EALFA Memo 040804A:
Data Processing Stream (mode a) for ALFA Drift Surveys
rg & mh – 11 August 2004

Definitions for data units were given in a previous memo ("Bookkeepping for E-ALFA Drift Surveys"). Here we briefly summarize a proposal for data processing stream within the IDL environment, to be used as a starting point for discussions.

- 0. A **Drift**, containing seven **Strips** each associated with one of the ALFA beams plus an eighth one possibly used for RFI monitoring, is obtained at the telescope and delivered as a single FITS file. This file gets stored in a (perhaps two?) raw data directory at NAIC and transferred on a regular basis to data processing sites (mode tbd).
- 1. FITS files are decoded and converted into **s** structures. An **s** structure is a Strip of consecutive **Strip Segments**, each taken with the same ALFA beam, at the same Declination, during the same observing session. An **s** structure has the same format as PP's **m** structures in the AO/IDL package: 4 dimensions (2 pols, n_{rec} records or R.A. samples, n_s strip segments of 600 (or less) seconds, n_{chn} spectral channels.
- 2. **Noise calibration** and **Bandpass subtraction** are applied next, one **s** structure at a time and sequentially for each segment in the strip. Immediately after, the output is **Baselined**. This result we shall refer to as an $\mathbf{s}_1$ structure and will constitute the unit of Level 1 data products. The noise calibration and bandpass subtraction is automatic and blind. The baselining is interactive, and will provide the first visual inspection of the data and its quality validation. These processes produce a set of ancillary files:
  1. a set of **cals** structures, which retain memory of the noise calibration;
  2. a set of **cont** structures, which contain the fluxes of continuum sources encountered along the strip (separately for polarization and for several spectral bands, to yield some indication of spectral index);
  3. a set of **bl** structures, which contain coefficients of polynomial baselines for each spectrum;
  4. the **bp** bandpasses applied to each strip segment,
  5. a **bad** structure, which flags pixels in position–frequency maps excluded from the estimate of the bandpasses.

  These structures are stored in adjacent directories, labelled and logged as described in the Bookkeeping memo.
- 3. Once all the data within a tile have been acquired, all strip segments with data within that tile are combined to produce an **m** structure of the tile, possibly by stitching together four **quadrants** of the tile, which may be processed separately. Continuum maps of the tile are produced and comparison of fluxes extracted from the continuum maps with those in a flux catalog are used to (i) re–calibrate the flux scale of the tile and, for a multiple pass tile, (ii) identify variable continuum sources. Regridded and compressed versions of the tile are produced — Level 2 data products — for public access.
- 4. Automatic signal extraction of spectral sources is carried out on the tile **m** structure, via a matched filter approach that can separate spurious features with unphysical characteristics (for a cosmic signal), producing a first–look catalog of sources in the tile, with an automatic quality index assigned. Candidate detections are visually inspected and assigned a quality index by an observer. Cross–referencing with other extragalactic catalogs and other tools within the NVO environment are used.

- 5. In a multiple pass survey, steps 1–4 are repeated for each pass. Catalogs of candidate detections are cross–referenced, expected shifts in frequency in the geocentric rest frame are tested. Tiles are regridded and signal extraction is carried out again, to take advantage of higher sensitivity. Coadded tiles form Level 3 data products. Final catalogs of detections constitute Level 4 data products.