An ultrafast algorithm for ultrafast time-resolved coherent Raman spectroscopy

Mazza, Francesco; van den Bekerom, Dirk

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

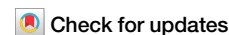# An ultrafast algorithm for ultrafast time-resolved coherent Raman spectroscopy

Check for updates

Francesco Mazza [1] & Dirk van den Bekerom [2] ✉

Time-resolved coherent Raman spectroscopy (CRS) is a powerful non-linear optical technique for quantitative, in-situ analysis of chemically reacting flows, offering unparalleled accuracy and exceptional spatiotemporal resolution. Its application to large polyatomic molecules, crucial for understanding reaction dynamics, has thus far been limited by the complexity of their rotational-vibrational Raman spectra. Progress in developing comprehensive spectral codes for these molecules, a longstanding goal, has been hindered by prohibitively long computation times required for their spectral synthesis. Here, we present an algorithm that achieves a million-fold improvement in computation time compared to existing methods. The algorithm demonstrates remarkable accuracy, with an approximation error below 0.1% across all tested probe delays, at both room temperature (296 K) and elevated temperatures (1500 K). This result could greatly expand the application of time-resolved CRS, particularly in plasma research, as well as in broader atmospheric and astrophysical sciences.

The advancement of laser technology over the last six decades has thrusted the development of optical diagnostics able to measure chemo-physical phenomena on molecular scales, moving from the rotational and vibrational nuclear motions on the pico- (ps) and femtosecond (fs) timescale[1], to the motion of electrons on the attosecond one[2]. The increasingly shorter duration and correspondingly higher peak power of pulsed lasers has furthermore enabled the development of time-resolved nonlinear optical spectroscopy[3]. Coherent Raman spectroscopy (CRS) in particular, has benefitted from the application of fs laser pulses to impulsively generate coherent rotational and vibrational wave packets, maximizing the quantum coherence and thus the signal yield[4].

CRS holds enormous potential to study chemical reactions in gas-phase media, and it is presently regarded as the gold standard for quantitative, in situ measurements in chemically reacting flows[5,6]. Bohlin and Kliewer, for example, have demonstrated hyperspectral imaging thermometry and wideband chemical detection in combustion environments[7,8]. Recently, CRS spectroscopy has also been employed to perform state-selective measurements of the collisional dynamics and energy transfer of molecular superrotors[9,10], and to detect the angular momentum transfer between roton pairs in superfluid helium[11].

Quantitative CRS measurements have nonetheless been limited to a notably small set of well-characterized chemical species—typically diatomic molecules, e.g., N_2, H_2, and CO, whose rotational and vibrational motions are well described by simple quantum mechanical models[12]. Numerical CRS codes for the vibrational spectra of a handful of triatomic molecules, such as

CO_2, N_2O, and H_2O[13,14], are also available. The extension of these codes to simulate the spectra of larger polyatomic molecules is complicated by the substantial number of vibrational modes to consider, as well as by the possibility of mode coupling (e.g., Coriolis coupling[15]). Theoretical models have been developed to perform vibrational spectroscopy on some highly symmetric polyatomics like CH_4[16], SF_6[17], and H_2CO[18], while pure-rotational spectroscopy has been applied to C_2H_6[19] and C_2H_4[20].

The application of CRS to larger polyatomic molecules requires computer codes that can handle their vast vibrational complexity. In order to harness the full diagnostic potential of CRS one would thus need: (i) spectroscopic databases, providing the Raman frequencies and transition polarizabilities of the target molecules, and (ii) a way of rapidly synthesizing the CRS spectra from these databases.

On the one hand, recent advancements in computational chemistry have prompted the creation of large spectral databases for molecular species relevant to atmospheric, planetary, and astrophysical studies[21,22]. Quantum chemistry methods furthermore allow for the vibrational mode assignment and the calculation of the polarizability of significantly more complex molecules, e.g. organosilicon compounds[23].

On the other hand, the synthesis of the CRS spectra from large spectral databases represents an unaddressed bottleneck. A recent study by ref. 24 provides an illustrative example: the authors employed the MeCaSDa spectral database[25] (accessible at http://vamdc.icb.cnrs.fr/PHP/methane.php) to simulate the symmetric C–H stretch spectrum of CH_4, they had to filter the ~1.7 million (M) spectral line list, reducing the number of considered

[1]Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, Delft, 2629HS, The Netherlands. [2]Energy & Materials Transition, Netherlands Organization for Applied Scientific Research (TNO), Urmonderbaan 22, Geleen, 6167RD, The Netherlands. ✉e-mail: dirk.vandenbekerom@tno.nl

transitions to ~6000 at elevated temperature. They furthermore employed a high-performance computing cluster to calculate the resulting spectra[24]. Mazza et al. used the same database to compute the H–C–H bend spectrum of CH_4: when employing the whole line list (~11 M lines), the computation of a single spectrum required more than 8 h on a commercial laptop[26]. This being state-of-the-art, the widespread use of CRS for quantitative spectroscopy on large polyatomics, especially in hot environments where one would possibly need to consider many billions (B) of spectral lines[21], is demonstrably hindered by the lack of fast spectral algorithms.

In the present work, we introduce an algorithm for the rapid synthesis of time-resolved CRS spectra, demonstrating a reduction of the computation time by a factor of 1 M. We refer to this algorithm as "ultrafast", in view of the parallel between this result and the shift from nanosecond to femtosecond lasers, which enabled ultrafast time-resolved spectroscopy[1]. The algorithm is based on the discrete integral transform developed by van den Bekerom and Pannier[27], who demonstrated a 2–3 orders-of-magnitude reduction in the computational cost of simulating the absorption spectrum of CO_2, as compared to the legacy spectral synthesis code in RADIS[28]. Here, the accuracy and computational performance of the ultrafast algorithm are assessed against the numerical code of Mazza et al., which models the spectrum of CH_4 in the dyad region of its Raman spectrum[26]. We furthermore discuss the potential implications of these results for the application of time-resolved CRS to investigate the gas-phase chemistry of large polyatomic molecules.

Results
Integral transform formulation
The time-domain description[29,30] of CRS is based on the sudden approximation[31], whereby the CRS signal arises from the interaction of a probe pulse with a molecular wave packet, impulsively created by the fs pump pulse[32]. The creation of such a wave packet need not be described in detail, but rather a phenomenological model for the nonlinear optical response of the medium can be used[3]. According to this model, the intensity of the CRS polarization field, as a function of its angular frequency ω, is given as:

$$I_{CRS}(\omega) = \left| \widehat{P}_{CRS}(\omega) \right|^2, \quad (1)$$

with $\hat{P}_{CRS}(\omega) = \sum_i \hat{P}_i(\omega)$ the total polarization, and the polarization for a single Raman line $\hat{P}_i(\omega)$ given by:

$$\widehat{P}_i(\omega) = \mathcal{F}\left\{ \chi_i^{(3)}(t_1) E_{pr}(t_1 - \tau) \right\}, \quad (2)$$

where $\chi_i^{(3)}(t) = W_i \exp(j\omega_i - \Gamma_i)t$ is the third-order nonlinear optical susceptibility for an individual line with ω_i the transition's line position and Γ_i being the corresponding damping coefficient, E_{pr} is the probe laser field, and t_1 is the coherence time relative to the pump/Stokes laser.

The Fourier transform (FT) can be centered around the center of the probe intensity profile at τ by replacing $t_1 \rightarrow t + \tau$, with t the time relative to the center of the probe pulse. Equation 2 then becomes:

$$\widehat{P}_i(\omega) = \exp(-j\omega\tau)\chi_i^{(3)}(\tau)\mathcal{F}\left\{ \exp\left[j(\omega_i + j\Gamma_i)t\right]E_{pr}(t) \right\}. \quad (3)$$

Here the complex phase common for all lines, $\exp(-j\omega\tau)$, may be omitted from Eq. 3 as it will disappear after squaring in Eq. 1. Furthermore, using the frequency shift property of the FT and applying a forward shift of $\omega_i + j\Gamma_i$, the total polarization can be written as:

$$\widehat{P}_{CRS}(\omega) = \sum_i \chi_i^{(3)}(\tau)\widehat{E}_{pr}(\omega - \omega_i - j\Gamma_i), \quad (4)$$

with $\widehat{E}_{pr}(\omega) = \mathcal{F}\{E_{pr}(t)\}$ the FT of the probe field. Equation 4 shows that the total polarization may be expressed directly as the sum of lineshapes $\widehat{E}_{pr}(\omega)$ that are scaled by a (complex) factor $\chi_i^{(3)}(\tau)$ and are "shifted" by $\omega_i + j\Gamma_i$, where convergence of the integral transform is guaranteed as $\Gamma_i > 0$ for all lines. In this expression only the term $\chi_i^{(3)}(\tau)$ is a function of the pulse-probe delay τ, and as such it is fully responsible for the well-known effect of coherence beating[33].

Since the total polarization is thus essentially a sum of lineshapes, $\widehat{P}_{CRS}(\omega)$ can be expressed as an integral kernel over the space of lineshape parameters ω and Γ by following the approach of van den Bekerom and Pannier[27]:

$$\widehat{P}_{CRS}(\omega) = \iint \widehat{X}^{(3)}(\omega', \Gamma'; \tau)\widehat{E}_{pr}(\omega - \omega' - j\Gamma')d\omega'd\Gamma'. \quad (5)$$

Here $\hat{X}^{(3)}$ (and $\hat{W}$) is the distribution function of the third-order nonlinear susceptibility (and line strength, respectively), that contains all spectral information:

$$\widehat{X}^{(3)}(\omega, \Gamma; \tau) = \widehat{W}(\omega, \Gamma)\exp[(j\omega - \Gamma)\tau], \quad (6a)$$

$$\widehat{W}(\omega, \Gamma) = \sum_i W_i\delta(\omega - \omega_i)\delta(\Gamma - \Gamma_i). \quad (6b)$$
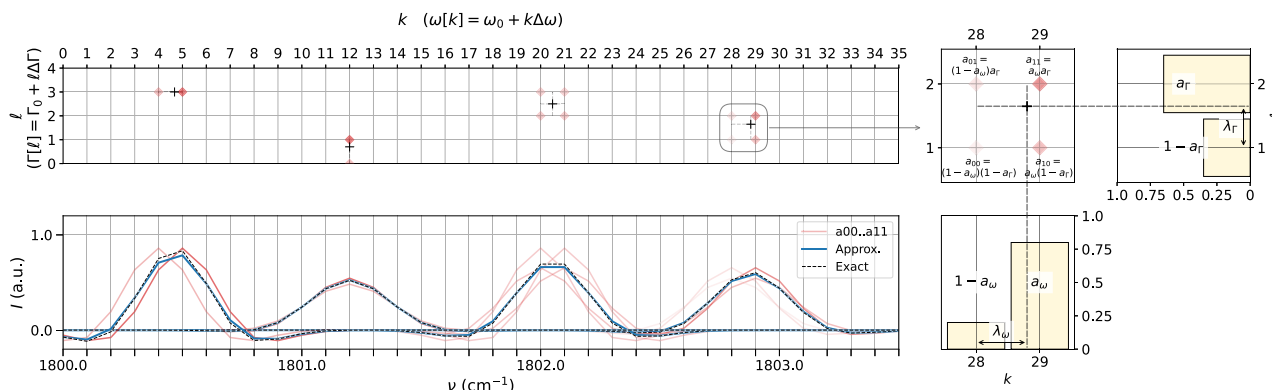

Fig. 1 | Approximation scheme in the frequency-linewidth space. (Top-left) Schematic representation of line strength distribution W[k, l]; crosses represent exact lines; red diamonds represent constituent lines with parameters on the grid. (Bottom-left) Lineshapes corresponding to the lines in top graph, demonstrating the constituent lineshapes in red, the approximated lineshape in blue, and the exact lineshape in dashed black. The opacity of the red lineshapes indicates the weight with which they contribute to the approximated lineshape. (Right) Detailed schematic of the last line, showing the weights of the individual position and damping contributions as bar plots. In this figure the grid spacing is here taken coarser than would be typical in order to highlight discrepancies.

Equation 5 expresses $\hat{P}_{CRS}(\omega)$ as a convolution over ω, which can equivalently be written as a product in the time domain. A compact expression for the time-domain polarization $P_{CRS}(t)$ is thus obtained:

$$P_{CRS}(t) = E_{pr}(t) \int \mathcal{F}^{-1}_{t \leftarrow \omega} \left\{ \hat{X}^{(3)}(\omega, \Gamma'; \tau) \right\} \exp(-\Gamma' t) d\Gamma', \; t \geq -\tau \quad (7)$$

Here the condition $t \geq -\tau$ enforces the physical constraint that no CRS polarization field is generated before the impulsive excitation of the Raman coherence: $P_{CRS}(t)$ is zero for $t < -\tau$. By transforming $P_{CRS}(t)$ back to the frequency domain and plugging into Eq. 1, the ultrafast computation of the CRS spectrum is achieved.

Equations 6a and 7 are discretized by replacing the continuous FT by a discrete FT (DFT) and replacing the integral over Γ by a summation, as summarized in Methods and explained in more detail in van den Bekerom and Pannier[27]. Figure 1 presents a graphical summary of the discretization method. In discretizing Eq. 6a, weights must be assigned to reproduce function values that are not on gridpoints. The choice of weights offers a degree of freedom that allows for the complete elimination of the approximation error at a single point on the t or ω axis. With the polarization signal strongly concentrated around the probe pulse center, elimination of the error at $t = 0$ s seems appropriate. Under this choice, the weights for position and damping for the ith spectral line, $a_{\omega,i}$ and $a_{\Gamma,i}$ respectively, are given by:

$$a_{x,i} = \frac{\exp(2\lambda_{x,i}\theta_{\Delta x}) - 1}{\exp(2\theta_{\Delta x}) - 1} = \frac{\sinh(\lambda_{x,i}\theta_{\Delta x})}{\sinh(\theta_{\Delta x})} \exp\left[(\lambda_{x,i} - 1)\theta_{\Delta x}\right] \quad (x = \omega, \Gamma),$$

(8)

with the dimensionless grid sizes $\theta_{\Delta\omega} = j\tau\Delta\omega/2$ and $\theta_{\Delta\Gamma} = -\tau\Delta\Gamma/2$. The factors $\lambda_{\omega,i} = (\omega_i - \omega[k_i])/\Delta\omega$ and $\lambda_{\Gamma,i} = (\Gamma_i - \Gamma[l_i])/\Delta\Gamma$, with values between 0 and 1, describe the relative position of the line with respect to the gridpoints. The step-by-step derivation of Eq. 8, as well as strategies for its efficient numerical evaluation, can be found in the Supplementary Information, section "Weight derivation and evaluation".

The discretization error introduced by the choice of weights in Eq. 8 can be derived analytically under relatively weak assumptions (see Supplementary Information, section "Approximation error—frequency-domain approximation"). The error component due to the interpolation of the spectral position, ω_i (or damping coefficient, Γ_i), is evaluated as a function of the following dimensionless variables: complex frequency, $z = (\omega - \omega_i - j\Gamma_i)\Delta t_{pr}$; probe pulse duration-to-delay ratio, $\theta_{\Delta t} = \Delta t_{pr}/\tau$; and the previously defined dimensionless grid size, $\theta_{\Delta\omega}$ (or $\theta_{\Delta\Gamma}$). It is possible to show (proof given in Supplementary Information, section "Approximation error—frequency-domain approximation") that the approximation error has an upper bound of:

$$\left| \frac{\hat{P}^{err}_{x,i}(\omega; \theta_{\Delta x}, \theta_{\Delta t})}{\hat{P}_i(\omega = \omega_i)} \right| \lesssim \frac{1}{2} \theta_{\Delta t} |\theta_{\Delta x}|^2 \left| \frac{d\hat{E}_{pr}(z)}{dz} \right| / \left| \hat{E}_{pr}(j \text{Im}(z)) \right|, \quad (9)$$

which is valid whenever $|\theta_{\Delta x}| \ll 1$ and $\theta_{\Delta t}|\theta_{\Delta x}| \ll 1$ (see Supplementary Information, section "Frequency-domain approximation"). With $\theta_{\Delta t}$ and the probe laser pulse shape and duration set by the experiment, the error in Eq. 9 only depends on the grid size chosen to discretize the position and damping space, which can be set arbitrarily small. For example, if the probe delay were increased from 5 to 100 ps, the number of spectral points would need to be increased by a factor of $\sqrt{20} \approx 4.5$ to keep the same accuracy. For a large number of lines (>1 M), this will not affect performance because the computation is bottlenecked by populating the distribution function (Eq.7), as shown by the linear increase in Fig. 2. The algorithm thus allows for arbitrary accuracy in the simulation of CRS spectra, where the only limitation is represented by the tradeoff between accuracy and computational performance. Nonetheless, the presented approach provides a dramatic performance improvement even at only a slight concession in accuracy.

Improvement in computational performance

Figure 2 illustrates the reduction in the computational cost of different numerical implementations of the algorithm detailed above. The performance of the conventional time-domain modeling based on Eq. 3 (named: *ref*) is compared to that of the ultrafast algorithm based on Eq. 7 (named: *ufa*). For both cases, a Python and C++ implementation (-py and -cpp, respectively) is tested; more details are available in the Methods section. The MeCaSDa spectral database[25] is rescaled according to the procedure of Mazza et al.[26] to calculate the weights W_i of the ~10^7 individual lines in the H–C–H bend (v_2 mode) CRS spectrum of CH_4. To the best of the knowledge of the authors, this represents the largest experimentally verified line list available for CRS spectroscopy. The computation time for an increasing number of spectral lines included in the spectrum is measured for the different numerical implementations. As detailed in Methods, a total of nine spectra at different conditions is computed at each benchmarking point in Fig. 2, with the markers in the figure representing the average of the nine computation times.

While implementing the conventional time-domain model in a compiled language (*ref-cpp*) makes the simulation of the CRS spectrum a couple of times faster than in an interpreted language such as Python (*ref-py*), this improvement is insignificant compared to the improvement gains of the ultrafast spectral algorithm. Even considering the slower, interpreted implementation of the fast algorithm (*ufa-py*), the computation time for a single spectrum employing the full database is abated from 4 h to 5 s, an improvement of more than three orders-of-magnitude. The fast algorithm implemented in C++ (*ufa-cpp*) includes multithreading, making use of all the cores, as well as full code vectorization through the implementation of single instruction multiple data (SIMD) intrinsics in places of code bottlenecks. This implementation reduces the computation time to 14 ms, representing an improvement by a factor of 1 M over the present state-of-the-art. We note that compared to *ufa-py*, *ufa-cpp* does not make any new approximations, and their outputs are nominally the same. The demonstrated improvement in computational performance enables the on-the-fly evaluation of experimental data for quantitative CRS (see Supplementary Information, section "On-the-fly fitting of experimental data"), which could lead to widespread commercialization of CRS spectrometers for gas-phase chemical analysis. The application code described there is included in the supplementary files as Supplementary Software 1.

We furthermore explore the possibility of employing even more extensive spectral databases, pushing above the tens of millions of Raman lines, by synthetically increasing the size of the spectral database (see Methods) as shown in Fig. 2. When a total of 100 M spectral lines is considered, the conventional modeling time for a single CRS spectrum is estimated to be more or less an entire day, depending on the language employed. The same computation, on the other hand, takes less than a single second by employing the ultrafast algorithm (97 ms). Our ultrafast spectral algorithm is, therefore, the only approach currently available to fully employ even larger line lists, possibly containing tens of billions of spectral lines (e.g. the YT34to10 line list[34] contains 35 B transitions for high-temperature CH_4), without the need for line reduction schemes. The estimated computation time required by conventional codes to simulate a single CRS spectrum employing 10 B lines is in the order of many months. In contrast, the *ufa-cpp* implementation would perform the same task in less than 10 s. Finally, we remark that already, for a small number of lines, the new algorithm beats the reference case. A typical spectrum for a diatomic molecule will contain between 100 and 1000 lines, such that the presented algorithm provides a factor 30-300 improvement: fast enough to perform on-the-fly fitting. The data underlying Fig. 2 is appended to the paper as supplementary data in Supplementary Data 1.

Accuracy of the ultrafast algorithm in the frequency domain

In addition to enabling the rapid computation of the CRS spectra, the integral transform approach is also highly accurate. The error introduced in the simulation of the CH_4 v_2 mode spectrum by the ultrafast algorithm is represented in Fig. 3. The error, evaluated for the *ufa-cpp* implementation
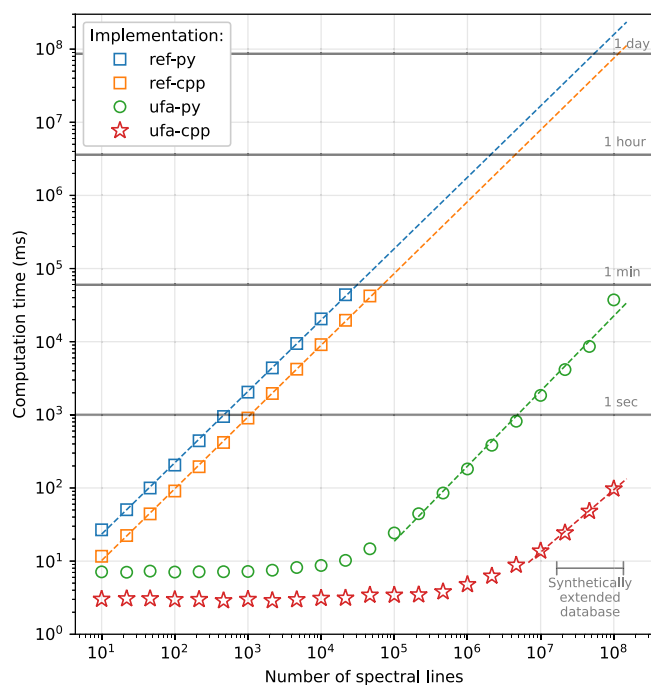
Fig. 2 | Computational performance of the ultrafast algorithm. Benchmark of different approaches/implementations for synthesizing time-resolved CRS spectra. *ref* refers to the reference approach based on Eq. 3, implemented as Python (*-py*) and C++ code (*-cpp*); *ufa* refers to the ultrafast approximation based on Eq. 7, also implemented as both Python (*-py*) and C++ (*-cpp*) codes. The fully optimized implementation (*ufa-cpp*) includes multithreading and full SIMD vectorization, which improves computation time by a factor of 1 M compared to the conventional approach (*ref-py*) for a large number of spectral lines. The performance in the 10^7–10^8 lines range is evaluated by synthetically expanding the number of lines in the MeCaSDa database (see Methods).

benchmarked in the section above, is relative to the *ref-py* implementation and is normalized to the maximum intensity of the spectrum. Figure 3a, b show a map of the error as a function of the frequency and probe pulse delay, simulating a pump-probe CRS experiment, for $T = 296$ K and $T = 1500$ K, respectively.

A clear beating pattern is observed for all spectral lines in both Fig. 3a, b, matching the coherence beating of the unresolved, Coriolis-split ro-vibrational transitions[26]. This shows that the dominating error components in the simulated CRS spectra arise from the coalescence of the approximation errors of the individual unresolved lines. The frequency-domain error is observed to be weakly dependent on the probe delay, with the error slightly increasing for larger probe delays. Nonetheless, for all delays between 20 and 200 ps, the error remains below 0.1%. In Fig. 3c, three sample spectra at probe delays of $\tau = 50$, 100, and 150 ps and a temperature 296 K are depicted, with their residuals plotted in Fig. 3e. In all cases, a discrepancy between the *ufa-cpp* (solid, colored) and *ref-py* (dashed, black) implementations is imperceptible at this scale. The magnitude of the approximation error for $\tau = 50$ and 100 ps is less than 0.05% for all the different spectral branches, while for $\tau = 150$ ps, the error in the Q-branch and the O(6) line at ~1655 cm^{-1} is increased to about 0.1% in magnitude. To put this result into perspective, such an approximation error is equivalent to the stochastic error in a hypothetical CRS experiment, where the signal-to-noise ratio (SNR) exceeded 30 dB, which is beyond the dynamic range of most detectors used in commercial spectrometers.

Figure 3b, d, f show the relative error for the frequency-domain spectra computed at $T = 1500$ K, approximately the reference temperature for the calculation of the line intensities in the MeCaSDa database[35]. At elevated temperatures, the higher vibrational states become significantly populated and contribute to the CRS spectrum. This is underscored by the stronger

coherence beating of all the spectral features in Fig. 3b, as well as by the larger number of P-branch lines appearing in Fig. 3d. Nevertheless, it is important to remark that the magnitude of the relative error is the same as in the room temperature simulation. The ultrafast algorithm thus allows for the inclusion of a large number of higher vibrational states in the spectral computation, with no additional detriment to its accuracy.

Accuracy of the ultrafast algorithm in the time domain

The possibility of performing time-resolved measurements, by delaying the probe pulse relative to the excitation pulse in pump-probe experiments, enables the use of CRS to study the decoherence of the molecular wave packet due to the interaction with its environment[9,10,36]. The synthetic modeling of these "probe delay scans" is even more computationally demanding than the frequency-domain spectral calculations described in the previous section, as Eqs. 1 and 3 need to be evaluated for all the values of τ in the delay series[24]. The integral transform formulation of the third-order susceptibility can also be employed here to derive an ultrafast algorithm for the time-domain synthesis of the CRS signal (details in Supplementary Information, section "Ultrafast probe delay scan"):

$$I_{\text{PDS}}(\tau) = \left| E_{\text{pr}}(-\tau) \right|^2 * \left| \chi^{(3)}(\tau) \right|^2, \tag{10a}$$

$$\chi^{(3)}(\tau) = \int \mathcal{F}^{-1}_{\tau \leftarrow \omega} \left\{ \widehat{W}(\omega, \Gamma') \right\} \exp(-\Gamma' \tau) d\Gamma', \ \tau \geq 0 \tag{10b}$$

Analogous to Eq. 7, the condition $\chi^{(3)}(\tau < 0) = 0$ introduces the physical consideration that there is no Raman coherence before the inter-action of the molecular ensemble with the pump/Stokes laser pulse. The approximation in Eq. 10b is coarser than the one in the frequency domain, because the weights given in Eq. 8 eliminate the error only at a single point in the temporal grid. Moreover, the integral transform is predicated on the fact that neighboring lineshapes in the grid differ only slightly, which is not the case in the time domain where the prototype lineshapes rapidly oscillate (see Supplementary Information, section "Approximation error—time-domain approximation").

In Fig. 4, we compare the dephasing of the integrated Q-branch signal over a delay of 200 ps, as simulated by the conventional and ultrafast spectral algorithms. In Fig. 4a, b, we compare three different approaches for the synthesis of the probe delay scan, including the 250 most intense lines in the Q-branch spectrum: (i) *ref-py-w*: integration of $I(\omega)$ from Eq. 4, resulting in a computation time of 92.4 s; (ii) *ufa-cpp-w*: integration of $I(\omega)$ from Eq. 7 (computation time: 1.0 s); and (iii) *ufa-cpp-t*: direct computation of $I(\tau)$ using Eq. 10a (computation time: 4.8 ms). The error is shown to be minimal for the *ufa-cpp-w* case, but significantly larger for the direct approach implemented as *ufa-cpp-t*, showing a clear parabolic structure as expected (see Supplementary Information, section "Approximation error—time-domain approximation").

In Fig. 4c, d, we employ the two ultrafast approaches, based on the frequency-domain integration (*ufa-cpp-w*) and direct time-domain calculation (*ufa-cpp-t*), to compute the dephasing of the Q-branch sig-nal, including all the corresponding 2.5 M Raman lines. The perfor-mance of the time-domain implementation is assessed by employing three different combinations of grid resolution for the frequency and damping axes. Increasing the number of gridpoints has a clear benefit on the time-domain computation, reducing the absolute relative error to less than 1%.

Finally, Fig. 4e, f presents the collisional dephasing of the Q-branch at temperatures $T = 296$, 800, and 1500 K, synthesized by *ufa-cpp-t* for $N_\omega = 40{,}000$ and $N_\Gamma = 4$. For all the tested temperatures, the error with respect to the frequency-domain integration is less than 2% and appears to be insensitive to temperature. The calculation time for the three traces is below 7 ms on average. We furthermore remark that the comparison between the two implementations of the ultrafast algorithm is repre-sentative of the accuracy of *ufa-cpp-t* with respect to the spectrally-
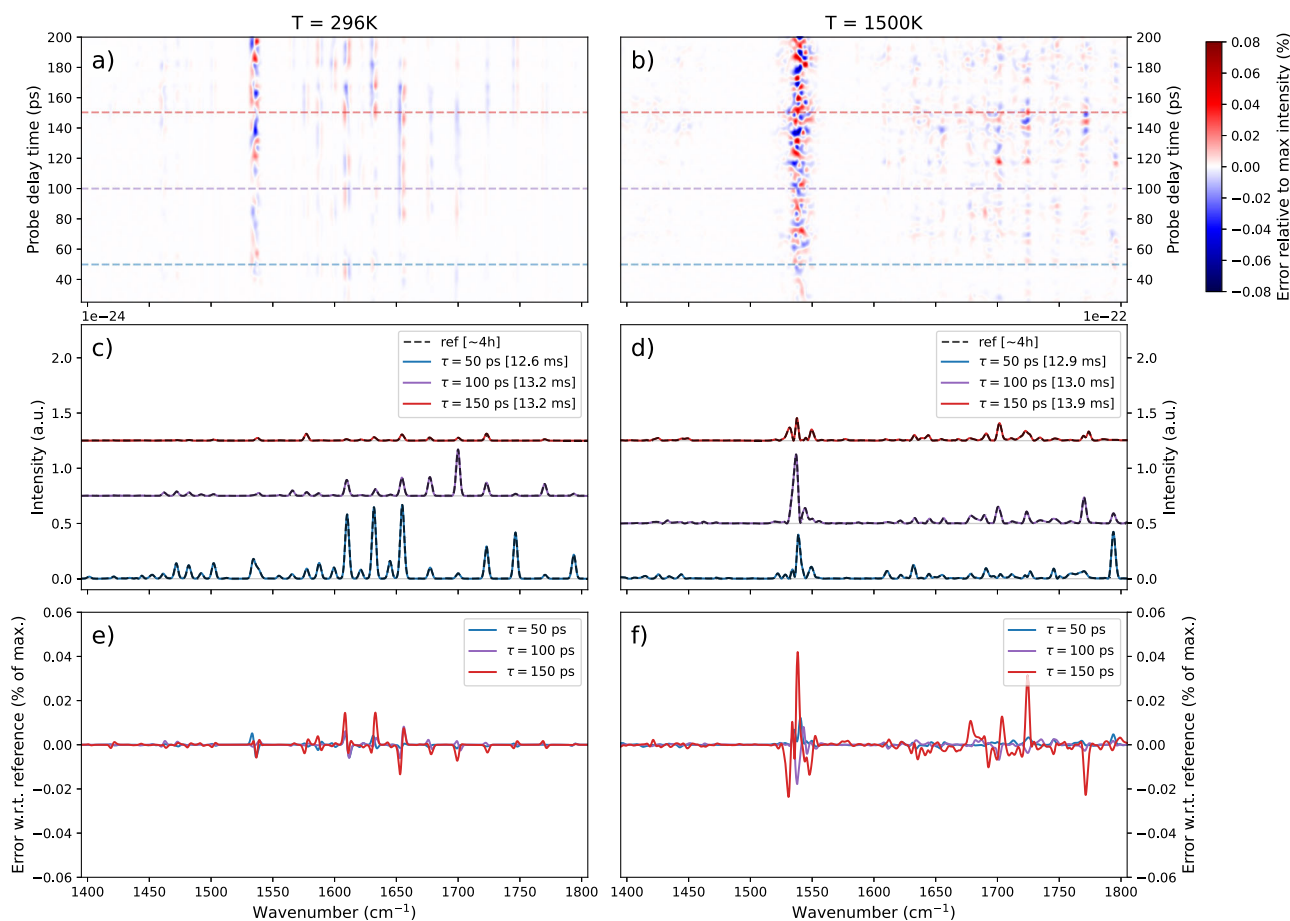
Fig. 3 | Accuracy of the ultrafast algorithm. a, b Map of the approximation error of the frequency-domain implementation of the ultrafast algorithm (*ufa-cpp*) w.r.t the reference (*ref-py*), evaluated at room temperature ($T = 296$ K, **a**) and elevated temperature ($T = 1500$ K, **b**), respectively, for a range of probe delays from 20 to 200 ps. The error is taken relative to the largest peak in the spectrum. Regardless of temperature and probe delay, the error never exceeds 0.1%. For probe delays of 50, 100, and 150 ps, representative spectra are plotted in figure **c** for $T = 296$ K and **d** for $T = 1500$ K, as indicated by the dashed lines in (**a**, **b**). The residuals, imperceptible in the (**c**, **d**) figures, are plotted in figures (**e**, **f**), respectively.

integrated conventional time-domain CRS model (*ref-py-w*). We assessed this by spectrally integrating the error maps in Fig. 3a, b, the resulting approximation error in *ufa-cpp-w* was found to be <0.03% at all probe delays, for both temperatures.

Discussion

In summary, we have introduced an algorithm for the ultrafast synthesis of time-resolved CRS spectra from large spectral databases, demonstrating a reduction in the computational cost by six orders-of-magnitude, while maintaining the approximation error below 0.1% in the frequency domain. The achieved performance increase allows for on-the-fly evaluation of the CRS signal, capable of live monitoring gas-phase chemical reactions. With spectral computation times of ~15 ms, fitting a spectrum typically takes under a second instead of many hours.

Besides the considerable improvement over existing spectral codes, the present algorithm has significant potential to enable the fruitful use of quantitative CRS spectroscopy in currently unexplored research domains. Here, we suggest a couple of prospect applications in the context of extraterrestrial and plasma chemistry. Gaseous CH_4 is one of the key components in the atmosphere of astronomical objects such as exoplanets and brown dwarves[34], as well as natural satellites, Titan being a fine example[37]. As such, CH_4 plays a key role in the physical (e.g., light absorption) and chemical (e.g., photochemistry) processes that shape these atmospheres, and large spectral line lists are available to investigate its rotational-vibrational spectrum at temperatures as high as 2000 K[34].

To the best of our knowledge, the ultrafast algorithm presented here is the only tool available to handle such large spectral databases without the need for additional simplifications, such as quasi-continuum cross-sections, which introduce arbitrary data cutoffs. This makes time-resolved CRS ideal for the laboratory investigation of CH_4 chemistry under the extreme conditions found, e.g., in the hypersonic entry in Titan's atmosphere[38] or, possibly, in the prebiotic synthesis of complex organic molecules[39].

Another likely application is thermometry in methane plasma discharges, which are rapidly gaining attention as a potential electric alternative for contemporary fossil-powered chemical processes[40,41]. The challenging conditions in the plasma make CRS an attractive diagnostic technique, but due to the high temperature in the plasma (>3000 K), as well as possible non-equilibrium thermodynamics effects, reducing the number of lines included in the synthetic spectrum comes at the risk of losing of spectral accuracy. With the ultrafast algorithm, no such tradeoff needs to be made anymore.

While CH_4 is the simplest hydrocarbon, polycyclic aromatic hydrocarbons (PAHs) are large organic molecules, which are also abundant in the interstellar medium and play a significant role in the hydrogen and carbon chemistry in planetary nebulae[42]. PAHs are also the precursors of soot in hydrocarbon combustion and are thus the subject of widespread laboratory research. The application of CRS in the fingerprint region[26] of the Raman spectrum can combine the high chemical specificity required to identify and measure different PAHs[43], with
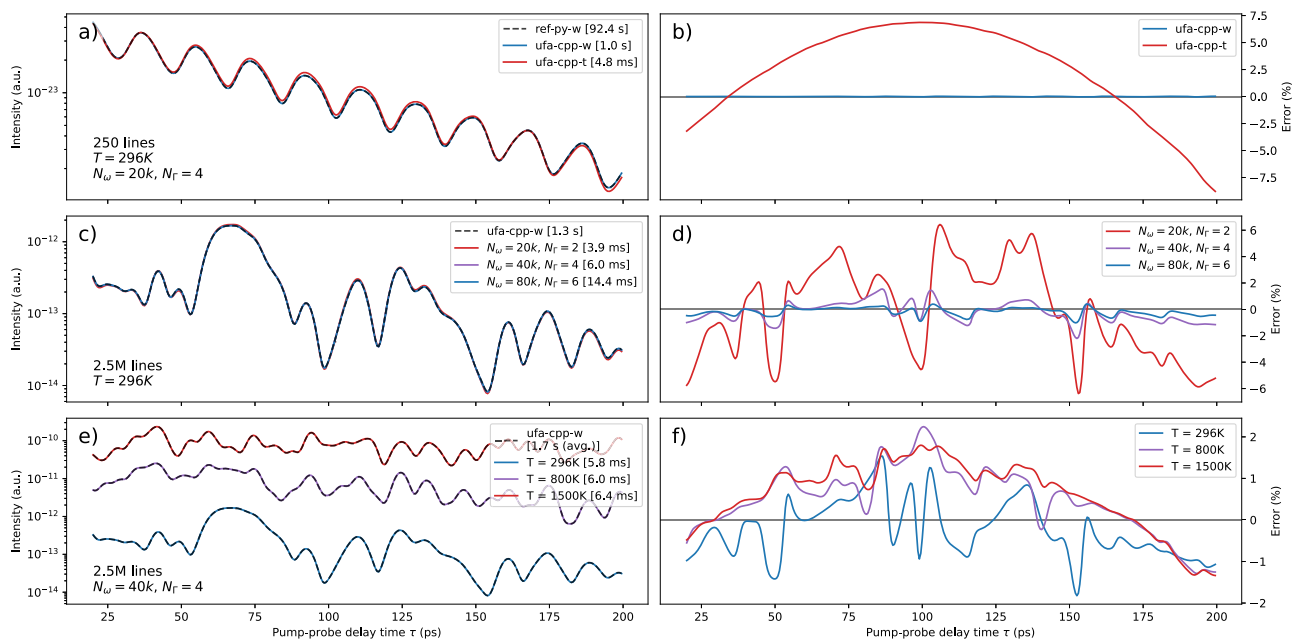
Fig. 4 | Accuracy of the fast algorithm in the time domain. a The fast approximation is compared against the reference for the 250 most intense lines in the Q-branch. Integration in frequency space (-*w*) yields accurate results in either case, so *ufa-cpp-w* is used as a fast reference for the other plots. **b** The error for the direct time-domain (*ufa-cpp-t*) is somewhat larger, ±7.5%. A parabolic profile is clearly visible in the error. **c** Error evaluated for the direct approach (*ufa-cpp-t*) for different numbers of gridpoints N_ω and N_Γ. **d** the error in the time domain is significantly reduced when a larger number of points is considered. **e** the coherence decay is evaluated for $N_\omega = 40\,000$ and $N_\Gamma = 4$ at different temperatures. At $T = 296$ K, $\delta = 1$ and $n = 0$ were used. At $T > 296$ K, the values $\delta = 0.47$ and $n = 2.67$ were taken from ref. 24 (see Methods). **f** The error remains within 2% for all values of T.

the ability to resolve their specific vibrational dynamics in the ground electronic state, thus providing important complementary information to time-resolved XUV-IR spectroscopy[44]. Currently, however, comprehensive line lists for such large molecules are unavailable. For example, a line-by-line database for the much smaller ethylene molecule, the ECaSDa database[25], (accessible at https://vamdc.icb.cnrs.fr/PHP/ecasda.php), contains only 6791 Raman transitions. It is our hope and expectation that removing the barrier for computation of such complex spectra will provide a large impetus to the development of line lists for these larger molecules.

Methods
Discretization
For use in numerical programs, the expression for $P_{CRS}(t)$ must be discretized. This is achieved by replacing the continuous FT by a DFT, and replacing the integral over Γ by the summation:

$$P_{CRS}[n] = E_{pr}[n] \cdot \sum_l \text{IDFT}_{n \leftarrow k} \left\{ \widehat{X}^{(3)}[k, l] \right\} \exp(-\Gamma[l] t[n]), \quad (11)$$

with $t[n] = t_0 + n\Delta t$ and $\omega[k] = \omega_0 + k\Delta\omega$ the discretized time/frequency axes, $\Gamma[l] = \Gamma_0 + l\Delta\Gamma$ the discretized damping axis, and

$$\widehat{X}^{(3)}[k, l] = \widehat{W}[k, l] \exp\left(j\omega[k] - \Gamma[l]\right)\tau, \quad (12)$$

the discretized distribution function, with the discrete line strength distribution $\widehat{W}[k, l]$ to be determined.

The key to discretizing $W(\omega, \Gamma)$, and indeed to the presented approach in general, is that the exact nonlinear susceptibility of an individual line, $\widehat{\chi}_i^{(3)}(\omega) = \mathcal{F}\{\chi_i^{(3)}(t)\}$, can be approximated by the weighted sum of susceptibilities with line positions $\omega[k]$ and damping coefficients $\Gamma[l]$ on the discrete grid. The contribution to the distribution function by a single line i

for the approximated line position and damping coefficients is given, respectively, by:

$$\widehat{X}_i^{(3)}(\omega_i, \Gamma_i; \tau) \approx W_i\left(1 - a_{\omega,i}\right) \exp\left(j\omega[k_i] - \Gamma_i\right)\tau \\ + W_i a_{\omega,i} \exp\left(j\omega[k_i + 1] - \Gamma_i\right)\tau, \quad (13a)$$

$$\widehat{X}_i^{(3)}(\omega_i, \Gamma_i; \tau) \approx W_i\left(1 - a_{\Gamma,i}\right) \exp\left(j\omega_i - \Gamma[l_i]\right)\tau \\ + W_i a_{\Gamma,i} \exp\left(j\omega_i - \Gamma[l_i + 1]\right)\tau, \quad (13b)$$

with $a_{\omega,i}$ and $a_{\Gamma,i}$ the weights for line position and damping coefficient, respectively, with a value between 0 and 1, and $k_i = \left\lfloor \frac{\omega_i - \omega_0}{\Delta\omega} \right\rfloor$ and $l_i = \left\lfloor \frac{\Gamma_i - \Gamma_0}{\Delta\Gamma} \right\rfloor$ the indices of the closest gridpoints before (i.e., "to the left of") the molecular line.

Equations 13a and 13b can be combined to approximate the exact nonlinear susceptibility as the weighted sum of 2 × 2 susceptibilities, each with line parameters only at discrete gridpoints:

$$\widehat{X}_i^{(3)}(\omega_i, \Gamma_i; \tau) \approx W_i \cdot \sum_{\delta k, \delta l \in 0,1} a_{\delta k, \delta l}^i \exp\left\{ \left(j\omega[k_i + \delta k] - \Gamma[l_i + \delta l]\right)\tau \right\}, \quad (14)$$

where δk and δl (not to be confused with the Kronecker delta's δ_{ij}) each can take the values 0 or 1, and the weights $a_{\delta k, \delta l}^i$ are simply the products of the individual weights in Eq. 14, i.e.:

$$a_{\delta k, \delta l}^i = f_{\delta k}(a_{\omega,i}) f_{\delta l}(a_{\Gamma,i}), \quad (15a)$$

$$f_n(a) = \begin{cases} 1 - a & \text{if } n = 0 \\ a & \text{if } n = 1 \end{cases} \quad (15b)$$

The numerical schemes employed for a computationally efficient evaluation of the position ($a_{\omega,i}$) and damping ($a_{\Gamma,i}$) weights are detailed in Supplementary Information, section "Weight derivation and evaluation".

The discrete distribution of the line strength $\widehat{W}[k, l]$ is then constructed by every molecular line contributing its weighted intensity to the corresponding 2×2 gridpoints. The intensity W_i is distributed over the gridpoints by applying the weights in Eq. 8, resulting in the following expression for $\widehat{W}[k, l]$:

$$\widehat{W}[k, l] = \sum_i \sum_{\delta k, \delta l \in 0,1} W_i \, a_{\delta k, \delta l}^i \, \delta_{k, k_i + \delta k} \, \delta_{l, l_i + \delta l}. \tag{16}$$

The entire procedure is summarized schematically in Fig. 1.

Benchmarked implementations

Three implementations are benchmarked on a mid-end consumer laptop with an 8-core AMD Ryzen PRO 5850U processor (1901 MHz) on Windows 10. The reference case is implemented purely in Python (*ref-py*) and C++ (*ref-cpp*). The compiled C++ version of the reference, compiled with MinGW, benefitted from a couple of straightforward optimizations, that resulted in about doubling the computational performance.

For the ultrafast approximation, two implementations are benchmarked: a pure Python implementation (*ufa-py*); and a C++ implementation (*ufa-cpp*), which implements multithreading through OpenMP pragmas as well as single instruction multiple data (SIMD) vectorization. The FT evaluations in the C++ implementation are performed using the PocketFFT library[45], which is also the basis for the FFT implemented in SciPy. In the multi-threaded code, the computation of the line strength distribution is split up into a number of parallel threads. Each thread processes a number of lines that are separated by a "chunk size", which is 2048 as a default to prevent multiple threads from adding lines in the same wavelength region. In case the threads start to overlap, a larger chunk size or a smaller number of parallel threads should be used. Full vectorization is implemented through AVX2 (Advanced Vector eXtensions 2) SIMD intrinsics, which enables simultaneous processing of 4 spectral lines on top of the multithreading parallelization.

The *ref-py* and *ref-cpp* codes were implemented from scratch based on the approach by ref. 26 and demonstrated a frequency shift of $0.01115 \, \text{cm}^{-1}$ compared to the original code. After comparison to the nominal Raman frequencies calculated in the MeCaSDa database, the discrepancy was attributed to a numerical artifact due to the FFT algorithm of the original code. Like the *ref-py* and *ref-cpp* codes, *ufa-py* and *ufa-cpp* reproduce the correct line positions. In all the benchmarked implementations, the Fourier-conjugated time and frequency axes were discretized using $N_\omega = N_t = 40{,}000$ points, resulting in a resolution of $\Delta\omega/2\pi c = 0.02 \, \text{cm}^{-1}$ and $\Delta t = 40 \, \text{fs}$, with the time axis spanning from -814 to 814 ps. Additionally, the "ufa" implementation employed $N_\Gamma = 4$ points to discretize the linewidth space. The same settings were used for both Figs. 2, 3.

The benchmarking is performed by evaluating a total of 3×3 spectra, with $T = 296, 800, 1500 \, \text{K}$, and $\tau = 20, 50, 100 \, \text{ps}$; the evaluation is carried out for an increasing number of spectral lines considered in the computation, from 10 to the ~11 M lines of the whole MeCaSDa database. To test the performance of the ultrafast algorithm up to 100 M spectral lines, the MeCaSDa database is synthetically extended by duplicating every spectral line in the database 10 times. The computation times reported in Results are the average of the nine spectral evaluations performed at each tested point, the standard deviation in the computation times for the nine evaluations is in each case around ~20% for computation times below ~5 ms and drops to ~2% for a larger computation time.

Linewidths model

The width of the Raman lines considered in the MeCaSDa database are here calculated according to ref. 26 The main dephasing mechanism for the Raman coherence is assumed to be the rotational energy transfer in molecular collisions, which is estimated by the modified exponential gap (MEG) scaling law:

$$\Gamma_i = \sum_{k \neq l} \gamma_{kj}, \tag{17a}$$

$$\gamma_{kj} = \alpha p \left(\frac{T_0}{T}\right)^n \left(\frac{1 + aE_j/k_B T \delta}{1 + aE_j/k_B T}\right)^2 \exp\left[\frac{-\beta\left(E_k - E_j\right)}{k_B T}\right], \tag{17b}$$

$$\gamma_{jk} = \frac{2J_j + 1}{2J_k + 1} \gamma_{kj} \exp\left(\frac{E_k - E_j}{k_B T}\right). \tag{17c}$$

With γ_{kj} and γ_{jk}, respectively, the upward and downward collisional transition rates; $T_0 = 296 \, \text{K}$ the reference temperature; p, the gas pressure (assumed atmospheric); E_k and E_j, the ro-vibrational energy in the upper and lower vibrational states; k_B, Boltzmann's constant; and fitting parameters α, β, δ, a and n. The species-specific parameter a is set to 2, while $\alpha = 4.45 \cdot 10^{-2}$ and $\beta = 1.52$ according to ref. 26. The temperature-related coefficients n and δ, on the other hand, are not reported in the literature for the ν_2 mode spectrum of CH_4. To a first approximation, the values reported by ref. 24 for the ν_1 mode spectrum are here adopted to simulate the ν_2 mode one: $\delta = 0.47$ and $n = 2.67$.

Data availability

The data used in Fig. 2 is attached as Supplementary Data 1. Any remaining datafiles are available from the corresponding author upon request.

Code availability

All code generated in the current study is attached as Supplementary Software 1 and is additionally available on the Github repository: https://github.com/dcmvdbekerom/ultrafast-crs. The remaining code files are available from the corresponding author upon request.

References

1. Zewail, A. H. Laser femtochemistry. *Science* **242**, 1645–1653 (1988).
2. Corkum, P. B. & Krausz, F. Attosecond science. *Nat. Phys.* **3**, 381–387 (2007).
3. Lavorel, B. et al. Femtosecond Raman time-resolved molecular spectroscopy. *C. R. Phys.* **5**, 215–229 (2004).
4. Scully, M. O. et al. FAST CARS: engineering a laser spectroscopic technique for rapid identification of bacterial spores. *Proc. Natl. Acad. Sci. USA* **99**, 10994–11001 (2002).
5. Roy, S., Gord, J. R. & Patnaik, A. K. Recent advances in coherent anti-Stokes Raman scattering spectroscopy: fundamental developments and applications in reacting flows. *Prog. Energy Combust. Sci.* **36**, 280–306 (2010).
6. Lempert, W. R. & Adamovich, I. V. Coherent anti-Stokes Raman scattering and spontaneous Raman scattering diagnostics of nonequilibrium plasmas and flows. *J. Phys. D* **47**, 433001 (2014).
7. Bohlin, A. & Kliewer, C. J. Diagnostic imaging in flames with instantaneous planar coherent Raman spectroscopy. *J. Phys. Chem. Lett.* **5**, 1243–1248 (2014).
8. Bohlin, A. & Kliewer, C. J. Direct coherent Raman temperature imaging and wideband chemical detection in a hydrocarbon flat flame. *J. Phys. Chem. Lett.* **6**, 643–649 (2015).
9. Milner, A. A., Korobenko, A., Hepburn, J. W. & Milner, V. Effects of ultrafast molecular rotation on collisional decoherence. *Phys. Rev. Lett.* **113**, 1–5 (2014).
10. Chen, T. Y., Steinmetz, S. A., Patterson, B. D., Jasper, A. W. & Kliewer, C. J. Direct observation of coherence transfer and rotational-to-vibrational energy exchange in optically centrifuged CO_2 super-rotors. *Nat. Commun.* **14**, 3227 (2023).

11. Milner, A. A. & Milner, V. Controlled excitation of rotons in superfluid helium with an optical centrifuge. *Phys. Rev. Lett.* **131**, 166001 (2023).

12. Hollas, M. J. *High Resolution Spectroscopy* (Butterworths, 1982).

13. Hall, R. J. & Stufflebeam, J. H. Quantitative CARS spectroscopy of CO_2 and N_2O. *Appl. Opt.* **23**, 4319–4327 (1984).

14. Greenhalgh, D. A., Hall, R. J., Porter, F. M. & England, W. A. Application of the rotational diffusion model to the CARS spectra of high-temperature, high-pressure water vapour. *J. Raman Spectrosc.* **15**, 71–79 (1984).

15. Jahn, H. A. A new Coriolis perturbation in the methane spectrum I. Vibrational-rotational Hamiltonian and wave functions. *Proc. R. soc. Lond. Ser. A* **168**, 469–495 (1938).

16. Jourdanneau, E. et al. CARS methane spectra: experiments and simulations for temperature diagnostic purposes. *J. Mol. Spectrosc.* **246**, 167–179 (2007).

17. Akhmanov, S. A. et al. CARS thermometry of polyatomic gases: SF_6. *IEEE J. Quantum Electron.* **20**, 424–428 (1984).

18. Walser, A. M. et al. Time-resolved investigation of the v_1 ro-vibrational Raman band of H_2CO with fs-CARS. *J. Raman Spectrosc.* **38**, 147–153 (2007).

19. Hosseinnia, A., Nordström, E., Bood, J. & Bengtsson, P.-E. Ethane thermometry using rotational coherent anti-Stokes Raman scattering (CARS). *Proc. Combust. Inst.* **36**, 4461–4468 (2017).

20. Hosseinnia, A., Brackmann, C. & Bengtsson, P.-E. Pure rotational coherent anti-Stokes Raman spectroscopy of ethylene, experiments and modelling. *J. Quant. Spectrosc. Radiat. Transf.* **234**, 24–31 (2019).

21. Tennyson, J. et al. The ExoMol database: molecular line lists for exoplanet and other hot atmospheres. *J. Mol. Spectrosc.* **327**, 73–94 (2016).

22. Richard, C., Boudon, V. & Rotger, M. Calculated spectroscopic databases for the VAMDC portal: new molecules and improvements. *J. Quant. Spectrosc. Radiat. Transf.* **251**, 107096 (2020).

23. Carteret, C. & Labrosse, A. Vibrational properties of polysiloxanes: From dimer to oligomers and polymers. 1. structural and vibrational properties of hexamethyldisiloxane $(CH_3)_3SiOSi(CH_3)_3$. *J. Raman Spectrosc.* **41**, 996–1004 (2010).

24. Chen, T. Y., Kliewer, C. J., Goldberg, B. M., Kolemen, E. & Ju, Y. Time-domain modelling and thermometry of the CH_4 v_1 Q-branch using hybrid femtosecond/picosecond coherent anti-Stokes Raman scattering. *Combust. Flame* **224**, 183–195 (2021).

25. Awa Ba, Y. et al. MeCaSDa and ECaSDa: methane and ethene calculated spectroscopic databases for the virtual atomic and molecular data centre. *J. Quant. Spectrosc. Radiat. Transf.* **130**, 62–68 (2013).

26. Mazza, F. et al. The ro-vibrational v_2 mode spectrum of methane investigated by ultrabroadband coherent Raman spectroscopy. *J. Chem. Phys.* **158**, 094201 (2023).

27. van den Bekerom, D. C. M. & Pannier, E. A discrete integral transform for rapid spectral synthesis. *J. Quant. Spectrosc. Radiat. Transf.* **261**, 107476 (2021).

28. Pannier, E. & Laux, C. O. RADIS: a nonequilibrium line-by-line radiative code for CO_2 and HITRAN-like database species. *J. Quant. Spectrosc. Radiat. Transf.* **222–223**, 12–25 (2019).

29. Prince, B. D., Chakraborty, A., Prince, B. M. & Stauffer, H. U. Development of simultaneous frequency- and time-resolved coherent anti-Stokes Raman scattering for ultrafast detection of molecular Raman spectra. *J. Chem. Phys.* **125**, 044502 (2006).

30. Pestov, D. et al. Optimizing the laser-pulse configuration for coherent Raman spectroscopy. *Science* **316**, 265–268 (2007).

31. Mankoč-Borštnik, N., Fonda, L. & Bortnik, B. Coherent rotational states and their creation and time evolution in molecular and nuclear systems. *Phys. Rev. A* **35**, 4132 (1987).

32. Friedrich, B. & Herschbach, D. Alignment and trapping of molecules in intense laser fields. *Phys. Rev. Lett.* **74**, 4623–4626 (1995).

33. Faeder, J., Pinkas, I., Knopp, G., Prior, Y. & Tannor, D. J. Vibrational polarization beats in femtosecond coherent anti-Stokes Raman spectroscopy: a signature of dissociative pump-dump-pump wave packet dynamics. *J. Chem. Phys.* **115**, 8440–8454 (2001).

34. Yurchenko, S. N., Amundsen, D. S., Tennyson, J. & Waldmann, I. P. A hybrid line list for CH_4 and hot methane continuum. *Astron. Astrophys.* **605**, A95 (2017).

35. Butterworth, T. D. et al. Quantifying methane vibrational and rotational temperature with Raman scattering. *J. Quant. Spectrosc. Radiat. Transf.* **236**, 106562 (2019).

36. Owschimikow, N. et al. Cross sections for rotational decoherence of perturbed nitrogen measured via decay of laser-induced alignment. *J. Chem. Phys.* **133**, 044311 (2010).

37. Niemann, H. B. et al. The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* **438**, 779–784 (2005).

38. Schipper, A. M. & Lebreton, J.-P. The Huygens probe-space history in many ways. *Acta Astronaut.* **59**, 319–334 (2006).

39. Miller, S. L. & Urey, H. C. Organic compound synthesis on the primitive earth. *Science* **130**, 245–251 (1959).

40. Fincke, J. R. et al. Plasma thermal conversion of methane to acetylene. *Plasma Chem. Plasma Process.* **22**, 105–136 (2002).

41. Mašláni, A. et al. Pyrolysis of methane via thermal steam plasma for the production of hydrogen and carbon black. *Int. J. Hydrog. Energy* **46**, 1605–1614 (2021).

42. García-Hernández, D. A. et al. Formation of fullerenes in H-containing planetary nebulae. *Astrophys. J. Lett.* **724**, 39–43 (2010).

43. Faccinetto, A. et al. Evidence on the formation of dimers of polycyclic aromatic hydrocarbons in a laminar diffusion flame. *Commun. Chem.* **3**, 112 (2020).

44. Lee, J. W. L. et al. Time-resolved relaxation and fragmentation of polycyclic aromatic hydrocarbons investigated in the ultrafast XUV-IR regime. *Nat. Commun.* **12**, 6107 (2021).

45. Reinecke, M. Pocketfft. *GitHub.* https://github.com/mreineck/pocketfft (2025).

Acknowledgements

Author contributions

F.M.: Writing—original draft, validation, writing—review and editing. D.v.d.B.: Conceptualization, methodology, software, validation, formal analysis, writing—review and editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42004-024-01397-8.

Correspondence and requests for materials should be addressed to Dirk van den Bekerom.

Peer review information *Communications Chemistry* thanks Pieter H. Neethling and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.