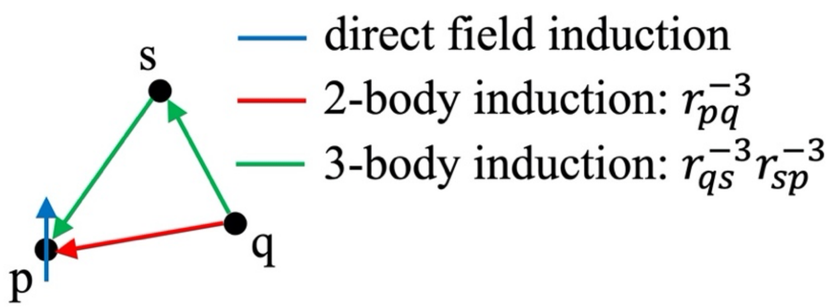
On the Cartesian representation of the molecular polarizability tensor surface by polynomial fitting to *ab initio* data

Oluwaseun Omodemi[1], Sarah Sprouse[1], Destyni Herbert[1], Martina Kaledin,[1,*] Alexey L. Kaledin[2,*]

[1] *Department of Chemistry & Biochemistry, Kennesaw State University, 370 Paulding Ave NW, Box # 1203, Kennesaw, GA 30144*

[2] *Cherry L. Emerson Center for Scientific Computation, Emory University, 1515 Dickey Drive, Atlanta, Georgia, 30322*

Abstract: We describe an approach to constructing an analytic Cartesian representation of the molecular dipole polarizability tensor surface in terms of polynomials in interatomic distances with a training set of *ab initio* data points obtained from a molecular dynamics (MD) simulation or by any other available means. The proposed formulation is based on a perturbation treatment of the unmodified point dipole polarizability model of Applequist [*J. Am. Chem. Soc.* **1972**, *94*, 2952] and is shown here to be, by construction (*i*) free of short-range or other singularities or discontinuities, (*ii*) symmetric and translationally invariant, and (*iii*) non-reliant on a body-fixed coordinate system. Permutational invariance of like nuclei is demonstrated to be readily applicable, making this approach useful for highly fluxional and reactive systems. Derivation of the method is described in detail, adding brief didactic numerical examples of H_2 and H_2O, and concluding with an MD simulation of the Raman spectrum of $H_5O_2^+$ at 300K with the polarizability tensor fitted to CCSD(T)/aug-cc-pVTZ data obtained using the HBB-4B potential [*J. Chem. Phys.* **2005**, *122*, 044308].



TOC graphics

1. INTRODUCTION

Highly accurate potential energy surface (PES) representations of small, moderate and, recently, fairly large polyatomic molecules, by way of fitting an analytic parameterized polynomial function to a set of *ab initio* data, have become routine in classical MD simulations of reaction dynamics and quantum mechanical studies of vibrational structure, provided one has access to a representative (training) set of high-level *ab initio* data.[1-20] Similarly, in pursuing studies of molecular vibrational spectroscopy, for example in the infrared regime, generation of high quality dipole moment surfaces (DMS) in terms of polynomials has also become equally routine,[21-25] given the same training set prerequisite. It has been noted that the requirement for a dipole moment training set may be at a lower level of *ab initio* theory than that of the PES.[4,26]

Extending this principle of high precision PES/DMS polynomial fitting of *ab initio* data to that of high precision polarizability tensor surface (PTS) polynomial fitting – dipole polarizability in the present treatment – has proven to be quite more challenging, if judging by the lack of reports of such in the literature. One reason for this could lie in the absence of a universal definition of dipole polarizability as a function of atomic Cartesian positions, in clear contrast to the case of a PES and a DMS, both well-defined scalar and vector quantities, respectively. Of course, it is possible to bypass this issue by either doing direct electronic structure theory dynamics, where the polarizability is well-defined, as has been done recently for Raman spectra in liquid phase using plane wave DFT,[27-29] or in special situations by treating the polarizability tensor components as scalar quantities within a molecular frame of reference and applying to them the same fitting methods as for a PES. For example, a well-utilized fit of a water molecule's polarizability tensor, with the components projected onto a body-fixed coordinate system, was done by Avila at a CCSD level with a quadruple-ζ-polarized quality basis set using a polynomial power series representation.[30] It is to be pointed out that even for a simple triatomic molecule such as H_2O, as is similarly the case for a generic polyatomic molecule, the latter technique may be problematic due to ambiguities in defining a body-fixed coordinate system. This is true specifically when like nuclei are present given that the directions of the molecule-defined space-fixed Cartesian axes become arbitrary, i.e., defined up to a sign. Thus, the sign of the in-plane off-diagonal polarizability element of H_2O may happen to depend on the input order of the two Hydrogens, physically indistinguishable, in an unphysical way. In other words, like-nuclei permutational invariance of the tensor in such a representation is not strictly enforced. Nevertheless, despite a good deal of numerical challenges, earlier and as well as more recent theoretical efforts[31-37] have demonstrated major advancement in polarizability representation at high-levels of *ab initio* theory and with impressive large-scale numerical applications.[38-40]

Another common approach to polarizability representation over the many years has been to devise approximate models based on polarizable point dipole theories,[41-43] and to parameterize them, i.e. the isotropic atomic polarizabilities, for the purpose of reproducing key molecular properties. This is currently the state of the art approach in standard polarizable force fields, and these approaches, which use classical molecular mechanics force field definitions, have gathered strong attention from the computational and modeling community.[44-49] More advanced treatments of analytic polarizability representation of complex systems beyond classical force field definitions, particularly those involving liquid water, have been reported.[38,39] Specifically, to simulate Raman spectra of bulk water, Medders and Paesani[38] used a high-level 'local' polarizability tensor function of H_2O to build a global water polarizability tensor using a many-body formulation with high level *ab initio* data. Further improvements of the point dipole model have been suggested more recently by Harczuk *et al.* who considered higher-order polarization effects to describe dipole polarizability and hyperpolarizability, with an application to large water clusters, and provided evidence of a systematically improvable polarizability model based entirely on electronic structure theory.[40] The latter work has far reaching implications for prospective development of molecular polarizability tensor representations of complex systems.

In the present work, detailed in the sections below, we present a strategy of combining the two aforementioned approaches, namely, (*i*) the polynomial power series expansion (with permutational symmetry of like nuclei) fitting to extensive high quality *ab initio* data, currently being the state-of-the-art approach for PES/DMS representation, and (*ii*) the point dipole model, in an effort to construct a continuous (differentiable everywhere), permutationally invariant polarizability tensor surface (PTS) suitable for classical molecular dynamics and quantum mechanical simulations.

2. COMPUTATIONAL METHODS

For the sake of perspective, we restate that representing a dipole, quadrupole or any higher-order non-scalar multipoles using Cartesian coordinates of atoms is formally trivial as these quantities have strict mathematical definitions.[50] The challenge with the polarizability tensor is exactly due to its definition as a linear response of the dipole to an applied electric field. This response as a function of the nuclear coordinates and molecular spatial orientation is not known *a priori* without first solving the electronic Schrodinger equation. Nevertheless, if one knows the exact scalar atomic polarizabilities α_p within the molecule, including their supposed dependence on the geometry, a closed form expression for an approximate molecular tensor $\boldsymbol{\alpha}$ may be obtained by formally applying a uniform field $\boldsymbol{F}$ and summing over the resultant induced dipoles at all the atomic sites. This is known as the polarizable point dipole model originally examined for polyatomic molecules by Applequist *et al*.[41] In this model the induced dipole at atom p is given by

$$\boldsymbol{\mu}_p = \alpha_p \left[\boldsymbol{F} - \sum_{q \neq p}^{N} \mathbf{T}_{pq} \boldsymbol{\mu}_q \right] \qquad (1),$$

where N is the number of atoms, $\boldsymbol{F}$ is the applied uniform electric field and $\mathbf{T}_{pq}$ is the dipole field 'direction cosines' tensor,[41-43]

$$\mathbf{T}_{pq} = -\frac{3}{r^5} \begin{bmatrix} x^2 - r^2/3 & xy & xz \\ yx & y^2 - r^2/3 & yz \\ zx & zy & z^2 - r^2/3 \end{bmatrix} \qquad (2)$$

with r being the p-q interatomic distance with the space-fixed Cartesian components x, y, z. In the above, $\mathbf{T}_{pp}$=0. Recasting Eq. 1 as a matrix equation, a closed form expression for the polarizability in a symmetric $3N$x$3N$ representation may be obtained as

$$\mathbf{G} = (\mathbf{A}^{-1} + \mathbf{T})^{-1} \qquad (3)$$

where $\mathbf{A}$ is a diagonal matrix of the atomic polarizabilities α_p, and $\mathbf{T}$ contains the $\mathbf{T}_{pq}$ pair blocks. Reducing $\mathbf{G}$ to a 3x3 form only involves summing over the atoms for each Cartesian pair $i,j = x, y, z$,[41-43]

$$\alpha_{ij} = \sum_{p,q}^{N} \left[\mathbf{G}_{pq}\right]_{ij} \qquad (4)$$

At this stage one may take Eqs 3-4 as a functional form for the polarizability and parameterize the diagonal of $\mathbf{A}$ using high-quality *ab initio* data, including a proper molecular geometry dependence of some sort, as will be shown below. However, the problem from a computational standpoint is that $\mathbf{G}$ has singularities, previously noted as originating from the model's neglect of electron exchange, whenever an interatomic distance r_{pq} approaches $4\left(\alpha_p \alpha_q\right)^{1/6}$.[41-43] For large systems the matrix inversion procedure becomes unmanageable, and as a result several functional fixes based on Thole's 'dipole smearing' model,[42] which requires the atomic polarizabilities α_p to be constant, have been proposed for this issue over the years.[40,41,49,51] For the purposes of the present study, we allow α_p to be functions of geometry and thus find it much more efficient to extend the old prescription of Applequist *et al*. and to approximate $\mathbf{G}$ using perturbation theory.[52] In this way the singularities are removed systematically and without loss of generality or computational flexibility.

Consider the asymptotic behavior of $\mathbf{G}$ as the molecule tends to the limit of a collection of non-interacting atoms, or $\{r_{pq}\} \to \infty$ for all *p,q*: $\mathbf{T} \to 0$ or $\mathbf{TA} \ll \mathbf{1}$, and therefore $\mathbf{G} \to \mathbf{A}$. In this limit, Eq. 3 can be expanded in the convergent series

$$\mathbf{G} = \mathbf{A} - \mathbf{ATA} + \mathbf{ATATA} - \cdots \quad (5)$$

with $\mathbf{T}$ as an off-diagonal perturbation. Truncating Eq. 5 after the *n*-th term leaves an error of the order $(\mathbf{A}^{-1} + \mathbf{T})\mathbf{G}_n - \mathbf{1} = (\mathbf{TA})^n$. A form of Eq. 5 was used by Applequist *et al*. in earlier calculations of optical rotatory parameters of adamantanes,[52] possibly as a way of making the calculations more easily tractable. However, we will show that Eq. 5 has important merits beyond purely numerical considerations.

Obviously, Eq. 5, and its contracted form in Eq. 4, is an excellent approximation to the polarizability tensor in the weak interaction limit, provided one has properly parameterized the scalar atomic polarizabilities α_p to decay to their free atom values. But what happens to Eq. 5 when chemical bonds start to form and therefore strong induction effects begin to take place? Even a high-order expansion will inevitably fail when bonds become short enough and the $\mathbf{TA} < \mathbf{1}$ condition no longer holds. Yet, if we treat Eq. 5 as simply a well-behaved, if approximate, functional form with which to represent the true *ab initio* polarizability, in the long and short interatomic distance regimes and everywhere in between, we do not need to worry about the mathematical validity of the expansion series, and only need to enforce its physical validity. To proceed, we introduce the following ansatz for a polarizability tensor function at a molecular configuration of *N* atoms in Cartesian coordinates $\mathbf{r} = (\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)$,

$$\widetilde{\mathbf{G}}_n(\mathbf{r}) = \lambda_1(\mathbf{r})\mathbf{A}(\mathbf{r}) - \lambda_2(\mathbf{r})\mathbf{A}(\mathbf{r})\mathbf{TA}(\mathbf{r}) + \lambda_3(\mathbf{r})\mathbf{A}(\mathbf{r})\mathbf{TA}(\mathbf{r})\mathbf{TA}(\mathbf{r}) - \cdots \quad (6)$$

where $\lambda_n(\mathbf{r})$ are the introduced geometry dependent correction factors. We require that $\mathbf{A}(\mathbf{r}) \to \mathbf{A}_0$ and $\lambda_n(\mathbf{r}) \to 1$ in the limit of non-interacting atoms, which also guarantees $\widetilde{\mathbf{G}}_n(\mathbf{r}) \to \mathbf{A}_0$, where $\mathbf{A}_0$ contains the free atom polarizabilities on the diagonal. In other words, whatever errors have been introduced by replacing the exact point dipole polarizability model by Eq. 5 are to be corrected by properly parameterizing the isotropic atomic polarizabilities $\mathbf{A}(\mathbf{r})$ and the correction factors $\lambda_n(\mathbf{r})$ so that *ab initio* data are reproduced with a tolerable level of accuracy. By way of fitting Eq. 6 to *ab initio* data, these corrections factors must contain information of the quantum electronic effects missing in the original point dipole model (Eqs. 1-4). Presently, we truncate the full expansion after the third term and let $\lambda_1(\mathbf{r}) = 1$ producing a third order approximation to the exact expansion.

We continue by 'contracting' Eq. 6 to its 3x3 form,

$$\alpha_{ij}^{(3)} = \delta_{ij} \sum_{p} \alpha_p - \lambda_2 \sum_{p \neq q} \alpha_p \alpha_q T_{pq}^{ij} + \lambda_3 \sum_{p,q} \alpha_p \alpha_q \sum_{s(\neq p,q)} \alpha_s \sum_{k=x,y,z} T_{ps}^{ik} T_{sq}^{kj} \quad (7)$$

with the superscript (3) indicating the three-body level of dipole-dipole inductions, and indices *p,q,s* representing the atomic sites and *i,j,k* their Cartesian components. Namely, the leading term is the sum of configuration-dependent isotropic atomic polarizabilities (first order effects); the second term is the sum over all direct 2-body dipole-dipole inductions (second order effects); and the third term is the sum over the 2-body and 3-body induced dipole-dipole inductions (third order effects). Note that contribution to polarizability's anisotropy starts with the 2-body term. The higher order terms, if included, would increase the flexibility of the model by contributing 4-,5-,...

body inductions and adding the extra parameters to the fit, but are likely to be progressively less important than the three leading terms due to the inverse cubic distance scaling of $\mathbf{T}$. Importantly, inspection of Eq. 7 reveals that $\alpha_{ij}^{(3)}$ is symmetric and invariant under translation since the T_{pq}^{ij} matrices possess both these properties. Moreover, it has a physically correct long-range behavior and is otherwise well-behaved everywhere in the interatomic configuration space.

All electronic structure calculations reported below were done with Gaussian16[53] and MOLPRO 2019.2[54] suites.

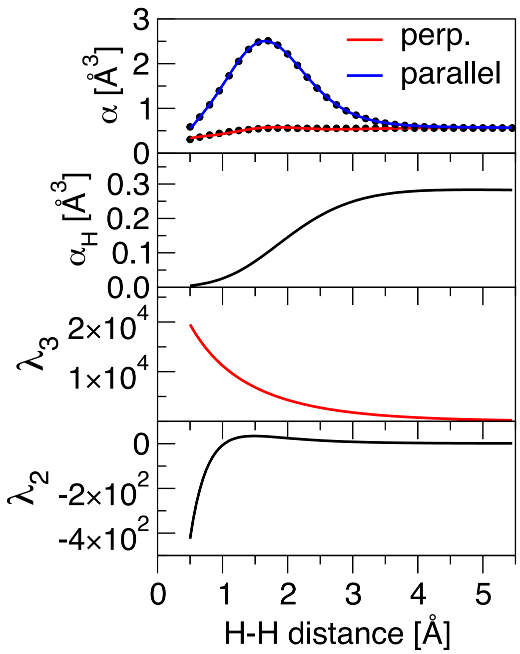


Figure 1. A fit of Eq. 6 to CCSD/cc-pVTZ data (shown by dots) for the parallel and perpendicular components of the H_2 polarizability tensor using a polynomial of 7th power. The experimental values for parallel and perpendicular components are 0.93 and 0.71 Å^3, respectively.[55] The isotropic polarizability α_H and the scaling factors λ_n are also shown. A set of 21 linear parameters was used in the fit. The absolute and relative RMS of the fit are 0.015 Å^3 and 1.4%, respectively.

3. RESULTS

3.1. H_2 polarizability. A good illustration of Eq. 7 functional form is a special case of a homonuclear diatomic, e.g. H_2, with bond distance *r* and oriented along the z-axis for clarity of presentation,

$$\alpha_{xx}^{(3)}(r) = \alpha_{yy}^{(3)}(r) = 2\alpha_{H}(r) - \lambda_2(r)\frac{\alpha_{H}^{2}(r)}{r^3} + 2\lambda_3(r)\frac{\alpha_{H}^{3}(r)}{r^6} \qquad (8a)$$

$$\alpha_{zz}^{(3)}(r) = 2\alpha_{H}(r) + 2\lambda_2(r)\frac{\alpha_{H}^{2}(r)}{r^3} + 8\lambda_3(r)\frac{\alpha_{H}^{3}(r)}{r^6} \qquad (8b)$$

where $\alpha_{H}(r)$, $\lambda_2(r)$ and $\lambda_3(r)$ are the distance dependent isotropic polarizability, the second-order and the third-order scaling functions, respectively. Note that spatial orientation of H_2 is completely arbitrary since all six independent components of the polarizability tensor are fit in the same manner to the corresponding *ab initio* data with the final result independent of the choice of axes, as we describe below for a non-trivial example. Both the dipole-dipole and dipole-induced-dipole contributions to the diagonal elements show up with the proper respective long-range distance dependence, r^{-3} and r^{-6}. Elsewhere, the expressions in Eq. 8 are continuous and well-behaved, except at the united atom limit, which is naturally avoided due to the high potential energy. If one takes a polynomial representation, aside from a constant, for instance $\alpha_{H}(r) = \alpha_{H}^{0} + c_{11}y + \cdots$, $\lambda_2(r) = 1 + c_{21}y + \cdots$, $\lambda_3(r) = 1 + c_{31}y + \cdots$, with $y = e^{-r/r_0}$, the transformed internuclear distance used throughout this paper, and r_0 being the range parameter and α_{H}^{0} the free Hydrogen polarizability, one only needs to fit the parameters to a set of *ab initio* data to have a properly behaving polarizability function both near the equilibrium and at long range. Presently we used 100 CCSD/cc-pVTZ points on a uniform grid with the H atom polarizability α_{H}^{0}=0.279 Å^3. The range parameter was r_0 = 1.2 Å. Convergence was achieved with 21 parameters and the polynomial power of 7th order, as can be seen in Figure S1. The final results are shown in Figure 1. The H-H distance dependence of the two tensor components is simple. The long range tails between 3 and 6 Å are described with little variation in the α_H function and the λ_2, λ_3 scaling factors, that is, the intrinsic form of Eq. 7 is a very good approximation to H_2 polarizability in this range (cf. Figure S2). At shorter range α_H tends to small values to offset the inverse distance spike, while the scaling factors change rapidly, especially in the repulsive wall region $r < 0.75$ Å, to regulate the functional forms. Additional discussion of the performance of various orders of approximation for H_2 can be found in the SI.

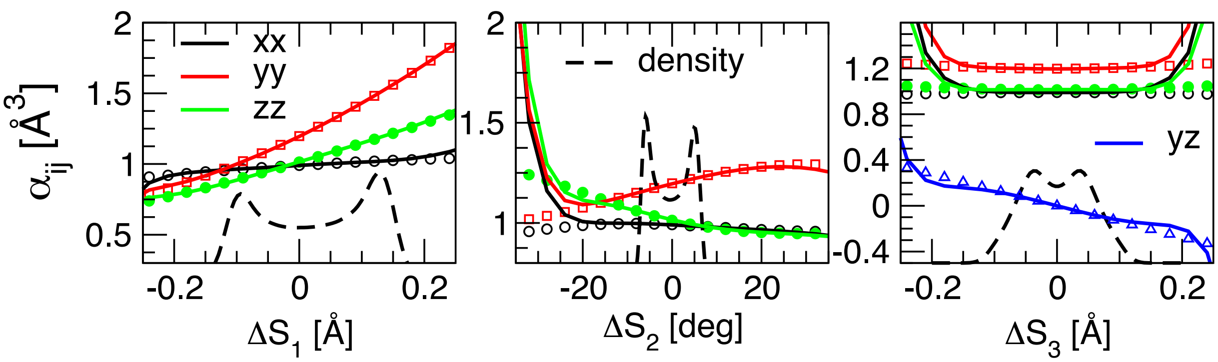


Figure 2. A test of the water molecule's polarizability tensor components (thick solid lines) as a function of the symmetry displacement coordinates, $\Delta S_1/\Delta S_3$ symmetric/antisymmetric stretch and ΔS_2 bend, from the equilibrium geometry of OH=0.9648 Å, HOH=103.9 deg. Circles, squares and triangles of corresponding color represent the testing set of *ab initio* points. The dashed line is the coordinate distribution density of the training set. In this example, H_2O lies in the YZ plane with Z bisecting the HOH angle. The experimental values for α_{xx}, α_{yy} and α_{zz} are 1.415, 1.528 and 1.468 Å^3, respectively.[56] The fit to *ab initio* data was generated using a 5th power invariant/covariant polynomial function with 128 linear coefficients on a training set of 5000 structures generated by propagating a classical trajectory at the total energy of 4669 cm^{-1}, corresponding to the harmonic ZPVE. The level of theory for the training and testing sets is B3LYP/6-31++G(d,p). The RMS error of the training fit is ~0.004 Å^3, or 0.3% in relative value. The corresponding testing set RMS values are 0.078 Å^3 and 11%.

3.2. Extension to high dimensions. In the general case of many dimensions, three atoms and more, the choice of fitting functions for α_p and λ_n is arbitrary, but presently we find it instructive that both the (*i*) like nuclei permutational symmetry and (*ii*) asymptotic behavior of weakly interacting atoms be enforced. Following the well-established Braams-Bowman formalism,[25] we express the atomic polarizabilities and the correction factors using polynomials in internuclear distances (see above), as

$$\alpha_p(\mathbf{r}) = \alpha_p^{(0)} + \sum_{m=1}^{M_1} c_m^{(g_p)} u_{p,m}^{\mathrm{cv}}\{y_{ab}\} \quad (9a)$$

$$\lambda_n(\mathbf{r}) = 1 + \sum_{m=1}^{M_2} c_{n,m}^{(\lambda)} u_m^{\mathrm{iv}}\{y_{ab}\} \quad (9b)$$

where $u_{p,m}^{\mathrm{cv}}$ and u_m^{iv} are the covariant and invariant polynomial combinations, respectively, that assure that the total polarizability does not change upon a permutation of any pair of like nuclei. Presently, we use a straightforward symmetrization scheme of common power polynomial terms although the more computationally advanced treatment using the invariant polynomial theory[25,26] may also be applied. The linear expansion coefficients for the atomic polarizabilities $c_m^{(g_p)}$ are the same for atoms within same nuclear group g_p. As a side note, we remind that in the like-atom permutational space, all like-nucleus atomic isotropic polarizabilities must transform as effective charges do in the commonly accepted dipole representation ansatz,[25] thus, covariantly symmetrized polynomials in Eq. 9a are employed. However, unlike in the dipole representations that use effective charges and are constrained to have the correct behavior for the dipole moment under translation, there are no additional constraints imposed on Eq. 9.

Despite the linear dependence of the isotropic atomic polarizabilities and scale factors on the expansion coefficients, the molecular polarizability tensor depends on them in a non-linear way, requiring use of function minimization methods while leading to possibilities of unstable solutions or stable but multiple local solutions. This is the main bottleneck of the approach compared to the trivially solvable linear least squares problem encountered in PES/DMS fitting.[25] To mitigate these issues to some degree, the unknown coefficients $\{c_m\}$ are searched for by a large-scale L-BFGS minimization engine[57] applied to a least-squares function, with the starting point of a non-interacting system of atoms (**c** = 0). To further improve the performance of this approach, i.e. to shorten computation time of iterations necessary for convergence, we evaluate the least-squares function along with its gradient using shared-memory parallelization. (See SI for additional details.)

3.3. Application to H_2O. A common, yet highly relevant polyatomic example to test the performance of our approach on is a water molecule, for which many models and fits have been published with various degrees of accuracy,[39] as discussed in the Introduction. Notwithstanding its simplicity, H_2O has a like-nuclei permutational symmetry which has been explicitly considered for the potential energy and dipole moment representations,[4] and here this property will be incorporated in the polarizability representation. Strictly for demonstration purposes, we consider making a fit to a training set of *ab initio* points obtained at a low-level of electronic structure theory, and secondly the configurations are to be sampled directly from 'laboratory frame' 3N Cartesian coordinates instead of internal coordinates. To this end we run a 20 ps long direct trajectory with the total energy of 4669 cm^{-1} (the harmonic ZPE at the B3LYP/6-31++G(d,p) level of theory) and a time step of 1 fs to generate 5000 geometries, sampled every 4th step, at which the full polarizability tensor is calculated. This training set should sample configurations energetically accessible in MD simulations at most classical conditions. After several exploratory calculations, a close fit to this training set was achieved with fifth order polynomials for Eqs. 9a and 9b, containing 32 terms for each of α_H, α_O, λ_1 and λ_2 resulting in a total of 128 independent terms. For the free atom polarizabilities we found it more efficient to use $\alpha_O^{(0)} = 3.04$ Å^3 from O^{2-} calculated at B3LYP/6-31++G(d,p), and $\alpha_H^{(0)} = 0$ from H^+, since the training set does not contain configurations with radical fragments. Similar treatment of atomic polarizabilities was done by Stillinger and David for H_2O.[58] We note that inclusion of the diffuse functions into the basis set has a profound effect on the O^{2-} ion polarizability, increasing it by a factor of ~10. The highest polynomial power of Eq. 7 in this representation is therefore 20. With the range parameter r_0 = 1.5 Å , a value within the range suggested in previous studies,[25] the absolute and relative RMS errors of the fit are 0.004 $Å^3$ and 0.3%, respectively. The SI contains additional details of the fitting.

In Figure 2 we examine fidelity of the fit by using a testing set of points that were not included in the training set. The testing set is simply a scan along the symmetry coordinates about the equilibrium configuration and sampling regions away from the training set. The latter is demarcated with the coordinate distribution function in each of the cuts. One can see that the fit goes through the points nearly exactly within the density boundaries, which is expected. Outside the density boundaries, along the symmetric stretch (ΔS_1) the fit stays very close to the data in both directions of the stretch until some deviation begins near -0.25 Å where the untrained OH repulsive wall regions are being sampled. Understandably, there is almost no deviation in the ΔS_1 outer direction since the model is designed to do better in the atomization limit. Along the antisymmetric stretch (ΔS_3) the deviations outside the training set are more pronounced for three of the four non-zero tensor components, α_{xx} , α_{yy} and α_{zz}, in part because the trained region is narrower than in ΔS_1. The α_{yz} component remains close to the data well outside the training set. Along the bending coordinate (ΔS_2) the fit performs very well beyond the trained region in the angle-opening direction ($\Delta S_2 > 0$), apparently due to weak dependence of the polarizability on H-H induction. In the angle closing direction outside the trained region, the agreement is very good for the three non-zero components until ΔS_2 = -20 degrees, or H-H distance of ~1.3 Å. Overall, the above test of the trained model for H_2O appears to show very good results beyond the ZPE energies sampled by the trajectory used for training, as is clearly illustrated in Figure S5. This is an encouraging result moving forward to more challenging applications. (See Section S-3 for definitions of H_2O coordinates.)

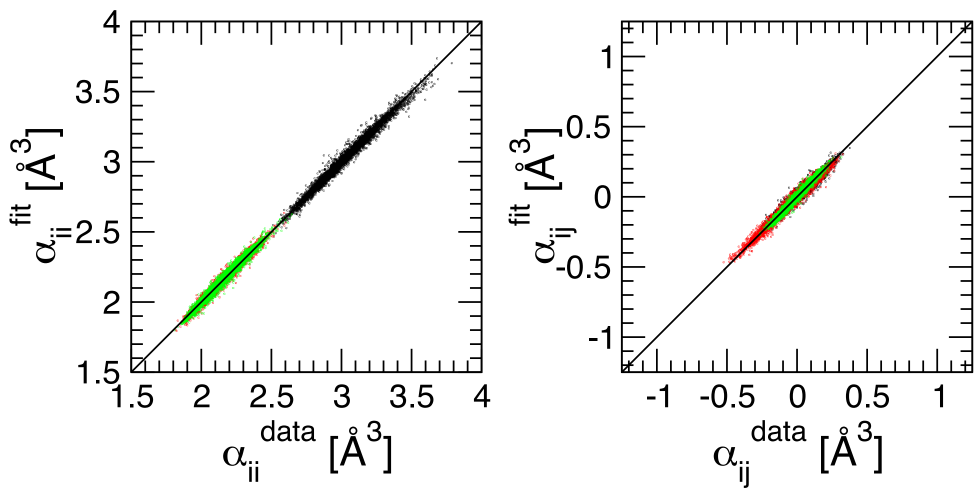


Figure 3. A fit of the full 6-th order polarizability tensor of $H_5O_2^+$, plotted on a 2.5 Å^3 range for both the diagonal and off-diagonal components, with a total of 1120 variable coefficients, to the CCSD(T)/aug-cc-pVTZ(O:*spd*,H:*sp*) training set data described in Section 3.4. The three diagonal elements xx, yy, zz and the three unique off-diagonal elements xy, xz and yz are shown using respective black, red and green color codes. The fit has a 0.037 Å^3 absolute RMS error (defined by Eq. S1) and a 1.2% relative RMS error (defined by equation S2).

3.4. Application to $H_5O_2^+$. To further demonstrate practical applicability of the proposed approach on a large system, one without a currently published polarizability hypersurface, we have intentionally taken a system with fluxional nuclei, $H_5O_2^+$, and for which also a high-quality analytic PES is available (HBB-4B).[26] This system has a number of low isomerization barriers, among which a shared proton exchange transition state has the energy E_{ex}~4000 cm^{-1}.[26,59] It has been noted that incorporation of permutational symmetry into the fit is critical for a proper description of the potential and the dipole functions in the vicinity of this transition state.[26]

Here, we shall limit the fit to be useful for classical MD simulations of Raman spectra for temperatures up to 1000 K. (Higher temperatures, and especially quantum applications will require additional extensive training of the fit that is out of the scope of the present work.) To this end, we generate a training set by propagating three 100 ps trajectories with a 1 fs time step, using the HBB-4B PES, with the total energies of 1000 cm^{-1} (low E), 4000 cm^{-1} (medium E) and 10000 cm^{-1} (high E). These trajectories sample configurations near the global minimum, the low energy isomerization transition states, the HH exchange barrier and higher energy stationary points, but just below the H_3O^+ + H_2O dissociation limit of ~12000 cm^{-1}.[26] For example, the high energy trajectory was found to contain 6 events of shared proton exchange with the outer hydrogens of both oxygens. We proceed to prune each set of the trajectories by taking every 80-th (low), 40-th (medium) and 16-th (high) point, respectively. This creates a training set of 10000 configurations $\{\mathbf{r}_n\}$ at which we calculate the polarizability tensor $\alpha_{ij}(\mathbf{r}_n)$ using central differences of the dipole moment. The energy distribution of the pruned set is shown in Figure S6. We use the CCSD(T) level of theory[60] as implemented in MOLPRO, same as the one used for generating the HBB-4B PES but in conjunction with a moderately reduced basis set, namely, aug-cc-pVTZ(O:*spd*,H:*sp*), with the f and d functions removed from O and H, respectively. (We note that calculation of polarizability by finite field dipole derivatives requires six calculations of the energy gradient at the linear response CCSD(T) level, as described in the MOLPRO manual.) Basis set reduction was found to produce negligible errors in the polarizability (0.014 Å^3 or 0.3% error calculated for $H_5O_2^+$ at its global minimum configuration), relative to the full aug-cc-pVTZ basis, while substantially reducing the computation time (see SI for details). The training set generation was done in 'parallel' by splitting the 10000 points into several blocks.

Table 1. Comparison of the non-zero polarizability tensor components, in Å^3, at several critical structures on the $H_5O_2^+$ HBB-4B surface for the fit (**PTS6**) against the *ab initio* values (**CCSD**) calculated at the CCSD(T)/aug-cc-pVTZ(O:*spd*,H:*sp*) level of theory. The number of imaginary frequencies N_{imag} and the energies E (cm^{-1}), reported previously by Huang *et al.*,[26] are shown for convenience. In the present definition, the x-axis is aligned with the O-O axis.

	N_{imag}	E	α_{xx}	α_{yy}	α_{zz}	α_{xy}
C2-MIN-CCSD	0	0	2.978	2.058	2.089	-0.123
C2-MIN-PTS6			2.981	2.073	2.077	-0.111
Cs-INV-CCSD	1	164	2.972	2.066	2.063	0.055
Cs-INV-PTS6			2.975	2.061	2.061	0.057
C2h-Trans-CCSD	1	213	2.985	1.911	2.223	-0.117
C2h-Trans-PTS6			2.983	1.878	2.270	-0.084
C2v-Cis-CCSD	1	434	2.999	2.225	1.905	0
C2v-Cis-PTS6			3.009	2.235	1.896	0
D2d-CCSD	2	524	2.997	2.044	2.044	0
D2d-PTS6			2.999	2.035	2.035	0
D2h-CCSD	3	918	2.994	2.212	1.859	0
D2h-PTS6			3.007	2.224	1.834	0
Cs-HH-CCSD	1	3944	2.750	2.158	2.136	-0.083
Cs-HH-PTS6			2.768	2.147	2.048	-0.069
C2v-CCSD	2	5042	2.781	2.142	2.099	0
C2v-PTS6			2.742	2.109	2.119	0
C2v-CCSD	3	6303	2.703	2.282	1.982	0
C2v-PTS6			2.752	2.209	2.005	0

For the fit we generated a full sixth order polynomial (**PTS6**) using r_0 = 1.5 Å with a total of 1120 independent terms with the highest power of 24. In this representation, 280 terms were used for each of the α_O, α_H, λ_2, and λ_3 functions, as shown by Eq. 9. The free atom polarizabilities were defined by treating H as H^+ and O as a closed shell O^{2-} resulting in $\alpha_H^{(0)} = 0$ and $\alpha_O^{(0)} = 5.5$ Å^3, the latter calculated at the CCSD(T)/aug-cc-pVTZ(O:*spd*) level of theory. Minimization of the least squares function yielded a fit with a 0.037 Å^3 RMS, which is a 1.2% relative error (see SI for the details). We note that choosing the free atom polarizabilities corresponding to H(^{2}S) and O(^3P_g) produced a fit of a very similar quality. In fact, a few other limiting cases for $\alpha_H^{(0)}$ and $\alpha_O^{(0)}$ were considered with all of them leading to RMS of 1.2-1.3%, suggesting that the polynomial order, i.e.

the number of variable coefficients, is more important than the choice of the isolated atom constant. A visual representation of the quality of the fit is shown in Figure 3 and also given in Table 1, where we show a comparison of the CCSD(T)/aug-cc-pVTZ(O:*spd*,H:*sp*) and the fitted polarizabilities at several critical points on the $H_5O_2^+$ HBB-4B surface, which incidentally are not included in the training set. One may see a close agreement of the fit with the *ab initio* data even for the structures that are seldom sampled in the training set, *i.e.*, the high energy **C2v** stationary points with 2 and 3 imaginary frequencies. For additional illustration of the **PTS6** surface, we show polarizability elements along various internal coordinate displacements in Figures S7-S9.

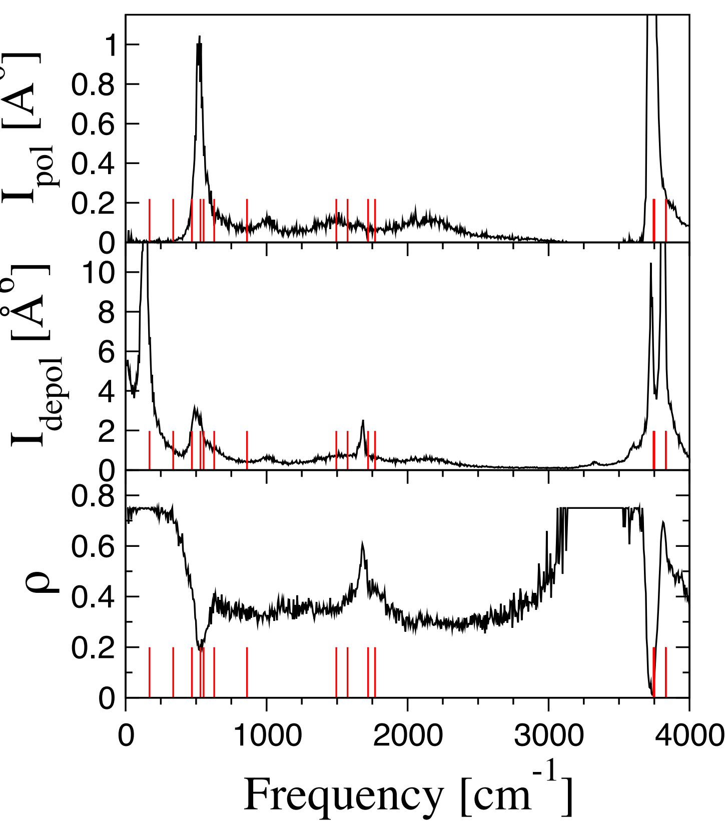


Figure 5. Raman spectra of $H_5O_2^+$ at 300K, the polarized (I_{pol}) and depolarized (I_{depol}) components, calculated using HBB-4B PES and the present **PTS6**. The harmonic frequencies are given as red sticks. The depolarization ratio is defined as the ratio of perpendicular to parallel scattered light (see SI for details).

3.5. Calculation of $H_5O_2^+$ Raman spectra. To culminate the discussion, we probe the fitted polarizability tensor hypersurface of $H_5O_2^+$ by simulations of Raman spectra. Following the usual prescription of decomposing the full tensor into a spherical part $\overline{\alpha} \equiv (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})/3$ and a traceless anisotropic part $\beta_{ij} \equiv \alpha_{ij} - \delta_{ij}\overline{\alpha}$, for *i,j=x,y,z*, the respective polarized and depolarized components of the Raman spectrum are given by[55]

$$I_{\text{pol}}(\omega) = \frac{1}{\pi}\int_0^{\infty} dt\, e^{-i\omega t}\langle\overline{\alpha}(0)\overline{\alpha}(t)\rangle \qquad (10a)$$

$$I_{\text{depol}}(\omega) = \frac{1}{\pi}\int_0^{\infty} dt\, e^{-i\omega t}\langle\sum_{ij}\beta_{ij}(0)\beta_{ij}(t)\rangle, \quad i,j = x,y,z \qquad (10b)$$

We run extensive MD simulations using the HBB-4B $H_5O_2^+$ PES. The polarizability data are calculated using **PTS6** and are post-processed by ensemble averaging of the bracketed correlation functions in Eq. 10. Details of the MD simulations, as well as additional data from double harmonic approximation, can be found in the SI. Figure 5 shows the Raman spectra computed at 300K. A thorough discussion of $H_5O_2^+$ vibrational spectra using the HBB-B4 PES is reported in an earlier publication,[61] thus here we limit ourselves to a qualitative discussion of the Raman activity. One can identify three main regions of activity in both polarized and depolarized scattering regimes: the OH stretch region on 3600-4000 cm^{-1}, the shared proton region near 1500-1800 cm^{-1}, and the low frequency region 0-800 cm^{-1}. There is a consistent correspondence between the MD Raman activity and the positions of the harmonic frequencies. However, one can see that the shared proton features extend substantially to the higher frequency region (a broad structure on 2000-2200 cm^{-1}) in the MD spectrum, that is, well beyond the harmonic limit. Similarly, there is a well pronounced activity at 1000 cm^{-1} in the MD spectrum, previously assigned to a resonance between the shared proton O-H-O stretch and a combination band involving one quantum of the O-O stretch and two quanta of the water wag,[61] a strongly non-harmonic feature. Inspection of the depolarization ratio sheds some light onto the symmetry properties. We note that the low frequency modes, those below ~500 cm^{-1}, i.e. the floppy torsional-motion-coupled local minima, display a steady ratio of about $\rho \simeq 0.75$, a strong indication of their non-totally-symmetric nature. As the frequency increases beyond 500 cm^{-1}, the ratio drops to $\rho \simeq 0.4$, an indication of more symmetric motion, e.g. the O-O stretch vibration. One can identify, however, spikes in the depolarization ratio, corresponding to the out-of-phase H_2O bend near 1700 cm^{-1} and the out-of-phase components of the OH quartet at ~3800 cm^{-1}. The symmetric component of the OH group, at 3700 cm^{-1}, is identifiable by the sharp dip in the ratio. (The high value of the ratio in the 3000-3500 cm^{-1} region is provisionally attributed to the spurious amplitudes in the Fourier transform of the polarized signal.)

4. SUMMARY

We have described a novel method for fitting a molecular dipole polarizability tensor surface to an extensive set of *ab initio* data. To represent the tensor in 3*N* Cartesians we expand the well-known point dipole model of Applequist *et al.* in a Taylor series and then express two types of scalar quantities in terms of polynomials in internuclear distances, namely the isotropic atomic polarizabilities α_p and the scaling factors λ_n using conventional function fitting methods. In this model, permutational invariance of like nuclei, Hermitian symmetry and translational invariance of the polarizability tensor are all included by construction. An exploratory test fit (with a ~0.004 Å^3 or 0.3% RMS error) produced for H_2O using low level *ab initio* data and a 5th order polynomial function exhibits a high degree of fidelity, suggesting the present model is both mathematically flexible and physically meaningful for the description of molecular dipole

polarizability. Further, we examined the utility of the present method on a more challenging $H_5O_2^+$ system. Using a previously published HBB-4B analytical PES (CCSD(T)/aug-cc-pVTZ quality), we generated an extensive training set at which we computed polarizability tensor also at the CCSD(T)/aug-cc-pVTZ level of theory. A fit of the PTS using a 6th order polynomial (1120 independent terms) was achieved with a ~0.037 Å^3 or 1.2% RMS error. Extensive MD simulations for $H_5O_2^+$ Raman spectra at 300K are presented, which to the best of our knowledge is for the first time in the literature at such a high level of electronic structure theory.

Extending the present approach of PTS representation to (*i*) describing dissociation limits, (*ii*) quantum vibrational applications and (*iii*) strong field driven MD and Stimulated Raman Excitation simulations[62] is straightforward and is only contingent on the quality of the training set and the order of the fitting polynomial. Fortunately, fitting of PES is computationally a much easier task, and for many relevant gas phase molecules high quality analytic potential surfaces are either available or may be calculated as needed using new robust machine learning technology,[11,14,19,20] which allows one to readily generate appropriate high quality training sets for PTS fitting. Further improvements of the approach include examination of prospective usefulness of higher order expansions in the Applequist model (Eq. 6) of $\tilde{\mathbf{G}}_n(\mathbf{r})$ for $n > 3$, and incorporation of Harczuk's *et al*. method[40] to account for local (atomic) hyperpolarizability.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. CHE–1855583. This work was also supported in part by research computing resources and technical expertise via a partnership between Kennesaw State University's Office of the Vice President for Research and the Office of the CIO and Vice President for Information Technology.[63]

ALK acknowledges the use of computational resources of the Cherry L. Emerson Center.

ASSOCIATED CONTENT

Supporting Information

Least-squares optimization procedure; details of H_2 and H_2O polarizability representation; training sets used for $H_5O_2^+$ polynomial fitting; details of electronic structure and MD simulations of $H_5O_2^+$.

This information is available free of charge via the Internet at http://pubs.acs.org

AUTHOR INFORMATION

Corresponding Authors

*E-mail: mkaledin@kennesaw.edu (MK)

*E-mail: akaledi@emory.edu (ALK)

ORCID

Alexey L. Kaledin; http://orcid.org/0000-0003-3112-3989

Martina Kaledin; http://orcid.org/0000-0003-1763-3552

Sarah Sprouse; https://orcid.org/0000-0001-6367-3067

Oluwaseun Omodemi; https://orcid.org/0000-0002-6489-271X

REFERENCES

(1) Xie, Z.; Braams, B. J.; Bowman, J. M. *Ab initio* global potential-energy surface for $H_5^+ \rightarrow H_3^+ + H_2$. *J. Chem. Phys.* **2005**, *122*, 224307:1-9.

(2) Park, W. K.; Park, J.; Park, S. C.; Braams, B. J.; Chen, C.; Bowman, J. M. Quasiclassical trajectory calculations of the reaction $C+C_2H_2 \rightarrow$ *l*-C_3H, *c*-C_3H+H, C_3+H_2 using full-dimensional triplet and singlet potential energy surfaces. *J. Chem. Phys.* **2006**, *125*, 081101:1-5.

(3) Jin, Z.; Braams, B. J.; Bowman, J. M. An *ab Initio* Based Global Potential Energy Surface Describing $CH_5^+ \rightarrow CH_3^+ + H_2$. *J. Phys. Chem.* A **2006**, *110*, 1569-1574.

(4) Huang, X.; Braams, B. J.; Bowman, J. M. *Ab Initio* Potential Energy and Dipole Moment Surfaces of $(H_2O)_2$. *J. Phys. Chem.* A **2006**, *110*, 445-451.

(5) Huang, X.; Braams, B. J.; Bowman, J. M.; Kelly, R. E. A.; Tennyson, J.; Groenenboom, G. C.; van Der Avoird, A. New *ab initio* potential energy surface and the vibration-rotation-tunneling levels of $(H_2O)_2$ and $(D_2O)_2$. *J. Chem. Phys.* **2008**, *128*, 034312:1-9.

(6) Huang, X.; Braams, B. J.; Carter, S.; Bowman, J. M. Quantum Calculations of Vibrational Energies of $H_3O_2^-$ on an *ab Initio* Potential. *J. Am. Chem. Soc.* **2004**, *126*, 5042-5043.

(7) Christoffel, K. M.; Jin, Z.; Braams, B. J.; Bowman, J. M. Quasiclassical Trajectory Study of the $CH_3^+ + HD \rightarrow CH_2D^+ + H_2$ Reaction. *J. Phys. Chem.* A **2007**, *111*, 6658-6664.

(8) Shepler, B. C.; Braams, B. J.; Bowman, J. M. Quasiclassical Trajectory Calculations of Acetaldehyde Dissociation on a Global Potential Energy Surface Indicate Significant Non-transition State Dynamics. *J. Phys. Chem.* A **2007**, *111*, 8282-8285.

(9) Sharma, A. R.; Braams, B. J.; Carter, S.; Shepler, B. C.; Bowman, J. M. Full-dimensional ab initio potential energy surface and vibrational configuration interaction calculations for vinyl. *J. Chem. Phys.* **2009**, *130*, 174301:1-9.

(10) Heazlewood, B. R.; Jordan, M. J. T.; Kable, S. H.; Selby, T. M.; Osborn, D. L.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. Roaming is the dominant mechanism for molecular products in acetaldehyde photodissociation. *Proc. Nat. Acad. Sci.* **2008**, *105*, 12719-12724.

(11) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. Δ-machine learning for potential energy surfaces: A PIP approach to bring a DFT-based PES to CCSD(T) level of theory. *J. Chem. Phys.* **2021**, *154*, 051102:1-8.

(12) Dawes, R.; Thompson, D. L.; Guo, Y.; Wagner, A. F.; Minkoff, M. Interpolating moving least-squares methods for fitting potential energy surfaces: Computing high-density potential energy surface data from low-density ab initio data points. *J. Chem. Phys.* **2007**, *126*, 184108:1-11.

(13) Dawes, R.; Passalacqua, A.; Wagner, A. F.; Sewell, T. D.; Minkoff, M.; Thompson, D. L. Interpolating moving least-squares methods for fitting potential energy surfaces: Using classical trajectories to explore configuration space. *J. Chem. Phys.* **2009**, *130*, 144107:1-9.

(14) Jiang, B.; Li, J.; Guo, H. Potential Energy Surfaces From High Fidelity Fitting of Ab Initio Points: The Permutation Invariant Polynomial-Neural Network Approach. *Int. Rev. Phys. Chem.* **2016**, *35*, 479-506.

(15) Unke, O. T.; Meuwly, M. Toolkit for the Construction of Reproducing Kernel-Based Representations of Data: Application to Multidimensional Potential Energy Surfaces.
J. Chem. Inf. Model. **2017**, *57*, 1923-1931.

(16) Manzhos, S.; Wang, X. G.; Dawes, R.; Carrington, T. A Nested Molecule-Independent Neural Network Approach for High-Quality Potential Fits. *J. Phys. Chem. A* **2006**, *110*, 5295-5304.

(17) Manzhos, S.; Carrington, T. Using redundant coordinates to represent potential energy surfaces with lower-dimensional functions. *J. Chem. Phys.* **2007**, *127*, 014103:1-10.

(18) Bowman, J. M.; Carrington, T.; Meyer, H. D. Variational quantum approaches for computing vibrational energies of polyatomic molecules. *Mol. Phys.* **2008**, *106*, 2145-2182.

(19) Meuwly, M. Machine Learning for Chemical Reactions. *Chemical Reviews* **2021**, *121*, 10218-10239.

(20) Manzhos, S.; Carrington, T. Neural Network Potential Energy Surfaces for Small Molecules and Reactions. *Chemical Reviews* **2021**, *121*, 10187-10217.

(21) Brown, A.; McCoy, A. B.; Braams, B. J.; Jin, Z.; Bowman, J. M. Quantum and classical studies of vibrational motion of CH_5^+ on a global potential energy surface obtained from a novel *ab initio* direct dynamics approach. *J. Chem. Phys.* **2004**, *121*, 4105-4116.

(22) Huang, X.; McCoy, A. B.; Bowman, J. M.; Johnson, L. M.; Savage, ; Dong, F.; Nesbitt, D. J. Quantum Deconstruction of the Infrared Spectrum of CH_5^+. *Science* **2006**, *311*, 60-63.

(23) McCoy, A. B.; Braams, B. J.; Brown, A.; Huang, X.; Jin, Z.; Bowman, J. M. *Ab Initio* Diffusion Monte Carlo Calculations of the Quantum Behavior of CH_5^+ in Full Dimensionality. *J. Phys. Chem. A* **2004**, *108*, 4991-4994.

(24) Huang, X.; Johnson, L. M.; Bowman, J. M.; McCoy, A. B. Deuteration Effects on the Structure and Infrared Spectrum of CH_5^+. *J. Amer. Chem. Soc.* **2006**, *128*, 3478-3479.

(25) Braams. B. J.; Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577-606.

(26) Huang, X.; Braams, B. J.; Bowman, J. M. *Ab initio* potential energy and dipole moment surfaces for $H_5O_2^+$. *J. Chem. Phys.* **2005**, *122*, 044308:1-12.

(27) Luber, S.; Iannuzzi, M.; Hutter, J. Raman spectra from *ab initio* molecular dynamics and its application to liquid S-methyloxirane. *J. Chem. Phys.* **2014**, *141*, 094503:1-9.

(28) Putrino, A.; Parrinello, M. Anharmonic Raman Spectra in High-Pressure Ice from *Ab Initio* Simulations. *Phys. Rev. Lett.* **2002**, *88*, 176401:1-4.

(29) Pagliai, M.; Cavazzoni, C.; Cardini, G.; Erbacci, G.; Parrinello, M.; Schettino, V. Anharmonic Infrared and Raman Spectra in Car–Parrinello Molecular Dynamics Simulations. *J. Chem. Phys.* **2008**, *128*, 224514:1-7.

(30) Avila, G. *Ab initio* Dipole Polarizability Surfaces of Water Molecule: Static and Dynamic at 514.5 nm. *J. Chem. Phys.* **2005**, *122*, 144310:1-10.

(31) Dykstra, C. E. Efficient calculation of electrically based intermolecular potentials of weakly bonded clusters. *J. Comput. Chem*. **1988**, 9, 476–487.

(32) Laidig, K. E.; Bader, R. F. W. Properties of atoms in molecules: Atomic polarizabilities. *J. Chem. Phys*. **1990**, 93, 7213–7224.

(33) Stout, J. M.; Dykstra, C. E. A Distributed Model of the Electrical Response of Organic Molecules. *J. Chem. Phys A*. **1998**, 102, 1576–1582.

(34) Kirtman, B.; Dykstra, C. E.; Champagne, B. Major intermolecular effects on nonlinear electrical response in a hexatriene model of solid state polyacetylene. *Chem. Phys. Lett*. **1999**, 305, 132–138.

(35) Gagliardi, L.; Lindh, R.; Karlström, G. Local properties of quantum chemical systems: The LoProp approach. *J. Chem. Phys*. **2004**, 121, 4494–4500.

(36) Guillaume, M.; Champagne, B. Modeling the electric field third-order nonlinear responses of an infinite aggregate of hexatriene chains using the electrostatic interaction model. *PCCP*. **2005**, 7, 3284-3289.

(37) Macchi, P.; Krawczuk, A. The polarizability of organometallic bonds. *Comput. Theor. Chem*. **2015**, 1053, 165–172.

(38) Medders, G. R.; Paesani, F. Infrared and Raman Spectroscopy of Liquid Water through "First-Principles" Many-Body Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1145−1154.

(39) Babin, V.; Leforestier, C.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *J. Chem. Theory Comput.* **2013**, *9*, 5395−5403.

(40) Harczuk, I.; Vahtras, O.; Ågren, H. Hyperpolarizabilities of extended molecular mechanical systems. *Phys. Chem. Chem. Phys.* **2016**, *18*, 8710-8722.

(41) Applequist, J.; Carl, J. R.; Fung, K.-K. An Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and Determination of Atom Polarizabilities. *J. Am. Chem. Soc.* **1972**, *94*, 2952-2960.

(42) Thole, B. T. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem. Phys.* **1981**, *59*, 341-350.

(43) van Duijnen, P. T.; Swart, M. Molecular and Atomic Polarizabilities: Thole's Model Revisited. *J. Phys. Chem. A* **1998**, *102*, 2399-2407.

(44) Caldwell, J. W.; Kollman, P. A. Structure and Properties of Neat Liquids Using Nonadditive Molecular Dynamics: Water, Methanol, and N-Methylacetamide. *J. Phys. Chem.* **1995**, *99*, 6208-6219.

(45) Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. Combined fluctuating charge and polarizable dipole models: Application to a five-site water potential function. *J. Chem. Phys.* **2001**, *115*, 2237-2251.

(46) Burnham, C. J.; Xantheas, S. S. Development of transferable interaction models for water. III. Reparametrization of an all-atom polarizable rigid model (TTM2–R) from first principles. *J. Chem. Phys.* **2002**, *116*, 1500-1510.

(47) Burnham, C. J.; Xantheas, S. S. Development of transferable interaction models for water. IV. A flexible, all-atom polarizable potential (TTM2-F) based on geometry dependent charges derived from an *ab initio* monomer dipole moment surface *J. Chem. Phys.* **2002**, *116*, 5115-5124.

(48) Burnham, C. J.; Anick, D. J.; Mankoo, P. K.; Reiter, G. F. The vibrational proton potential in bulk liquid water and ice. *J. Chem. Phys.* **2008**, *128*, 154519:1-20.

(49) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933-5947.

(50) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, 1996.

(51) Burnham, C. J.; Li, J. C.; Xantheas, S. S.; Leslie, M. The parametrization of a Thole-type all-atom polarizable water model from first principles and its application to the study of water clusters (n=2–21) and the phonon spectrum of ice Ih. *J. Chem. Phys.* **1999**, *110*, 4566-4581.

(52) Applequist, J.; Rivers, P.; Applequist, D. E. Theoretical and Experimental Studies of Optically Active Bridgehead-Substituted Adamantanes and Related Compounds. *J. Amer. Chem. Soc.* **1969**, *91*, 5705-5711.

(53) Gaussian 16, Revision A.03, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar,

J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

(54) MOLPRO, version 2019.2, a package of ab initio programs, H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. J. Bennie, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, S. J. R. Lee, Y. Liu, A. W. Lloyd, Q. Ma, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, T. F. Miller III, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and M. Welborn, see https://www.molpro.net.

(55) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.

(56) Murphy, W. F. The Rayleigh depolarization ratio and rotational Raman spectrum of water vapor and the polarizability components for the water molecule. *J. Chem. Phys.* **1977**, *67*, 5877-5882.

(57) Liu, D.; Nocedal, J. On the Limited Memory BFGS Method for Large Scale Optimization. *Mathematical Programming B* **1989**, *45*, 503-528.

(58) Stillinger, F. H.; David, C. W. Polarization model for water and its ionic dissociation products. *J. Chem. Phys.* **1978**, *69*, 1473-1484.

(59) Wales, D. J. Rearrangements and tunneling splittings of protonated water dimer. *J. Chem. Phys.* **1999**, *110*, 10403-10409.

(60) Hampel, C.; Peterson, K. A.; Werner, H.-J. A comparison of the efficiency and accuracy of the quadratic configuration interaction (QCISD), coupled cluster (CCSD), and Brueckner coupled cluster (BCCD) methods. *Chem. Phys. Lett.* **1992**, *190*, 1-12.

(61) Kaledin, M.; Kaledin, A. L.; Bowman, J. M.; Ding, J.; Jordan, K. D. Calculation of the Vibrational Spectra of $H_5O_2^+$ and Its Deuterium-Substituted Isotopologues by Molecular Dynamics Simulations. *J. Phys. Chem. A* **2009**, 72671-7677.

(62) Pierre-Jacques, D.; Tyler, C.; Dyke, J.; L. Kaledin, A. L.; Kaledin, M. A polarizability driven ab initio molecular dynamics approach to stimulating Raman activity: Application to C_{20}. *Mol. Phys.* **2021**, *119*, e1939453:1-9.

(63) T. Boyle and A. Ramazan, "Kennesaw State University HPC Facilities and Resources" Digital Commons Training Materials, 10 (2021). https://digitalcommons.kennesaw.edu/training/10.

Supplementary Information

On the Cartesian Representation of the Molecular Polarizability Tensor Surface by Polynomial Fitting to *Ab Initio* Data

Oluwaseun Omodemi[1], Sarah Sprouse[1], Destyni Herbert[1], Martina Kaledin[1,*], Alexey L. Kaledin[2,*]
[1] *Department of Chemistry & Biochemistry, Kennesaw State University, 370 Paulding Ave NW, Box # 1203, Kennesaw, GA 30144*
[2] *Cherry L. Emerson Center for Scientific Computation, Emory University, 1515 Dickey Drive, Atlanta, Georgia, 30322*

S-1. Least-squares optimization

We define the least-squares function for polarizability fitting, taken from Eq. 7, in the following way,

$$W(\mathbf{c}) = \frac{1}{2K}\sum_{k=1}^{K}\sum_{i,j=1}^{3}\left(\alpha_{ij}^{(3)}(\mathbf{r}_k;\mathbf{c}) - \alpha_{ij}^{\mathrm{DFT}}(\mathbf{r}_k)\right)^2 \qquad (S.1)$$

where k runs over the total of K configurations; i, j are the polarizability tensor's Cartesian components X, Y, Z. "DFT" refers to the data obtained in an electronic structure calculation, which may be of any level of theory, via trajectory propagation or any other comparable sampling methods. In this form the coefficients of all six independent tensor components are simultaneously optimized. This formulation is advantageous. The absolute RMS error of the fit is $\sqrt{W}$, and the relative RMS error is defined as

$$\mathrm{RMS}(\%) = 100\left[\frac{\sum_{k=1}^{K}\sum_{i,j=1}^{3}\left(\alpha_{ij}^{(3)}(\mathbf{r}_k;\mathbf{c}) - \alpha_{ij}^{\mathrm{DFT}}(\mathbf{r}_k)\right)^2}{\sum_{k=1}^{K}\sum_{i,j=1}^{3}\left(\alpha_{ij}^{\mathrm{DFT}}(\mathbf{r}_k)\right)^2}\right]^{\frac{1}{2}} \qquad (S.2)$$

To carry out an actual optimization, we differentiate W with respect to $\mathbf{c}$ analytically and pass both the function and its gradient

$$\nabla W(\mathbf{c}) = \frac{1}{K}\sum_{k=1}^{K}\sum_{i,j=1}^{3}\left(\alpha_{ij}^{(3)}(\mathbf{r}_k;\mathbf{c}) - \alpha_{ij}^{\mathrm{DFT}}(\mathbf{r}_k)\right)\frac{\partial\alpha_{ij}^{(3)}(\mathbf{r}_k;\mathbf{c})}{\partial\mathbf{c}} \qquad (S.3)$$

to the BFGS optimizer. With ∇W being highly non-linear, in a difficult case a typical number of iterations to reach a tight convergence, e.g. RMS(%) < 1, can be $O(10^5)$ with CPU times measured in days on a single processor workstation. Obviously, this is a major bottleneck when the total polynomial order and size of the training set need to be increased. However, noticing that both W and ∇W can be calculated by partitioning the training set K into any number of independent blocks, we apply the usual MPI 'send-receive' parallelization routine to Eqs. S.1 and S.3 and nest it inside the main BFGS driver. With the overhead of a BFGS step being small, the parallelization is highly scalable, permitting us to run efficient optimizations on a 56-core CPU Intel Xeon "Gold 6132" 2.6GHz node.

S-2. Details of H_2 fitting

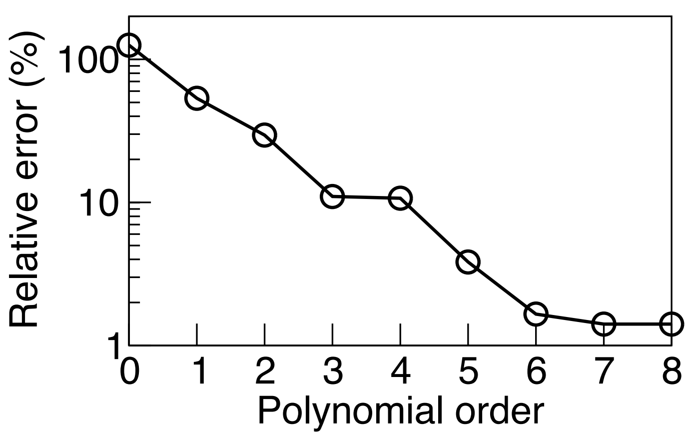


Figure S1. Convergence of the relative fitting error for H_2 polarizability tensor with the highest order of the Morse-type polynomial in Eq. 6.

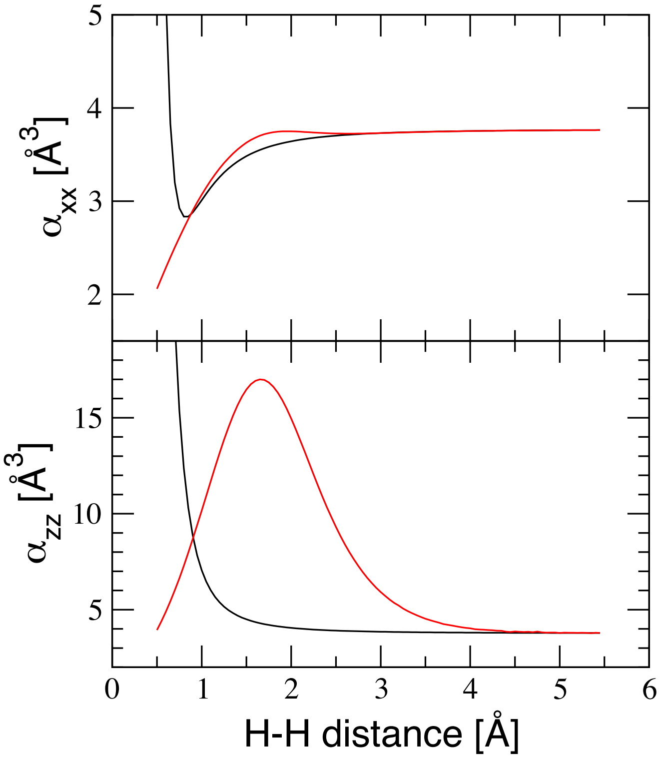


Figure S2. Comparison of pure-form Eq. 6 with λ_2=λ_3=1 and α_H = 0.279 Å^3 (black line) with the CCSD data (red lines) for H_2 polarizability.

S-3. Details of H_2O fitting

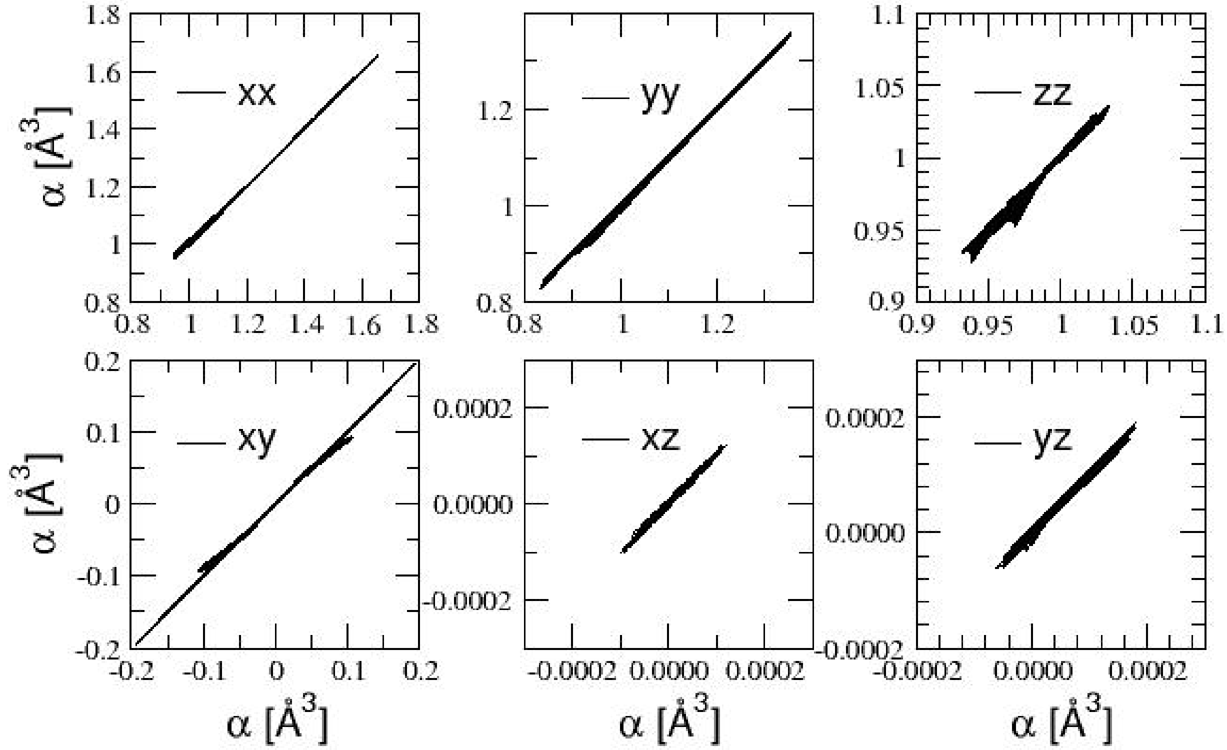


Figure S3. Correlation diagrams of the results of fitting H_2O polarizability tensor **α** six independent components (xx, yy, zz, xy, xz, yz) in the space-fixed frame to a training set of 5000 structures generated by propagating a classical trajectory at the total energy of ~4669 cm^{-1}, corresponding to the harmonic ZPVE. The level of theory for the training set is B3LYP/6-31++G(d,p). The RMS of the fit is ~0.004 Å^3.

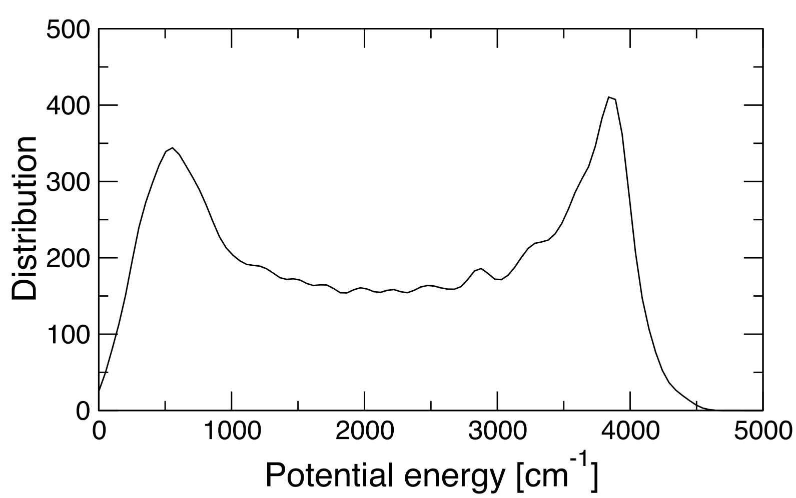


Figure S4. Distribution of the H_2O potential energies of the training set relative to the minimum.

The following symmetry displacement coordinates were used for testing the fit (Figure 2),

$$\Delta S_1 = \frac{\Delta R_1 + \Delta R_2}{\sqrt{2}} \qquad (S.4a)$$

$$\Delta S_2 = \Delta\theta_{\mathrm{HOH}} \qquad (S.4b)$$

$$\Delta S_3 = \frac{\Delta R_1 - \Delta R_2}{\sqrt{2}} \qquad (S.4c)$$

with R_1, R_2 the two OH distances and θ the HOH angle. The equilibrium values are R_1=R_2=0.9648 Å, θ_{HOH}=103.9 deg.

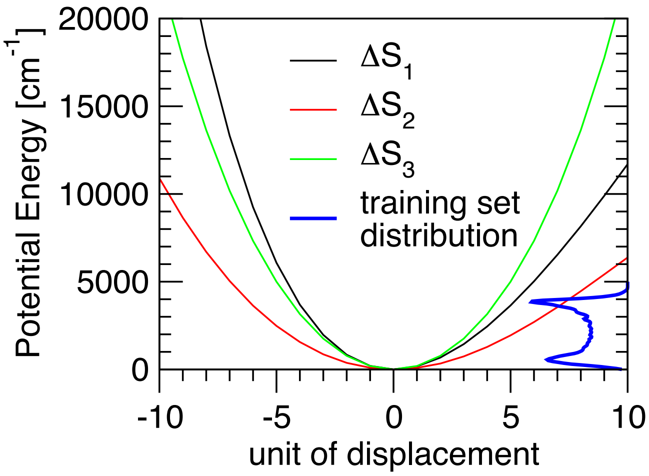


Figure S5. Comparison of the H_2O potential energies above the minimum accessed in the 1D symmetric cuts with the corresponding potential energy distribution in the training set. One unit of displacement corresponds to 0.03 Å for ΔS_1, ΔS_3 and 4 degrees for ΔS_2. As is evident, the training set potential energies do not exceed 5000 cm^{-1}.

S-4. Details of $H_5O_2^+$ fitting

We first examined a 5th order polynomial, same order as was used for H_2O with very good results. A non-linear search in a 564 parameter space yielded a fit with absolute and relative RMS of 0.048 Å^3 and 2.1%, respectively. Some of the polarizabilities at the geometries with higher energies were not reproduced well enough, and thus we examined the effect of a complete 6th order representation, with 1120 parameters. This brought the RMS down to 0.037 Å^3 and 1.2% and improved the previously problematic high energy structures rendering an overall satisfactory fit.

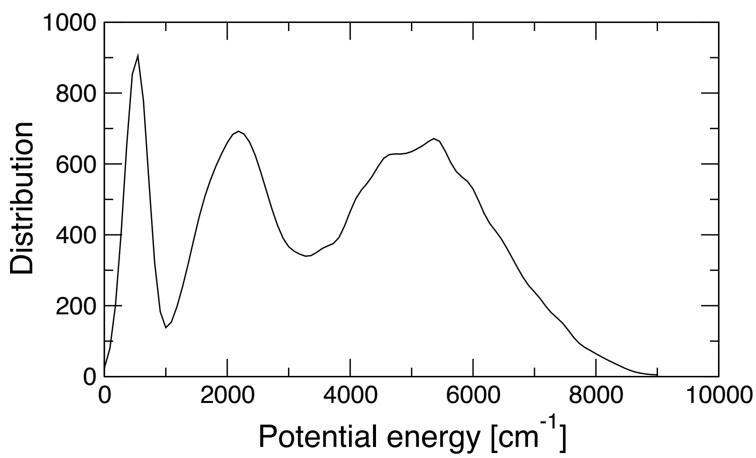


Figure S6. Distribution of the $H_5O_2^+$ potential energies of the training set relative to the minimum.

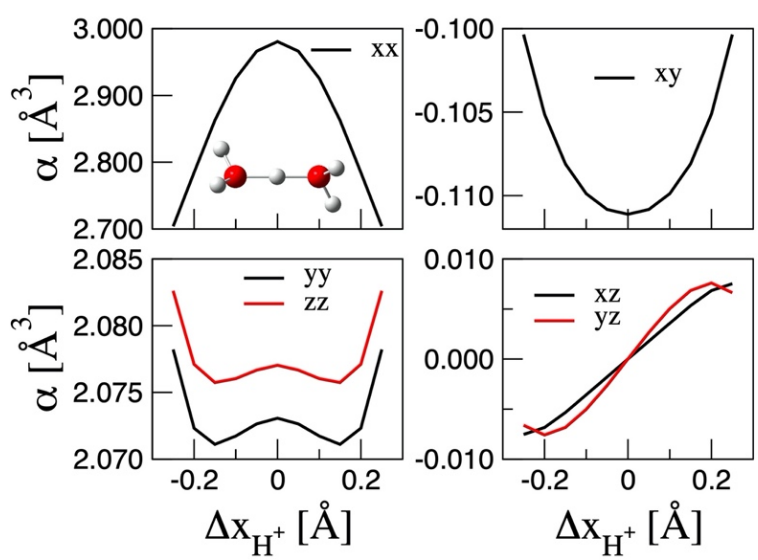


Figure S7. An illustration of the **PTS6** polarizability tensor elements of $H_5O_2^+$, as functions of the central Hydrogen atom (H^+) displacement from its C_2 equilibrium position, as calculated with HBB-PES, along the x-axis (O-O axis); z is defined as the C_2 axis.

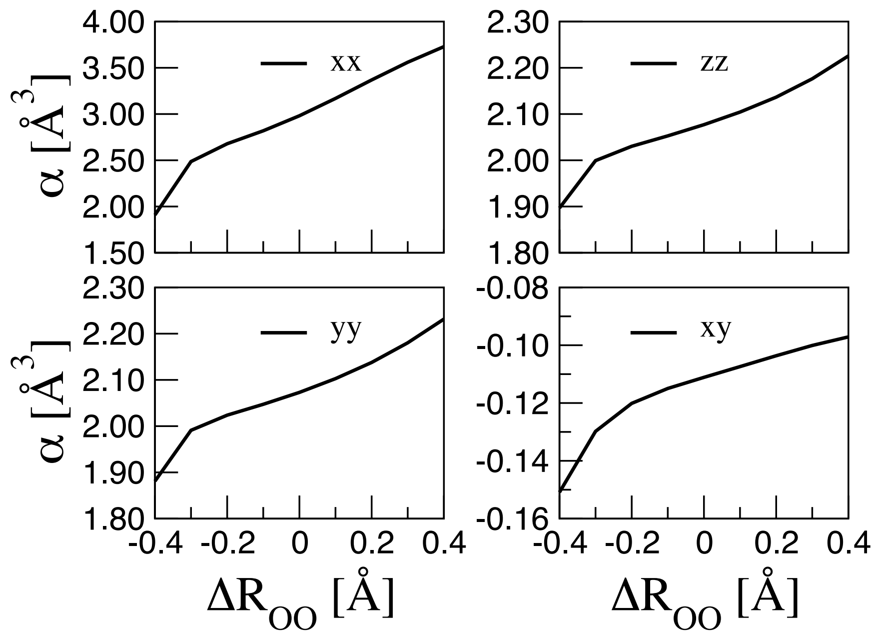


Figure S8. An illustration of the **PTS6** non-zero polarizability tensor elements of $H_5O_2^+$, as functions of the OO stretch deviation from its C_2 equilibrium position, as calculated with HBB-PES, along the x-axis (O-O axis); z is defined as the C_2 axis.

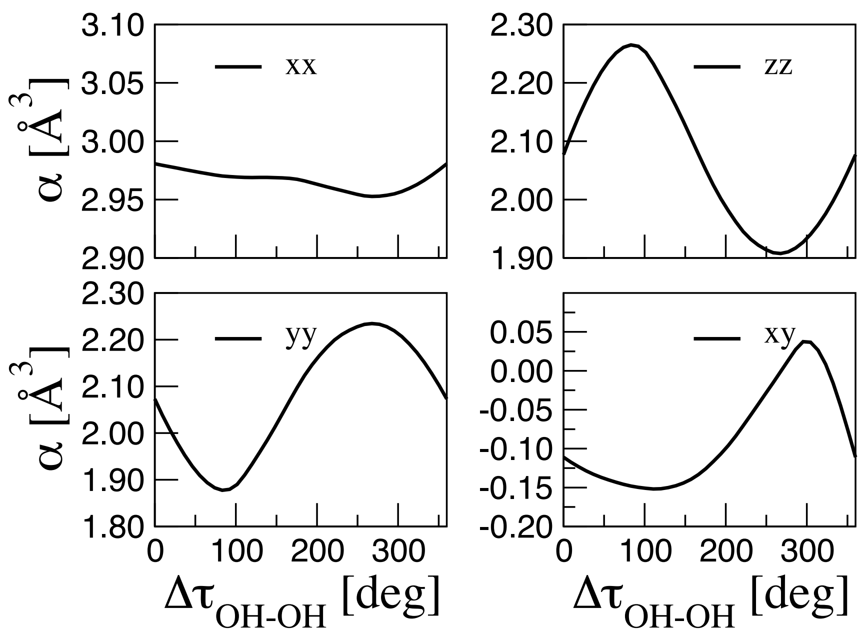


Figure S9. An illustration of the **PTS6** non-zero polarizability tensor elements of $H_5O_2^+$, as functions water monomer torsion displacement from its C_2 equilibrium position, as calculated with HBB2-PES, along the x-axis (O-O axis); z is defined as the C_2 axis.

S-5. Details of $H_5O_2^+$ electronic structure calculations and MD simulations

Table S1: Comparative analysis of the quantum mechanical methodologies in calculating the polarizability tensor, α. The root-mean-square errors, RMSE are given in Å^3 units, and the corresponding root-mean-square percent errors are RMSPE values. The reference point is defined at the CCSD(T)/aug-cc-pVTZ level of theory. MP2, CCSD, and CCSD(T) methods were tested with the basis sets aug-cc-pVDZ (AVDZ), cc-pVTZ (VTZ), aug-cc-pVTZ excluding *d* functions on hydrogen and *f* functions on oxygen atoms, respectively (AVTZ-tr), and aug-cc-pVTZ (AVTZ). All polarizabilities were evaluated at the $H_5O_2^+$ the minimum geometry optimized at the CCSD(T)/AVTZ level of theory. Computer times are reported for the single point calculation on a single processor.

Method/Basis set	Time (sec)	RMSE (Å^3)	RMSPE (%)
MP2/AVDZ	15	0.006	0.40
MP2/VTZ	82	0.157	11.1
MP2/AVTZ-tr	50	0.010	0.71
MP2/AVTZ	326	0.008	0.56
CCSD/AVDZ	121	0.023	1.62
CCSD/VTZ	723	0.168	11.96
CCSD/AVTZ-tr	397	0.016	1.15
CCSD/AVTZ	2871	0.021	1.46
CCSD(T)/AVDZ	349	0.007	0.50
CCSD(T)/VTZ	2063	0.157	11.17
CCSD(T)/AVTZ-tr	1300	0.005	0.33
CCSD(T)/AVTZ	9847	0.000	0.00

Root-mean-square error, RMSE in Å^3:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\alpha_i - \alpha_i^{ref})^2} \qquad (\mathrm{S.}\,5)$$

Corresponding root-mean-square percent error, RMSPE:

$$\mathrm{RMSPE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\alpha_i-\alpha_i^{ref}}{\alpha_i^{ref}}\right)^2} \times 100\,\% \qquad (\mathrm{S.}\,6)$$

Table S2: $H_5O_2^+$ harmonic vibrational frequencies (cm^{-1}, global minimum, C_2 symmetry), IR intensities (km/mol), Raman intensities (Å^4/amu), and depolarization ratios, ρ calculated at the MP2/aug-cc-pVTZ level of theory.

Labels	Frequencies	IR intensities	Raman intensities	ρ
ν_1 (A)	170.9	40.4	0.7	0.750
ν_2 (B)	367.1	275.5	0.2	0.750
ν_3 (A)	461.7	130.4	0.6	0.668
ν_4 (B)	534.6	74.2	0.2	0.750
ν_5 (A)	535.6	60.3	0.2	0.743
ν_6 (A)	623.5	0.0	4.5	0.176
ν_7 (B)	911.4	3025.5	0.4	0.750
ν_8 (B)	1473.0	267.7	0.5	0.750
ν_9 (A)	1550.0	98.2	2.3	0.468
ν_{10} (A)	1705.6	2.6	0.7	0.458
ν_{11} (B)	1761.2	963.0	0.6	0.750
ν_{12} (B)	3733.5	241.0	9.8	0.750
ν_{13} (A)	3741.3	8.9	134.8	0.010
ν_{14} (B)	3837.3	240.7	15.9	0.750
ν_{15} (A)	3837.7	336.5	17.0	0.732

MD trajectories for $H_5O_2^+$ were propagated at constant energy (NVE) corresponding to temperature 300 K and zero total angular momentum. In total, 10 trajectories were generated randomly and propagated up to 100 ps using the velocity-Verlet integrator with a time step of 0.2 fs. The Raman spectra were calculated by the Fourier transform of the polarizability correlation functions (Eq. 10*a* and 10*b* main text) recorded along the trajectories and time-averaged (the signal length was 6.5 ps) to yield a better converged spectrum. To better describe peak intensity in the higher frequency regions, in which proton motion is also involved, the classically derived spectral functions, I (Eq. 10*a* and 10*b* main text) were corrected by a quantum mechanical frequency-dependent factor equal to $\omega/[1-\exp(-\omega/kT)]$, where ω is the frequency.[Ref. 44 in main text]

We use McQuarrie's derivation of the polarized and depolarized Raman intensity components in terms of parallel and perpendicular scattered light, (Ref. 43 in main text)

$$I_{\text{pol}} = I_{\parallel} - \frac{4}{3} I_{\perp} \qquad (S.7a)$$

$$I_{\text{depol}} = 10 I_{\perp} \qquad (S.7b)$$

Following Wilson's definition of the depolarization ratio as the ratio of the perpendicular to the parallel scattered light,[1] the expression used for constructing Figure 5 of main text is

$$\rho \equiv \frac{I_{\perp}}{I_{\parallel}} = \frac{3I_{\text{depol}}}{4I_{\text{depol}} + 30I_{\text{pol}}} \qquad (S.8)$$

With this definition, the ratio is bound on the (0,0.75] range with the 0 corresponding to a totally symmetric motion and 0.75 to motion having none of the totally symmetric component.

References

[1] Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular vibrations*; Dover: New York, 1980.