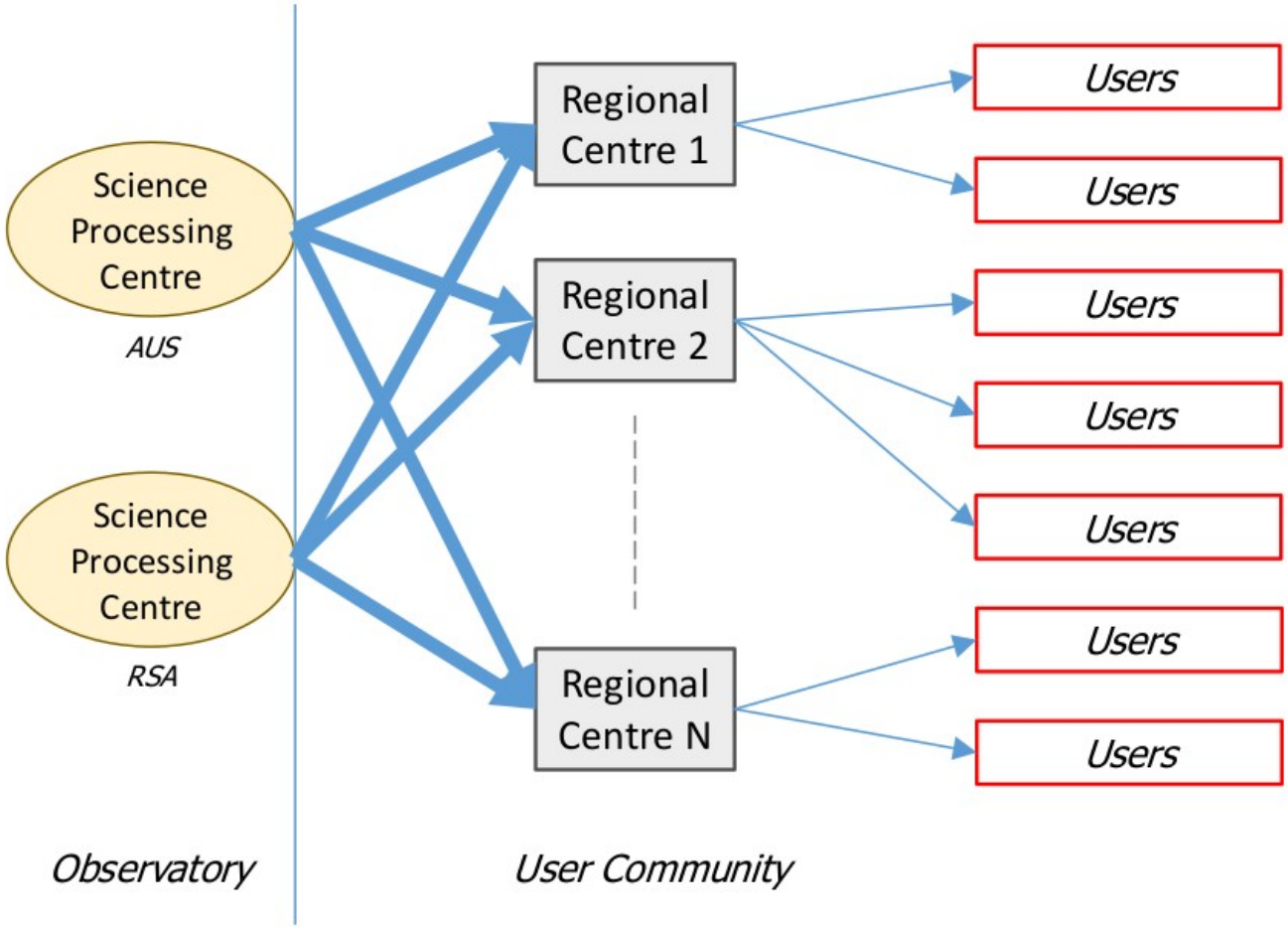
Plans and Progress towards an Indian SRC

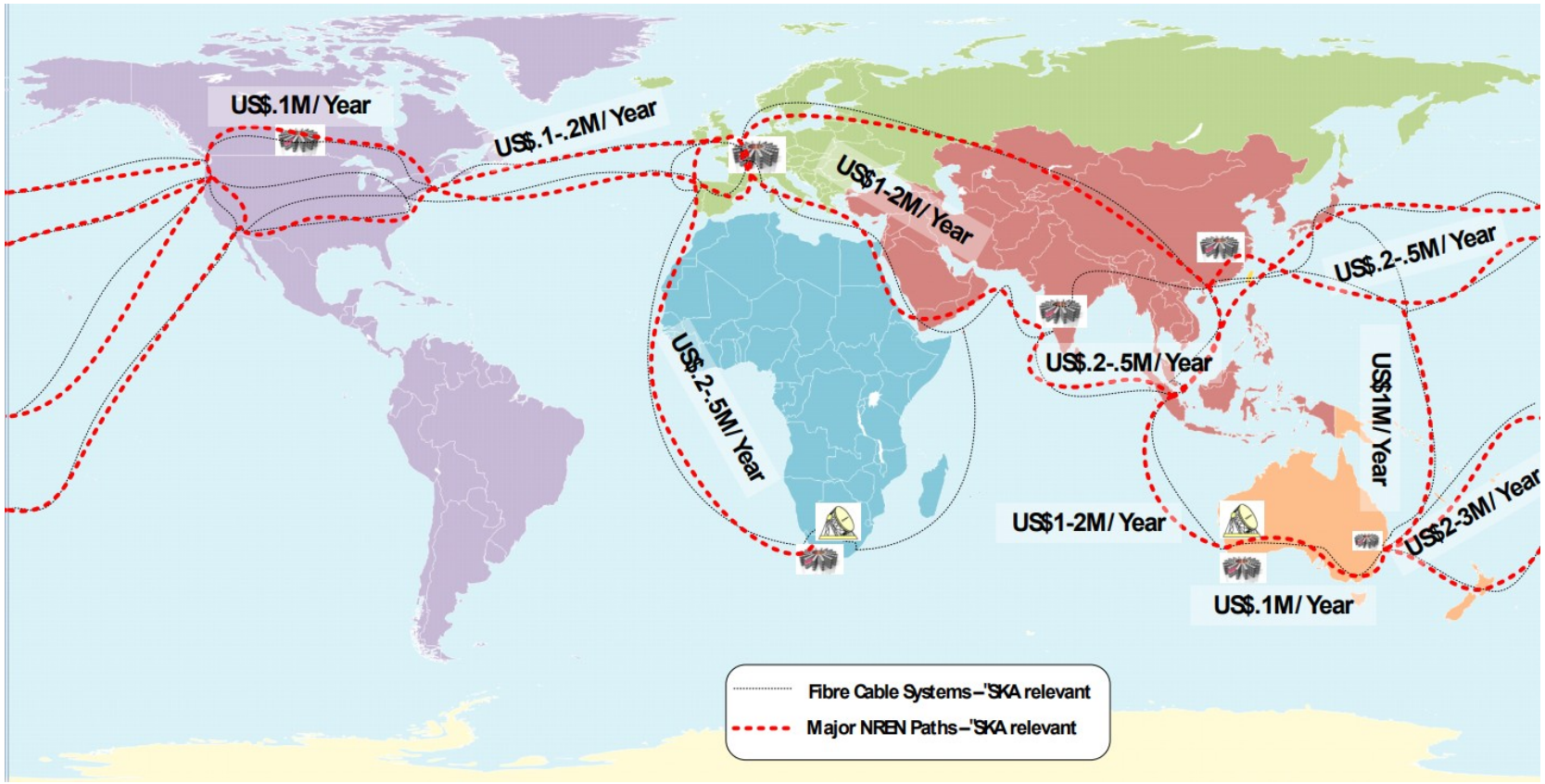
Yogesh Wadadekar
(NCRA- TIFR)

NCRA • TIFR

SKA Regional Centres



An SRC in western India?



Indian SRC Plans 2019-23

- Funding proposal requesting for about 5.5 MEuro over a 4 year period 2019-2023 for a prototype SRC submitted to the government.
- We expect the first tranche of funding (about 700k Euros) to arrive by December.
- We intend to get started by buying some storage and compute hardware to be used for pipeline processing of SKA precursor data.
- **Prototype SRC activities will initially focus on development and deployment of pipelines for processing of all legacy GMRT/uGMRT data and MWA observations of the Sun (with Divya Oberoi).**
- We will transfer the MWA solar observations (about 4.5 PB at present) via a dedicated high speed internet connection between the Pawsey Centre in Perth and NCRA in cooperation with NKN in India and the AARnet in Australia.

NCRA • TIFR

Imaging the GMRT archive

- All interferometric GMRT observations taken since Cycle 1 in 2002 have been archived over many generations of tape and disk
- The GMRT archive on NAPS now hosts about **250 TB** of interferometric data which are served to users as visibilities. But a majority of GMRT observations are unpublished.
- We were looking to provide users with “first look” (worst case) and “science ready” (best case) images for all GMRT observations. A “certifiably bad” tag on data is also useful.
- We currently use the SPAM pipeline (by Huib Intema) which has been thoroughly tested for the TGSS ADR1 data release at 150 MHz and with many other GMRT datasets at 150, 235, 325, 610 MHz bands.
- The entire processing architecture has been designed in such a way that it will be straightforward to migrate to any other processing pipeline in the future.

NCRA • TIFR

About SPAM

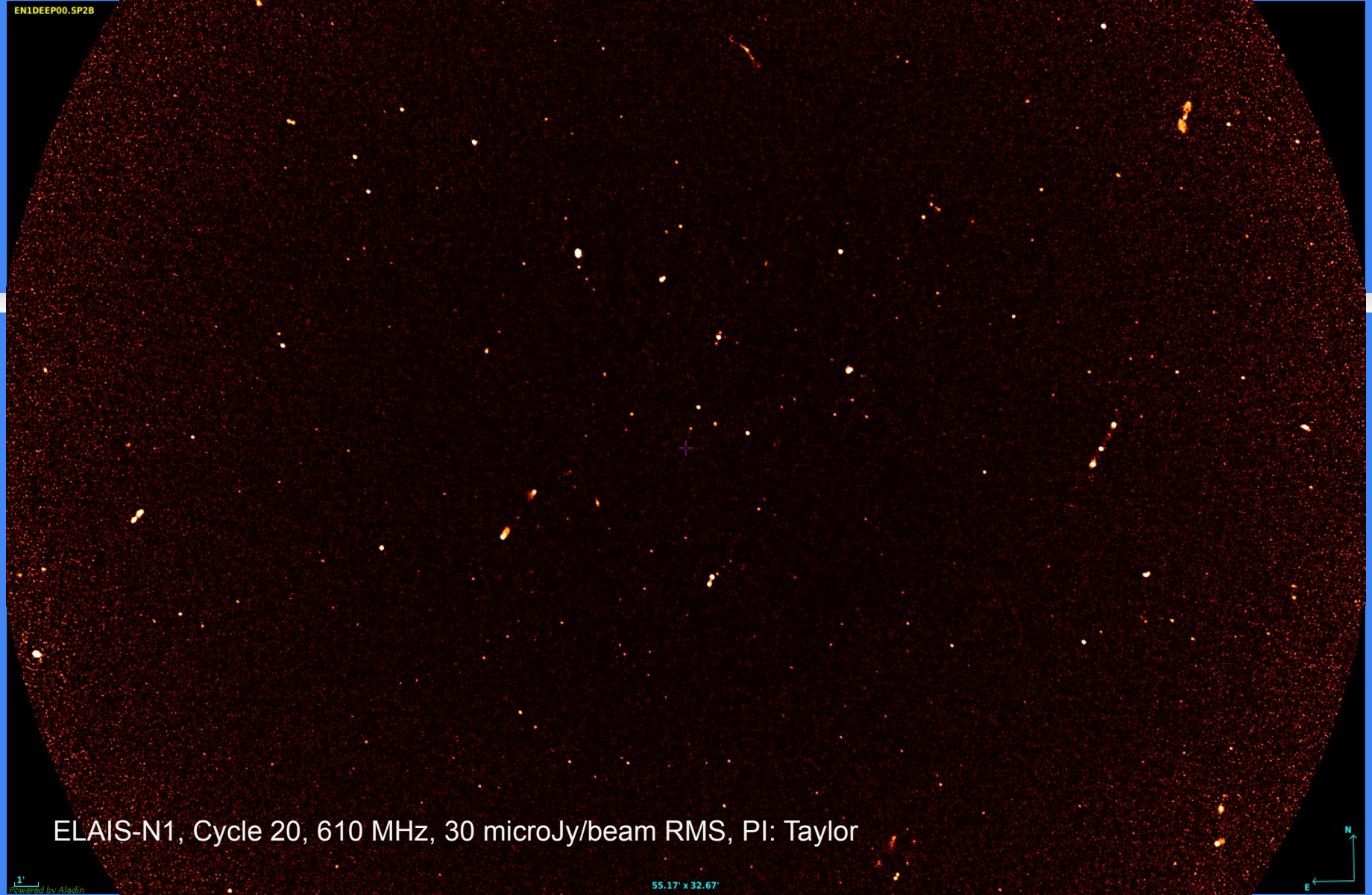
- SPAM is a Python module that provides an interface to AIPS via ParselTongue and ObitTalk
- ParselTongue provides access to AIPS tasks, data files (images & visibilities) and tables.
- Data reductions are carried out by well-tested Python scripts in SPAM that execute AIPS tasks directly or via high-level functions that make multiple AIPS or ParselTongue calls.
- SPAM now also includes a fully automated pipeline for reducing legacy GMRT observations at 150, 235, 325 and 610 MHz.
- **Entire software stack has been packaged as a Docker container**

NCRA • TIFR

The processing factory

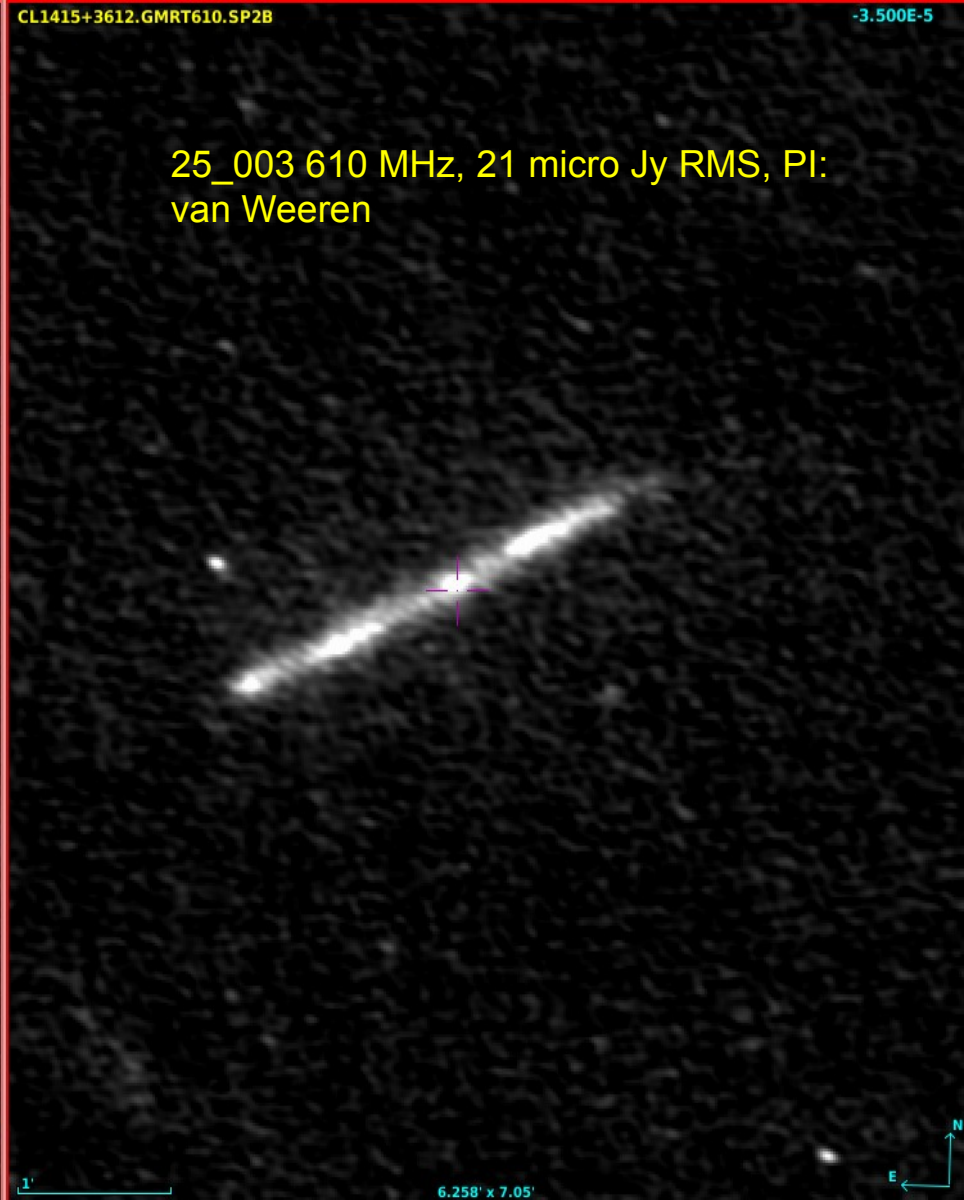
- A cluster of desktops configured with the SPAM software stack
- Input visibilities and output images stored on a 1 Petabyte sized DELL Isilon storage system.
- All processing tracked via custom Python scripts that use and store processing information into a Postgres database.
- The first hardware setup and scripting was designed by a group of undergraduate computer engineering students. We have now **refactored the software in collaboration with an industry partner (Persistent Systems Ltd)** on the production version of the system.

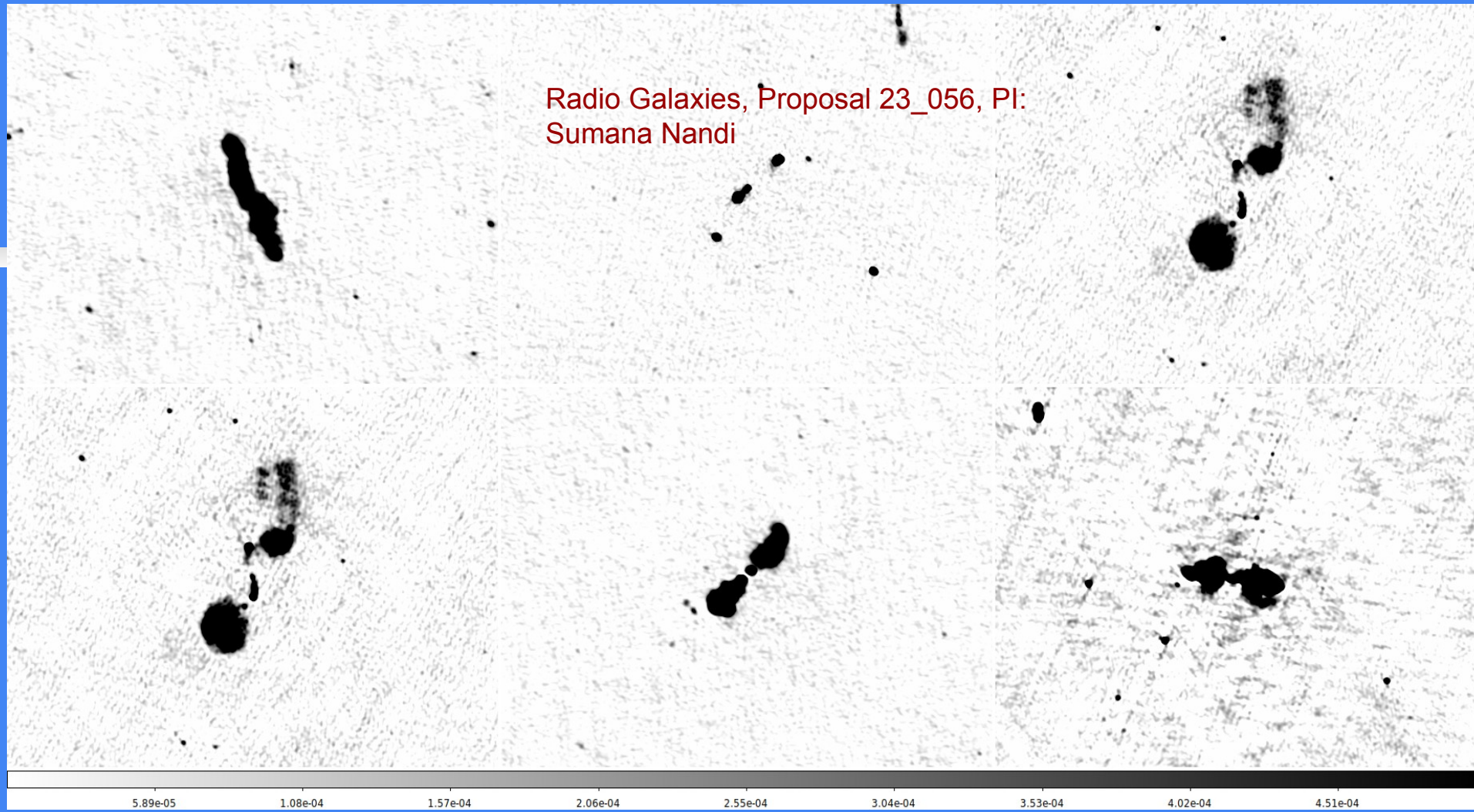
NCRA • TIFR

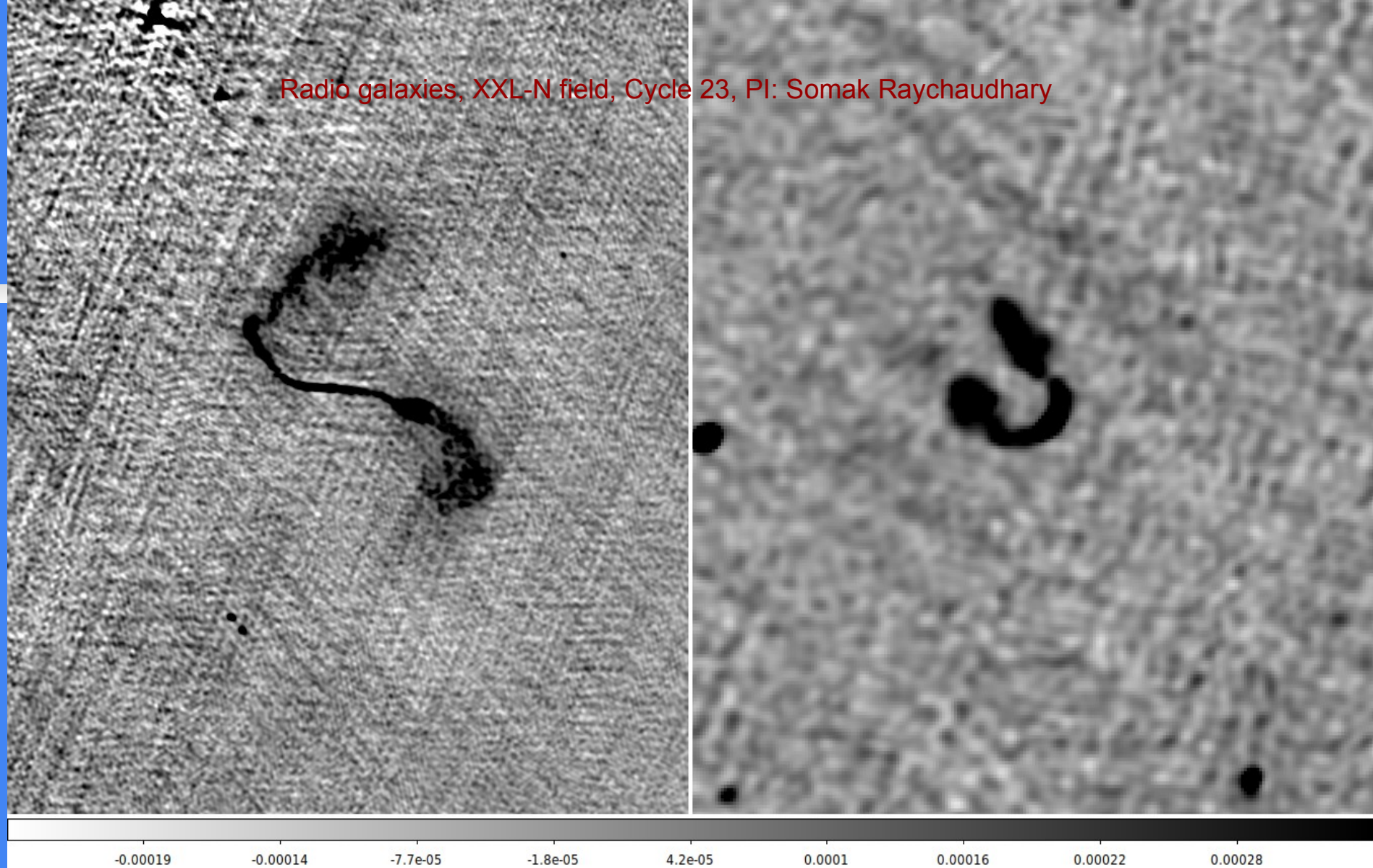


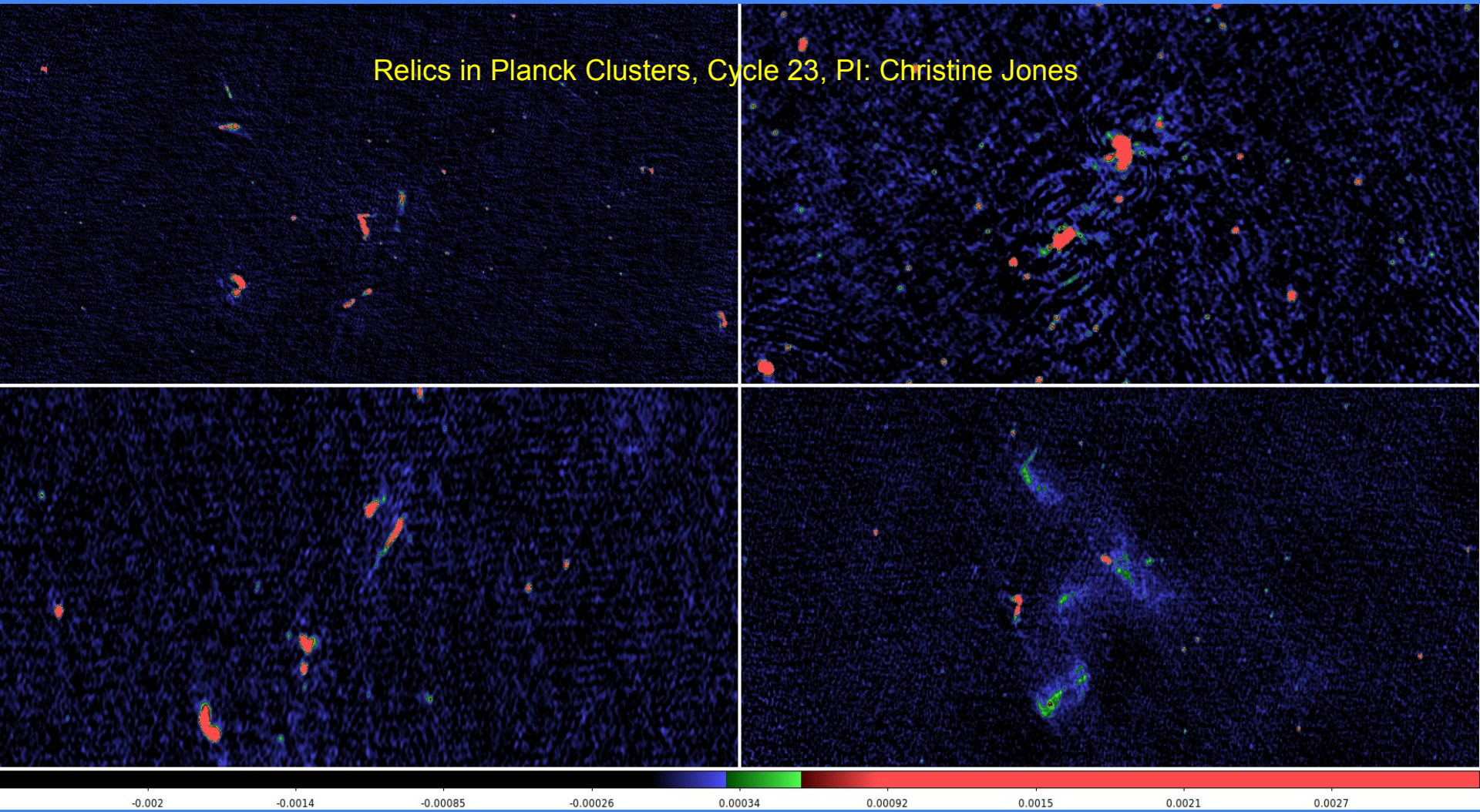
ELAIS-N1, Cycle 20, 610 MHz, 30 microJy/beam RMS, PI: Taylor

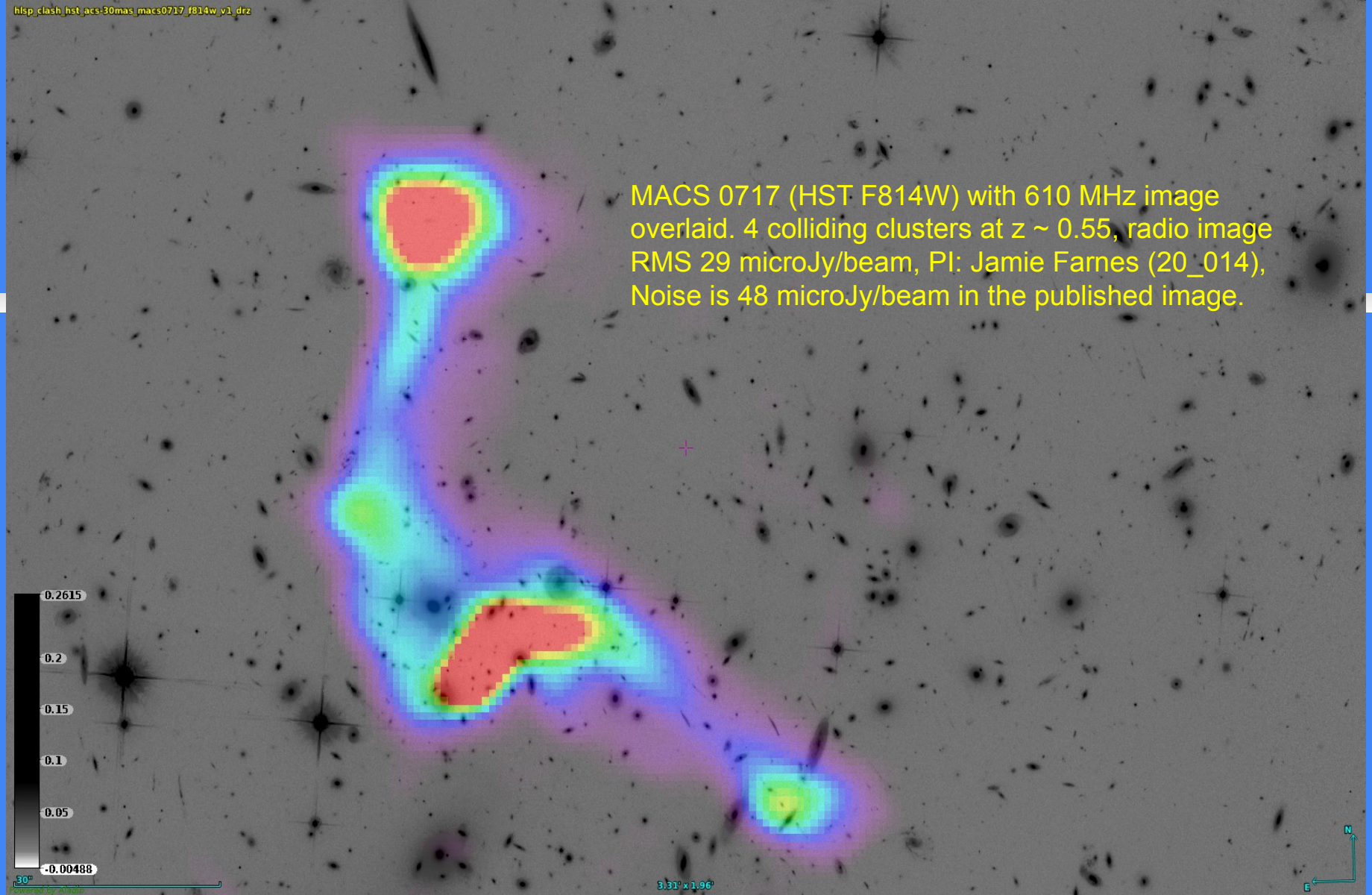

CDS/P/PanSTARRS/DR1/color-z-zg-g
1'
6.258' x 7.05'
N
E


CL1415+3612.GMRT610.SP2B
-3.500E-5
25_003 610 MHz, 21 micro Jy RMS, PI: van Weeren
1'
6.258' x 7.05'
N
E

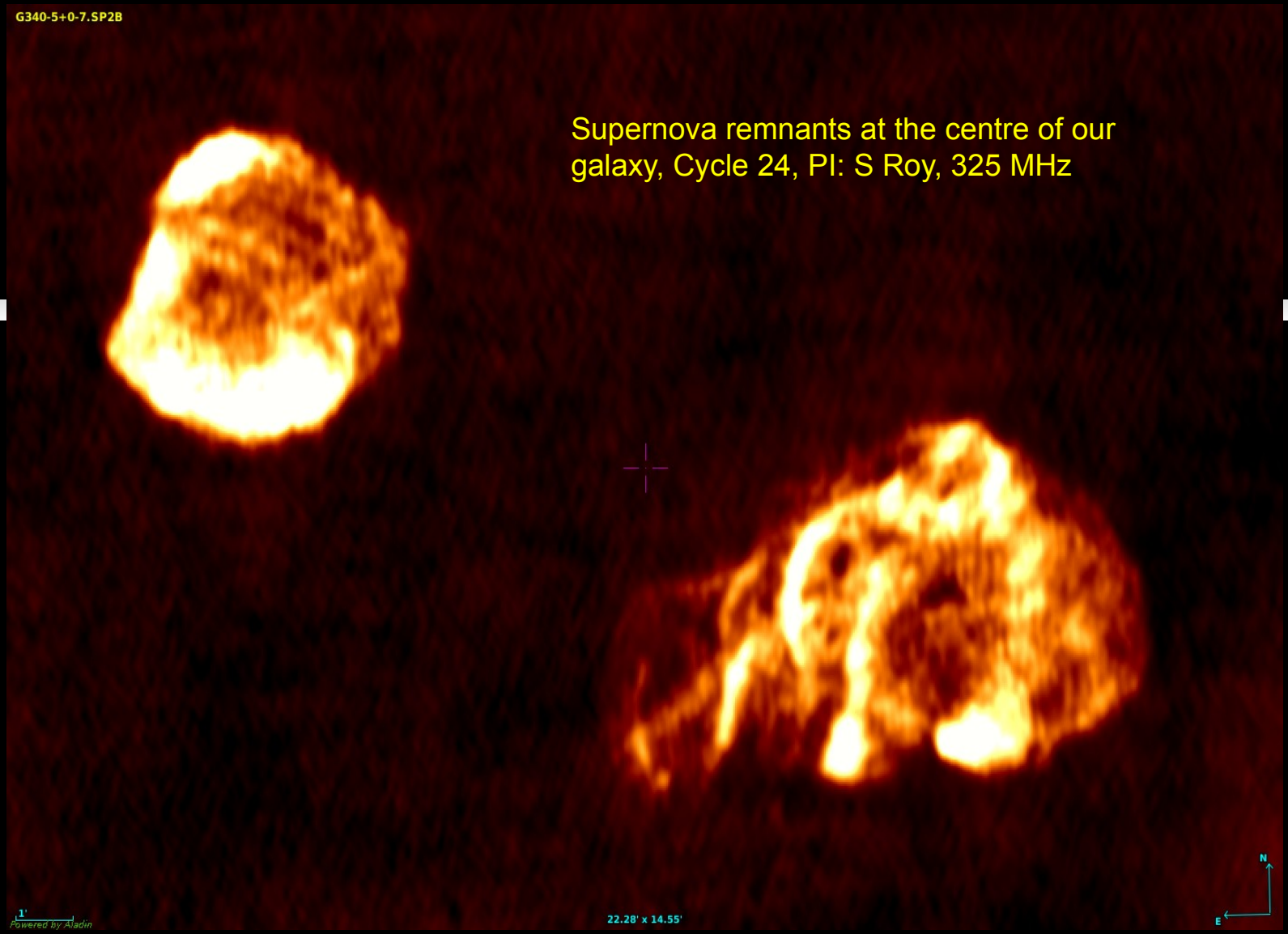

Radio Galaxies, Proposal 23_056, PI:
Sumana Nandi
5.89e-05
1.08e-04
1.57e-04
2.06e-04
2.55e-04
3.04e-04
3.53e-04
4.02e-04
4.51e-04


Radio galaxies, XXL-N field, Cycle 23, PI: Somak Raychaudhary
-0.00019
-0.00014
-7.7e-05
-1.8e-05
4.2e-05
0.0001
0.00016
0.00022
0.00028

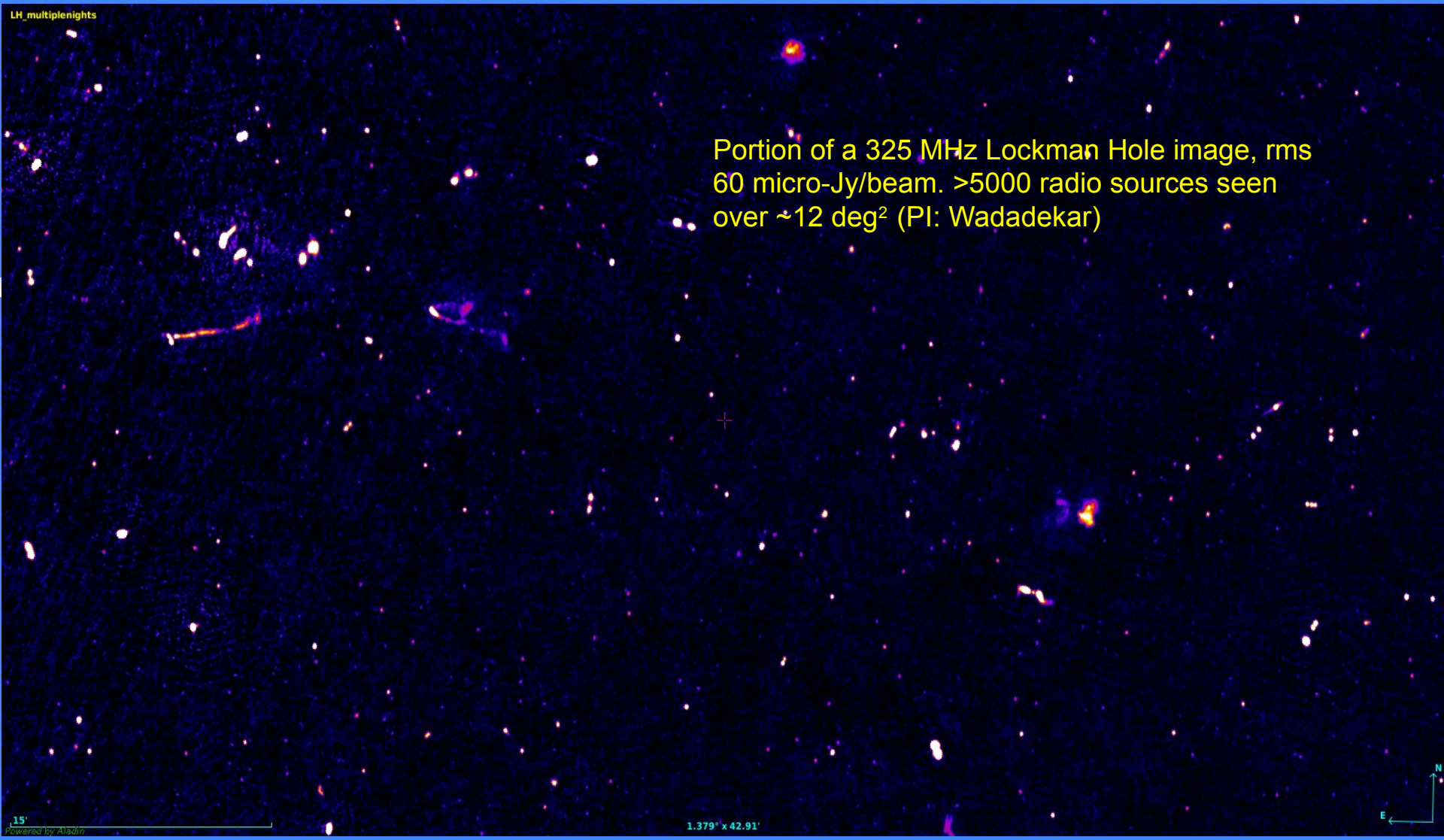

Relics in Planck Clusters, Cycle 23, PI: Christine Jones
-0.002
-0.0014
-0.00085
-0.00026
0.00034
0.00092
0.0015
0.0021
0.0027



MACS 0717 (HST F814W) with 610 MHz image overlaid. 4 colliding clusters at z ~ 0.55, radio image RMS 29 microJy/beam, PI: Jamie Farnes (20_014), Noise is 48 microJy/beam in the published image.

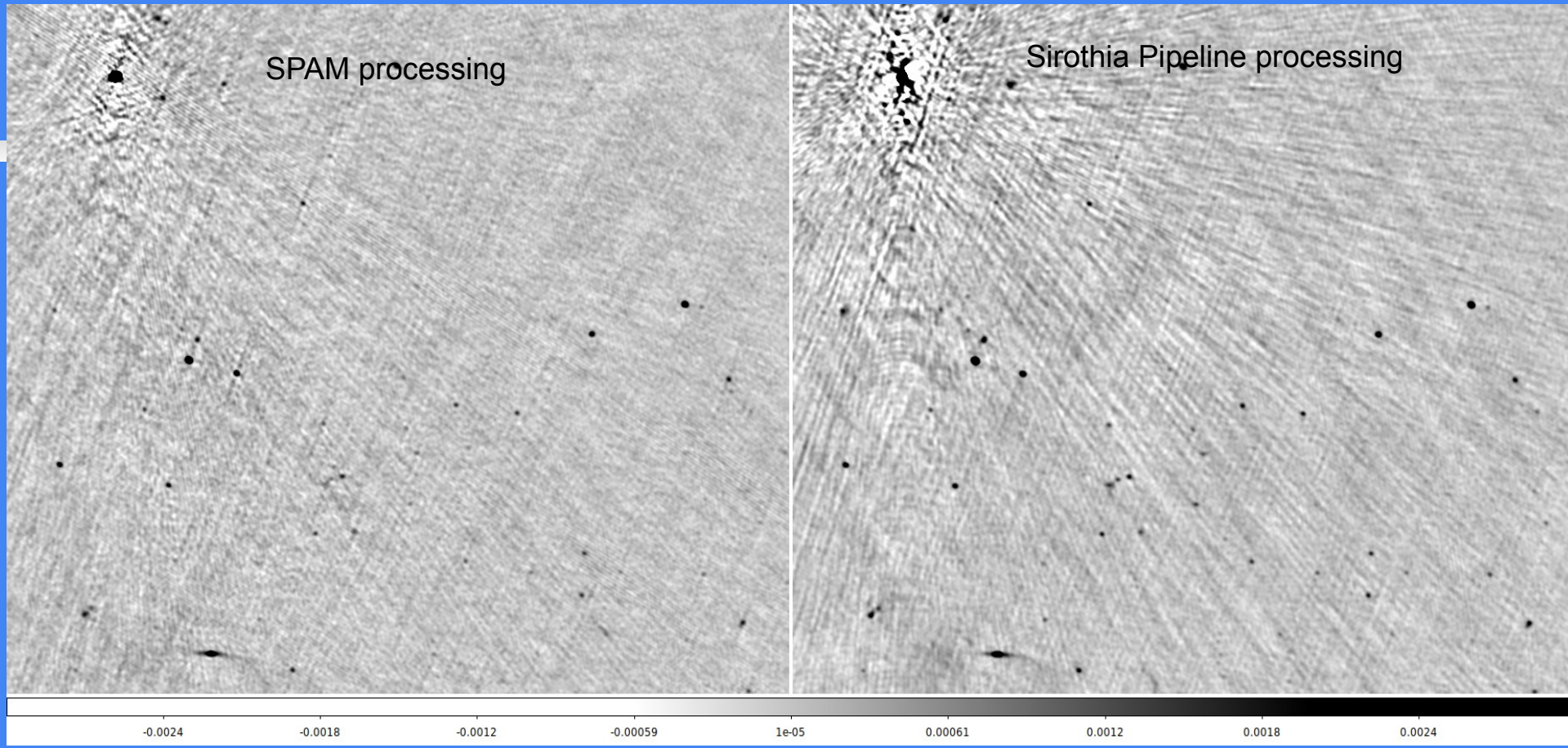


Supernova remnants at the centre of our galaxy, Cycle 24, PI: S Roy, 325 MHz



Portion of a 325 MHz Lockman Hole image, rms 60 micro-Jy/beam. >5000 radio sources seen over ~12 deg^2 (PI: Wadadekar)

XMMLSS field, 325 MHz, Cycle 20, PI: Wadadekar



What we are doing now

- Processing all cycles with GHB and GSB data one cycle at a time
- Processing data in 150, 240, 325, 610 MHz bands; skipping L band for now
- Only processing data with bandwidth >= 16 MHz
- **Doing manual tests with a variety of images in all bands compared to images prepared by hand or with other pipelines to verify flux calibration and image quality**
- Developing Automated Quality Control with some quality flags to warn users
- Generating PyBDSF component catalogs
- Producing HiPS images in each band for easy visualisation
- **Integrate all outputs into the NAPS system which will deliver raw visibilities and imaging data via the same search interface.** Pilot release of Cycle 25 and Cycle 26 images will happen soon
- Some testing and refactoring of CASA based uGMRT pipelines has commenced.

NCRA • TIFR

Characterise telescope and ionosphere

- Since SPAM does does direction dependent calibration, it produces a ionospheric phase screen model along different lines of sight. Divya Oberoi and Huib working on this aspect.
- Our pipeline processes data in a uniform and reproducible manner. The thousands of SPAM logs that our processing factory will produce can be mined to answer questions like: How has the RFI environment changed in the last 15 years in each band? How strong is the diurnal variation? What was the impact of moving from GHB to GSB on the sensitivity?

NCRA • TIFR

What we will do in the future

- Once we have a working uGMRT pipeline and enough compute resources, we want to image data as soon as they are taken.
- **Within a few days of a GMRT observation, authorised users will begin to see images of their sources in the NAPS system**
- These images will be made public at the same time as the raw visibilities – currently 18 months after an observation for regular cycles.

NCRA • TIFR

Thank You!

Comments and questions welcome

NCRA • TIFR