The Journal of Open Source Software

WrightTools: a Python package for multidimensional spectroscopy

Blaise J. Thompson[1], **Kyle F. Sunden**[1], **Darien J. Morrow**[1], **Daniel D. Kohler**[1], **and John C. Wright**[1]

1 University of Wisconsin–Madison

DOI: 10.21105/joss.01141

Software

- Review
- Repository
- Archive

Submitted: 16 December 2018
Published: 17 January 2019

Introduction

"Multidimensional spectroscopy" (MDS) is a family of analytical techniques that record the response of a material to multiple stimuli—typically multiple ultrafast pulses of light. This approach has several unique capabilities;

- resolving congested states (Donaldson et al., 2008; W. Zhao & Wright, 1999),
- extracting spectra that would otherwise be selection-rule disallowed (Boyle, Neff-Mallon, & Wright, 2013; Boyle, Neff-Mallon, Handali, & Wright, 2014),
- resolving fully coherent dynamics (Pakoulev et al., 2009),
- measuring coupling (Wright, 2011),
- and resolving ultrafast dynamics (Czech et al., 2015; Smallwood & Cundiff, 2018).

In our view, the most exciting aspect of these techniques is the vast number of different approaches that scientists can take to learn about material quantum states. Often, a number of these experiments can be accomplished with a single instrument. The diversity of related-but-unique approaches to interrogating quantum systems is an important strength of MDS.

Advancements in optics and laser science are bringing ultrafast multidimensional spectroscopy to more and more laboratories around the world. At the same time, increasing automation and computer control are allowing traditionally "one-dimensional" spectroscopies to be recorded against other dimensions.

Due to its diversity and dimensionality, MDS data is challenging to process and visualize. The tools that scientists develop to process one experiment may not work when different experimental variables are explored. Historically, MDS practitioners have developed custom, one-off data processing workflows that need to be radically changed when new experiments are undertaken. These changes take time to implement, and can become annoyances or opportunities for error. Even worse, the challenge of designing a new processing workflow may dissuade a scientist from creatively modifying their experimental strategy, or comparing their data with data taken from another instrument. This limit to creativity and flexibility defeats one of the main advantages of the MDS "family approach".

`WrightTools` is a new Python package that is made specifically for multidimensional spectroscopy. It aims to be a core toolkit that is general enough to handle all MDS datasets and processing workloads. Being built for and by MDS practitioners, `WrightTools` has an intuitive, high-level, object-oriented interface for spectroscopists. To our knowledge, `WrightTools` is the first MDS-focused toolkit to be freely avaliable and openly licensed.

Challenges and Implementation

There are several recurring challenges in MDS data processing and representation:

- There are no agreed-upon file formats. Files generated by researchers may have inconsistent internal conventions, and they often fail to be fully self-describing.
- There is a great diversity of dataset types. The same instrument is capable of producing datasets with many different combinations of scanned hardware.
- Dataset size may be large enough to run into computer memory limits.
- Dataset dimensionality is large enough to represent challenges in human interaction and visualization.

The excellent Scientific Python ecosystem is well suited to address all of these challenges (Travis E. Oliphant, 2007). Numpy supports interaction with and manipulation of multidimensional arrays (Travis E Oliphant, 2006). Matplotlib supports one, two, and even three-dimensional plotting (Hunter, 2007). h5py (Collette, 2013) interfaces with hdf5 (The HDF Group, 1997), allowing for storage and memory-safe access to large multidimensional arrays in a binary format that can be accessed from a variety of different popular languages, including MATLAB and Fortran. `WrightTools` does not intend to replace or reimplement these core libraries. Instead, `WrightTools` offers an interface that impedance-matches multidimensional spectroscopy and Scientific Python.

`WrightTools` defines a universal MDS data format: the `wt5` file. These are simply hdf5 files with certain internal conventions that are designed for MDS. These internal conventions enable the flexibility and ease-of-use that we discuss in the rest of this section. Instances of WrightTools's classes dynamically interact with the multidimensional spectroscopic data within these files. These classes are children of h5py classes. `WrightTools` offers a variety of functions that try hard to convert data stored in various other formats to `wt5`.

`WrightTools` defines a unique and flexable strategy of storing and manipulating MDS datasets. A single instance of the `WrightTools.Data` class is implemented as a group containing many separate arrays. There are two principle multidimensional array classes: `Channel` and `Variable`. Conceptually, these correspond to independent (scanned) dimensions—"variables"—and dependent (measured) signals—"channels". Channels typically contain measured signals from all of the different sensors that are employed simultaneously during a MDS experiment. Variables contain coordinates of different light manipulation hardware that are scanned against each-other to make up an MDS experiment. All variables are recorded, including coordinates for hardware that are not actually moved during that experiment (an array with one unique value) or other independent variables, such as lab time.

There can be many variables that change in the context of a single MDS experiment. The typical spectroscopist only really cares about a small subset of these variables, but exactly what subset matters may change as different strategies are used to explore the dataset. Furthermore, it is often useful to "combine" multiple variables using simple algebraic relationships to exploit the natural symmetry of many MDS experiments and to draw comparisons between different members of the MDS family (Neff-Mallon & Wright, 2017). In light of these details, `WrightTools` provides a high-level `Axis` class that allows users to transparently define which variables, variable relationships, and unit conventions are important to them for representation and manipulation. Each `Axis` contains an `expression`, which dictates its relationship with one or more variables. Given 5 variables with names `['w1', 'w2', 'wm', 'd1', 'd2']` , example valid expressions include `'w1'`, `'w1=wm'`, `'w1+w2'`, `'2*w1'`, `'d1-d2'`, and `'wm-w1+w2'`. Users may treat axes like multidimensional arrays, using `__getitem__` syntax and slicing, but axes do not themselves

contain arrays. Instead, the appropriate axis value at each dataset coordinate is computed on-the-fly using the given expression. Users may at any time change their axes by simply calling `transform` with new expressions.

`WrightTools` offers a suite of data manipulation tools with MDS in mind. Users can access portions of their data using high-level methods like `chop`, `split`, and `clip`. They can process their data using simple mathematical operations or more specific tools like `level`, `gradient`, `collapse`, and `smooth`. Users can even `join` multiple datasets together, creating higher-dimensional datasets when appropriate. All of these operations refer to the self-describing internal structure of the `wt5` file wherever possible. Users are not asked to refer to the specific shape and indicies of their data arrays. Instead, they deal with simple axis expressions and unit-aware coordinates.

`WrightTools` offers a set of "artists" to quickly draw typical representations. These make it trivial to make beautiful Matplotlib representations of MDS datasets. Again, the self-describing internal structure is capitalized upon, auto-filling labels (including units, symbols, and expressions) and auto-scaling axes. For higher-than-two dimensional datasets, `WrightTools` makes it easy to plot many separate figures that can be looped through using an image viewer or stitched into a looping animated gif. A convenience function, `interact2D`, allows users to explore a complete dataset using matplotlib's built-in widgets.

Availability

`WrightTools` is hosted on GitHub and archived on Zenodo (Thompson et al., 2018). `WrightTools` is distributed using pip and conda (through conda-forge). Documentation is available at wright.tools.

Impact

`WrightTools` has directly enabled no fewer than eleven publications (Chen et al., 2017; Czech et al., 2015; J. D. Handali, Sunden, Kaufman, & Wright, 2018; J. D. Handali et al., 2018; Horak et al., 2018; Kohler, Thompson, & Wright, 2017, 2018; Morrow, Kohler, & Wright, 2017; Morrow, Kohler, Czech, & Wright, 2018; Neff-Mallon & Wright, 2017; K. Sunden, Thompson, & Wright, 2018). Many of these publications have associated open datasets and `WrightTools`-based processing scripts which enhance the scientific community's ability to audit and reproduce the published work. Although these publications span several different MDS family members and instruments, the common usage of `WrightTools` makes it trivial to download and immediately interact with the raw and processed datasets, and (when applicable) simulations that comprise these publications. These practices are not yet common in the MDS community.

Though still relatively uncommon, MDS is an increasingly important family of analytical techniques used by Chemists and Physists to interrogate especially complex systems and to answer especially challenging questions. By abstracting away common array manpulation, file management, and data visualization tasks, `WrightTools` promises to increase the productivity and creativity of MDS practitioners. We hope that `WrightTools`, and the universal `wt5` file format, will become a useful open source core technology for this growing community. We are particularly excited about ongoing projects that build on top of `WrightTools`, including packages for data acquisition and simulation (K. Sunden et al., 2018; Thompson, Sunden, Morrow, & Neff-Mallon, 2018).

Acknowledgements

The development of `WrightTools` has been supported by by the National Science Foundation Division of Chemistry under Grant No. CHE–1709060. D.J.M acknowledges support from the Link Foundation.

References

Boyle, E. S., Neff-Mallon, N. A., & Wright, J. C. (2013). Triply Resonant Sum Frequency Spectroscopy: Combining Advantages of Resonance Raman and 2D-IR. *The Journal of Physical Chemistry A*, *117*(47), 12401–12408. doi:10.1021/jp409377a

Boyle, E. S., Neff-Mallon, N. A., Handali, J. D., & Wright, J. C. (2014). Resonance IR: A Coherent Multidimensional Analogue of Resonance Raman. *The Journal of Physical Chemistry A*, *118*(17), 3112–3119. doi:10.1021/jp5018554

Chen, J., Morrow, D. J., Fu, Y., Zheng, W., Zhao, Y., Dang, L., Stolt, M. J., et al. (2017). Single-crystal thin films of cesium lead bromide perovskite epitaxially grown on metal oxide perovskite (SrTiO3). *Journal of the American Chemical Society*, *139*(38), 13525–13532. doi:10.1021/jacs.7b07506

Collette, A. (2013). *Python and hdf5*. O'Reilly.

Czech, K. J., Thompson, B. J., Kain, S., Ding, Q., Shearer, M. J., Hamers, R. J., Jin, S., et al. (2015). Measurement of ultrafast excitonic dynamics of few-layer MoS2 using state-selective coherent multidimensional spectroscopy. *ACS Nano*, *9*(12), 12146–12157. doi:10.1021/acsnano.5b05198

Donaldson, P. M., Guo, R., Fournier, F., Gardner, E. M., Gould, I. R., & Klug, D. R. (2008). Decongestion of methylene spectra in biological and non-biological systems using picosecond 2DIR spectroscopy measuring electron-vibration–vibration coupling. *Chemical Physics*, *350*(1-3), 201–211. doi:10.1016/j.chemphys.2008.02.050

Handali, J. D., Sunden, K. F., Kaufman, E. M., & Wright, J. C. (2018). Interference and phase mismatch effects in coherent triple sum frequency spectroscopy. *Chemical Physics*, *512*, 13–19. doi:10.1016/j.chemphys.2018.05.023

Handali, J. D., Sunden, K. F., Thompson, B. J., Neff-Mallon, N. A., Kaufman, E. M., Brunold, T. C., & Wright, J. C. (2018). Three dimensional triply resonant sum frequency spectroscopy revealing vibronic coupling in cobalamins: Toward a probe of reaction coordinates. *The Journal of Physical Chemistry A*. doi:10.1021/acs.jpca.8b07678

Horak, E. H., Rea, M. T., Heylman, K. D., Gelbwaser-Klimovsky, D., Saikin, S. K., Thompson, B. J., Kohler, D. D., et al. (2018). Exploring electronic structure and order in polymers via single-particle microresonator spectroscopy. *Nano Letters*, *18*(3), 1600–1607. doi:10.1021/acs.nanolett.7b04211

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. doi:10.1109/mcse.2007.55

Kohler, D. D., Thompson, B. J., & Wright, J. C. (2017). Frequency-domain coherent multidimensional spectroscopy when dephasing rivals pulsewidth: Disentangling material and instrument response. *The Journal of Chemical Physics*, *147*(8), 084202. doi:10.1063/1.4986069

Kohler, D. D., Thompson, B. J., & Wright, J. C. (2018). Resonant third-order susceptibility of PbSe quantum dots determined by standard dilution and transient

grating spectroscopy. *The Journal of Physical Chemistry C*, *122*(31), 18086–18093. doi:10.1021/acs.jpcc.8b04462

Morrow, D. J., Kohler, D. D., & Wright, J. C. (2017). Group- and phase-velocity-mismatch fringes in triple sum-frequency spectroscopy. *Physical Review A*, *96*(6). doi:10.1103/physreva.96.063835

Morrow, D. J., Kohler, D. D., Czech, K. J., & Wright, J. C. (2018). Communication: Multidimensional triple sum-frequency spectroscopy of MoS2 and comparisons with absorption and second harmonic generation spectroscopies. *The Journal of Chemical Physics*, *149*(9), 091101. doi:10.1063/1.5047802

Neff-Mallon, N. A., & Wright, J. C. (2017). Multidimensional spectral fingerprints of a new family of coherent analytical spectroscopies. *Analytical Chemistry*, *89*(24), 13182–13189. doi:10.1021/acs.analchem.7b02917

Oliphant, T. E. (2006). *A guide to numpy*. Trelgol Publishing.

Oliphant, T. E. (2007). Python for Scientific Computing. *Computing in Science & Engineering*, *9*(3), 10–20. doi:10.1109/MCSE.2007.58

Pakoulev, A. V., Rickard, M. A., Kornau, K. M., Mathew, N. A., Yurs, L. A., Block, S. B., & Wright, J. C. (2009). Mixed frequency-/time-domain coherent multidimensional spectroscopy: Research tool or potential analytical method? *Accounts of Chemical Research*, *42*(9), 1310–1321. doi:10.1021/ar900032g

Smallwood, C. L., & Cundiff, S. T. (2018). Multidimensional coherent spectroscopy of semiconductors. *Laser & Photonics Reviews*, 1800171. doi:10.1002/lpor.201800171

Sunden, K., Thompson, B., & Wright, J. (2018). WrightSim: Using PyCUDA to simulate multidimensional spectra. In *Proceedings of the 17th python in science conference*. SciPy. doi:10.25080/majora-4af1f417-00c

The HDF Group. (1997). Hierarchical Data Format, version 5.

Thompson, B. J., Sunden, K. F., Morrow, D. J., & Neff-Mallon, N. A. (2018). Pycmds. Zenodo. doi:10.5281/zenodo.1481203

Thompson, B. J., Sunden, K. F., Morrow, D. J., Kohler, D. D., Neff-Mallon, N. A., Czech, K. J., Kaufman, E. M., et al. (2018). Wrighttools. Zenodo. doi:10.5281/zenodo.1198904

Wright, J. C. (2011). Multiresonant coherent multidimensional spectroscopy. *Annual Review of Physical Chemistry*, *62*(1), 209–230. doi:10.1146/annurev-physchem-032210-103551

Zhao, W., & Wright, J. C. (1999). Spectral Simplification in Vibrational Spectroscopy Using Doubly Vibrationally Enhanced Infrared Four Wave Mixing. *Journal of the American Chemical Society*, *121*(47), 10994–10998. doi:10.1021/ja9926414