Funded by the National Science Foundation, Grant #2127548

Surging into the Cloud and Grid:

Insights into keeping up with the processing workload for the

Very Large Array Sky Survey

by David Halstead

The March Toward the Clouds: MF Perspectives

March 2, 2022

CI Compass Cyberinfrastructure for NSF Major Facilities Workshop

USC Viterbi School of Engineering Information Sciences Institute
INDIANA UNIVERSITY
UNIVERSITY OF NOTRE DAME
THE UNIVERSITY OF UTAH
TEXAS TECH UNIVERSITY
renci

Surging into the Cloud and Grids: Insights into keeping up with the processing workload for the Very Large Array Sky Survey

David Halstead,
CIO, NRAO

Panel at the 2022 NSF Cyberinfrastructure for Major Facilities Workshop

The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc

Funded by the National Science Foundation, Grant #2127548

The Very Large Array: New Mexico Desert Plains of San Agustin

VLA construction started in 1972,dedicated in 1980, upgraded to digital infrastructure 2001-2012

Each of the VLA's 28 antennas is an 25m dish with 8 cryogenically cooled receivers, covering frequencies from 1.0 GHz (30cm) to 50 GHz (0.6 cm) in a "Y" configuration.

Three times a year, rail based "Transporters" pick up each dish and moves them, one at a time, farther down their track. Over the course of 16 months, the VLA cycles each of its legs from 2/3 to 23 miles long.

Traditionally scientists had to wait many months for a requested Observation to be taken PROVIDED their proposal is accepted in a highly over-subscribed request process….

UNTIL NOW!

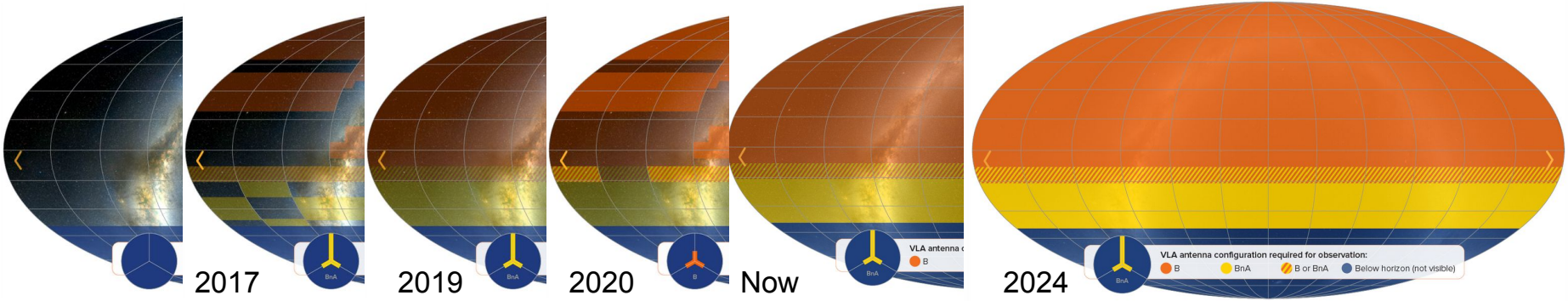
The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc

Funded by the National Science Foundation, Grant #2127548

Very Large Array Sky Survey (VLASS): Mapping the Radio Universe

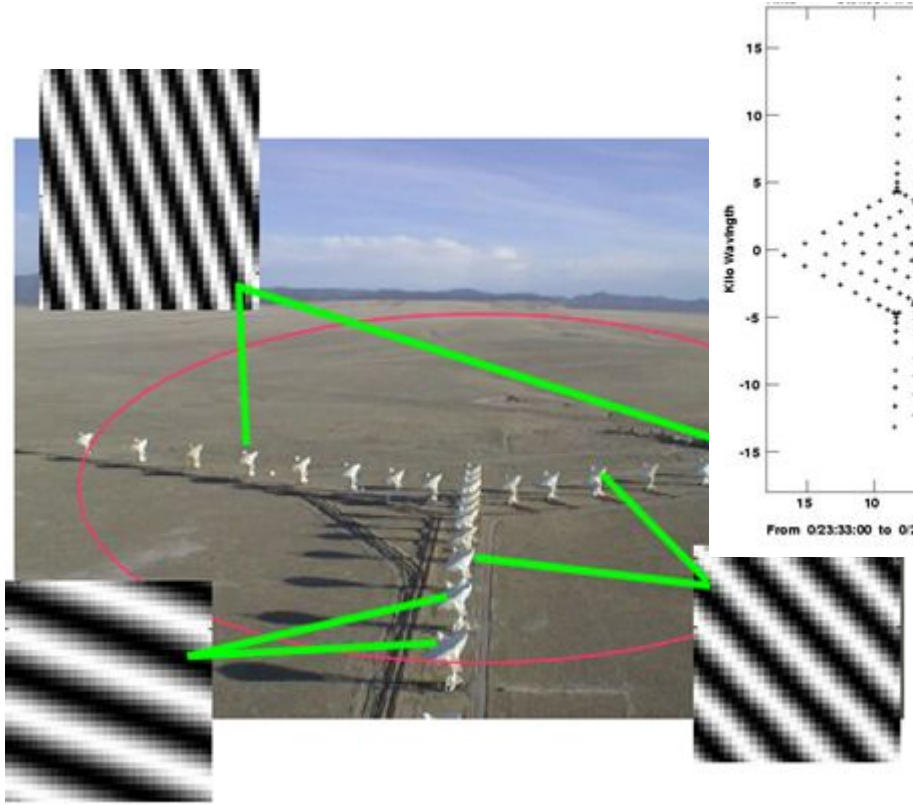
In 2017, the VLA pointed its antennas toward the northern sky and began one of the largest all-sky radio observations ever undertaken. VLASS will:

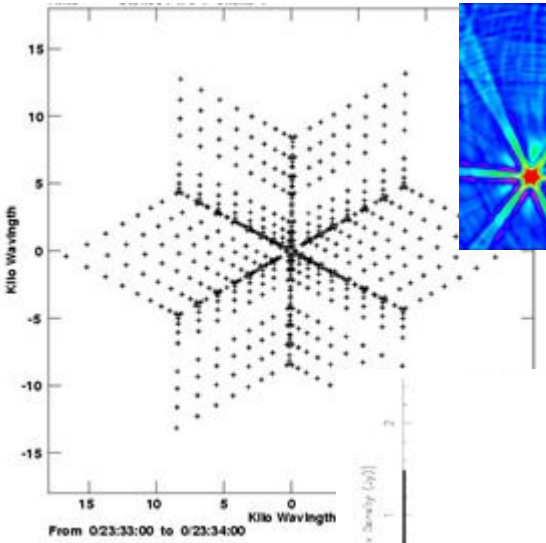
- Map 80 percent of the sky in three phases over seven years
- Produce ~100,000 high resolution data “cubes”
- Catalog approximately 10 million radio sources
- Be “multi-epoch” to reveal hidden explosions
- Cover 2–4GHz which complements (does not compete with) previous and proposed surveys
- ~5500 hours of VLA time over 7 years (6 observing cycles) generating ~2.2PetaBytes of archive quality data (Current *total* NRAO Archive is 4.2 PetaBytes x2 copies)

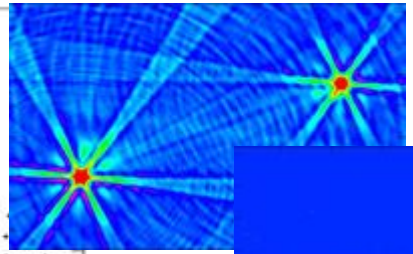


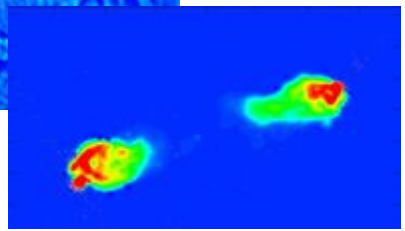
Funded by the National Science Foundation, Grant #2127548

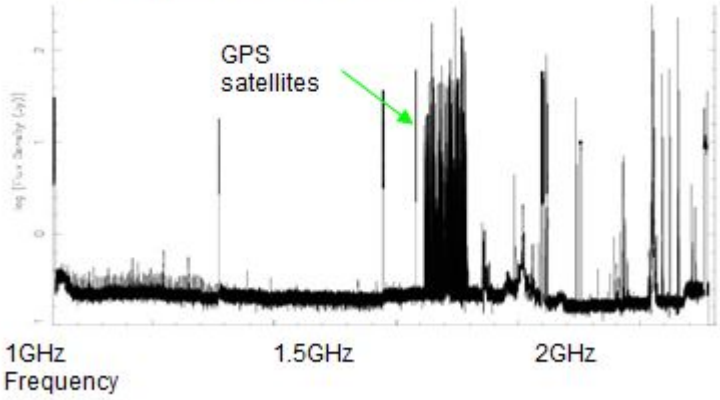
Aperture Synthesis Interferometry











Funded by the National Science Foundation, Grant #2127548

Processing needs (CapEx or OpEx)

Each of the 3 VLASS full sky epochs have processing spread over ~28 months. In-house 30 node cluster (24-core, 512 GB RAM) is sufficient for “Quick look” and “Continuum” processing over ~7 years of survey.

BUT: All-sky high quality data “Cubes” would take 16 years!

AWS cloud resources have been strategically employed, to manage peak load from re-processing.

BUT: AWS costs ~$60k/month, compared to $300k up-front in 2019, and $30k/year Data Center hosting. Even astronomers appreciate an order of magnitude problem

HTCondor is being adopted in collaboration with the University of Wisconsin–Madison and MSI Universities in STEM partnerships to be supported under **NSF Includes Project Radial**

Funded by the National Science Foundation, Grant #2127548

Processing needs (CapEx or OpEx)

spread over ~28 months.
fficient for “Quick look” and

ears!

d, to manage peak load from