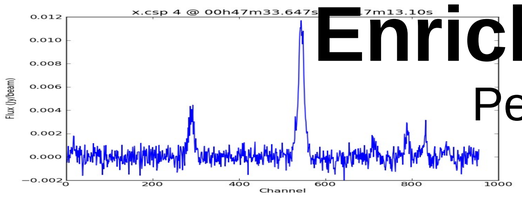
Enriching the ALMA Archive with Science Products

Peter Teuben[1], Marc Pound[1], Lee Mundy[1], Jorge García Yus[2], Felix Stoehr[3]

1) University of Maryland, College Park, MD

2) JAO, Santiago, Chile

3) ESO, Garching, Germany



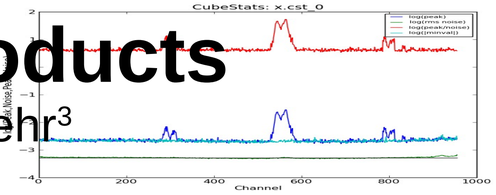


Figure 1: Example "spectra" used in the LineID task in ADMIT.
On the left a CubeSpectrum, on the right a CubeStats spectrum.

What happened in 2020 again?

The shutdown of ALMA due to the COVID-19 crisis provided an opportunity to reprocess all archived ALMA data and derive a uniform set of Science Products using **ADMIT** (ALMA Data Mining Toolkit).

Users will be able to download these as an additional data product from the ALMA Archive. These data are typically very small, containing only figures and tables. In a related project (see Teuben et al. 2020) these data can be queried in a python environment.

See also poster P8-67 (Stoehr) describing the ARI-L project (74,000 science FITS files from Cycles 2-4).

How did we do this?

- A script AAP (ADMIT After Pipeline) processes a series of pipeline product and runs ADMIT in native and a 16 channel binned version, both at a 5 sigma signal clipping
- A recipe estimates the redshift 'z' for each FITS file (see **VLSR** box below)
- A total of 500,000 FITS files taking 137 TB in 8591 MOUS ("science observation") from Cycle 3-7
- So far, we only tested this procedure on several thousand Cycle7 data
- We are developing a sqlite interface to this data to allow users to query data on the science meta-data
- The new CASA6 environment allows for interfaces via Jupyter notebooks

What kind of Science Products?

- Continuum Maps:
  - point sources, fluxes, rms
- Line Cubes
  - VLSR
  - Line detections and identifications
  - Line density count
  - Line cubes, moment maps, fluxes, rms

Combined ALMA and ADMIT Results Table

ALMA Science Archive Results						ADMIT Results		
index	Project Code	Source Name	RA	Dec	...	SPW	Tasks	...
0	2017.1.005.S	CenA	12:34:56	01:02:03	...	23,25,27,29	6	...
1	2018.3.040.S	NGC 1234	16:17:18	-19:20:21	...	1,2,3,4,5,6	8	...

Spectral Window Table

index	ALMA index	SPW	# lines	# sources	CubeStats	CubeSpectrum	PV Slice	...
10	0	23	4	10				...
11	0	25	1	1				...

Line Table

index	Spectral Win. index	Frequency	Formula	Transition	velocity	startchan	endchan	...
9	10	110.20137	13CO	1-0	238.238	786	812	...
8	11	102.54798	CH3CCHv=0	6(0)-5(0)	179.740	529	586	...

Source Table

index	Spectral Win. index	Line index	Ra	Dec	Peak	Flux	S/N	...
41	10	9	12:34:56	01:02:03	12.3	20.1	30	...
42	10	5	12:34:49	01:03:00	3.6	7.5	5	...

Figure 3: Examples of the metadata tables we create from ALMA and ADMIT products for sophisticated science queries. The tables are linked via their indices (color highlights), allowing combined queries like "Sources with 13CO detected at S/N >10." Note we can also insert visual cues for the user such as thumbnail images of ADMIT products [Spectral Window Table].

VLSR

- Knowledge of the redshift (VLSR) is crucial for LineID
- ALMA observing files do not always have VLSR
- VLSR is not carried to the FITS header in the imaging pipeline
- A recipe has been developed to get reliable VLSR where this is possible, resulting in a very high success rate
  - If VLSR given, use it
  - If source name found in *a lookup table,* use it
  - If a Line_ID given, use it to compute VLSR
  - If source name found in NED/SIMBAD, use it
  - If nothing found, run ADMIT marking all lines U(nknown)
  - Optionally, the RESTFREQ and mid-band FREQ can also provide a VLSR, but this can result in failure
- Some projects (e.g. high-z spectral search) will need sophisticated finger printing and a priori and a posteriori knowledge