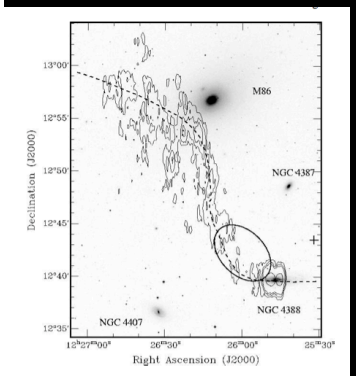
Source Finding: HIPASS, Duchamp and beyond . . .

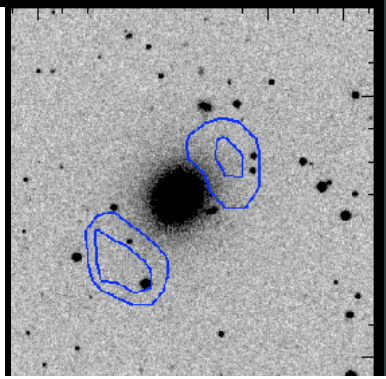
and beyond . . .

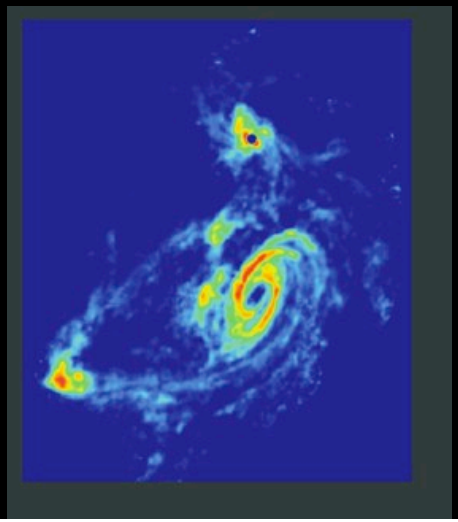
Virginia Kilborn
Kathrin Wolfinger
Swinburne University

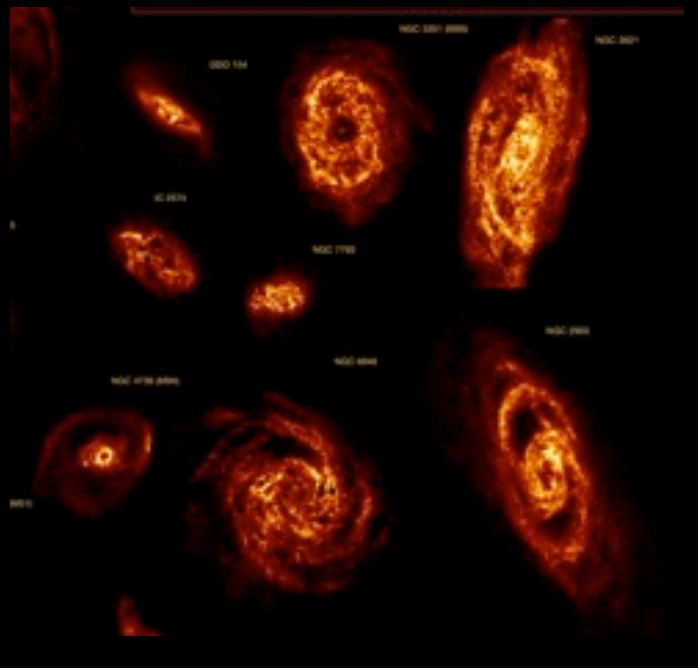
What do we want from a source finder?

- Produce a catalogue of sources with known characteristics (completeness/reliability)
- Provide accurate parameters for these sources (positions, fluxes, velocities etc)
- (Match the HI source to previously known objects).
- May need more than one algorithm to achieve these aims





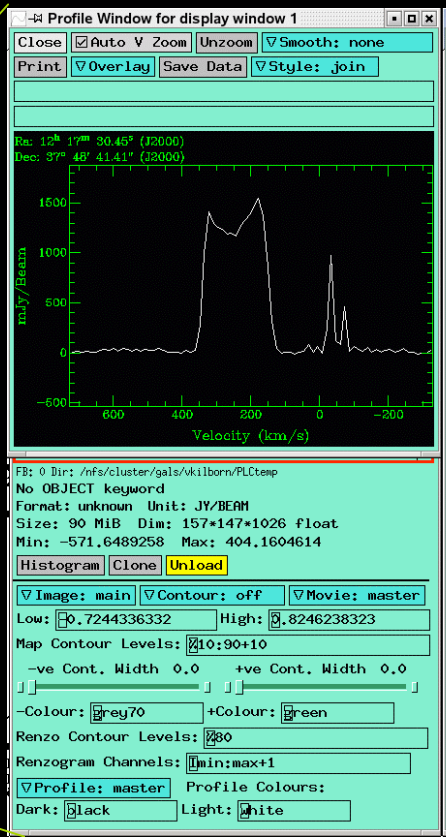


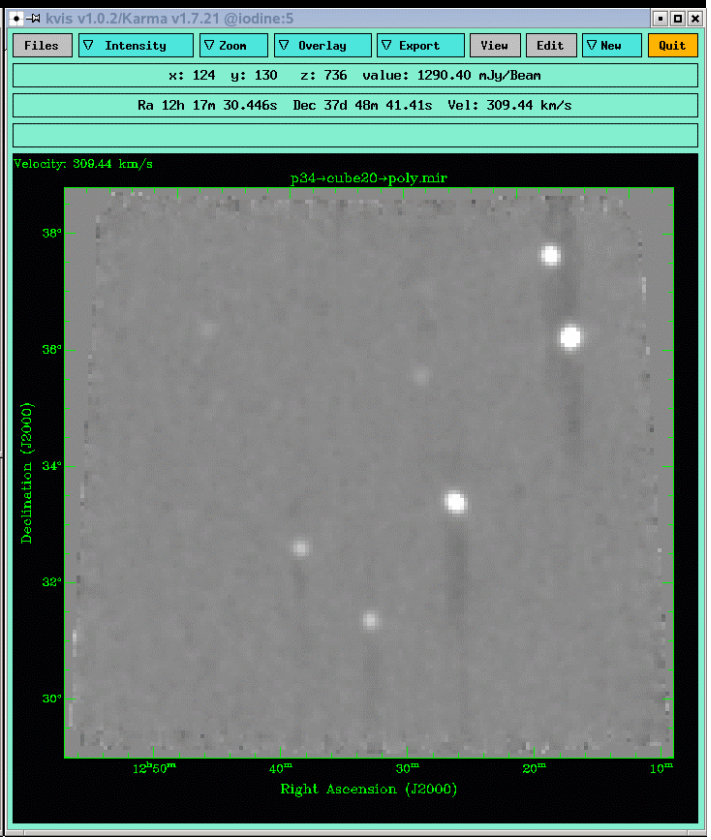


Galaxy finding in HIPASS

- Version 1: Graduate student powered
  (e.g. kvis, searching by eye)

  – OK for small datasets
  – Not robust for larger datasets





HIPASS galaxy finding v2: Automated algorithms

- 2 galaxy finding algorithms used for HIPASS:
  - MULTIFIND a peak flux detection algorithm utilising the MIRIAD task “IMSAD”
  - TOPHAT - a wavelet detection algorithm that used 40 filters to detect sources in the data
- Both algorithms needed to produce final source list as neither finder detected all sources
  - MULTIFIND ~83%
  - TOPHAT ~90%

Automated interference removal

- Step 2 in the HIPASS galaxy finding was to remove known sources of interference
  - Detections with v_{hel}<300 km/s removed (catalogued separately in HVC catalogue)
  - Detections with small velocity widths, at known frequencies of interference and recombination lines removed (this will also need to be done for WALLABY/DINGO - the increased velocity resolution should make this easy).

Meyer et al 2004

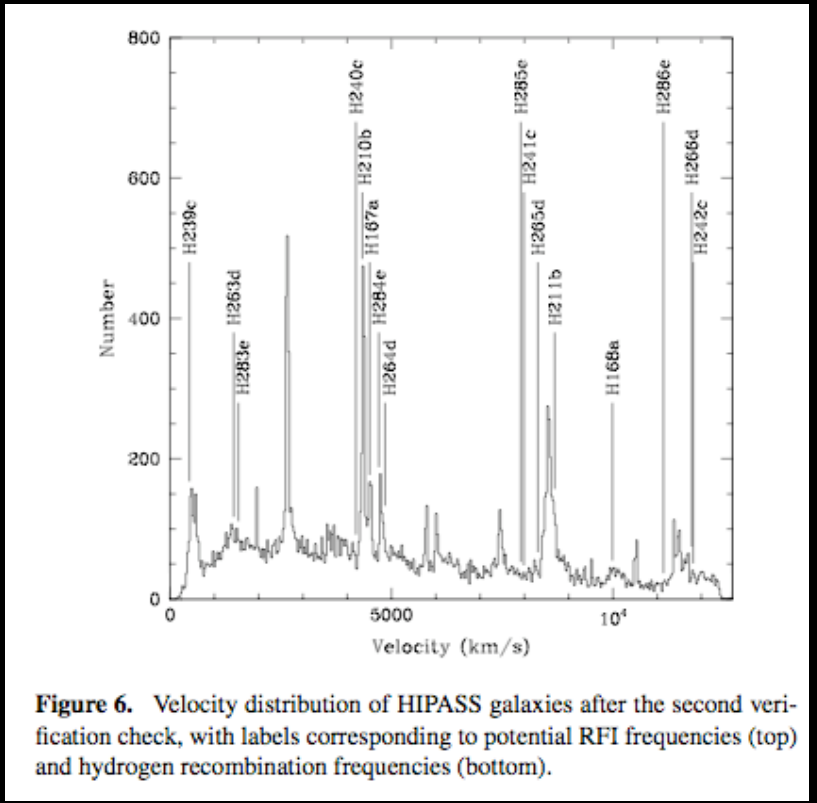


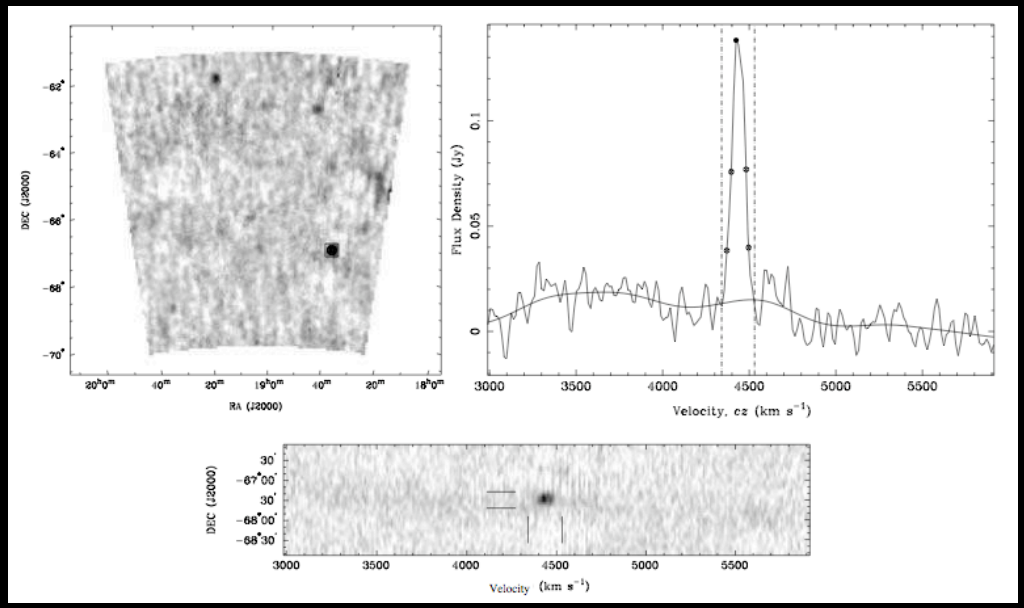
Figure 6. Velocity distribution of HIPASS galaxies after the second verification check, with labels corresponding to potential RFI frequencies (top) and hydrogen recombination frequencies (bottom).

HIPASS catalogue production

- Combined list from MULTIFIND and TOPHAT, produced a list of ~140,000 candidates that were inspected by eye…
- The final catalogue had 4,315 sources
- (the same ratio for WALLABY, with expected source count of 500,000 would give ~16,000,000 candidates to sort through !!)

HIPASS Parameterisation

- Interactive algorithm to parameterise HIPASS sources
  - Gaussian baseline fitted to spectra
  - Extended and confused sources flagged to be dealt with separately



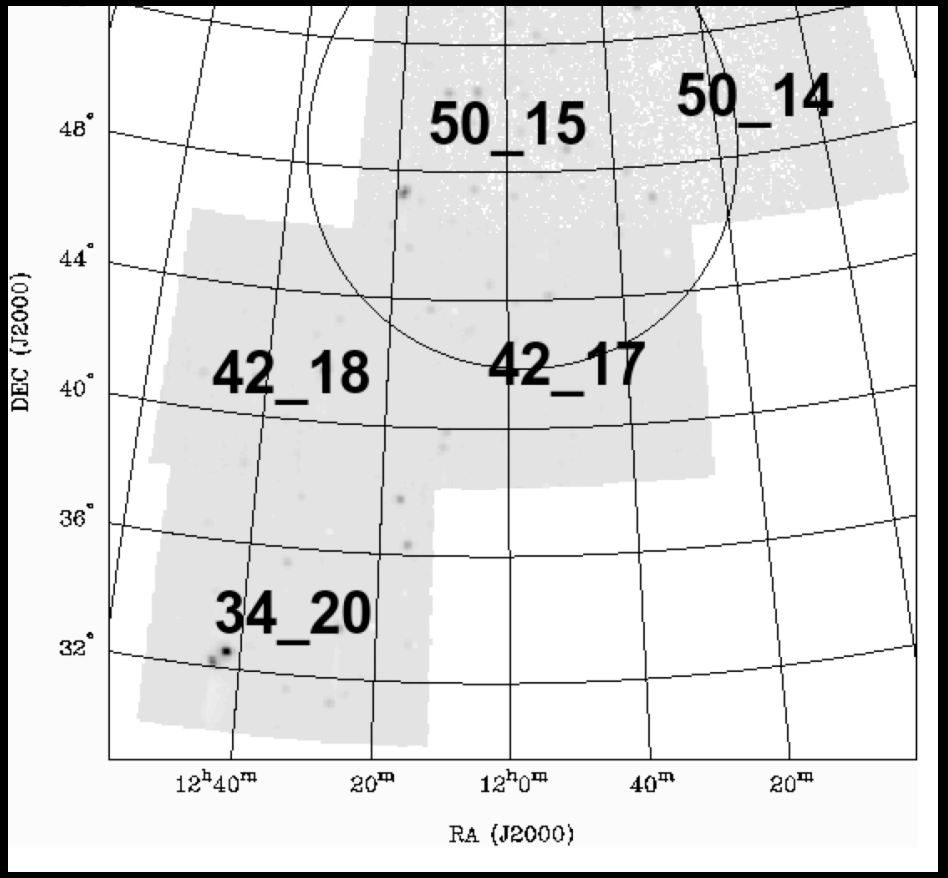
WALLABY/DINGO source finder needs to automatically parameterise the catalogue (?)

Lessons from HIPASS

- Need a smarter algorithm that produces less false detections (32 false to 1 real is too high!).
- We will have to deal with recombination lines, however the high frequency resolution of WALLABY/DINGO should help
- Parameterisation of the sources - preferably automatic
- Extended sources/doubles may need to be dealt with separately.

Duchamp: Test results

- Five cubes of Ursa Major region used for tests
- Data from Jodrell Bank Multibeam receiver
- Cubes have similar rms noise levels to each other



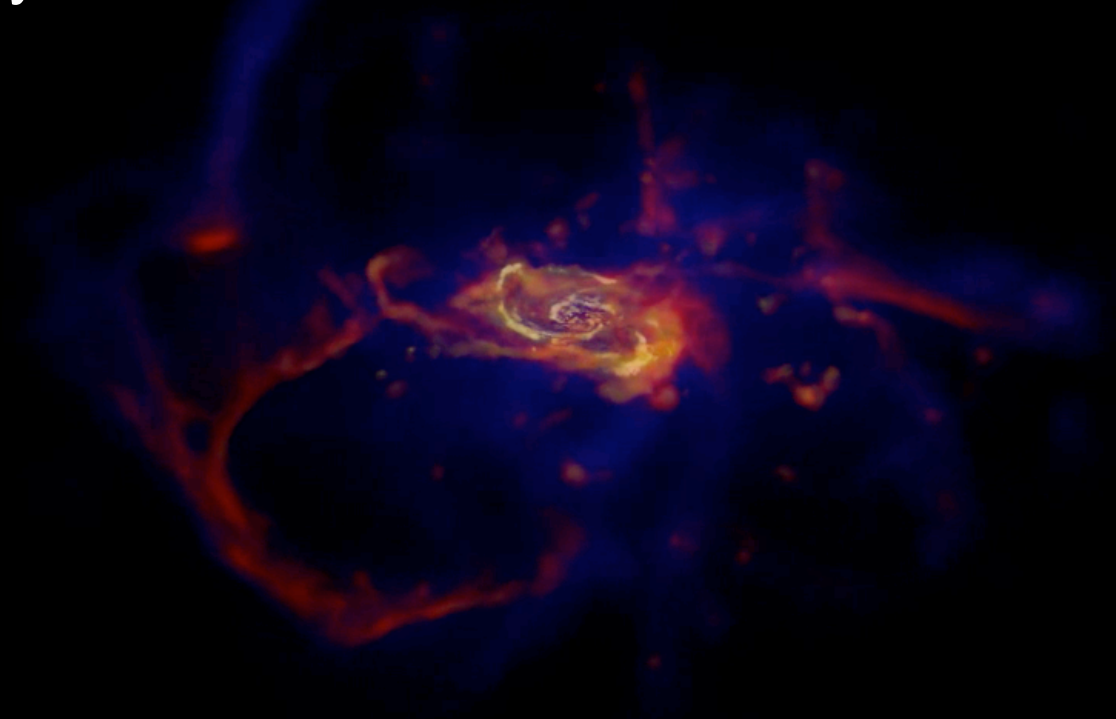
Wolfinger et al (in prep)

Test Method

- “Fake” Gaussian sources inserted into the cubes - with varying peak fluxes and widths
- Duchamp run multiple times, varying input options
- “Best” input options chosen according to number of fake sources detected, and total number of sources returned

Aside: Realistic test sources

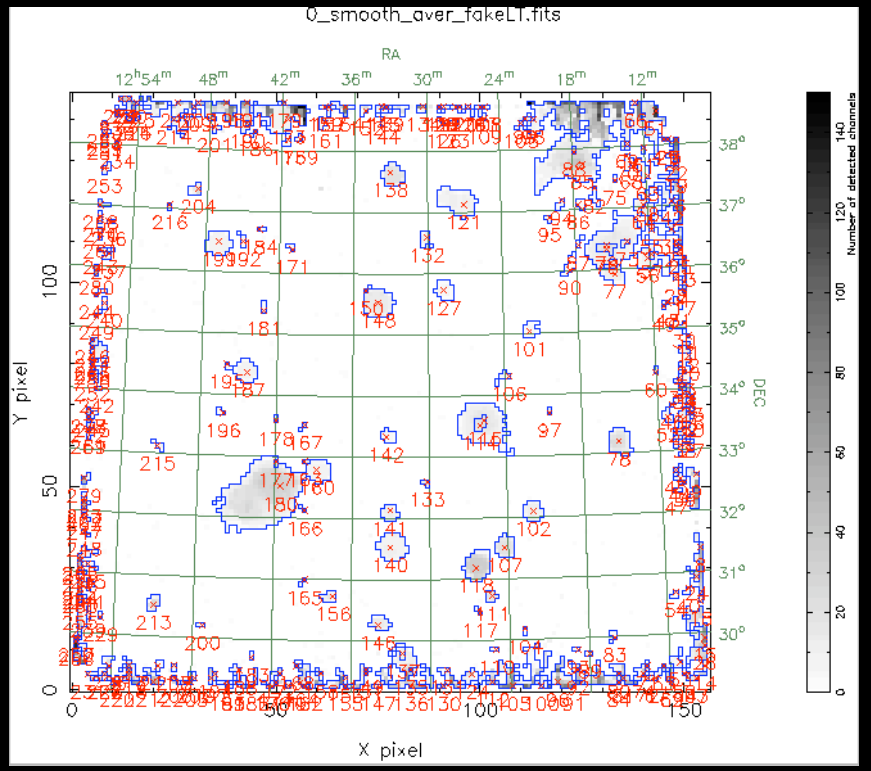
- Use simulated galaxies (or real data) in cubes - scaled for distance and varying orientation to determine recovery rate from the finders



Crain et al. 2010

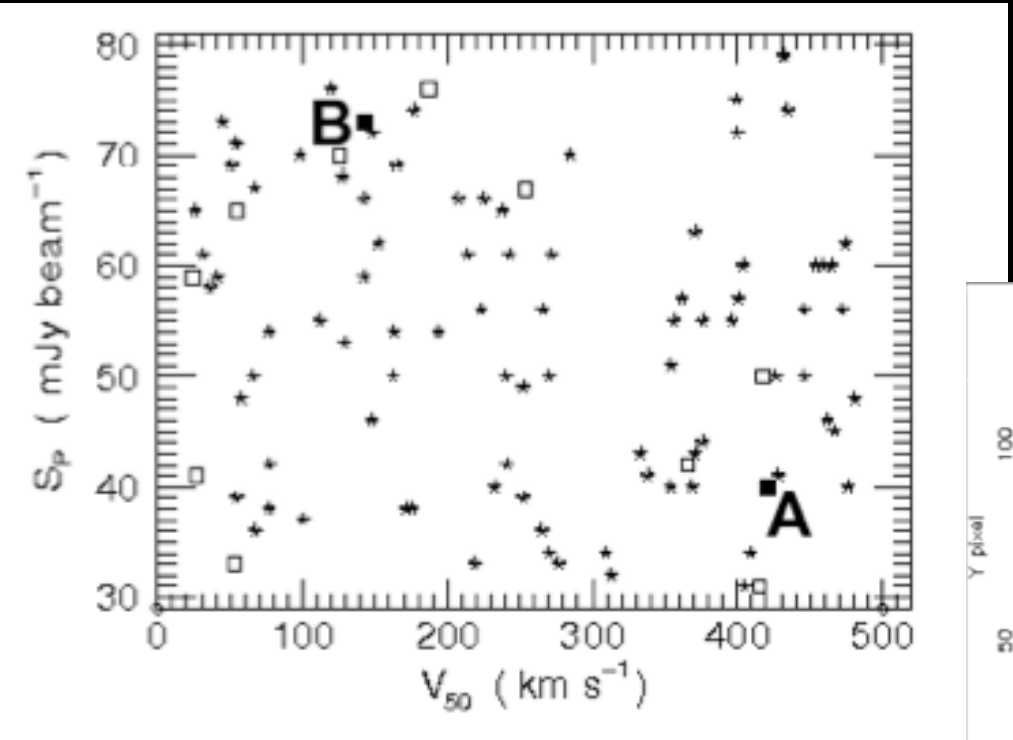
Final Duchamp search parameters

- Statistics: FDR, robust, mean/rms
- Spectral Smoothing: Hanning width = 7, threshold = 26 mJy
- A trous reconstruction: minimum channels=2, threshold = 16 mJy

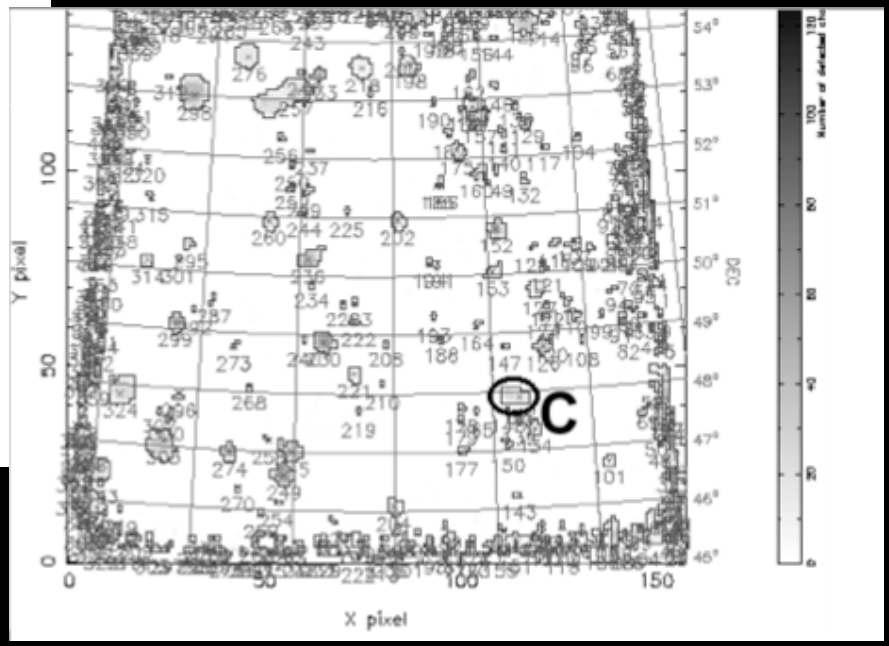


Results: Completeness

- We found 100% recovery rate of fake sources above 5σ peak flux - those missed were only due to confusion



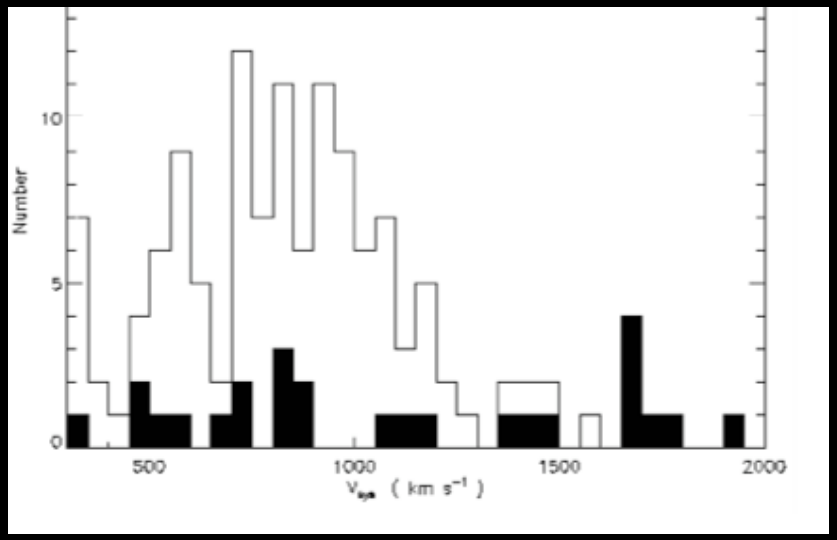
Stars: matched, squares unmatched

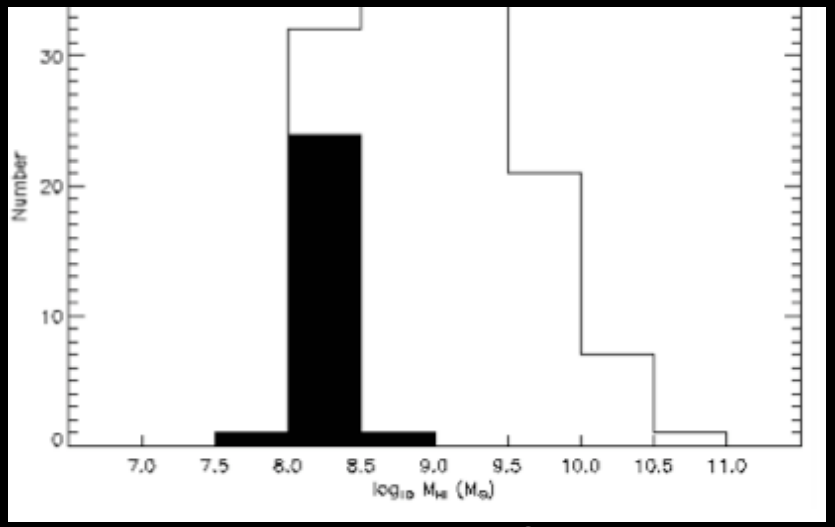


Wolfinger et al (in prep)

Results: Real data

- Duchamp run again on the 5 cubes using the *best* input parameters
- 124 HI detections found in the region - when matched with NED, 26 detections were not matched to previously catalogued optical counterparts with redshifts
- Reliability will be tested with observations of a subset of these 26 sources at the GMRT in August.

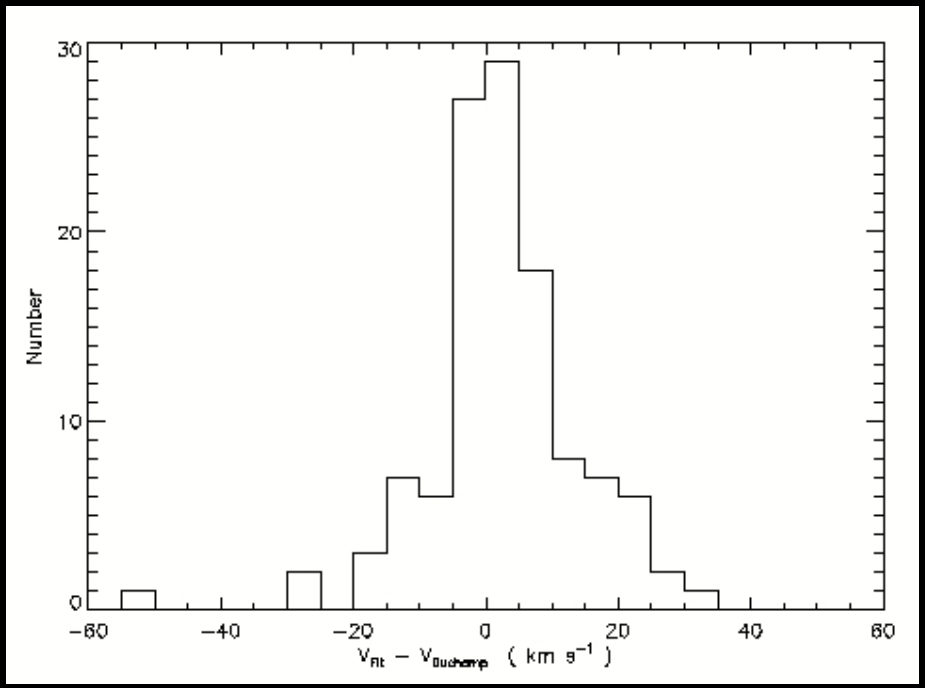




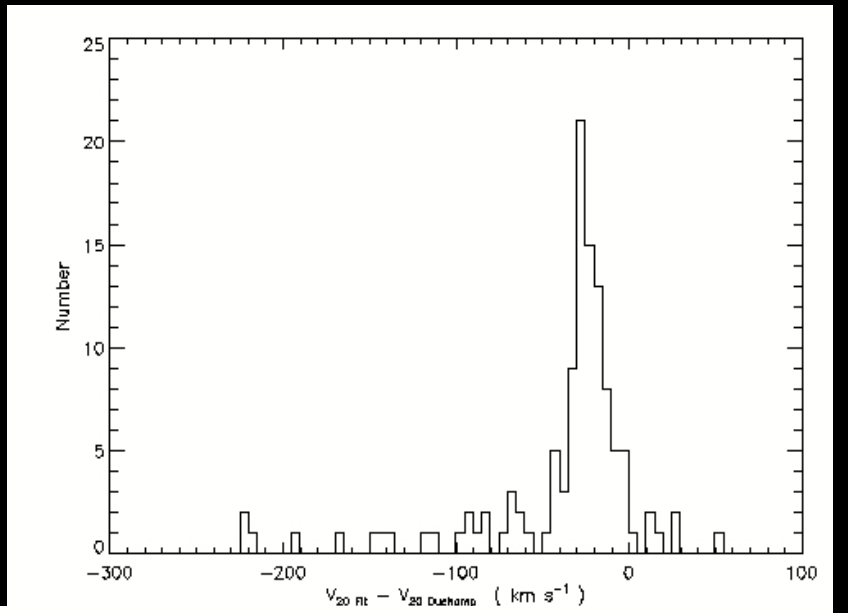
Wolfinger et al (in prep)

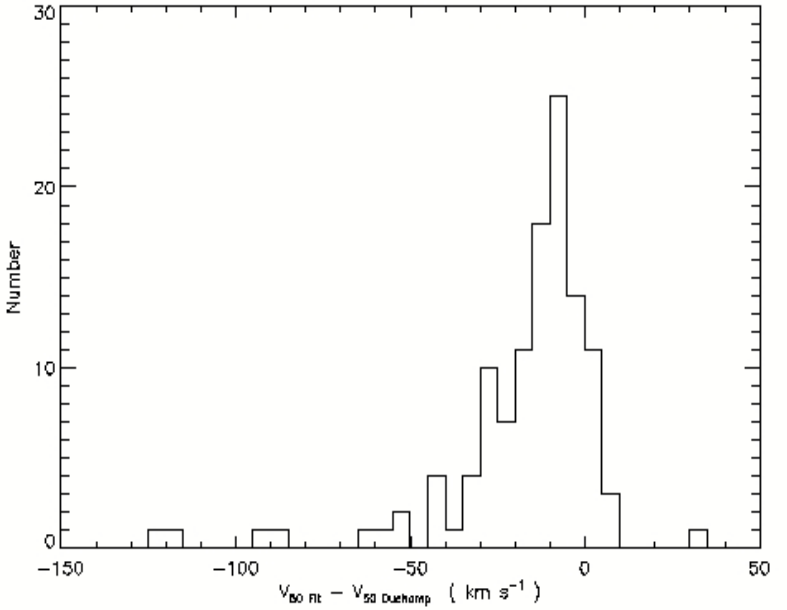
Results: Duchamp Parameterisation

- All of the “real” sources were re-parameterised using an interactive algorithm, then compared with the original Duchamp output.

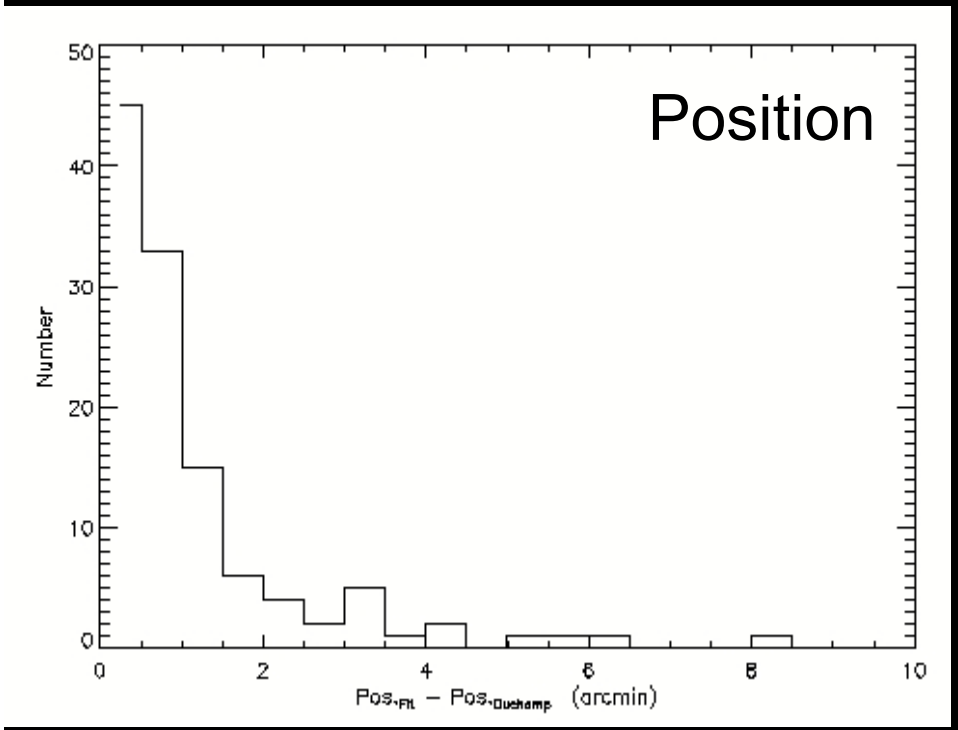


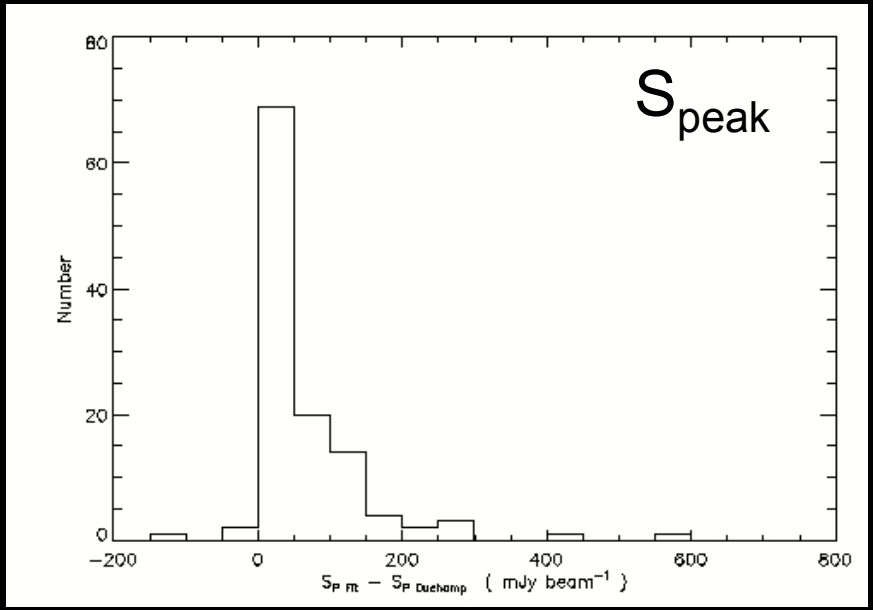
Wolfinger et al (in prep)

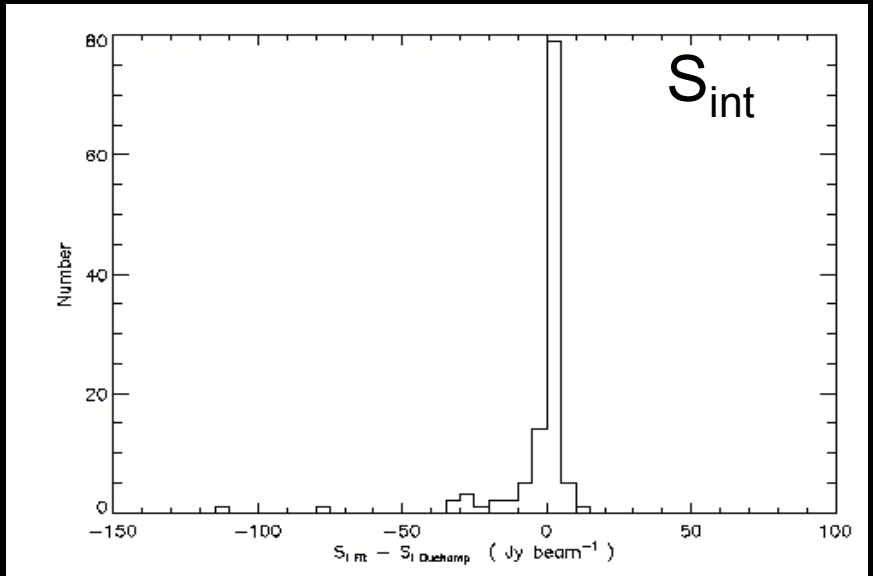




Duchamp Parameterisation- position and flux comparison



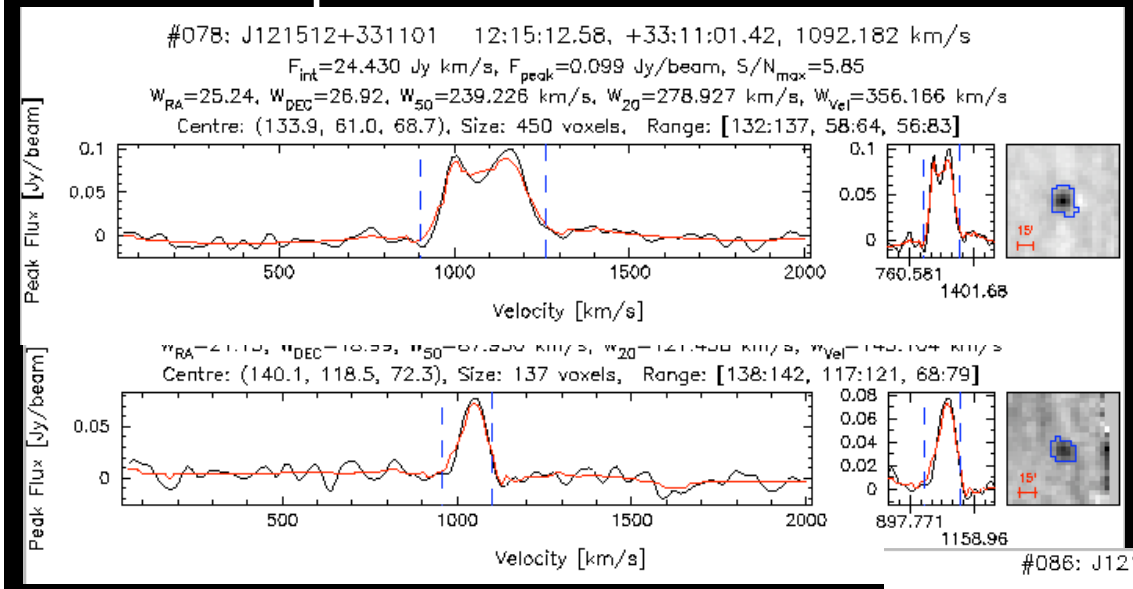




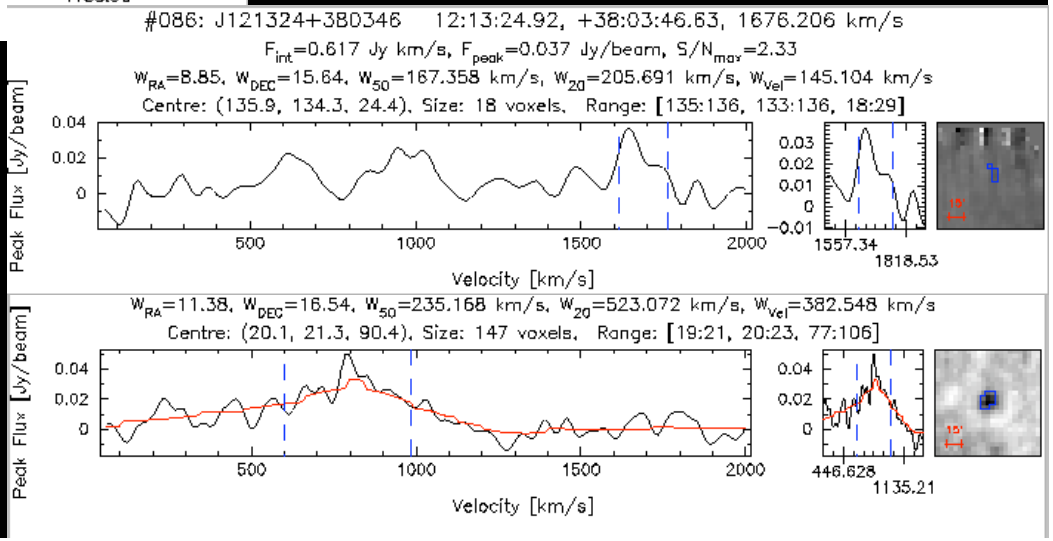
Wolfinger et al (in prep)

Example output spectra from Duchamp

Well-fit spectra



Poorly-fit spectra



Challenges for source finding in ASKAP

- Robust parameterisation (part of source finding program, or separate ?)
- Dealing with merged/confused sources

- Resolved sources
- Diffuse emission

- Local group and nearby galaxies

- Do we need to eyeball every source ??
- Re-observation of a subset of detections to determine reliability ?