PySpecKit: Where's the Bandgap? (PHYS265)

PySpecKit: Where's the Bandgap? Enoch Oluwaboro - May 18th, 2025.

PySpecKit is a python toolkit designed for the analysis and interpretation of spectroscopic data across optical, infared, and radio wavelengths. Its primary goal is to provide a general purpose framework for spectroscopic line analysis that is compatible with a wide range of data formats. PySpecKit enables line fitting procedures like Gaussian and Voigt profile fitting as well as baseline subtraction and interactive data visualization. It also offers tools for producing publication-quality plots with LaTeX rendered axis labels. PySpecKit was originally developed for astrophysics applications, but its flexible design makes it useful for fields such as materials science where it can be used to fit photoluminescence or absorbance spectra to extract physical properties like the optical bandgap for semiconductor materials.

I picked this package because I am in the Materials Science and Engineering major and wanted a project that could bridge the topics between what I've learned this semester in PHYS265 and my core coursework in materials science. Spectroscopy is important in astrophysics and matsci so I thought his would be a fun project.

PySpecKit was first released around 2011, developed by Adam Ginsburg and several collaborators. Its development was influenced by the need to automate and improve spectral line fitting in large radioastronomy datasets. Some related codes are specutils, pysynphot, and astropy's modeling framwork, which PySpecKit uses internally.

It appears to still be actively maintained by Adam Ginsburg and contributions can be made on the toolkit's GitHub. It is also encouraged to contact them at pyspeckit@gmail.com or post to issues.

PySpecKit was incredibly easy to install and use, I simply used pip install pyspeckit and it downloaded in a matter of seconds.

The source code is available at github.com/pyspeckit/pyspeckit

It does not appear that the code is used by other packages, but there is documentation that PySpecKit is used in over 60 papers that clearly cite the package.

PySpecKit is primarily used as a python module and is imported into Jupyter notebooks or python scripts.

For my example usage, I wanted to do a spectral fitting to extract peak positons that would correlate to the bandgap energy of silicon, an extremely popular semiconductor.

The package uses matplotlib to generate figures.
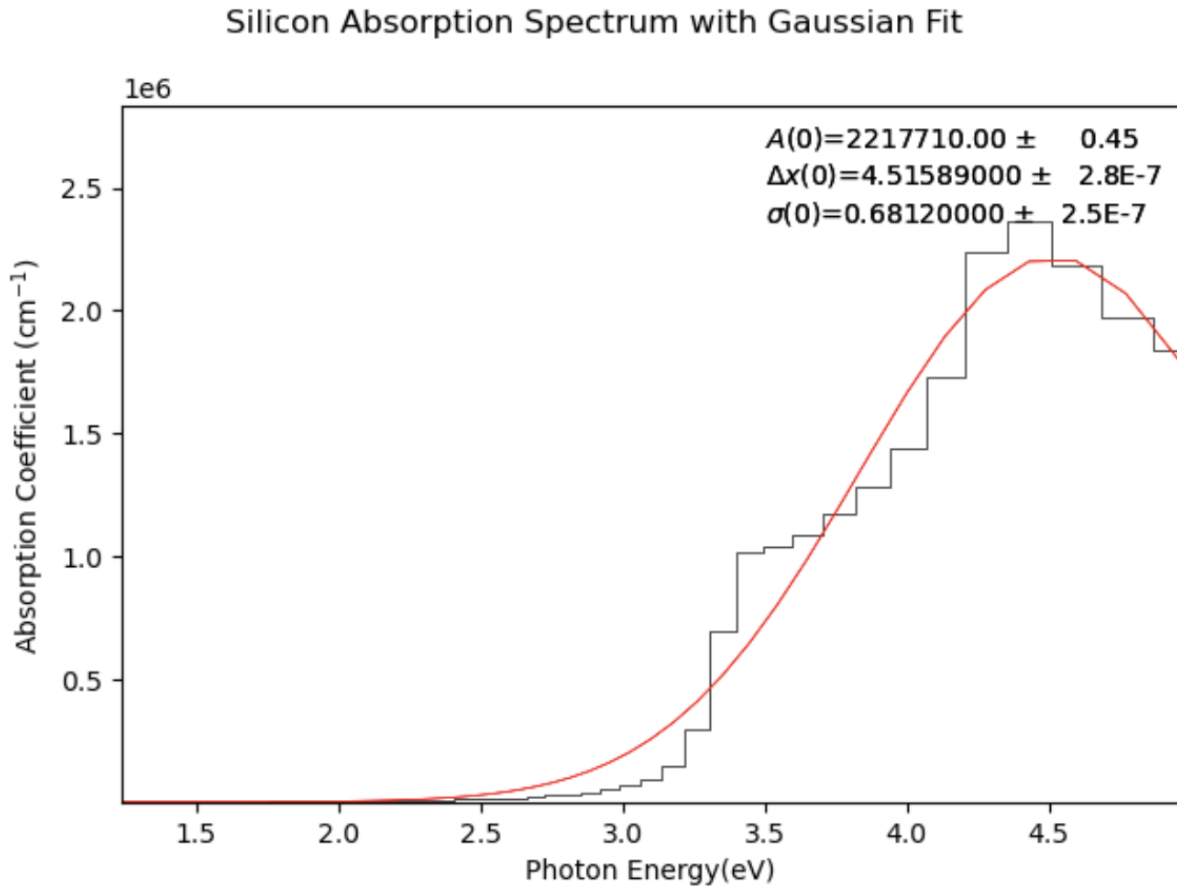
Figure 1: Gaussian Fit to Silicon Absorption Spectrum.

This plot shows the absorption coefficient of silicon as a fucntion of photon energy derived from data from Green and Keevers' research in 1995. A Gaussian profile was fit using PySpecKit to model the absorption edge. The peak's center provides an estimate of the optical bandgap energy of silicon, and the width represents the spread in transition energies.

The package is pure python and built on top of popular scientific libraries such as NumPy, SciPy, Astropy, and Matplotlib. The input for the package is a 1D spectrum, with an array of x-values and corresponding y-values. PySpecKit does support loading Flexible Image Transport System (FITS) files and other formats directly. The primary output of the package includes fitted model parameters, diagnostic plits with overlaid model fits, and residuals and goodness of fit statistics. They are accessible as python objects.

PySpecKit includes a set of unit tests located in the pyspeckits/tests/ directory of the repository on GitHub.

I'm confident that PySpecKit can produce a reliable result because of a few factors. A test suite that validates core functionality is easily acessible, as well as the usage and maintenance from researchers within the fields of spectroscopy and astronomy. It also uses established fit libraries

liek lmfit, and I can compare the fitted results to known models. Also, for this particular project, the Gaussian fit appears reasonable and matches an expected spectral shape.

PySpecKit relies on numpy, matplotlib, astropy, lmfit, and pyspeckit.spectrum.models. These were founnd in the repository in the setup.py file.

PySpecKit provides documentation in the form of a GitHub README, a documentation website (https://pyspeckit.readthedocs.io/en/latest/index.html) with examples that demonstrate typical use cases.

The preferred citation method is to cite their ads, 'Pyspeckit: A Spectroscopic Analysis and Plotting Package'

References

PySpecKit preferred citation: https://ui.adsabs.harvard.edu/abs/2022AJ....163..291G/abstract
Spectral Data for silicon: https://www.pveducation.org/pvcdrom/materials/optical-properties-of-silicon

I did not learn new python methods to use this package. Everything I used was taught in class except for this line because the data was incomplete.

data = data[~np.isnan(data).any(axis=1)]

I knew where to get the data and what to look for, but everything else was completely new to me.