Spectroscopic Techniques

Supplemental Material

Vibrational Stark Spectroscopy of Fluorobenzene Using Quantum Cascade Laser Dual Frequency Combs

Urszula Szczepaniak[1], Samuel Hayes Schneider[2], Raphael Horvath[1], Jacek Kozuch[2], Markus Geiser[1]

1 IRsweep AG, Laubisruetistrasse 44, 8712 Staefa, Switzerland

2 Department of Chemistry, Stanford University, Stanford, California 94305-5012, USA

Corresponding author email: urszula.szczepaniak@irsweep.com

(1) S/N Evaluation

We benchmark the DCS with FT-IR, therefore we use S/N as in FT-IR using the following formula: , where the numerator corresponds to the signal intensity in the measurement (i.e., the maximum difference signal from the fit described in the manuscript), and due to the different spectral characteristics of the two spectrometers, is calculated differently for DCS and FT-IR. In DCS, each spectrum consists of a set of points that correspond to the narrow laser lines that are individually detected with their standard deviation. We describe different ways of DCS quantification below. For FT-IR, is equal to the RMSE (root mean squared error), calculated for a linear fit (to zero-line) of the baselined to zero spectral region without spectral features below 1200 cm^{-1}. The obtained signal-to-noise ration (S/N) results are collected in Table S2. Of note, there are several processes taking place during the experiment, that may influence the spectral response and influence the analysis of the noise, e.g., ice formation, nitrogen bubbling, etc. Furthermore, the 2-MeTHF solvent is observed to possess a small Stark contribution in the spectral region analyzed.

Thanks to monitoring of laser line intensity in the DCS system, we can quantify the standard deviation for each spectral point, as presented in the manuscript and visualized in Figure 4 (main text) and Figure S1 below. We use the standard deviation values as the base in the following (sections 1.1–1.3) analysis.

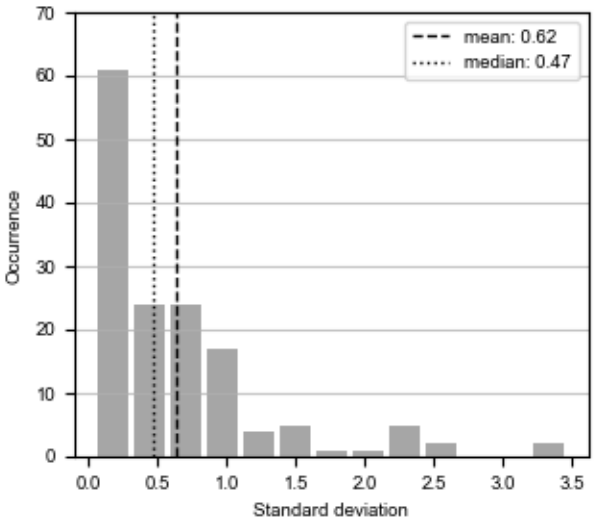

Figure S1. Histogram of DCS standard deviation values. The analysis is across all spectral elements.

1.1. DCS - Mean

The mean value of standard deviation, i.e., its arithmetic average, is the simplest way of treating the set, although not the most statistically representative. Nevertheless, this is the value that could be thought of as the most similar to FT-IR derived value. The mean of DCS noise is 0.62.

1.2. DCS - Median

The median represents how the population is distributed and is more representative of populations that are not normal. The median value for DCS noise is 0.47.

1.3. DCS - Mean of (signal over standard deviation)

Another approach can be to calculate S/N at each data point and then take the mean value of it. This will be more affected by very low noise values. A simple arithmetic average of S/N for DCS gives the value of 22.

1.4. FT-IR

From the fitting of a spectrum to zero (see above) made possible due to using the flat noise characteristics in a spectral region without spectral features, one obtains rmse = 0.32.

1.5. Summary

The fitted signal values for DCS and FT-IR are 4.53 and 5.04, respectively. For FT-IR it results in S/N of ~16. FT-IR and DCS values for different approaches are collected in Table S1. To account for the improvement of S/N coming from using the DCS instrument, we are comparing the S/N values in the form of a ratio: .

Table S1. The results of S/N analysis for spectra obtained with DCS and FT-IR instruments.

Approach				
1.1	7	16	0.44	7.5
1.2	10	16	0.63	10.7
1.3	22	16	1.4	23.6

Differences in the values collected in Table S1 highlight the difficulty of assessing S/N for a system that does not have uniformly distributed noise and that has spectral regions with very low noise levels. We decided to use noise level as determined as the median of standard deviation of signal for the DCS system. This value was used to further estimate S/N values for potentially more demanding experiments. The results are presented in Table S2.

From Figure 4 it can be seen that the C-F stretching band is located in the low-noise region of the DCS spectrum. However, it should be noted that this is not the origin of the observed improvement of S/N obtained thanks to DCS system, since the noise was calculated for the whole spectral window covered by DCS. The location of the low noise region only affects the estimated value (the Stark tuning rate) and its uncertainty. As laser development continues, more homogeneous power distributions will be achieved, leading to less wavelength-dependent noise levels throughout the covered spectral range.

Table S2. Comparison of performances of two setups used in this study. All predictions assume, unless stated otherwise, 100 mM solution of fluorobenzene. S/N were estimated using a square root dependence of the S/N on time and a linear dependence on concentration. The base for S/N estimations are the values from row 1.2 of Table S1.

Feature	FT-IR	Dual-comb spectrometer
S/N per 1 min acquisition time	6.3	62
Acquisition time to obtain S/N = 6.3	1 min	16 ms
Acquisition time to get S/N = 6.3 at 1 mM	167 h	159 s
Noise averaged over the spectrum	✓ [1]	✗
Possibility of weighted noise analysis	✗	✓

(1) Representation how the weighted fit was implemented – realized using Python

Requirements: numpy (as np) and scipy.optimize (as optimization).

1 Only valid when no filters are introduced in the beam path, e.g., to reduce illumination of the detector.

a. Define the fit formula (separate derivative and pseudo-Voigt formulas not shown). At first, declare A, mu, sig, and alfa, as the values you found by fitting the absorbance spectrum (a separate step).

```
def fit(x, c1, c2, c3):
    return c1*pVoigt(x, A, mu, sig, alfa)+c2*firstDer(x, A, mu, sig, alfa)+c3*secondDer(x, A, mu, sig, alfa)
```

b. Fit the data to the sum of derivatives. Inputs: fit – as defined above, WavenumberValues – *x*-axis, AbsorbanceValues – *y*-axis, x0 – initial guess of parameters (here: x0 = np.array([0, 0, 0])).

```
params, params_covariance = optimization.curve_fit(fit, WavenumberValues, AbsorbanceValues, x0, sigma=STD, absolute_sigma=True) # sigma=STD - standard deviation of each laser line intensity; absolute_sigma=True refers to the fact that exact values of standard deviation are used, without additional normalization; for FT-IR, sigma and absolute_sigma are not specified and used as default. See below.
"""sigma : None or M-length sequence or MxM array, optional
    Determines the uncertainty in ydata. If we define residuals as r = ydata - f(xdata, *popt), then the interpretation of sigma depends on its number of dimensions:
        A 1-d sigma should contain values of standard deviations of errors in ydata. In this case, the optimized function is chisq = sum((r / sigma) ** 2).
        A 2-d sigma should contain the covariance matrix of errors in ydata. In this case, the optimized function is chisq = r.T @ inv(sigma) @ r.
        New in version 0.19.
        None (default) is equivalent of 1-d sigma filled with ones.
    """

yfitDCS = fit(WavenumberValues, params[0], params[1], params[2])
```

c. Find the Stark tuning rate and its uncertainty

Calculate the Stark tuning rate:

```
DmuDCS  = np.sqrt(10*params[2])
```

Calculate the uncertainty on the Stark tuning rate:

```
paramsErr = np.sqrt(10*np.sqrt(np.diag(params_covariance)))
```

The Stark tuning rate is DmuDCS +/- paramsErr[2]

d. For S/N analysis, one needs to obtain signal value from the fit:

```
sigDCS = max(yfitDCS)-min(yfitDCS) # signal DCS – from the fit not to rely on noisy data points
```

(2) **The Full FT-IR Spectrum**

The full recorded FT-IR spectrum is presented in Figure S2.

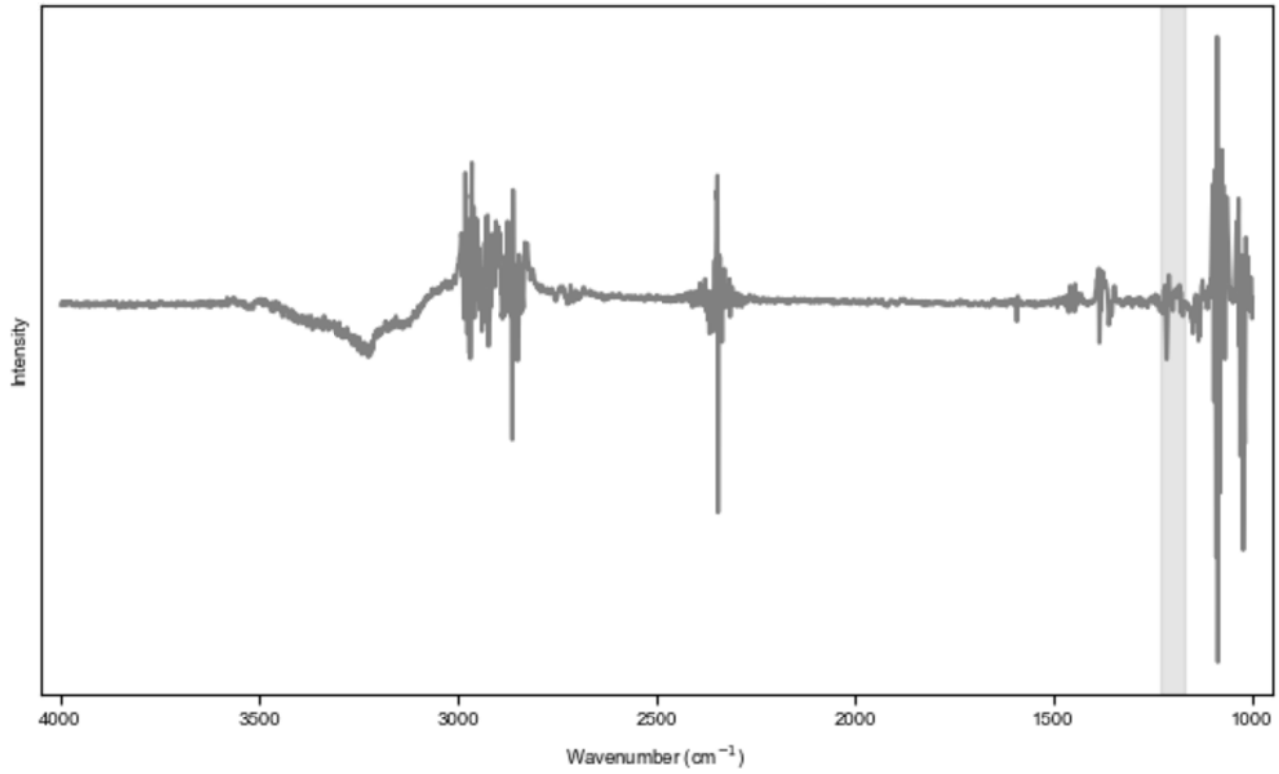

Figure S2. Full FT-IR spectrum. Note features corresponding to, e.g., ice formation. Vertical bar corresponds to the window of analysis for the Stark tuning rate determination.