Develop and Implement data analyses pipeline for time resolved absorption spectroscopy and make a comparative study

R. J. Dwayne Miller

NSERC Undergraduate Research Position - Summer 2023

Time resolved spectroscopic techniques play an important role in spectrally and temporally resolving[1] the photoinduced non-equilibrium dynamics to help elucidate the fundamental mechanism of light-driven processes spanning from photosynthesis to energy harvesting to understanding the initial steps of vision. With the advent of modern laser systems, chemical dynamics in the range of a few femtoseconds to milliseconds (10^{-15}-10^{-6} s) can routinely be monitored and investigated. Experimental difficulties to obtain time-resolved optical signal is often augmented by the extensive data analysis procedure that follows. Several methods have been employed to extract the underlying kinetic model that best represents the dynamics of the molecular system after photoexcitation.

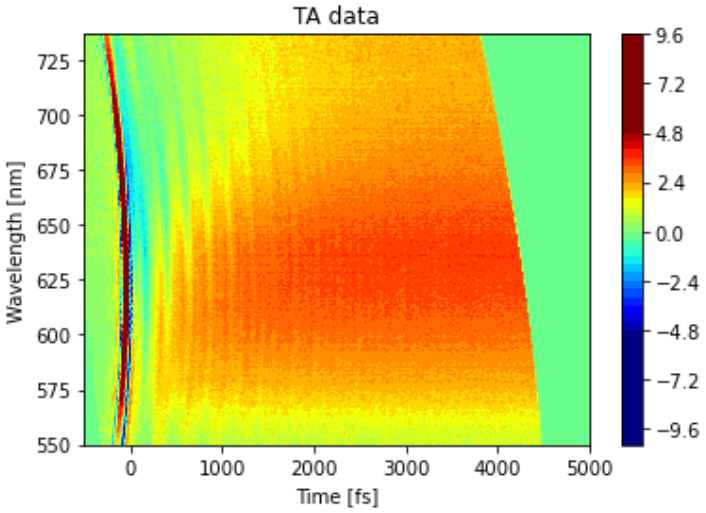

Fig 1: Prototype of a 2-dimensional transient absorption data

Many methods[2,3] have been employed to unequivocally assign the individual spectral components from the overlapping time-resolved data. Ranging from multivariate global analysis to regression analysis. The pipeline for each of these schemes, however, is different. Some of them require *a priori* knowledge on the investigated model as inputs to the data analysis pipeline. Thus, for each method that is in use, it may involve the risk of ambiguities depending on the user choices. In this project, we will carry out transient absorption spectroscopy experiments with ultrafast lasers (70-300 fs, 1-10 mJ/pulse) on a suitable sample both in the solid crystalline and solution phase. Then we'll try to implement different data analysis schemes to a 2D transient absorption dataset and compare

$$\boldsymbol{X}(\lambda, t) = (\boldsymbol{S_1}(\lambda), \boldsymbol{S_2}(\lambda), \dots) \times \begin{pmatrix} \boldsymbol{C_1}(t) \\ \boldsymbol{C_2}(t) \\ . \\ . \\ . \end{pmatrix}$$

Fig 2: Experimental data $\boldsymbol{X}(\lambda, t)$ can be represented as the summation of basis spectra $\boldsymbol{S_i}(\lambda)$weighted by their amplitude $\boldsymbol{C_i}(t)$

their advantages and disadvantages. The followings are the pipelines for data analysis procedure that we plan to implement in this project and compare:

1. Single Value Decomposition of the 2D time-wavelength dataset to extract the spectral and kinetic basis spectra.
2. Global/Target analysis (with a kinetic model)
3. Kinetic model free analysis of transient absorption spectra (2D correlation analysis, Markov chain Monte-Carlo based analysis[4])

Expected outcome of the project:

1. Learn the basic principle of time-resolved spectroscopic techniques in the pump-probe scheme.
2. Sample preparation and carrying out transient absorption experiment in both solution and solid (crystalline) phase.
3. Developing algorithm for three different scheme of time-resolved data analysis (SVD-Global analysis, model free 2D correlation analysis, Markov Chain Monte-Carlo based transient absorption data analysis)
4. Compare the advantage/disadvantage of each of these schemes.

The project is interdisciplinary (combination of physics, chemistry and biology) and well suited for a 3rd / 4th year undergrad student. Basic familiarity with programming is helpful. For questions and inquiries please contact Nita Ghosh at : nita.ghosh@mail.utoronto.ca

Ref:

1. Gruebele, M. & Zewail, A. H. Femtosecond wave packet spectroscopy: Coherences, the potential, and structural determination. *J. Chem. Phys.* **98**, 883 (1998).

2. Kollenz, P., Herten, D. P. & Buckup, T. Unravelling the Kinetic Model of Photochemical Reactions via Deep Learning. *J. Phys. Chem. B* **124**, 6358–6368 (2020).

3. Beechem, J. M. [2] Global analysis of biochemical and biophysical data. *Methods Enzymol.* **210**, 37–54 (1992).

4. Ashner, M. N., Winslow, S. W., Swan, J. W. & Tisdale, W. A. Markov Chain Monte Carlo Sampling for Target Analysis of Transient Absorption Spectra. *J. Phys. Chem. A* **123**, 3893–3902 (2019).