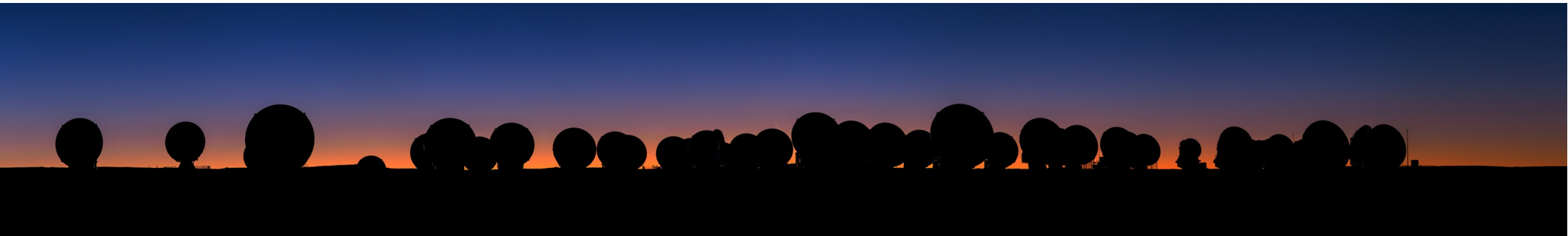
ADMIT:
ALMA Data Mining Toolkit

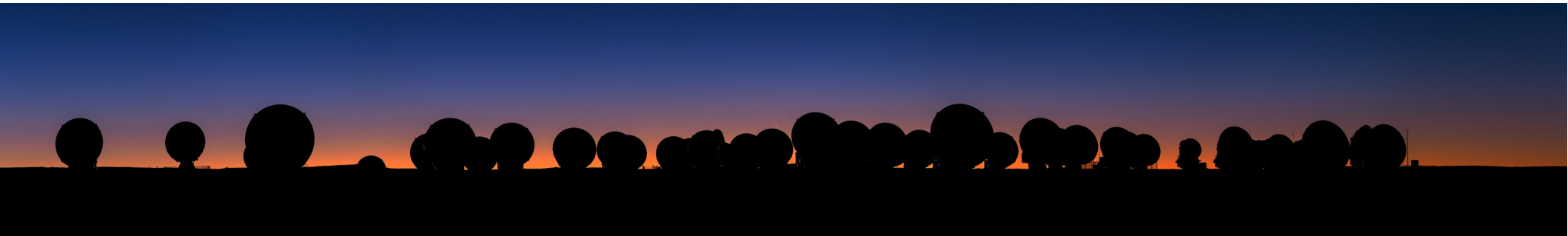
Peter Teuben
University of Maryland

ADASS
Oct 19, 2016



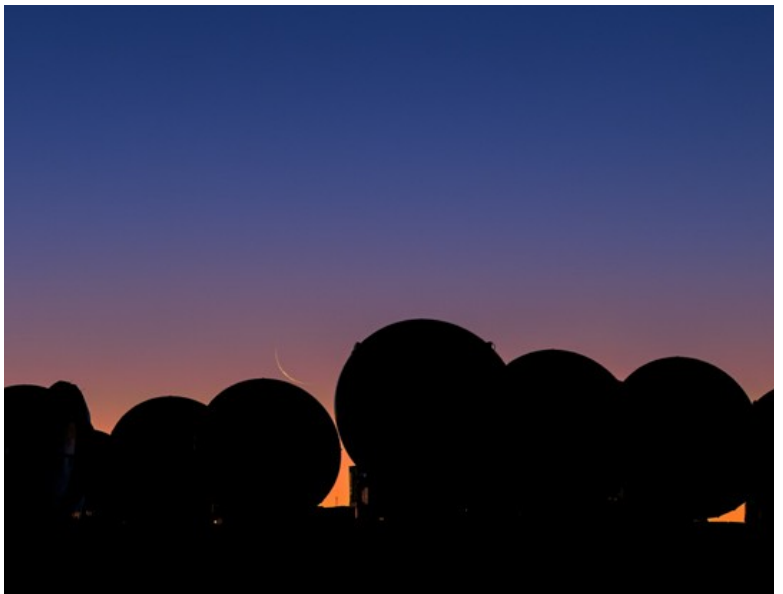
ALMA

- A schizophrenic telescope
- Software: **CASA**
  - Calibration “align the phases; scale amplitudes”
  - Mapping or Imaging “that fourier transform + deconvolution”
  - Image/Cube Analysis “separate the signal from the noise”
- Two telescope models (implemented as set of normalized Tables)
  - AlmaScienceDataModel (**ASDM**) for ALMA (and eVLA)
  - MeasurementSet (**MS**) for CASA
- ALMA Development Projects and Studies
  - **CARTA** (casaviewer replacement) and **ADMIT** (post-pipeline analysis) 2014-2016 - 2yr
  - **ASTUTE** (study - prepared for ADMIT) (2013 - 1yr)
  - **TP2VIS** (study how to combine single dish (“Total Power”) & interferometer data (2017 - 1yr)



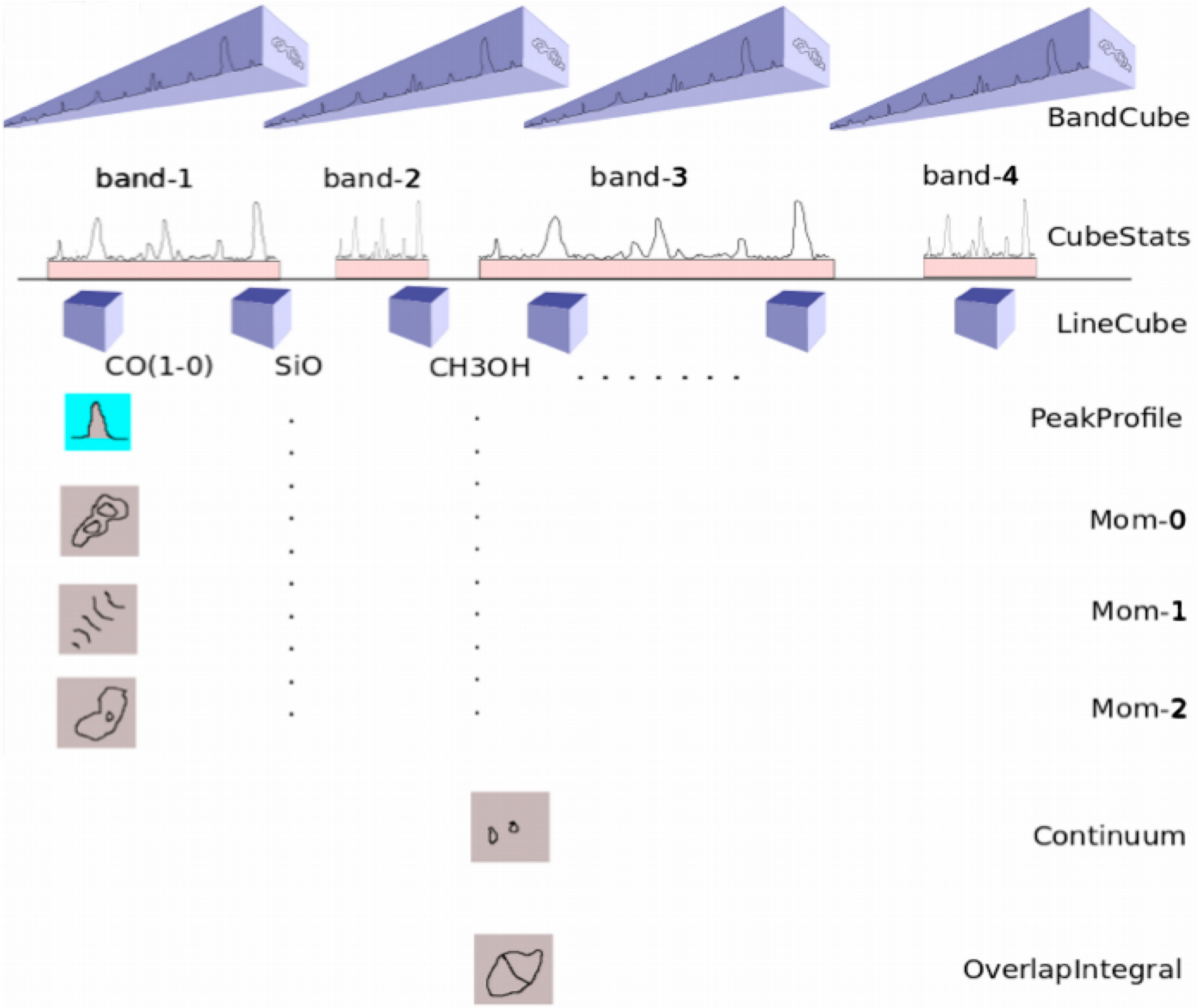
ADMIT team

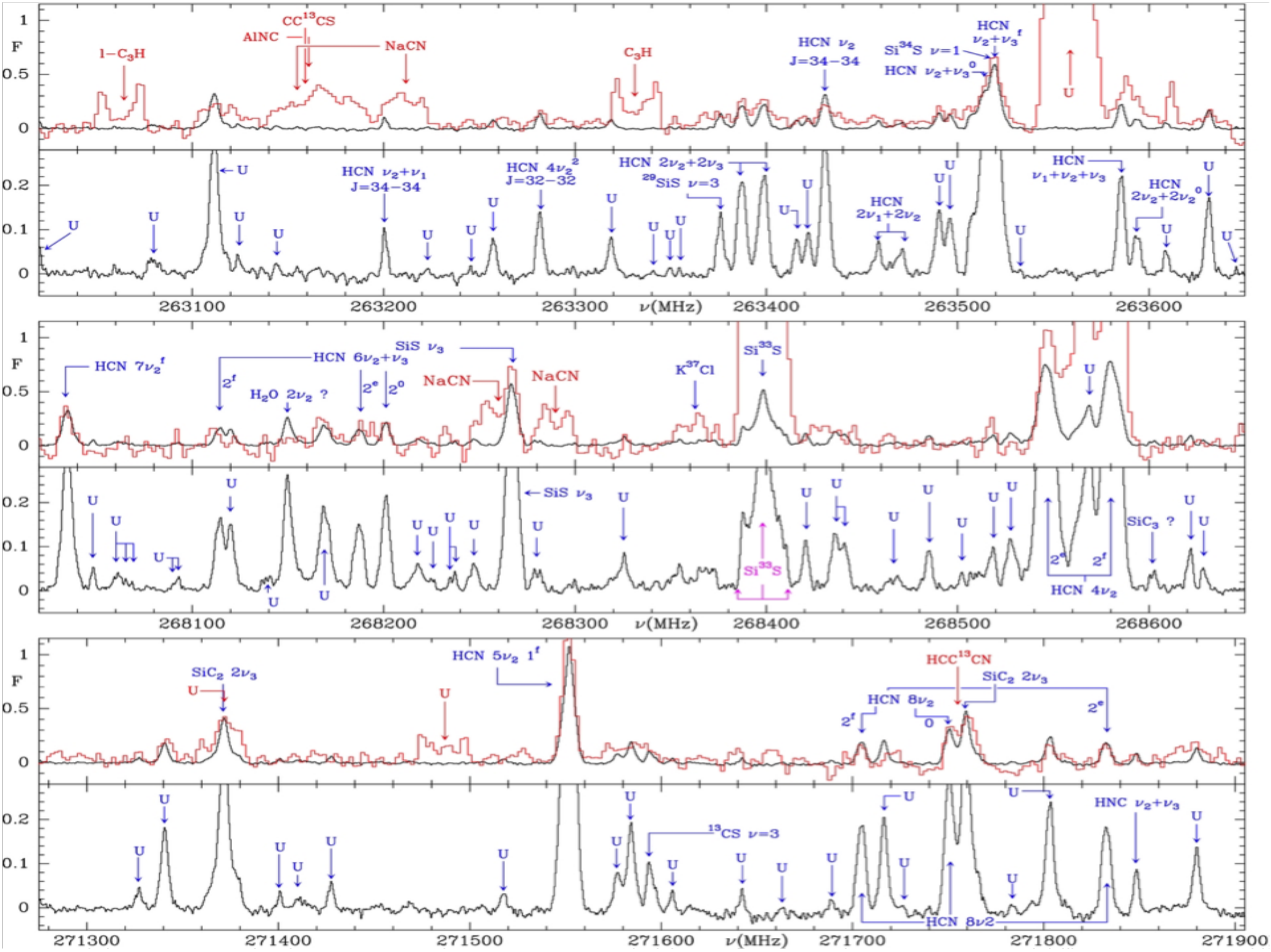
- PI: Lee Mundy (UMD)
- UMD: **Peter Teuben**, Marc Pound, Kevin Rauch
- UIUC: Leslie Looney, Doug Friedel, Lisa Xu, Robert Harris
- NRAO: Jeff Kern, Mark Lacy, John Hibbard (+...)

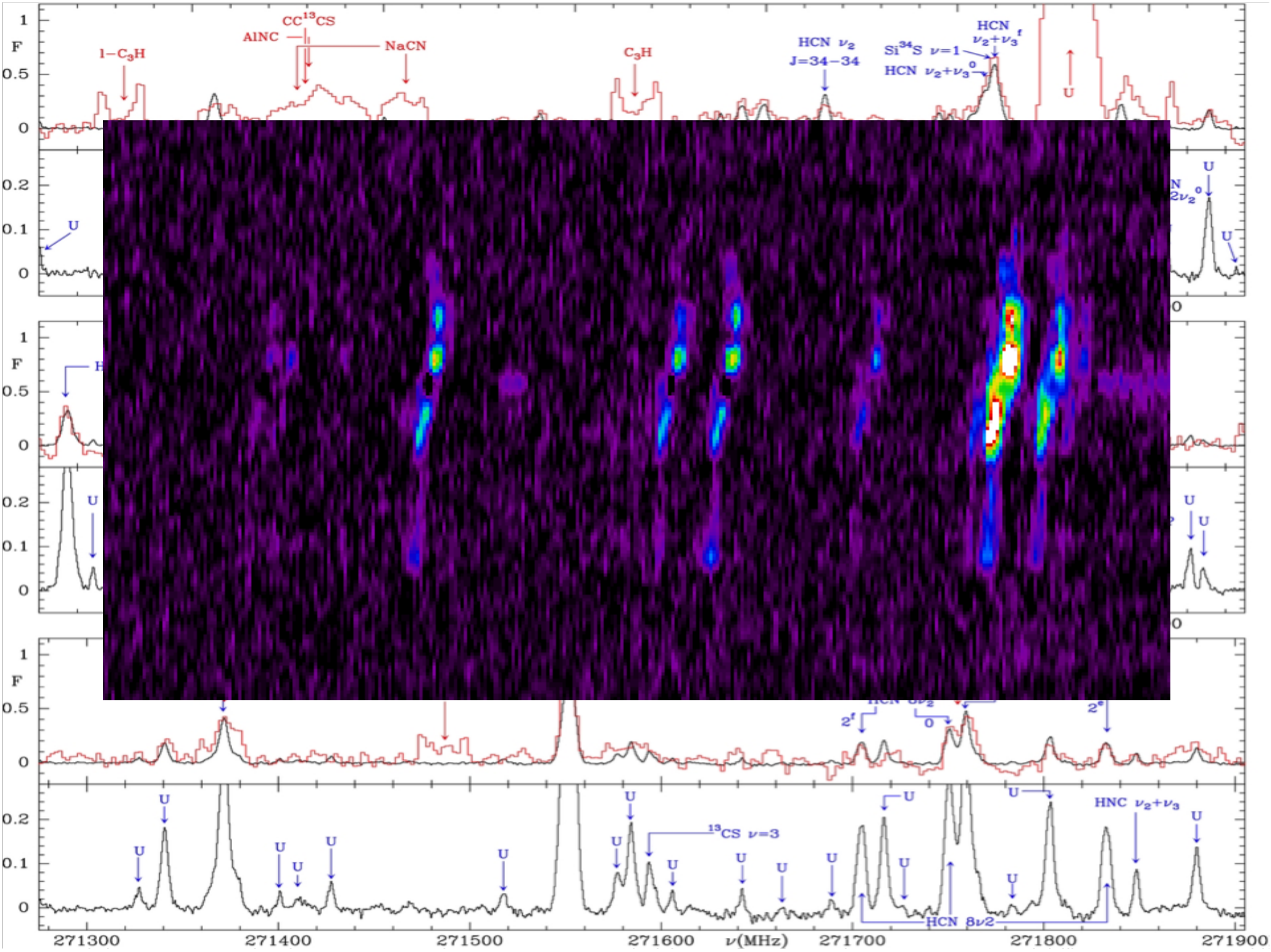


See also

- Artemix – ALMA RemoTE MIning eXperiment – P1.34
- Machine learning … spectral lines - P2.2
- ALMA Science Archive – P8.25
- ALMAWebQL v2 – D7
-


BandCube
band-1
band-2
band-3
band-4
CubeStats
LineCube
CO(1-0)
SiO
CH3OH
. . . . . .
PeakProfile
Mom-0
Mom-1
Mom-2
Continuum
OverlapIntegral


F
l-C3H
CC13CS
AlNC
NaCN
C3H
HCN ν2 J=34-34
Si34S ν=1
HCN ν2+ν3 0
HCN ν2+ν3 f
U
HCN ν2+ν1 J=34-34
HCN 4ν2 2 J=32-32
HCN 2ν2+2ν3
29SiS ν=3
HCN 2ν1+2ν2
HCN ν1+ν2+ν3
HCN 2ν2+2ν2 0
ν(MHz)
263100
263200
263300
263400
263500
263600
HCN 7ν2 f
HCN 6ν2+ν3
SiS ν3
2f
H2O 2ν2 ?
2e
2o
NaCN
K37Cl
Si33S
SiS ν3
SiC3 ?
HCN 4ν2
268100
268200
268300
268400
268500
268600
SiC2 2ν3
HCN 5ν2 1f
HCC13CN
HCN 8ν2
13CS ν=3
HNC ν2+ν3
HCN 8ν2
271300
271400
271500
271600
271700
271800
271900


F
1-C3H
CC13CS
AlNC
NaCN
C3H
HCN ν2
J=34-34
Si34S ν=1
HCN ν2+ν3 0
HCN ν2+ν3 f
U
2f
0
2e
13CS ν=3
HCN 8ν2
HNC ν2+ν3
271300
271400
271500
271600
ν(MHz)
271700
271800
271900

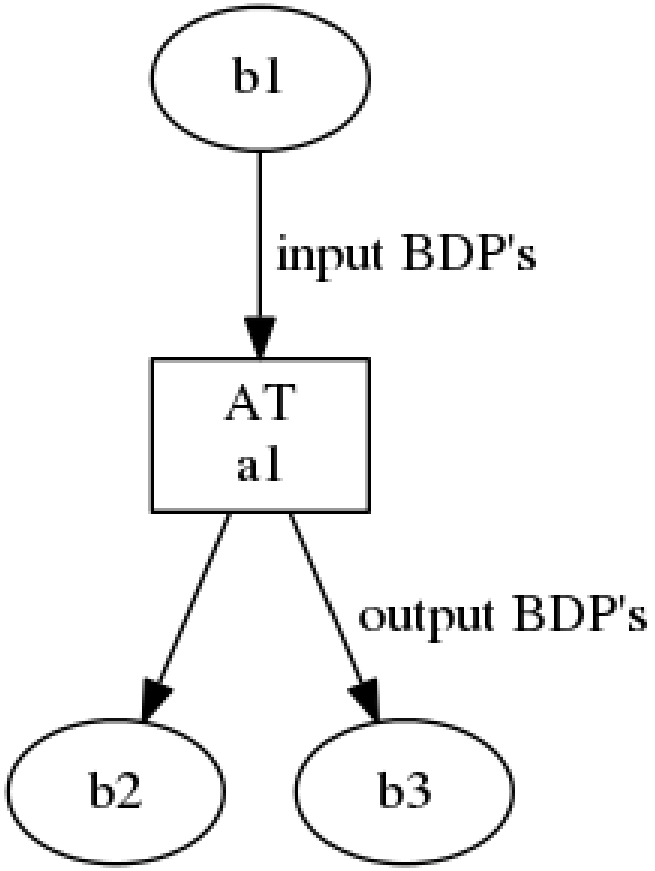
ADMIT:

Extract and add interesting science data products to the ALMA archive

- Python toolkit (generic – not CASA dependent)
- “flow” with reflow and dependencies (à la “make”)
  - **AT** = ADMIT Task
  - **BDP** = Basic Data Product (xml wrapper)
- AT’s can be implemented using any module(s), e.g. CASA or radio-astro-tools or MIRIAD
  - Current version: using CASA tasks and tools + Numpy/Scipy
- BDP implement tables, but wrap images
- “MultiFlow” : combine flows

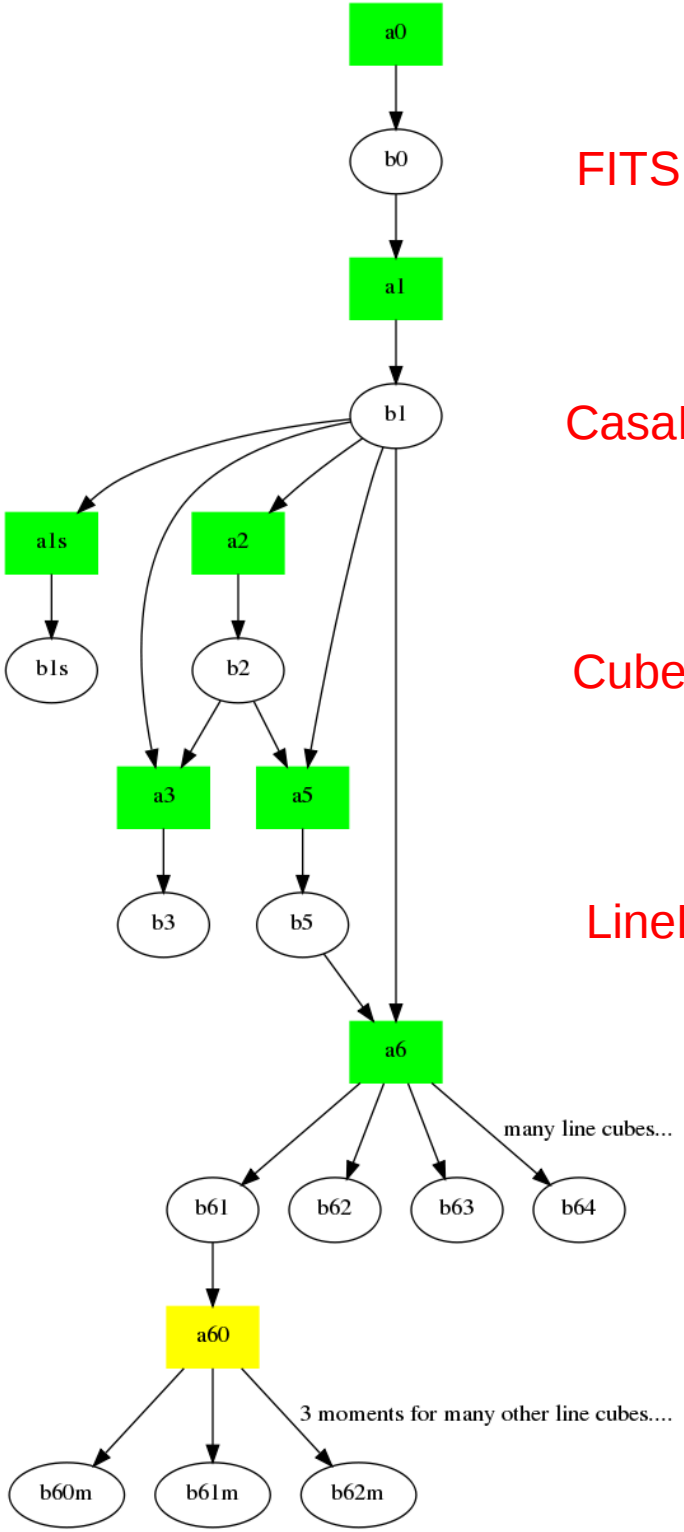
essential classes: AT and BDP

- AT:
  - BDP_in[]
  - BDP_out[]
  - Parameters
- BDP
  - AT
  - BDP_parents[]
  - BDP_children[]



ADMIT:
AT's and BDP's

- BDP:
  - Data wrapper (in XML)
  - Small data are inside the XML file, e.g. tables, numbers
  - Large data (fits files, png files) are XLINK'd
  - A BDP remembers how it was produces (AT w/ Parameters)
- AT:
  - Compute wrapper
    - can be implement in CASA, MIRIAD, NEMO etc.
    - can contain many calls to tasks in CASA/MIRIAD/NEMO etc.
  - Takes # input BDP(s)
  - Produces # output BDP(s)
    - an M to N mapping (M can be 0)
  - Has optional input parameters (“keyword=value”) to control the task



FITS

CasaImage

CubeStats

LineList

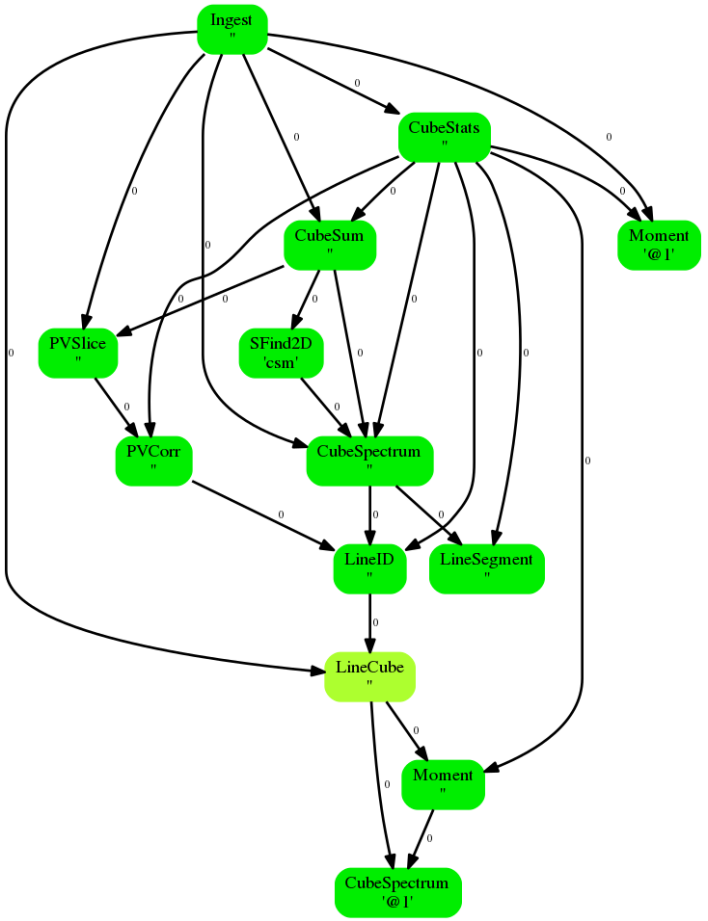
LineCube

LineCubeMoment0,...

ADMIT flow diagram AT and BDP's

A simple ADMIT flow:

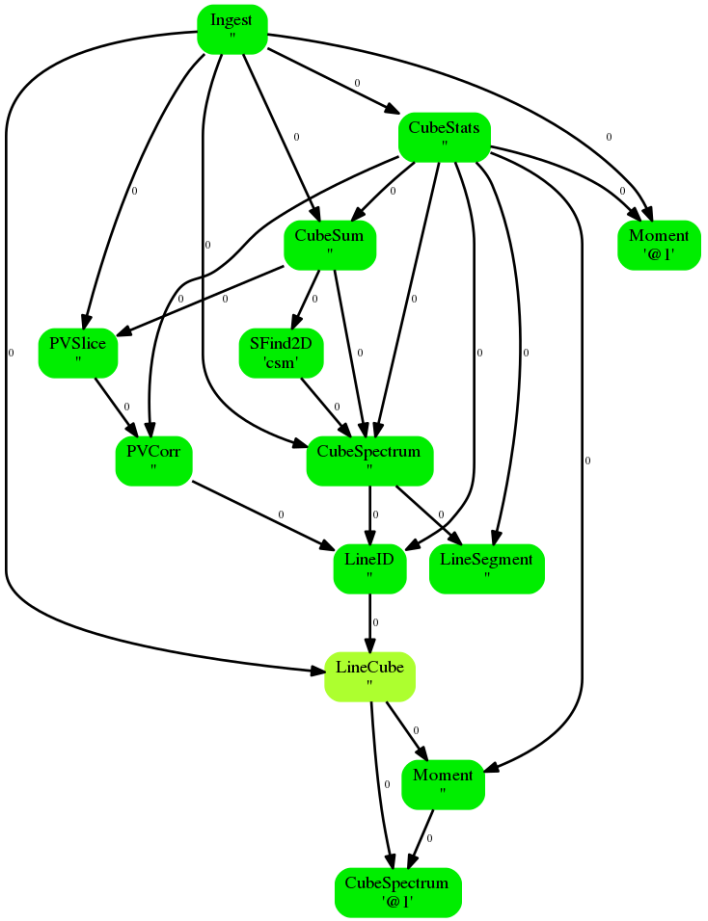
cube statistics, spectra, line-id, line-cubes

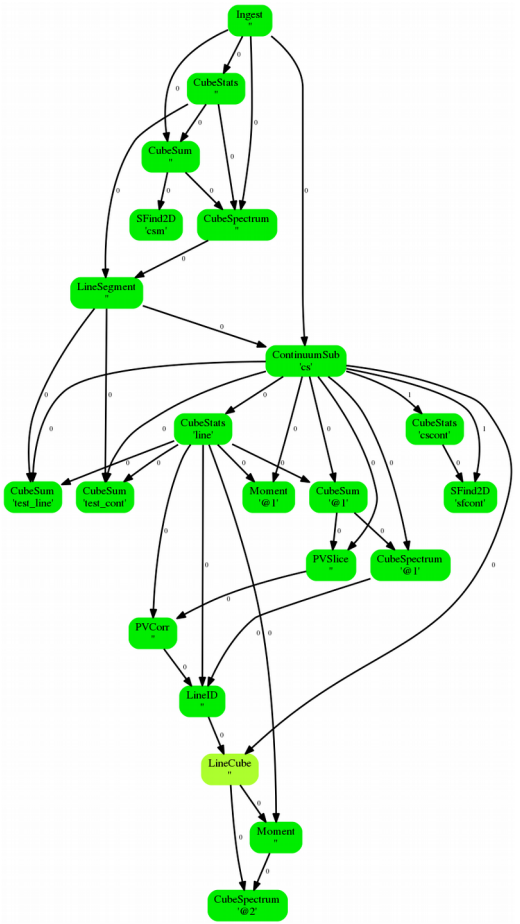


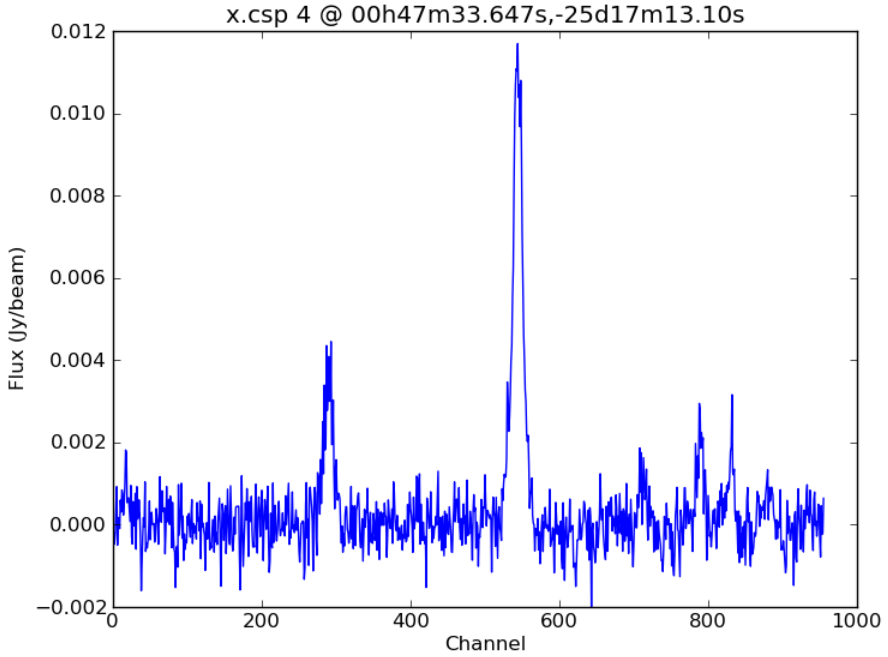
Automagically generated and updated during a flow

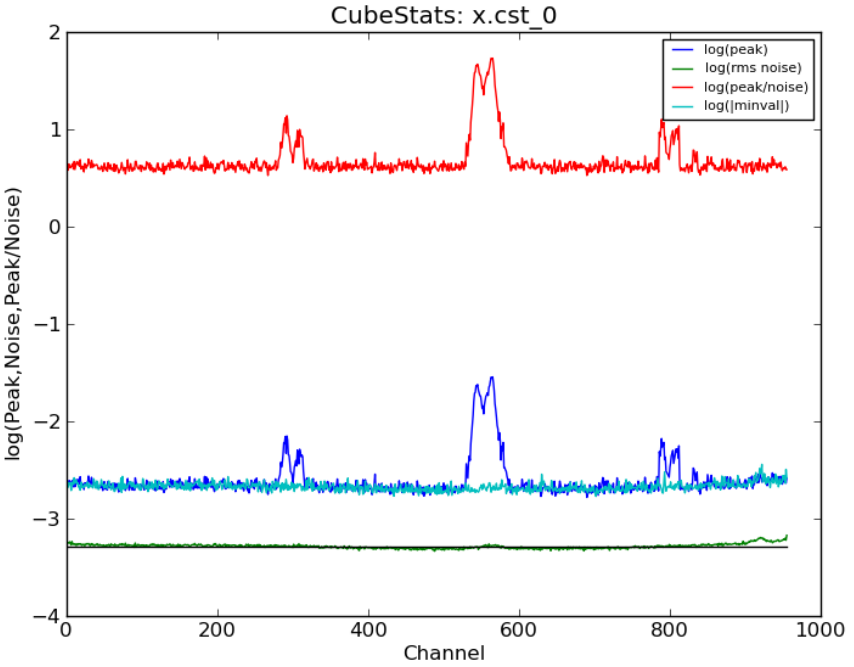
A simple ADMIT flow:

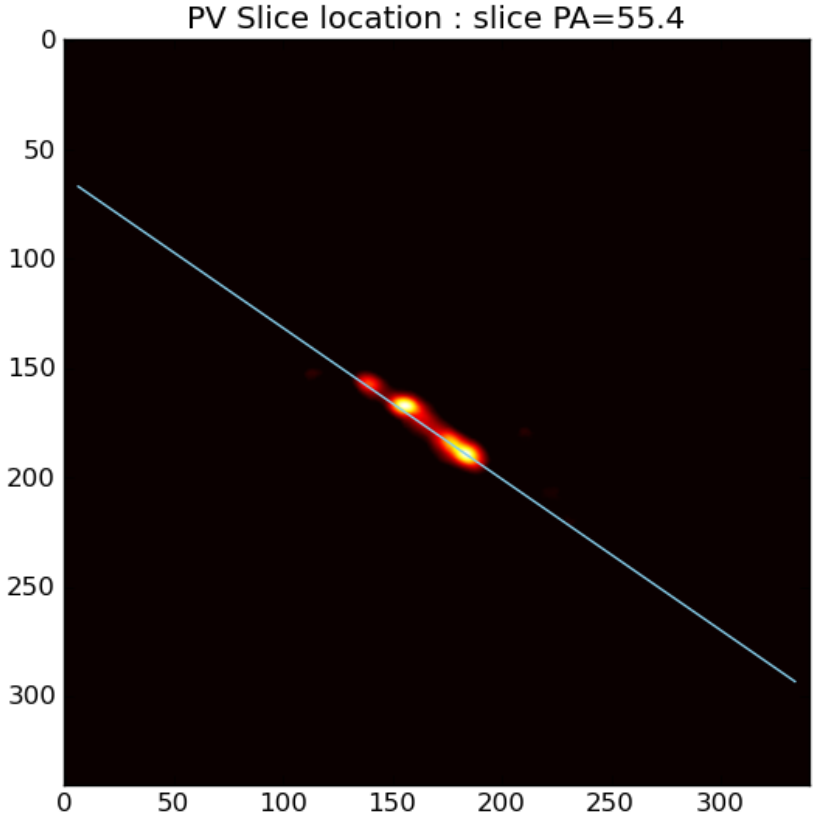
with added continuum subtraction

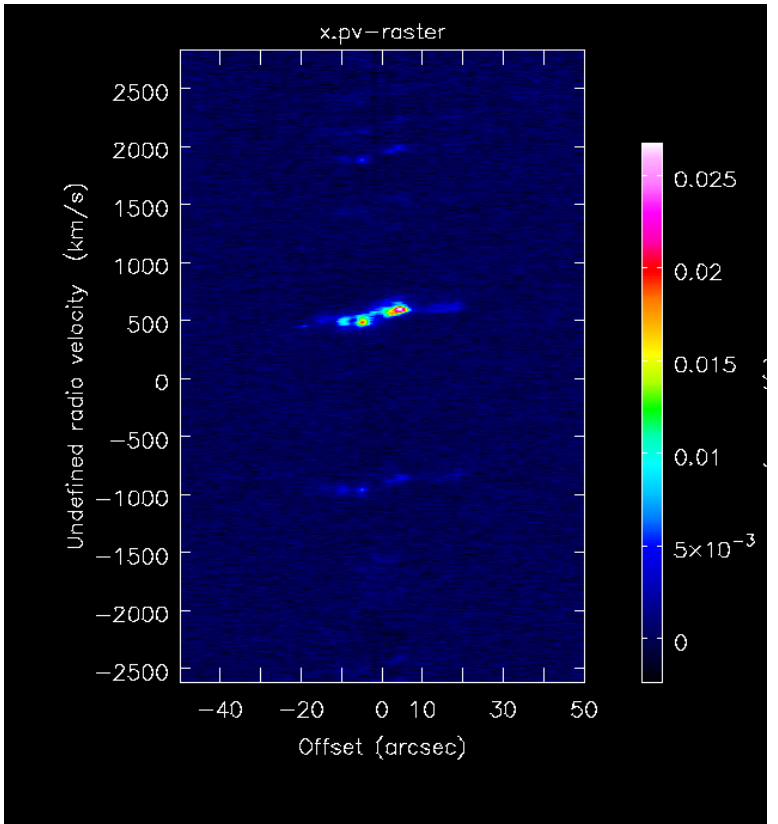




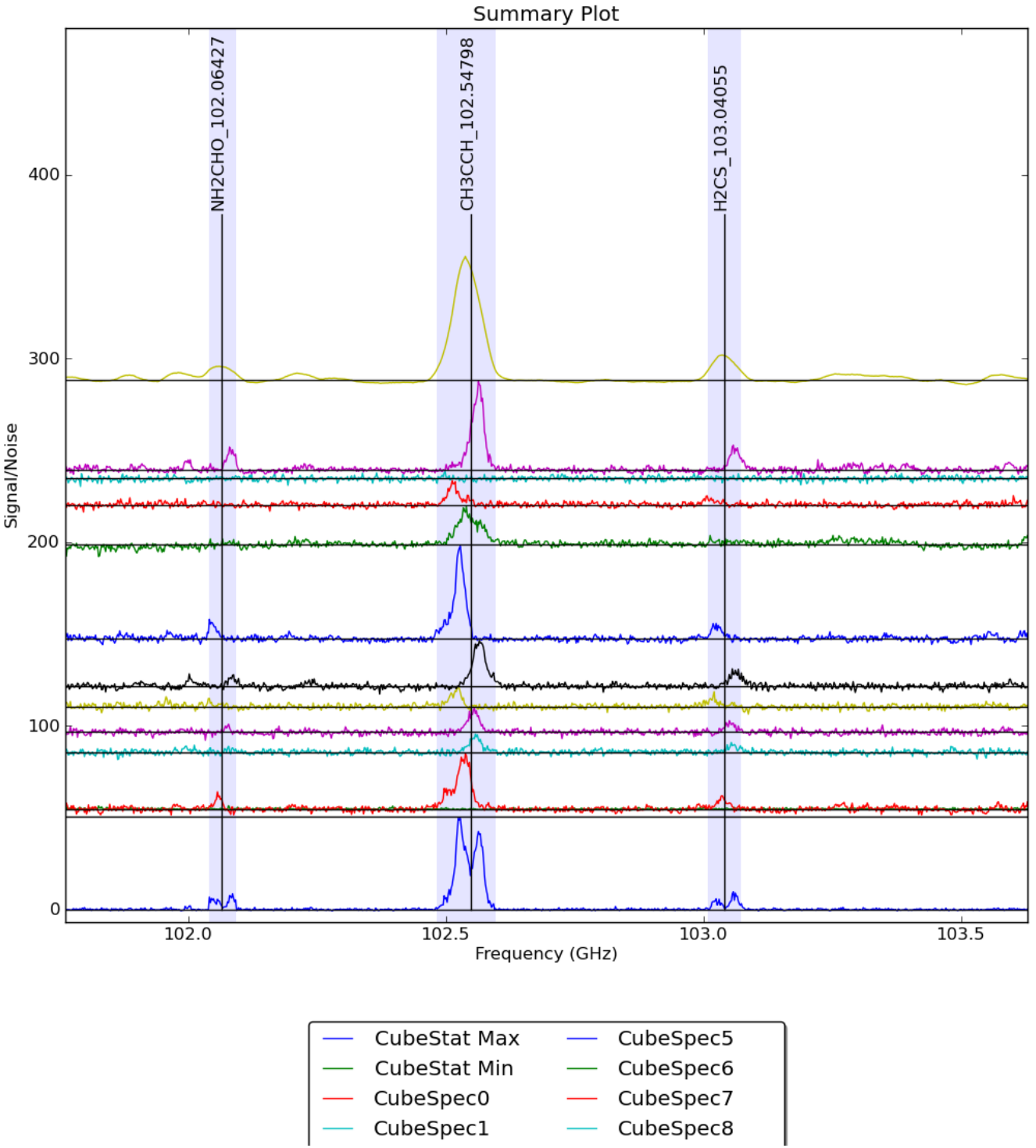

x.csp 4 @ 00h47m33.647s,-25d17m13.10s
Flux (Jy/beam)
Channel
0.012
0.010
0.008
0.006
0.004
0.002
0.000
−0.002
0
200
400
600
800
1000


CubeStats: x.cst_0
log(Peak,Noise,Peak/Noise)
Channel
log(peak)
log(rms noise)
log(peak/noise)
log(|minval|)
2
1
0
−1
−2
−3
−4
0
200
400
600
800
1000


PV Slice location : slice PA=55.4
0
50
100
150
200
250
300


x.pv−raster
Undefined radio velocity (km/s)
Offset (arcsec)
2500
2000
1500
1000
500
0
−500
−1000
−1500
−2000
−2500
−40
−20
0
10
30
50
0.025
0.02
0.015
0.01
5×10−3
0

c2h_2
c2h_1
hn13c
sio
h13co+
hco
h13cn
so_1
ch3ch2cn
87.5
87.0
86.5
86.0


Summary Plot
NH2CHO_102.06427
CH3CCH_102.54798
H2CS_103.04055
Signal/Noise
400
300
200
100
0
102.0
102.5
103.0
103.5
Frequency (GHz)
CubeStat Max
CubeStat Min
CubeSpec0
CubeSpec1
CubeSpec5
CubeSpec6
CubeSpec7
CubeSpec8

ADMIT Output for test0.admit

Flow View | Form View | LineID Editor | ADMIT Log | ADMIT documentation

TASKS LISTED IN ID NUMBER ORDER - CLICK ON TASK NAME TO SEE ITS OUTPUTS. COLORS AND ICONS INDICATE THE STATUS OF EACH TASK:

✔ TASK RAN NORMALLY (GREEN) | ⚠ TASK NEEDS TO BE RE-RUN (ORANGE) | ⊘ TASK IS DISABLED (PINK) | TASK HAS CRASHED (RED)

Flow Diagram for test0.admit

Ingest_AT (taskid=0)	file=x.fits mask=True		✔
CubeStats_AT (taskid=1)	robust=medabsdevmed ppp=True		✔
CubeSum_AT (taskid=2)	numsigma=4.0 sigma=0.00113133 smooth=[]		✔
SFind2D_AT (taskid=3)	nsigma= 6.0 sigma=2.57344 region= robust=['hin', 1.5] snmax= 35.0		✔
CubeSpectrum_AT (taskid=4)	pos=[(68, 63), (69, 63), ('10h27m51.227s', '-43d54m18.44s'), ('10h27m50.243s', '-43d54m24.36s')]	x.im	✔
LineSegment_AT (taskid=5)	numsigma=5.0 minchan=4 maxgap=3 segment=ADMIT smooth=[]		✔
PVSlice_AT (taskid=6)	slice=['6.00', '51.28', '121.00', '73.02'] width=5		✔
PVCorr_AT (taskid=7)	numsigma=3.0 range=[16,32]		✔
LineID_AT (taskid=8)	numsigma=5.0 minchan=4 maxgap=3 recomb=shallow smooth=[] tier1width=0.0 csub=[0, 0] iterate=True		✔
LineCube_AT (taskid=9)	pad=5 equalize=False		✔
Moment_AT (taskid=10)	moments=[0, 1, 2] numsigma=[2.0] mom0clip=2 chans=all	x.CO_115.27120	✔
CubeSpectrum_AT (taskid=11)	pos=[(68, 63)]	x.CO_115.27120	✔
Moment_AT (taskid=12)	moments=[0] numsigma=[3.0] mom0clip=2 chans=all	x-@1.mom	✔

ADMIT Output for test0.admit

Flow View | Form View | LineID Editor | ADMIT Log | ADMIT documentation

TASKS LISTED IN ID NUMBER ORDER - CLICK ON TASK NAME TO SEE ITS OUTPUTS. COLORS AND ICONS INDICATE THE STATUS OF EACH TASK:

✔ TASK RAN NORMALLY (GREEN) | ⚠ TASK NEEDS TO BE RE-RUN (ORANGE) | ⊘ TASK IS DISABLED (PINK) | TASK HAS CRASHED (RED)

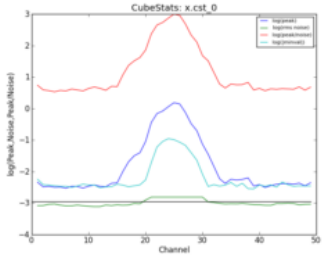
Flow Diagram for test0.admit

Ingest_AT (taskid=0) file=x.fits mask=True ✔

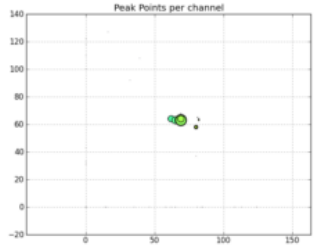
CubeStats_AT (taskid=1) robust=medabsdevmed ppp=True ✔

CubeStats_AT computes image-plane robust statistics on datacubes. These statistics are particularly useful for identifying spectral lines in images where the noise varies as a funnction of frequency.

CASA image	x.im
RMS method	medabsdevmed
RMS value	1.131E-03
Dynamic range	1.361E+03
Data mean	8.325E-04



Emission characteristics as a function of channel, as derived by CubeStats_AT (cyan: global rms, green: noise per channel, blue: peak value per channel, red: peak/noise per channel).



Peak point plot: Locations of per-channel peaks in the image cube x.im

CubeSum_AT (taskid=2) numsigma=4.0 sigma=0.00113133 smooth=[] ✔

SFind2D_AT (taskid=3) nsigma= 6.0 sigma=2.57344 region= robust=['hin', 1.5] snmax= 35.0 ✔

CubeSpectrum_AT (taskid=4) pos=[(68, 63), (69, 63), ('10h27m51.227s', '-43d54m18.44s'), ('10h27m50.243s', '-43d54m24.36s')] x.im ✔

LineSegment_AT (taskid=5) numsigma=5.0 minchan=4 maxgap=3 segment=ADMIT smooth=[] ✔

ADMIT Task Inputs for test0.admit

Flow View | Form View | LineID Editor | ADMIT Log | ADMIT documentation

TASKS LISTED IN EXECUTION ORDER - CLICK ON TASK NAME TO SEE ITS OUTPUTS. COLORS AND ICONS INDICATE THE STATUS OF EACH TASK:

✔ TASK RAN NORMALLY (GREEN) | ⚠ TASK NEEDS TO BE RE-RUN (ORANGE) | ⊘ TASK IS DISABLED (PINK) | TASK HAS CRASHED (RED)

- Ingest_AT (taskid=0) ✔
- CubeStats_AT (taskid=1) ✔
- CubeSum_AT (taskid=2) ✔
- SFind2D_AT (taskid=3) ✔
- CubeSpectrum_AT (taskid=4) ✔
- LineSegment_AT (taskid=5) ✔
- PVSlice_AT (taskid=6) ✔
- PVCorr_AT (taskid=7) ✔
- LineID_AT (taskid=8) ✔
- LineCube_AT (taskid=9) ✔
- Moment_AT (taskid=10) ✔
- CubeSpectrum_AT (taskid=11) ✔
- Moment_AT (taskid=12) ✔

Update ADMIT flow state (dry run)

Re-run ADMIT flow

ADMIT Task Inputs for test0.admit

Flow View | Form View | LineID Editor | ADMIT Log | ADMIT documentation

LineID_AT (taskid=8) ✔

On-line documentation of LineID_AT

Parameters

Input BDPs

Index	Type	File name
0	CubeSpectrum_BDP	x.csp
1	CubeStats_BDP	x.cst
2	PVCorr_BDP	x.pvc

Output BDPs

Index	Type	File name
0	LineList_BDP	x.ll

Keywords

numsigma	5.0
force	[]
tier1width	0.0
online	False

ADMIT Task Inputs for test0.admit

Flow View | Form View | LineID Editor | ADMIT Log | ADMIT documentation

mincnan	4
pattern	AUTO
smooth	[]
recalcnoise	False
vlsr	-999999.99
maxgap	3
reject	[]
method	{'PeakFinder': {'thresh': 0.0}}
mode	ONE

LineCube_AT (taskid=9) ✔

Moment_AT (taskid=10) ✔

CubeSpectrum_AT (taskid=11) ✔

Moment_AT (taskid=12) ✔

Update ADMIT flow state (dry run)

Re-run ADMIT flow

ADMIT Output for test0.admit

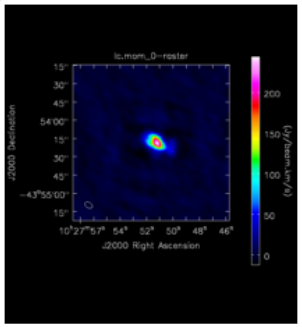
Flow View | Form View | LineID Editor | ADMIT Log | ADMIT documentation

LineCube_AT (taskid=9) pad=5 equalize=False ✔

Moment_AT (taskid=10) moments=[0, 1, 2] numsigma=[2.0] mom0clip=2 chans=all x.CO_115.27120 ✔

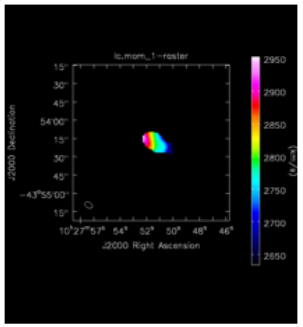
Moment_AT creates moment maps using custom clip levels.

Moment_AT output for x.CO_115.27120/lc.im



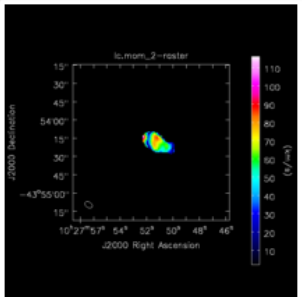
Carbon Monoxide Moment 0 map of Source NGC3256

View in CASA



Carbon Monoxide Moment 1 map of Source NGC3256

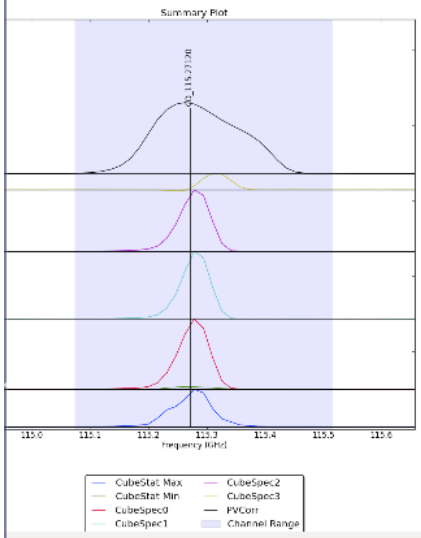
View in CASA



/home/teuben/ADMIT/data1/M3/run0/test0.admit

resources
x.CO_115.27120
x.csm
x.im
x.pv
x.pv.smooth
x-@1.mom_0
admit.dot
admit.log
admit.png
admit.xml
admit0.py
admit1.py
form.html
imstat0.logfile
imstat1.logfile
index.html
lineIDedit.html
llfiles.json
lltable.8.json
log.html
pvoverlay.png
test0.fits
test0.fits.apar
testCubeSpectrum.tab
testCubeStats.tab
testPVCorr.tab
x.csm.bdp
x.csm.png
x.csm_histo.png
x.csp.bdp
x.csp_0.png
x.csp_1.png
x.csp_2.png
x.csp_3.png
x.cst.bdp
x.cst_0.png
x.cst_00.png
x.cst_1.png
x.cst_2.png
x.cst_5.png
x.cst_6.png
x.cst_ppp.png
x.im.bdp
x.ll.bdp
x.ll_pvspec.png
x.ll_pvspec.svg
x.ll_pvspec.svg.html
x.ll_spec000.svg
x.ll_spec001.png
x.ll_spec001.svg.html
x.ll_spec001.svg
x.ll_spec002.png
x.ll_spec002.svg.html
x.ll_spec002.svg
x.ll_spec003.png
ll_statspec0.png
x.ll_statspec0.svg
x.ll_statspec0.svg.html
x.ll_statspec1.png
x.ll_statspec1.svg
x.ll_statspec1.svg.html
x.ll_summary.png
x.lseg.bdp
seg_spec002.png
x.lseg_spec003.png
x.lseg_statspec0.png
x.lseg_statspec1.png
x.lseg_summary.png
x.pv.bdp
x.pv.png
x.pvc.bdp
-@1.mom_0.png
x-@1.mom_0_histo.png
x-csm.ds9.reg
x-csm.sl.bdp
x-csm.sl.png

_summary.png



pixels 83.9 kB 37% 32 / 44

"x.ll_summary.png" selected (83.9 kB), Free space: 12.2 GB

Technical Details

- CVS → git : https://github.com/astroumd/admit
- Quasi-Agile
  - Prototyped (ASTUTE, BDP centric → AT centric)
  - Rapid development cycle
  -
- Tests: unit, integration, regression (buildbot)
- Documentation: sphinx
- Runtime reporting: bootstrap
- Integration into CASA

Example

example install

1% git clone https://github.com/astroumd/admit

2% cd admit

3% autoconf

4% ./configure --with-casa-root=/opt/casa-release-4.7.0-el6

5% source admit_start.sh

example usage

6% runa1 test253_spw3.fits

7% aopen test253_spw3.admit

example CASA

% casa

Import admit

p = admit.project('foo.admit')

Timeline / Future

- ADMIT 1.0 “delivered” May 2016
  - ADMIT 1.1 delivered Nov 2016 (handover to NRAO)
  - Acceptance Test TBD
  - ADMIT 2.0 in a next development proposal?
- Future Features
  - Desktop vs. Pipeline usage
  - Expand from the current 2 (or 5) Recipes
  - Expand from the current ~20 AT’s
  - Ingest_AT uses “noise-flat” cubes