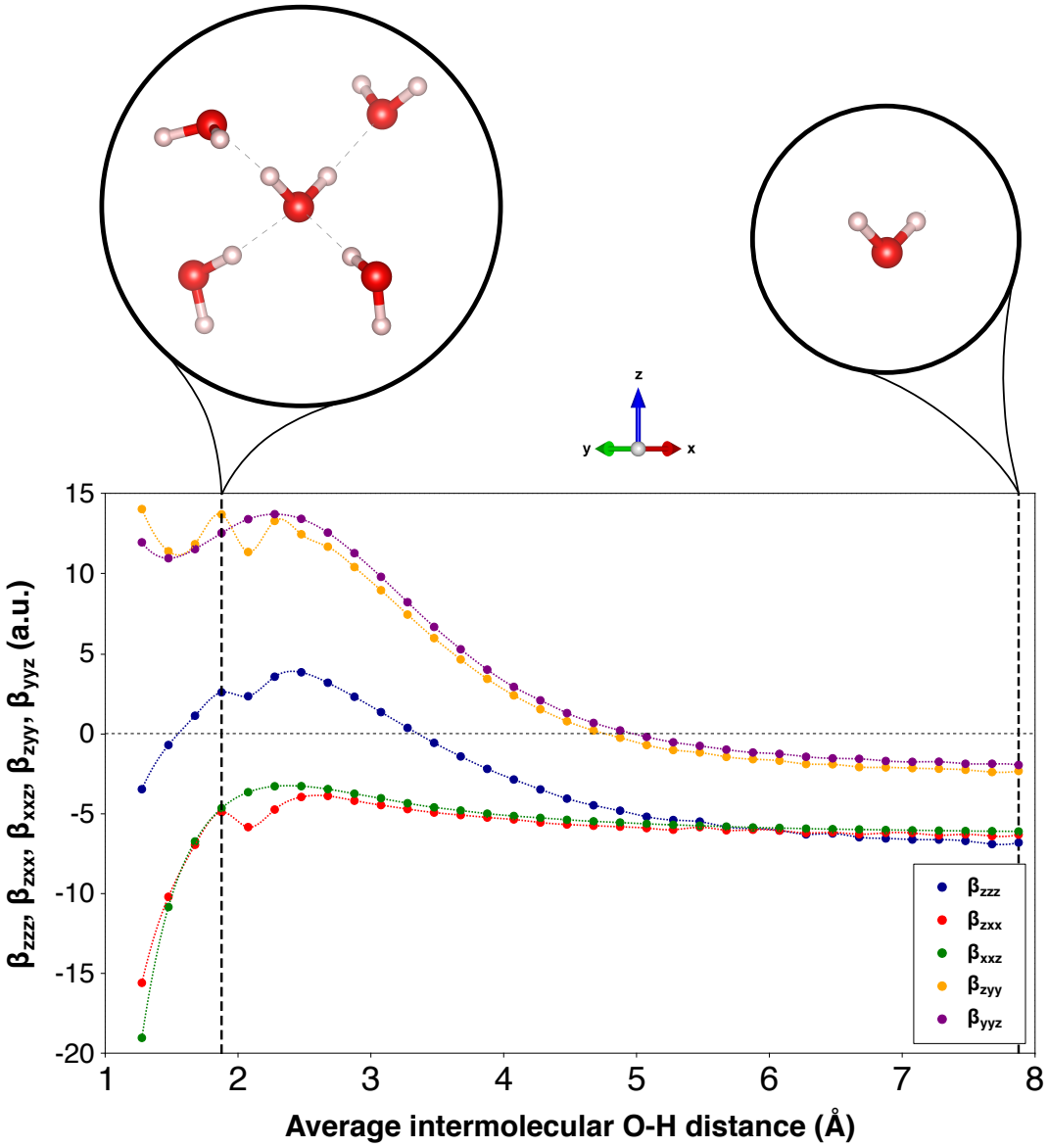

Fig. S1 Variation of the non-zero components of the hyperpolarisability as a function of the average length of the hydrogen bonds around the considered molecule.

2 Determining the second-order electric susceptibility from experimental SHG signals

$$I_S(\gamma) = \left| a_1 \left(\frac{\chi^{(2)}_{s,xxz} + \chi^{(2)}_{s,yyz}}{2} \right) \sin(2\gamma) \right|^2 \tag{S1}$$

$$I_P(\gamma) = \left| \left(a_2 \left(\frac{\chi^{(2)}_{s,xxz} + \chi^{(2)}_{s,yyz}}{2} \right) + a_3 \left(\frac{\chi^{(2)}_{s,zxx} + \chi^{(2)}_{s,zyy}}{2} \right) + a_4 \chi^{(2)}_{s,zzz} \right) \cos^2(\gamma) + a_5 \left(\frac{\chi^{(2)}_{s,zxx} + \chi^{(2)}_{s,zyy}}{2} \right) \sin^2(2\gamma) \right|^2 \tag{S2}$$

$$I_{45}(\gamma) = \left| a_1 \left(\frac{\chi^{(2)}_{s,xxz} + \chi^{(2)}_{s,yyz}}{2} \right) \sin(2\gamma) + \left(a_2 \left(\frac{\chi^{(2)}_{s,xxz} + \chi^{(2)}_{s,yyz}}{2} \right) + a_3 \left(\frac{\chi^{(2)}_{s,zxx} + \chi^{(2)}_{s,zyy}}{2} \right) + a_4 \chi^{(2)}_{s,zzz} \right) \cos^2(\gamma) + a_5 \left(\frac{\chi^{(2)}_{s,zxx} + \chi^{(2)}_{s,zyy}}{2} \right) \sin^2(2\gamma) \right|^2 \tag{S3}$$

3 Theoretical background

In nonlinear optics experiments, the molecules located at the interface are embedded in an applied optical field that corresponds to an external electric field.[1,2] This latter consistently induces a polarisation, $\mathbf{P}(t)$, which, if we assume that the polarisation at time t solely depends on the instantaneous value of the electric field strength, can be expressed under its general form as a power:[1,3,4]

$$\mathbf{P}(t) = \mathbf{P}^{(1)}(t) + \mathbf{P}^{(2)}(t) + \mathbf{P}^{(3)}(t) + \cdots \tag{S4}$$

where $\mathbf{P}^{(n)}(t)$ is the n-order polarisation term. Since second-order nonlinear phenomena such as SHG mostly result from second-order effects, the higher-order terms can be neglected when investigating the aforementioned techniques. Therefore, the induced polarisation can be expressed under a general form for second-order phenomena as follows:

$$\mathbf{P}(t) = \mathbf{P}^{(1)}(t) + \mathbf{P}^{(2)}(t). \tag{S5}$$

Although both the above-mentioned polarisation terms depend on the electric field strength, $\mathbf{E}(t)$, the dependence of $\mathbf{P}^{(1)}(t)$ upon the electric field strength is linear while

this of $\mathbf{P}^{(2)}(t)$ is nonlinear. For a lossless and dispersion-less medium, this latter term can also be written as a function of frequencies:[1,2,4]

$$\mathbf{P}^{(2)}(\omega_3) = \varepsilon_0 \chi^{(2)}(\omega_3, \omega_2, \omega_1)\mathbf{E}(\omega_1)\mathbf{E}(\omega_2) \tag{S6}$$

where ε_0 is the vacuum permittivity and $\chi^{(2)}$ is the second-order nonlinear optical susceptibility. Generally speaking, in nonlinear optics, ω_3 is a linear combination of ω_1 and ω_2, where ω_3 is the frequency of the induced polarisation and ω_1 and ω_2 are the frequencies of the incident beams.[2] In SHG, $\omega_1 = \omega_2$ and $\omega_3 = 2\omega_1$, which allows one to express the global induced polarisation under a SHG-specific form, including the linear term, as:

$$\mathbf{P}(2\omega) = \varepsilon_0 \chi^{(1)}\mathbf{E}(\omega) + \varepsilon_0 \chi^{(2)}\mathbf{E}(\omega)\mathbf{E}(\omega) \tag{S7}$$

where $\chi^{(1)}$ is the linear susceptibility. This second-rank tensor represents the constants of proportionality relating linearly the field amplitude to the polarisation. Besides, the 27 components of the third-rank $\chi^{(2)}$ tensor link the product of field amplitudes to the nonlinear polarisation. Restrictions resulting from various symmetries can be applied to the 27 components of $\chi^{(2)}$, therefore reducing significantly the number of independent components of $\chi^{(2)}$. First, for second-order phenomena, the two last indices of χ^2_{ijk}, j and k, can be interchanged in the second-order polarisation $\chi^2_{ijk}(\omega_n + \omega_m, \omega_m, \omega_n)E_k(\omega_m)E_j(\omega_n)$ as long as the last two frequency arguments, n and m, are also interchanged. This symmetry, known as intrinsic permutation symmetry, results from the fact that the two fields can equally be first and second in the product $E_j(\omega_n)E_k(\omega_m)$. Hence, since $\chi^{(2)}_{ijk} = \chi^{(2)}_{ikj}$, the number of independent components of $\chi^{(2)}$ is reduced to 18.[1,3,4] Furthermore, for a lossless medium, all the components of $\chi^{(2)}$ are real and, therefore, the full permutation symmetry can be applied. This condition states that all the frequency indices of $\chi^{(2)}$ can be interchanged as long as the Cartesian indices are also interchanged. Albeit the components of $\chi^{(2)}$ generally depend on the frequencies of the applied fields, they can be considered as constant under the assumption of the instantaneous response made for nonlinear optics, which corresponds to the Kleinman symmetry.[1,5] Combined with the full permutation symmetry, it allows one to write $\chi^{(2)}_{ijk} = \chi^{(2)}_{jki} = \chi^{(2)}_{kij} = \chi^{(2)}_{ikj} = \chi^{(2)}_{jik} = \chi^{(2)}_{kji}$, which reduces the number of independent components of the nonlinear second-order susceptibility to 10 whenever the Kleinman symmetry condition is valid, *i.e.*, if the dispersion of the susceptibility can be neglected. Discussions about this assumption can be found further. Additionally, in the case of an interface along the *(Oxy)* plane, the x and y directions are equivalent in average and, consistently, the z-axis is a fourfold axis of symmetry. Those additional symmetries further reduce the number of independent components of $\chi^{(2)}$.

In SHG experiments, $\chi^{(2)}$ links the polarisation angle of the incident light, γ, to the S-, P-, and 45-polarised components of the harmonic intensity, which is the measurable quantity. Experimentally, considering the isotropy of the surface (located along the *(Oxy)* plane) and for a water/air interface, $\chi^{(2)}$ contains only three nonzero components, namely, $\chi^{(2)}_{zzz}$, $\chi^{(2)}_{zxx}$, and $\chi^{(2)}_{xxz}$. $\chi^{(2)}$ therefore links the microscopic properties,

which can be investigated by means of atomistic simulations, to the macroscopic SHG intensities, which correspond to the experimental responses. Considering Eq. S7, $\chi^{(2)}$ can be determined by applying a series of external electric fields to a system and subsequently calculating the induced polarisation, **P**, which can be easily related to the electric dipole moment, **p**:

$$\mathbf{P} = \frac{\mathbf{p}}{V} \tag{S8}$$

where V is the volume cell. For practical convenience, authors generally rather consider the induced surface polarisation,[6–9] which is defined as follows:

$$\mathbf{P_s} = \frac{\mathbf{p}}{S} \tag{S9}$$

where S is the area of the interface. In what follows, we will only use the surface polarisation and, according to Eq. S7, the surface first- and second-order electric susceptibilities, $\chi_s^{(1)}$ and $\chi_s^{(2)}$, respectively. Using a surface polarisation and surface electric susceptibilities allowed comparing the responses from an interface to another when the values of the components of the electric susceptibilities were used instead of their ratios. For a finite system like a molecule, the electric dipole moment **p** of a collection of i undifferentiated point charges (positive or negative), q_i, at positions $\mathbf{r}_i$ is defined as:

$$\mathbf{p} = \sum_i q_i \mathbf{r}_i. \tag{S10}$$

Provided that the negative and positive charges of the system are differentiated, Eq. S10 can be also expressed as a function of an ionic and an electronic contributions:

$$\mathbf{p} = \mathbf{p}_{ion} + \mathbf{p}_{elec} \tag{S11}$$

The ionic contribution, which calculation is quite trivial, is related to the positions of ions in space, $\mathbf{R}_i$, and to the net charge that they carry, Z_i. This latter corresponds to their atomic number in the case of all-electron calculations or to their valency charge in the case of pseudo-potentials calculations:

$$\mathbf{p}_{ion} = \sum_i Z_i \mathbf{R}_i \tag{S12}$$

Besides, obtaining the electronic term, $\mathbf{p}_{elec}$, in Eq. S11 is not straightforward since, unlike nuclei, electrons cannot be, *a priori*, considered as localised points.[10] Among the few methods that can be used to obtain $\mathbf{p}_{elec}$ in the Modern Theory of Polarisation,[11–13] the Maximally Localised Wannier Functions (MLWF) represent the less computational-costly and the most accurate option. The Wannier functions, which correspond to the localised, *i.e.* the Fourier-transform, equivalent of the Bloch functions, can be defined, for the electron band n, as:[14–16]

$$|w_{n\mathbf{R}}\rangle = \frac{V_{cell}}{(2\pi)^3} \int_{BZ} e^{-i\mathbf{k}\cdot\mathbf{R}} |\psi_{n\mathbf{k}}\rangle d^3k \tag{S13}$$

Since the Wannier functions are localised, the average position of the electrons in a given electronic band n, which is called the Wannier centre, can be calculated as follows:

$$\bar{\mathbf{r}}_n = \frac{V_{cell}}{(2\pi)^3} \int_{BZ} \langle u_{n\mathbf{k}} | i\nabla_{\mathbf{k}} u_{n\mathbf{k}} \rangle d^3k \tag{S14}$$

Consistently, the Wannier centres can be seen as electronic localised point charges in this formalism.[17] Hence, the electronic contribution to the electric dipole moment (Eq. S11) can be expressed as a function of the Wannier centres (Eq. S15): each Wannier centre describes an occupied Wannier function, *i.e.*, an electron band which, in the case of a water molecule, carries consistently a net charge exactly equal to -2:

$$\mathbf{p}_{elec} = -2 \sum_{n}^{occ} \bar{\mathbf{r}}_n \tag{S15}$$

The total electric dipole moment for a given molecule can therefore be calculated using the positions of the Wannier centres, $\bar{\mathbf{r}}_n$, the positions of the ions, x_i, and their net charges, q_i, as follows:[17,18]

$$\mathbf{p} = \sum_i (q_i x_i)^{ions} - 2 \sum_{n}^{occ} (\bar{\mathbf{r}}_n)^{WFs}. \tag{S16}$$

4 Detailed method

4.1 Determining the interface dipole contribution to the surface polarisation

For the two considered simulation boxes, snapshots were regularly extracted from the CMD simulations. On these snapshots, the positions of the ions were potentially first relaxed through a sequence of wave function optimisations using the above-described settings. The results about the influence of the geometry optimisation are available and discussed in SI. Then, regardless of this eventual geometry relaxation, an external electric field was applied to the system using the Berry phase.[19,20] The wave function was optimised in presence of the electric field, without however relaxing the ions positions. The MLWFs were subsequently determined and their corresponding Wannier centres localised.[21] This procedure was performed for the 25 different external electric fields applied on each snapshot (see Table S1 and Fig. 1b). Among those 25 fields, some were collinear to a given box simulation axis while some were not collinear, *i.e.*, presenting two non-zero components. This allowed determining the interaction terms existing between the different components of the electric fields. For instance, $\chi^{(2)}_{s,xxz}$ is the coefficient that links $P_{s,x}$ to E_xE_z and its determination requires that both E_x and E_z are non-zero. Moreover, on each direction (collinear to the main axes or not), several intensities of the electric field were applied to (1) maximise the accuracy of the determination of the interaction coefficients and (2) to correctly determine the quadratic coefficients (*e.g.* $\chi^{(2)}_{s,zxx}$ or $\chi^{(2)}_{s,zzz}$). On overall, the 25 different applied external electric fields finely sampled all the domain of fields, as displayed in Fig. 1b, which provided the conditions for an efficient least-squares fitting. After the 25 different DFT calculations were carried out on a given snapshot, each side of the water slab along the z-axis was considered separately to avoid the influence of the second interface (which presence is due to the periodicity of the simulation box) on the global surface polarisation. For that, water molecules were divided in two groups based on the z-coordinate of the molecular referential point (Step 2 in Fig. S2). The influence of the selected reference on the final responses was thoroughly investigated in section SI. In the whole study, when nothing is specified, the molecular referential point is defined as the oxygen atom of the molecule.

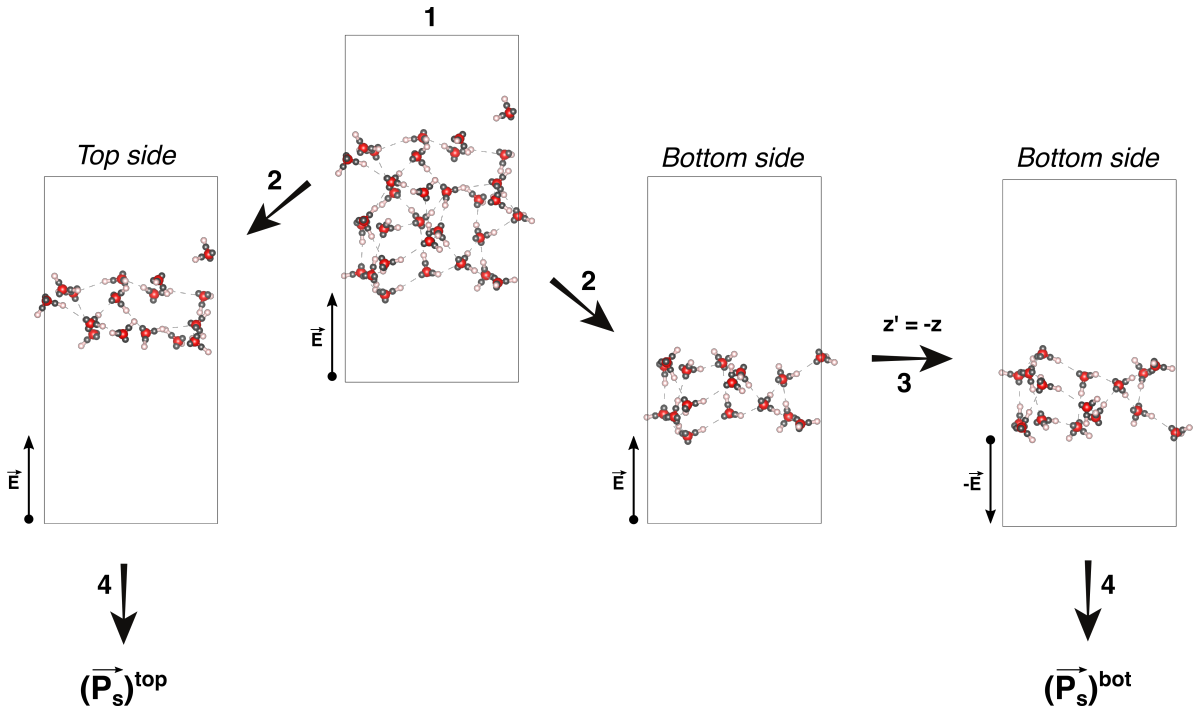

Fig. S2 Procedure to calculate the 3 independent non-zero components of $\chi_s^{(2)}$ from a given snapshot. (1) The wave function was optimised in presence of an electric field to determine the coordinates of the Wannier centres; (2) The water slab was cut in two parts with regard to the z-coordinate of the molecular referential point of the molecules; (3) The bottom part was returned by applying $z' = -z$ on all the ions and Wannier centres coordinates and by applying $E_z^{'} = -E_z$ on the electric field; (4) The surface polarisation of each side was calculated, based on the coordinates of the ions and Wannier centres.

For the bottom part, the coordinates of the atoms as well as the electric field were returned along the z-axis by multiplying the z-coordinate and the z-component of the electric field by -1 (Step 3 in Fig. S2). Then, on each part of the water slab (bottom and top interfaces), the electric dipole moment of each water molecule was calculated based on the ions and Wannier centres coordinates as defined in Eq. S17:

$$\mathbf{p} = \sum_i (q_i x_i)^{ions} - 2\sum_n^{occ} (\bar{\mathbf{r}}_n)^{WFs}. \tag{S17}$$

The electric dipole moments were then summed over the water molecules to obtain the global electric dipole moment of each side of the water slab. This quantity was subsequently divided by the area of the interface to obtain the surface polarisation, $\mathbf{P}_s$, for each side of the water slab (Step 4 in Fig. S2). Considering the theoretical background (see section 3 of SI), Eq. S7 can be re-written using surface polarisation and surface electric susceptibilities as follows:

$$\mathbf{P_s} = \mathbf{P_s^0} + \varepsilon_0 \chi_s^{(1)} \mathbf{E} + \varepsilon_0 \chi_s^{(2)} \mathbf{EE} \tag{S18}$$

where $\mathbf{P_s^{(0)}}$, ε_0, $\mathbf{E}$, $\chi_s^{(1)}$, and $\chi_s^{(2)}$ represent the permanent surface polarisation, the vacuum permittivity, the electric field, the first-order surface electric susceptibility, and

the second-order surface electric susceptibility, respectively. In Eq. S18, $\chi_s^{(2)}$ represents the quantity to be calculated since it is exactly equal to the $\chi_s^{(2)}$ determined by SHG experiments. Eq. S18 can be written under developed forms along *x*-, *y*-, and *z*-axes, for a given applied external electric field, as follows:

$$P_{s,x} = P_{s,x}^{(0)} + \varepsilon_0 \sum_{i=y,z} \chi_{s,xi}^{(1)} E_i + \varepsilon_0 \sum_{i=y,z} \sum_{j=y,z} \chi_{s,xij}^{(2)} E_i E_j \tag{S19}$$

$$P_{s,y} = P_{s,y}^{(0)} + \varepsilon_0 \sum_{i=x,z} \chi_{s,yi}^{(1)} E_i + \varepsilon_0 \sum_{i=x,z} \sum_{j=x,z} \chi_{s,yij}^{(2)} E_i E_j \tag{S20}$$

$$P_{s,z} = P_{s,z}^{(0)} + \varepsilon_0 \sum_{i=x,y} \chi_{s,zi}^{(1)} E_i + \varepsilon_0 \sum_{i=x,y} \sum_{j=x,y} \chi_{s,zij}^{(2)} E_i E_j \tag{S21}$$

The three components of $\mathbf{P^{(0)}}$, the nine components of $\chi_s^{(1)}$, and the 27 components of $\chi_s^{(2)}$ were fitted by the least-squares method using a Levenburg-Marquardt gradient for the minimisation.[22,23] The fitting was only constrained by the intrinsic permutation symmetry, induced by SHG, existing on the components of $\chi_s^{(2)}$, which implied that $\chi_{s,ijk}^{(2)} = \chi_{s,ikj}^{(2)}$ and, therefore, that only 18 components of $\chi_s^{(2)}$ were independent. Additionally, the permanent surface polarisation, $\mathbf{P_s^{(0)}}$, *i.e.*, a surface polarisation in absence of any applied electric field, existed due to the non-random orientation of water molecules at the interface[6,24] and was included in the fitting procedure. For the fitting, the *y*-values were constituted by the calculated surface polarisation values while the *x*-values were the different external electric fields applied during the quantum mechanics calculations. The fitting was performed separately on each side of the interface (top and bottom) of each snapshot using the 25 values of (P_x, P_y, P_z, E_x, E_y, E_z). Then, the fitted components (three components of $\mathbf{P^{(0)}}$, nine components of $\chi_s^{(1)}$, and 18 components of $\chi_s^{(2)}$) were averaged over the two sides of the water slab and over all the selected snapshots. The SHG responses were simulated by injecting the calculated $\chi_{s,xxz}^{(2)}$, $\chi_{s,zxx}^{(2)}$, and $\chi_{s,zzz}^{(2)}$ in Eq. S1, S2, and S3 to calculate I_S, I_P, and I_{45}, respectively, as a function of the incident polarisation angle, γ. To consider the fourfold symmetry axis existing along the normal to the interface, *i.e.*, the $x \leftrightarrow y$ symmetry, the average value of $\chi_{xxz}^{(2)}$ and $\chi_{yyz}^{(2)}$ as well as the average value of $\chi_{zxx}^{(2)}$ and $\chi_{zyy}^{(2)}$ were used in Eq. S1, S2, and S3. The intensities of the three curves were normalised to the P-polarised curve (that generally exhibits the highest intensity). When various SHG responses were compared, the two sets of curves were normalised to the maximum of their P-polarised curves.

4.2 The electric field discontinuity

In the literature, the importance of the discontinuity of the electric field at the interface has been extensively pointed out.[7,8] Indeed, E_z varies rapidly across the interface layer (Fig. S3), which implies that $P_s^{(2)}$ in the interface layer can be significantly different from that in the bulk. Moreover, the external electric field, applied using the

Berry phase and corresponding to a Maxwell electric field, can be significantly different from the local electric field as defined in the vacuum phase. Two options can allow overcoming this problem. First, instead of using E_z, many authors[7,8] rather considered the z-component of the electric displacement, D_z, which is continuous and can be regarded as constant across the interface layer. The z-component of the electric displacement can be expressed as follows:

$$D_z = E_z + 4\pi P_z \tag{S22}$$

where P_z is the component of the polarisation along the z-axis of the whole simulation box as defined in Eq. S8. The above-described fitting procedure has then been carried out considering E_x, E_y, and D_z since they are all constant along x, y, and z-axes, respectively. Nonetheless, by using the electric displacement, we considered the field just inside the water slab, which did not correspond to the external electric field applied on the whole slab. Therefore, after the fitting, we corrected the determined values of $\chi^{(1)}_{s,app}$ as follows:

$$\chi^{(1)}_{s,ij} = (\varepsilon_r)^n \chi^{(1)}_{s,ij,app} \tag{S23}$$

where $n = 0$ if $j \neq z$ and $n = 1$ if $j = z$, and $\varepsilon_r = 1.77689$. Values of $\chi^{(2)}_{s,app}$ were corrected in a similar way:

$$\chi^{(2)}_{s,ijk} = (\varepsilon_r)^n \chi^{(2)}_{s,ijk,app} \tag{S24}$$

where $n = 0$ if $j \neq z$ and $k \neq z$, $n = 1$ if $j = z$ or $k = z$, and $n = 2$ if $j = z$ and $k = z$, and $\varepsilon_r = 1.77689$.

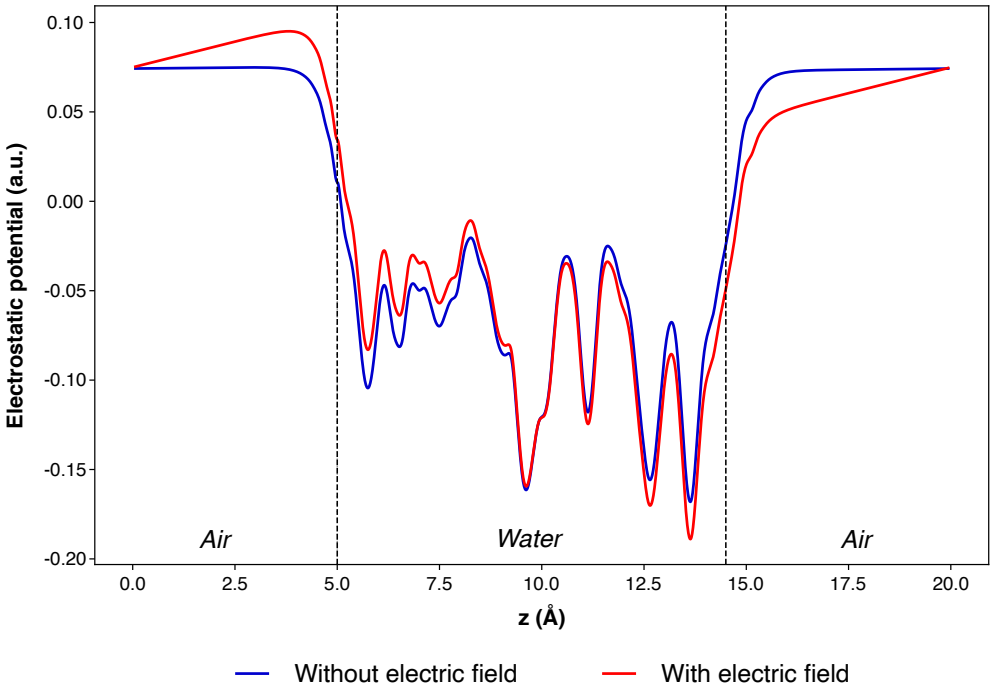

Fig. S3 Electrostatic potential along the z-axis in the cell comprising 32 water molecules on a selected snapshots, with (red curve) and without (blue curve) applying an external electric field. The dashed lines indicate the position of the Gibbs dividing surface.

4.3 The bulk quadrupole contribution

Although SHG spectroscopy is valid only in the electric dipole approximation, the higher-order contribution from the induced electric quadrupole can be significant since this latter is not forbidden in centrosymmetric medium.[25] Hence, despite the low intensity of the induced electric quadrupole contribution compared to the interface induced electric dipole term, the difference in volume between the interface and the bulk generally makes the bulk contribution comparable to the surface signal.[26] As early as in 1962, Bloembergen suggested a significant bulk electric quadrupole contribution from the centrosymmetric medium.[27,28] Since that, this topic has been intensively debated in SHG and the scientific consensus was that the bulk electric quadrupole contribution is significant in the SHG signal, which is indicated by the Kleinman symmetry breaking,[29–34] as it has also been demonstrated for other materials.[35] However, recently, a further sophisticated theory has been proposed to explicit the role of the electric quadrupole contribution in the total SHG/SFG signals: although this latter is always significant in the SHG/SFG signal, its significance cannot be explained by the Kleinman symmetry breaking.[26] Furthermore, the same authors demonstrated that, in the case of SHG applied on a water/air interface, the bulk electric quadrupole not only contributes significantly to the SHG signal but dominates it.[25] The significance of the electric quadrupole is still unclear since it could also depend on the characteristics of the surface structure[8] as well as on the value of the optical dielectric constant.[7] In particular, in media with a large optical dielectric constant, it has been demonstrated that the bulk contribution (quadrupole) is an order of magnitude lower than the surface non-linearity.[7]

According to these discussions and considering the failure of the sole dipole contribution to simulate a consistent SHG signal in our approach, we decided to include the bulk electric quadrupole contribution in our method. Generally speaking, the induced nonlinear polarisation in the multipole expansion is given as follows:[25,26,36]

$$P_i^{(2)} = \mu_i - \frac{\partial q_{il}}{\partial l} \tag{S25}$$

where μ_i and q_{il} correspond to the induced dipole and quadrupole per volume, respectively, where i and l represent the laboratory-fixed coordinates (*i.e.*, X, Y, and Z). The contribution of induced magnetic dipole and magnetic quadrupole was neglected considering their low significance for water molecules. The two terms in Eq. S25 can be expressed respectively as follows:

$$\mu_i = \varepsilon_0 \sum_{j,k} \chi_{ijk}^{dipole} E_{1j} E_{2k} + \varepsilon_0 \sum_{j,k,l} (\chi_{ijkl}^{quad1} \frac{\partial E_{1j}}{\partial l} E_{2k} + \chi_{ijkl}^{quad2} E_{1j} \frac{\partial E_{2k}}{\partial l}) \tag{S26}$$

$$q_{il} = \varepsilon_0 \sum_{j,k} \chi_{ijkl}^{quad3} E_{1j} E_{2k} \tag{S27}$$

Eq. S26 and S27 introduce two main quadrupole contributions, which are commonly described by researchers.[25,36] The first term, related to χ^{quad1} and χ^{quad2}, originates from the electric field gradient localised at the interface (as presented on Fig. S3) and

induces a dipole term, which is therefore included in the dipole moment, as mentioned in Eq. S26. Besides, the second term in Eq. S25 corresponds to the quadrupole polarisation, q_{IL}, which is induced by the electric field (and not by the gradient of the electric field) in the bulk. While the first term is included in the dipole moment calculated in our approach, the second term is excluded from the dipole term since it corresponds to an induced quadrupole polarisation.

In the *ab initio* approach developed here, we demonstrated that the sole consideration of the induced dipole moment, μ_i in Eq. S25, did not conduct to satisfactory results. Therefore, we included the quadrupole contribution, *i.e.*, q_{il}, in the total induced surface polarisation, which allowed to take the bulk quadrupole terms into account in the total SHG response. For that, the molecular quadrupole moment of each molecule was calculated as follows:

$$Q_{ij}^{mol} = \frac{1}{2}\left[\sum_{k}^{Ions+WFC} q_k(R_i^k - R_i^{ref})(R_j^k - R_j^{ref}) - \sum_{l}^{WFC} f_l(Q_{ij})^l\right] \tag{S28}$$

where Q_{ij}^{mol} is the (i,j) component of the molecule quadrupole moment matrix, Q^{mol}, $\mathbf{R^k}$ is the coordinate of the k^{th} ion or Wannier centre, $\mathbf{R^{ref}}$ is the coordinate of the molecular referential point, Q_{ij}^{l} is the (i,j) component of the second-moment of the l^{th} Wannier centre, determined in the MLWF approach, and f_l is the occupancy of the l^{th} orbital ($f_l = 2$ in all cases). The molecular referential point was the same criterion used to classify water molecules along the z-axis to cut the water slab in two parts. The molecular quadrupole moments were then summed over all the molecules comprised in the slice that was selected to calculate the total bulk quadrupole moment:

$$p_{iz}^{Quad} = \frac{\sum_n Q_{iz}^{mol}}{\delta_{app}} \tag{S29}$$

where δapp was the apparent slice thickness:

$$\delta_{app} = \frac{nM}{N_A \rho L_x L_y} \tag{S30}$$

where n was the number of water molecules considered to calculate P_{iz}^{Quad}, M was the molar mass of water, N_A was Avogadro number, ρ was the density of water (assumed to be equal to 1 g cm^{-3}), and L_x and L_y were the simulation box dimensions along the x- and y-axes, respectively. Then, the fitting procedure was carried out on the surface dipole polarisation that included the above-described calculated quadrupole contribution:

$$P_{s,i} = \frac{p_i^{Dip} + p_{iz}^{Quad}}{S} \tag{S31}$$

where $i = x, y, z$, p_i^{Dip} was the dipole moment calculated using Eq. S17, and $P_{s,i}$ was the surface polarisation (including quadrupole terms) that can be used for the fitting procedure (Eq. S19,S20, and S21).

4.4 The external electric fields applied on the system

The laser used in our experiments, which was a typical apparatus for SHG experiments, had an average power of 3 W, *i.e.*, 3 $J.s^{-1}$ with 100 fs pulses at a repetition rate of 80 MHz. Therefore, the laser provided a power per pulse of 3.75 x 10^{-8} $J.pulse^{-1}$, which corresponded to 3.75 x 10^{5} W for a pulse considering the duration of 100 fs for a pulse. By assuming a beam cross section area of around 100 μm^2 to 1 mm^2, the total surface power of the laser was between 3.75 x 10^{13} $W.m^{-2}$ and 3.75 x 10^{11} $W.m^{-2}$, respectively. Those values, divided by the speed of light (3 x 10^{8} $m.s^{-1}$) and the permittivity of vacuum (8.85 x 10^{-12} $m^{-3}.kg^{-1}.s^{4}.A^{2}$) corresponded, after taking their root, to electric fields of 3.75 x 10^{8} $V.m^{-1}$ and 3 x 10^{7} $V.m^{-1}$, respectively, which, in atomic units, corresponded to 7.5 x 10^{-4} and 7.5 x 10^{-5} a.u. We selected electric fields with slight higher intensities to be sure that the second-order phenomena would be observed (Table S1).

Table S1 External electric fields applied on the system using the Berry phase[19,20] during the wave function optimisation and the localisation of the Wannier functions. Each of the 25 below-mentioned fields corresponded to a distinct calculation and, therefore, to a distinct value of $\mathbf{P_s}$. The collinear fields are in the left part of the table while the non-collinear ones are in the right part of the table.

	E_x (a.u.)	E_y (a.u.)	E_z (a.u.)		E_x (a.u.)	E_y (a.u.)	E_z (a.u.)
1	0.000	0.000	0.000	14	0.000	0.005	0.005
2	0.000	0.000	0.005	15	0.000	-0.005	0.005
3	0.000	0.000	-0.005	16	0.005	0.000	0.005
4	0.000	0.000	0.010	17	-0.005	0.000	0.005
5	0.000	0.000	-0.010	18	0.000	0.005	0.005
6	0.000	0.005	0.000	19	0.000	-0.005	-0.005
7	0.000	-0.005	0.000	20	0.005	0.000	-0.005
8	0.000	0.010	0.000	21	-0.005	0.000	-0.005
9	0.000	-0.010	0.000	22	0.005	0.005	0.000
10	0.005	0.000	0.000	23	-0.005	0.005	0.000
11	-0.005	0.000	0.000	24	-0.005	-0.005	0.000
12	0.010	0.000	0.000	25	0.005	-0.005	0.000
13	-0.010	0.000	0.000				

5 Classical Molecular Dynamics Simulations Results

First, the density was calculated on the last 500 ps of the *NPT* simulation, which was performed without vacuum. We found a density of 0.996 g.cm^{-3}, which was in good accordance with the experimental and theoretical values reported in the literature.[37] The vapor pressure was also calculated on the *NVT* simulation by averaging the number of water molecules in the gas phase over the last nine nanoseconds (out of ten) of the simulation, to exclude a thermalization period of one nanosecond. The ideal gas law allowed determining a vapor pressure of 3.9 kPa, which was also in reasonable accordance with the reported values of around 3.2 kPa at 25°C.[38–41] Besides, the number density distributions of the water molecules, averaged on the whole trajectory (nine nanoseconds), are presented for the two studied systems (containing 32 and 240, respectively) in Fig. S4. For the system containing 32 water molecules (Fig. S4a), clear oscillations could be observed with three maxima and a period of around 2.95 Å, comparable to the O-O distance. Consistently, those maxima corresponded to the different water molecules layers in the slab, which, considering the small number of molecules, were significantly discretised. Therefore, the water slab contained three water layers, among which only one may be considered as a bulk layer. Besides, for the system containing 240 molecules (Fig. S4b), the aforementioned oscillations were still observable albeit they displayed a significantly lower intensity and period, of around 2.75 Å. Overall, the density curve presented six distinct maxima that corresponded to six water layers, which was consistent with the fact that we doubled the number of water molecules along the z-direction compared to the previous system.

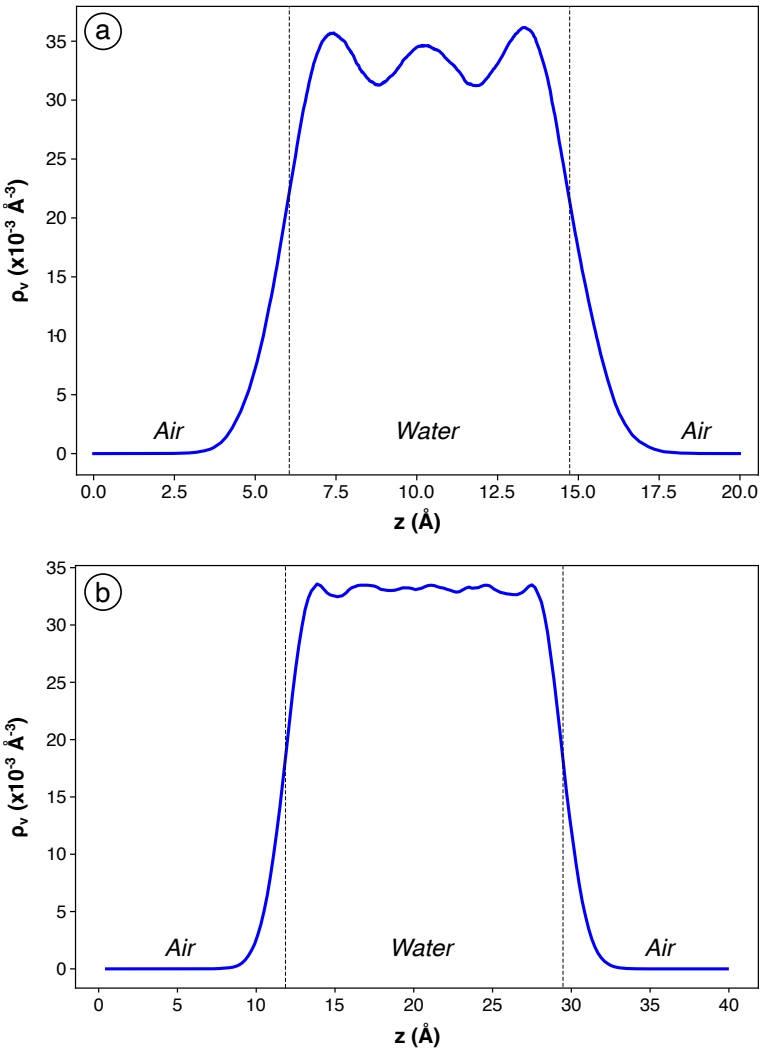

Fig. S4 Number density profiles of water molecules for the water/air interfaces containing 32 water molecules (a) and 240 water molecules (b). The dashed lines indicate the position of the Gibbs dividing surface.

To assess the accuracy of the water inter-molecular interactions, we also determined the radial distribution function for the O-O and the O-H distances (Fig. S5), for the system with 240 water molecules, which represented the best compromise between realism and computational cost. For clarity, only the inter-molecular O-H pairs were represented in Fig. S5. The $g_{OO}(r)$ displayed a first maximum at 2.75 along with a first minimum at 3.35 , which was in good agreement with both experimental and theoretical reported values.[42–44] A second smooth peak at 4.53 as well as a bulging at 6.86 could also be observed on the $g_{OO}(r)$. Besides, the $g_{OH}(r)$ exhibited a first maximum and a first minimum at 1.78 and 2.43 , respectively, with a second maximum peaks at 3.26 , which was also in good accordance with the literature.[18,42–44]

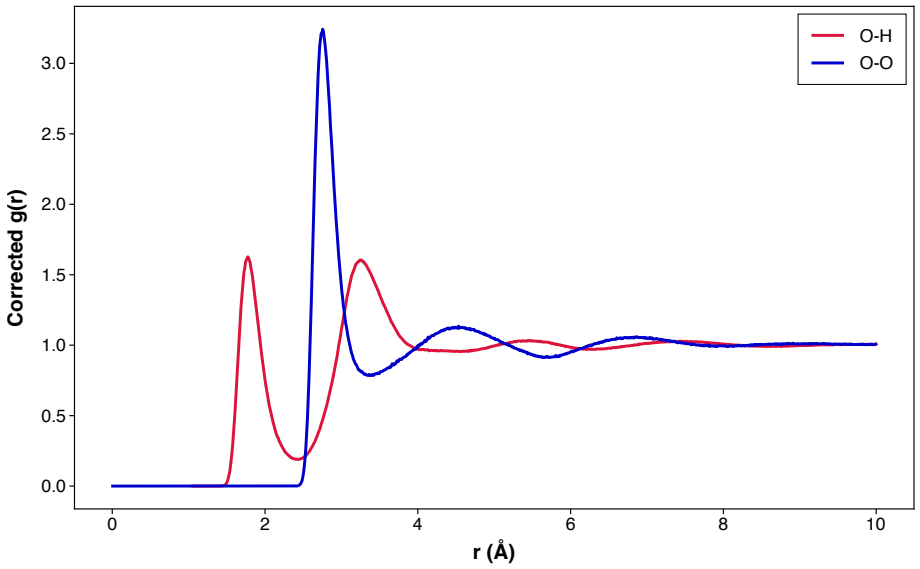

Fig. S5 O-O and inter-molecular O-H radial distribution functions in the water slab from the classical molecular dynamics simulation conducted on the cell comprising 240 water molecules. The radial distribution functions have been corrected to account for the vacuum part of the cell.

The interfacial orientation and structure are essential to characterise accurately the SHG signal since it is well admitted that the non-random molecule orientation at the interface is responsible for the SHG signal. Therefore, we calculated P_1 and P_2, which represent the first and second Legendre moment variations defined as:

$$P_1(z) = \langle \cos(\theta) \rangle \tag{S32}$$

$$P_2(z) = \frac{3}{2}\langle \cos^2(\theta) \rangle - \frac{1}{2} \tag{S33}$$

where θ is the angle between the interface normal, oriented towards the vacuum, and the water molecule dipole moment vector. The brackets mean that $\cos(\theta)$ and $\cos^2(\theta)$ were averaged over time and molecules in each 0.5 Å-thick chunk along the z-axis. θ can vary from 0° to 180°, which indicate, for the upper interface, that the molecule is oriented with its hydrogen atoms towards the vacuum or towards the water slab, respectively. P_1, which represents the average orientation of water molecules, was zero in the bulk, slightly negative near the interface, and became positive in the vacuum phase at the vicinity of the interface (Fig. S6a). This indicates that the water molecules had their hydrogen atoms oriented towards the water bulk when they were near the interface, which was in agreement with the literature[6,24]. Moreover, the positive P_1 in the vacuum phase (near the interface) could be attributed to the fact that molecules were more favourably ejected from the interface when their hydrogen atoms were oriented towards the vacuum. In the vacuum phase, P_1 exhibited a random distribution corresponding to the cosine of water molecules in the vacuum phase with, however, very few water molecules, which induced large variations. In the water bulk, P_1 was zero, which represented the average of the cosine between 0° and 180°. This was consistent with a random distribution of water

molecules orientation in the bulk phase. Besides, P_2 was zero (with slight oscillations) in the water bulk but was clearly negative on the two sides of the interface (Fig. S6b), which could be attributed to an alignment of water molecules more favourable in the plane of the interface.

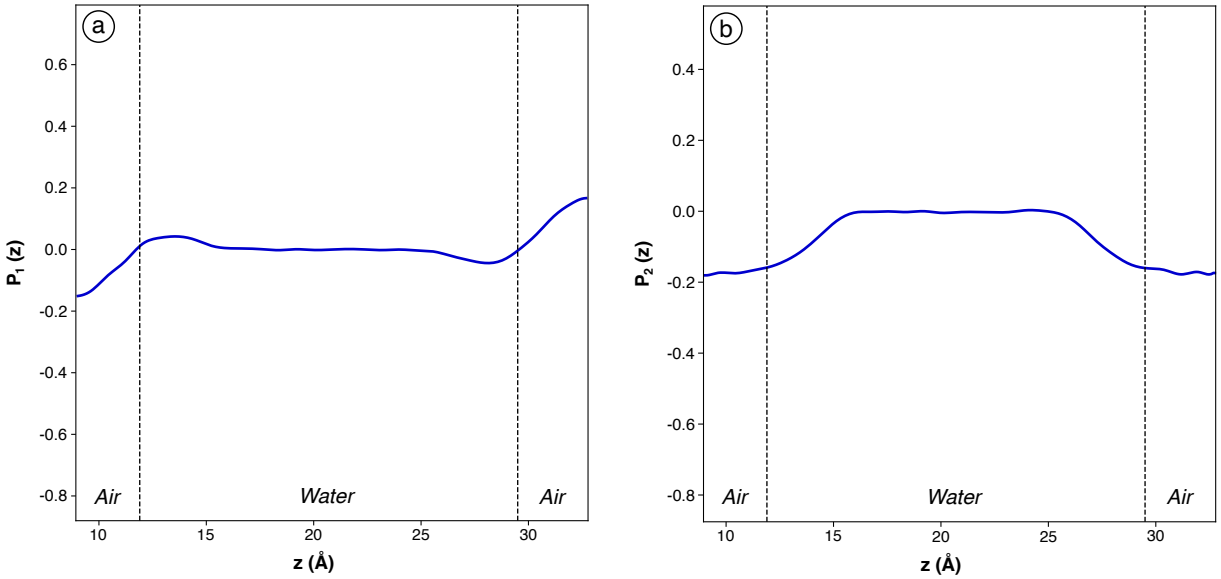

Fig. S6 The two order parameters P_1 and P_2 representing the first and second Legendre moment variations as defined in Eq. S32 and Eq. S33, respectively, from the classical molecular dynamics simulation conducted on the cell comprising 240 water molecules. The dashed lines indicate the position of the Gibbs dividing surface.

6 Variability of the SHG values and validation of the predictions

To assess the accuracy of our predictions, we compared the simulated SHG signals, calculated using the determined $\frac{\chi^{(2)}_{s,zzz}}{\chi^{(2)}_{s,xxz}}$ and $\frac{\chi^{(2)}_{s,zxx}}{\chi^{(2)}_{s,xxz}}$ ratios, to our experimental SHG signals as well as to various reported experimental SHG signals for a water/air interface, including their average. We gathered most experimental values encountered in the literature for the water/air interface (Table S2). Only the values from Goh and co-workers[29] had to be excluded since no information was available about the Fresnel coefficients that were used for the calculation of $\chi^{(2)}_{s,zzz}$, $\chi^{(2)}_{s,zxx}$, and $\chi^{(2)}_{s,xxz}$.

Since significant variations of the two considered ratios are not necessarily related to important variations of the SHG responses, we compared directly the SHG signals, calculated using Eq. S1, S2, and S3 using the ratios presented in Table S2. We then calculated the root-mean-square errors (RMSE) between our simulated SHG signals and our experimental SHG signal to assess the prediction accuracy of the developed method. We also determined the RMSE between the various reported experimental SHG signals and our experimental signals, to assess the variability of experiments from an apparatus to another (Fig. S7). For our simulated responses, RMSE can be regarded as the mean error committed on the simulated response by our prediction. On overall, the simulated responses exhibited similar RMSE with regards to our ex-

Table S2 The different experimental $\frac{\chi_{zzz}}{\chi_{xxz}}$ and $\frac{\chi_{zxx}}{\chi_{xxz}}$ ratios for a water/air interface from the literature.

$\frac{\chi^{(2)}_{s,zzz}}{\chi^{(2)}_{s,xxz}}$	$\frac{\chi^{(2)}_{s,zxx}}{\chi^{(2)}_{s,xxz}}$	Reference
3.22	0.48	Fordyce *et al.*[31]
2.56	0.30	Zhang *et al.*[34]
2.58	0.29	Pham *et al.*[6]
2.52	0.29	Wei *et al.*[45]
2.56	0.41	Tamburello-Luca *et al.*[32]
2.54	0.27	Bian *et al.*[46]
2.53	0.29	Bian *et al.*[47]
2.64	0.33	Average
3.18	0.23	This work - Experiments
2.94	0.46	This work - Simulations

perimental values than the experimental SHG responses from the literature.

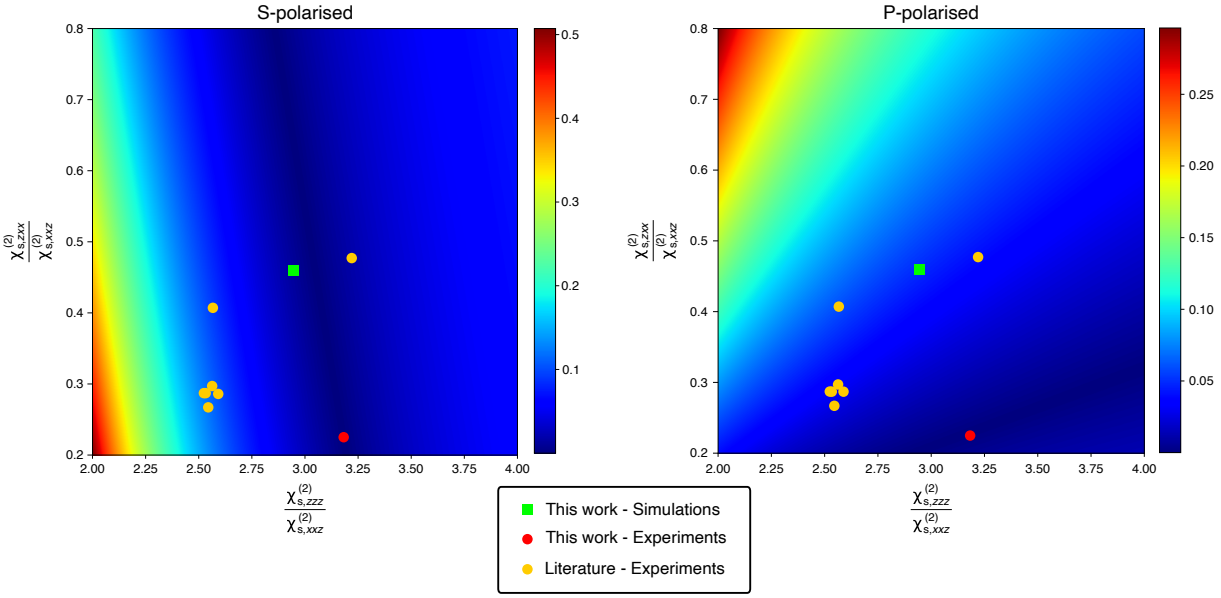

Fig. S7 RMSE between the simulated SHG signal determined in this work (green) and the experimental SHG signal measured in this work (red) and between various experimental SHG signals from literature (orange) and the experimental SHG signal measured in this work, for the P-polarised and the S-polarised curves.

7 Quantum mechanics results

7.1 Dipole moments

To assess the validity of our method, the dipole moment norms in absence of any applied electric field were plotted on a histogram (Fig. S8a). The average dipole moment norm of the water molecules, calculated on 901 snapshots, was 3.11 D, which was consistent with the values reported for bulk water in the literature.[18,48] The

dipole moments followed a Gaussian distribution, probably induced by a significant influence of the chemical environment of the water molecule, particularly the number of hydrogen bonds, on its dipole moment.[18,48] This phenomenon could also be observed on Fig. S8b: the dipole moment norms were significantly higher in the water bulk, at the centre of the slab, than near the interfaces. The water molecules located near the interfaces exhibited, indeed, less hydrogen bonds than the water molecules in the bulk, which induced lower dipole moment norms, as demonstrated in the literature.[18] The molecules going from the bulk to the vacuum phase (yellow zone in Fig. S8b) presented even lower dipole moment norms, albeit a little number of molecules were found in this configuration.

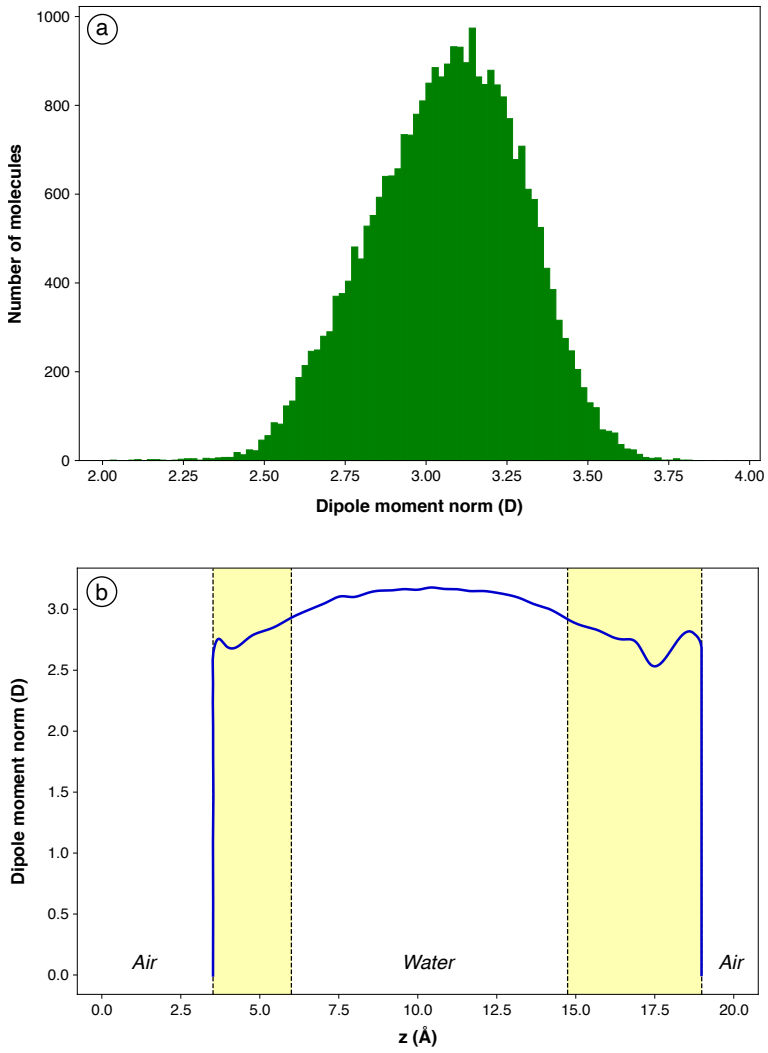

Fig. S8 Distribution of the molecular dipole moment norms (a) and average molecular dipole moment norms on 0.5Å-thick slices along the *z*-axis (b) for the cell containing 32 molecules and calculated on 901 snapshots, in absence of any applied electric field. The dipole moments were calculated using Eq. S17 using BLYP exchange-correlation functional.

7.2 Detailed results

All the determined coefficients are summarised in Table S3 for the systems comprising 32 and 240 molecules, with and without the inclusion of the quadrupole contribution. The presented values were averaged on 901 snapshots for the system containing 32 water molecules and on 144 snapshots for the system comprising 240 water molecules. All the calculations were performed using BLYP exchange-correlation functional, with a prior geometry optimisation for the system containing 32 water molecules and without any geometry optimisation for the system comprising 240 molecules. First, the values of $P_{s,x}^{(0)}$ and $P_{s,y}^{(0)}$ were significantly lower than $P_{s,z}^{(0)}$, which was consistent with the fact that, on average, $P_{s,x}^{(0)}$ and $P_{s,y}^{(0)}$ should be zero. Here, their non-zero values were due to fluctuations of the interface but could actually be neglected considering their very low values. Interestingly, the difference between $P_{s,z}^{(0)}$ and the components along *x*- and *y*-axes of $P_s^{(0)}$ increased by one order of magnitude when the quadrupole contribution was included. The values of -1.2468 x 10^{-2} a.u. and -1.1987 x 10^{-2} a.u. for $P_{s,z}^{(0)}$ of the two different cells corresponded, when divided by ε_0, to electric potentials of -1.5668 x 10^{-1} and -1.5064 x 10^{-1} a.u., respectively. When converted in SI units, those values corresponded to -4.2634 V and -4.0991 V, respectively, which was in good agreement with the values ranging from 3.5 V to 4.5 V reported in the literature for the inner potential of water at an air/water interface.[49] In particular, they were in accordance with the value of 3.87 V calculated by Remsing and co-workers[50] using DFT and with the value of 4.48 V determined experimentally by Yesibolati and co-workers.[49] The correct inner electric potential of water at an air/water interface could only be obtained by including the bulk quadrupole contribution, which therefore contributed not only in the second-order term, but also in the permanent surface polarisation $P_{s,z}^{(0)}$.

Component	32 molecules		240 molecules	
	Without Q	With Q	Without Q	With Q
$P_{s,x}^{(0)}$ (10^{-2} a.u.)	0.0230	0.0204	-0.0090	-0.0076
$P_{s,y}^{(0)}$ (10^{-2} a.u.)	-0.0340	-0.0345	0.0908	0.0919
$P_{s,z}^{(0)}$ (10^{-2} a.u.)	-0.1130	-1.2468	-0.0630	-1.1987
$\chi_{s,xx}^{(1)}$ (a.u.)	6.5900	6.5916	14.0733	14.0727
$\chi_{s,xy}^{(1)}$ (a.u.)	-0.0034	-0.0031	-0.0017	-0.0021
$\chi_{s,xz}^{(1)}$ (a.u.)	0.0015	0.0017	-0.0024	-0.0025
$\chi_{s,yx}^{(1)}$ (a.u.)	-0.0035	-0.0032	-0.0017	-0.0021
$\chi_{s,yy}^{(1)}$ (a.u.)	6.5832	6.5841	14.0508	14.0504
$\chi_{s,yz}^{(1)}$ (a.u.)	-0.0006	-0.0008	-0.0027	-0.0026
$\chi_{s,zx}^{(1)}$ (a.u.)	0.0013	0.0006	-0.0018	-0.0013
$\chi_{s,zy}^{(1)}$ (a.u.)	-0.0007	-0.0006	-0.0014	-0.0018
$\chi_{s,zz}^{(1)}$ (a.u.)	7.5921	7.5838	14.3573	14.3571
$\chi_{s,xxx}^{(2)}$ (a.u.)	0.0503	0.0452	-0.0966	-0.0929
$\chi_{s,xxy}^{(2)}$ (a.u.)	0.0030	0.0048	0.0847	0.0870
$\chi_{s,xxz}^{(2)}$ (a.u.)	-0.6963	-2.6906	-0.9763	-4.0427
$\chi_{s,xyy}^{(2)}$ (a.u.)	0.0086	0.0054	-0.0472	-0.0492
$\chi_{s,xyz}^{(2)}$ (a.u.)	-0.0021	-0.0006	0.0107	0.0094
$\chi_{s,xzz}^{(2)}$ (a.u.)	0.0071	-0.0029	-0.0126	-0.0039
$\chi_{s,yxx}^{(2)}$ (a.u.)	0.0012	0.0034	0.0886	0.0904
$\chi_{s,yxy}^{(2)}$ (a.u.)	0.0074	0.0042	-0.0452	-0.0455
$\chi_{s,yxz}^{(2)}$ (a.u.)	0.0006	0.0015	0.0148	0.0130
$\chi_{s,yyy}^{(2)}$ (a.u.)	-0.0968	-0.0942	0.2407	0.2470
$\chi_{s,yyz}^{(2)}$ (a.u.)	-0.7031	-2.7027	-0.9045	-3.9617
$\chi_{s,yzz}^{(2)}$ (a.u.)	-0.0230	-0.0300	0.0507	0.0592
$\chi_{s,zxx}^{(2)}$ (a.u.)	0.3642	-1.0464	0.5906	-1.8704
$\chi_{s,zxy}^{(2)}$ (a.u.)	-0.0019	-0.0010	0.0070	0.0062
$\chi_{s,zxz}^{(2)}$ (a.u.)	0.0038	-0.0050	-0.0047	0.0048
$\chi_{s,zyy}^{(2)}$ (a.u.)	0.3620	-1.0540	0.6219	-1.8295
$\chi_{s,zyz}^{(2)}$ (a.u.)	-0.0125	-0.0189	0.0243	0.0336
$\chi_{s,zzz}^{(2)}$ (a.u.)	-2.5750	-9.4952	-2.9815	-11.8847
$\frac{\chi_{s,zzz}^{(2)}}{\chi_{s,xxz}^{(2)}}$	3.6980	3.5290	3.0537	2.9400
$\frac{\chi_{s,zxx}^{(2)}}{\chi_{s,xxz}^{(2)}}$	-0.5230	0.3889	-0.6049	0.4627

Table S3 The independent components of $P_s^{(0)}$, $\chi_s^{(1)}$, and $\chi_s^{(2)}$ and the corresponding $\frac{\chi_{zzz}}{\chi_{xxz}}$ and $\frac{\chi_{zxx}}{\chi_{xxz}}$ ratios calculated with and without the inclusion of the quadrupole contribution ("With Q" and "Without Q", respectively)

Besides, only the diagonal terms of $\chi_s^{(1)}$, *i.e.*, $\chi_{s,xx}^{(1)}$, $\chi_{s,yy}^{(1)}$, and $\chi_{s,zz}^{(1)}$ were non-zero, which was consistent with the intrinsic definition of the first-order electric susceptibility. In addition, these latter were all positive, which was also in agreement with the physical response of water under an applied electric field. The $x \leftrightarrow y$ symmetry was verified for the two different systems, with and without the inclusion of the quadrupole contribution. On overall, the inclusion of the bulk quadrupole contribu-

tion did not affect the values of the components of $\chi_s^{(1)}$. $\chi_{s,zz}^{(1)}$ was higher than $\chi_{s,xx}^{(1)}$ and$\chi_{s,yy}^{(1)}$, particularly for the system containing 32 water molecules, which probably resulted from the small number of water molecules in bulk conditions for this system. On the cell containing 240 water molecules, this difference was lower but still present. This indicated a higher first-order electric susceptibility of water molecules located near the interface and an influence of the local chemical environment on this property. Since $\chi^{(1)}$ is a volume quantity, the surface $\chi_{s,xx}^{(1)}$, $\chi_{s,yy}^{(1)}$, and $\chi_{s,zz}^{(1)}$ allowed obtaining the electric susceptibility of bulk water, χ_{water}^{e} as follows:

$$\chi_{water}^{e} = \frac{\chi_{s,ii}^{(1)}}{\frac{1}{4}L_z} \tag{S34}$$

where L_z is the cell length along z-axis and the $\frac{1}{4}$ factor stands for the fact that $\chi_{s,ii}^{(1)}$ was calculated on the half of the water slab, itself representing half of the simulation box. For the system containing 32 molecules, we obtained $\chi_{water}^{e} \simeq 0.6972$ along x- and y-axes and $\chi_{water}^{e} = 0.8026$ along z-axis while, for the system containing 240 molecules, we obtained $\chi_{water}^{e} \simeq 0.7441$ along x- and y-axes and $\chi_{water}^{e} = 0.7597$ along z-axis. These values, particularly for the system containing 240 molecules, were in good agreement with the value of 0.77689 calculated from the refractive index of water (1.333). Finally, regardless the number of water molecules and the inclusion of the quadrupole contribution, most components of $\chi_{s,xxz}^{(2)}$ were zero or negligible compared to the non-zero components. Only $\chi_{s,xxz}^{(2)}$, $\chi_{s,zxx}^{(2)}$, $\chi_{s,yyz}^{(2)}$, $\chi_{s,zyy}^{(2)}$, and $\chi_{s,zzz}^{(2)}$ were significant, which was in accordance with the SHG theoretical background (see section 3 of SI). In addition, the $x \leftrightarrow y$ was verified for all the systems since $\chi_{s,xxz}^{(2)} \simeq \chi_{s,yyz}^{(2)}$ and $\chi_{s,zxx}^{(2)} \simeq \chi_{s,zyy}^{(2)}$ systematically. As highlighted in the main part of the article, without the inclusion of the bulk quadrupole contribution, $\chi_{s,xxz}^{(2)}$ and $\chi_{s,zzz}^{(2)}$ were negative while $\chi_{s,zxx}^{(2)}$ was positive. Nonetheless, when the quadrupole contribution was included, $\chi_{s,zxx}^{(2)}$ became negative and, therefore, all the components of $\chi_s^{(2)}$ were negative, which was in agreement with the reported observations.[25,51] Regardless the inclusion of the bulk quadrupole contribution in the fitting procedure, the Kleinman symmetry was systematically non-valid since $\frac{\chi_{s,zxx}^{(2)}}{\chi_{s,xxz}^{(2)}}$ was far from 1 (imposed by the Kleinman symmetry). Tamburello-Luca and co-workers[32] highlighted the fact that the difference between $\chi_{s,zxx}^{(2)}$ and $\chi_{s,xxz}^{(2)}$ is generally smaller for liquid-liquid interfaces than for a water-air interface and disappears totally when the dielectric permitivities of the two media are the same.[32] Therefore, as Goh and co-workers,[29] they suggested that the difference between $\chi_{s,zxx}^{(2)}$ and $\chi_{s,xxz}^{(2)}$ is mainly due to the bulk quadrupole contribution since this difference vanishes when the electric field gradient across the interface disappears.[29,32] However, following this theory, when the bulk induced quadrupole is not included in the total surface polarisation, $\chi_{s,zxx}^{(2)}$ and $\chi_{s,xxz}^{(2)}$ should be equal while it was not the case in our results (see Table S3). Our results rather were in agreement with more recent theories, which stated that the bulk quadrupole contribution is not

related to the Kleinman symmetry breaking.[25,34]

7.3 Summary of fits

The coefficient of determination, R^2, and the RMSE of the prediction of the surface polarisation were systematically calculated on each interface side of each snapshot. The RMSE can be regarded as the mean error committed on the response by the least-squares prediction. After the fitting, for each interface side of each snapshot, the 25 values of (P_x, P_y, P_z) were calculated from the 25 values of (E_x, E_y, D_z) using the parameters determined during the fitting procedure. Then, the R^2 and RMSE were calculated for P_x, P_y, and P_z by comparing the values predicted by the models to the exact values. For both interface sides and on all the considered snapshots, the R^2 was higher than 0.99. Besides, for the system containing 32 water molecules, the RMSE on P_x and P_y exhibited a clear Gaussian behaviour centred at 1.6×10^{-6} a.u while the RMSE on P_z was an asymmetric Gaussian centred at 1.1×10^{-6} a.u. (Fig. S9). The values of RMSE were an order of magnitude lower than the typical values of P_s. The residual error could originate from the third-order terms that were not considered here although they might have been present in the total surface polarisation. However, they were most probably very low compared to the zero-, first-, and second-order terms.

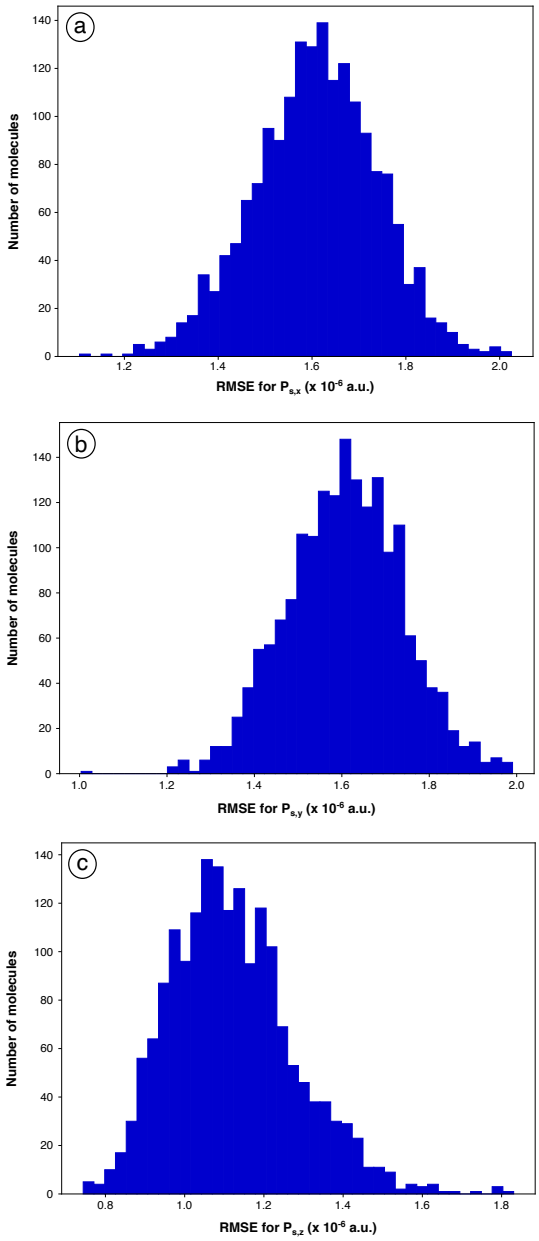

Fig. S9 Distributions of the root-mean-square errors between P_x (a), P_y (b), and P_z (c) predicted using the determined parameters and P_x, P_y, and P_z calculated using the Wannier centres (Eq. S17).

7.4 Influence of the exchange-correlation functional

The influence of the exchange-correlation functionals on the simulated SHG signal was assessed on 10 selected snapshots, for the system containing 32 water molecules, with the inclusion of the quadrupole contribution. We selected two widely used functionals in the GGA, namely, BLYP[52,53] and PBE,[54] along with one of the most employed hybrid-GGA functionals, B3LYP.[52,53,55,56] The geometry relaxation (when it was performed), the wave function optimisation, and the Wannier centres localisation were all conducted using the same functional. $\chi^{(2)}_{s,zzz}$, $\chi^{(2)}_{s,zxx}$, and $\chi^{(2)}_{s,xxz}$ were determined for each system, *i.e.*, with the three different functionals (with and without the prior geometry optimisation), and subsequently used to simulate the SHG signal. Although the ratios varied slightly between the three functionals, we observed no significant difference in terms of RMSE between the predicted values and our experimental values when the geometry was optimised prior the Wannier centres localisation (Table S4). Besides, when the geometry was not optimised, B3LYP afforded lower RMSE compared to BLYP or PBE, which is consistent with the fact that hybrid functionals generally provide more accurate results for the dipole polarisabilities.[57]

Table S4 $\frac{\chi_{zzz}}{\chi_{xxz}}$ and $\frac{\chi_{zxx}}{\chi_{xxz}}$ ratios, and the corresponding RMSE between the predicted and our experimental values as a function of the exchange-correlation functional, the prior geometry optimisation, and the number of water molecules in the system.

Functional	Geometry Opt.	Water Molecules	$\frac{\chi^{(2)}_{zzz}}{\chi^{(2)}_{xxz}}$	$\frac{\chi^{(2)}_{zxx}}{\chi^{(2)}_{xxz}}$	RMSE		
					P	S	45°
BLYP	No	32	3.3799	0.4316	0.0057	0.0944	0.0873
BLYP	No	240	3.0389	0.4737	0.0133	0.0648	0.0496
PBE	No	32	3.3788	0.4237	0.0071	0.0942	0.0879
B3LYP	No	32	3.3362	0.4150	0.0062	0.0888	0.0820
BLYP	Yes	32	3.4549	0.3729	0.0202	0.1069	0.1073
BLYP	Yes	240	2.9854	0.4448	0.0064	0.0657	0.0536
PBE	Yes	32	3.4657	0.3474	0.0218	0.1031	0.1043
B3LYP	Yes	32	3.4629	0.3649	0.0212	0.1064	0.1073
Experiments - This work			3.1781	0.2280	0	0	0

7.5 Influence of the geometry optimisation

The SPC/E model was used to describe water molecules during the CMD simulations, performed to extract snapshots for the quantum mechanical calculations. In this model, the O-H intramolecular bond length is set to 1 Å, which may introduce a lack of accuracy for the subsequent optimisation of the wave function and localisation of the Wannier centres. Therefore, on 10 snapshots, regularly extracted from the nine-nanoseconds CMD trajectory, the influence of the prior geometry optimisation on the simulated SHG response was assessed: $\chi^{(2)}_{s,zzz}$, $\chi^{(2)}_{s,zxx}$, and $\chi^{(2)}_{s,xxz}$ were determined for each system, *i.e.*, with or without a prior geometry optimisation, and subsequently used to simulate the SHG signal using Eq. S1, S2, and S3. Since the exchange-correlation functional did not affect significantly the prediction, this assessment was performed using BLYP. Besides, we considered the systems comprising 32 and 240 water molecules, to additionally assess the influence of the cell size on the effect of the geometry optimisation.

For the cell comprising 32 water molecules, the prior geometry optimisation slightly increased $\frac{\chi^{(2)}_{s,zzz}}{\chi^{(2)}_{s,xxz}}$ and decreased $\frac{\chi^{(2)}_{s,zxx}}{\chi^{(2)}_{s,xxz}}$ (Table S4). Moreover, the prior geometry optimisation increased significantly the RMSE between the predicted values and the experimental values for the P-polarised curve, *i.e.*, it decreased the quality of the prediction (Table S4), while it exhibited little influence on the 45°- and S-polarised curves. Hence, for the system comprising 32 water molecules, the geometry optimisation was not necessary and even not recommended since it decreased the quality of the prediction. Besides, for the cell containing 240 water molecules, the prior geometry optimisation decreased significantly the RMSE for the P-polarised curve while the other RMSE were roughly equivalent. Since the P-polarised curves were normalised to their maximum, this indicates a better fitting of the P-polarised curve when the geometry was optimised prior to the Wannier centres localisation. The gain of accuracy was low considering the significantly higher computational costs induced by a prior geometry optimisation. We could therefore avoid any geometry optimisation since, for the cell comprising 32 water molecules, it decreased the prediction quality and, for the cell containing 240 water molecules, the prediction accuracy increased lowly. The different influence of the prior geometry optimisation on the SHG simulated response for 32 and 240 molecules could originate from the small cell size in the case of 32 water molecules, which could introduce some inaccuracies. Indeed, we showed on the density profiles (Fig. S4) that, for the system containing 32 water molecules, only three layers of water were present, among which two were in contact with vacuum. Overall, the little influence of the geometry optimisation highlighted the fact that the second-order electric properties of water are mainly controlled by the hydrogen bonds, which are accurately described by CMD simulations.

7.6 Influence of the number of snapshots

The influence of the number of snapshots selected for the calculation of $\frac{\chi^{(2)}_{s,zzz}}{\chi^{(2)}_{s,xxz}}$ and $\frac{\chi^{(2)}_{s,zxx}}{\chi^{(2)}_{s,xxz}}$ was assessed by calculating the values of these ratios on an increasing number of

snapshots randomly selected in the pool of the regularly-extracted snapshots. For the system comprising 32 water molecules, 11 and 21 snapshots were enough to obtain values in the 95% confidence interval located around the average value for two ratios, respectively (Fig. S10). This average was calculated on 901 snapshots, which were extracted every 10 ps of the nine-nanoseconds CMD trajectory. Furthermore, 221 and 81 snapshots were required to attain the 97.5% confidence interval for the first and second above-mentioned ratios, which was considerably higher than to attain the 95% confidence interval. From 400 to 901 snapshots, the values of the two ratios converged very closely to the average values, although attaining this number of snapshots was computationally costly considering the low gain of accuracy.

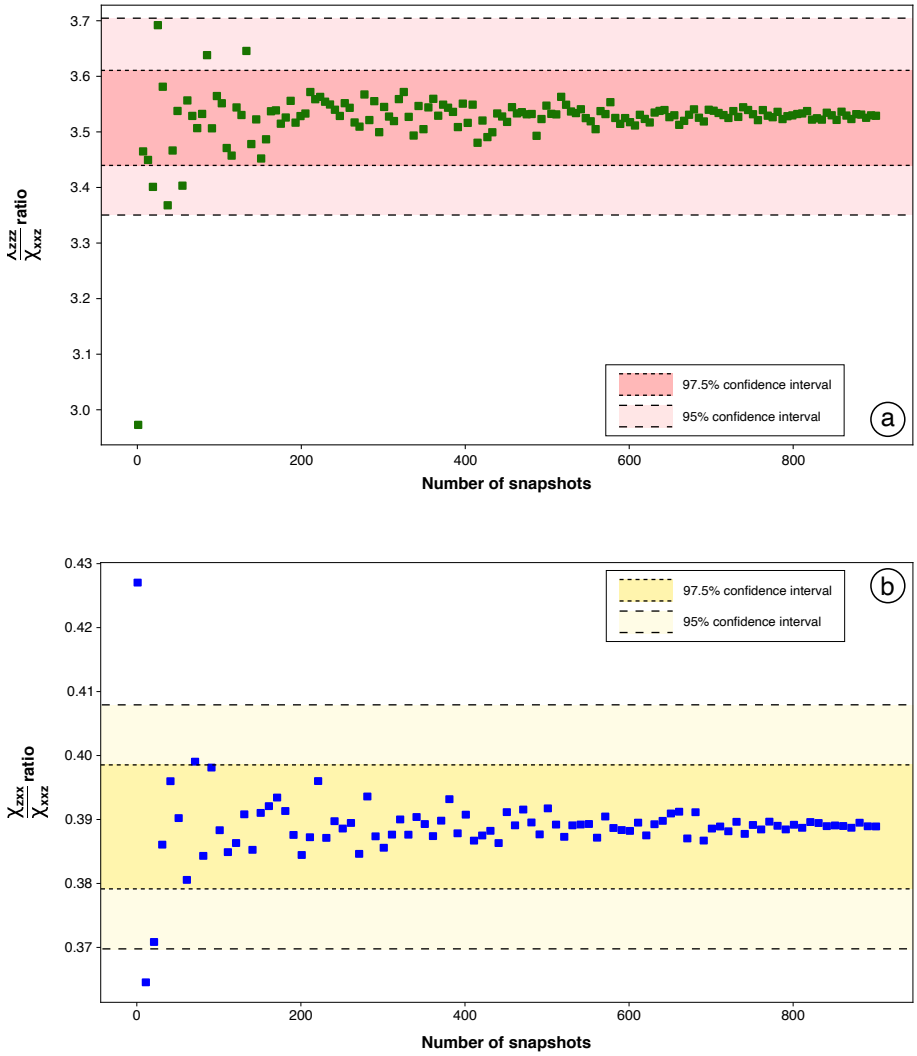

Fig. S10 $\frac{\chi^{(2)}_{s,zzz}}{\chi^{(2)}_{s,zxx}}$ (a) and $\frac{\chi^{(2)}_{s,zxx}}{\chi^{(2)}_{s,xxz}}$ (b) ratios as a function of the number of snapshots considered in the average, for the system containing 32 water molecules. The confidence intervals at 95% and 97.5% of the average value calculated on 901 snapshots are presented.

For the system containing 240 water molecules, 1 snapshot and 4 snapshots were enough to reach the 95% confidence interval (which is therefore not represented)

located around the average value for $\frac{\chi^{(2)}_{s,zzz}}{\chi^{(2)}_{s,xxz}}$ and $\frac{\chi^{(2)}_{s,zxx}}{\chi^{(2)}_{s,xxz}}$ ratios, respectively (Fig. S11). This average was calculated on 144 snapshots, which were selected every 62.5 ps of the nine-nanoseconds CMD trajectory. Besides, 34 and 13 snapshots were required to attain the 97.5% confidence interval for the two above-mentioned ratios, respectively, which was, also, considerably higher than to attain the 95% confidence interval. From 100 to 144 snapshots, the values of the ratios converged to the average value. On overall, a significant computational cost was required to increase slightly the degree of accuracy of the calculated values of the ratios and a value within the 95% confidence interval should constitute a satisfactory prediction.

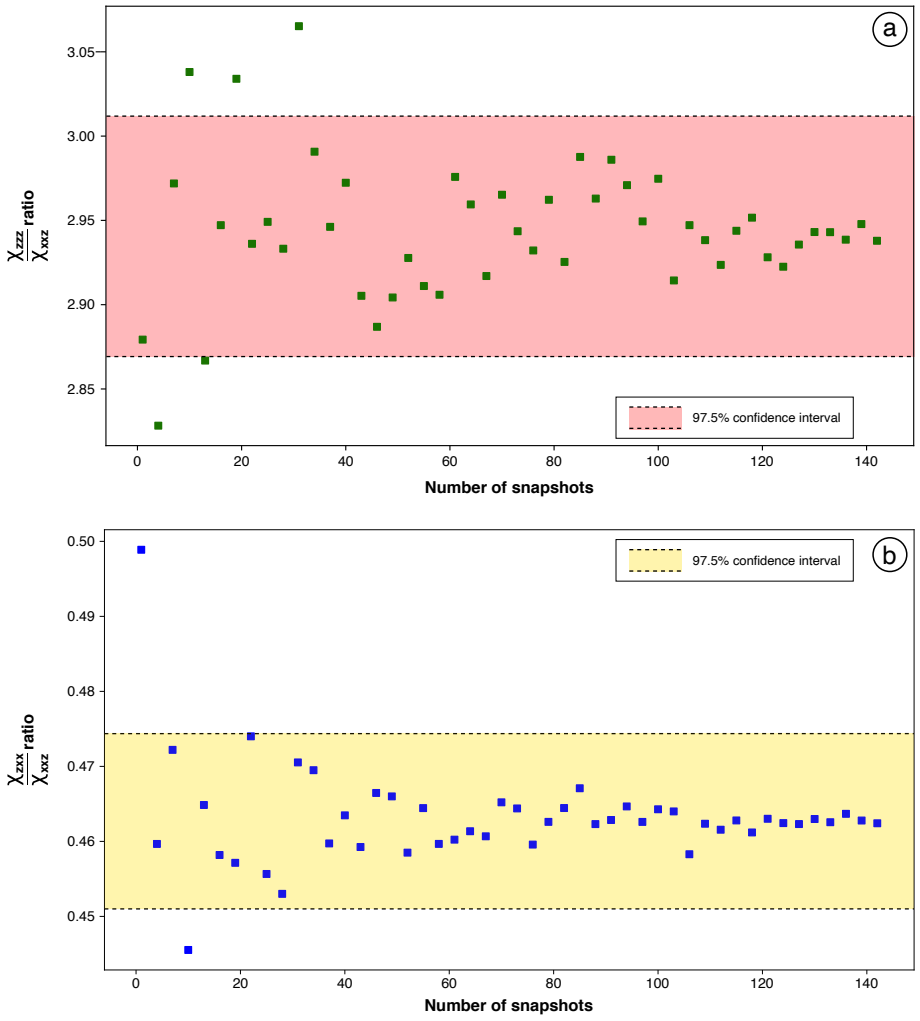

Fig. S11 $\frac{\chi^{(2)}_{s,zzz}}{\chi^{(2)}_{s,zxx}}$ (a) and $\frac{\chi^{(2)}_{s,zxx}}{\chi^{(2)}_{s,xxz}}$ (b) ratios as a function of the number of snapshots considered in the average, for the system containing 240 water molecules. The confidence interval at 97.5% of the average value calculated on 144 snapshots is presented.

7.7 Influence of the molecular referential point

In our approach, the molecular origin was used (1) to classify molecules along the z-axis to determine the molecules belonging to the upper side and those belonging to the lower side of the slab and (2) to calculate the molecular quadrupole moment (see Eq. 15 of the main part of the article). We assessed the influence of the molecular referential point on the dipole and quadrupole contributions by varying it along the bisector of the $\widehat{\mathrm{HOH}}$ angle of the water molecule, from the oxygen atom to the barycentre of the two hydrogen atoms (Fig. S12).

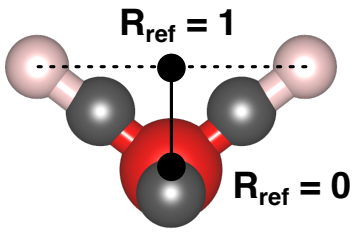

Fig. S12 Definition of the molecular referential point, *i.e.*, the reference coordinate used for classifying molecules along the z-axis and for calculating the molecular quadrupole moment. R_{Ref} corresponds to the fractional coordinate along the axis defined by the oxygen atom and the barycentre of the two hydrogen atoms. R_{Ref} here corresponds to the abscissa in Fig. S13a-d.

The influence of the molecular origin on the three non-zero independent components of $\chi_s^{(2)}$ is presented in Fig. S13a-d, calculated on 144 snapshots for the system comprising 240 water molecules. When the quadrupole contribution was not included, the z fractional coordinate of the molecular referential point exhibited a significant -roughly linear- influence on $\chi_{s,zzz}^{(2),D}$, $\chi_{s,zxx}^{(2),D}$, $\chi_{s,xxz}^{(2),D}$, as well as on $\frac{\chi_{s,zxx}^{(2),D}}{\chi_{s,xxz}^{(2),D}}$, with a slight influence on $\frac{\chi_{s,zzz}^{(2),D}}{\chi_{s,xxz}^{(2),D}}$. The three former quantities significantly increased while the two ratios decreased when the z-coordinate of the referential point was moved towards the barycentre of the two hydrogen atoms. When the quadrupole contribution was included, the influence of the referential point was decreased by two orders of magnitude and was nearly zero for the two ratios. On overall, the remaining influence was more related to physical fluctuations and did not follow anymore any trend. Therefore, the inclusion of the quadrupole contribution to the surface polarisation implied that the choice of the molecular referential point did not affect the responses (the values of the components of $\chi_s^{(2)}$ and their ratios). This was a satisfactory result since, with a physical point of view, the arbitrary choice of the molecular referential point should not influence the global responses. This constituted an additional proof to demonstrate that including the quadrupole contribution in the total surface polarisation was required to obtain physical sound results. This was in accordance with the theoretical formalism developed by Byrnes and co-workers[58], who stated that the molecular origin has no influence on $\chi_s^{(2)}$. Also, Shiratori and Morita[26] extended the

discussion of Byrnes *et al.*[58] to assess the influence of the molecular origin on each of the various dipole and quadrupole contributions to $\chi_s^{(2)}$. They demonstrated that, although the sum of the different interface contributions and the bulk contributions are invariant under the shift of the molecular origin, each interface contribution term is affected by this definition. This is also consistent with the results presented in Fig. S13.

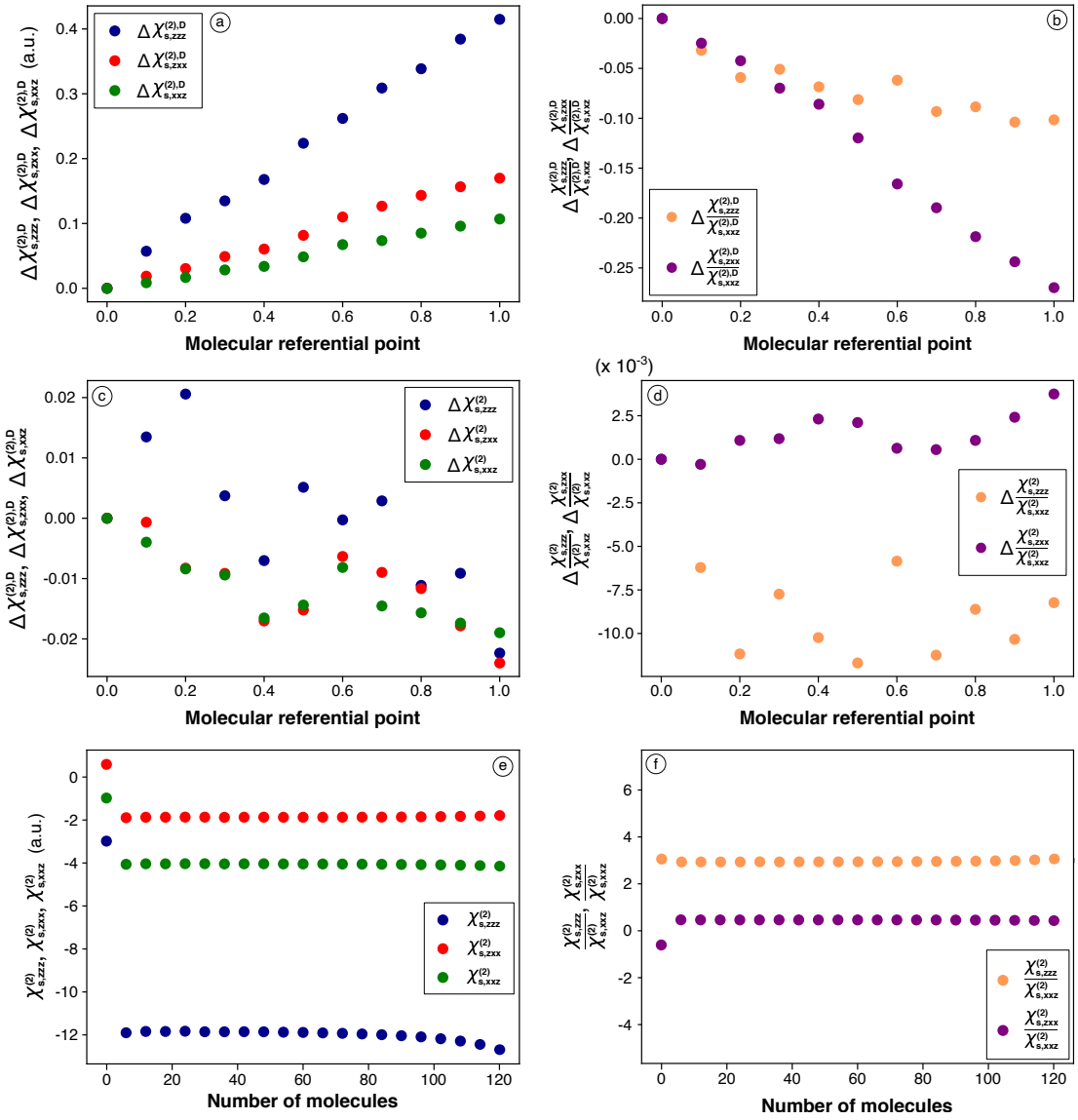

Fig. S13 Influence of the molecular referential point, on $\chi_{s,zzz}^{(2),D}$, $\chi_{s,zxx}^{(2),D}$, $\chi_{s,xxz}^{(2),D}$, $\frac{\chi_{s,zzz}^{(2),D}}{\chi_{s,xxz}^{(2),D}}$, and $\frac{\chi_{s,zxx}^{(2),D}}{\chi_{s,xxz}^{(2),D}}$ (a and b), *i.e.*, without including the quadrupole contribution, and on $\chi_{s,zzz}^{(2)}$, $\chi_{s,zxx}^{(2)}$, $\chi_{s,xxz}^{(2)}$, $\frac{\chi_{s,zzz}^{(2)}}{\chi_{s,xxz}^{(2)}}$, and $\frac{\chi_{s,zxx}^{(2)}}{\chi_{s,xxz}^{(2)}}$ (c and d), *i.e.* with including the quadrupole contribution and influence of the number of the -most buried- water molecules included in the calculation of the bulk quadrupole contribution on $\chi_{s,zzz}^{(2)}$, $\chi_{s,zxx}^{(2)}$, $\chi_{s,xxz}^{(2)}$, $\frac{\chi_{s,zzz}^{(2)}}{\chi_{s,xxz}^{(2)}}$, and $\frac{\chi_{s,zxx}^{(2)}}{\chi_{s,xxz}^{(2)}}$.

7.8 Influence of the thickness of the slice used for the bulk quadrupole contribution

The influence of the thickness of the slice (proportional to the number of water molecules) used for calculating the bulk quadrupole contribution was also assessed. This was carried out on 144 snapshots on the system comprising 240 water molecules, to have a large range of possible thicknesses. The considered molecules were the most buried molecules, *i.e.*, the closest molecules from the centre of the water slab. Fig. S13e-f presents the evolution of $\chi^{(2)}_{s,zzz}$, $\chi^{(2)}_{s,zxx}$, and $\chi^{(2)}_{s,xxz}$ as well as of $\frac{\chi^{(2)}_{s,zzz}}{\chi^{(2)}_{s,xxz}}$ and $\frac{\chi^{(2)}_{s,zxx}}{\chi^{(2)}_{s,xxz}}$ ratios as a function of the number of water molecules considered for the calculation of the bulk quadrupole contribution (following Eq. S28 and S29), from 0 to 120 molecules, this latter value corresponding to all the molecules of the interface side. Consistently, the point at 0 corresponded to no water molecule included in the quadrupole moment calculation, *i.e.*, no quadrupole contribution included in the total surface polarisation. Therefore, a significant gap was exhibited in Fig. S13e-f between the first point (that corresponded to no quadrupole contribution included) and the second point, for which 15 water molecules were comprised in the bulk quadrupole contribution. On overall, the number of water molecules considered exhibited very low influence on the three components of $\chi^{(2)}_s$ as well as on their ratios. A significant influence was observed when more than 100 molecules were considered, which corresponded to the inclusion of interfacial molecules in the calculation of the bulk quadrupole contribution and was therefore no meaningful. This result demonstrated that the water molecules displayed a similar quadrupole moment in the bulk regardless their position in the slab, as long as they were buried enough.

8 Additional References

[1] R. W. Boyd, *Nonlinear Optics*, Academic Press Inc, Amsterdam ; Boston, 3rd edn, 2008.

[2] K. B. Eisenthal, *Chemical Reviews*, 1996, **96**, 1343–1360.

[3] A. Alejo-Molina and H. Hardhienata, *IOP Conference Series: Earth and Environmental Science*, 2016, **31**, 012020.

[4] Y. R. Shen, *The principles of nonlinear optics*, Wiley-Interscience, Hoboken, N.J, Wiley classics library ed edn, 2003.

[5] D. A. Kleinman, *Physical Review*, 1962, **126**, 1977–1979.

[6] T. T. Pham, A. Jonchère, J.-F. Dufrêche, P.-F. Brevet and O. Diat, *The Journal of Chemical Physics*, 2017, **146**, 144701.

[7] P. Guyot-Sionnest, W. Chen and Y. R. Shen, *Physical Review B*, 1986, **33**, 8254–8263.

[8] P. Guyot-Sionnest and Y. R. Shen, *Physical Review B*, 1987, **35**, 4420–4426.

[9] V. Sokhan and D. Tildesley, *Molecular Physics*, 1997, **92**, 625–640.

[10] N. A. Spaldin, *Journal of Solid State Chemistry*, 2012, **195**, 2–10.

[11] R. Resta, *Europhysics Letters (EPL)*, 1993, **22**, 133–138.

[12] D. Vanderbilt and R. D. King-Smith, *Physical Review B*, 1993, **48**, 4442–4455.

[13] R. D. King-Smith and D. Vanderbilt, *Physical Review B*, 1993, **47**, 1651–1654.

[14] G. H. Wannier, *Physical Review*, 1937, **52**, 191–197.

[15] N. Marzari and D. Vanderbilt, *Physical Review B*, 1997, **56**, 12847–12865.

[16] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza and D. Vanderbilt, *Reviews of Modern Physics*, 2012, **84**, 1419–1475.

[17] P. L. Silvestrelli and M. Parrinello, *The Journal of Chemical Physics*, 1999, **111**, 3572–3580.

[18] P. L. Silvestrelli and M. Parrinello, *Physical Review Letters*, 1999, **82**, 3308–3311.

[19] D. Vanderbilt, *Berry phases in electronic structure theory*, Cambridge University Press, Cambridge, 2018.

[20] M. V. Berry, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 1984, **392**, 45–57.

[21] G. Berghold, C. J. Mundy, A. H. Romero, J. Hutter and M. Parrinello, *Physical Review B*, 2000, **61**, 10040–10048.

[22] K. LEVENBERG, *Quarterly of Applied Mathematics*, 1944, **2**, 164–168.

[23] D. W. Marquardt, *Journal of the Society for Industrial and Applied Mathematics*, 1963, **11**, 431–441.

[24] F. S. Cipcigan, V. P. Sokhan, A. P. Jones, J. Crain and G. J. Martyna, *Physical Chemistry Chemical Physics*, 2015, **17**, 8660–8669.

[25] K. Shiratori, S. Yamaguchi, T. Tahara and A. Morita, *The Journal of Chemical Physics*, 2013, **138**, 064704.

[26] K. Shiratori and A. Morita, *Bulletin of the Chemical Society of Japan*, 2012, **85**, 1061–1076.

[27] N. Bloembergen and P. S. Pershan, *Physical Review*, 1962, **128**, 606–622.

[28] N. Bloembergen, R. K. Chang, S. S. Jha and C. H. Lee, *Physical Review*, 1968, **174**, 813–822.

[29] M. C. Goh, J. M. Hicks, K. Kemnitz, G. R. Pinto, T. F. Heinz, K. B. Eisenthal and K. Bhattacharyya, *The Journal of Physical Chemistry*, 1988, **92**, 5074–5075.

[30] M. Cynthia Goh and K. B. Eisenthal, *Chemical Physics Letters*, 1989, **157**, 101–104.

[31] A. J. Fordyce, W. J. Bullock, A. J. Timson, S. Haslam, R. D. Spencer-Smith, A. Alexander and J. G. Frey, *Molecular Physics*, 2001, **99**, 677–687.

[32] A. A. T. Luca, P. Hébert, P. F. Brevet and H. H. Girault, *J. Chem. Soc., Faraday Trans.*, 1995, **91**, 1763–1768.

[33] Q. Du, R. Superfine, E. Freysz and Y. R. Shen, *Physical Review Letters*, 1993, **70**, 2313–2316.

[34] W.-k. Zhang, D.-s. Zheng, Y.-y. Xu, H.-t. Bian, Y. Guo and H.-f. Wang, *The Journal of Chemical Physics*, 2005, **123**, 224713.

[35] K.-D. Bauer and K. Hingerl, *Optics Express*, 2017, **25**, 26567.

[36] K. Matsuzaki, S. Nihonyanagi, S. Yamaguchi, T. Nagata and T. Tahara, *The Journal of Chemical Physics*, 2019, **151**, 064701.

[37] J. B. Patterson and E. C. Morris, *Metrologia*, 1994, **31**, 277–288.

[38] D. Ambrose and I. Lawrenson, *The Journal of Chemical Thermodynamics*, 1972, **4**, 755–761.

[39] *Steam tables: thermodynamic properties of water including vapor, liquid and solid phases ; English units*, ed. J. H. Keenan, Wiley, New York, 1969.

[40] N. Matsunaga and A. Nagashima, *International Journal of Thermophysics*, 1987, **8**, 681–694.

[41] S. Kurioka and K. Ikeda, *Journal of the Physical Society of Japan*, 1991, **60**, 4181–4198.

[42] G. Hura, D. Russo, R. M. Glaeser, T. Head-Gordon, M. Krack and M. Parrinello, *Physical Chemistry Chemical Physics*, 2003, **5**, 1981.

[43] A. K. Soper, *Journal of Physics: Condensed Matter*, 2007, **19**, 335206.

[44] J. Wang, G. Román-Pérez, J. M. Soler, E. Artacho and M.-V. Fernández-Serra, *The Journal of Chemical Physics*, 2011, **134**, 024516.

[45] Q. Wei, D. Zhou and H. Bian, *Physical Chemistry Chemical Physics*, 2018, **20**, 11758–11767.

[46] H.-t. Bian, R.-r. Feng, Y.-y. Xu, Y. Guo and H.-f. Wang, *Physical Chemistry Chemical Physics*, 2008, **10**, 4920.

[47] H.-t. Bian, R.-r. Feng, Y. Guo and H.-f. Wang, *The Journal of Chemical Physics*, 2009, **130**, 134709.

[48] F.-X. Coudert, F. Cailliez, R. Vuilleumier, A. H. Fuchs and A. Boutin, *Faraday Discuss.*, 2009, **141**, 377–398.

[49] M. N. Yesibolati, S. Laganà, H. Sun, M. Beleggia, S. M. Kathmann, T. Kasama and K. Mølhave, *Physical Review Letters*, 2020, **124**, 065502.

[50] R. C. Remsing, M. D. Baer, G. K. Schenter, C. J. Mundy and J. D. Weeks, *The Journal of Physical Chemistry Letters*, 2014, **5**, 2767–2774.

[51] S. Yamaguchi, K. Shiratori, A. Morita and T. Tahara, *The Journal of Chemical Physics*, 2011, **134**, 184705.

[52] C. Lee, W. Yang and R. G. Parr, *Physical Review B*, 1988, **37**, 785–789.

[53] A. D. Becke, *Physical Review A*, 1988, **38**, 3098–3100.

[54] J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, 1996, **77**, 3865–3868.

[55] A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648–5652.

[56] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *The Journal of Physical Chemistry*, 1994, **98**, 11623–11627.

[57] J. R. Hammond, N. Govind, K. Kowalski, J. Autschbach and S. S. Xantheas, *The Journal of Chemical Physics*, 2009, **131**, 214103.

[58] S. J. Byrnes, P. L. Geissler and Y. Shen, *Chemical Physics Letters*, 2011, **516**, 115–124.