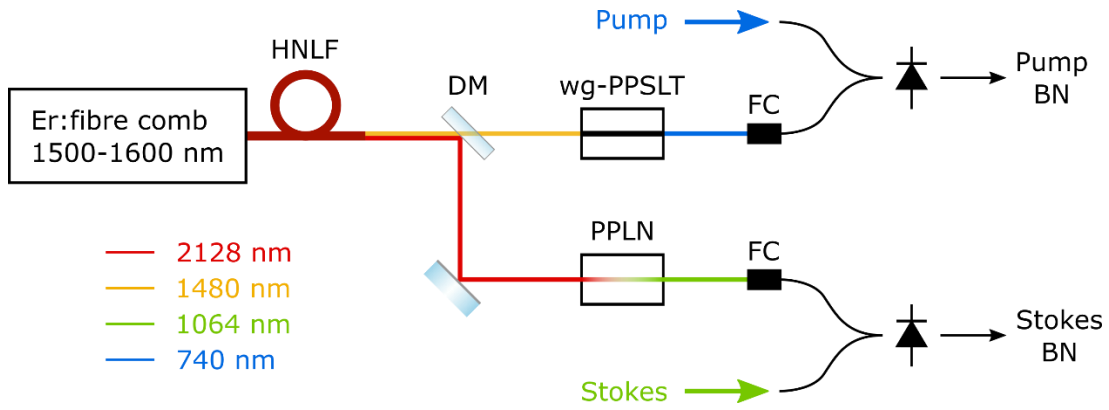
Supplementary Information for: Stimulated Raman Scattering Metrology of Molecular Hydrogen

Marco Lamperti[1,*], Lucile Rutkowski[2], Daniele Ronchetti[1,†], Davide Gatti[1], Riccardo Gotti[1,‡], Giulio Cerullo[1], Franck Thibault[2], Hubert Jóźwiak[3], Szymon Wójtewicz[3], Piotr Masłowski[3], Piotr Wcisło[3], Dario Polli[1] and Marco Marangoni[1]

[1] Dipartimento di Fisica - Politecnico di Milano and IFN-CNR, Via Gaetano Previati 1/C, 23900 Lecco, Italy
[2] Univ Rennes, CNRS, IPR (Institut de Physique de Rennes)-UMR 6251, F-35000 Rennes, France
[3] Institute of Physics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, Grudziadzka 5, 87-100 Torun, Poland
* marco.lamperti@uninsubria.it; currently at Dipartimento di Scienza e Alta tecnologia, Università degli studi dell'Insubria, 22100 Como, Italy
[†] Currently at Department of Physics, Universität Hamburg, Jungiusstrasse 9, 20355 Hamburg, Germany and Max Planck School of Photonics, Friedrich-Schiller University of Jena, Albert-Einstein-Str. 6, 07745 Jena, Germany
[‡] Currently at Dipartimento di Ingegneria Industriale e dell'Informazione, Università di Pavia, Via Ferrata 5, 27100 Pavia, Italy

Supplementary Note 1: Comb referencing of pump and Stokes lasers



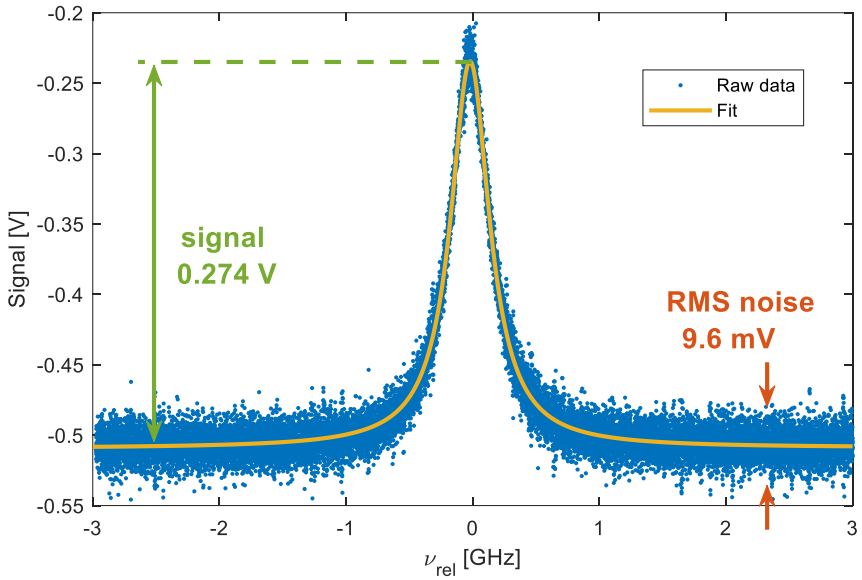
Supplementary Figure 1. Referencing scheme of the pump and Stokes lasers to the optical frequency comb. HNLF: highly nonlinear fibre; DM: dichroic mirror (crossover wavelength at 1550 nm); wg-PPSLT: optical waveguide realized on a periodically-poled stoichiometric lithium tantalate crystal; PPLN: periodically-poled lithium niobate crystal; FC: fibre coupler; BN: beat note.

Supplementary Figure 1 shows the optical setup used to calibrate the frequencies of the pump and Stokes fields. It makes use of an Er:fibre femtosecond oscillator whose repetition rate is locked to a 100 MHz signal obtained from a low-noise RF synthesizer, which is in turn referenced to a GPS-disciplined Rb clock. The oscillator output goes through a home-made Er:fibre amplifier and a highly nonlinear fibre to generate a coherent supercontinuum ranging from about 1040 to 2150 nm. The supercontinuum is split into two parts through a dichroic beam splitter with cut-off frequency at 1550 nm. The high-frequency part is coupled to a nonlinear optical waveguide embedded in a periodically-poled stoichiometric lithium tantalate crystal that provides efficient second harmonic generation around 740 nm. The output of the waveguide is fibre coupled and mixed with part of the pump radiation to obtain the corresponding beat note. The low-frequency part of the continuum is focused on a periodically-poled lithium niobate crystal to obtain by second harmonic generation a comb around 1064 nm: this is coupled into an optical fibre and mixed with the Stokes radiation to obtain the corresponding beat note. With this approach the absolute calibration of the spectral distance

between pump and Stokes lasers results independent of the carrier-envelope frequency f_{ceo} of the comb, which is left free running in our experiment. The spectral distance remains fully defined by the comb repetition frequency and by the difference between the pump and Stokes beat notes (see Methods in the main text for related equations).

Supplementary Note 2: Signal-to-noise ratio of SRS spectra

In this section we show that the stimulated Raman loss (SRL) was measured under conditions very close to the shot noise limit. To this aim, we measured the signal to noise ratio (SNR) in a single spectral scan and compared it to the shot noise level expected from the power impinging on the detection photodiode. Supplementary Figure 2 shows the voltage Raman signal and the corresponding SRL measured at a pressure of 1 bar, with a scan duration of 500 ms and an integration time of 10.24 µs for each spectral point. By fitting the raw data with a Lorentzian profile, a signal of 274 mV is extracted at the peak, where $\mathrm{SRL} = 3.7 \cdot 10^{-4}$, to be compared with a root mean square (RMS) noise of 9.6 mV measured on the raw data far from resonance. The experimental SNR is then $\mathrm{SNR_{exp}} = 28.5$.



Supplementary Figure 2. Single-scan measurement of the Q(1) line of H_2 at 1 bar of pressure. Points are measured data, yellow line is a Lorentzian fit.

The shot-noise limited SNR for a beam with optical power P_{opt} impinging on a photodiode with responsivity ρ over an integration time τ is independent on the gain of the detector chain, and is given by the square root of the number of photoelectrons N_e generated in the photodiode during the integration time τ:

$$\mathrm{SNR_s} = \sqrt{N_e} = \sqrt{\frac{\rho\, P_{\mathrm{opt}}\, \tau}{e}} = 9.5 \cdot 10^4$$

where $\rho\, P_{opt}\, \tau$ is the electric charge generated in the photodiode during the integration time and e is the electron charge. This equation assumes an impinging pump beam with Poisson photon statistics, whose power fully contributes to the measured signal. It has been quantified using the experimental parameters $P_{\mathrm{opt}} = 320\ \mu\mathrm{W}$, $\rho = 0.45\,\frac{\mathrm{A}}{\mathrm{W}}$. To compare $\mathrm{SNR_s}$ with $\mathrm{SNR_{exp}}$ one needs to divide the latter by the SRL peak, since the peak voltage signal of 0.274 V in Supplementary Figure 2 corresponds to an impinging average pump power that is 1/SRL times higher. The normalized value for $\mathrm{SNR_{exp}}$ is:

$$\mathrm{SNR_{exp,norm}} = \frac{\mathrm{SNR_{exp}}}{\mathrm{SRL}} = 7.7 \cdot 10^4$$

which is only 20% below $\mathrm{SNR_s}$.

Supplementary Note 3: Quantum-scattering calculations of collisional parameters.

Quantum scattering calculations are performed on the potential energy surface (PES) of the H_2-H_2 system reported in Ref. 1 within the Born-Oppenheimer approximation for the separation of electronic and nuclear motion. The PES is six-dimensional, i.e. it depends on the intermolecular distance, *R*, the three Jacobi angles, θ_1, θ_2 and $\phi = \phi_1 - \phi_2$, and intramolecular distances, r_1 and r_2 (see Ref. 1 for details). For the purpose of scattering calculations, the PES is expanded over bispherical harmonics, $I_{l_1 l_2 l_{12}}(\theta_1, \theta_2, \phi)$:

$$V(R, r_1, r_2, \theta_1, \theta_2, \phi) = \sum_{l_1, l_2, l_{12}} A_{l_1 l_2 l_{12}}(R, r_1, r_2) I_{l_1 l_2 l_{12}}(\theta_1, \theta_2, \phi),$$

where the bispherical harmonics are defined as

$$I_{l_1 l_2 l_{12}}(\theta_1, \theta_2, \phi) = \sqrt{\frac{2l_{12}+1}{4\pi}} \sum_m C^{l_1 l_2 l_{12}}_{m,-m,0} Y_{l_1,m}(\theta_1, \phi_1) Y_{l_2,m}(\theta_2, \phi_2).$$

Here, $C^{l_1 l_2 l_{12}}_{m_1, m_2, m_{12}}$ is the Clebsch-Gordan coefficient, and $Y_{lm}(\theta, \phi)$ denotes the spherical harmonic.

The expansion coefficients (radial coupling terms), $A_{l_1 l_2 l_{12}}(R, r_1, r_2)$, are averaged over internuclear coordinates of the active (r_1) and perturbing (r_2) molecules. The spectroscopically active molecule may be in the ground or first excited vibrational states while the perturbing one is always in the ground state (at least at room temperature). Thus, the average is performed by integration of the radial coupling terms with the weight corresponding to the squared modulus of the isolated H_2 wavefunction in the v=0, j=1 and v=1, j=1 rovibrational states (for the average over r_1) and v=0, j=0 (for the average over r_2 and collisions with *para*-H_2) or v=0, j=1 (for the average over r_2 and collisions with *ortho*-H_2). We use 15 radial coupling terms in the scattering calculations with l_1, l_2, l_{12}, up to 4,4,8, to solve the close-coupling equations using the MOLSCAT code[2]. Calculations are performed for 265 kinetic energies in range 0.1 - 6000 cm^{-1} with a step dependent on the range of kinetic energies, gradually increasing from 0.1 cm^{-1} (for E_{kin} < 10 cm^{-1}) to 500 cm^{-1} (for E_{kin} > 2000 cm^{-1}). Scattering S-matrices are used to calculate the generalized spectroscopic cross sections[3–5]:

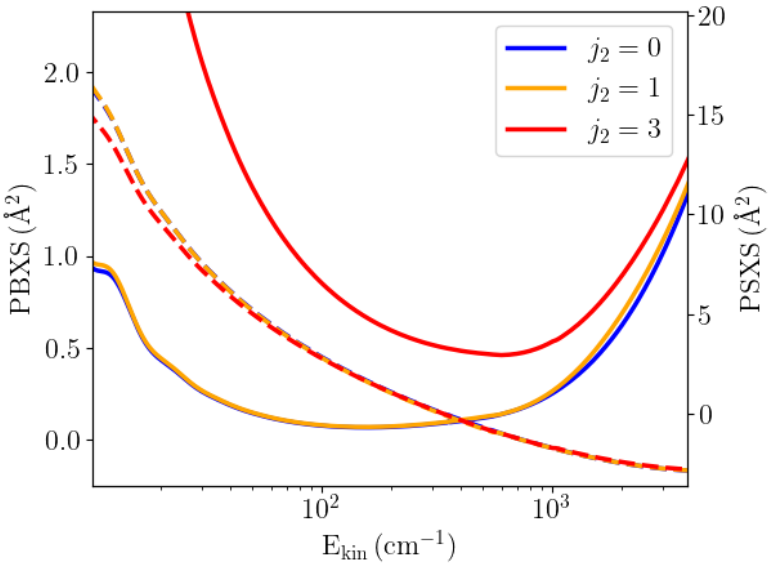
$$\sigma_\lambda^q(v_i, j_i, v_f, j_f, j_2; E_{kin}) = \frac{\pi}{k^2}(-1)^{\lambda+j_2} \sum_{j_2', l, l', \bar{l}, \bar{l}'} i^{l-l'-\bar{l}+\bar{l}'} (-1)^{l'-\bar{l}'-j_2'} \sqrt{[l, l', \bar{l}, \bar{l}']}$$
$$\times \begin{pmatrix} l & \bar{l} & \lambda \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l' & \bar{l}' & \lambda \\ 0 & 0 & 0 \end{pmatrix} \sum_{J, \bar{J}} [J, \bar{J}] \sum_{L, L', \bar{L}, \bar{L}'} \sqrt{[L, L', \bar{L}, \bar{L}']}$$
$$\times \begin{Bmatrix} \lambda & \bar{l} & l \\ j_2 & L & \bar{L} \end{Bmatrix} \begin{Bmatrix} \lambda & \bar{l}' & l' \\ j_2' & L' & \bar{L}' \end{Bmatrix} \begin{bmatrix} L & L' & j_f & j_f \\ \bar{L} & j_i & \bar{L}' & j_i \\ \lambda & \bar{J} & J & q \end{bmatrix}$$
$$\times \left[\delta_{j_2 j_2'} \delta_{ll'} \delta_{\bar{l}\bar{l}'} \delta_{LL'} \delta_{\bar{L}\bar{L}'} - S^J(E_{T_i})_{v_i j_i (j_2 l) L; v_i j_i (j_2' l') L'} S^{\bar{J}*}(E_{T_f})_{v_f j_f (j_2 \bar{l}) L; v_f j_f (j_2' \bar{l}') \bar{L}'} \right]$$

using the BSD code[6]. Here, q is the tensor rank of the spectral transition operator (q=0 for isotropic and q=2 for anisotropic Raman transitions), that drives the transition from the v_i=0, j_i=1 to the v_f=1, j_f=1 state in H_2. j_2 and j_2' denote the pre- and post-collisional rotational quantum number of the perturbing molecule, respectively, l is the end-over-end angular momentum of the colliding pair, and J is the total angular momentum. $\begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$ is the 3-j symbol, $\begin{Bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{Bmatrix}$ denotes the 6-j symbol, $\begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}$ is the 12-j symbols of

the second kind, and $[x_1, x_2, \ldots, x_n] = (2x_1 + 1)(2x_2 + 1)\ldots(2x_n + 1)$. The scattering S-matrix are calculated at different total energies, $E_{T_i} = E_{kin} + E_{v_i j_i} + E_{j_2}$, and $E_{T_f} = E_{kin} + E_{v_f j_f} + E_{j_2'}$, but at the same kinetic energy, E_{kin}. λ is the rank of the velocity tensor. For λ= 0, the real and imaginary parts of σ_λ^q correspond to the standard pressure broadening (PBXS) and shift (PSXS) cross sections, respectively. For λ= 1, the generalized spectroscopic cross section corresponds to the complex Dicke cross section associated with the motional narrowing. The angular momenta are coupled according to the scheme used by Ben-Reuven[7] (the same scheme applies to the primed, post collisional quantum numbers):

$$\vec{\mathbf{j_2}} + \vec{\mathbf{l}} = \vec{\mathbf{L}},\ \vec{\mathbf{j_i}} + \vec{\mathbf{L}} = \vec{\mathbf{J}};\ \vec{\mathbf{j_2}} + \vec{\vec{\mathbf{l}}} = \vec{\vec{\mathbf{L}}},\ \vec{\mathbf{j_f}} + \vec{\vec{\mathbf{L}}} = \vec{\vec{\mathbf{J}}}$$

Note that this differs from the scattering scheme implemented in MOLSCAT. For details regarding the transformation of the S-matrices between these two coupling schemes see Ref. 8.

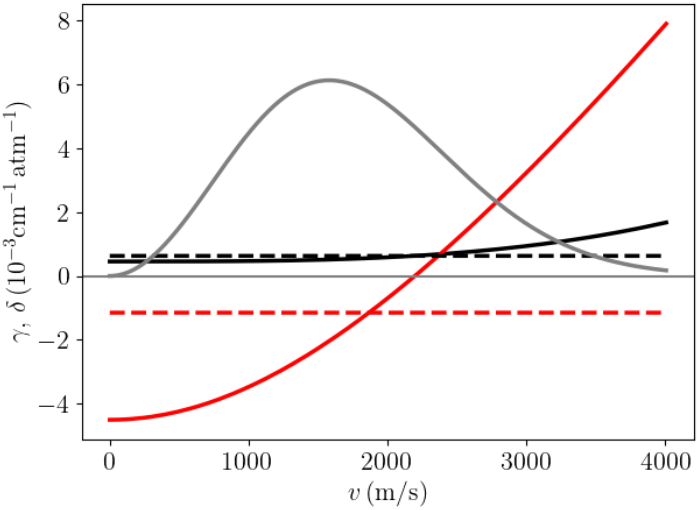


Supplementary Figure 3. Pressure broadening (solid lines) and pressure shift (dashed lines) cross sections of the self-perturbed Q(1) 1-0 line in H_2, for three chosen (out of six used in the calculations, j_2 = 0 to 5) rotational quantum numbers of the perturbing molecule.

Supplementary Figure 3 presents the calculated values of the PBXS and PSXS for several chosen values of the rotational quantum number of the perturber. The dependence of the calculated PSXS on j_2 is weak and becomes important only for E_{kin} < 50 cm^{-1}. The PBXS, on the other hand, exhibits a more complex dependence on j_2. For j_2 < 3, the PBXS is dominated by the dephasing contribution, until the relative kinetic energy is large enough to allow for the v_1=1, j_1 = 1, $j_2 \rightarrow v_1$=1, j_1=3, j_2' excitation. This corresponds to E_{kin} ≈ 558 cm^{-1} for j_2=j_2'=0 and j_2=j_2'=1, and to E_{kin} ≈ 203 cm^{-1} for j_2=2, j_2'=0 processes. As the kinetic energy surpasses these values, the inelastic events start contributing to the broadening (note that a *para*-H_2 molecule in j=0 state or an *ortho*-H_2 molecule in j=1 state cannot relax to a lower rotational state). The dephasing contribution is associated with elastic scattering, and its dependence on j_2 is negligible. For $j_2 \geq 3$, the inelastic contribution is non-zero in the whole range of kinetic energies. This is because even at the smallest kinetic energies considered, the j_1 = 1, $j_2 \rightarrow j_1$=3, j_2' , j_2' < j_2 deexcitation of the perturber is possible. The inelastic contribution to the PBXS is simply a half-sum of the inelastic state-to-state cross sections from the initial and final spectroscopic states:

$$\sigma^{\text{inel}} = \frac{1}{2}\left(\sum_{j_{i'} \neq j_i, j_2, j_{2'}} \sigma(v_i = 0, j_i = 1, j_2 \rightarrow v_{i'} = 0, j_{i'}, j_{2'}) + \sum_{j_{f'} \neq j_f, j_2, j_{2'}} \sigma(v_f = 1, j_f = 1, j_2 \rightarrow v_{f'} = 1, j_{f'}, j_{2'}) \right)$$

Its dependence on j_2 is more significant than the j_2-dependence of the dephasing contribution. The inelastic contribution to the broadening is the largest for j_2=3. As j_2 increases, the rotational spacing between the levels in H_2 becomes larger, and the inelastic contribution to the PBXS decreases.



Supplementary Figure 4. *Ab initio* speed dependences of collisional broadening, γ(solid black curve), and shift, δ (solid red curve), for the self-perturbed Q(1) Raman line in H_2. The speed-averaged values of the pressure broadening, γ_0, and shift, δ_0, coefficients are presented as black and red dashed curves, respectively. As a reference, we put the Maxwellian speed distribution at 303 K (grey curve).

The speed-averaged pressure broadening and shift coefficients, as well as the complex Dicke parameter are obtained by averaging the generalized spectroscopic cross sections over the Maxwell-Boltzmann distribution of relative kinetic energies (see Supplementary Figure 4) and by summing the resulting values over the relative populations of H_2 (the summation is truncated at $j_{2\,\max}$ = 5, which covers 99.5% of the total population of the H_2 molecule at the experimental temperature)

$$\gamma_0 + i\delta_0 = \frac{1}{2\pi c} n\langle v_r\rangle \sum_{j_2} p_{j_2}(T) \int \mathrm{d}x e^{-x} \sigma_0^q(v_\mathrm{i}, j_\mathrm{i}, v_\mathrm{f}, j_\mathrm{f}, j_2; E_\mathrm{kin} = x k_\mathrm{B} T)$$

$$\nu_\mathrm{opt} = \frac{1}{2\pi c} n\langle v_r\rangle M_2 \sum_{j_2} p_{j_2}(T)$$
$$\times \int \mathrm{d}x x e^{-x} \left(\frac{2}{3} x \sigma_1^q(v_\mathrm{i}, j_\mathrm{i}, v_\mathrm{f}, j_\mathrm{f}, j_2; E_\mathrm{kin} = x k_\mathrm{B} T) - \sigma_0^q(v_\mathrm{i}, j_\mathrm{i}, v_\mathrm{f}, j_\mathrm{f}, j_2; E_\mathrm{kin} = x k_\mathrm{B} T)\right)$$

Here, $\langle v_\mathrm{r}\rangle = \sqrt{8 k_\mathrm{B} T/\pi\mu}$ is the mean relative speed of the colliding pair at given temperature, k_B is the Boltzmann constant, μ is the reduced mass of the H_2-H_2 system and $M_2 = m_2/(m_1 + m_2)$. For the self-perturbed case, considered here, the mass of the perturber (m_2) and active molecule (m_1) is the same, hence, M_2 = 1/2. $p_{j_2}(T)$ is the population of the j_2 level at given temperature

$$p_{j_2}(T) = w_{j_2}(2j_2 + 1) e^{-E_{j_2}/k_\mathrm{B}T}/Z(T)$$

and *Z(T)* is the partition function

$$Z(T) = \sum_{j_2} w_{j_2}(2j_2 + 1) e^{-E_{j_2}/k_\mathrm{B}T}/Z(T)$$

The w_{j_2} factor is associated with nuclear spin statistics and equals 1 for even j_2 values (*para*-H_2) and 3 for odd j_2 values (*ortho*-H_2). E_{j_2} is the energy of the rovibrational state of the perturbing molecule.

From the kinetic energy dependent generalized spectroscopic cross sections, one may easily deduce their dependence on the speed of the active molecule at a given temperature[9–11]. This mathematical transformation is not repeated here. The speed dependence of collisional broadening and shift is particularly pronounced in hydrogen lines[12], see Supplementary Figure 4. The HTP model takes this effect into account by approximating the full speed-dependence of $\gamma$and δ with a quadratic function:

$$\gamma(v) + i\delta(v) = \gamma_0 + i\delta_0 + (\gamma_2 + i\delta_2)(v^2/v_{\mathrm{m}}^2 - 3/2)$$

where v_m is the most probable speed of the active molecule, and γ_2 and δ_2 are calculated by demanding that the slope of the quadratic function equals the slope of the actual speed dependence at $v = v_{\mathrm{m}}$. Explicit formulas for γ_2 and δ_2 can be found in Ref. 13.

Supplementary Note 4: Multi-pressure global fitting.

The first step of data analysis was to model the experimental spectra with the β-corrected Hartmann-Tran profile (βHTP), which is a spectral line profile optimized for H_2 isotopologues[14]. It accounts for all main collisional phenomena at play, including velocity-changing collisions and speed-dependent effects, while exhibiting an analytical representation that can be efficiently integrated into least-square fitting procedures. The βHTP is parametrized by 10 variables: the unperturbed (zero-pressure) transition frequency ν_0, the pressure-shift (γ_0) and pressure-broadening (δ_0) coefficients averaged over all molecular speeds, their quadratic corrections δ_2 and γ_2 that describe the speed dependence of the relaxation rates, the real and imaginary parts of the frequency of the velocity-changing collisions ν_{VC} and $\mathrm{Im}\{\nu_{\mathrm{VC}}\}$, which describe the Dicke narrowing effect, the sample temperature T and pressure p, and the correlation parameter η. To reduce the correlations between these parameters[12], we used measured values for p and T and performed a global fitting of multi-pressure datasets fixing a proper subset of parameters to the *ab initio* values extracted from the above discussed quantum-scattering calculations for the Q(1) 1-0 line. In a first fitting procedure we included in this subset only δ_2 and γ_2, which can be accurately predicted by theory and are known to be strongly correlated with the other parameters[12]. The results of the fitting are reported in Supplementary Table 1: except for δ_0 and $\mathrm{Im}\{\nu_{\mathrm{VC}}\}$, the fitted parameters show a very good agreement with the reference values from quantum-scattering calculations. This is a first important indication of the consistency between experimental spectra and the βHTP model seeded by *ab initio* collisional calculations. A second relevant indication in this direction is the very small penalty obtained in the fitting χ^2 value, by about 10%, upon including in the subset also the reference values for γ_0, ν_{VC} and $\mathrm{Im}\{\nu_{\mathrm{VC}}\}$, i.e. all remaining collisional parameters except the linear pressure shift δ_0, which is a critical parameter to be predicted. This is highlighted in Figure 2c and d of the main text by the residuals of these two types of fitting, in both cases exhibiting a structured profile at a level of 1% of the line maximum. For the parameter that is of main interest here, namely the line centre ν_0, the two types of fit produce instead largely different values, almost 2.5 MHz far apart. This is due to the inclusion of high-pressure spectra in the fit. While being of key importance for the retrieval of robust collisional parameters, they introduce a strong leveraging effect that amplifies the impact on the line centre of any imperfect assessment of collisional parameters. This effect is consistent with the already reported difficulty of βHTP to fully model the collisional physics of H_2 in such a large pressure range[12], as it emerges from the structured residuals.

Supplementary Table 1. Comparison between *ab initio* calculated collisional parameters (*Ab initio* column) and fitted collisional parameters. Column Fit 1 corresponds to a fit with only γ_2 and δ_2 fixed to reference (*ab initio*) values, while Fit 2 corresponds to a fit with all collisional parameters except δ_0 fixed to reference values. Units are 10^{-4} cm^{-1} bar^{-1} for γ_0 and γ_2, 10^{-3} cm^{-1} bar^{-1} for δ_0 and δ_2, 10^{-2} cm^{-1} bar^{-1} for ν_{VC}, 10^{-3} cm^{-1} bar^{-1} for $\mathrm{Im}\{\nu_{\mathrm{VC}}\}$ and MHz for $\nu_0 - \nu_{\mathrm{ref}}$. ν_{ref} is the theoretical value for the transition frequency. Uncertainties reported between parentheses are standard errors obtained from the covariance matrix for

fitted parameters, and from estimated contributions from the uncertainty of potential energy surfaces and computational parameters (basis set size, range of propagation) for ab initio values.

Parameter	*Ab initio*	Fit 1	Fit 2
$\boldsymbol{\gamma_0}$	6.29(63)	6.461(7)	*ab initio*
$\boldsymbol{\delta_0}$	-1.15(12)	-2.6892(6)	-2.6963(6)
$\boldsymbol{\gamma_2}$	1.05(10)	*ab initio*	*ab initio*
$\boldsymbol{\delta_2}$	2.27(22)	*ab initio*	*ab initio*
$\boldsymbol{\nu_{\mathrm{VC}}}$	4.25(21)	4.214(2)	*ab initio*
$\mathbf{Im}\{\boldsymbol{\nu_{\mathrm{VC}}}\}$	-1.46(146)	-3.05(3)	*ab initio*
$\boldsymbol{\nu_0 - \nu_{\mathrm{ref}}}$	0	1.77(7)	4.17(4)
$\boldsymbol{\chi^2_{\mathrm{r}}}$	/	1.2	1.3

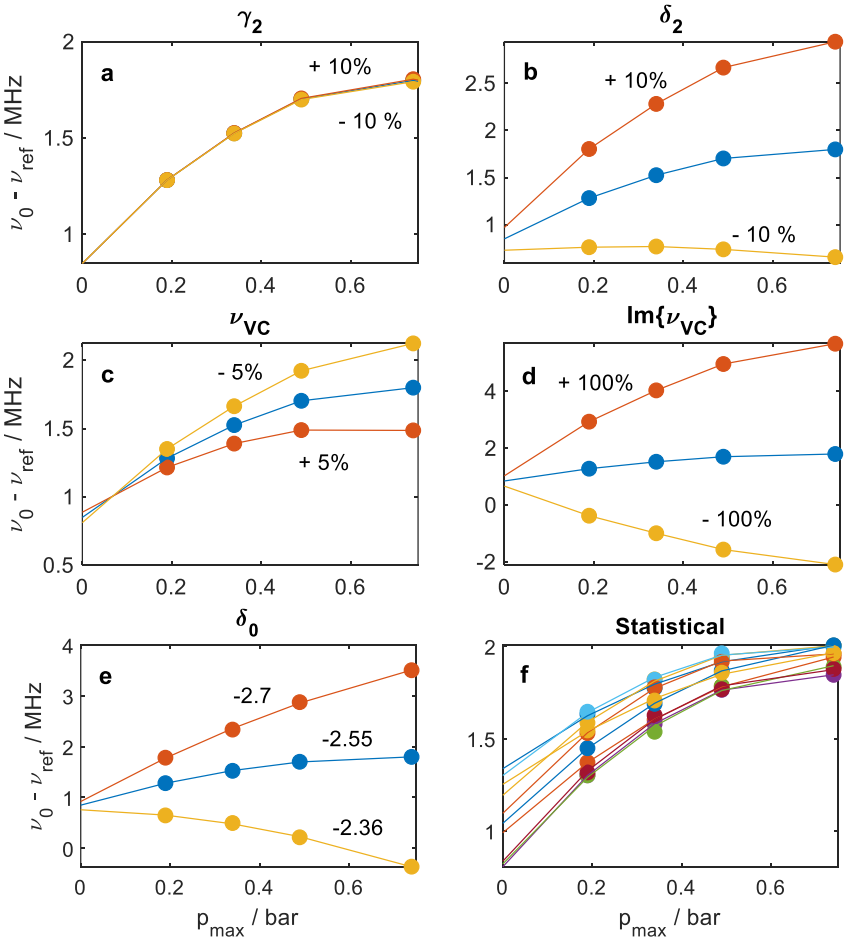
Supplementary Note 5: Uncertainty analysis

The extrapolated value of the transition frequency ν_0 to p_{max} = 0 critically depends on the fixed collisional parameters used in the fit function. These parameters, as obtained by the *ab initio* quantum scattering calculations described in Supplementary Note 3, suffer from uncertainties that can be propagated to the extrapolated transition frequency as follows:

$$\sigma_{\mathrm{sys}} = \sqrt{\sigma_{\gamma_2}^2 + \sigma_{\delta_2}^2 + \sigma_{\nu_{\mathrm{VC}}}^2 + \sigma_{\mathrm{Im}\{\nu_{\mathrm{VC}}\}}^2 + \sigma_{\delta_0}^2}$$

where σ_{sys} is the overall systematic uncertainty on the extrapolated ν_0 and σ_x^2 is the contribution to this uncertainty from the collisional parameter $x = \{\gamma_2, \delta_2, \nu_{\mathrm{VC}}, Im\{\nu_{\mathrm{VC}}\}, \delta_0\}$. We exclude from this budget γ_0 as its contribution is negligible.

To propagate the uncertainties on any single collisional parameter x to the extrapolated transition frequency, we repeated the extrapolation procedure explained in the main text three times for each collisional parameter considered: the first time with all parameters fixed to the corresponding reference values derived from *ab initio* calculations (or from fitting, for δ_0 only); the second (third) time with the value of x increased (decreased) by the estimated uncertainty on that parameter.



Supplementary Figure 5. Difference between fitted transition frequency ν_0 and theoretical determination ν_{ref} (from Komasa et al.[15]) as a function of the maximum pressure p_{max} considered in the global fitting. Points correspond to transition frequencies determined from the fitting, whereas lines represent their quadratic extrapolation to p_{max} = 0. **a-d,** blue dots and lines correspond to fits performed as per "Fit 2" column of Supplementary Table 1; red and yellow dots and lines correspond to fits with one collisional parameter increased or decreased, respectively, by the amount specified in the corresponding label; the panel title indicates the parameter whose value is changed. **e,** same as **a-d**, except that the actual values of δ_0 used in the fitting have been reported in the figure. **f,** each colour identifies a fit over a different spectral dataset; all fits have been performed as per "Fit 2" column of Supplementary Table 1.

The results of this procedure are presented in Supplementary Figure 5a-e: blue points correspond to transition frequencies obtained for different values of p_{max} when all collisional parameters but δ_0 fixed to the corresponding *ab initio* value (as in the "Fit 2" column of Supplementary Table 1), while the blue line represents a quadratic extrapolation of the transition frequency to p_{max} = 0. Red and yellow points represent transition frequencies obtained with a single collisional parameter increased or decreased, respectively, by an amount equal to its theoretical uncertainty (10% for γ_2 and δ_2, 5% for ν_{VC}, 100% for $\mathrm{Im}\{\nu_{\mathrm{VC}}\}$). Yellow and red lines correspond to quadratic extrapolations to p_{max} = 0. The systematic contribution to ν_0 for any x parameter is taken as half the maximum deviation of the extrapolated transition frequencies (i.e. half the distance between the red and yellow lines at p_{max} = 0).

For what concerns the pressure shift parameter δ_0, the *ab initio* value was too inaccurate to be fixed in the fit. If left free in the multi-pressure fits, we observed its fitted value to change in the interval from $-2.36 \cdot 10^{-2}$ to $-2.7 \cdot 10^{-2}\mathrm{cm}^{-1}\,\mathrm{bar}^{-1}$ by modifying p_{max} from its minimum (0.2 bar) to its maximum (4 bar): we took this as the confidence interval for δ_0 (corresponding to yellow and red points and lines in Supplementary Figure 5e), thereby fixing its most probable value to $-2.55 \cdot 10^{-2}\mathrm{cm}^{-1}\mathrm{bar}^{-1}$ (blue points and line in

Supplementary Figure 5e). The above interval did not change appreciably upon modifying the type and the number of the other collisional parameters taken fixed in the fit.

The overall uncertainty on ν_0 is obtained by quadrature sum of the systematic and statistical contributions:

$$\sigma_{\nu_0} = \sqrt{\sigma_{\mathrm{sys}}^2 + \sigma_{\mathrm{stat}}^2}.$$

The latter is obtained by repeating the extrapolation procedure to determine ν_0 over many spectral datasets that were purposedly assembled by mixing single spectral acquisitions obtained in different days and months. Supplementary Figure 5f represents the result of such a procedure: each colour corresponds to a different dataset, with points representing the fitted values of the transition frequency as a function of p_{max} and lines corresponding to the quadratic extrapolation to zero pressure. The statistical uncertainty is evaluated as the standard deviation of the extrapolated transition frequencies obtained over all mixed datasets.

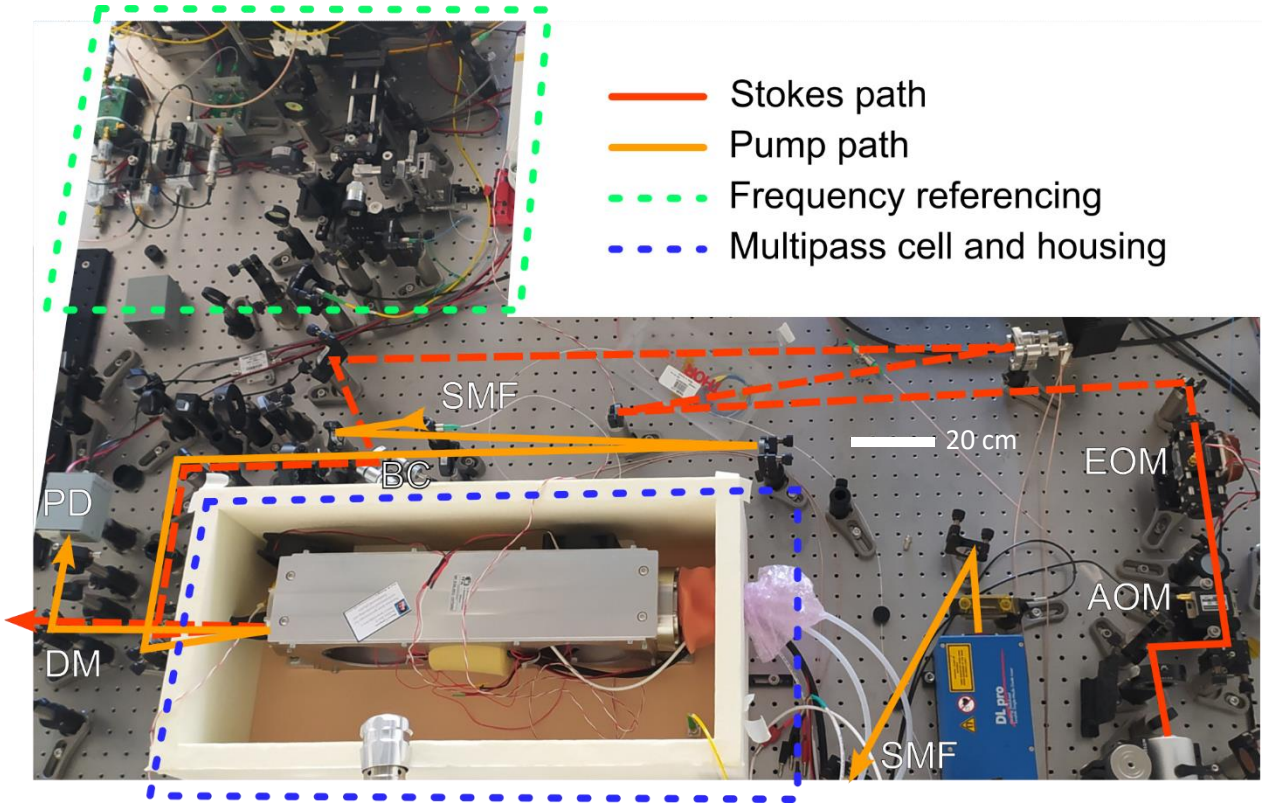
The uncertainty budget is given in Supplementary Table 2, with a systematic contribution (234 kHz) comparable to the statistical one (200 kHz). Their quadrature sum returns a total uncertainty of 310 kHz. It is worth noting that other systematic sources of uncertainty, such as the calibration of the frequency axis and of the pressure/temperature gauges, have been excluded from the budget as their impact is negligible as compared to the lineshape contribution.

Supplementary Table 2: **Error budget**. The total uncertainty given by the sum in quadrature of the two systematic errors (234 kHz) and the statistical error (200 kHz). All values are in kHz.

Type B	Systematic error (*ab initio* parameters)	220
	Systematic error (δ_0)	81
	Total line-shape error (σ_{sys})	234
Type A	Statistical error (σ_{stat})	202
A+B	Total uncertainty (σ_{ν_0})	310

Supplementary Note 6: Implementation of the spectrometer

The implementation of the spectrometer is shown in Supplementary Figure 6, where a photograph of the system is provided, together with annotations to identify the paths of laser beams. Stokes and pump laser paths are shown in red and orange, respectively. The Stokes beam passes through an acousto-optical modulator (AOM) for active power stabilization and a resonant electro-optical amplitude modulator (EOM), then through a reflective telescope for mode matching to the multipass cell. It is overlapped to the pump beam at a dichroic beam combiner (BC), sent to the multipass cell, and separated from the pump beam at a dichroic mirror (DM) after exiting the cell. The pump beam is first sent through a single mode fibre (SMF) for mode cleaning, then through a reflective telescope for mode matching, and combined with the Stokes beam before entering the cell. After exiting the cell, it is reflected by the dichroic mirror and sent to the detection photodiode (PD). Frequency referencing of the two lasers is performed in the region identified by the dashed green box, where the layout shown in Supplementary Figure 1 is implemented. The multipass cell is housed in a thermally insulated box, identified by the blue dashed box. White tubes exiting the cell housing connect the multipass cell with the pressure sensor and two flow controllers used to maintain a constant gas flow and to actively stabilize pressure.



Supplementary Figure 6. Optical setup of the spectrometer. AOM. acousto-optical modulator; EOM: electro-optical amplitude modulator; BC: beam combiner; DM: dichroic mirror; SMF: single mode optical fibre; PD: photodiode.

Supplementary References

1. Garberoglio, G., Jankowski, P., Szalewicz, K. & Harvey, A. H. Second virial coefficients of H_2 and its isotopologues from a six-dimensional potential. *J. Chem. Phys.* **137**, 154308 (2012).
2. Hutson, J. M. & Le Sueur, C. R. molscat: A program for non-reactive quantum scattering calculations on atomic and molecular collisions. *Comput. Phys. Commun.* **241**, 9–18 (2019).
3. Monchick, L. & Hunter, L. W. Diatomic–diatomic molecular collision integrals for pressure broadening and Dicke narrowing: A generalization of Hess's theory. *J. Chem. Phys.* **85**, 713–718 (1986).
4. Schaefer, J. & Monchick, L. Line broadening of HD immersed in He and H_2 gas. *Astron. Astrophys.* **265**, 859–868 (1992).
5. Thibault, F. *et al.* Rovibrational line-shape parameters for H_2 in He and new H_2-He potential energy surface. *J. Quant. Spectrosc. Radiat. Transf.* **202**, 308–320 (2017).
6. Thibault, F. Code for pressure broadening, shift and complex Dicke cross sections for 2 diatomics. https://data.mendeley.com/datasets/tdvgvwr2t7/2 (2021) doi:10.17632/tdvgvwr2t7.2.
7. Ben-Reuven, A. Impact Broadening of Microwave Spectra. *Phys. Rev.* **145**, 7–22 (1966).
8. Green, S. Rotational excitation in collisions between two rigid rotors: Alternate angular momentum coupling and pressure broadening of HCl by H_2. *Chem. Phys. Lett.* **47**, 119–122 (1977).
9. Berman, P. R. Speed-dependent collisional width and shift parameters in spectral profiles. *J. Quant. Spectrosc. Radiat. Transf.* **12**, 1331–1342 (1972).
10. Pickett, H. M. Effects of velocity averaging on the shapes of absorption lines. *J. Chem. Phys.* **73**, 6090–6094 (1980).
11. Stankiewicz, K., Stolarczyk, N., Jóźwiak, H., Thibault, F. & Wcisło, P. Accurate calculations of beyond-Voigt line-shape parameters from first principles for the He-perturbed HD rovibrational lines: A comprehensive dataset in the HITRAN DPL format. *J. Quant. Spectrosc. Radiat. Transf.* **276**, 107911 (2021).

12. Wcisło, P. *et al.* Accurate deuterium spectroscopy for fundamental studies. *J. Quant. Spectrosc. Radiat. Transf.* **213**, 41–51 (2018).
13. Wcisło, P. *et al.* The first comprehensive dataset of beyond-Voigt line-shape parameters from ab initio quantum scattering calculations for the HITRAN database: He-perturbed H_2 case study. *J. Quant. Spectrosc. Radiat. Transf.* **260**, 107477 (2021).
14. Konefał, M. *et al.* Analytical-function correction to the Hartmann–Tran profile for more reliable representation of the Dicke-narrowed molecular spectra. *J. Quant. Spectrosc. Radiat. Transf.* **242**, (2020).
15. Komasa, J., Puchalski, M., Czachorowski, P., Łach, G. & Pachucki, K. Rovibrational energy levels of the hydrogen molecule through nonadiabatic perturbation theory. *Phys. Rev. A* **100**, (2019).