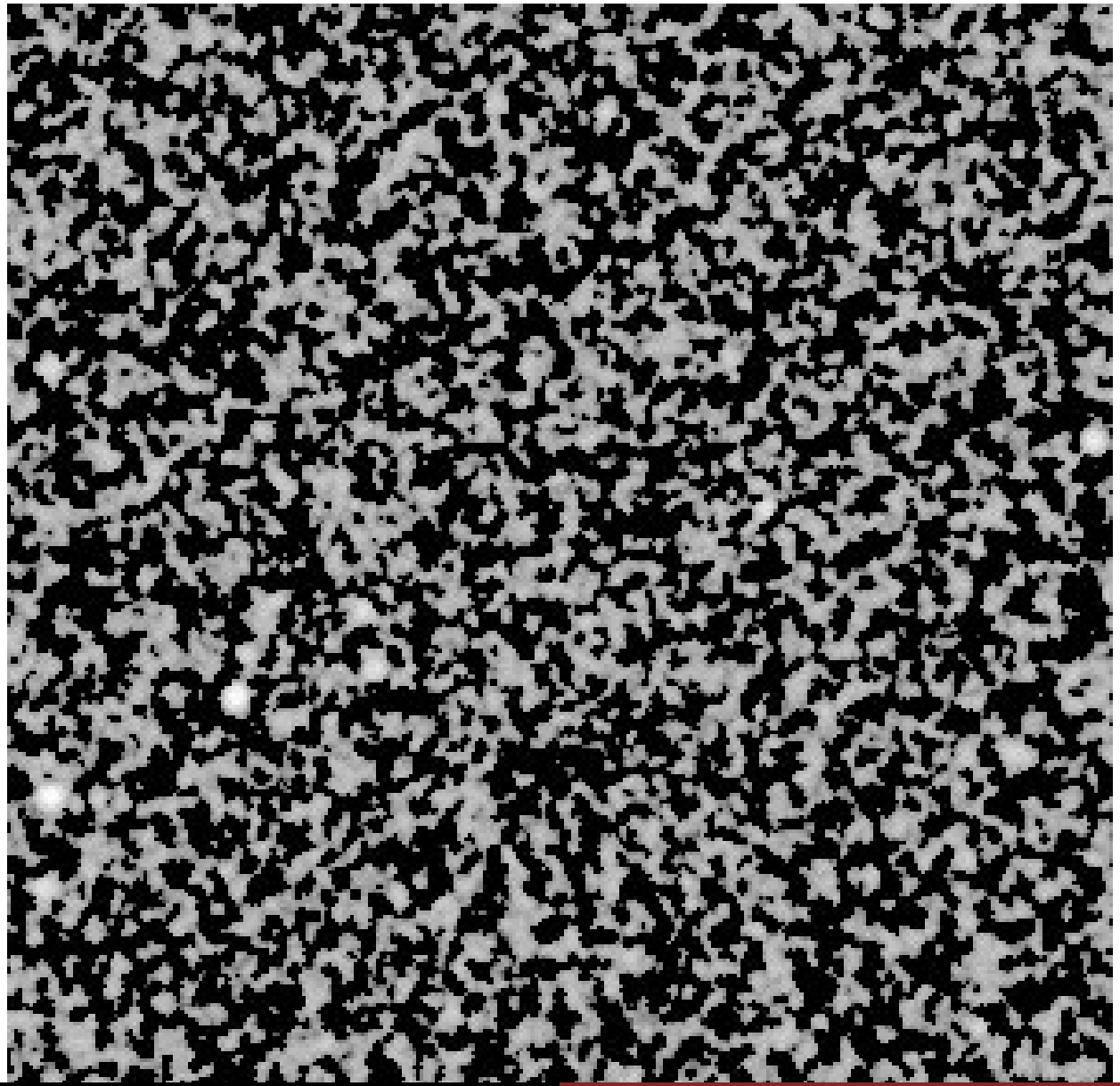
Accelerating the Astronomical Source Finding Process

Yaseen Hamdulay and Jarred de Beer

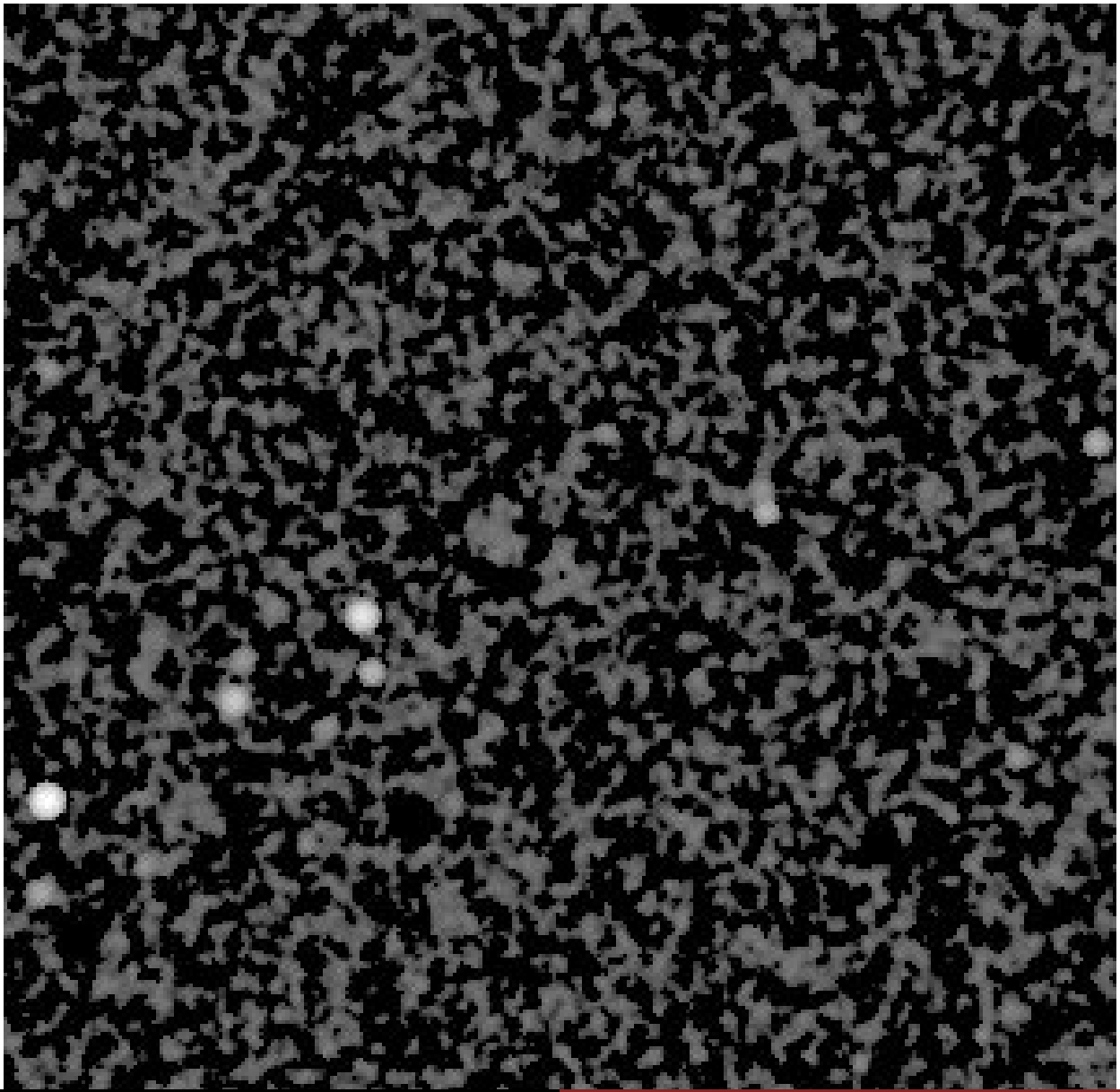
supervised by Michelle Kuttel and Sarah Blythe

26 May 2015

Survey Example: Noisy



Survey Example: Denoised



What is Radio Astronomy Source Finding?

- Process of identifying galaxies or other objects from blind surveys of the sky.
- It is made difficult by the amount of noise that gets detected.
- Traditionally done by astronomers by hand.

Automated Source Finders

- Source finders perform differently with respect to completeness and reliability as they often trade one off for the other.
- **Reliability** is the ratio of true positive detected sources to total sources.
- **Completeness** is the ratio of sources detected to actual sources.
- It is useful to have a variety of source finders depending on an astronomers work load.
- Accelerating two source finders: DUCHAMP and SoFiA's Source and Clip finder.

Research Questions

Yaseen: Can a GPU implementation of the A'trous Wavelet Reconstruction algorithm accelerate the DUCHAMP source finding process and how much speedup can be obtained?
Jarred: Can a GPU implementation of the S+C algorithm accelerate the SoFiA source finding process and how much speedup can be obtained?

GPU

- GPU's are low-cost highly parallel coprocessors.
- Difficult to program on due to highly parallel nature and unusual memory hierarchy

Automated Methods

- There exist automated source finders with DUCHAMP being the most well known.
- With the next generation of Radio Interferometers we are expecting current generation source finders to take between hours and days.
- We are proposing to use GPU's to accelerate the source finding process.

1. Implement single-threaded version of algorithm.
2. Correctness check.
3. Implement naive version of algorith on GPU.
4. Correctness check and performance comparison.
5. Accelerate.

Yaseen, DUCHAMP

- DUCHAMP is a complete source finding package that is well-known in the astronomy community
- According to Popping et al DUCHAMP performs the best in terms of completeness and reliability for point sources and one of the best for larger galaxies which makes it a good target for acceleration.

Pipeline overview

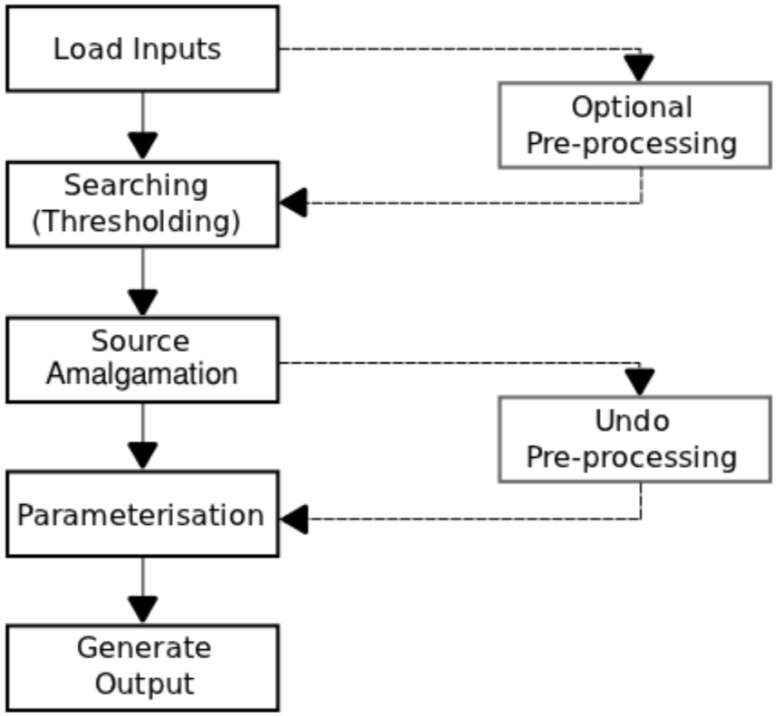


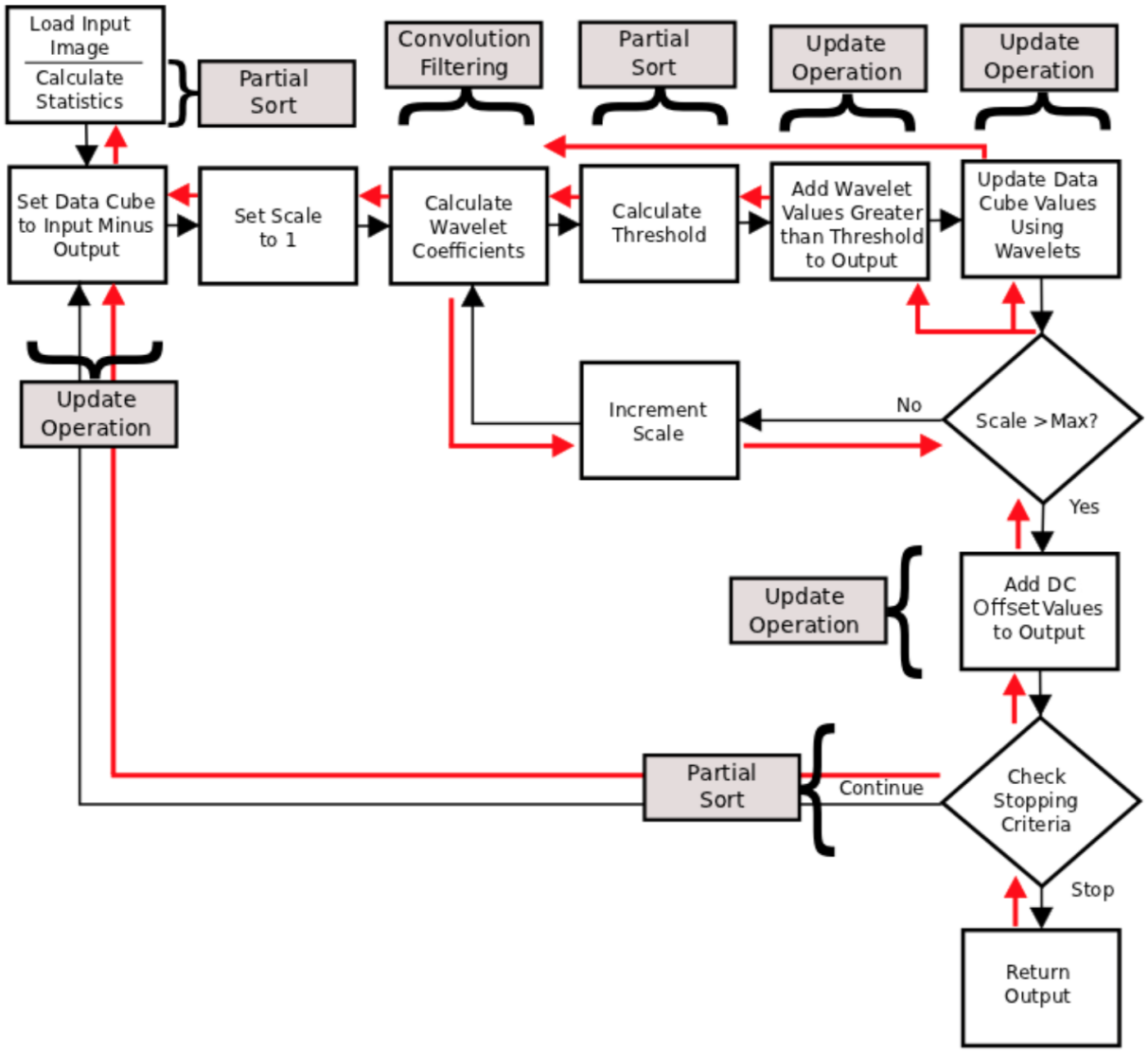
Figure : DUCHAMP pipeline

The DUCHAMP package takes a data cube and pushes it through a pipeline with data cube at one end and the parameterised sources at the other.

Preprocessing

- Between 65% and 92% of the source finding time.
- DUCHAMP uses the A'trous wavelet reconstruction algorithm.

A'trous Wavelet Reconstruction



If we have enough time

The merging algorithm is in $O(N^2)$, for data cubes with many sources this can quickly take lots of time. Implementing on GPU can potentially speed this up.

Jarred, SoFia: Smooth and Clip filter

Related Work

- Selavy, CPU parallel implementation of DUCHAMP.
- Badenhorst et al also did a CPU parallel implementation of the wavelet reconstruction filtering algorithm.
- Gary Resnick, a previous honours student, accelerated the searching (thresholding) part of the source finder.
- Parallel Gaussian Source Finder noise suppression has been ported to the GPU with massive performance improvements.

Evaluation

- The primary goal of this project is to accelerate the source finding process. Our most important metric is therefore execution speed.
- The execution time of the overall source finder depends on the data cube it is executing on. We should keep this constant or use predetermined data cubes.
- Compare against single threaded implementation.
- Ensuring correctness is of utmost importance.