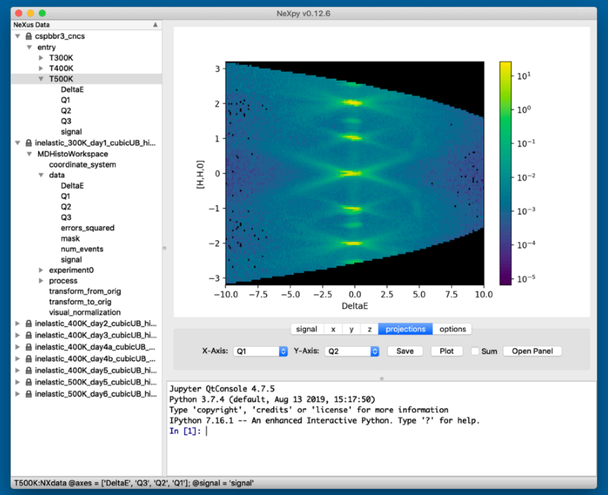
NeXpy

A GUI TOOLBOX FOR ANALYZING HDF5 DATA



RAYMOND OSBORN
Neutron & X-ray Scattering Group
Materials Science Division

https://nexpy.github.io/nexpy/

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Work supported by the U.S. Department of Energy,
Office of Science, Materials Sciences and Engineering Division.

BASIS OF NeXus Data Format

Semantic HDF5 files

- NeXus files are HDF5 files with the addition of some semantics.
  - Simple design rules to make the files easy to navigate.
  - A list of definitions that cover most experimental metadata.
- The purpose of these rules is to make them self-describing.
  - It is usually possible to understand their contents without referring to any documentation.
- NeXus uses a hierarchical design similar to a file system.
  - Hierarchy allows complex data to be stored in a readily accessible form.
    - Important data at a high level
    - Arcane details at a low level
- Base classes provide a glossary of terms required for most experiments.

STATUS OF NeXus

https://www.nexusformat.org

- The NeXus data format is now well established as an international standard for the storage of data at neutron and synchrotron x-ray facilities.
- It is the official archive format at a number of facilities.
  - Both spallation neutron sources (*e.g.*, SNS/ISIS) and synchrotron sources (*e.g.*, Diamond/ESRF).
  - It is also used by the μSR and, more recently, electron microscopy communities.
- There is active participation in the NeXus International Advisory Committee by nearly 20 facilities in Asia, Europe, and North America.
  - Official NIAC meetings take place every two years with code camps nearly every year.
  - Monthly online meetings (even before the pandemic) deal with maintenance issues.
- Dectris has worked with NIAC to adopt NeXus for detector storage.

NeXus INTERNATIONAL ADVISORY COMMITTEE

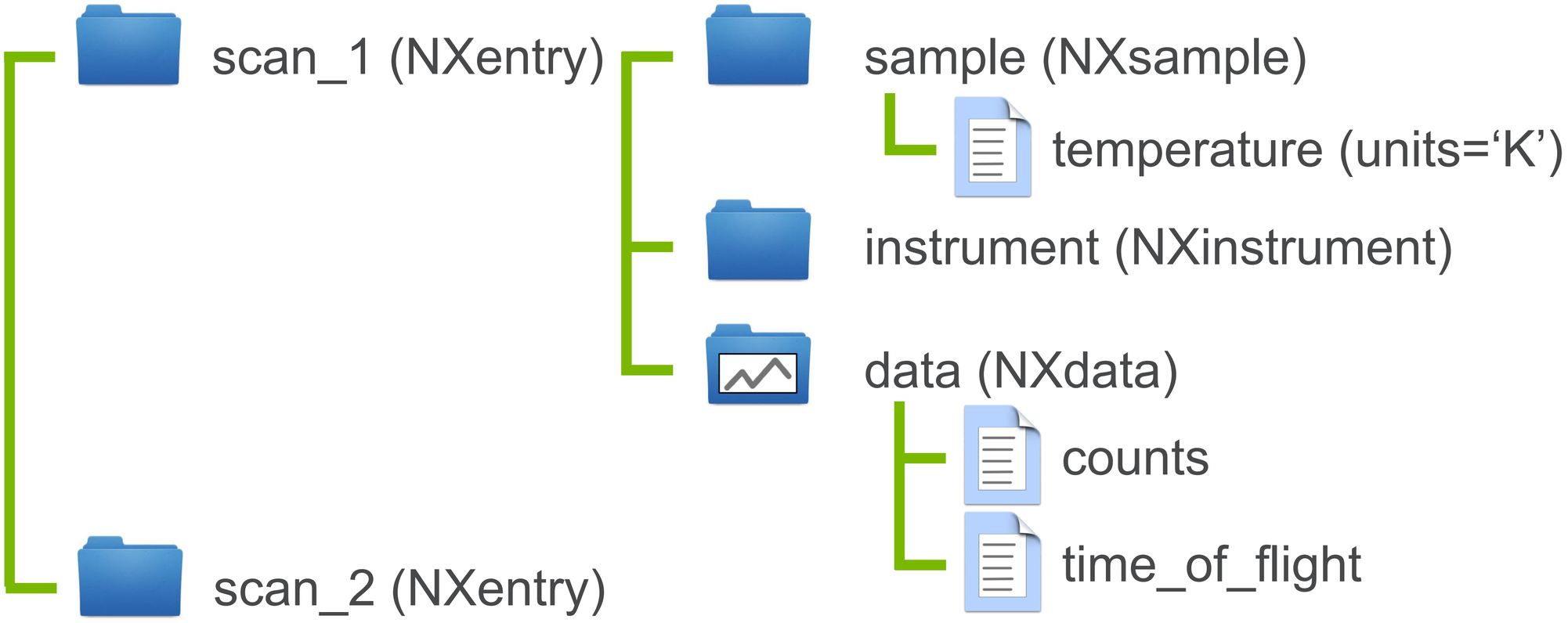
Chair: Aaron Brewster (Lawrence Berkeley Laboratory)

- Advanced Light Source, USA
- Advanced Photon Source, USA
- Bragg Institute, Australia
- Canadian Light Source, Canada
- Diamond/ISIS, UK
- European Synchrotron Radiation Facility, France
- European XFEL, Germany
- Extreme Light Infrastructure, Eastern Europe
- Helmholtz Zentrum Berlin, Germany
- J-PARC, Japan
- Los Alamos National Laboratory, USA
- NSLS-II, USA
- Spallation Neutron Source/HFIR, USA
- Spring8, Japan
- Swiss Light Source/SINQ, Switzerland
- Synchrotron Soleil, France

DESIGN RULES OF NeXus Data Format

Design Rules

- NeXus files contain three types of object.
  - Groups
  - Fields
  - Attributes



USING NeXus DATA IN PYTHON SHELLS

NeXus Format Python API

NeXpy uses the nexusformat package to read, manipulate, and write NeXus files.

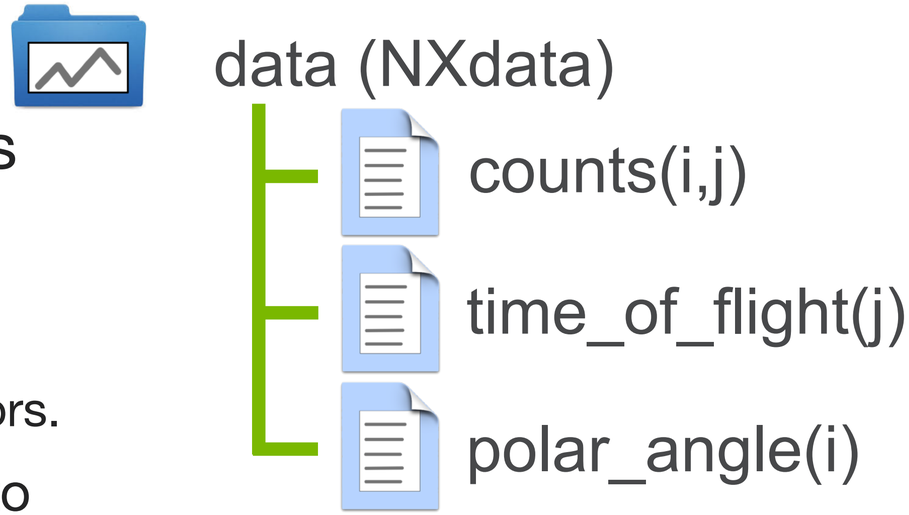
```
$ pip install nexusformat
$ conda install -c conda-forge nexusformat
```

It has the following features:

- Lazy loading of existing NeXus files (using h5py), creating a (kind of) memory map of the entire file.
  - ```
    >>> psycco = nxload('pycco_120K.nxs', 'rw')
    >>> print(pycco['entry/sample/temperature'])
    120.0
    ```
- Mapping of all NeXus objects (groups, fields, and attributes) into Python objects.
  - ```
    >>> sample = NXsample(temperature=120.0)
    >>> sample['temperature'].units='K'
    ```
- Intuitive creation of standard-conforming NeXus structures.
  - `>>> data = NXdata(z,(y,x))`
  - `>>> data.plot()`
- Normalization of data written using different conventions.
  - *e.g.*, variable-length Unicode strings *vs* size-1 fixed-length byte arrays

ENCAPSULATING PLOTTABLE DATA

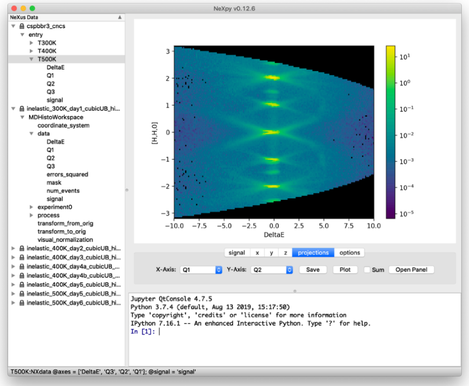
NXdata Groups



- NXdata groups are a key component of HDF5 files that could be useful to other types of data.
- It encapsulates everything needed for a plot.
  - *i.e.*, the signal and axes. (*cf* dimension scales), weights, and errors.
- In NeXpy, NXdata groups can be indexed and manipulated to generate new NXdata groups.
  - *e.g.*, `data[:,10:20], 2*data[:,5]`
- For 1D data, means, standard deviations and moments can also be calculated.
  - `data.mean(), data.std()`
- And, of course, NXdata groups can be plotted.
  - `data.plot()`

PURPOSE OF NeXpy

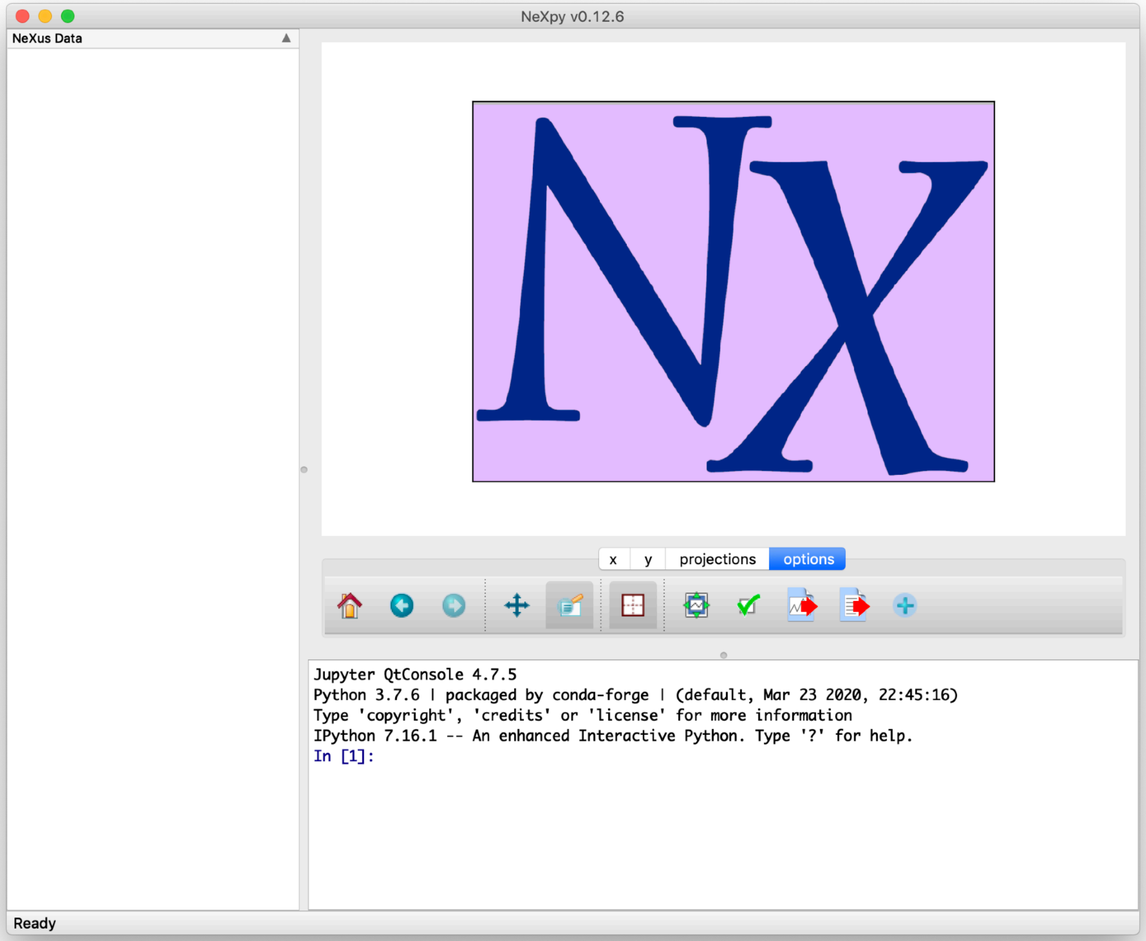
Restoring a scientist's control over their own data



- NeXpy is a GUI toolbox for analyzing and visualizing data stored in HDF5 files.
- Its original purpose was to handle neutron and x-ray scattering data stored in the NeXus format. However, it will open any HDF5 file and many of its features can be applied to any kind of data.
- The overarching goal is to make it easy to 'play' with the data.
  - Easy to inspect, visualize, manipulate, and fit the data.
  - Easy to compare data from multiple experiments and techniques.
  - Easy to develop new algorithms and modes of analysis.

INSTALLING NeXpy

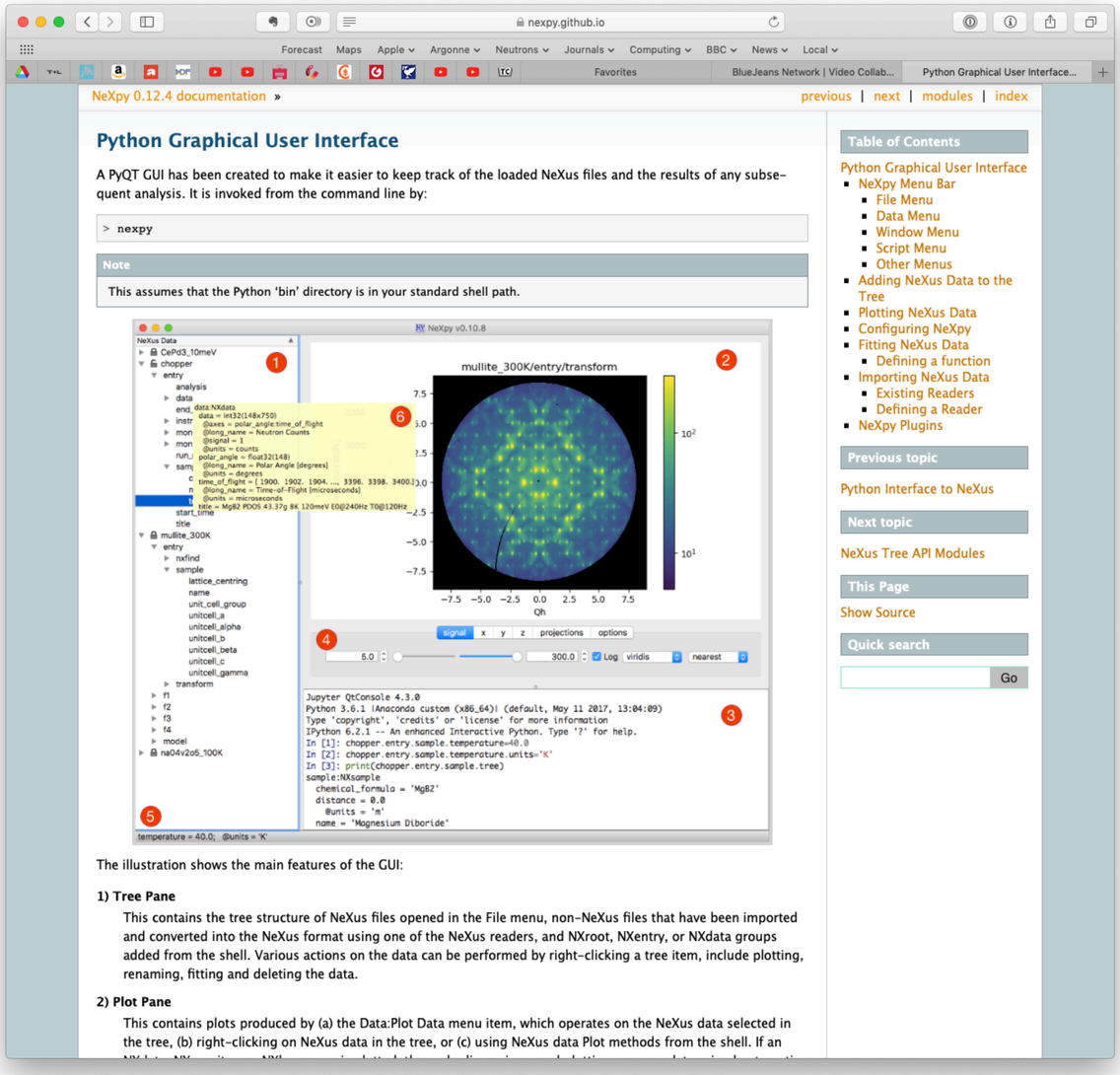
- NeXpy is a pure Python package.
- There are multiple ways to install it.
  - conda install -c conda-forge nexpy
  - pip install nexpy
  - git clone https://github.com/nexpy/nexpy.git
- Dependencies:
  - PyQt (PyQt5, PyQt6, PySide2, or PySide6)
  - IPython
  - Matplotlib
  - h5py
  - nexusformat



INSTALLING NeXpy

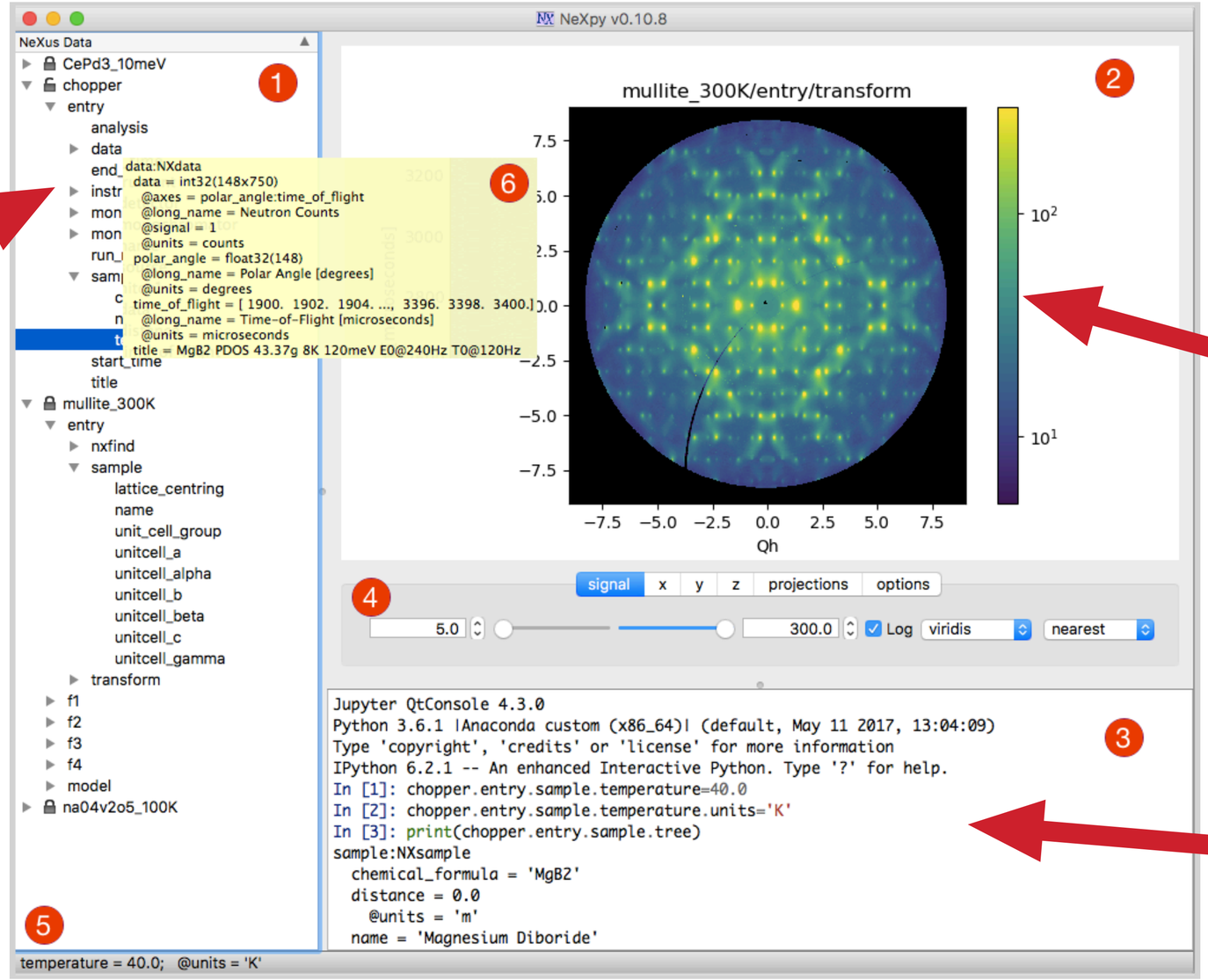
https://nexpy.github.io/nexpy/

- NeXpy is a pure Python package.
- There are multiple ways to install it.
  - conda install -c conda-forge nexpy
  - pip install nexpy
  - git clone https://github.com/nexpy/nexpy.git
- Dependencies:
  - PyQt (PyQt5, PyQt6, PySide2, or PySide6)
  - IPython
  - Matplotlib
  - h5py
  - nexusformat



ANATOMY OF NeXpy

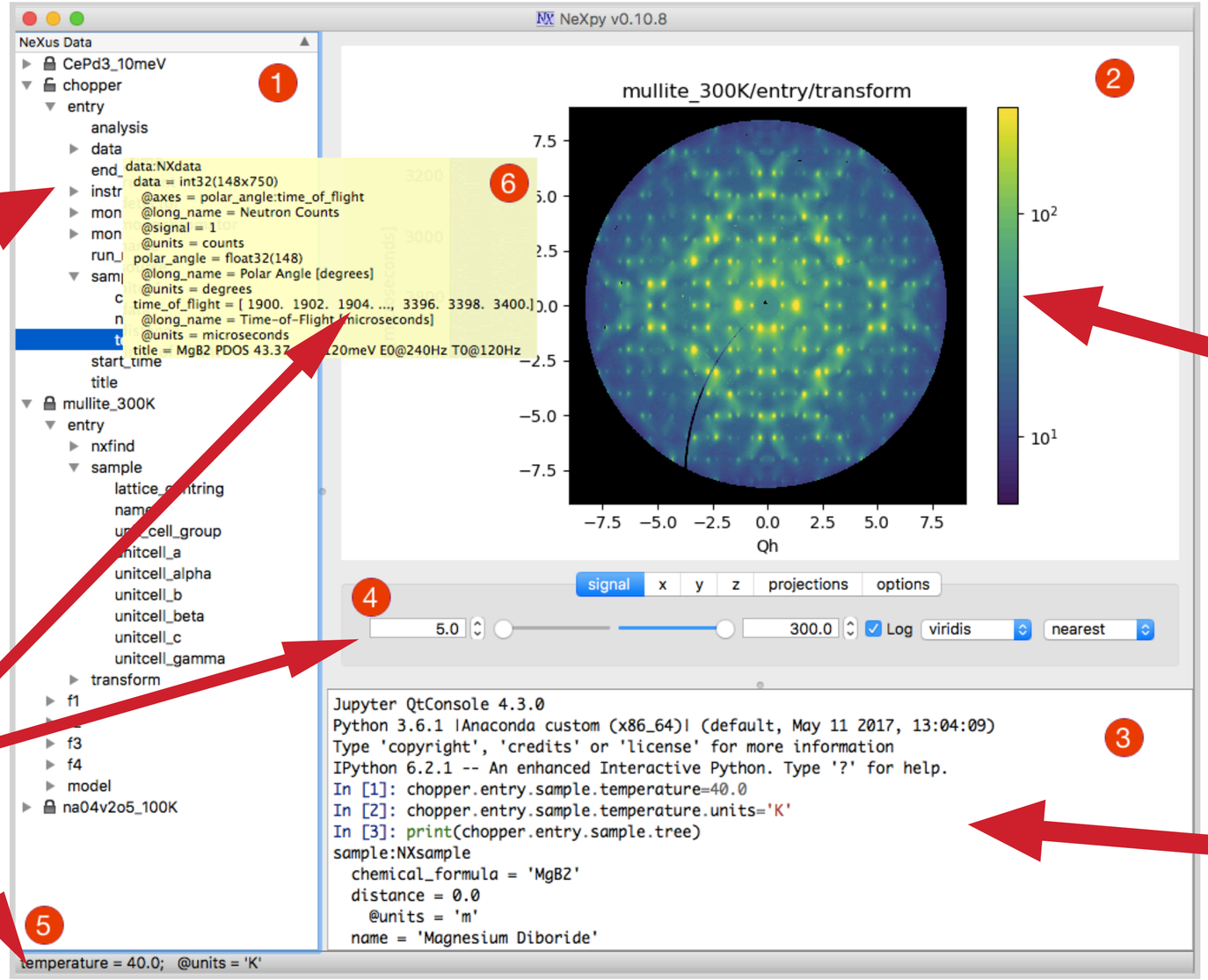
1) Tree Pane



2) Plot Pane

3) Shell Pane

ANATOMY OF NeXpy



1) Tree Pane

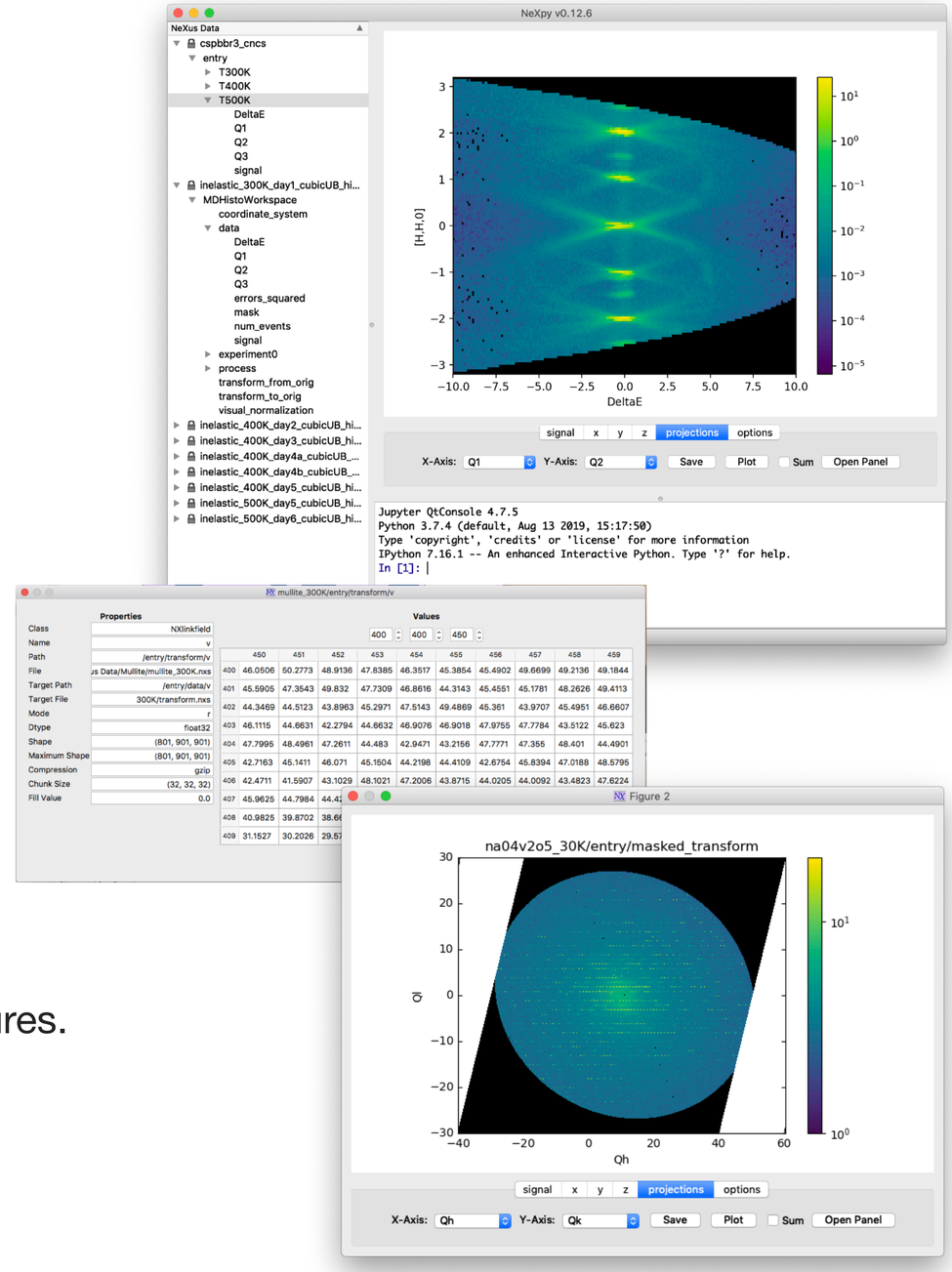
2) Plot Pane

3) Shell Pane

4) Axis Panels
5) Status Bar
6) Tooltips

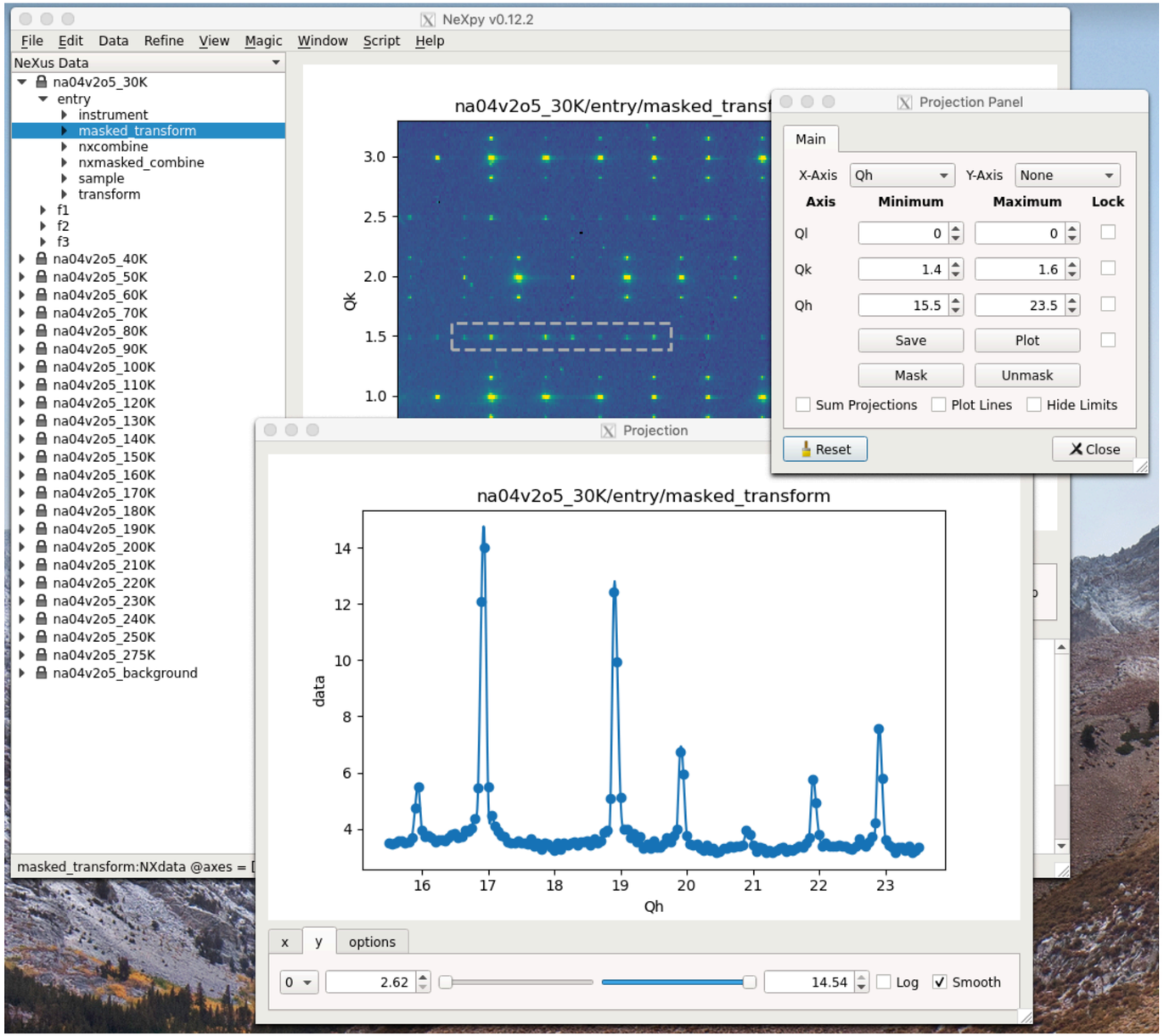
FEATURES OF THE NeXpy GUI

- NeXus data can be directly loaded into the tree using an Open File dialog or imported from other formats,.
  - *e.g.,* SPEC using *spec2nexus* (https://spec2nexus.readthedocs.io) or CBF using *Fabio* (https://fabio.readthedocs.io).
- The GUI allows data to be viewed and manipulated:
  - *e.g.*, plotted, viewed in a table, created, deleted, renamed, copied, and pasted.
- New NeXus data can be created, copied, and saved to a file.
- All groups and fields in the tree are accessible from the command line of the IPython shell, with all changes updated in the tree.
- Panels facilitate comparisons of data from multiple files.
  - Projection, Limits, Scan, and Fit Panels
- Specialized functionality can be implemented using a plugin architecture.
- As a bonus, NeXpy provides convenient GUI access to special Matplotlib features.
  - Skewed axes
  - Symmetric color plots
  - Smoothing in 1D and 2D
  - Reordering legends



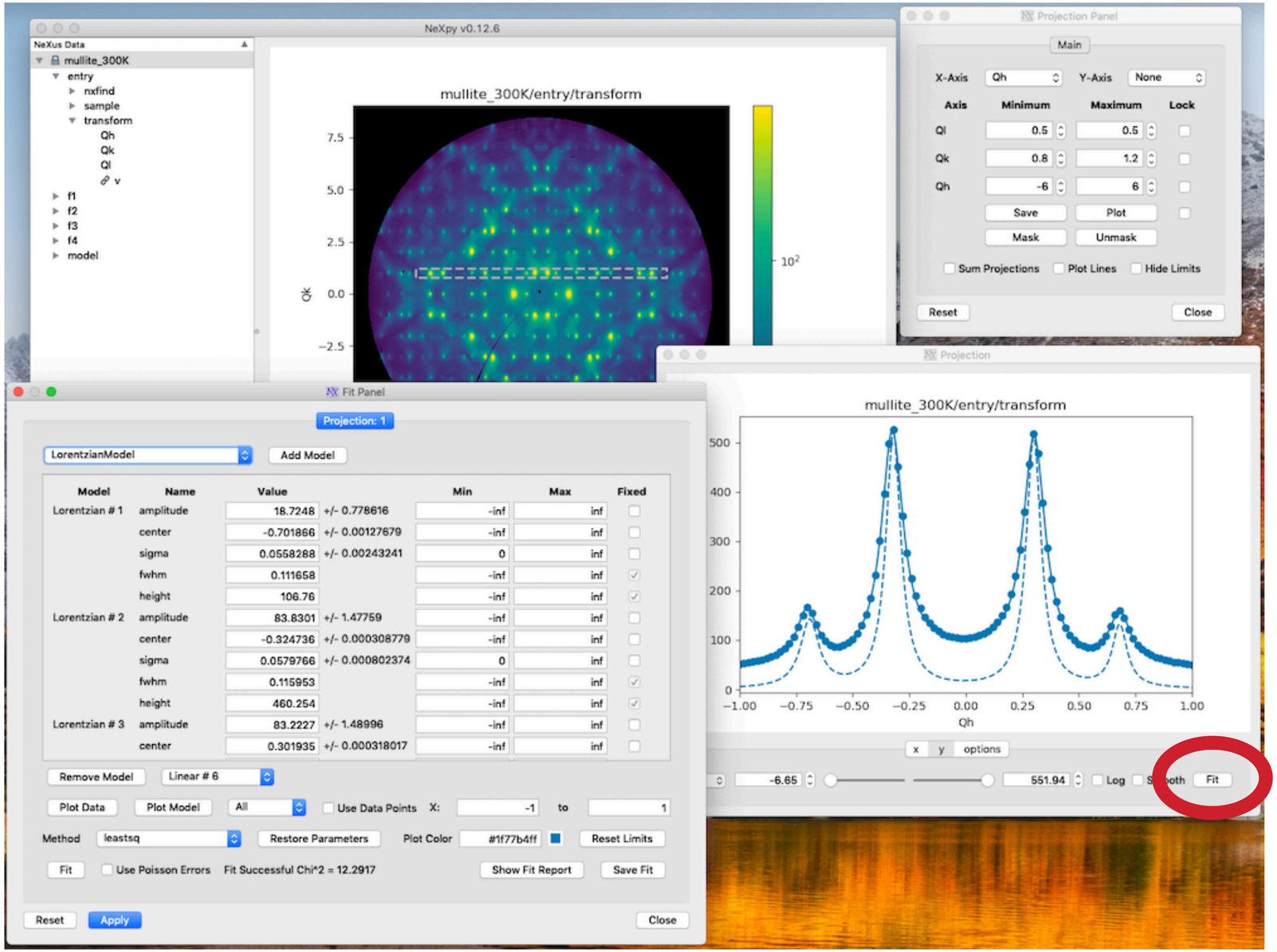
PROJECTION PANEL

- NeXpy makes it easy to plot arbitrary 1D and 2D projections through multidimensional data.
- 1D projections from different data slices can be over-plotted.
- The resulting plots can be saved, exported, or, for 1D data, fitted.
  - Using the ‘lmfit’ package (see later).



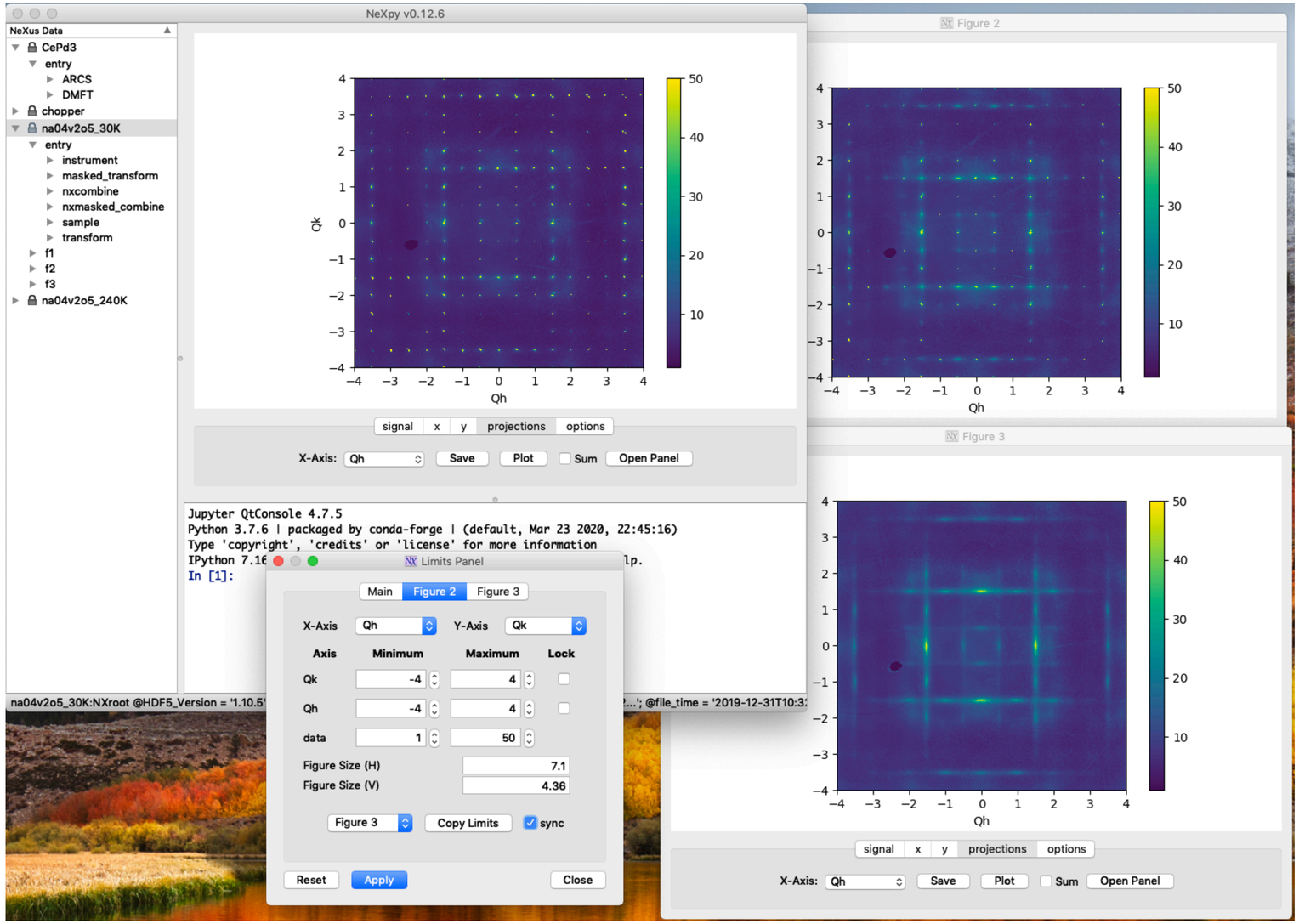
FIT PANEL

- NeXpy provides a GUI interface to the least-squares fitting package, ‘lmfit’.
- A ‘Fit’ button on every 1D plot invokes the Fit Panel.
- The ‘lmfit’ package has support for a wide range of lineshapes.
  - Gaussian, Lorentzian, DHO, pseudo-Voigt, LogNormal,…
  - It is easy to define your own.
- Fit results can be saved to a NeXus group or file.



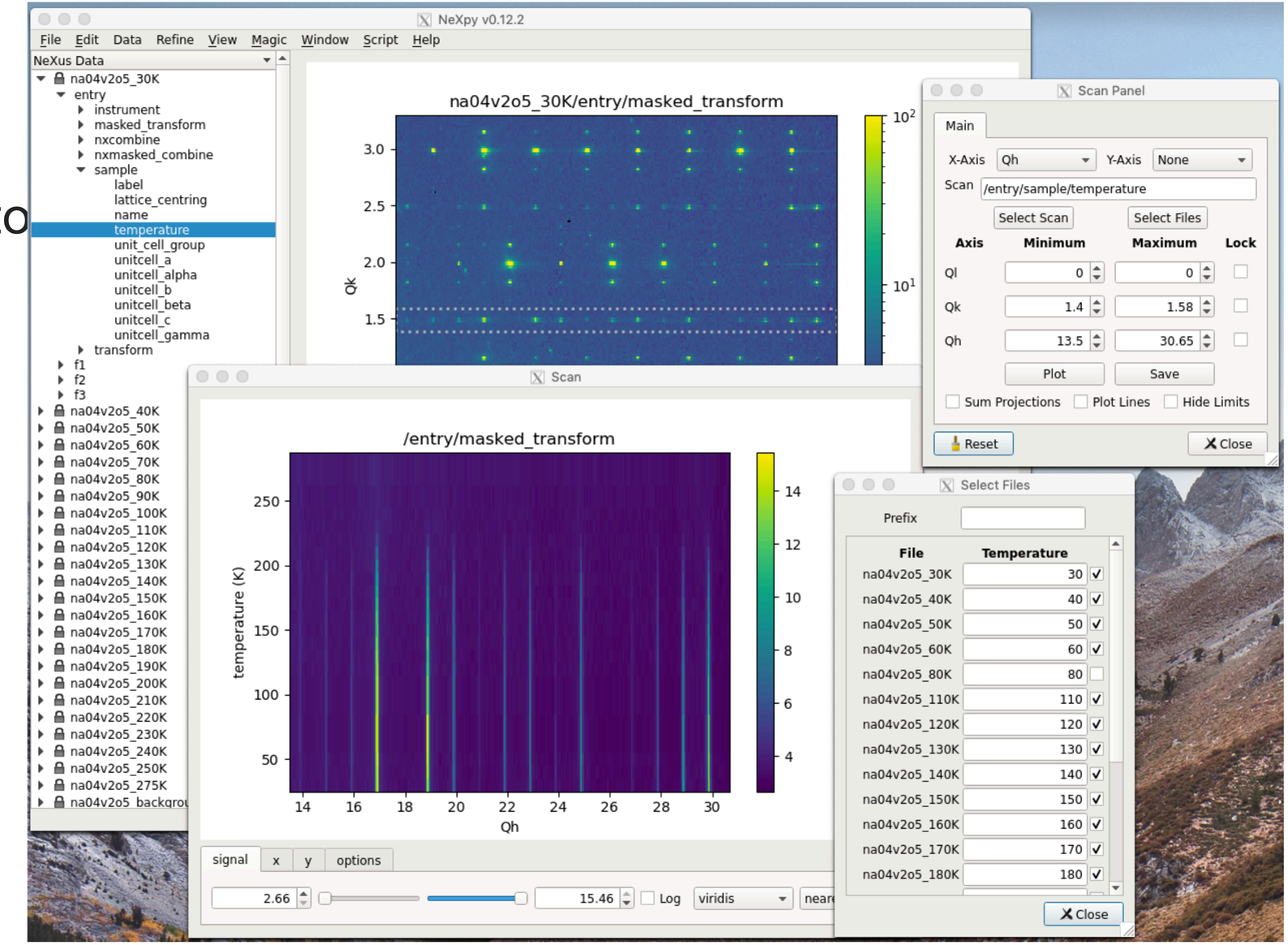
LIMITS PANEL

- The Limits Panel allows multiple plots to be synchronized automatically.
  - Changes to the parent plot are immediately propagated to the synchronized plots.
  - This includes the plotting axes and their limits, as well as other plotting options:
    - Log axes
    - Color maps
    - Aspect ratios
    - Skew angles
    - Smoothing



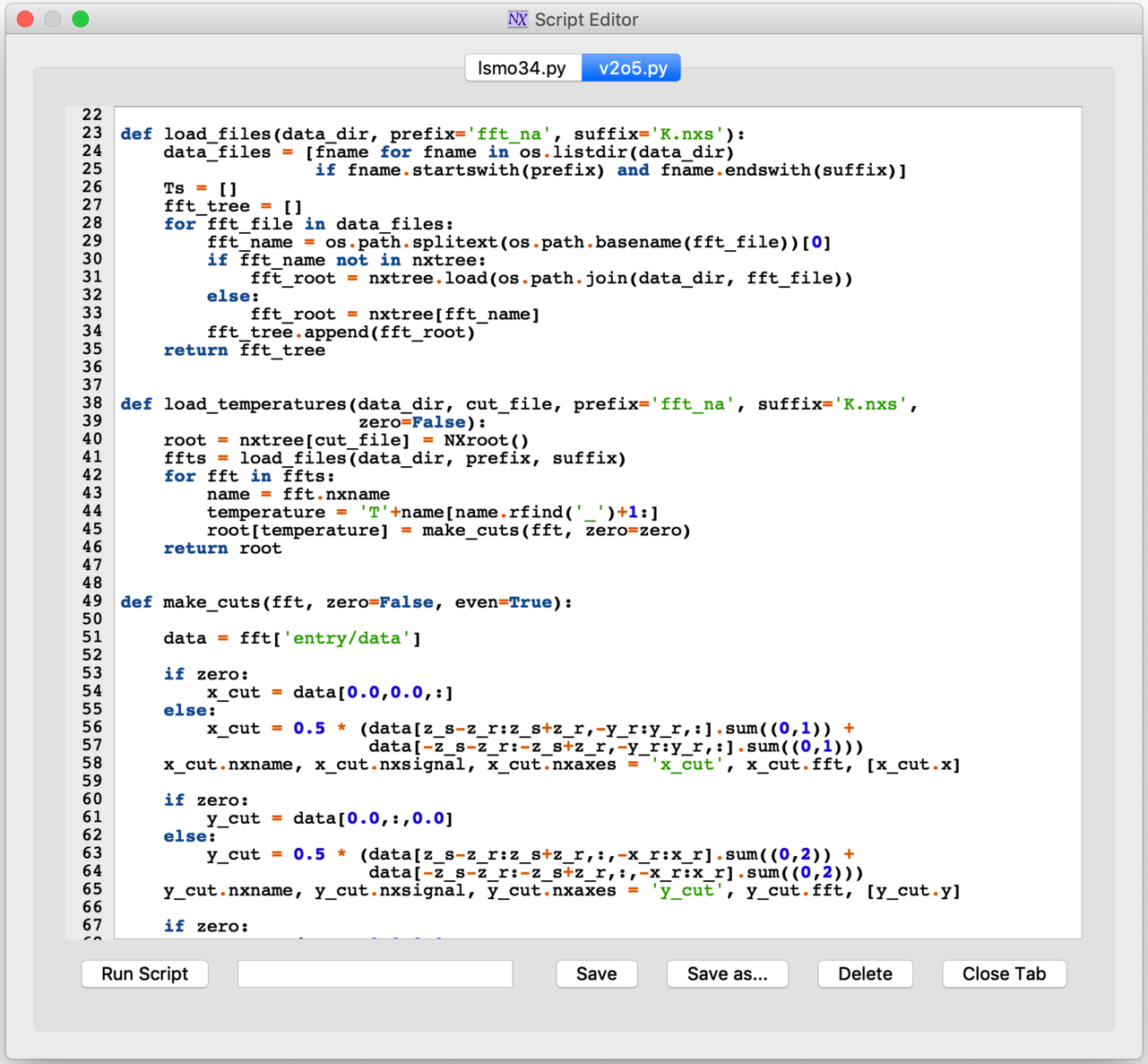
SCAN PANEL

- One goal is to make it easy to combine data from multiple files.
- A Scan Panel allows data to be plotted against a parametric variable that changes from file to file.
  - *e.g.*, temperature
- This uses HDF5 virtual datasets to expand the dimensionality without any increase in storage.



SCRIPT EDITOR

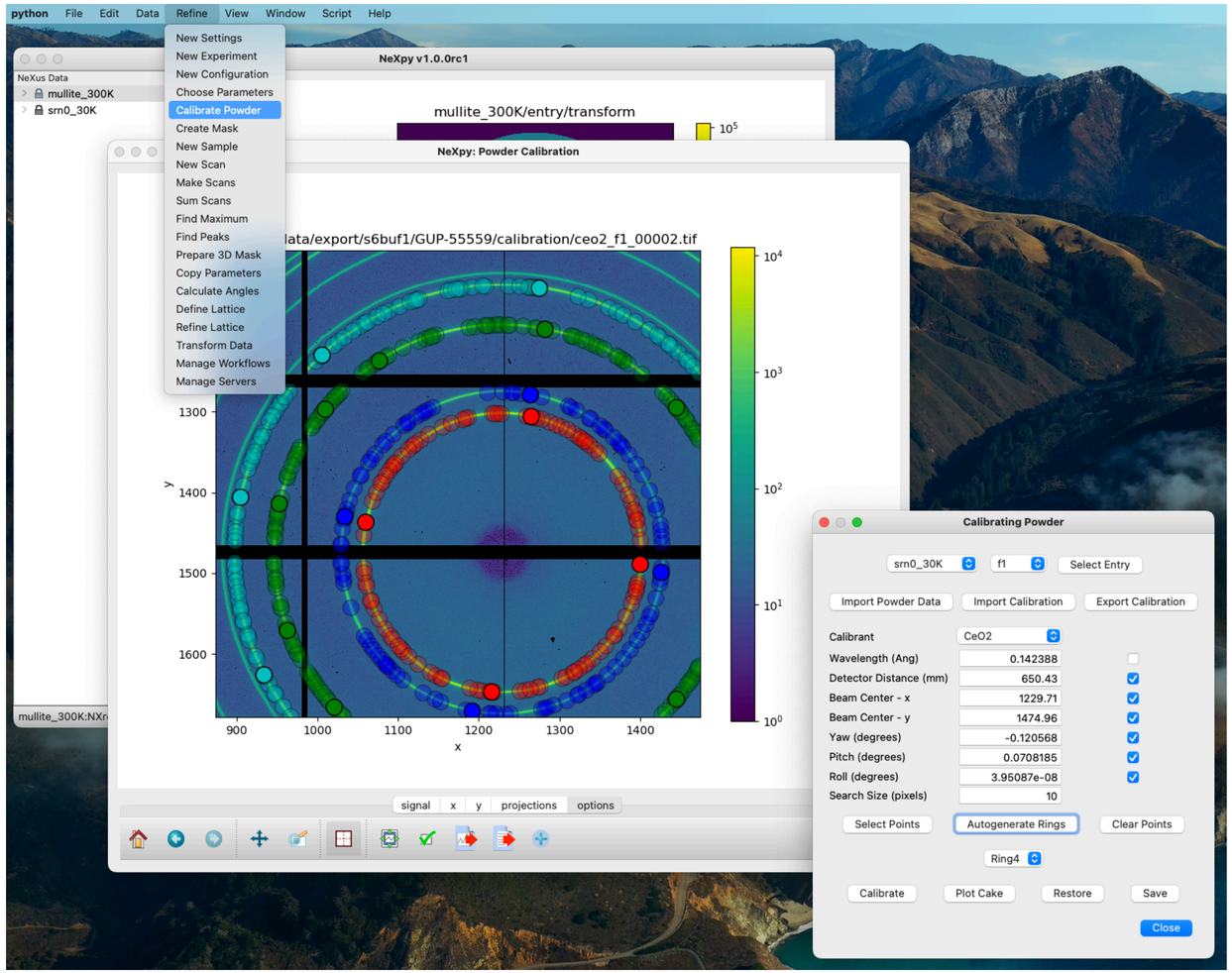
- NeXpy has a built-in editor for developing Python scripts.
- The code can be run immediately within the IPython shell.
  - For performing repetitive operations.
  - For developing complex algorithms.
- The script editor can be used to prototype new modes of data analysis.
  - *e.g.*, 3D-ΔPDF



EXTENDING NeXpy

Plugin Architecture

- Additional menu items can be added to extend NeXpy functionality for specialist applications.
- A simplified widget library allows sophisticated GUIs to be developed without expert knowledge of PyQt.
- The screenshot shows one part of a complete workflow for single crystal diffuse scattering implemented as NeXpy plugins.



Acknowledgement

ESRF The European Synchrotron

pyFAI
Fabio

SUMMARY

https://nexpy.github.io/nexpy/

- NeXpy provides a simple GUI and scripting interface to allow scientists to ‘play’ with their data.
- A number of features facilitate analyses that encompass multiple data files.
- A script editor allows new modes of data analysis to be prototyped.
- A plugin architecture allows the NeXpy GUI to provide a framework for any specialized applications with minimal knowledge of PyQt.
- There is extensive online help.
  - Installation instructions and descriptions of both the ‘nexusformat’ API and the NeXpy GUI.
  - Jupyter notebook to introduce the main concepts of the NeXus format and the Python API.