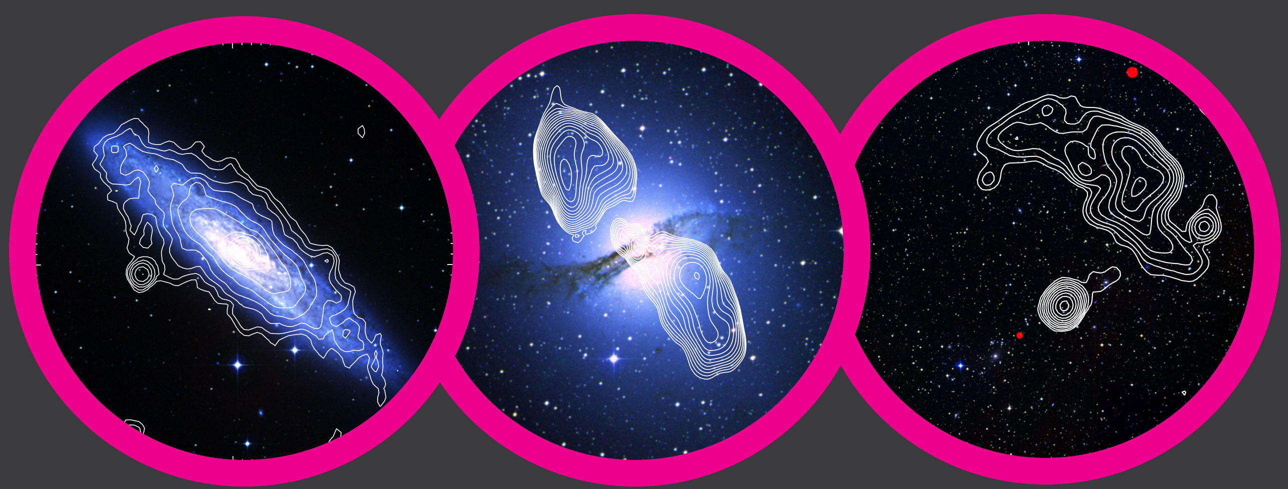
FINDING SOURCES IN THE NOISE.

The Square Kilometre Array will be the largest radio telescope ever built. Spanning Australia and South Africa, thousands of receivers will be joined together using radio interferometry to produce the most detailed radio images of our universe we've yet seen.

It will also produce huge amounts of data, upwards of an exabyte a day, every day. This poses a problem, because this can't all be stored. But radio images are mostly noise, so we can sift through to find just the interesting bits.

The problem is finding all the interesting sources in this noise, quickly and accurately.

1. WHAT'S OUT THERE?



The Sculptor Galaxy showing supernova remnants and other sources

Galactic 'lobes' of Centaurus A, produced by a supermassive black hole

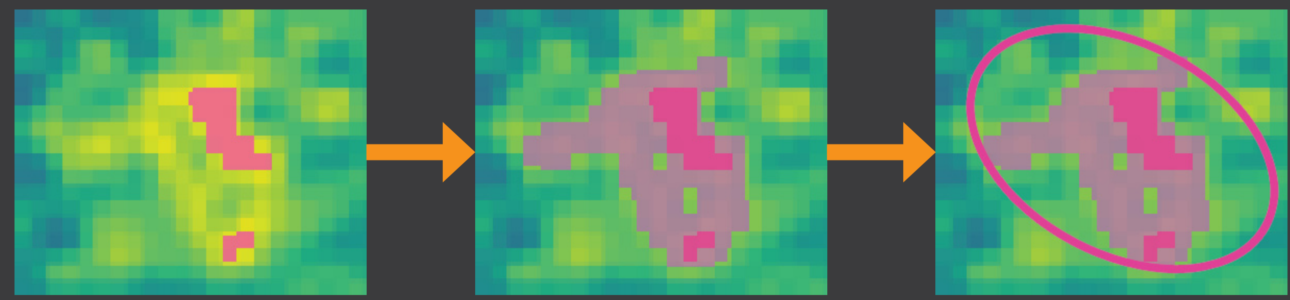
Relics from an ancient collision between two galaxy clusters on the periphery of Abell 3667

Radio telescopes operate at wavelengths much lower than optical telescopes. We see a different universe at these frequencies. At optical wavelengths (400–700nm), we see objects because they are hot and emit blackbody radiation.

At the lower frequencies of radio telescopes there are other emission processes at play. We see the otherwise invisible spectral emission lines of neutral hydrogen that extends throughout the Milky Way; we see synchrotron radiation emitted by charged particles that lets us see the effects of rotating black holes, or the long since forgotten collision of two galactic clusters.

2. ON THE THRESHOLD

Bright point sources like stars and galaxies aren't so hard to pick out. There are already source finding tools for these which use 'thresholding algorithms'.

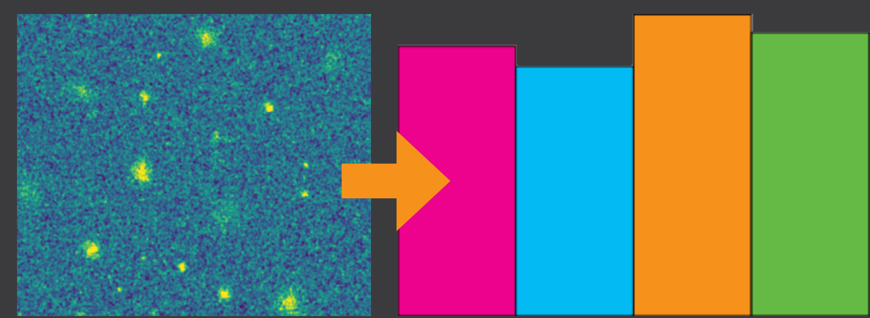


1. Find the brightest pixels in the image to create 'seeds'
2. Grow the seeds into islands by expanding them to a threshold brightness
3. Fit the islands with an elliptic gaussian to find things like the total flux.

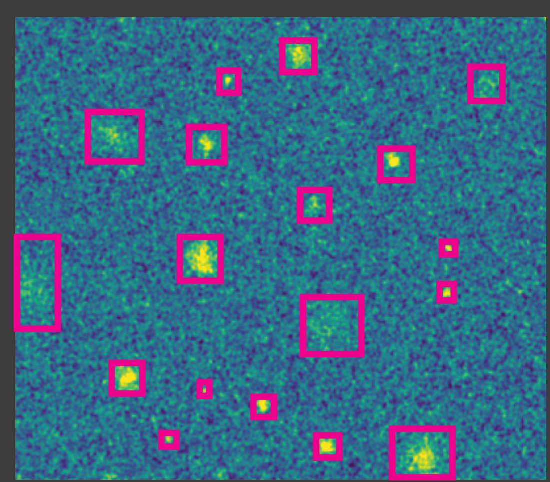
3. ENTER ODDITY

Diffuse, faint large scale objects—which often are the most interesting—are much harder to find. Tony Yeoman and supervisor Marcus Frean at Victoria University developed an algorithm for finding these types of sources using Baysian statistics, called 'Oddity'.

1. Assign each pixel value to a small number of 'bins' and create a histogram of the radio image.



2. Divide the image into a grid. For each tile, create a histogram of just its pixels, and subtract these from the overall histogram.
3. Using Baysian statistics and a 'multinomial generative model', find the likelihood that the tile and the overall background were generated by separate sources.



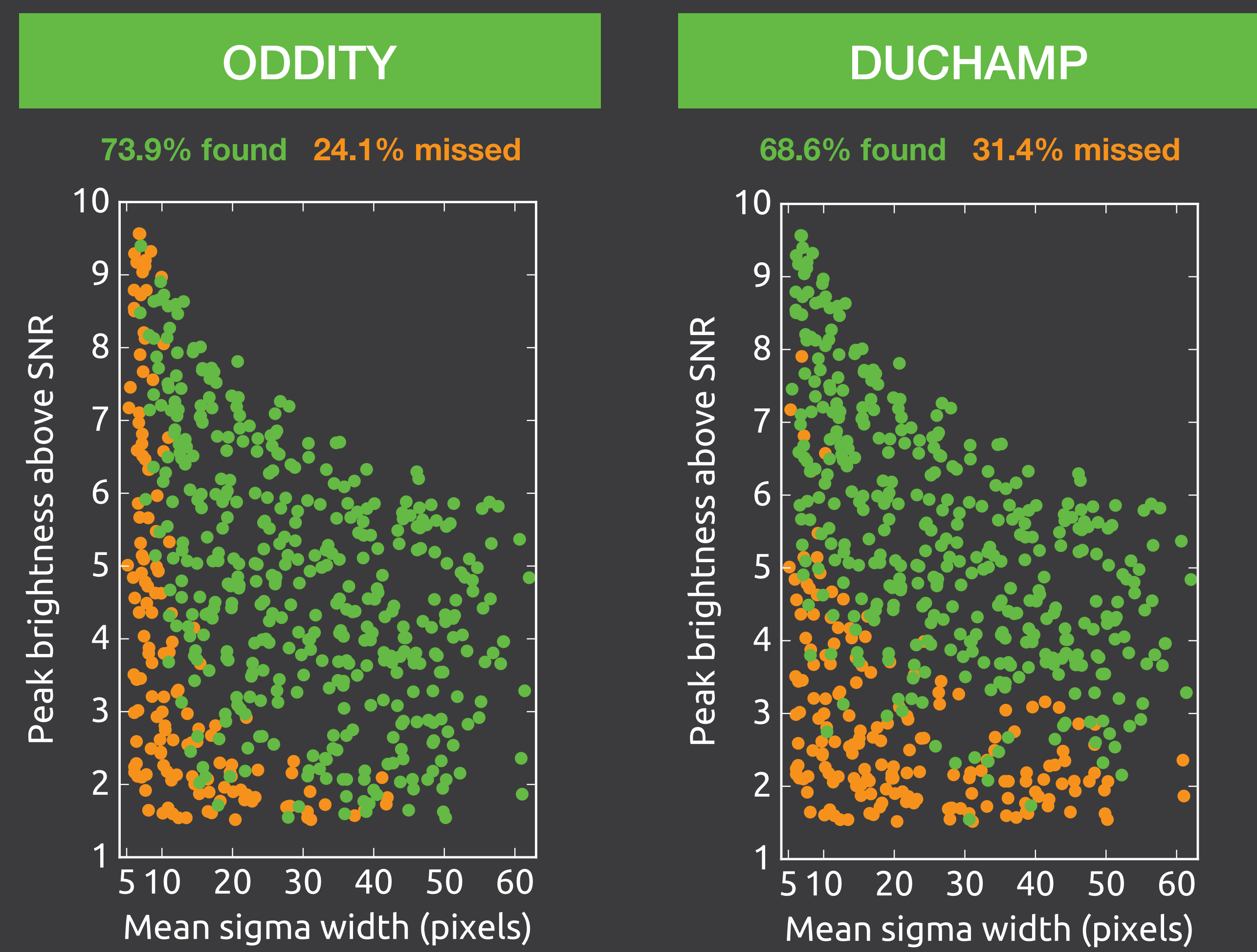
4. Here's the trick. If the the likelihood was 60%, then there's a 60% chance we leave the tile 'turned on', with its pixel count left subtracted from the background. In this way, we cycle through the whole grid hundreds of times flicking individual tiles on and off until we come to an optimal, steady state. This is called a Monte Carlo Markov Chain.
5. Finally, we merge nearby tiles that are 'on'. And then we tweak the location of their edges many thousands of times, trying to find an optimal fit around the source.

4. THE COMPETITION

Oddity does things very differently, but how does it stack up compared to the competition? We created a series of artifical radio images, with a range of elliptical gaussian sources, differing in their size and brightness. We then overlaid this with noise.

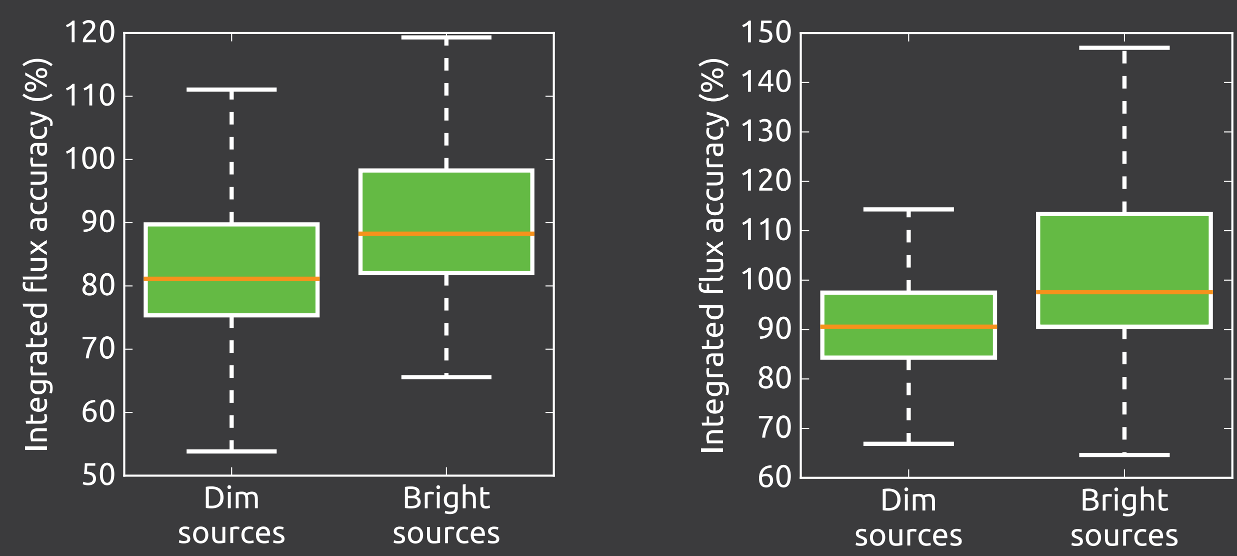
The rules were simple: find the largest number of sources with fewer than 1% false positives.

The leading contender was a thresholding algorithm called Duchamp. How did it compare?



Median integrated flux accuracy across all sources was 84% of the true value.

Median integrated flux accuracy across all sources was 94% of the true value.



Each algorithm has its strength. Oddity does well for finding large, diffuse and very dim sources, but struggles with small bright sources due to the lack of pixels required to make statistically meaningful inferences. On the other hand, threshold algorithms like Duchamp do well on bright images of all sizes, but struggle with large yet dim sources, since these lack the required 'seed' pixels.

Development of Oddity continues in preparation for the Square Kilometre Array.

Poster and research prepared by Torrance Hodgson, under supervision of Melanie Johnston-Holitt (School of Chemical and Physical Sciences) for the VUW Summer Research Scholarship program 2017.