BREAKTHROUGH LISTEN

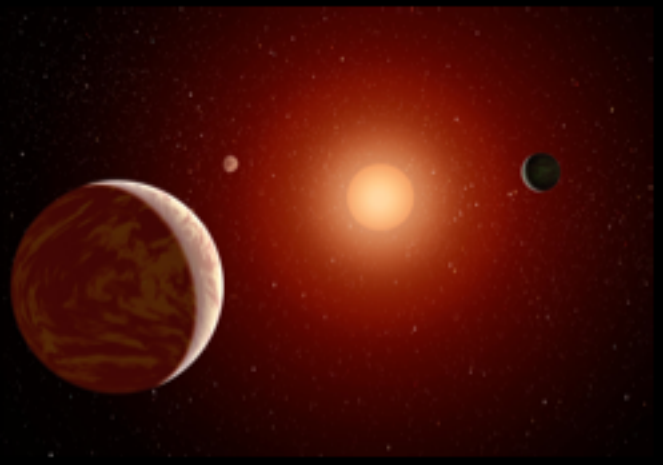
DANNY PRICE
UNIVERSITY OF CALIFORNIA BERKELEY
HDF WORKSHOP, 19 JUNE 2016

BERKELEY SETI
RESEARCH CENTER

Breakthrough Listen is the largest ever scientific research program aimed at finding evidence of civilizations beyond Earth.

THE BREAKTHROUGH LISTEN INITIATIVE:

OVERVIEW



1 Million Stars

Milky Way Galactic Plane Survey

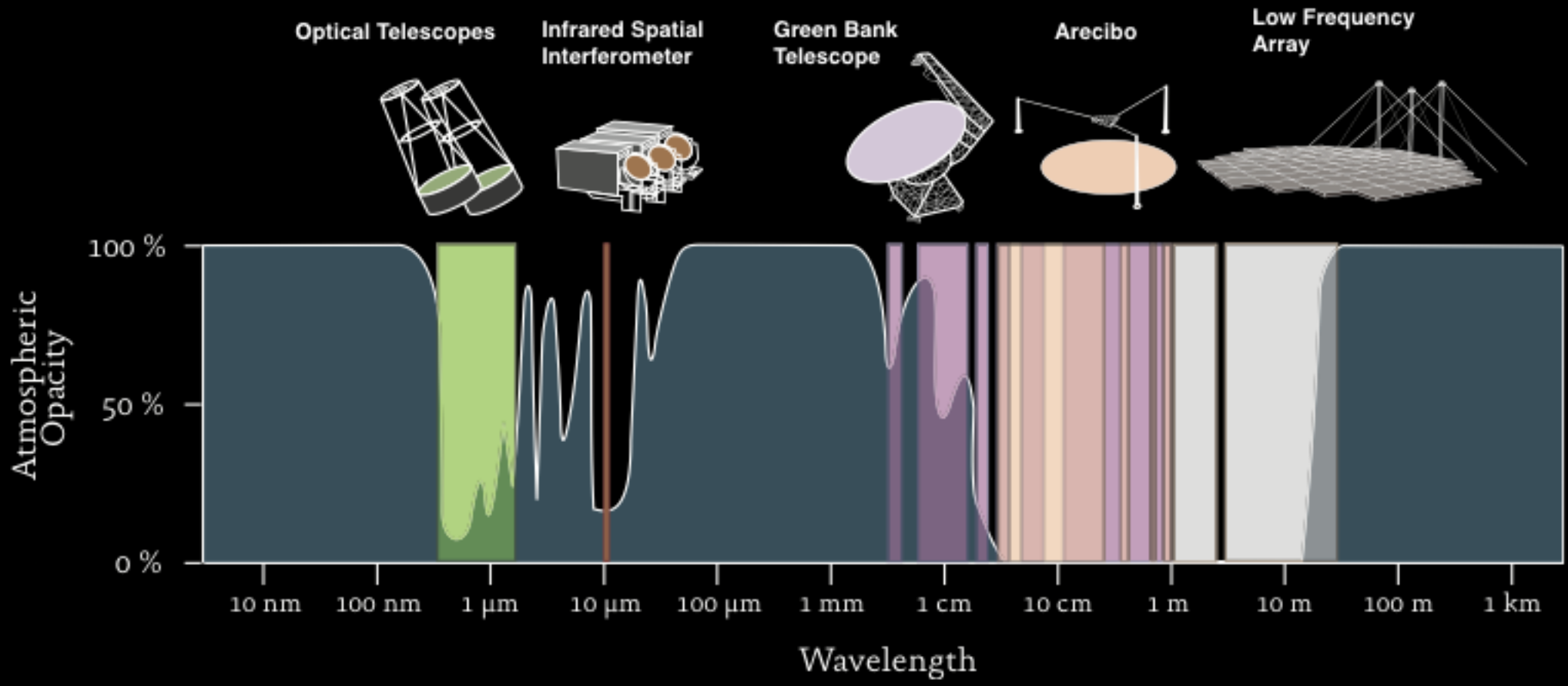
100 Galaxies

1 day of Breakthrough Listen = 1 year of any previous search

Many PB of scientific data to be made publicly available:
BL will be one of the largest public astronomy datasets

SETI.BERKELEY.EDU | BREAKTHROUGHINITIATIVES.ORG

Searching Across the Electromagnetic Spectrum



Using multiple telescopes, we can search across the electromagnetic spectrum for indicators of advanced technology.

The Breakthrough Listen Initiative: Telescopes

Automated Planet Finder (Lick Observatory)

- Search for extremely narrow emission lines from artificial lasers
- Extremely high resolution "Levy Spectrometer" 374 - 950 nm, $\lambda/\Delta\lambda = 10^5$

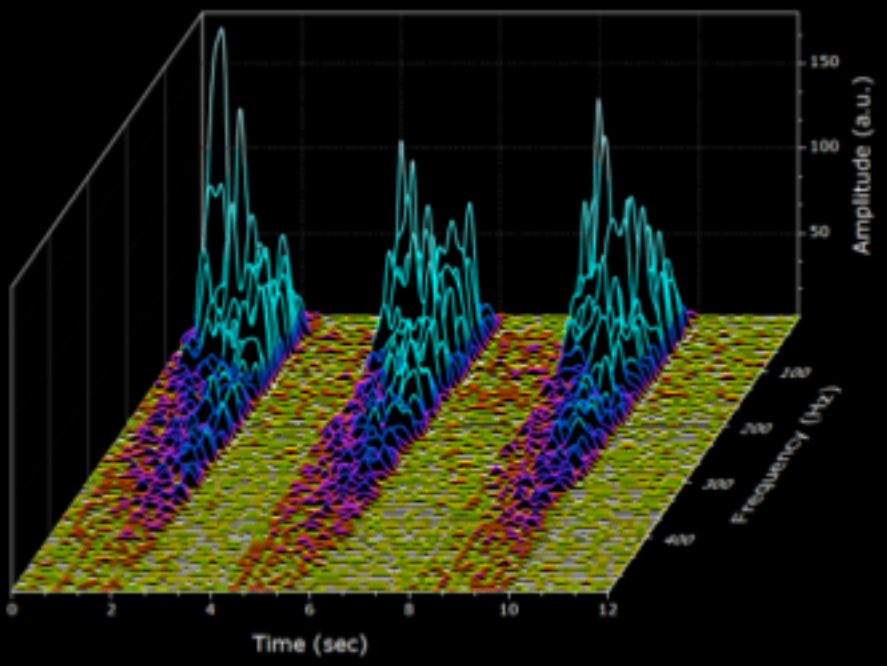
Green Bank Telescope (Green Bank, WV)

- Radio search focusing on targeted and raster observations
- Nearly continuous frequency coverage 300 MHz - 100 GHz
- Flexible IF system can deliver up to 10 GHz dual-pol analog bandwidth
- Extremely radio quiet (Federally protected radio quiet zone)

Parkes Telescope (NSW, Australia)

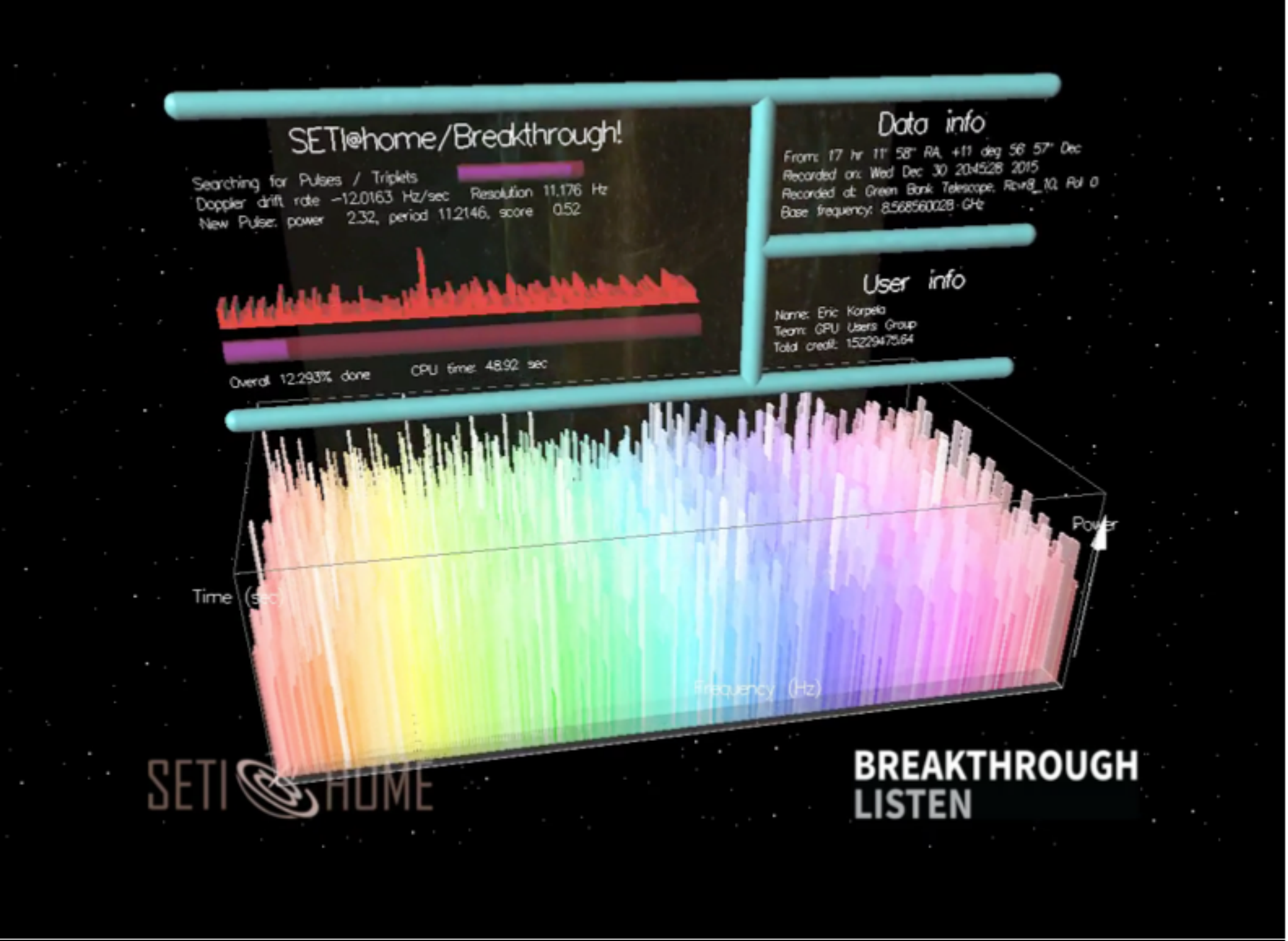
- Radio search focusing on surveys
- Southern hemisphere location gives great access to galactic plane
- Multi-beam receiver allows very efficient L-band (1.2 - 1.5 GHz) galactic plane surveys (Parkes Multibeam Receiver)
- Wide-band single-pixel and Phased Array Feed upgrades possible.

Data Overview



- Data are *N*-dimensional arrays of power spectra (PSD)
- Axes are (frequency, time) and each data point is power
- Telescope metadata also needs to be stored (where it's pointing!)
- Different resolutions in time and frequency are stored, and a subset of raw data is also stored.

image: originlab.com

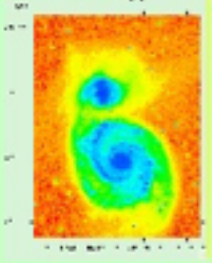

SETI@home/Breakthrough!
Searching for Pulses / Triplets
Doppler drift rate -12.0163 Hz/sec Resolution 11.176 Hz
New Pulse: power 2.32, period 11.2146, score 0.52
Overall 12.293% done
CPU time: 48.92 sec
Data info
From: 17 hr 11' 58" RA, +11 deg 56' 57" Dec
Recorded on: Wed Dec 30 20:45:28 2015
Recorded at: Green Bank Telescope, Rcvr8_10, Pol 0
Base frequency: 8.568560028 GHz
User info
Name: Eric Korpela
Team: GPU Users Group
Total credit: 152294756.64
Power
Time (sec)
Frequency (Hz)
SETI HOME
BREAKTHROUGH
LISTEN

Breakthrough Listen Data Rates: Green Bank

Raw Data	20 GB/sec
Reduced spectra	~660 GB/hr
Total on-site storage	2.2 PB

Need to make all data publicly available!


310
THE PROBLEM WITH FITS
©1979 ATARI INC

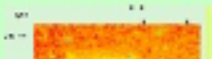


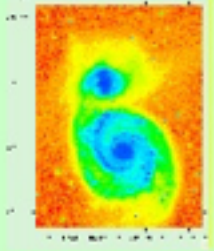
FITS

The Astronomical Image and Table Format

Flexible Image Transport System

HISTORY FITS was invented in 1979. At the time, there were limitations that
HISTORY seem a little archaic now. A good example is the FITS header:
HISTORY KEYWORDS are a maximum of 8 characters and are ALL CAPS.
HISTORY Each LINE must be shorter than 80 characters.
HISTORY UNICODE was invented a decade later so isn't supported.
HISTORY While here have been some minor additions over the years,
HISTORY such as the CONTINUE and HIERARCH keywords, these are
HISTORY pretty ugly solutions.





FITS

The Astronomical Image and Table Format

Flexible Image Transport System

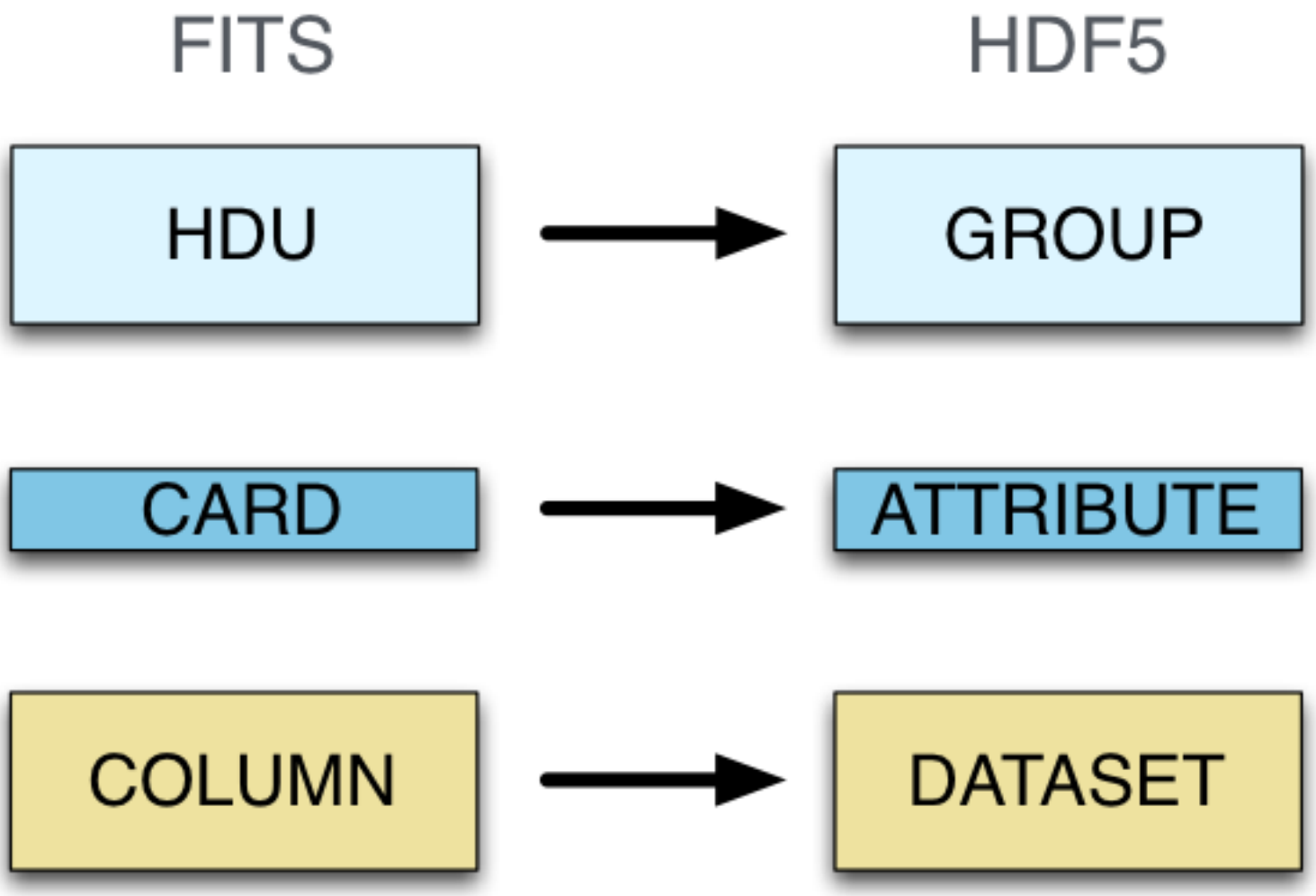
```
COMMENT   For the most part, FITS headers are irritating, not limiting.
COMMENT   A larger problem exists:
COMMENT   Datasets just keep getting bigger.
COMMENT   FITS is not well equipped to handle >1TB datasets.
COMMENT   Pretty damning summary of the issues in [1] and [2]
```

---

[1] B. Thomas et. al. "Significant Problems in FITS Limit Its Use in Modern Astronomical Research," ADASS XXIII, 2013

[2] B. Thomas and E. al, "The Future of Astronomical Data Formats I. Learning from FITS," Astronomy and Computing, pp. 1–14, Apr. 2014.

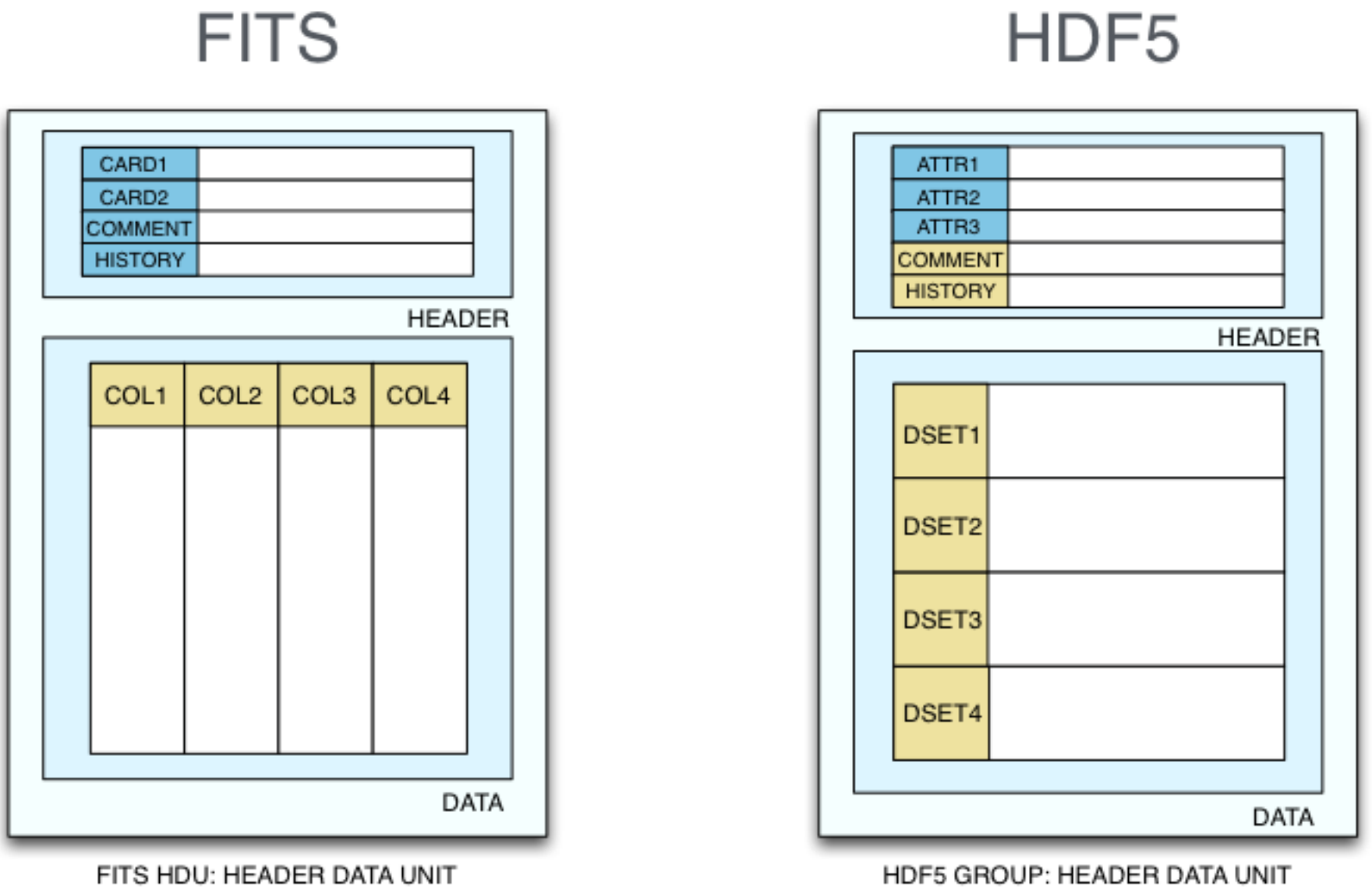
HDFITS: PORTING FITS DATA MODEL TO HDF5



github.com/telegraphic/fits2hdf

FITS2HDF

FITS2HDF



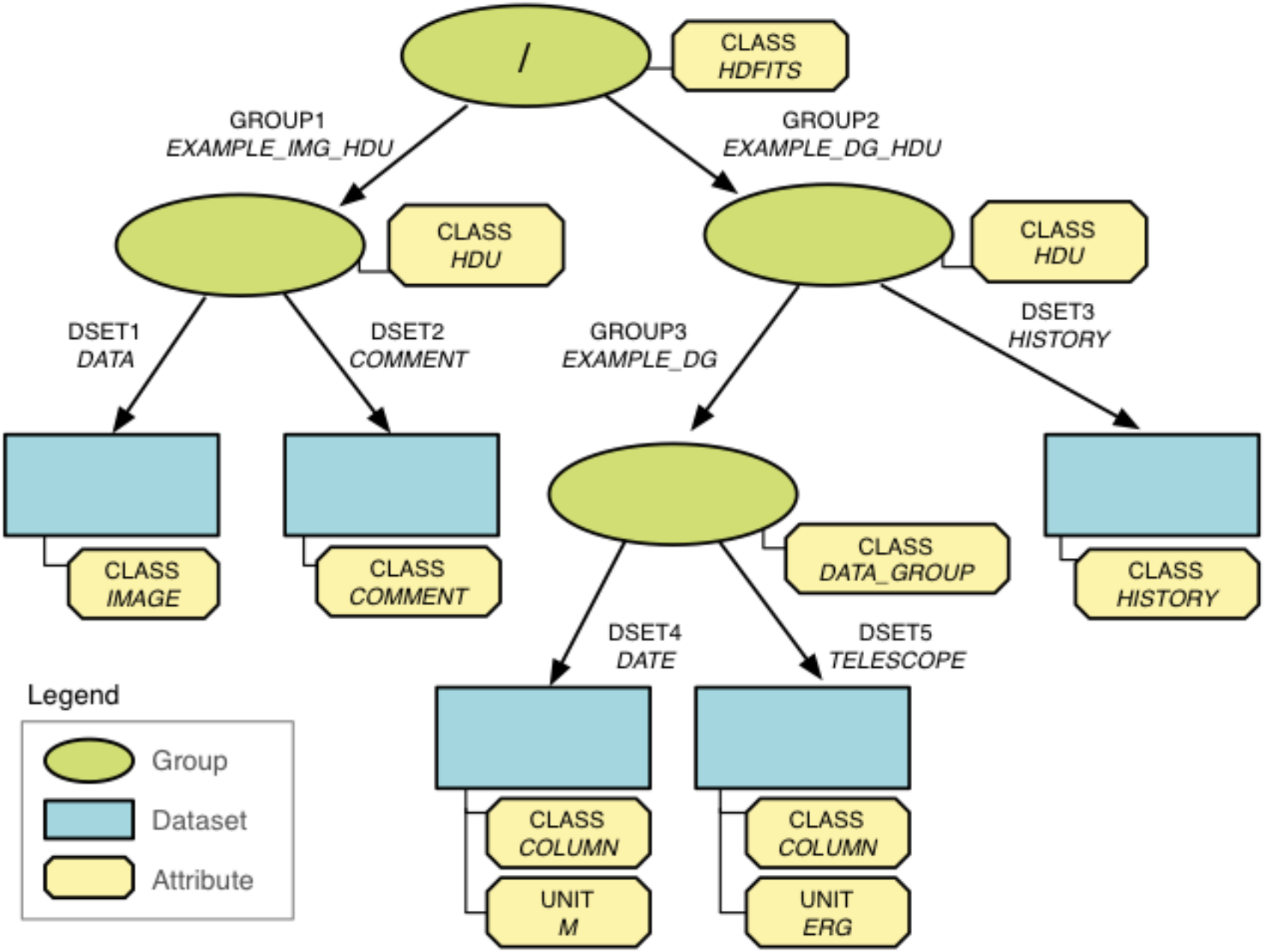
FITS HDU: HEADER DATA UNIT

HDF5 GROUP: HEADER DATA UNIT

github.com/telegraphic/fits2hdf

HDFITS: EXAMPLE TREE DIAGRAM

HDFITS: EXAMPLE TREE DIAGRAM



CONCLUDING REMARKS

Concluding Remarks

- Breakthrough Listen is the most comprehensive SETI experiment ever undertaken.
- With big searches comes big data challenges.
- HDF5 will be used instead of FITS!
- Open-source *fits2hdf* conversion utility available at github.com/telegraphic/fits2hdf
- Looking in to OpenDAP + Jupyter in the cloud to provide public data access.

SETI.BERKELEY.EDU | BREAKTHROUGHINITIATIVES.ORG

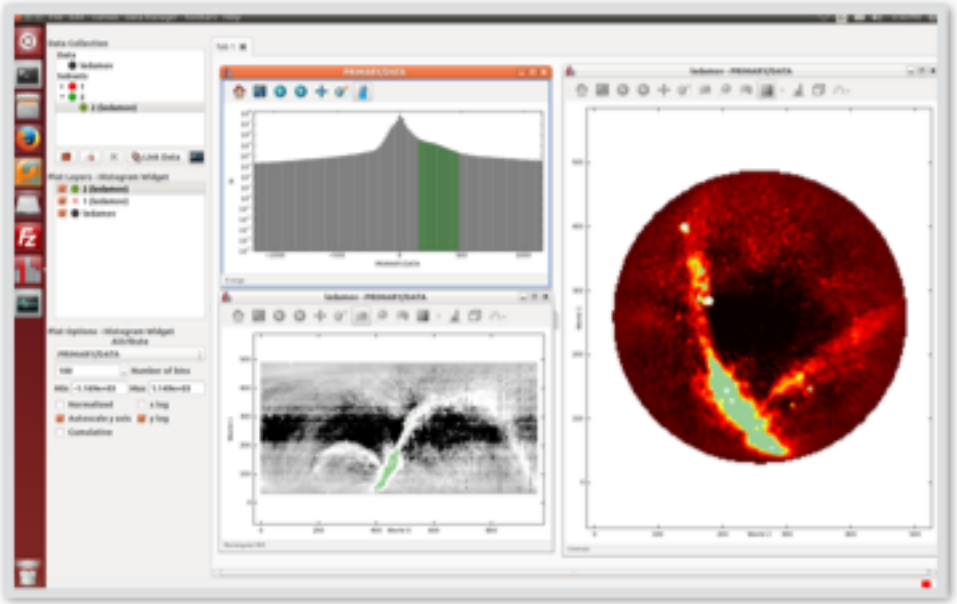
BACKUP SLIDES

BACKUP SLIDES

I/O Support: Image viewers

I/O support: image viewers

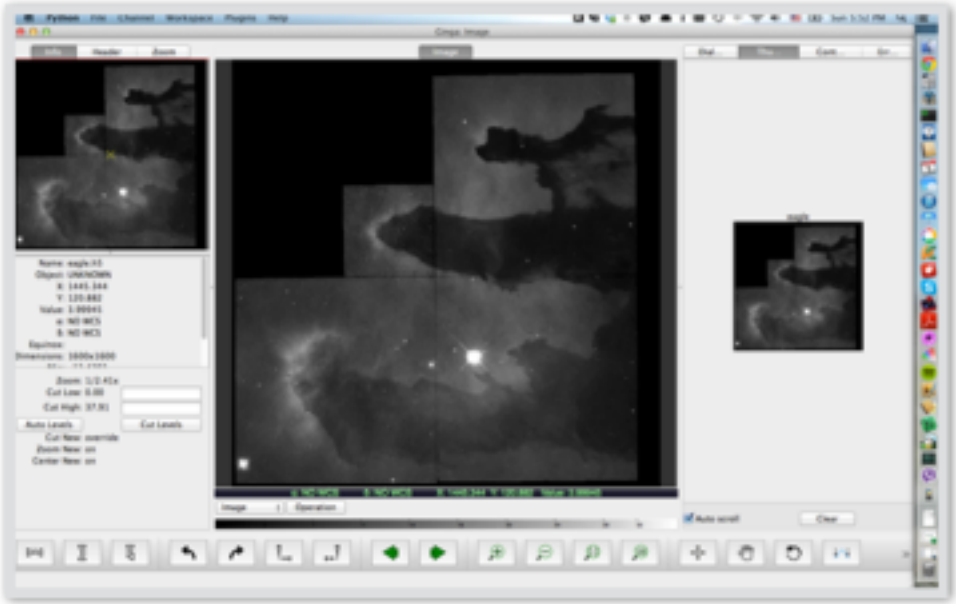
Proof of concept: reading HDFITS images into Python-based FITS viewers



GLUE

www.glueviz.org

2 lines of code changed in order
to read HDFITS (thanks to existing basic HDF5 support)



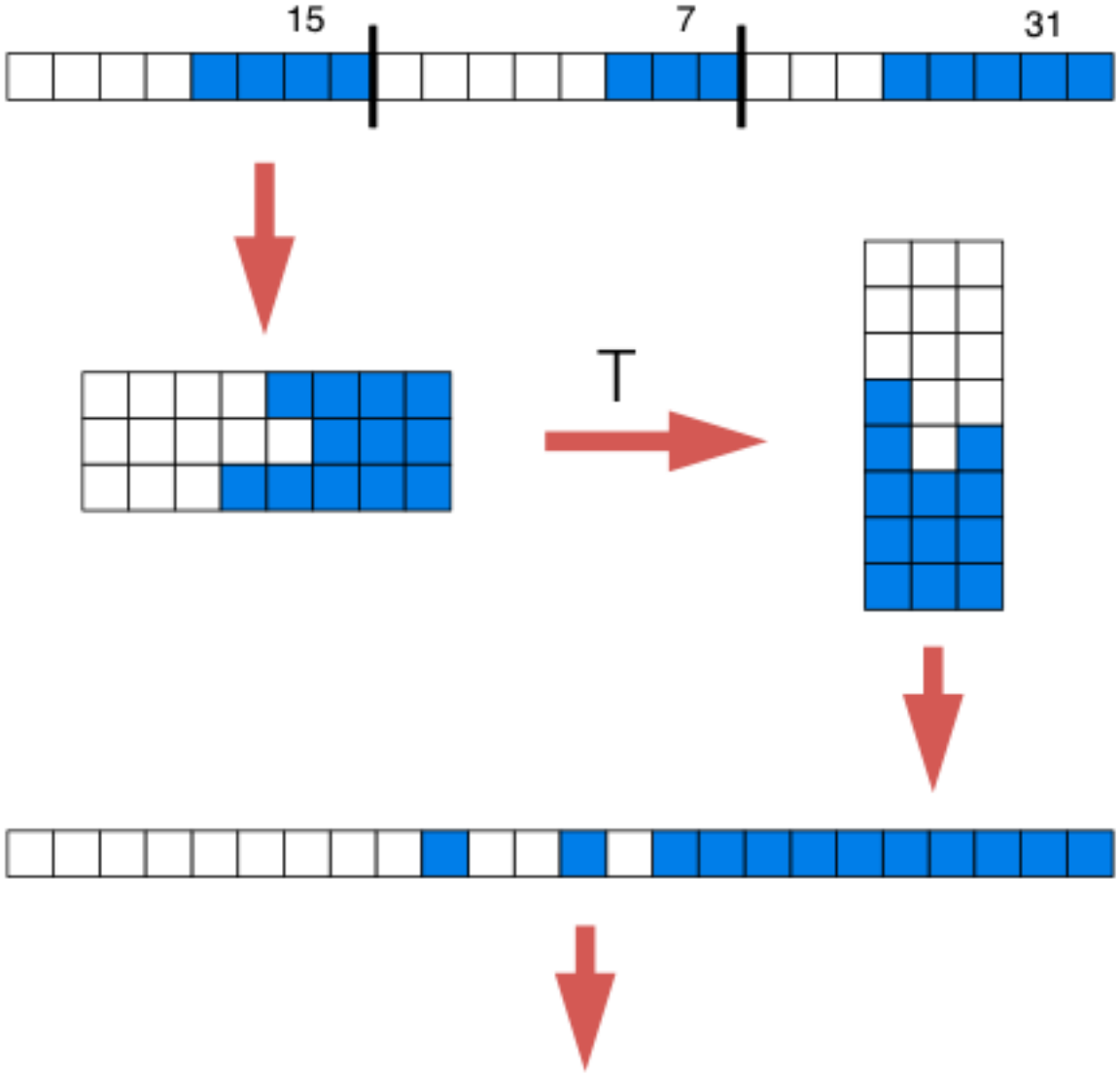
GINGA

ginga.readthedocs.org

10 lines of code changed / added in order
to read HDFITS (thanks to abstracted AstroImage class)

Bitshuffle: a HDF5 pre-compression filter for radio astronomy

for radio astronomy



Kiyo Masui (CHIME)
github.com/kiyo-masui/bitshuffle

Results: Compression

Results: Compression

Input file: LEDA visibility data (109 channel, 10 integration subset), in FITS-IDI format

	Size	Ratio	Write time	Read time	Weissman
(FITS)	1.16 GB	1.0 x	8.0 s	7.0 s	-
None	1.16 GB	1.0 x	9.0 s	7.5s	-
LZF	1.08 GB	1.08 x	13.5 s	8.0 s	1.05
bitshuffle	**705 MB**	**1.65 x**	**7.5 s**	**9.0 s**	**2.46**
GZIP	831 MB	1.40 x	53.0 s	10.0 s	1.00

$$W = \alpha \frac{r}{\overline{r}} \frac{\log \overline{T}}{\log T}$$

http://spectrum.ieee.org/view-from-the-valley/computing/software/a-madefortv-compression-metric-moves-to-the-real-world