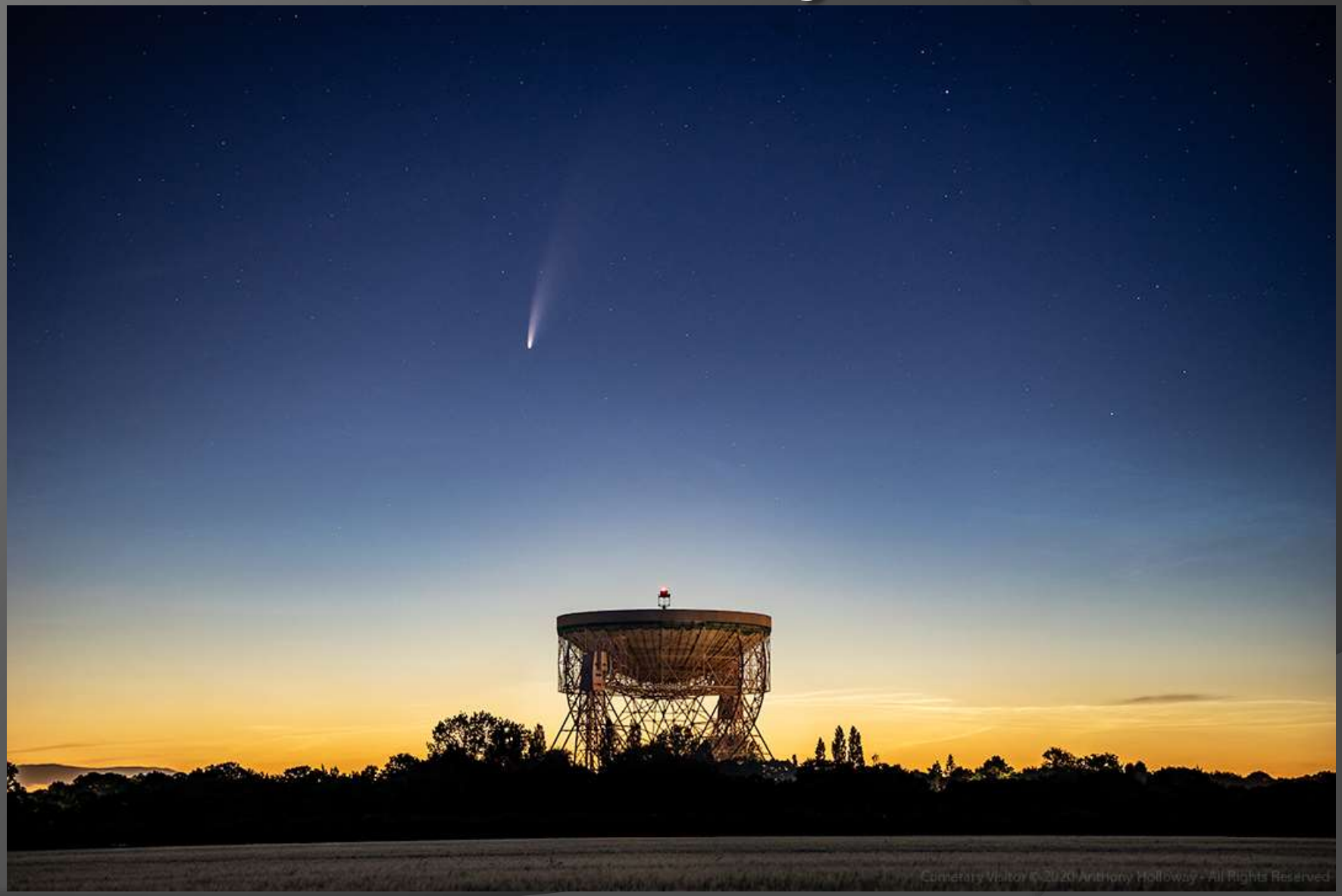
Radio Astronomy with IRIS



Dr Anthony Holloway, Jodrell Bank Centre for Astrophysics, University of Manchester

Jodrell Bank Centre for Astrophysics

e-MERLIN / VLBI National Facility - The UK's facility for high resolution radio astronomy observations, operated by JBCA and The University of Manchester for the Science and Technology Facilities Council.

ALMA UK Regional Centre, operated by JBCA and The University of Manchester for the Science and Technology Facilities Council.

Square Kilometre Array Organisation

MeerKAT

South African Radio Astronomy Observatory (SARAO)

People & User Communities

- e-MERLIN/VLBI National Facility
- ALMA UK Regional Centre
- JBCA SKA postdocs
- JBCA academic staff, postdocs, postgrads

- International Collaborations

2018/19

- 7 nodes
- Dell PowerEdge 640 machines
- Dual Xeon Gold 5122 CPU
- 4 core / 8 thread each => 8 C / 16 T
- 3.6 GHz (Turbo 3.7 GHz)
- 1.5 TB RAM
- 4 x 12 TB drives in RAID 10 configuration = 24 TB usable per machine

2019/20

- 10 Lenovo SR650 machines – 2U
- 4 x 12 TB drives in RAID 10 configuration = 24 TB usable per machine
- 2 x 240 GB SSD in RAID 1 for OS

- 9 of the nodes
- Dual Xeon Gold 5222 4C/8T (8C/16T) 3.8 GHz (Turbo 3.9 GHz)
- 1.5 TB RAM

- 1 node has
- Dual Xeon Gold 5215L 10C/20T (20C/40T) 2.5 GHz (Turbo 3.4 GHz)
- 3 TB RAM (Dual CPU's capable of 9 TB)

Blackett facility

- Blackett is an umbrella term for GridPP/IRIS/locally funded machines hosted by School of Physics and Astronomy in IT Services machine room
  - Provides the GridPP/WLCG Tier-2 Service, local capacity to HEP, and since 2018, IRIS capacity too
- CPU logical processor / job slot counts
  - 4992 Skylake including 3072 IRIS (“UKT0”) 192GB (3GB/proc)
  - 1280 Broadwell 64GB (2GB/proc)
  - 2352 Westmere 48GB (2GB/proc)
  - Interfaces ARC-CE/HTCondor and Vac; OpenStack in progress
- Storage 6.06PB total
  - Provided by 67 storage hosts
  - DOME exposing Grid protocols (GridFTP, xroot, WebDAV)
- Also various server machines hosting VM “pets” providing local/GridPP/IRIS services
- Currently dedicated 10Gb/s connectivity with a backup, directly to JISC network (previously Net Northwest)
  - ie bypasses the main University firewall, routers, etc

Interferometry

Interferometry is a powerful tool in astronomy that links together one or more pairs of radio antennas, even those thousands of kilometers apart, to create a new and vastly more powerful "virtual" telescope called an interferometer.

Interferometers harness the space between the antennas: the **larger the spacings, the higher the resolving power,** allowing it to see finer and finer details, like the zoom lens of a camera.

How Is This Done?

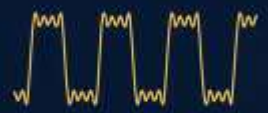
Astronomers reconstruct images of an object in space using interferometers, telescopes that observe the Fourier transform of an object's brightness pattern on the sky.

Inteference Pattern

Wave patterns from an interferometer are similar to the wave patterns created when light passes through a pair of slits. In radio astronomy, the antenna pair takes the place of the two slits, but the resulting patterns are similar.

Fourier Transform

The Fourier transform is a mathematical tool that deconstructs any signal into a sum of sine waves.



A sine wave in 2 dimentions looks like a set of stripes

More Antennas = Clearer Picture

Turning this pattern into an image takes many hours of observations. Like a time-lapse exposure, this slowly builds up an image of even a very dim source. It also allows Earth's rotation to, in effect, fill in the empty spaces in the array to produce a more complete picture.

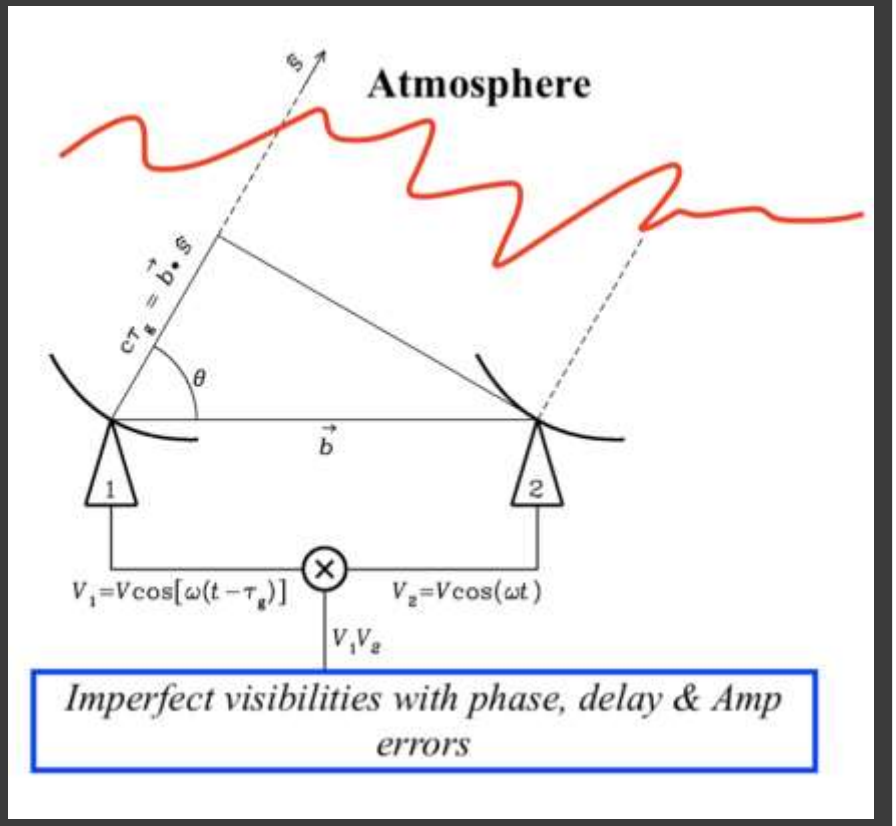
2 Antennas

The signals received at each antenna must be matched wave for wave, even for antennas that are half a world away. Atomic clocks at each site allow for their observations to be mathematically combined using a specialized supercomputer called a **correlator.**

Credit: NRAO/AUI/NSF; S. Dagnello

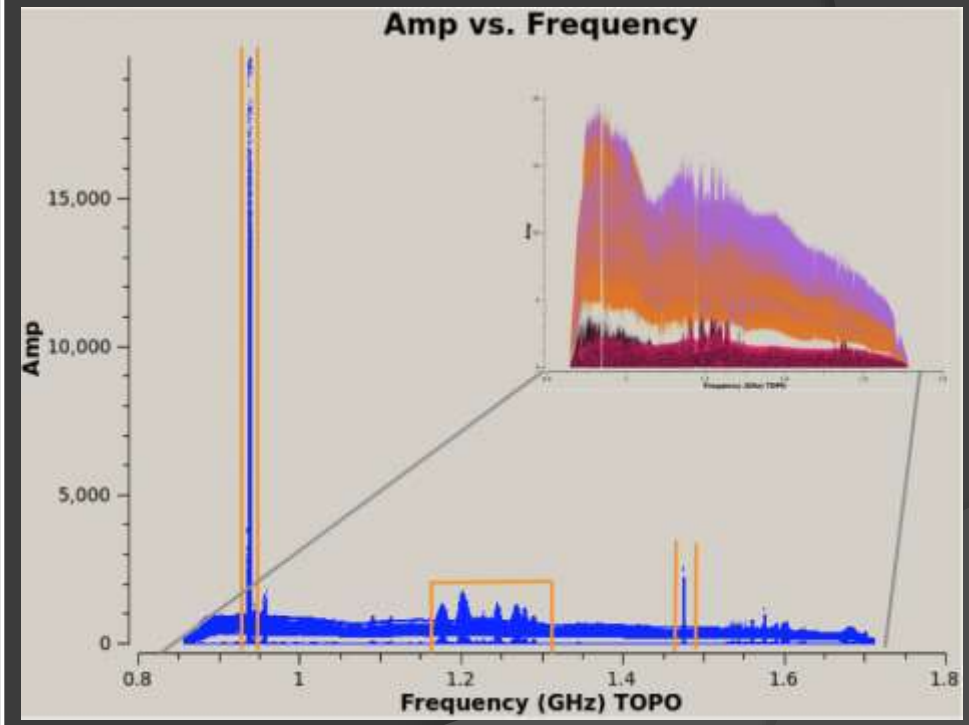
MEERKAT PROCESSING

Data received from an interferometer



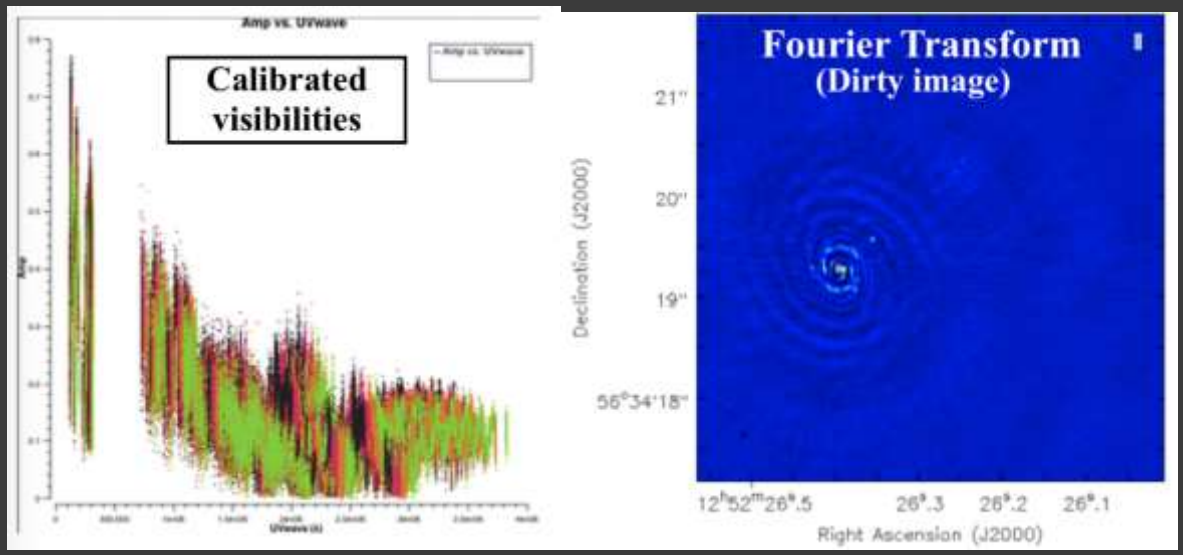
The data received is imperfect – Calibrate it

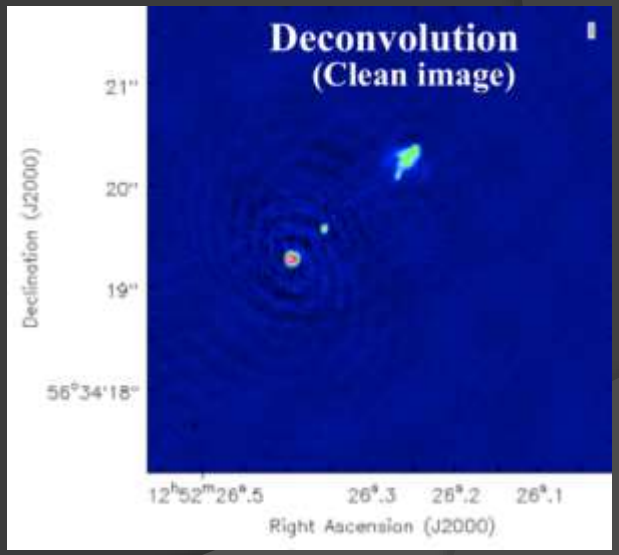
But first clean it - flagging



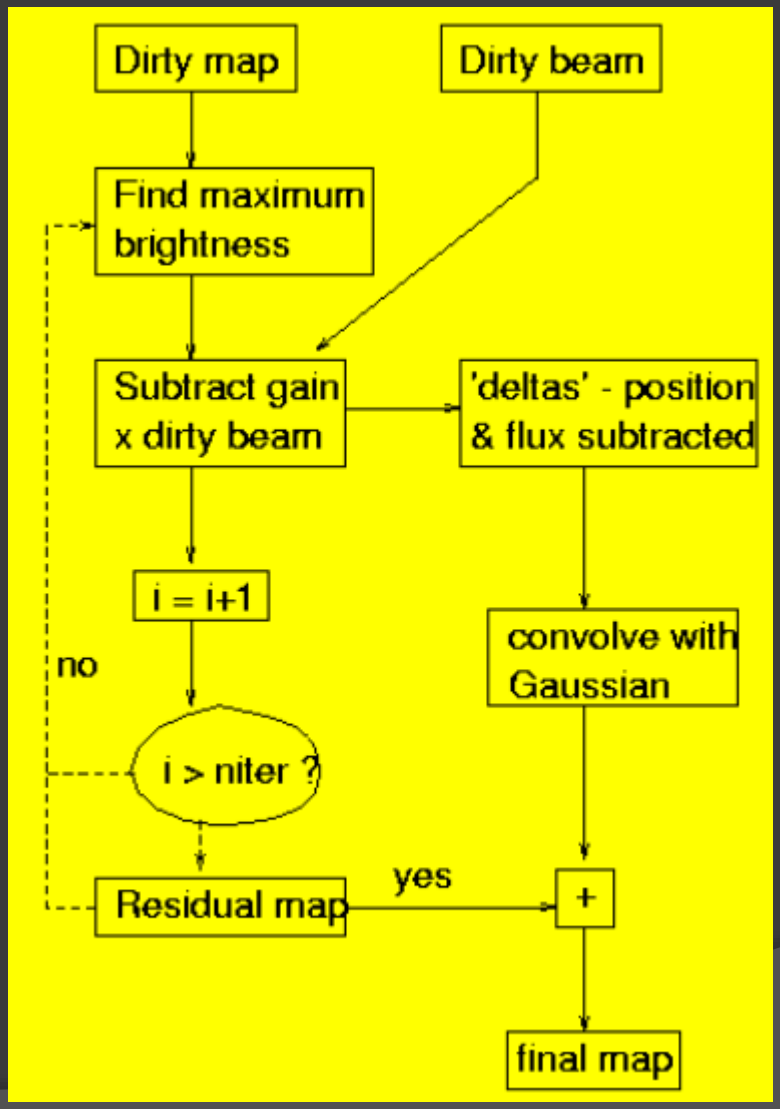
PRIMARY CALIBRATION

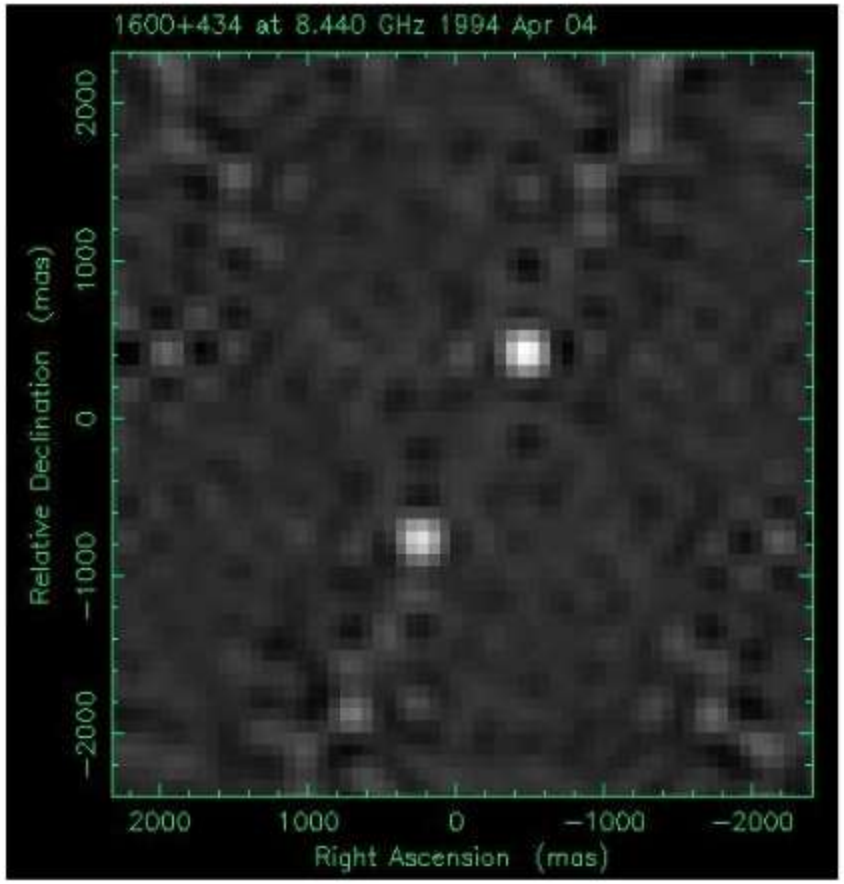
IMAGING



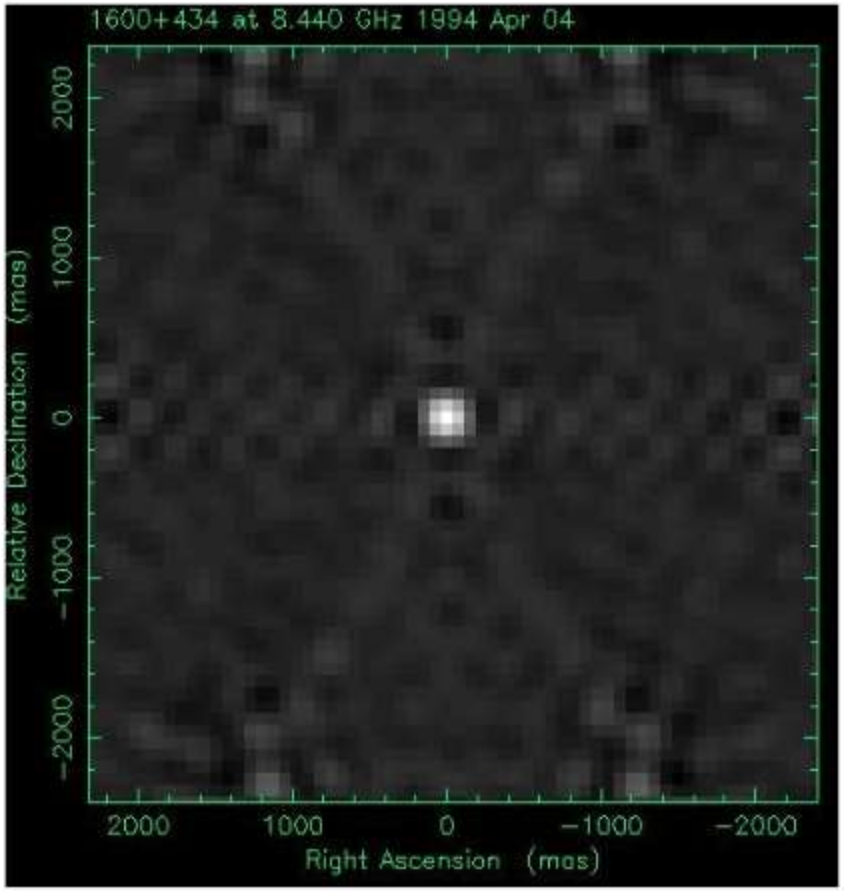


Deconvolution



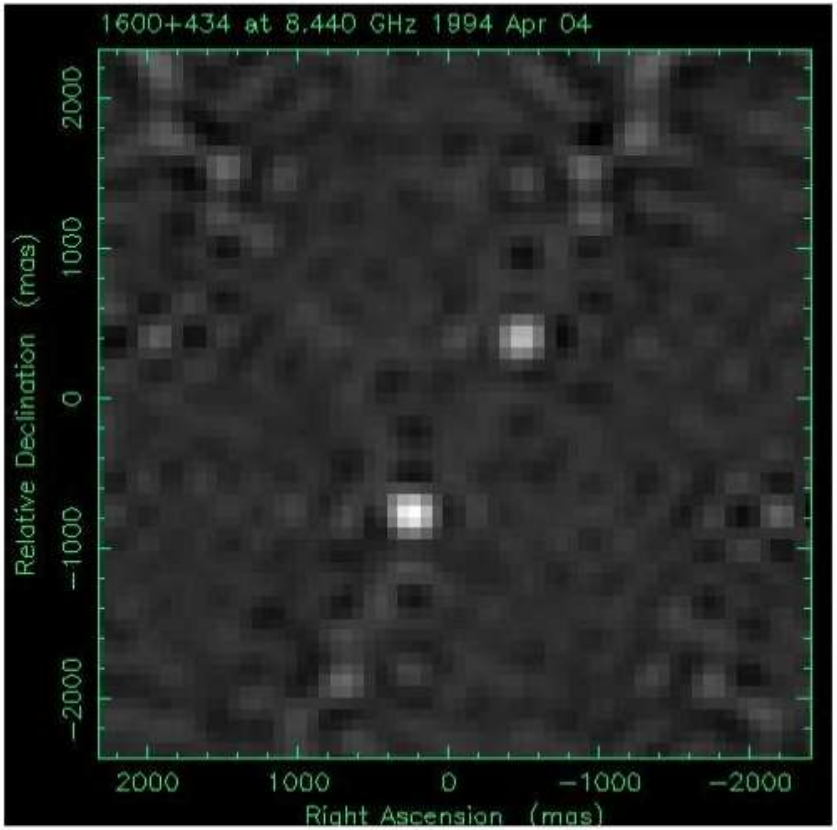


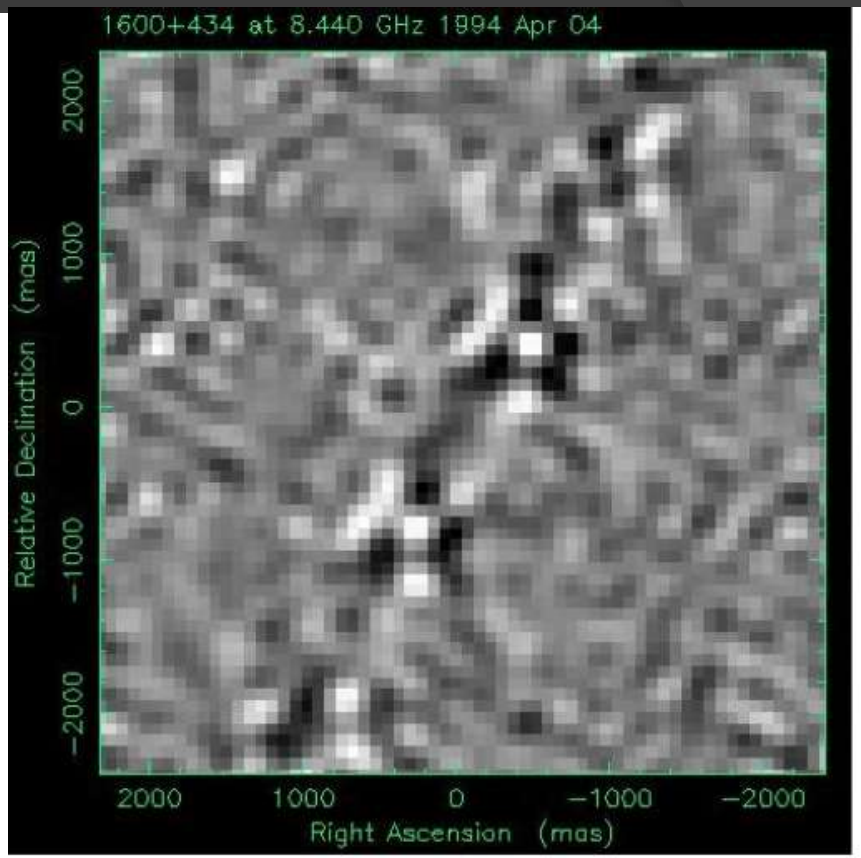
Dirty map



Dirty beam

Jackson/Muxlow

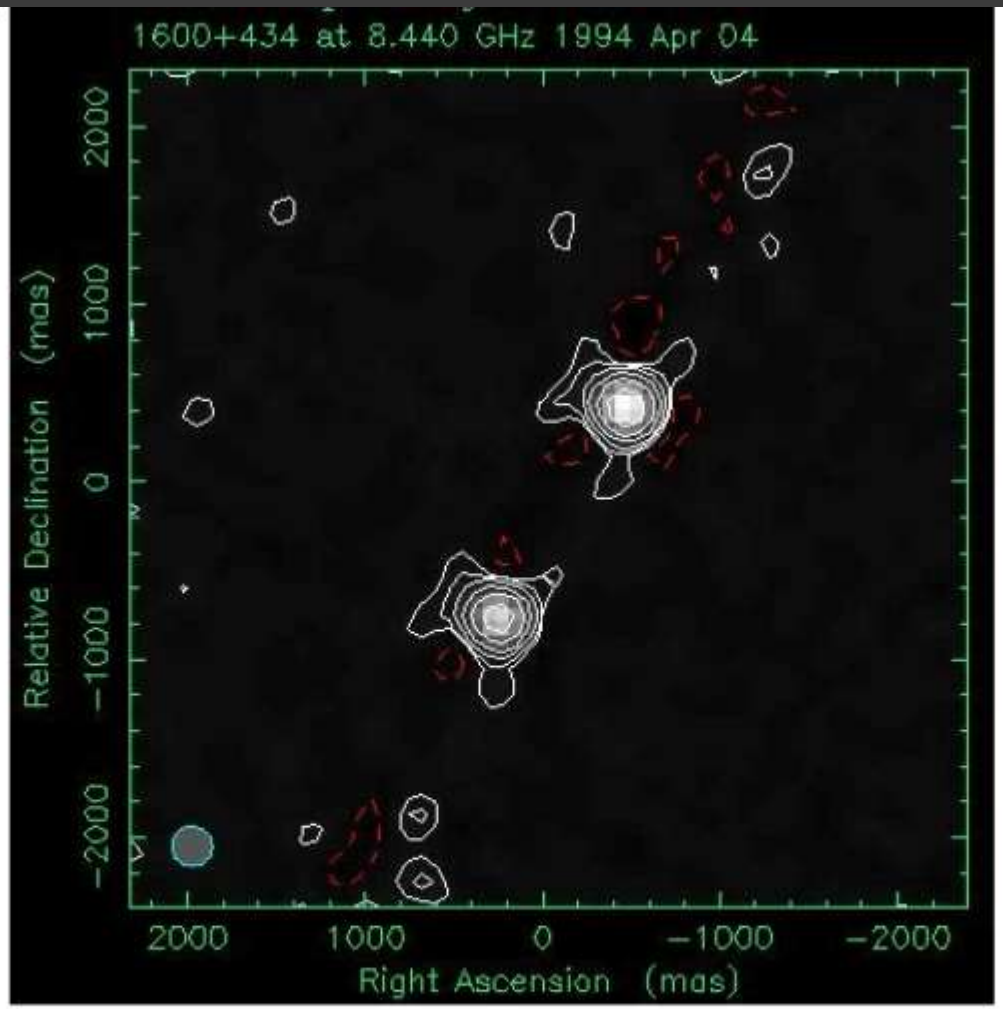




Residual after 1 CLEAN (gain 0.5)

Residual after 100 CLEANs (gain 0.1)

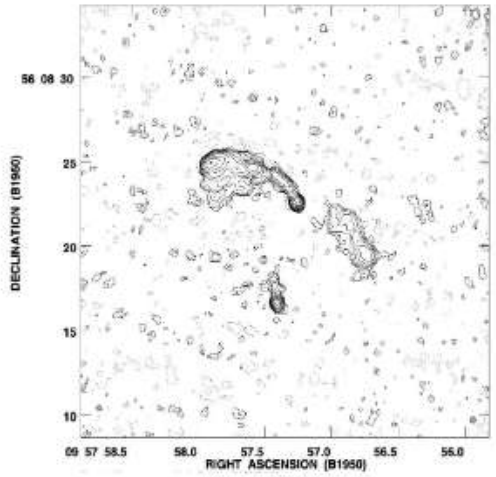
Jackson/Muxlow



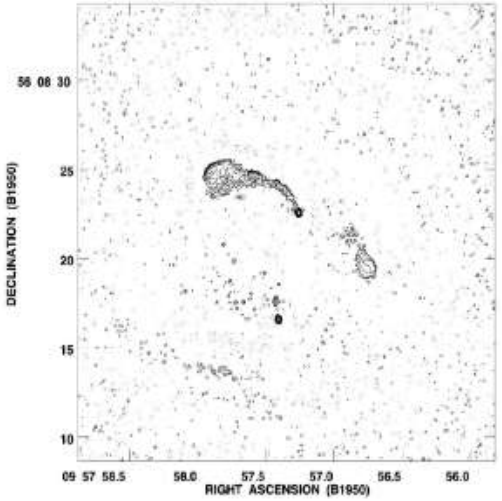
CLEAN map (residual+CCs) after 100 CLEANs (gain 0.1)

Jackson/Muxlow

Data weighting in the u-v plane



natural



uniform

Generally more u-v tracks on inner part

Can choose to

* weight all data equally (natural)
  - gives best S:N, less good beam

* weight all u-v grid points equally (uniform)
  - gives good resolution, less S:N

* Compromises possible
  - Briggs "robust" parameter -5 → 5

Jackson/Muxlow

Basic CLEAN

More advanced routines:-

- CASA tclean
- Multi scale clean
- Wide field clean
- RASCIL

CASA

National Radio Astronomy Observatory
Enabling forefront research into the Universe at radio wavelengths

Log On | Visit Public Website | Contact Us

Search NRAO... Go

Home | About NRAO | Science | Research Faciliti | Observing | Opportunities

- About CASA
- News & Events
- Download
- Documentation
- Getting Help
- Acknowledgements

CASA Docs: Official documentation

Help us improve CASA: short user survey

CASA logos/promotion

Search CASA

Go

Contact us

About CASA

CASA, the *Common Astronomy Software Applications* package, is being developed with the primary goal of supporting the data post-processing needs of the next generation of radio astronomical telescopes such as ALMA and VLA. The package can process both interferometric and single dish data. The CASA infrastructure consists of a set of C++ tools bundled together under an iPython interface as data reduction tasks. This structure provides flexibility to process the data via task interface or as a python script. In addition to the data reduction tasks, many post-processing tools are available for even more flexibility and special purpose reduction needs.

CASA
Common Astronomy Software Applications

CASA is developed by an international consortium of scientists based at the National Radio Astronomical Observatory (NRAO), the European Southern Observatory (ESO), the National Astronomical Observatory of Japan (NAOJ), the Academia Sinica Institute of Astronomy and Astrophysics (ASIAA), the CSIRO division for Astronomy and Space Science (CASS), and the Netherlands Institute for Radio Astronomy (ASTRON) under the guidance of NRAO.

- CASA Docs
- CASA Guides
- Newsletter
- Download
- Helpdesk
- Subscribe

OVERVIEW OF WORK ON IRIS HIGHMEM NODES

Contributions from Priyaa Thavasimani, Willice Obonyo, Iulia Cimpan & Naomi Asabre Frimpong.

Work on the high memory grid nodes is currently split into two projects:

1. Implementing the processMeerKAT pipeline for processing data from the SKA precursor telescope MeerKAT. This work focuses on having a fully operational pipeline available for users asap, which provides equivalent outputs to the South African pipeline.

2. Testing and developing the RASCIL SKA SDP software. This is a development project, working with the software intended to be implemented for the SKA itself and uses data from multiple radio telescopes.

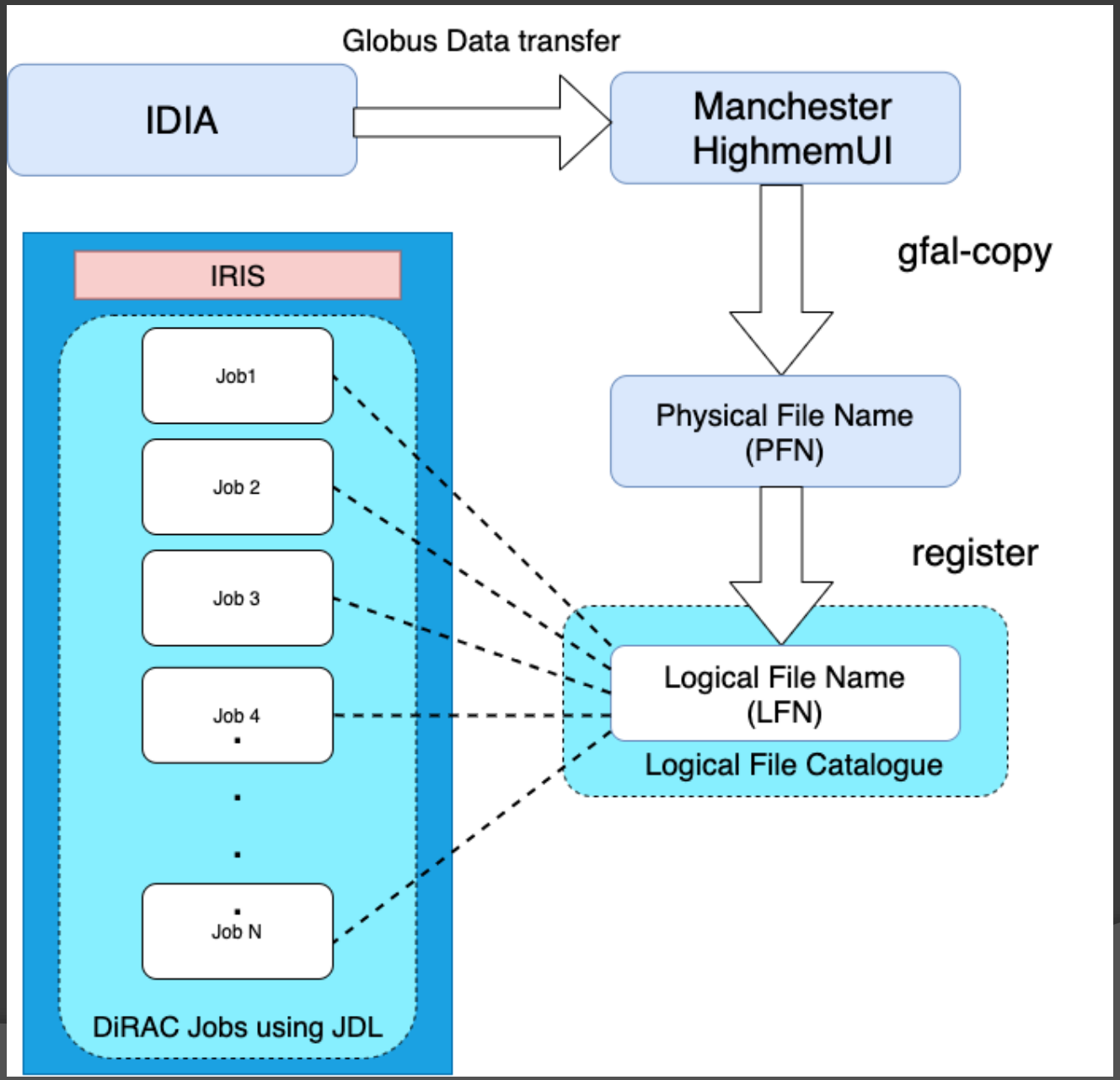
The next project for the high memory nodes will be to implement the LOFAR Long Baseline processing pipeline, currently running at SurfSARA.

THE MEERKAT TELESCOPE

- Located in the Karoo desert - SA
- A precursor for the SKA-mid
- An Array of 64 dishes
- Each dish has a diameter of 13.5m
- Maximum Baseline = 8 km (res ~6.0” at L)
- Available frequency bands receivers

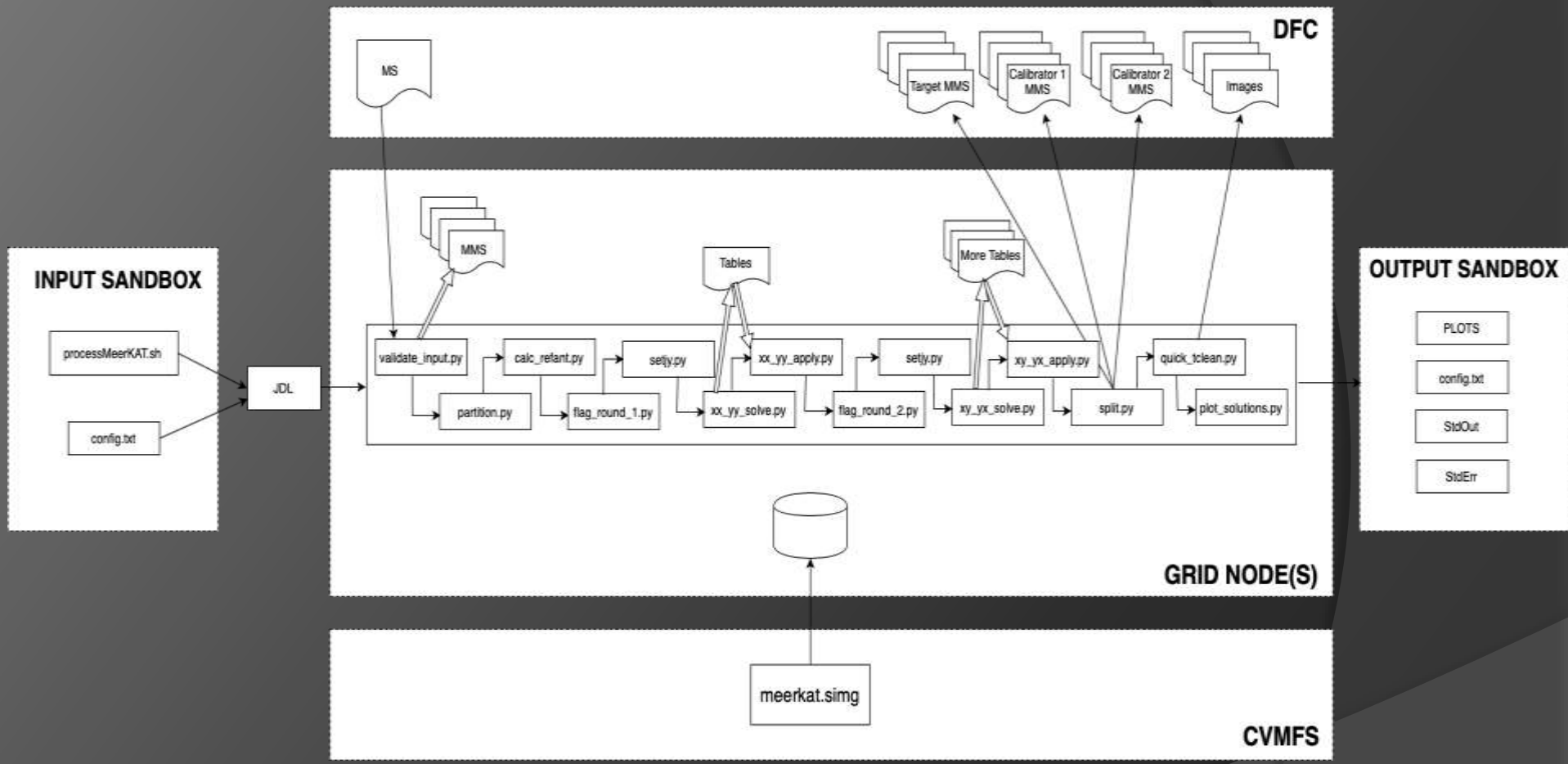
  0.58 – 1.015 GHz
  1 – 1.75 GHz
  8 – 14.5 GHz – not yet

IRIS IMPLEMENTATION OF MEERKAT PIPELINE

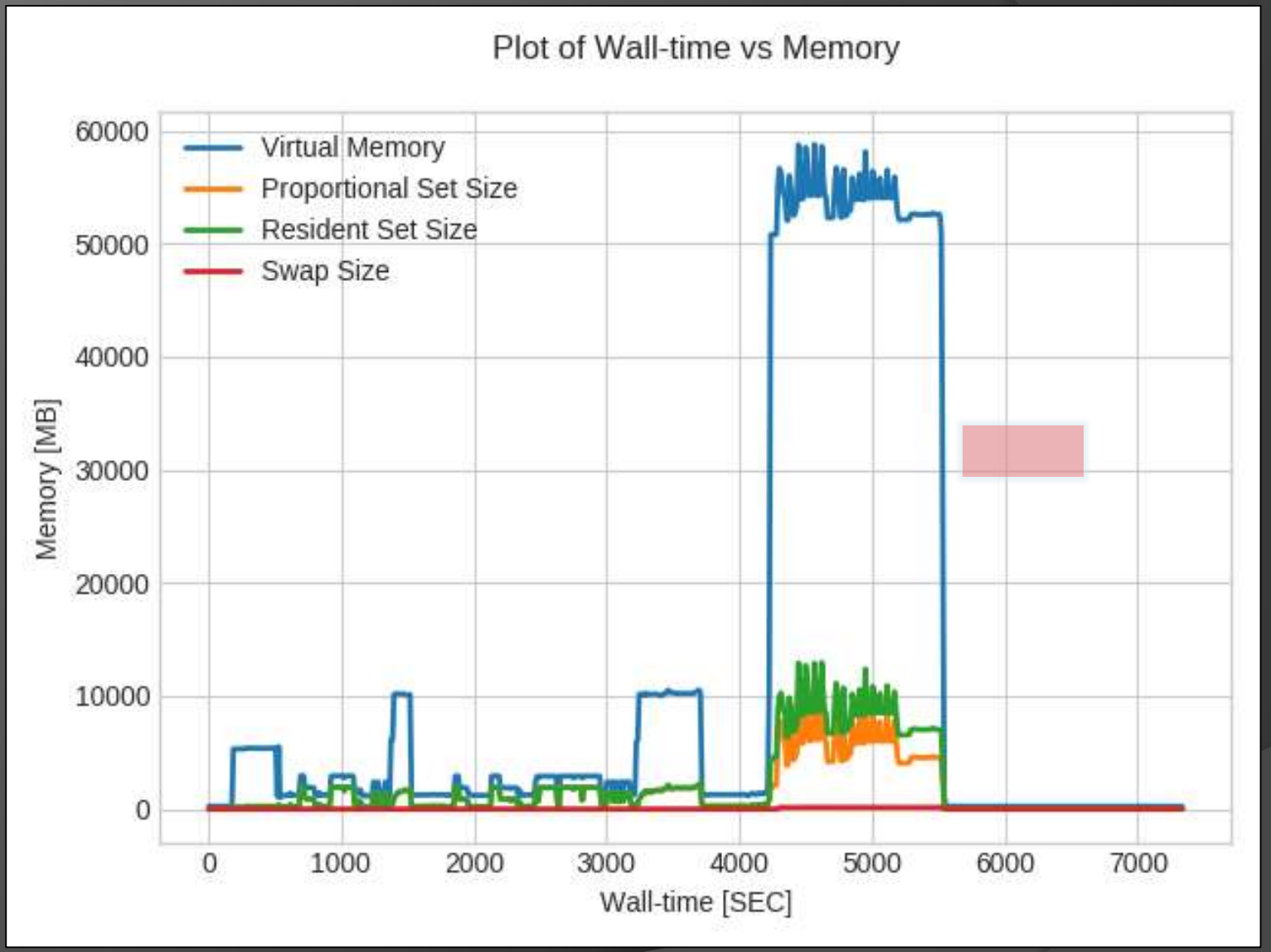


- Data are transferred using Globus end points in Cape Town and Manchester (takes ~12 hours per ~1TB dataset)
- Originally tried adding directly to the DIRAC filecatalog using a UI in Cape Town, but the transfer was very slow (3 days for a single dataset).
- Process now Globus followed by local registration using the DIRAC file catalog client.
- Intention to write stand alone DIRAC task for register-in-place.

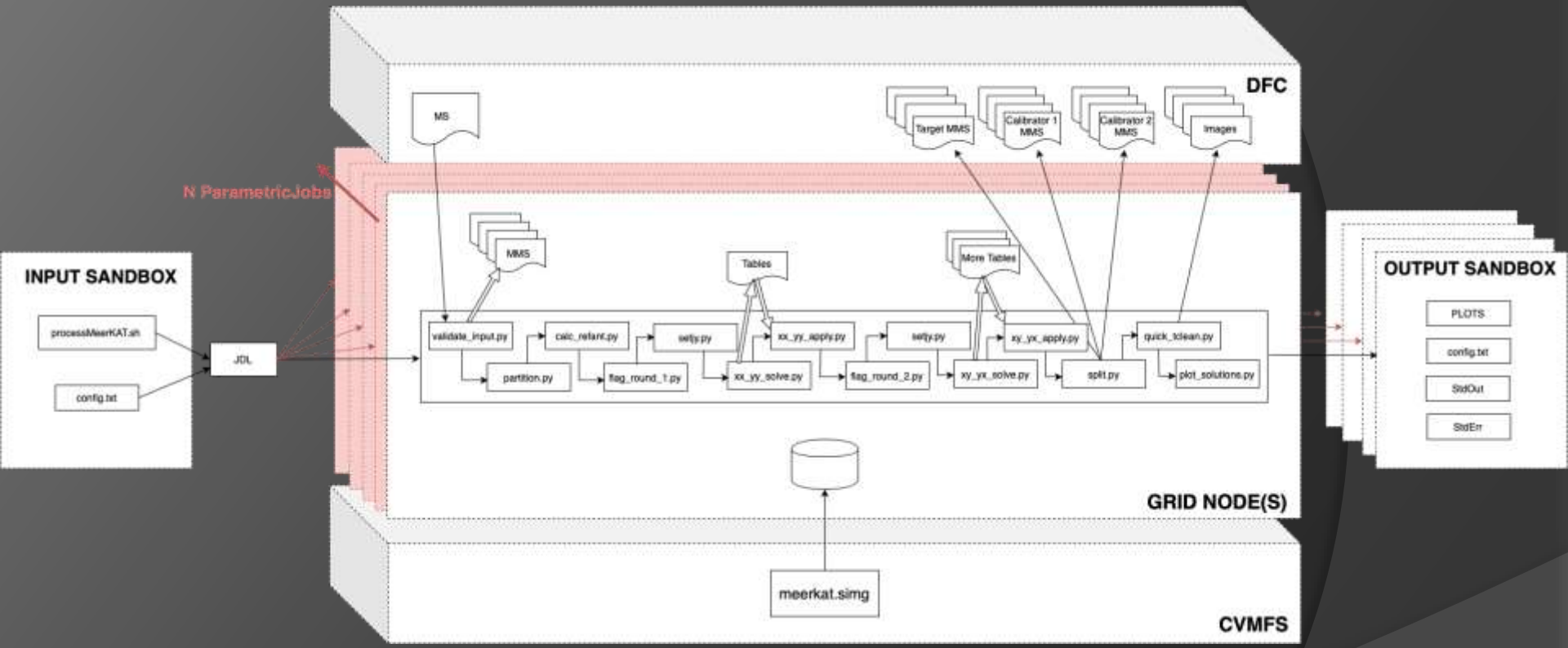
IRIS IMPLEMENTATION OF MEERKAT PIPELINE



IRIS IMPLEMENTATION OF MEERKAT PIPELINE



IRIS IMPLEMENTATION OF MEERKAT PIPELINE

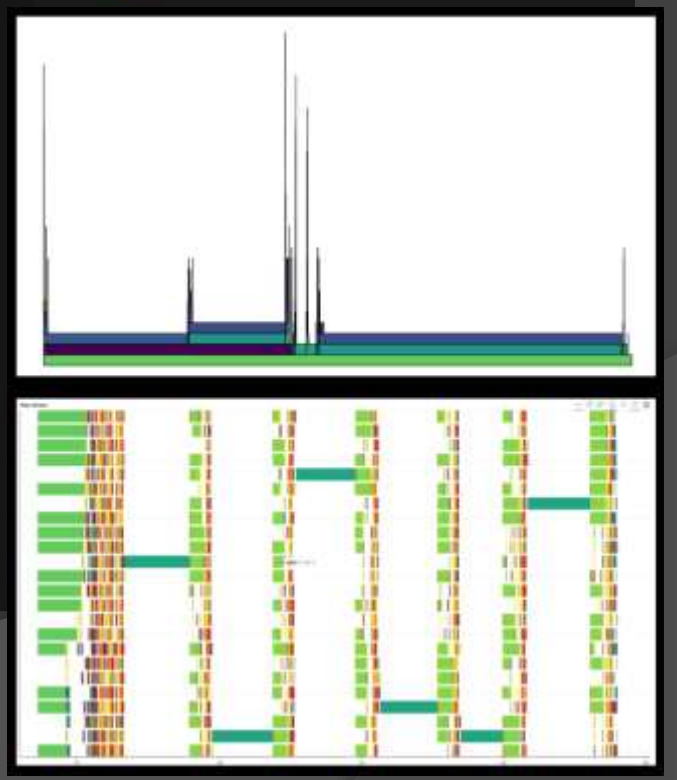


RASCIL

Radio Astronomy Simulation, Calibration and Imaging Library

Tim Cornwell

- Python & Numpy
- Can be run via jupyter notebooks
- Using threaded libraries for SMP and the Dask library for distributed processing



RASCIL (PACKAGE)

DATA MODELS PACKAGE.

These are for modelling data

PACKAGE CONTENTS.

buffer_data_models, data_model_helpers, memory_data_models, parameters, polarisation

PROCESSING COMPONENTS PACKAGE.

These are the processing components exposed to the Execution and Framework

PACKAGE CONTENTS.

arrays (package), calibration (package), flagging (package), fourier_transforms (package), griddata (package), image (package), imaging (package), simulation (package), skycomponent (package), skymodel (package), util (package)

visibility (package)

SUBMODULES.

atmospheric_screen, Base, chain_calibration, Cleaners, Coalesce, Configurations, convolution_functions, deconvolution, dft, fft_coordinates, fft_support, gather_scatter,, gradients, gridding, imaging_params, iterators, Jones, Kernels, Noise, Operations, Pointing, primary_beams, Rfi, simulation_helpers, Solvers, Surface, testing_support, timeslice_single, vis_select, visibility_fitting, visibility_geometry, Weighting, wstack_single

WORKFLOWS PACKAGE.

Algorithm Reference Library workflows

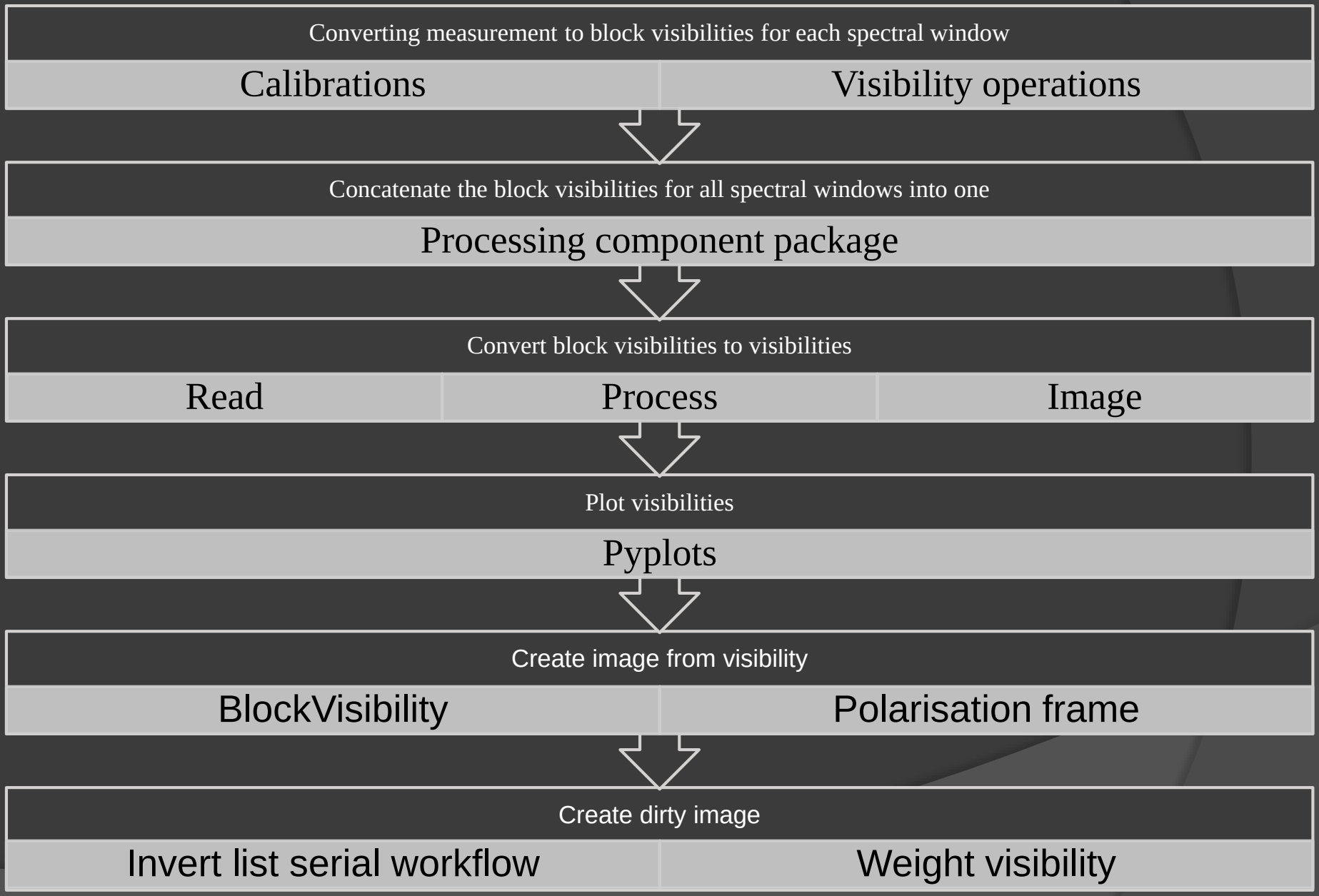
PACKAGE CONTENTS.

rsexecute (package), serial (package), shared (package)

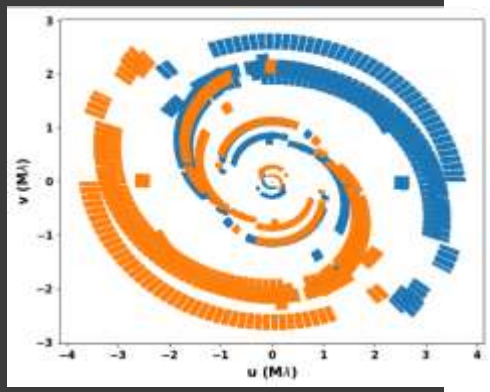
SUBMODULES.

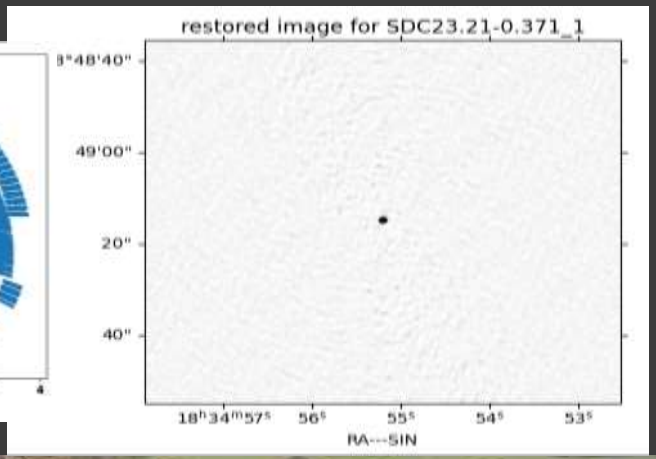
calibration, calibration_rsexecute, calibration_serial, execution_support, image, image_rsexecute, imaging, imaging_rsexecute, imaging_serial, imaging_shared, pipeline_mpccal_rsexecute, pipeline_rse

RASCIL WORKFLOW



- RASCIL testing and development using data from eMERLIN, ALMA & JVLA telescopes.

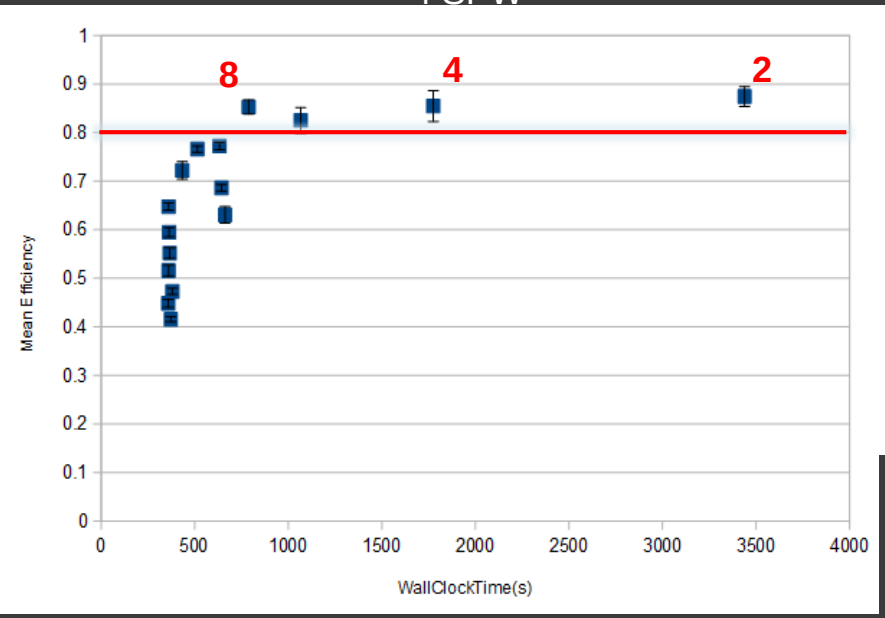




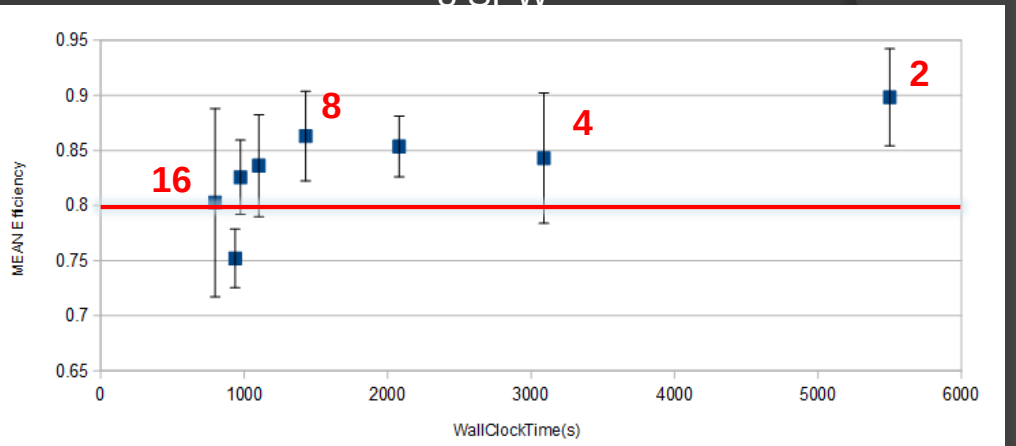
RASCIL PERFORMANCE PROFILING ON IRIS

e-MERLIN test dataset

4 SPW



8 SPW



- Profiling tests to set up template jobs given particular data specifications.
- Efficiency ~85%
- Memory profiling ongoing.

ALMA Atomium:

ALMA tracing the origins of molecules forming dust in oxygen-rich M-type stars

PI Leen Decin, KU Leuven, Belgium

CoPI Carl Gottlieb, CfA Harvard, USA

Avison, Adam; Baudry, Alain; Beck, Elvire De; Blitz, M.; Boulangier, Jels; Cannon, Emily; Carrillo, Sanchez J.D.; Danilovich, Taïssa; De Ceuster, Frederik; De Koter, Alex; El Mellah, Ileyk; Etoka, Sandra; Fabrice, Herpin; Gielen, Clio; Gobrecht, David; Gottlieb, Elaine; Gray, Malcolm; Heard, Dwayne; Hutton, Lewis; James, A.; Jeste, Manali; Keller, Denise; Kervella, Pierre; Khouri, Theo; Lagadec, Eric; Long Lee Kin, Kelvin; MacDonald, Iain; Mangan, Thomas; Menten, Karl; Millar, Tom; Montagès, Miguel; Müller, Holger; Nuth, Joseph A.; Pimpanuwat, Bannawit; Plane, John; Price, Daniel; Raghvendra, Sahai; Richards, A.M.S.; Sindel, J.P.; Van der Sande, Marie; Wallstrom, Sofia; Ward, Homan; Waters, Rens; West, Niclas; Wiegert, Joachim; Wong, Ka-Tat; Yates, Jeremy; Zijlstra, Albert

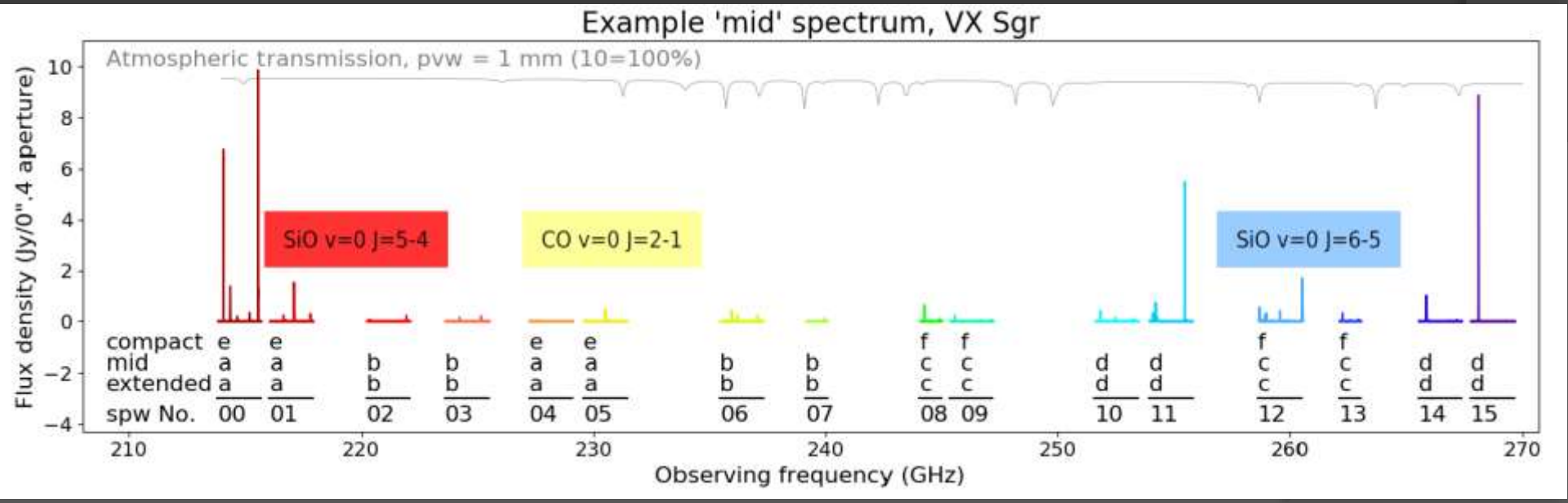
Based mostly in Europe and USA, originated from at least 5 continents

ATOMIUM sample

eas 2020 Virtual

Star	Type	D✷ mas
S Pav	SRa	12
T Mic	SRb	9
U Del	SRb	8
RW Sco	OH/IR	5
V PsA	SRb	13
SV Aqr	LPV	4
R Hya	Mira	23
U Her	Mira	11
pi1 Gru	SRb	21
AH Sco	Mira	6
R Aql	Mira	12
W Aql	Mira	11
GY Aql	Mira	20
IRC-10529	OH/IR	6
KW Sgr	RSG	4
IRC+10011	OH/IR	7
VX Sgr	RSG	9

8e-8 Increasing mass loss rate 6e-5 M⊙/yr

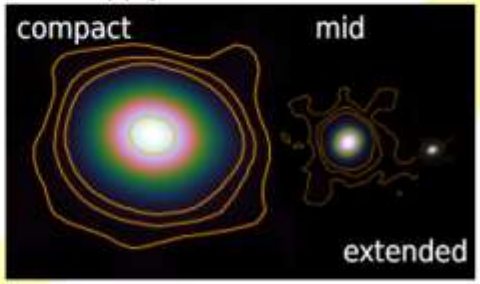


Data Reduction

ALMA observatory/ESO
Pipeline calibration
Initial cubes

UK ARC, Manchester
IRIS computing
Atoms access by VPN

Self-calibrate continuum
Apply to all data



Understand:
Dust formation
Wind driving
Shaping
of CSEs
& PNe

Extract spectra
Line ID (CDMS)
Codes e.g. Hydro
Phantom, AMVAC
Chem e.g. KROME
Masers...

Line cubes for
each config
~70,000 chans
per star.
Over 1 million
channels total

- 17 stars, 3 array spacings, 2 or 4 frequency tunings = 482 data sets - 48 TB download
- Current result is 200 GB per data set, will be final product of 3 TB per star

To capitalise on e-MERLIN's unique technical capabilities, the STFC have commissioned twelve long-term super-projects to help answer the big questions.

eMERGE (e-MERLIN Galaxy Evolution Survey) is an ambitious Legacy survey to exploit e-MERLIN's unique combination of sensitivity and spatial resolution to study the formation and evolution of star-forming galaxies and AGN out to redshifts of $z > 5$.

Principal Investigators: Tom Muxlow (JBCA), Ian Smail (Durham) and Ian McHardy (Southampton)

Time Allocation: 918 hours (e-MERLIN including Lovell Telescope).
Tier 1: 738 hours (360 hours at L-band, 378 hours at C-band). Tier 0: 180 hours (L-band)

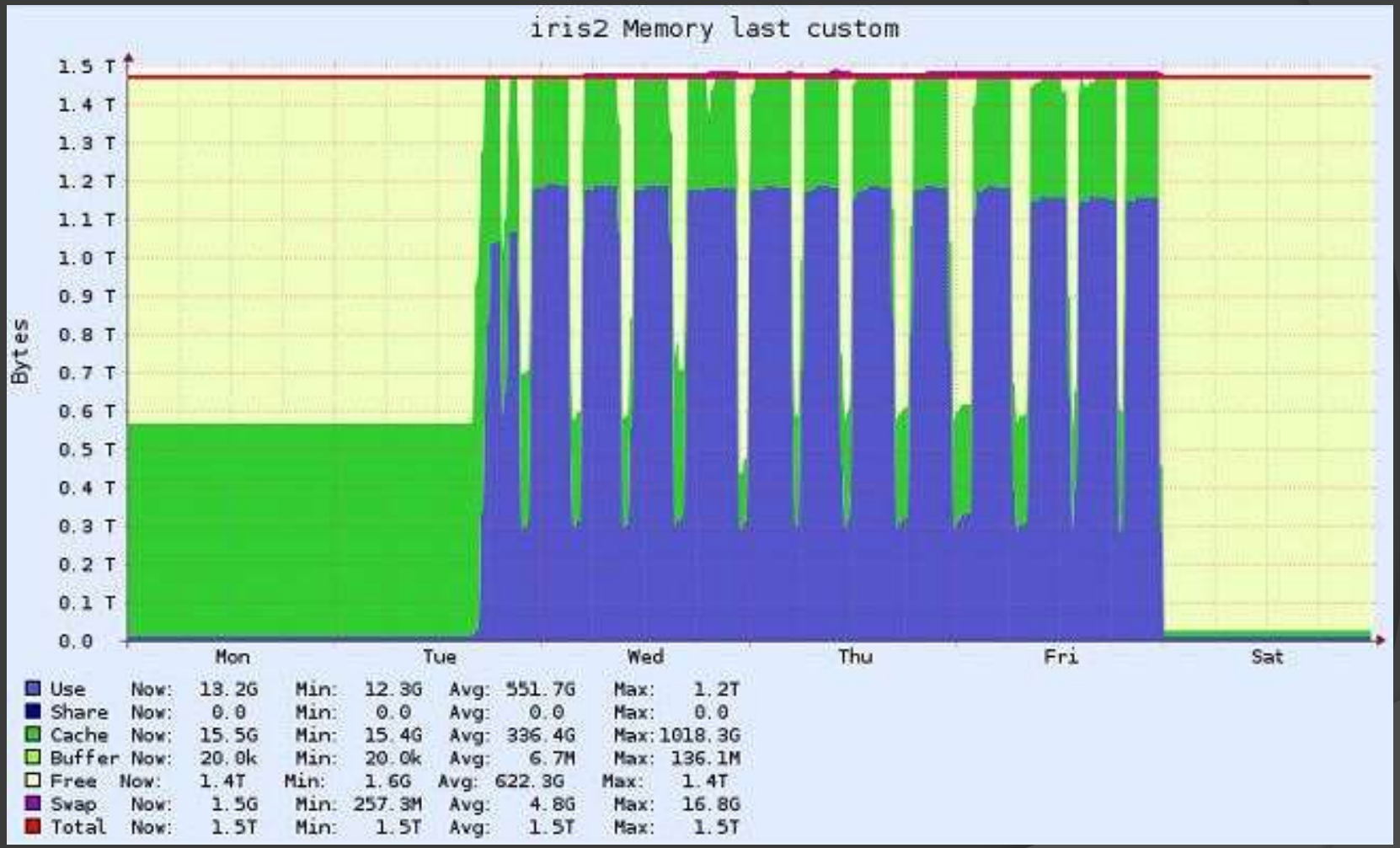
eMERGE DR-1

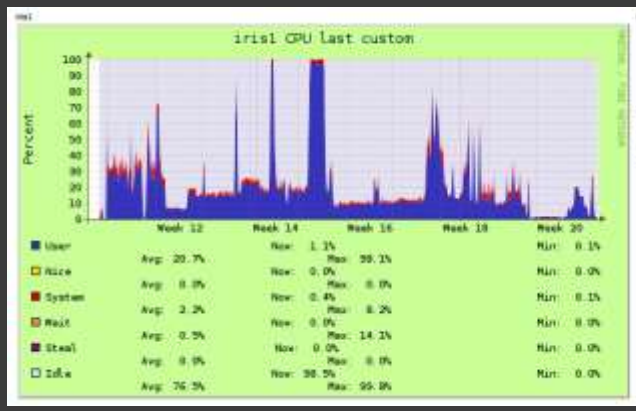
Covers the central Lovell-dominated field with ~25% of the e-MERLIN data (+JVLA data) which have been averaged

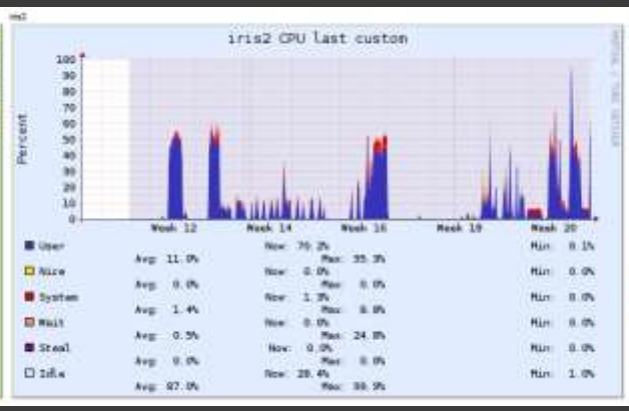
Credit: NRAO/AUI/NSF

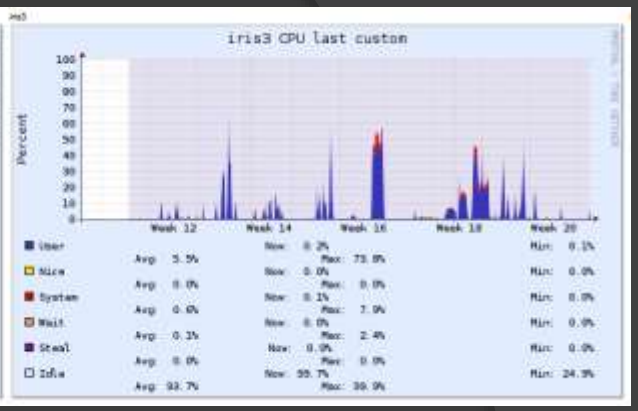
eMERGE

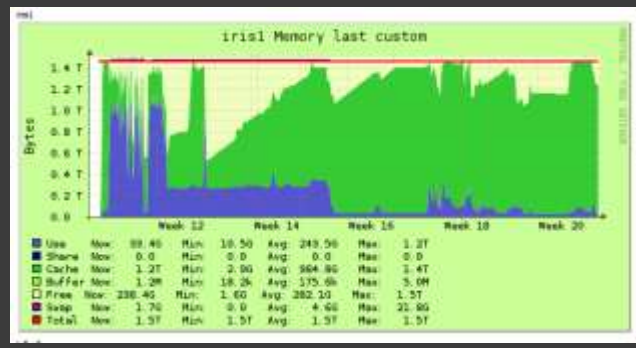
- Singularity image created for CASA and scripts for pipeline processing
- The DR1 dataset is ~2TB – IRIS himem nodes key to speeding up processing.
- The DR2 is expected ~ 10TB within the next 12 months.
- Need enough RAM to hold several 45k x 45k 2D FITS images in memory plus as much input data as possible
- Scaling tests to be conducted to plot curves of growth for memory usage as functions of (a) 2D image size and (b) dataset size in TB.
- DR-1 has 848 radio sources catalogued
- DR-2 will contain several thousand sources.
- DR-1 has required imaging utilising the ~1.5TB RAM, DR-2 will certainly require the full 3TB of RAM to image - together with parallelisation.

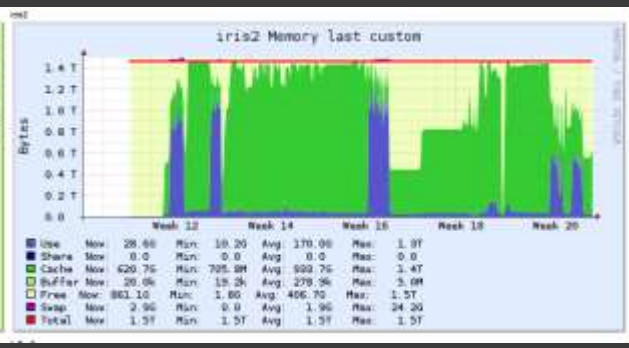

iris2 Memory last custom
Bytes
1.5 T
1.4 T
1.3 T
1.2 T
1.1 T
1.0 T
0.9 T
0.8 T
0.7 T
0.6 T
0.5 T
0.4 T
0.3 T
0.2 T
0.1 T
0.0
Mon
Tue
Wed
Thu
Fri
Sat
Use Now: 13.2G Min: 12.3G Avg: 551.7G Max: 1.2T
Share Now: 0.0 Min: 0.0 Avg: 0.0 Max: 0.0
Cache Now: 15.5G Min: 15.4G Avg: 336.4G Max: 1018.3G
Buffer Now: 20.0k Min: 20.0k Avg: 6.7M Max: 136.1M
Free Now: 1.4T Min: 1.6G Avg: 622.3G Max: 1.4T
Swap Now: 1.5G Min: 257.3M Avg: 4.8G Max: 16.8G
Total Now: 1.5T Min: 1.5T Avg: 1.5T Max: 1.5T

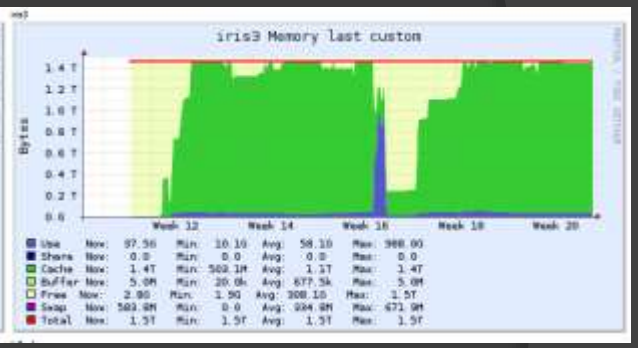

iris1 CPU last custom
Percent
Week 12
Week 14
Week 16
Week 18
Week 20
User
Nice
System
Wait
Steal
Idle


iris2 CPU last custom
Percent
Week 12
Week 14
Week 16
Week 18
Week 20
User
Nice
System
Wait
Steal
Idle


iris3 CPU last custom
Percent
Week 12
Week 14
Week 16
Week 18
Week 20
User
Nice
System
Wait
Steal
Idle


iris1 Memory last custom
Bytes
Week 12
Week 14
Week 16
Week 18
Week 20
Use
Share
Cache
Buffer
Free
Swap
Total


iris2 Memory last custom
Bytes
Week 12
Week 14
Week 16
Week 18
Week 20
Use
Share
Cache
Buffer
Free
Swap
Total

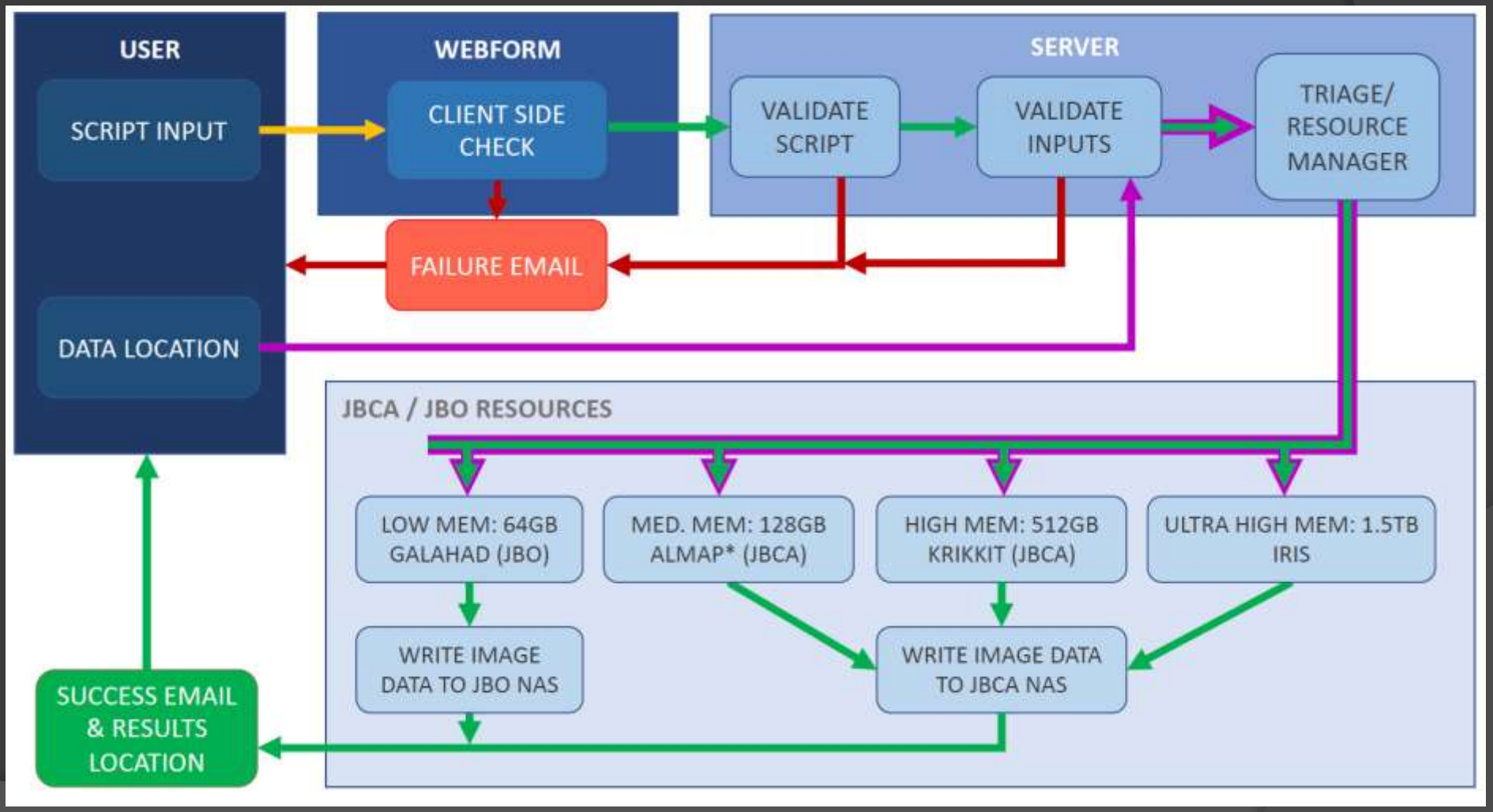

iris3 Memory last custom
Bytes
Week 12
Week 14
Week 16
Week 18
Week 20
Use
Share
Cache
Buffer
Free
Swap
Total


iris1
GB
Week 12
Week 14
Week 16
Week 18
Week 20
disk_free last custom
Now: 10.44k
Min: 10.42k
Avg: 14.94k
Max: 23.37k


iris2
GB
Week 12
Week 14
Week 16
Week 18
Week 20
disk_free last custom
Now: 6.26k
Min: 1.92k
Avg: 9.46k
Max: 23.37k


iris3
GB
Week 12
Week 14
Week 16
Week 18
Week 20
disk_free last custom
Now: 14.93k
Min: 14.90k
Avg: 18.59k
Max: 23.37k

Community – CLEAN service



Future

- CASA being developed to have more parallel operations
- RASCIL
- Better exploit multi-core CPU
- GPU imaging codes in testing
- Increases in dataset sizes