SUPPLEMENTARY MATERIAL

Accurate intensity calibration of multichannel spectrometers using Raman intensity ratios

Ankit Raj,[1,*] Chihiro Kato,[2] Henryk A. Witek,[1,3,†] and Hiro-o Hamaguchi[1,3,‡]

[1]*Department of Applied Chemistry and Institute of Molecular Science, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan*

[2]*Department of Chemistry, KISTEC, Ebina, Japan*

[3]*Center for Emergent Functional Matter Science, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan*

Contents

---

* ankit7540.ipm02g@nctu.edu.tw, ankit@nycu.edu.tw
† hwitek@mail.nctu.edu.tw
‡ hhama@nctu.edu.tw

Contents

List of Figures

List of Tables

List of symbols

$\delta\nu_s$	:	A set of wavenumber intervals (forming a vector) between the consecutive data points of the x-axis in wavenumbers	cm^{-1}
$\delta\nu_0$	:	Wavenumber interval at a reference point on the x-axis in wavenumbers, scalar	cm^{-1}
γ	:	polarizability anisotropy, $\gamma = \alpha_{\parallel} - \alpha_{\perp}$	[a.u. or other specified unit]
$\bar{\alpha}$	:	mean polarizability , $\bar{\alpha} = (\alpha_{\parallel} + 2\alpha_{\perp})/3$	[a.u. or other specified unit]
λ	:	Wavelength	[nm]
$\psi_{v,J}$	:	Wavefunction corresponding to specific vibrational state (v) and rotational state (J) in the ground electronic state	
$\langle\psi_{v,J}\|\Omega\|\psi_{v',J'}\rangle$	:	Expectation value (or matrix element) of an operator Ω	[a.u. or other specified unit]
F_J	:	Boltzmann population (in the ground vibrational state)	unitless
ν_0	:	Frequency of the laser (in absolute wavenumbers)	cm^{-1}
ν_s	:	Frequency of the scattered light	cm^{-1}

Sec S1. Additional details on the intensity calibration procedures

A. C_0 correction

The C_0 correction, given by $(\frac{\delta\nu_s}{\delta\nu_0})$, is expressed in terms of the spacing in the wavelength scale as, $(\frac{\delta\nu_s}{\delta\nu_0}) = (\frac{\nu_s}{\nu_0})^2(\frac{\delta\lambda_s}{\delta\lambda_0})$. The derivation for this expression is given below:

$$\nu = \frac{1}{\lambda} \tag{1}$$

$$\delta\nu = -\frac{1}{\lambda^2}\ \delta\lambda \tag{2}$$

$$\frac{\delta\nu_s}{\delta\nu_0} = \frac{\frac{1}{\lambda_s^2}\ \delta\lambda_s}{\frac{1}{\lambda_0^2}\ \delta\lambda_0} \tag{3}$$

$$\frac{\delta\nu_s}{\delta\nu_0} = \left(\frac{\nu_s}{\nu_0}\right)^2\ \left(\frac{\delta\lambda_s}{\delta\lambda_0}\right), \tag{4}$$

where ν_0 is some reference wavenumber, and ν_s is any other wavenumber in the spectra. (Here, ν_0 is a scalar, while ν_s is a vector.) In Eqn. (4), the factor $(\frac{\nu_s}{\nu_0})^2$ corrects for the number of photons of the spectra measured across a linear wavelength scale. The term $(\frac{\delta\lambda_s}{\delta\lambda_0})$ is the small correction arising from the non-linearity of wavelength over the array of pixels.

In practice, the C_0 correction is simply computed using the wavenumber step (or spacing) in the wavenumber array representing the x-axis. Thus, $\delta\nu_s$, a vector corresponding to the wavenumber step for every data-point of the wavenumber axis is determined by computing the difference in adjacent data-points in the wavenumber axis, $\delta\nu_s = \nu_i - \nu_{i+1}$. For the very last datapoint, the wavenumber spacing can be approximated using extrapolation. The denominator, $\delta\nu_0$, is a scalar representing the similar wavenumber spacing at certain reference wavenumber datapoint, which can be conveniently chosen, for example, at the center of x-axis. The C_0 correction is specific to the monochromator, type of the grating used, and also on the studied spectral region.

B. C_1 correction

The spectra of a broadband white-light source used for obtaining the C_1 correction, must be first corrected using the C_0 curve. Next, the fit for the white-light spectra using the expression for the number of photons per wavenumber is done, wherein temperature is a fit parameter. In this process, if the approximate temperature of the lamp is known, then this can be used the fixed parameter in the fit.

• If the C_1 correction is obtained using a calibrated reference, such as a light source or a fluorophore, then (C_0/C_1) correction for relative Raman intensities may be sufficient for regular experiments where high accuracy in the relative Raman intensities is not required. To facilitate the determination of (C_0/C_1) correction, programs have been made available online (procedure developed for IgorPro at `gen_correction.ipf` in GitHub repository[1], Python function at `gen_correction.py` in the GitHub repository associated with this work Ref. [2]). These programs are based on the principle of using broadband white-light spectra from a tungsten lamp assumed to be a black-body emission source for C_1 determination. The input data required for usage these programs are: raw spectra of white-light (as a 1D or 2D array), the x-axis vector (in relative wavenumbers) alongwith the information on the laser wavelength.

Sec S2. Determination of C_2 correction using Raman intensities

A. General methodology

Consider a pair of (x, y) data, described as $y = f(x)$, which is perturbed in terms of the y-values by a perturbation defined as $p(x)$ to produce say $c(x)$. The perturbation $p(x)$ is equivalent to the wavelength dependent sensitivity of the spectrometer. The objective is to determine $p(x)$ when the y-values for $f(x)$ and $c(x)$ are known for sparse pairs of x-points.

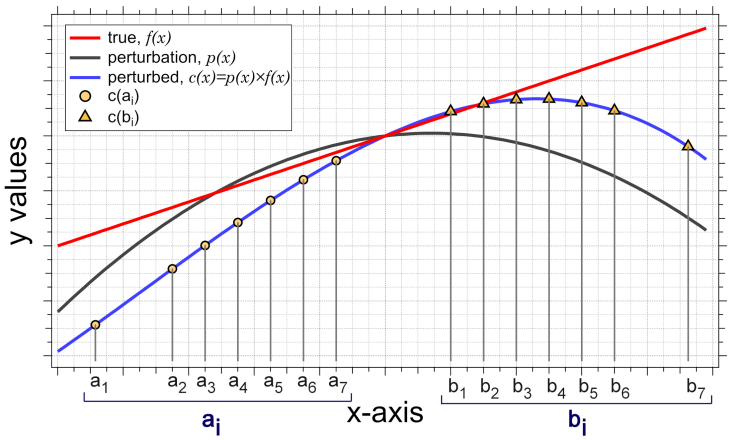

FIG. S1: Plot describing the concept of the fitting used in the present work for determining the correction factor C_2 derived from Raman intensities. Ratio of y-values at pair of x-points (a_i, b_i), in the true, $f(x)$ and perturbed functions, $c(x)$ are used to determine the perturbation, $p(x)$ modeled as a polynomial. The choice of x-positions, a_i and b_i is arbitrary and these can be randomly located.

The ratio of the y-values at two x-points (a_i, b_i) in the true data, $f(x)$ expressed as $R^{true}_{a_i,b_i}$ is affected by the perturbation to yield $R^{pertb.}_{a_i,b_i}$. This change is directly related to the perturbation function by the ratio of $p(x)$ at a_i and b_i which can be given as $R^{p(x)}_{a_i,b_i}$.

$$\frac{R^{pertb.}_{a_i,b_i}}{R^{true}_{a_i,b_i}} = R^{p(x)}_{a_i,b_i} \ , \tag{5}$$

where $R^{pertb.}_{a_i,b_i}$=$c(a_i)/c(b_i)$ and $R^{true}_{a_i,b_i}$=$f(a_i)/f(b_i)$. The perturbation, $p(x)$ can be modeled as a polynomial function with coefficients k_n to obtain,

$$\frac{R^{pertb.}_{a_i,b_i}}{R^{true}_{a_i,b_i}} = \frac{1 + k_1 a_i + k_2 a_i^2 + k_3 a_i^3 + ...}{1 + k_1 b_i + k_2 b_i^2 + k_3 b_i^3 + ...} \tag{6}$$

The difference of the LHS to RHS in Eqn. (6) can be used in a least squares minimization procedure to obtain the coefficients k_n.

$$\text{Residual} = \sum_i (\text{LHS} - \text{RHS})^2 \tag{7}$$

$$\textbf{Min}(\text{ Residual }) : k_n \text{ as fit parameters} \tag{8}$$

Error in the LHS of Eqn. (6) is given as

$$\sigma_{\text{LHS}_i} = \frac{c(a_i)}{c(b_i)} \sqrt{\left(\frac{\sigma_{c(a_i)}}{c(a_i)}\right)^2 + \left(\frac{\sigma_{c(b_i)}}{c(b_i)}\right)^2} \ , \tag{9}$$

where $c(a_i)$ and $c(b_i)$ are the values of $c(x)$ at a_i and b_i, respectively. Similarly, $\sigma_{c(a_i)}$ and $\sigma_{c(b_i)}$ represent the error in the values of $c(x)$ at a_i and b_i, respectively. Weights, computed as $w_i = 1/\sigma^2_{\mathrm{LHS}_i}$, can then be applied to the elements followed by least squares minimization of the residual,

$$\text{Residual} = \sum_i w_i (\text{LHS} - \text{RHS})^2 \ , \tag{10}$$

where LHS and RHS refer to the terms in Eqn. (6).

The above general procedure can be used with ratios of the y-values for all available x-points, i.e. to obtain $R^{\text{pertb.}}_{p,q}$ and $R^{\text{true}}_{p,q}$, for all available p and q points. In the present work, we have utilized the Raman band intensities for determining the perturbation as the C_2 correction. Matrices are used for describing sets of intensity ratios ($\mathbb{R}^{\text{true}}$ for reference intensity ratios and $\mathbb{R}^{\text{expt}}$ for intensity ratios obtained from experimental data), for a clearer discussion of such intensity ratios in context of the vibration-rotation bands of H_2, HD and D_2 (see Fig. 1 and 7 in the main document).

B. Present scheme

Steps of the algorithm are described below :

1. The experimental data (available as array of band areas, corresponding errors and the Raman wavenumbers) is used to generate the $\mathbb{R}_{\text{obs}}$ matrix. Using the errors in band areas, the respective weights are generated.

2. The analogous reference data computed, at a given temperature, is used to generate the $\mathbb{R}_{\text{true}}$ matrix (see Section Sec S2 C for more details).

3. Next, using the band positions and initial coefficients of the polynomial and temperature, the matrix $\mathbb{S}$ is generated.[1]

4. The dimensions of the four matrices ($\mathbb{R}_{\text{obs}}$, $\mathbb{R}_{\text{true}}$, $\mathbb{D}$ and $\mathbb{S}$) are checked.

5. Difference matrix, $\mathbb{D}$, (for each species) is generated using the $\mathbb{R}_{\text{obs}}$, $\mathbb{R}_{\text{true}}$ and $\mathbb{S}$ matrix (in our work, we had three $\mathbb{D}$ matrices, one each for H_2, HD and D_2).

6. Elements of the difference matrix ($\mathbb{D}_{i,j}$) are weighted using the corresponding elements of the weight matrix ($\mathbb{W}_{i,j}$).

7. Squares of the Frobenius norm of the difference matrices is then computed. The sum of such norms defined as E (obtained from each individual difference matrix), is minimized, by varying the coefficients (c_n) of the polynomial ($1 + \sum c_n \nu^n$) representing the wavenumber-dependent sensitivity, and the temperature (T_{analysis}). This involves re-computing the reference matrix ($\mathbb{R}_{\text{true}}$), and the sensitivity matrix ($\mathbb{S}$) using the updated values of the parameters: temperature and the coefficients of the polynomial. This minimization is carried out by employing standard algorithm(s) for optimization available in FORTRAN, accessible via library calls in SciPy[3]. In this operation, the change of E (i.e. the gradient) is evaluated at every subsequent iteration, for particular values of the fitted parameters, until the gradient decreases below a threshold value.

8. The optimized coefficients of the polynomial are used to generate the C_2 correction. After total correction of the Raman spectral intensities, the accuracy of the intensity calibration procedure is checked (e.g., using temperature determined from Raman intensities).

---

[1] The polynomial function has the form : $1 + \sum c_n \nu^n$. Initial guess for the coefficients (c_n) and the temperature (T_{analysis}) are required in this analysis.

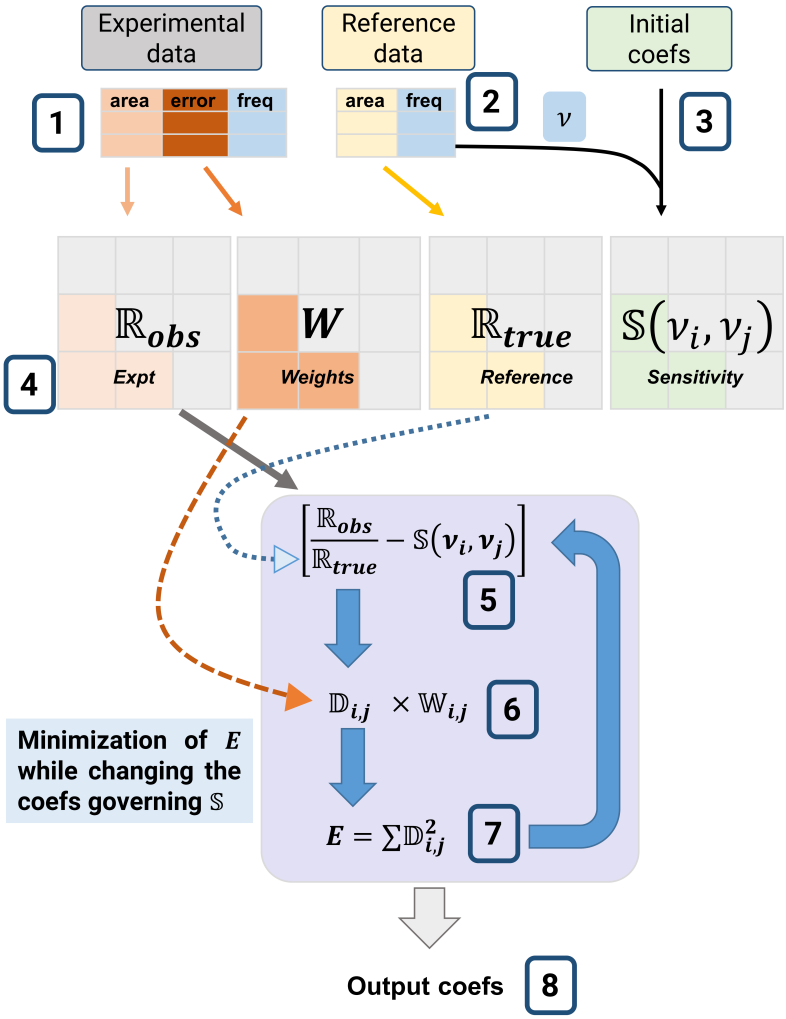

FIG. S2: Scheme for obtaining the coefficients of the polynomial used to model the wavenumber dependent sensitivity. From the top, band area and its error is used to obtain the $\mathbb{R}_{obs}$ matrix and the weight matrix, $\mathbb{W}$. Reference intensities computed from known values of polarizability invariants are used to obtain the $\mathbb{R}_{true}$ matrix. Next, using initial guess values of the coefficients for an appropriately chosen polynomial (dependent on wavenumber), matrix $\mathbb{S}$ is generated. Difference matrix, $\mathbb{D}$, is then obtained using $\mathbb{R}_{obs}$, $\mathbb{R}_{true}$, and $\mathbb{S}$, as described in the main document. Weights are applied to elements of the difference matrix followed by a least squares minimization of the norm of $\mathbb{D}$ with coefficients of the polynomial and the temperature, as the target optimized variables.

C. Computation of $\mathbb{R}_{\text{true}}$

Raman intensity ratios among the O, Q and S can be grouped into 6 categories which are shown below.

$$\text{a} = \frac{I_{O(J_i)}}{I_{O(J'_i)}} \qquad \text{d} = \frac{I_{Q(J_i)}}{I_{S(J'_i)}}$$

$$\text{b} = \frac{I_{O(J_i)}}{I_{Q(J'_i)}} \qquad \text{e} = \frac{I_{Q(J_i)}}{I_{Q(J'_i)}}$$

$$\text{c} = \frac{I_{O(J_i)}}{I_{S(J'_i)}} \qquad \text{f} = \frac{I_{S(J_i)}}{I_{S(J'_i)}}$$

FIG. S3: The ratios of Raman intensities among the O, Q and S bands.

In general, these intensity ratios have the form

$$\mathbb{R}_{\text{true},ij} = \frac{I_i}{I_j}_{\text{true}} = \frac{\nu_s^3}{\nu_{s'}^3}\frac{F_J}{F_{J'}}\frac{f(\langle\bar{\alpha}\rangle^2, \langle\bar{\gamma}\rangle^2, b_J^{\Delta J=0,\pm 2})}{f'(\langle\bar{\alpha}\rangle^2, \langle\bar{\gamma}\rangle^2, b_{J'}^{\Delta J=0,\pm 2})} \tag{11}$$

where ν_s is the frequency of the scattered photon (in absolute wavenumbers), F_J is the Boltzmann population for the initial state J, and the term $f(\langle\bar{\alpha}\rangle^2, \langle\bar{\gamma}\rangle^2, b_J^{\Delta J=0,\pm 2})$ represents a function of the matrix element of mean polarizability and anisotropy, with the Placzek-Teller factor.[2] A distinction is made between the terms in the numerator and denominator using the prime ($\prime$) symbol.

Details on the computation of the individual terms in Eqn. 11 are discussed below:

1. ν_s is available from the Raman transition frequencies (obtained either from experiments or computed).

2. The Boltzmann population F_J (in the ground vibrational state) for diatomic molecules is given by

$$F_J = \frac{g_J(2J+1)e^{\frac{-E_J}{k_B T}}}{\mathbf{Q}}, \tag{12}$$

where g_J is weight from the nuclear spin statistics, $(2J+1)$ is the rotational degeneracy, E_J is the rotational energy for state, k_B is the Boltzmann constant, T is the temperature (in Kelvin), and $\mathbf{Q}$ is the ro-vibrational partition function (also known as the sum of states).

With regard to the present work, accurate energies of the ro-vibrational states ($E_{v,J}$) of H_2, HD and D_2 were obtained from the accurate data set on the dissociation energies of these molecules reported by Komasa et al.[4] and Pachucki et al.[5]. Subsequently, the numerical value of $\mathbf{Q}$ was computed to high accuracy by including a large set of ro-vibrational states in the summation (the ro-vibrational states up to 30,000 cm^{-1} (relative to the $v=0, J=0$ state) were included in this computation).[3]

$$\mathbf{Q} = \left(\sum_J^{J_{\max}} g_J(2J+1)e^{\frac{-E_{v,J}}{k_B T}}\right)_{v=0} + \left(\sum_J^{J_{\max}} (2J+1)e^{\frac{-E_{v,J}}{k_B T}}\right)_{v=1} + \\ \ldots + \left(\sum_J^{J_{\max}} g_J(2J+1)e^{\frac{-E_{v,J}}{k_B T}}\right)_{v=v_{\max}} \tag{13}$$

[2] This term varies for O, Q and S Raman transitions and also for parallel and perpendicular detection. Hence, a general form is presented.

[3] The total number of ro-vibrational states included in Eqn. 13 during the computation of $\mathbf{Q}$ were: 197 for H_2, 268 for HD and 365 for D_2. The partition function so constructed was tested for computation of the Boltzmann populations up to T=4000 K.

3. The term comprising the ro-vibrational matrix elements of polarizability invariants and the Placzek-Teller factor $f(\langle\bar{\alpha}\rangle^2, \langle\bar{\gamma}\rangle^2, b_J^{\Delta J=0,\pm 2})$ is specific for the O, Q and S Raman transitions and also for parallel and perpendicular detection. Expressions for intensity of parallel polarized Raman scattering are placed in the main document. For the perpendicularly polarized case refer to Sec S6. Numerical values of the ro-vibrational matrix elements of polarizability invariants for various excitation wavelengths have been reported earlier.[6, 7] Values for 532 nm laser excitation, pertinent to this study, are listed in Sec S5. Numerical uncertainties in these values were estimated to be within 1% (see Sec S5 A).

D. Non-linear optimization for the determination of C_2 correction

The non-linear optimization process starts with initial coefficients provided by the user. The number of these coefficients depend on the degree of the polynomial used to model the C_2 correction. Additionally, temperature ($T_{analysis}$) is also an initial parameter to be provided.

In the present work, we started the analysis for C_2 correction with the linear model for wavenumber dependent sensitivity. Thus, two initial coefficients were required to be provided (defined as c_1^{init} and $T_{analysis}^{init}$)[4]. Three distinct runs were performed : c_1^{init}=0.0, $T_{analysis}^{init}$ = 300 ; $c_1^{init} = 0.2$, $T_{analysis}^{init} = 300$; and $c_1^{init} = -0.2$, $T_{analysis}^{init} = 300$. These three runs gave the same results[5] as $c_1 = -0.63$, $T_{analysis} = 292$.

The c_1 coefficient from linear polynomial was then used as the initial guess for the quadratic fit. The additional coefficient (c_2) introduced here was checked with three guesses: 0.0, 0.4 and -0.4 in three separate runs.

This iterative process was continued for further polynomials tested. A random set of initial coefficients were also used in a separate analysis, which gave similar optimized results, within numerical uncertainty, as obtained earlier. These tests confirmed the convergence of our optimization analysis.

The C_2 curves corresponding to the tested polynomials are shown below.

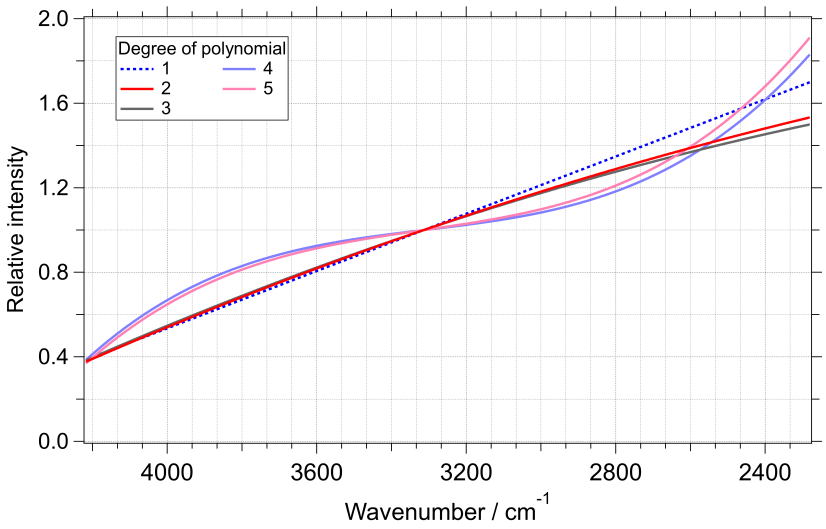

FIG. S4: Polynomial solutions modeling the C_2 correction obtained using the analysis of relative experimental and theoretical Raman intensities from H_2 and isotopologues.

In the next step, the obtained C_2 correction curves were used for rectifying the Raman spectra of $CHCl_3$ measured separately (see main document for more details). The Stokes and anti-Stokes Raman bands of $CHCl_3$ were used for testing the accuracy of performed calibration (and hence validity of the obtained C_2 correction was checked).

---

[4] Coefficients of the polynomial are listed in small caps, for example, c_1, c_2.

[5] Note that the wavenumber points used in this analysis were rescaled using a scalar ν' to yield $\nu_s - \nu'$. Thus, the actual magnitude of the final coefficients is not visible in the above discussion.

E. Details on the programs available in the online repository

The online repository[2] includes various programs with examples developed in Python to perform the described analysis.

1. The module `compute_spectra` is useful is computing the reference intensities for the present analysis. This module reads pre-calculated data on the matrix elements of polarizability invariants from .dat files (standard ASCII files). Boltzmann populations at a given temperature are computed using a separate module, `boltzmann_popln`, which is called by the main module when required.

2. In the present work, vibration-rotation Raman intensities were used for the analysis, where the Raman intensity ratios from common initial states were given preference. This kind of analysis for the pure rotational Raman bands has been demonstrated earlier (which was based on only using the intensity ratios from common rotational states).[8, 9] A similar analysis for the pure rotational Raman bands is also possible as shown in the present work, where all bands intensity ratios are used and some of the terms are weighted. Programs for such an analysis have been made available in the online repository.

3. The present analysis can also be performed where temperature of the sample is fixed (i.e. temperature is not a fit parameter) during determination of the coefficients of the polynomial(s). Programs implementing such a scheme have been made available on the Github repository.

4. The measure of the difference between experimental and reference matrices, can be defined in various ways. The net error (or residual) is expressed in terms of the elements of the difference matrices. In the available programs, three versions for the net error have been implemented:

   (a) Sum of absolute values of elements : $E = \sum_i \sum_j |d_{i,j}|_{\mathrm{H_2}} + \sum_i \sum_j |d_{i,j}|_{\mathrm{HD}} + ...$

   (b) Frobenius norm : $E = \sqrt{\sum_i \sum_j |d_{i,j}|^2_{\mathrm{H_2}}} + \sqrt{\sum_i \sum_j |d_{i,j}|^2_{\mathrm{HD}}} + ...$

   (c) Square of the Frobenius norm : $E = \sum_i \sum_j |d_{i,j}|^2_{\mathrm{H_2}} + \sum_i \sum_j |d_{i,j}|^2_{\mathrm{HD}} + ...$

   One can choose the specific definition for an analysis.

5. It is emphasized that the obtained solution (i.e. polynomial for the C_2 correction) from the present scheme must be verified by using the intensities of the calibrated spectrum employing an independent technique. This is necessary to verify physical correctness and determining accuracy.

Sec S3. Details of the Raman Spectrometer

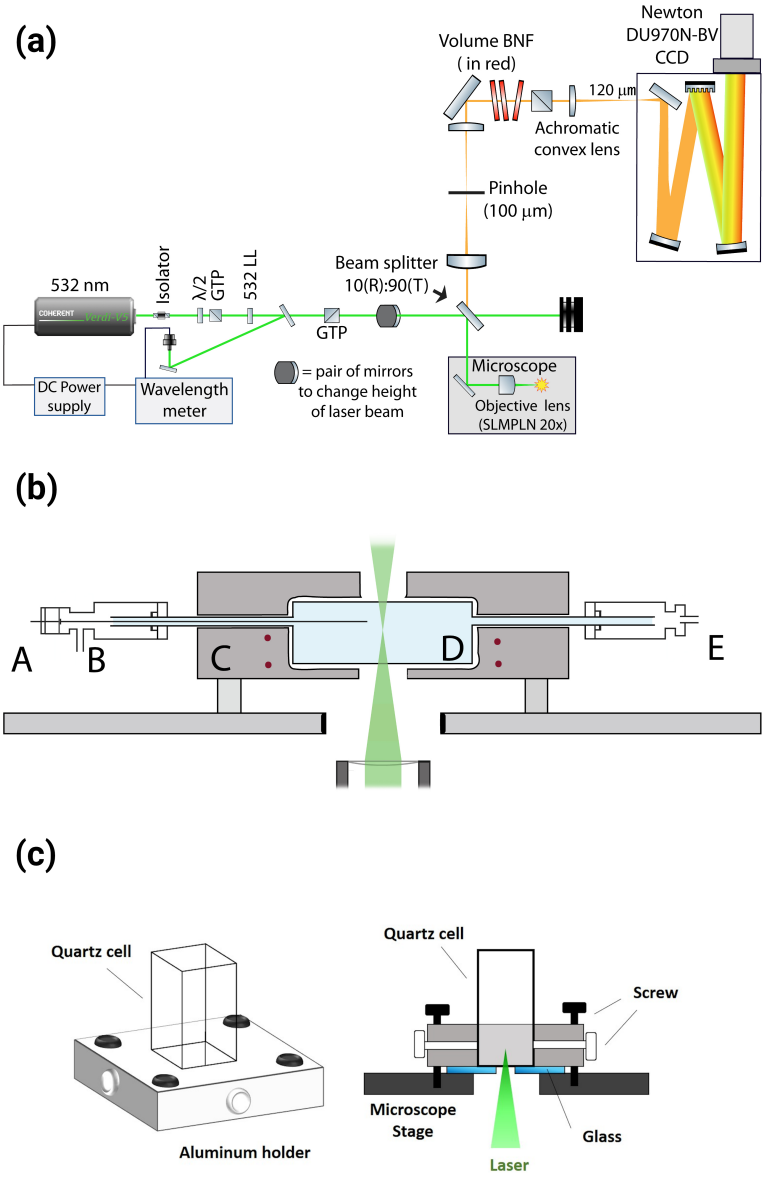

FIG. S5: Layout of the Raman spectrometer used in the present work. Sub-figure (a) shows the overall view from the top where GTP refers to Glan–Taylor prism, 532 LL is the laser line filter for 532 nm, and volume BNF refers to volume Bragg notch filters (OptiGrate). Gas samples have been measured using a gas cell assembly placed on the microscope stage, (b) the gas cell assembly, where (A) thermocouple inlet for measurement of temperature inside the cell, (B) inlet port for gases, (C) metal casing with resistive heaters, (D) quartz cell, and (E) inlet port for gases and/or vacuum manifold, and (c) the quartz cuvette placed on the microscope stage using the custom-designed mount.

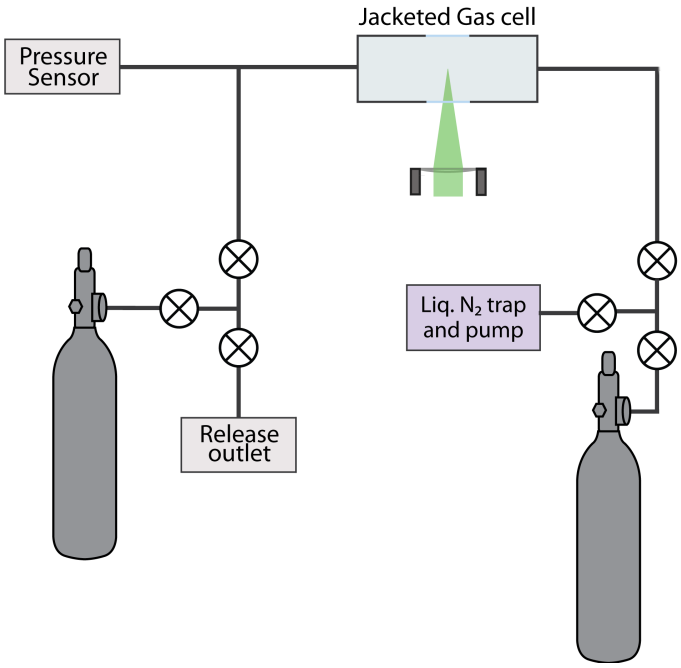

FIG. S6: Overall design of the sampling setup for gases. This setup is based on stainless steel 316 tubes for connections and valves for transferring gas samples from cylinders to the sample chamber. The pressure sensor used is an absolute NIST traceable pressure transducer from Omega (PX409-AUSBH).

Sec S4. Wavenumber calibration

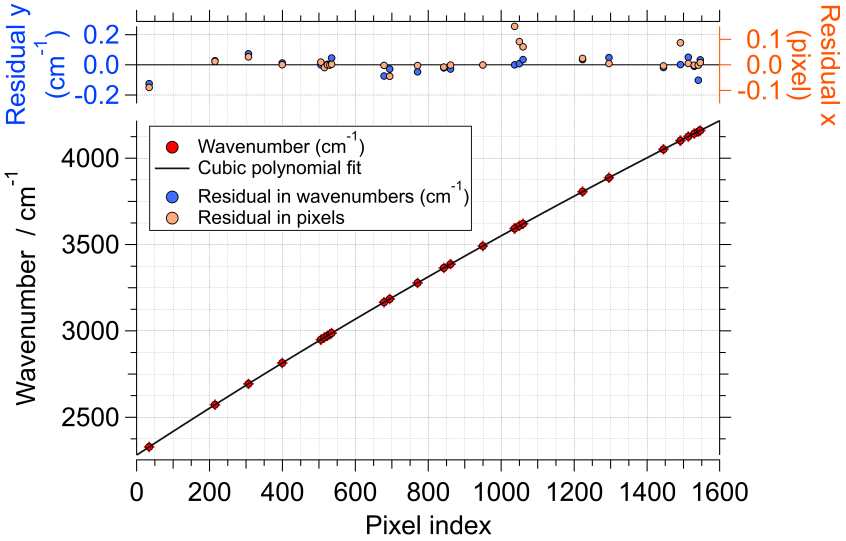

FIG. S7: Fit of the reference wavenumbers (in cm^{-1}) over the band positions (in pixel), for wavenumber calibration using the vibration-rotation bands from H_2, HD and D_2. The Q1-band of N_2 was also included, and a total of 26 data points were used.

Sec S5. Excitation wavelength dependent ro-vibrational matrix elements of mean polarizability and anisotropy of H_2, HD and D_2 used in the present work

TABLE S8: Ro-vibrational matrix elements and absolute Raman cross-sections of vibration-rotation Raman transitions in H_2 and HD, for parallel polarized Raman detection. (λ =532.2 nm)

H_2

Transition		ν	$\langle\tilde{\alpha}\rangle$	$\langle\tilde{\alpha}\rangle^2$	$\langle\gamma\rangle$	$\langle\gamma\rangle^2$	$d\sigma/d\Omega$ [a,b]	Boltzmann population
		cm^{-1}	a.u.	$C^2V^{-2}m^4$	a.u.	$C^2V^{-2}m^4$	$cm^2\ sr^{-1}\ molecule^{-1}$	T=293.15 K
O(4)	$v=0, J=2 \rightarrow v=1, J=4$	3329.04	0.531 90	7.6910×10^{-83}	0.563 25	8.6244×10^{-83}	1.9149×10^{-32}	$3.818\,77 \times 10^{-3}$
O(3)	$v=0, J=1 \rightarrow v=1, J=3$	3568.22	0.600 85	9.8142×10^{-83}	0.587 11	9.3706×10^{-83}	1.7869×10^{-32}	$8.657\,36 \times 10^{-2}$
O(2)	$v=0, J=0 \rightarrow v=1, J=2$	3806.79	0.671 49	1.2258×10^{-82}	0.612 64	1.0203×10^{-82}	1.4433×10^{-32}	$1.155\,07 \times 10^{-1}$
Q(4)	$v=0, J=4 \rightarrow v=1, J=4$	4102.58	0.791 48	1.7030×10^{-82}	0.673 16	1.2319×10^{-82}	1.2977×10^{-30}	$3.818\,77 \times 10^{-3}$
Q(3)	$v=0, J=3 \rightarrow v=1, J=3$	4125.87	0.786 86	1.6831×10^{-82}	0.665 33	1.2034×10^{-82}	1.2767×10^{-30}	$8.657\,36 \times 10^{-2}$
Q(2)	$v=0, J=2 \rightarrow v=1, J=2$	4143.47	0.783 36	1.6682×10^{-82}	0.659 45	1.1822×10^{-82}	1.2622×10^{-30}	$1.155\,07 \times 10^{-1}$
Q(1)	$v=0, J=1 \rightarrow v=1, J=1$	4155.25	0.781 03	1.6583×10^{-82}	0.655 53	1.1682×10^{-82}	1.2603×10^{-30}	$6.617\,21 \times 10^{-1}$
Q(0)	$v=0, J=0 \rightarrow v=1, J=0$	4161.17	0.779 85	1.6533×10^{-82}	0.653 56	1.1612×10^{-82}	1.2243×10^{-30}	$1.315\,19 \times 10^{-1}$
S(0)	$v=0, J=0 \rightarrow v=1, J=2$	4497.84	0.671 49	1.2258×10^{-82}	0.612 64	1.0203×10^{-82}	6.2631×10^{-32}	$1.315\,19 \times 10^{-1}$
S(1)	$v=0, J=1 \rightarrow v=1, J=3$	4712.90	0.600 85	9.8142×10^{-83}	0.587 11	9.3706×10^{-83}	3.2978×10^{-32}	$6.617\,21 \times 10^{-1}$
S(2)	$v=0, J=2 \rightarrow v=1, J=4$	4917.01	0.531 90	7.6910×10^{-83}	0.563 25	8.6244×10^{-83}	2.4901×10^{-32}	$1.155\,07 \times 10^{-1}$
S(3)	$v=0, J=3 \rightarrow v=1, J=5$	5108.40	0.464 73	5.8712×10^{-83}	0.540 93	7.9543×10^{-83}	2.0397×10^{-32}	$8.657\,36 \times 10^{-2}$
S(4)	$v=0, J=4 \rightarrow v=1, J=6$	5285.59	0.399 41	4.3367×10^{-83}	0.520 00	7.3508×10^{-83}	1.7303×10^{-32}	$3.818\,77 \times 10^{-3}$

HD

Transition		ν	$\langle\tilde{\alpha}\rangle$	$\langle\tilde{\alpha}\rangle^2$	$\langle\gamma\rangle$	$\langle\gamma\rangle^2$	$d\sigma/d\Omega$ [a,b]	Boltzmann population
		cm^{-1}	a.u.	$C^2V^{-2}m^4$	a.u.	$C^2V^{-2}m^4$	$cm^2\ sr^{-1}\ molecule^{-1}$	T=293.15 K
O(4)	$v=0, J=2 \rightarrow v=1, J=4$	3004.52	0.523 18	7.4408×10^{-83}	0.528 47	7.5920×10^{-83}	1.7940×10^{-32}	$2.390\,76 \times 10^{-2}$
O(3)	$v=0, J=1 \rightarrow v=1, J=3$	3185.22	0.578 86	9.1091×10^{-83}	0.547 70	8.1546×10^{-83}	1.6754×10^{-32}	$1.040\,49 \times 10^{-1}$
O(2)	$v=0, J=0 \rightarrow v=1, J=2$	3365.09	0.635 68	1.0985×10^{-82}	0.568 12	8.7740×10^{-83}	1.3541×10^{-32}	$2.731\,95 \times 10^{-1}$
Q(4)	$v=0, J=4 \rightarrow v=1, J=4$	3593.87	0.730 71	1.4515×10^{-82}	0.614 27	1.0257×10^{-82}	1.2245×10^{-30}	$2.390\,76 \times 10^{-2}$
Q(3)	$v=0, J=3 \rightarrow v=1, J=3$	3609.11	0.727 45	1.4386×10^{-82}	0.608 81	1.0076×10^{-82}	1.2103×10^{-30}	$1.040\,49 \times 10^{-1}$
Q(2)	$v=0, J=2 \rightarrow v=1, J=2$	3620.61	0.725 00	1.4289×10^{-82}	0.604 71	9.9408×10^{-83}	1.2007×10^{-30}	$2.731\,95 \times 10^{-1}$
Q(1)	$v=0, J=1 \rightarrow v=1, J=1$	3628.30	0.723 35	1.4224×10^{-82}	0.601 98	9.8511×10^{-83}	1.2016×10^{-30}	$3.923\,70 \times 10^{-1}$
Q(0)	$v=0, J=0 \rightarrow v=1, J=0$	3632.16	0.722 53	1.4192×10^{-82}	0.600 61	9.8064×10^{-83}	1.1692×10^{-30}	$2.026\,59 \times 10^{-1}$
S(0)	$v=0, J=0 \rightarrow v=1, J=2$	3887.68	0.635 68	1.0985×10^{-82}	0.568 12	8.7740×10^{-83}	6.1056×10^{-32}	$2.026\,59 \times 10^{-1}$
S(1)	$v=0, J=1 \rightarrow v=1, J=3$	4052.19	0.578 86	9.1091×10^{-83}	0.547 70	8.1546×10^{-83}	3.2932×10^{-32}	$3.923\,70 \times 10^{-1}$
S(2)	$v=0, J=2 \rightarrow v=1, J=4$	4209.96	0.523 18	7.4408×10^{-83}	0.528 47	7.5920×10^{-83}	2.5445×10^{-32}	$2.731\,95 \times 10^{-1}$
S(3)	$v=0, J=3 \rightarrow v=1, J=5$	4359.94	0.468 69	5.9717×10^{-83}	0.510 34	7.0802×10^{-83}	2.1301×10^{-32}	$1.040\,49 \times 10^{-1}$
S(4)	$v=0, J=4 \rightarrow v=1, J=6$	4501.21	0.415 46	4.6922×10^{-83}	0.493 25	6.6140×10^{-83}	1.8442×10^{-32}	$2.390\,76 \times 10^{-2}$

a: For perpendicularly polarized detection, use the depolarization ratio (ρ) with ($d\sigma/d\Omega$). See Table S16 and Ref. [10]
b: For ($\parallel + \perp$) detection, use $(1+\rho) \times (d\sigma/d\Omega)$ with the appropriate ρ values.
See Ref. [6] and online repository[7] for programs to compute the ro-vibrational matrix elements at other excitation wavelengths in the range of 182–1320 nm

TABLE S9: Ro-vibrational matrix elements and absolute Raman cross-sections of vibration-rotation Raman transitions in D_2, for parallel polarized Raman detection. (λ =532.2 nm)

D_2								
Transition		ν	$\langle\bar{\alpha}\rangle$	$\langle\bar{\alpha}\rangle^2$	$\langle\gamma\rangle$	$\langle\gamma\rangle^2$	$d\sigma/d\Omega$ [a,b]	Boltzmann population
		cm^{-1}	a.u.	$C^2V^{-2}m^4$	a.u.	$C^2V^{-2}m^4$	$cm^2\ sr^{-1}\ molecule^{-1}$	T=293.15 K
O(4)	$v=0, J=2 \rightarrow v=1, J=4$	2572.64	0.502 77	6.8716×10^{-83}	0.481 85	6.3118×10^{-83}	1.6173×10^{-32}	$9.055\,34 \times 10^{-2}$
O(3)	$v=0, J=1 \rightarrow v=1, J=3$	2693.97	0.543 94	8.0430×10^{-83}	0.496 07	6.6897×10^{-83}	1.5084×10^{-32}	$1.123\,63 \times 10^{-1}$
O(2)	$v=0, J=0 \rightarrow v=1, J=2$	2814.55	0.585 75	9.3273×10^{-83}	0.511 02	7.0990×10^{-83}	1.2172×10^{-32}	$3.850\,00 \times 10^{-1}$
Q(6)	$v=0, J=6 \rightarrow v=1, J=6$	2949.76	0.659 80	1.1835×10^{-82}	0.551 79	8.2770×10^{-83}	1.1303×10^{-30}	$5.577\,70 \times 10^{-3}$
Q(5)	$v=0, J=5 \rightarrow v=1, J=5$	2962.19	0.656 85	1.1729×10^{-82}	0.546 88	8.1303×10^{-83}	1.1175×10^{-30}	$1.310\,47 \times 10^{-2}$
Q(4)	$v=0, J=4 \rightarrow v=1, J=4$	2972.61	0.654 38	1.1641×10^{-82}	0.542 79	8.0091×10^{-83}	1.1070×10^{-30}	$9.055\,34 \times 10^{-2}$
Q(3)	$v=0, J=3 \rightarrow v=1, J=3$	2980.99	0.652 40	1.1571×10^{-82}	0.539 51	7.9126×10^{-83}	1.0989×10^{-30}	$1.123\,63 \times 10^{-1}$
Q(2)	$v=0, J=2 \rightarrow v=1, J=2$	2987.29	0.650 91	1.1518×10^{-82}	0.537 05	7.8405×10^{-83}	1.0938×10^{-30}	$3.850\,00 \times 10^{-1}$
Q(1)	$v=0, J=1 \rightarrow v=1, J=1$	2991.51	0.649 91	1.1483×10^{-82}	0.535 40	7.7927×10^{-83}	1.0969×10^{-30}	$2.074\,16 \times 10^{-1}$
Q(0)	$v=0, J=0 \rightarrow v=1, J=0$	2993.62	0.649 42	1.1465×10^{-82}	0.534 58	7.7688×10^{-83}	1.0690×10^{-30}	$1.854\,28 \times 10^{-1}$
S(0)	$v=0, J=0 \rightarrow v=1, J=2$	3166.36	0.585 75	9.3273×10^{-83}	0.511 02	7.0990×10^{-83}	5.6926×10^{-32}	$1.854\,28 \times 10^{-1}$
S(1)	$v=0, J=1 \rightarrow v=1, J=3$	3278.52	0.543 94	8.0430×10^{-83}	0.496 07	6.6897×10^{-83}	3.1498×10^{-32}	$2.074\,16 \times 10^{-1}$
S(2)	$v=0, J=2 \rightarrow v=1, J=4$	3387.26	0.502 77	6.8716×10^{-83}	0.481 85	6.3118×10^{-83}	2.4941×10^{-32}	$3.850\,00 \times 10^{-1}$
S(3)	$v=0, J=3 \rightarrow v=1, J=5$	3492.09	0.462 29	5.8097×10^{-83}	0.468 34	5.9626×10^{-83}	2.1374×10^{-32}	$1.123\,63 \times 10^{-1}$
S(4)	$v=0, J=4 \rightarrow v=1, J=6$	3592.57	0.422 52	4.8532×10^{-83}	0.455 46	5.6393×10^{-83}	1.8918×10^{-32}	$9.055\,34 \times 10^{-2}$

a: For perpendicularly polarized detection, use the depolarization ratio (ρ) with ($d\sigma/d\Omega$). See Table S16 and Ref. [10]
b: For ($\parallel + \perp$) detection, $(1+\rho) \times (d\sigma/d\Omega)$ with the appropriate ρ values.
See Ref. [6] and online repository[7] for programs to compute the ro-vibrational matrix elements at other excitation wavelengths in the range of 182–1320 nm

A. Error in the matrix elements of polarizability invariants

In our earlier report[6], a clamped-nuclei potential for H_2 reported by Wolniewicz[11] was employed, to first compute the relevant ro-vibrational wavefunctions. These wavefunctions were then used with *ab-initio* inter-nuclear and wavelength dependent polarizability invariants computed using response theory reported by Christiansen et al.[12–14] to obtain the required ro-vibrational matrix elements of polarizability invariants specific to certain excitation wavelength.

In this calculation, errors can be introduced from many sources which were identified and the magnitudes were estimated (the following discussion concerns with the numerical procedure used in Ref. [6] for computation of the polarizability invariants).

1. The potential energies and the corrections were interpolated using cubic spline interpolation procedure while solving the radial nuclear equation. Error introduced by the cubic spline interpolation was estimated by removing each data point one by one and generating new interpolated data which was compared to the original value, thus giving information about the maximal oscillation of the cubic spline in the absence of a data point. The maximal shift in the numerical value of the potential, and corrections for the potential was found to be $\sim 10^{-7}$ a.u. A similar analysis on the polarizability data points over distance showed a maximal shift of $\sim 10^{-6}$ a.u.

2. Any effect of random noise in the potential energy surface was investigated by introducing artificial white noise oscillating within $\pm 1 \times 10^{-6}$ having random statistical distribution. Wavefunctions obtained by such perturbed potentials were used to calculate the matrix elements. This was done for 31 random noise vectors for statistical analysis and the 3σ standard deviation in the values of the matrix elements is shown in Table S10. The net effect on the matrix element was below 10^{-5} a.u.

TABLE S10: Effect of random noise (oscillating within $\pm 10^{-6}$ a.u.) in the potential energy on the matrix elements (static)

Matrix element	3σ
$\langle \psi_{0,0} \vert \gamma \vert \psi_{0,0} \rangle$	8.7×10^{-6}
$\langle \psi_{0,0} \vert \bar{\alpha} \vert \psi_{0,0} \rangle$	7.1×10^{-6}
$\langle \psi_{0,0} \vert \gamma \vert \psi_{0,2} \rangle$	8.8×10^{-6}
$\langle \psi_{0,0} \vert \bar{\alpha} \vert \psi_{1,0} \rangle$	2.7×10^{-6}

3. During generation of the **H**-matrix, the reduced nuclear mass(μ) of the molecule was required. The masses of proton and deuteron have their respective uncertainties which propagates to the reduced mass and then to the elements of the **H**-matrix. In order to estimate the maximal error due to uncertainty in nuclear masses, two **H**-matrices were setup, one with the error added to both the nuclei and second, where the error was subtracted. The resulting wavefunctions were used to calculate the matrix elements, from which the maximal error in respective matrix elements was determined. These are shown in Table S11 where it is seen that the error due to uncertainty of nuclear masses is of the order of 10^{-11} a.u. Among the three isotopologues, due to the smaller nuclear mass of a proton the contribution of error due to nuclear mass is more in H_2, which decreases in HD and is the lowest in D_2.

TABLE S11: Maximal error in the matrix elements (static) due to the uncertainty in nuclear mass

	$\Delta\langle \psi_{0,0} \vert \bar{\alpha} \vert \psi_{0,0} \rangle$	$\Delta\langle \psi_{0,0} \vert \gamma \vert \psi_{0,0} \rangle$
H_2	2.5×10^{-11}	3.6×10^{-11}
HD	1.2×10^{-11}	1.5×10^{-11}
D_2	6.3×10^{-12}	6.3×10^{-12}

4. Finite differences were used to model the 1-D Schrödinger equation (see Ref. 6 and the associated supplementary material), in which the **H**-matrix was set up using an array of inter-nuclear distance of certain step size. The choice of step size (h) was crucial to obtain resulting wavefunctions of sufficient accuracy. With the step (h) of 0.004 a.u., the truncation error when keeping the Taylor series expansion up to the 4^{th}-order derivative was estimated (see more details in Ref. 6), giving a net maximal error of

the order of 10^{-8} a.u. Effect of the step size and numerical rounding off in computation on the matrix element was estimated by calculating matrix elements of $\bar{\alpha}$ and γ using the wavefunctions obtained from using different step sizes. From this analysis, uncertainty of the order of 10^{-8} a.u. due to above mentioned numerical errors and approximations was found for any matrix element when the step, h=0.004 a.u. was used.

5. When computing the wavelength dependent matrix elements of the polarizability invariants, the exact wavelength (say λ_i) at which the matrix element is sought might differ by up to a few nanometers with respect to the wavelength at which the invariant over distance is available (say $\lambda_i \pm q_i$). In the present work, q_i is up to $\sim$5 nm. In such case, there are two ways to compute the matrix element. First, the polarizability invariant can be interpolated over wavelength to obtain the required invariant as a function of distance for λ_i followed by computation of the matrix element. Secondly, the matrix elements can be computed at $\lambda_i + q_i$ (or $\lambda_i - q_i$) followed by interpolation (of the matrix elements) to obtain an interpolated value at λ_i. Both of these methods were tested to compute the respective matrix element at a given wavelength and the difference was found to be less than 1.0×10^{-7} a.u. revealing the low error in spline interpolation procedure achievable due to the densely spaced polarizability data points over wavelength.

The combined effect of all the errors discussed above, namely error due to spline interpolation, random noise in the potential, uncertainty in nuclear mass, the step size when generating the **H**-matrix, the uncertainty due to numerical rounding, and interpolation of matrix elements over wavelength, is no more than $\mathbf{1 \times 10^{-3}}$ **a.u.** for any matrix element, and this is regarded as the maximal numerical error. The error in the square of matrix elements listed in Table S8–S9, estimated via error propagation, is around **0.3%**.

Sec S6. Analysis for perpendicularly polarized Raman spectra

The C_0 correction is derived from the wavenumber axis of the Raman spectrum. Thus, it is same for the parallel as well as the perpendicularly polarized dataset. (The C_0 correction is shown in Fig. 5 of the main document.)

The C_1 correction obtained from the perpendicularly polarized broadband white light spectrum is shown below.

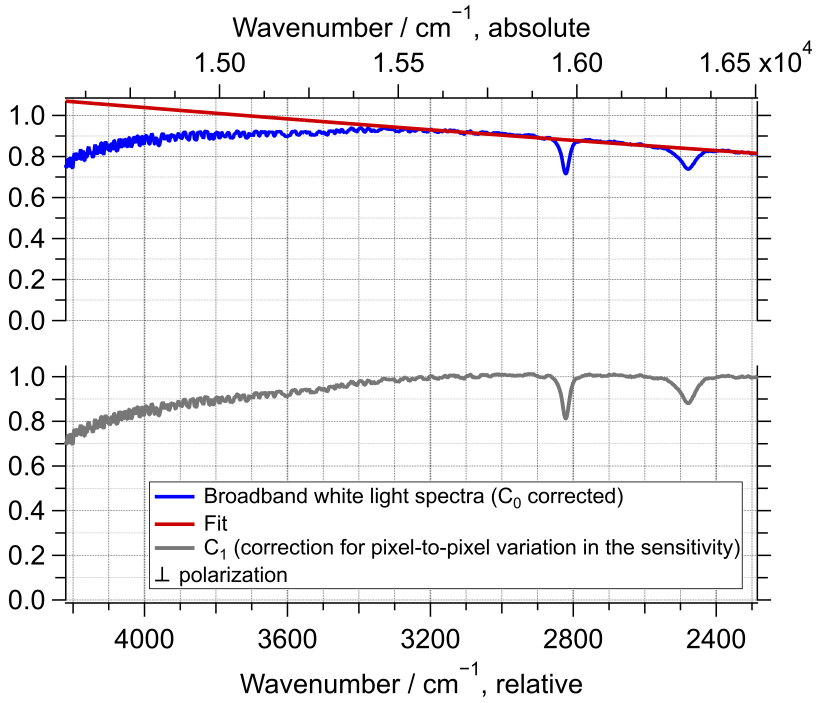

FIG. S12: The C_1 correction for the pixel-to-pixel difference in the wavenumber-dependent sensitivity in the studied spectral region. This correction (black line) was obtained by a ratio of the observed perpendicularly polarized broadband white-light spectrum (blue line) to the fit, assuming black-body emission (red line).

Next, the C_0 and C_1 corrections were applied to the perpendicularly polarized Raman spectrum of the vibration-rotation bands of H_2, HD and D_2 (shown in Fig. S13). In Fig. S13(d)–(f), the $Q(0)$ band for the three gases should ideally be negligible (they have depolarization ratio of zero, see Table S16), however a small feature was observed in our experiments. This indicates that the polarization response of the spectrometer is not perfect. This is attributed to (*i*) the incident laser not being linearly polarized, (*ii*) polarization aberration due to the microscope objective lens, and (*iii*) the limited extinction of the analyzer in the collection path, of the Raman spectrometer.
Nonetheless, we proceeded with determination of the C_2 correction for this dataset using the band areas of the observed Raman bands, shown in Fig S13, to evaluate the magnitude of errors encountered in this analysis.

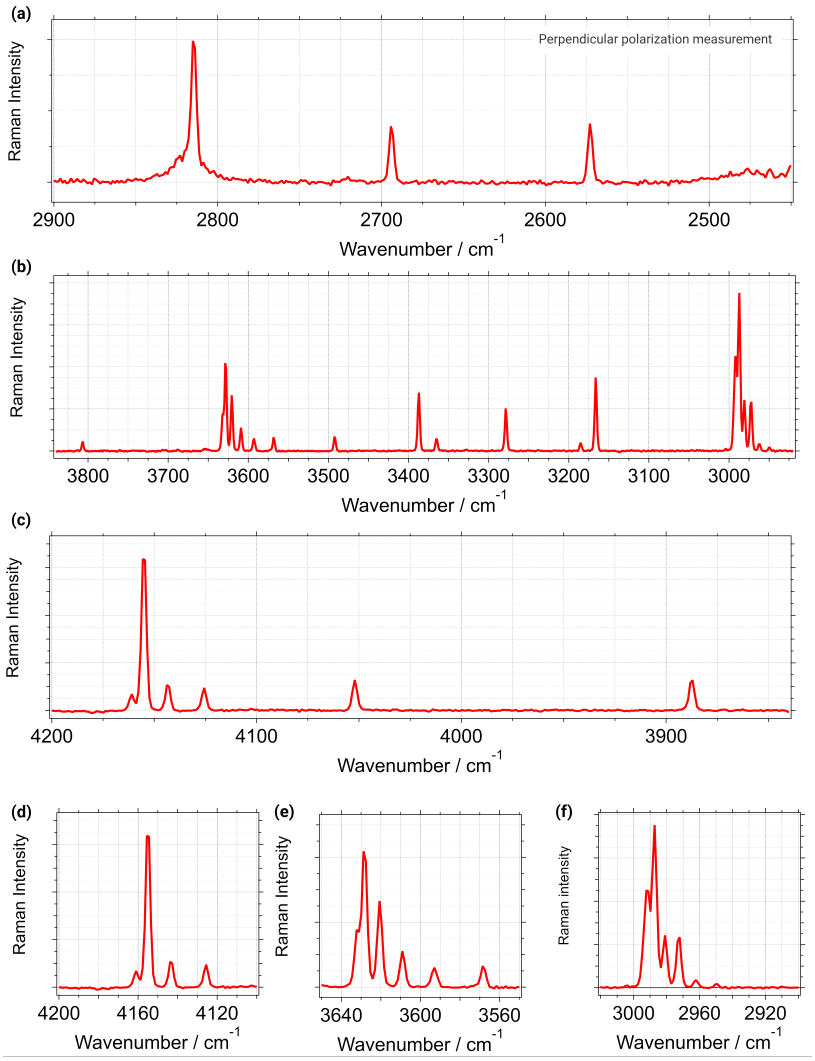

FIG. S13: Perpendicularly polarized vibration-rotation Raman spectra of H_2, HD and D_2. Sub-fig (a)–(c) show the full spectrum divided into three sections. Sub-fig (d)–(f) show the zoomed up Q-branches of H_2, HD and D_2, respectively.

Intensities for the perpendicularly polarized vibration-rotation Raman bands from diatomic molecules are given by[15]

$$I_{O(J)} = I_0\, k_\nu\, \nu_o \nu_s^3 N F_J \left(\frac{1}{10} \frac{J(J-1)}{(2J-1)(2J+1)} \langle \gamma \rangle^2_{v=0,J;v=1,J+2} \right). \tag{14}$$

$$I_{Q(J)} = I_0\, k_\nu\, \nu_o \nu_s^3 N F_J \left(\frac{1}{15} \frac{J(J+1)}{(2J-1)(2J+3)} \langle \gamma \rangle^2_{v=0,J;v=1,J-2} \right),$$

$$I_{S(J)} = I_0\, k_\nu\, \nu_o \nu_s^3 N F_J \left(\frac{1}{10} \frac{(J+1)(J+2)}{(2J+1)(2J+3)} \langle \gamma \rangle^2_{v=0,J;v=1,J-2} \right),$$

where I_0 is the incident laser flux, $\nu_o \nu_s^3$ is the frequency factor, N is the number of molecules, F_J is the fraction of molecules populating the initial state, and the term in brackets includes the ro-vibrational matrix elements of polarizability invariants along with the Placzek-Teller factor, $b_J^{\Delta J=0,\pm 2}$.

The C_2 correction determined from this analysis is shown in Fig. S14, along with the corresponding C_2 correction obtained from dataset on the parallel polarized spectra. **Temperature simultaneously determined from this analysis was 294.7 K (the temperature recorded from inside the gas cell was 299.5 K).**

Thus, a larger deviation of around 5K was found, in comparison to the analysis on the parallel polarized data where this deviation was around 1.3 K.

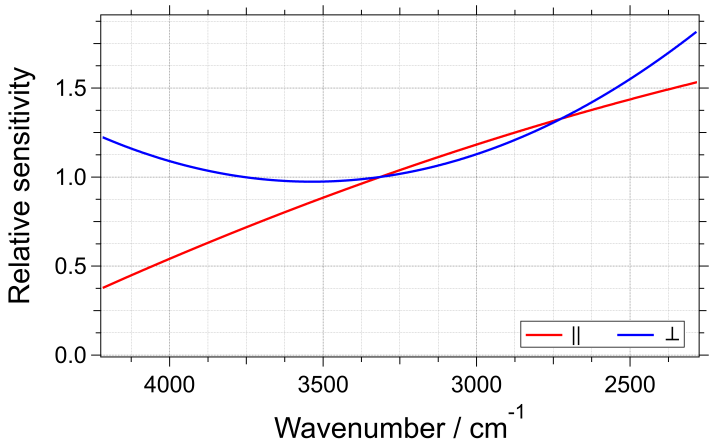

FIG. S14: The C_2 correction determined in our work for parallel and perpendicular detection measurements. A quadratic polynomial was used for modeling the C_2 correction for parallel polarized detection, while a cubic polynomial was used for the perpendicularly polarized detection.

Further, a correction for polarization was performed using the O and S bands which are depolarized ($\rho = 0.75$). The constant correction factor for this purpose was determined by comparing the observed depolarization ratio after the intensity calibration procedure (see Fig. S15). After this correction, the depolarization values of the O and S bands were within 0.71–0.78. Following this correction, the depolarization ratios of the polarized Q bands were checked, which showed significant deviations from the reference values obtained from theory[6] and experiments[10]. This comparison is shown in Table. S16. The observed deviations are likely due to the imperfect polarization response of our Raman spectrometer (as discussed earlier). Other possible reasons include the large uncertainty in the determination of band area of the unresolved Q-branch, and the much weaker bands in the perpendicularly polarized spectra compared to the parallel polarized case. This comparison also demonstrates the limit of the present results, in terms of the accuracy of the depolarization ratios.

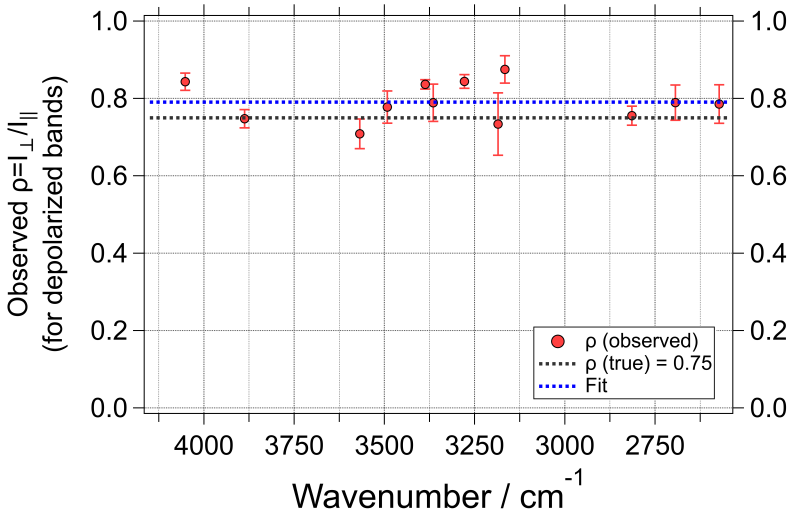

FIG. S15: The observed ratios of perpendicular to parallel polarized Raman intensities of vibration-rotation Raman bands of H_2, HD, and D_2. All of these bands are depolarized with an anticipated true depolarization ratio of 0.75 (shown in dashed black line). This feature is used for correcting the polarization dependence of the observed Raman intensities. The fit shown in broken blue color is used for obtaining the required correction for the perpendicular polarized spectra.

TABLE S16: Depolarization ratios (ρ) of the Q1(J=0–3) bands of H_2, HD and D_2, obtained after intensity calibration (including correction for polarization) of the Raman spectra. Corresponding reference values from experiments and our previous calculations are also listed.

H_2				
J	ν (cm^{-1})	ρ (present expt.)	ρ (expt. Ref. [10])	ρ (our calc[6])
0	4161.17	0.0308 (0.02)	0.0000	0.0000
1	4155.25	0.0518 (0.03)	0.0177(0.0006)	0.0183
2	4143.47	0.0416 (0.03)	0.0133(0.0006)	0.0133
3	4125.87	0.0365 (0.02)	0.0125(0.0006)	0.0125
HD				
J	ν (cm^{-1})	ρ (present expt.)	ρ (expt. Ref. [10])	ρ (our calc[6])
0	3632.16	0.0384 (0.02)	0.0000	0.0000
1	3628.30	0.0497 (0.03)	0.0177(0.0006)	0.0180
2	3620.61	0.0470 (0.03)	0.0126(0.0006)	0.0130
3	3609.11	0.0469 (0.02)	0.0121(0.0006)	0.0122
D_2				
J	ν (cm^{-1})	ρ (present expt.)	ρ (expt. Ref. [10])	ρ (our calc[6])
0	2993.62	0.0191 (0.01)	0.0000	0.0000
1	2991.51	0.0476 (0.03)	0.0174(0.0006)	0.0177
2	2987.29	0.0373 (0.03)	0.0118(0.0006)	0.0127
3	2980.99	0.0504 (0.02)	0.0112(0.0004)	0.0120

Final remarks:
Results obtained from the perpendicularly polarized dataset show an increased deviation of the temperature obtained from the analysis ($T_{analysis}$) in comparison to the true temperature (T_{expt}): $|T_{analysis} - T_{expt}|$ = 5K.

The depolarization ratios of the depolarized vibration-rotation bands after correction were between 0.71–0.78. The depolarization ratios of the polarized Q bands were larger by $\sim$0.03 compared to the true values. Using the depolarization ratios of the Q bands to test the polarization response of the Raman spectrometer is a rather stringent test, which needs a careful control of geometrical parameters in the excitation and collection of Raman photons to get accurate results.[10]

Sec S7. Band areas of the studied Raman transitions

In our analysis, the areas of the studied Raman bands were determined by band fitting. The O- and S-bands were well resolved, and these bands were fit individually. The Q-bands were partially resolved in the observed Raman spectra, and hence they were fit using a combination of Gaussian functions[6] whose centers were fixed to the known Raman transition frequencies known earlier.[4, 5] Fig. S17–S18 show the fit of the Q-branches for H_2, HD and D_2 with the individual peaks and the residuals, respectively.

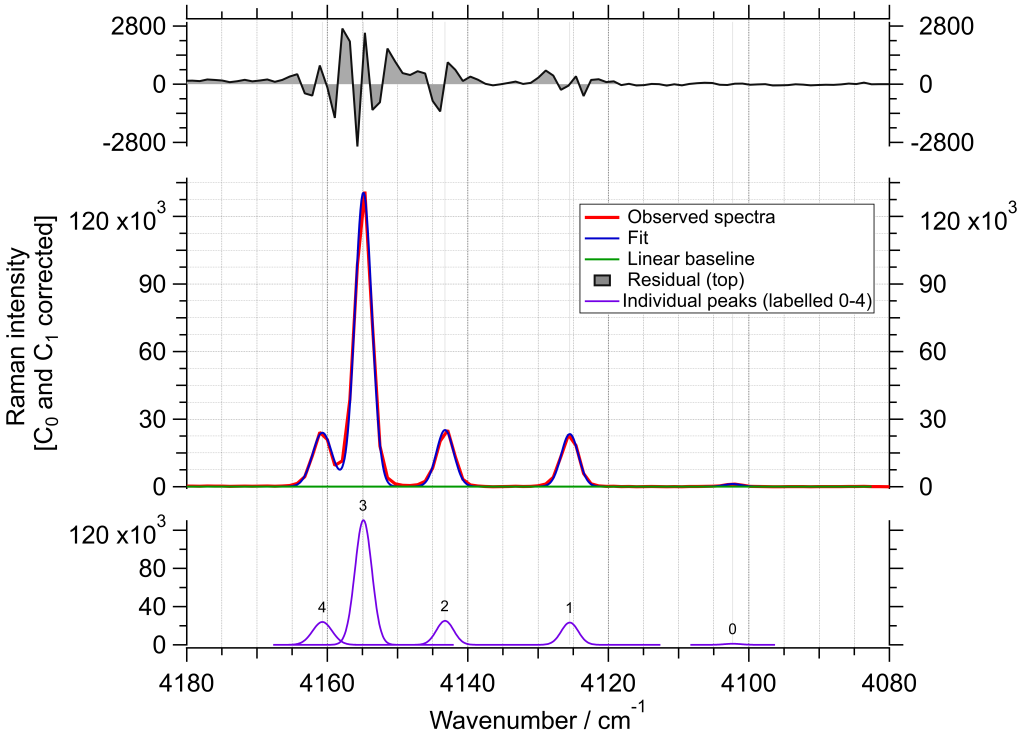

FIG. S17: Plot showing fit of the Q-branch of H_2 to a combination of 5 Gaussian functions. Band areas of the individual peaks was determined as the integral of the Gaussian peaks.

[6] Lorentzian and Voigt functions were also tested for fitting. When using Lorentzian functions, larger residuals were obtained. Fitting with Voigt function showed prominent contribution from the Gaussian term relative to the Lorentzian term. Hence, Gaussian functions were eventually used.

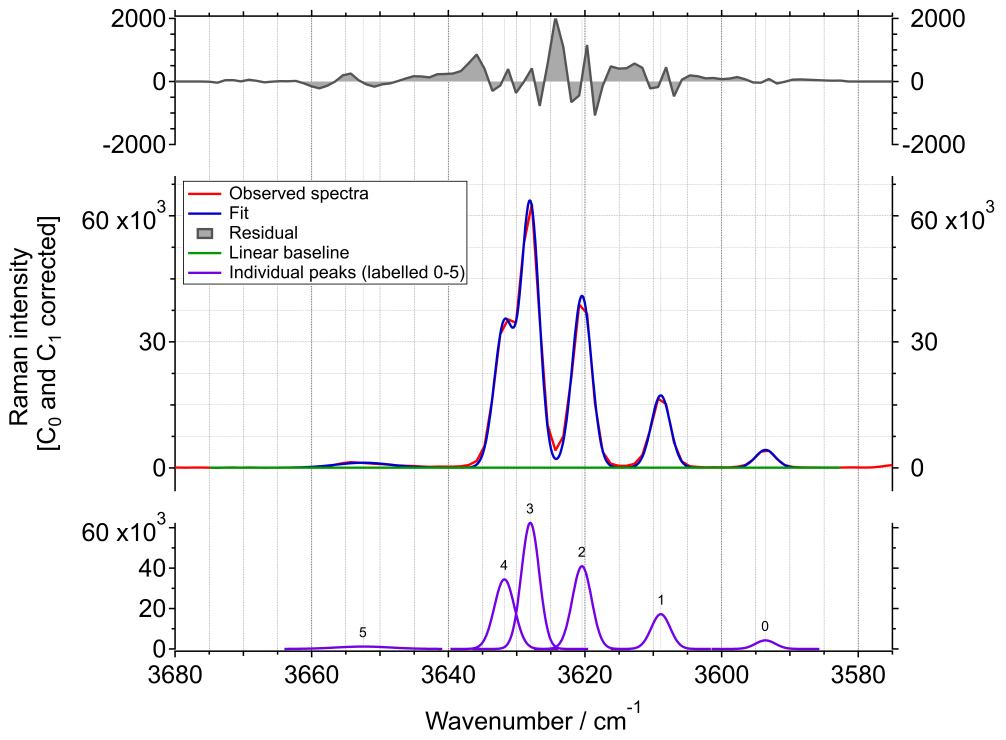

FIG. S18: Plot showing fit of the Q-branch of HD to a combination of 5 Gaussian functions. Here, peak 0–4 correspond to the Q-bands, while peak 5 is from H_2O. Band areas of the individual peaks was determined as the integral of selected the Gaussian peaks.

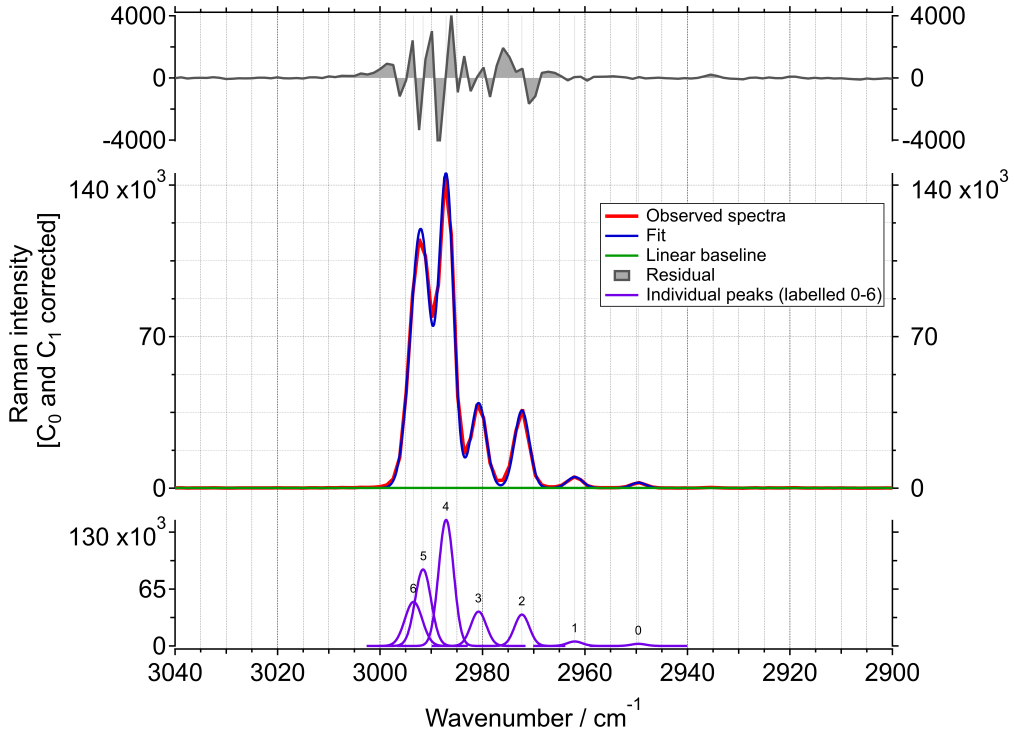

FIG. S19: Plot showing fit of the Q-branch of D_2 to a combination of 7 Gaussian functions. Band areas of the individual peaks was determined as the integral of the Gaussian peaks.

Sec S8. Additional details on error analysis

A. Error in $\mathbb{R}_{obs}$

• Error in individual experimental Raman intensity was estimated using, (***i***) statistical error, and (***ii***) error from peak fitting.

(***i***) Statistical error was determined from previous experiments where Raman spectra from the mixture of gases were acquired at various exposure times to yield Raman peaks with varying signal-to-noise ratios. Each of these sets of spectra were analyzed for the standard deviations in the band intensities (determined as band areas via peak fitting). The obtained profile is shown in Fig. S20.

In the actual spectra analyzed for the determination of the C_2 correction, the statistical error in the O and S bands ranged from 3% to 10%, while for the Q bands it was up to 5%.

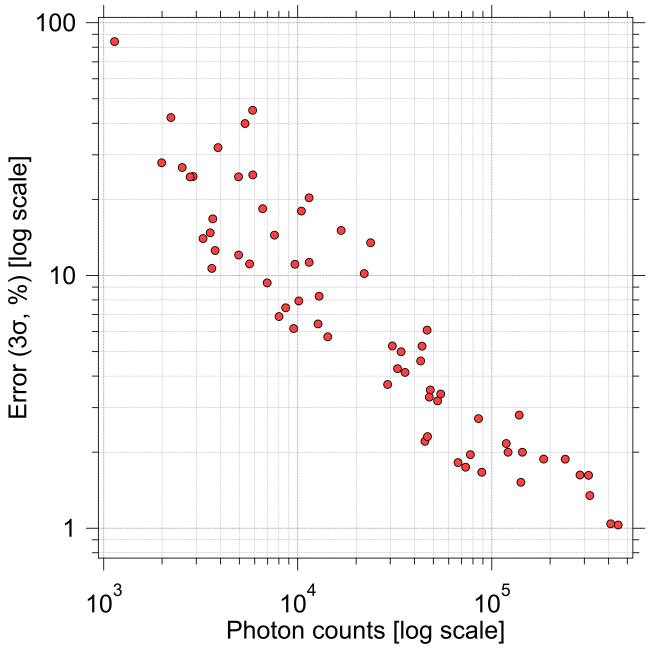

FIG. S20: Plot of experimental error over the magnitude of band area (both shown in log scales). The actual experimental error (3σ) is obtained from set of Raman spectra of H_2, HD and D_2 measured with varying exposures each having 6 spectra. Band area was determined using fitting. This set of result is specific to the present spectrometer and largely depends on the CCD, its measurement parameters and the overall spectrometer alignment.

(***ii***) Error encountered in the peak fitting was checked by looking at the residuals, and noting the approximated uncertainty in the band area reported by the fitting program (MultiPeak fit in IgorPro[16]). This uncertainty ranged from 1 to 12% for the Q bands, depending on intensity of the specific Q band and whether the band was resolved. For the O and S bands, the reported uncertainty during the peak fitting process ranged from 2 to 4%.

The net error in the determined band area of a peak was then estimated as, $\sqrt{\sigma^2_{\text{statistical}} + \sigma^2_{\text{peak fitting}}}$.

B. Error in $\mathbb{R}_{true}$

• The net error in the computed elements of the $\mathbb{R}_{true}$ matrix was below 0.5%. We did not consider the contribution of this error in the non-linear optimization analysis. The origin of this error is from the ro-vibrational matrix elements of the polarizability invariants for H_2, HD and D_2, and the uncertainties in the Boltzmann populations of the initial states relevant to the studied Raman transitions. See Sec S2 C and Sec S5 A for more details.

Bibliography

[1] A repository containing procedures developed in IgorPro for Raman data analysis., https://github.com/ankit7540/RamanSpec_BasicOperations (2021), accessed: 2021-07-15.

[2] A repository containing Python modules for intensity calibration using band intensities of Raman bands., https://github.com/ankit7540/IntensityCalbr (2021), accessed: 2021-07-12.

[3] P. Virtanen, , R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt, SciPy 1.0: fundamental algorithms for scientific computing in python, Nature Methods **17**, 261 (2020).

[4] J. Komasa, K. Piszczatowski, G. Łach, M. Przybytek, B. Jeziorski, and K. Pachucki, Quantum Electrodynamics Effects in Rovibrational Spectra of Molecular Hydrogen, J. Chem. Theory Comput. **7**, 3105 (2011).

[5] K. Pachucki and J. Komasa, Rovibrational levels of HD, Phys. Chem. Chem. Phys. **12**, 9188 (2010).

[6] A. Raj, H. Hamaguchi, and H. A. Witek, Polarizability tensor invariants of H_2 , HD and D_2, J. Chem. Phys. **148**, 104308 (2018).

[7] A repository containing a FORTRAN program, a python module along with polarizability data and rovibrational wavefunctions, for the calculation of rovibrational matrix elements of polarizability for H_2, HD and D_2., https://github.com/ankit7540/H2-PolarizabilityMatrixElements (2018), accessed: 2021-07-15.

[8] A. Raj, C. Kato, H. A. Witek, and H. Hamaguchi, Toward standardization of raman spectroscopy: Accurate wavenumber and intensity calibration using rotational raman spectra of H_2, HD, D_2, and vibration–rotation spectrum of O_2, J. Raman Spectrosc. **51**, 2066 (2020).

[9] A repository containing Igor procedures and Python modules for wavenumber and intensity calibration using Raman spectra from H_2, HD, D_2 and O_2., https://github.com/ankit7540/RamanSpecCalibration (2020), accessed: 2020-02-28.

[10] T. M. James, M. Schlösser, S. Fischer, M. Sturm, B. Bornschein, R. J. Lewis, and H. H. Telle, Accurate depolarization ratio measurements for all diatomic hydrogen isotopologues, J. Raman Spectrosc. **44**, 857 (2013).

[11] L. Wolniewicz, Relativistic energies of the ground state of the hydrogen molecule, J. Chem. Phys. **99**, 1851 (1993).

[12] O. Christiansen, A. Halkier, H. Koch, P. Jørgensen, and T. Helgaker, Integral-direct coupled cluster calculations of frequency-dependent polarizabilities, transition probabilities and excited-state properties, J. Chem. Phys. **108**, 2801 (1998).

[13] T. Helgaker, P. Jørgensen, and J. Olsen, Exact and approximate wave functions, in *Molecular Electronic-Structure Theory* (John Wiley & Sons, Ltd, 2000) Chap. 4, pp. 107–141, https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119019572.ch4.

[14] T. Helgaker, P. Jørgensen, and J. Olsen, Coupled-cluster theory, in *Molecular Electronic-Structure Theory* (John Wiley & Sons, Ltd, 2000) Chap. 13, pp. 648–723, https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119019572.ch13.

[15] D. A. Long, *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules.* (John Wiley & Sons Ltd., Baffins Lane, Chichester, England, 2002).

[16] Igor Pro, WaveMetrics, Lake Oswego, OR, USA, A scientific data analysis software with numerical computing environment and a programming language.