

# ?Astute?

## Unleashing ALMA: Science Mining and Visualization of Large Data Cubes

**DATA(NX,NY,NZ,NS) → ms[i].pdf**

- UM-astro: **Lee** Mundy, **Peter** Teuben, Alberto Bolatto, Marc Pound, Kevin Rauch, Alex Fitts
- UM-cs: Amitabh Varshney, **Horace** Ip, **Jon** Weisbaum
- UIUC-astro: **Leslie** Looney, **Doug** Friedel
- NRAO-cv: Adam Leroy
- NRAO-soc: Jeff Kern

# Proposed Study Focus

## 1. Software Exploration

- Access patterns, Image format, Meta Data
- Model vs. Observations

## 2. Community Input

- NAASC scientists and cycle 0&1 observers

## 3. Team Building

- Supply infrastructure to build on

## 4. Study Deliverables

- Memo's

# Worked on:

- Image segmentation (ADASS poster)
  - model galactic disk w/ molecular clouds
  - N253 ALMA data
- CVS module for Astute (called: **astute**)
  - RunIDL (GDL), RunSH, ClumpFind
  - \$ASTUTE/opt: dendro, clfind
- GLUE (linked data / visualisation)
- CasaCore (also for MIRIAD)
  - **carmafiller**, casa::Lattice
- **image-io.pdf** (CARMA memo in prep.)

# CASA

- **casapy** : ipython frontend to the CASA code
- CASA (formerly AIPS++) has two code trees:
  - **casacore** (e.g. `casa::Lattice` and `casa::Image`)
  - **casa** (science interfaces, including python)
- How to integrate alien codes?
  - Direct import (“**from astute import clumpfind**”)
  - Client-server
  - Pickle (python's binary In/Out)

-

# Python

- <http://enthought.com> (binary, free academic)
- SciPy / NumPy / Matplotlib / Pandas
  - Annual SciPy conference (June 2013 @ Austin; w/ mini A&A)
  - STSCI initiatives (e.g. pyraf)
- AstroPy (community packaging)
  - FITS, WCS, VO, TableIO, asciitable,
- GLUE (linked data toolkit; Beaumont et al.)
- yt (theory visualization; Turk et al.)

# Python

- Pros:
  - Rapid code development
  - Powerful data structures
  - Fast in-core usage (little disk I/O)
- Cons:
  - Interpreted (though Numpy/SciPy can be fast)
  - Limited distributed memory [?]

# Scaling laws

- How do our implementations scale as function of  $(NX, NY, NZ, NS)$

# Interesting Benchmarks

- **HANNING** (in NZ; slower if cube is stored NX first)
  - MIR::hanning 6.6 vs. 210 sec
  - NEMO::ccdsMOOTH 2.7 vs. 4.9 sec [double]
  - CASA::ia.hanning() 3.1 vs. 10.3 sec [perm err?]
- **FITS**
  - MIR::fits 3.8 sec
  - NEMO::fitsccd 7.0 sec [double/matdef]
  - CASA::importfits() 5.8 sec
  - pyfits::readdata() 2.3 sec



# ADASS chat w/ Friedel

- Create meta data from feature ID/clumpfind
  - Tool plugin comparison [koribalski, starlink]
- Use splatalogue type access to identify lines
- Use mol spectra (per pos on sky) to map P,T,rho,species
- Theory:
  - RAD/LVG interfaces
  - Wolfire code
- Use meta data to mask emission and create more meaningful vel and vel disp maps
- Observing general theory data

# Time Scales

- 18 Apr 2013: NRAO Development Workshop
  - spring: call for proposal Development Studies
- August 2013 ? spring ?
  - Call for proposal for Development Projects