

UCLA HEDP Experimental Plasma Physics Analysis Package Documentation

May 28, 2020

Contents

1	Overview	3
2	The Common Data Format (CDF)	3
2.1	Requirements	3
2.2	Conventions	4
2.3	Considerations for Large Datasets	4
3	Structure of the Package	6
3.1	tools	6
3.1.1	csv.py	6
3.1.2	dataset.py	6
3.1.3	hdf.py	6
3.1.4	math.py	6
3.1.5	pos.py	6
3.1.6	util.py	6
3.2	load	6
3.2.1	lapd.py	6
3.2.2	hrr.py	6
3.2.3	imgdir.py	6
3.2.4	ascii.py	6
3.3	process	6
3.3.1	bdot.py	6
3.3.2	langmuir.py	6
3.3.3	tdiode.py	6
3.3.4	interferometer.py	6
3.3.5	imgseq.py	6
3.3.6	scope.py	6
3.3.7	process.py	6
3.4	analyze	6
3.4.1	filter.py	6
3.4.2	polarization.py	6
3.4.3	spectral.spectrogram.py	6
3.4.4	thomson.thomson_scattering.py	6
3.5	user	6
3.6	dataview	6

1 Overview

2 The Common Data Format (CDF)

2.1 Requirements

The UCLA HEDP package is designed around a standardized structured data format, hereafter referred to as common data file or CDF. A CDF object is an HDF file that conforms to a standard structure as defined in the function `tools.dataset.validDataset`. In particular, each CDF file must contain

- A dataset called "data" with n dimensions. This array must have an attribute "dimensions" which holds a string array of list n which provides a name for each dimension in the dataset (by convention, all lowercase). A second attribute "unit" defines the units of the data in this array as a string that can be interpreted by the `astropy.units` string-to-unit parser.

In the future this could be generalized so that the CDF could contain multiple data arrays with different names (but all with the same shape and the same axes). This would be conducive to datasets such as density and electron temperature planes from Langmuir measurements which share a set of axes and belong conceptually together. In this case, a new attribute array should be added to the parent group listing the names of all of the data arrays.

- For each entry in the "dimensions" attribute there must exist a dataset with the same name which contains the axis values for that dimension. This dataset must also include a mandatory "unit" attribute as defined above.

This format was designed to contain all of the information necessary to make plots of the dataset.

2.2 Conventions

2.3 Considerations for Large Datasets

Large datasets are not stored on consecutive sections of computer memory but are rather broken into "chunks" that are stored separately. Accessing data spread across multiple chunks is much more computationally costly than accessing the same amount of data from a single chunk. The HDF format and the **h5py** package allow control over how a dataset is broken into chunks in order to maximize efficiency. Where possible, the UCLA HEDP package attempts to chunk datasets in a way that minimizes read times for common operations. For example, a dataset containing time arrays at a variety of spatial locations may chose to chunk the time axis together, anticipating

Defined in

3 Structure of the Package

3.1 tools

3.1.1 csv.py

3.1.2 dataset.py

3.1.3 hdf.py

3.1.4 math.py

3.1.5 pos.py

3.1.6 util.py

3.2 load

3.2.1 lapd.py

3.2.2 hrr.py

3.2.3 imgdir.py

3.2.4 ascii.py

3.3 process

3.3.1 bdot.py

3.3.2 langmuir.py

3.3.3 tdiode.py

3.3.4 interferometer.py

3.3.5 imgseq.py

3.3.6 scope.py

3.3.7 process.py

3.4 analyze

3.4.1 filter.py

3.4.2 polarization.py

3.4.3 spectral.spectrogram.py

3.4.4 thomson.thomson_scattering.py

3.5 user

3.6 dataview