DARK MATTER, RADIO ASTRONOMY, AND HIGH-THROUGHPUT COMPUTING

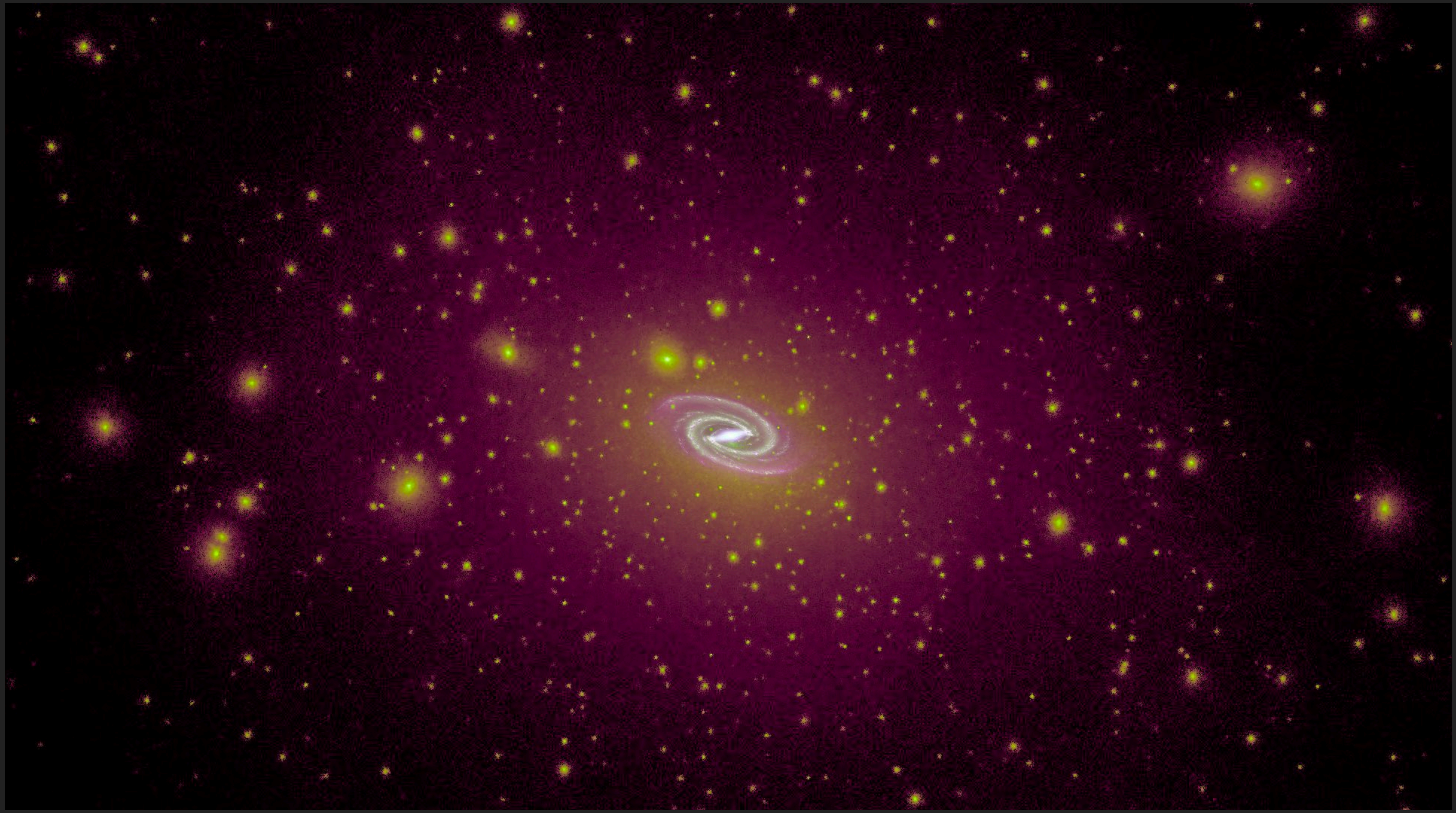
Michael N. Martinez, UW Madison Dept. of Physics

Department of Physics
UNIVERSITY OF WISCONSIN–MADISON

INTRODUCTION: MY RESEARCH

DARK MATTER

- What we see as a galaxy in space is only a small part of the total mason that region
- Invisible "dark matter" envelops galaxies and provides the backbone of structure in the universe
- Enables formation of galaxies as we know them

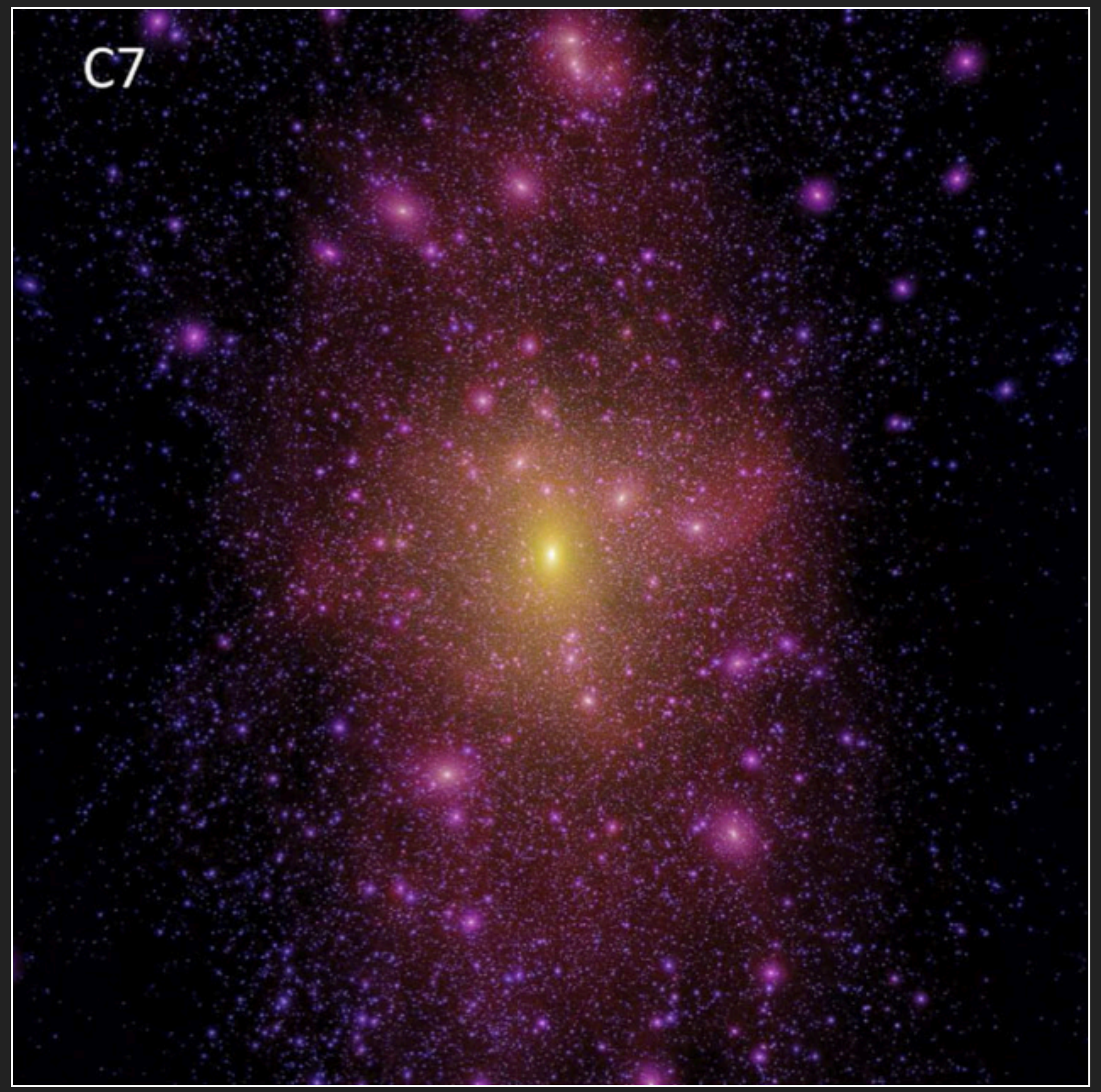


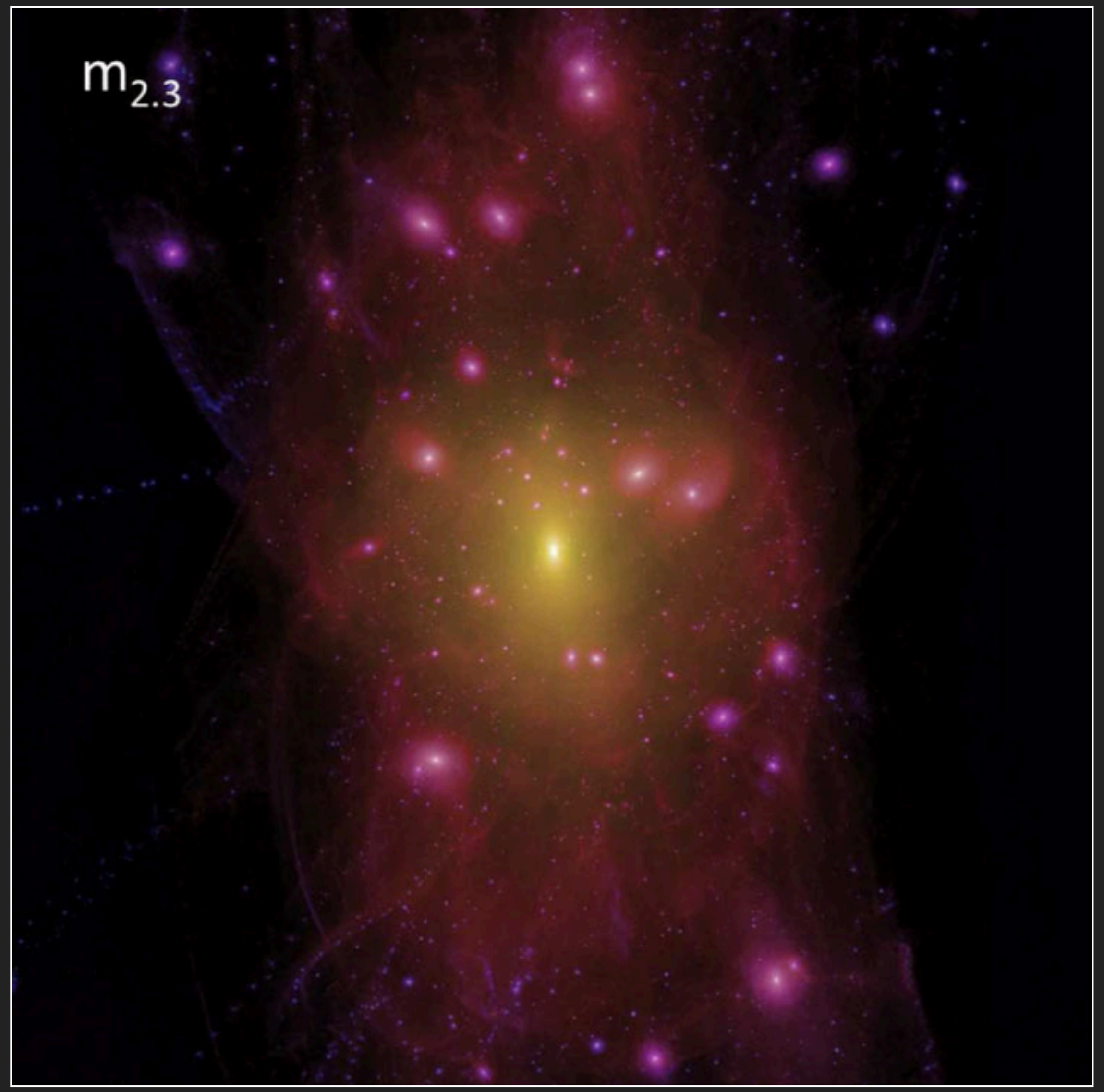
NASA (edited)

WHAT IS DARK MATTER?

- 25% of the universe's mass/energy
- Some massive particle that interacts gravitationally
- Clumps, or "halos" of dark matter, form the basic structures of the universe
- The microscopic properties of dark matter affect its macro-scale distribution!
    - Working from large to small we can learn about particle properties

"Cold Dark Matter": Halos form down to very small mass scales





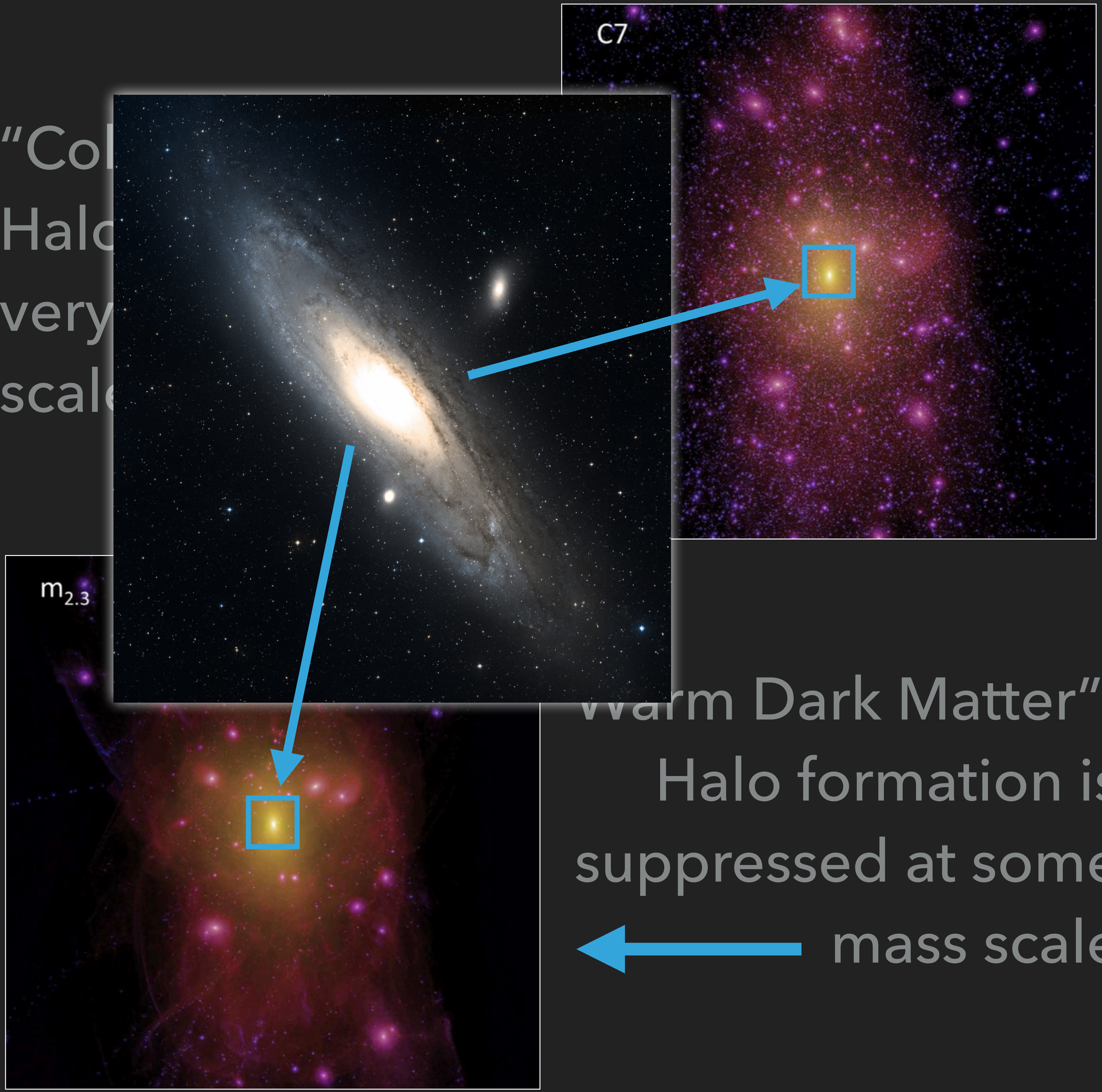
"Warm Dark Matter": Halo formation is suppressed at some mass scale

Simulations from Lovell et al. 2012

WHAT IS DARK MATTER?

- 25% of the universe’s mass/energy
- Some massive particle that interacts gravitationally
- Clumps, or “halos” of dark matter, form the basic structures of the universe
- The microscopic properties of dark matter affect its macro-scale distribution!
    - Working from large to small we can learn about particle properties

“Col
Halo
very
scale



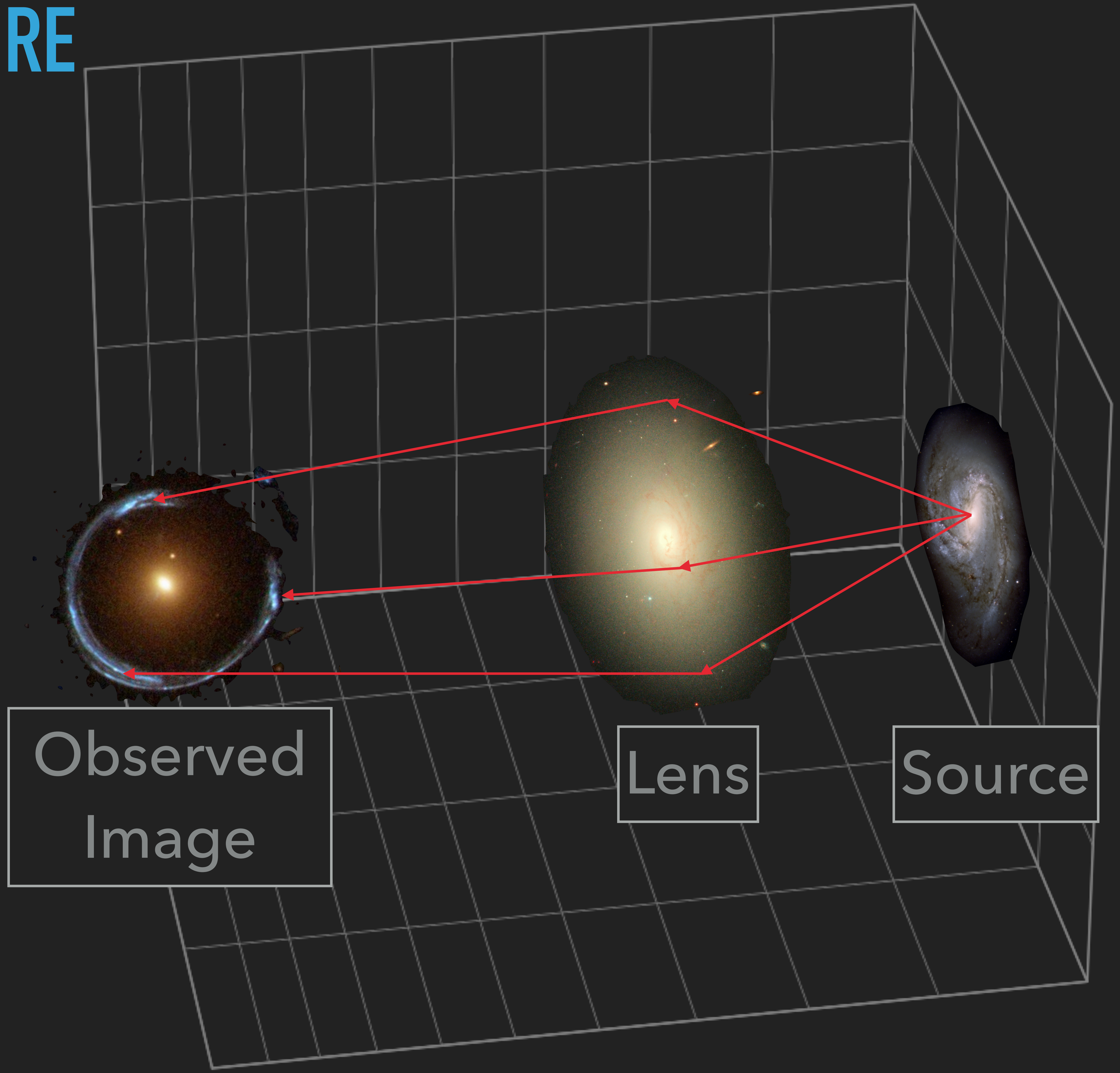
…arm Dark Matter”: Halo formation is suppressed at some mass scale

M31 image from Digitized Sky Survey

Simulations from Lovell et al. 2012

STRONG LENSING AND DARK SUBSTRUCTURE

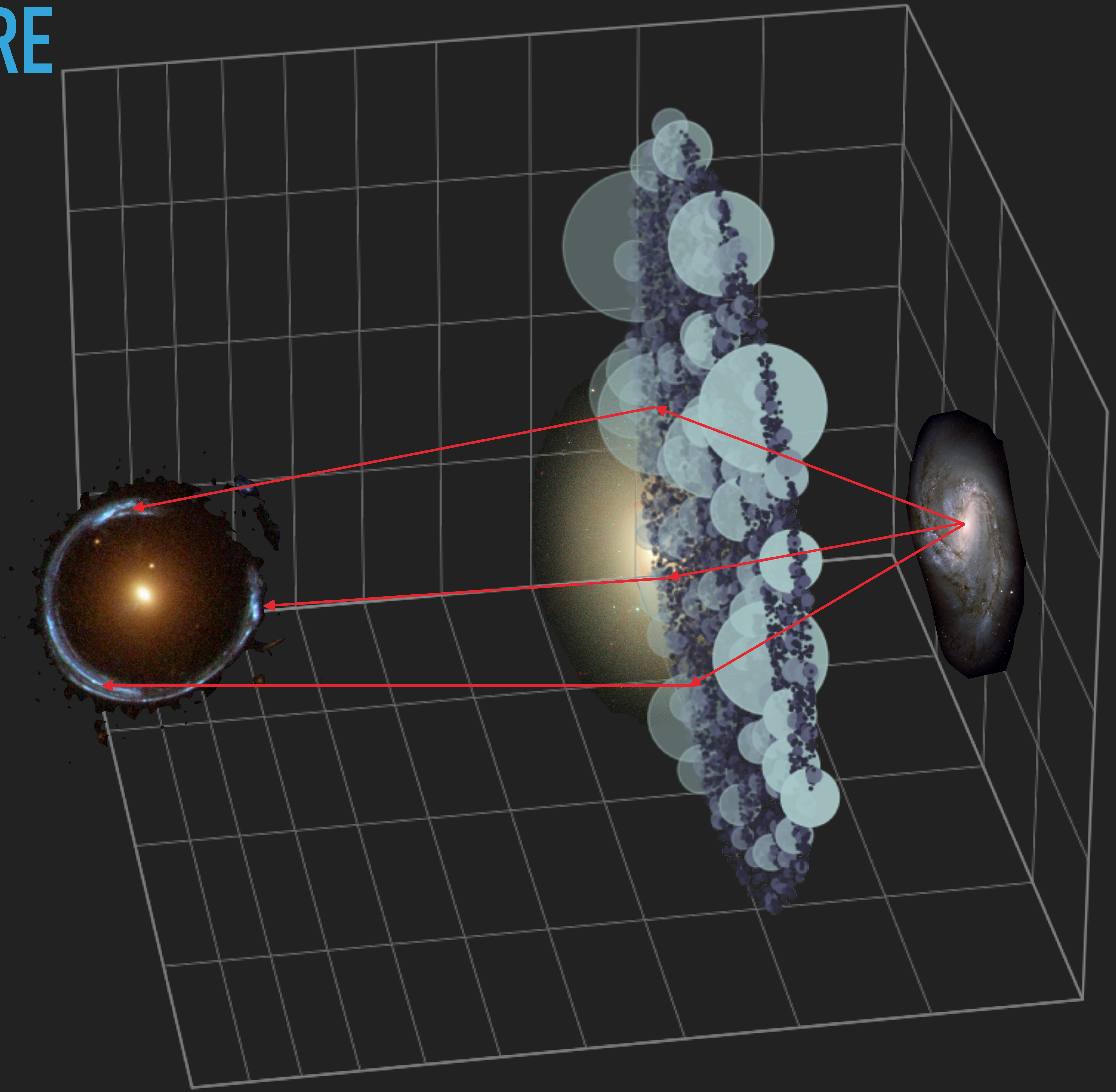
- Massive object splits a background source into multiple observed images
- Distortion of extended sources produces arc shapes around the lens
- “Cosmic telescope” with many applications



NGC 4150 image: NASA/ESA HST; NGC 3627 image: ESO; Cosmic Horseshoe image: NASA/ESA HST

STRONG LENSING AND DARK SUBSTRUCTURE

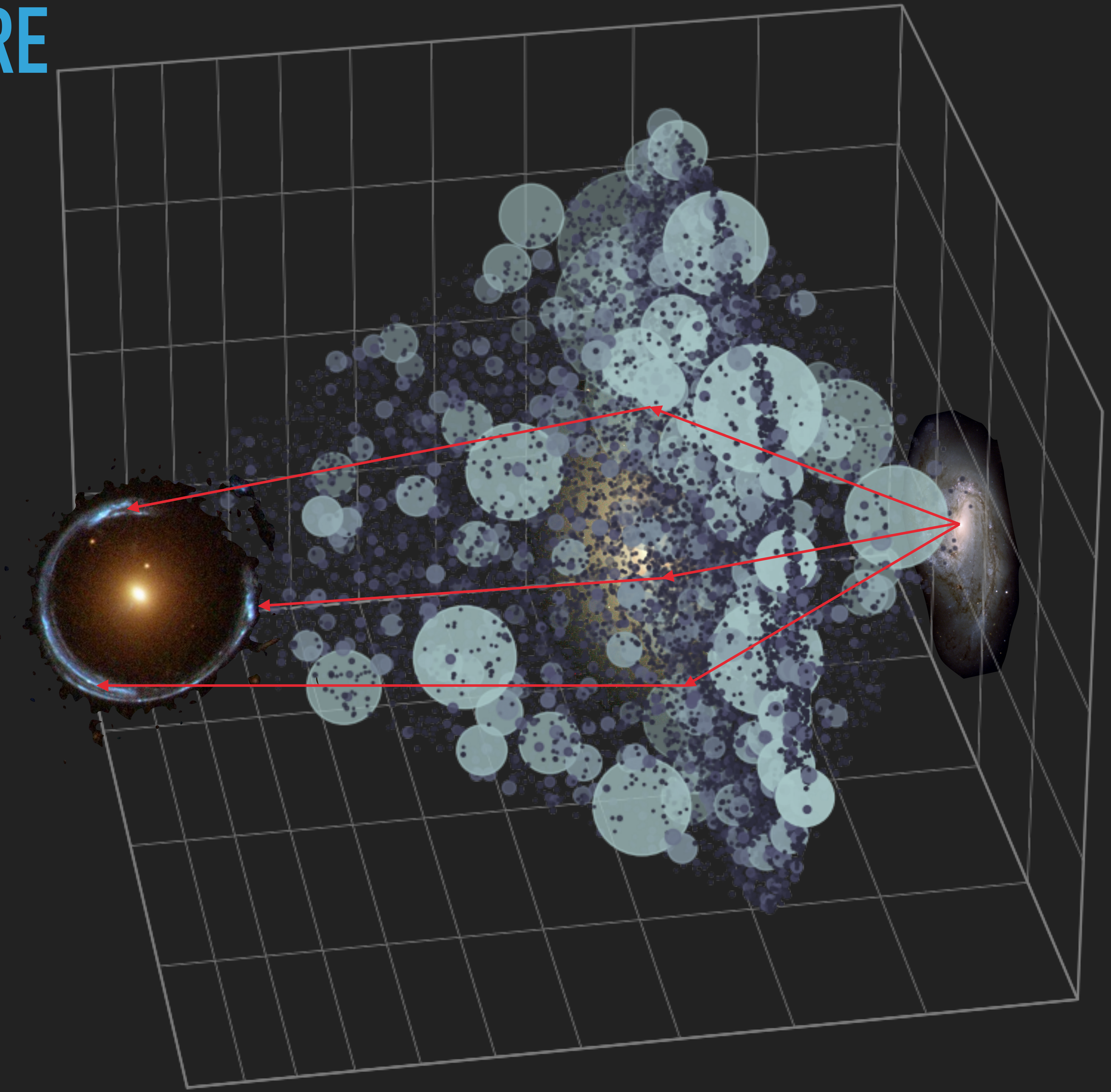
- The lensing effect is sensitive to *all* mass along the light's path
- Smaller dark matter halos can affect image positions, brightness, etc
- These halos are typically too small to form stars, making them invisible



NGC 4150 image: NASA/ESA HST; NGC 3627 image: ESO; Cosmic Horseshoe image: NASA/ESA HST

STRONG LENSING AND DARK SUBSTRUCTURE

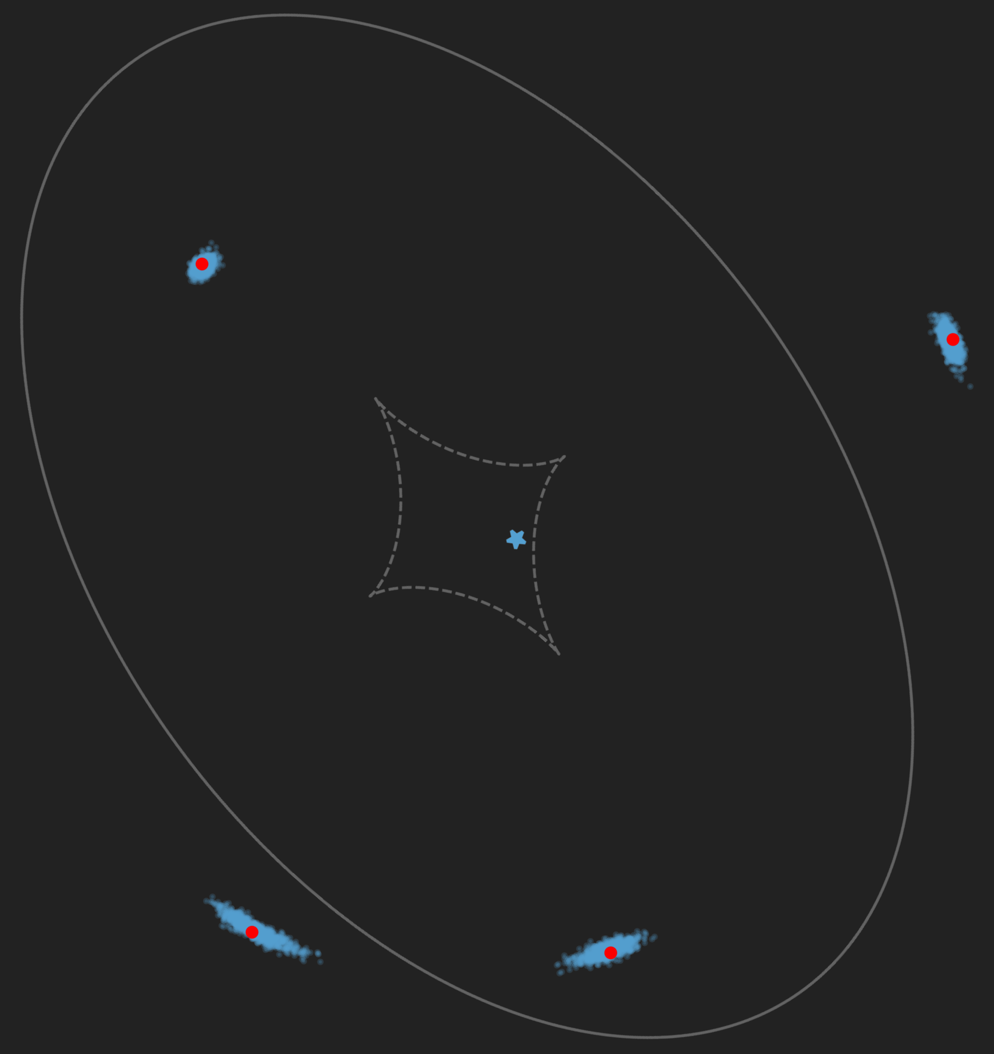
- The lensing effect is sensitive to *all* mass along the light's path
- Smaller dark matter halos can affect image positions, brightness, etc
- These halos are typically too small to form stars, making them invisible
- Other halos unaffiliated with the main lens also contribute!

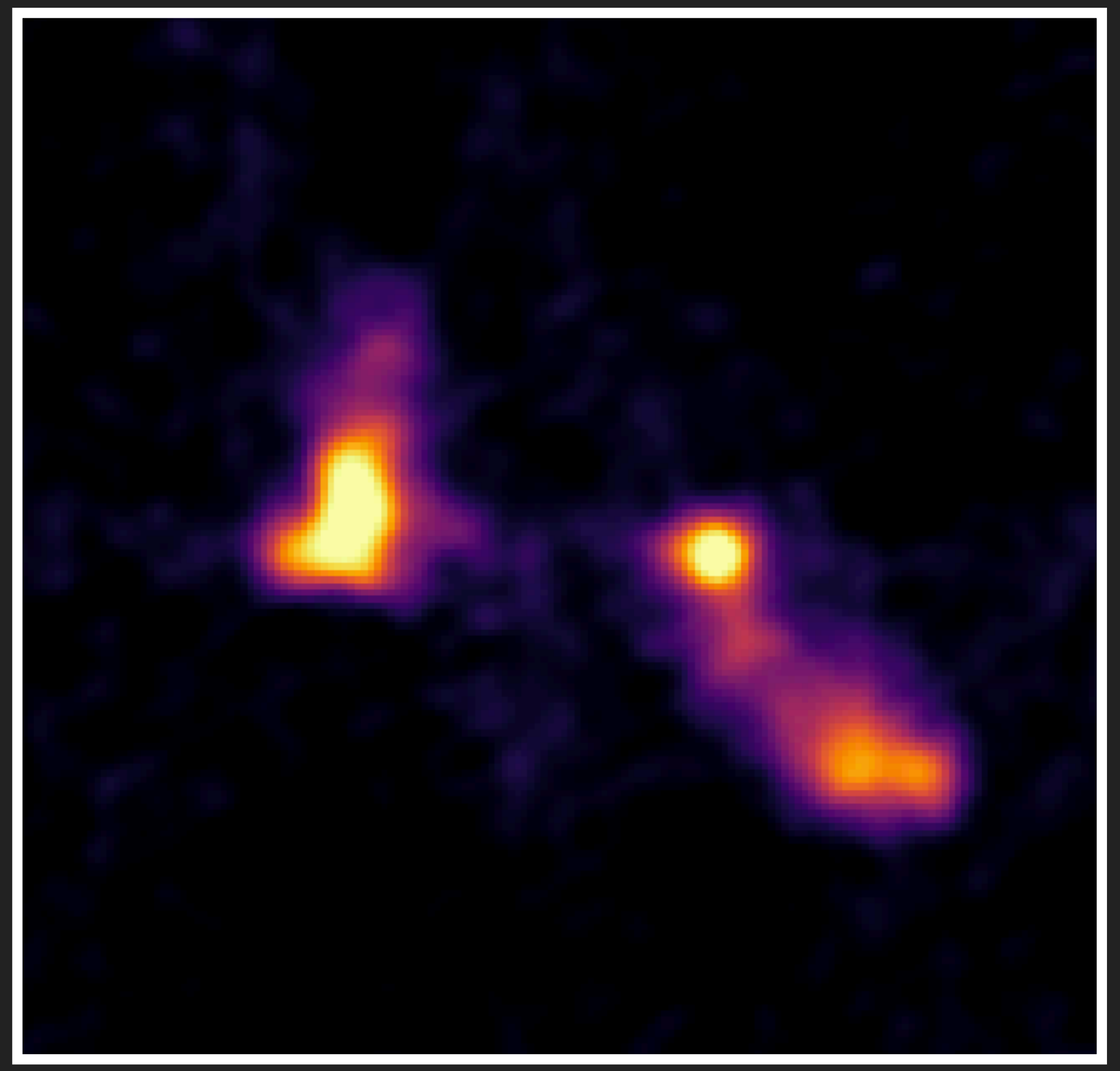


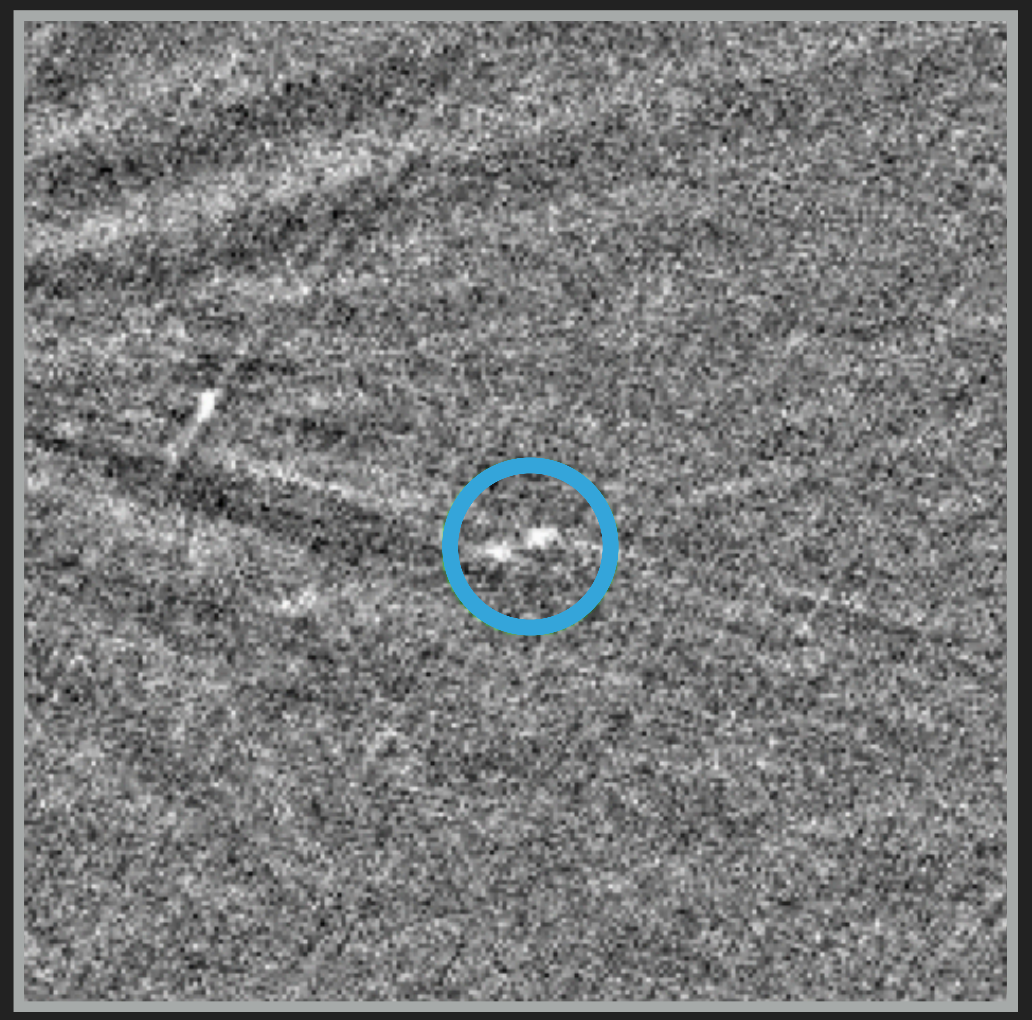
NGC 4150 image: NASA/ESA HST; NGC 3627 image: ESO; Cosmic Horseshoe image: NASA/ESA HST

COMPUTATION IN MY RESEARCH

- Simulations: How do lenses look given different models of dark matter?
- Radio imaging: Processing data from real gravitational lenses
- Image stacking: Finding more lenses by increasing sensitivity of a radio survey



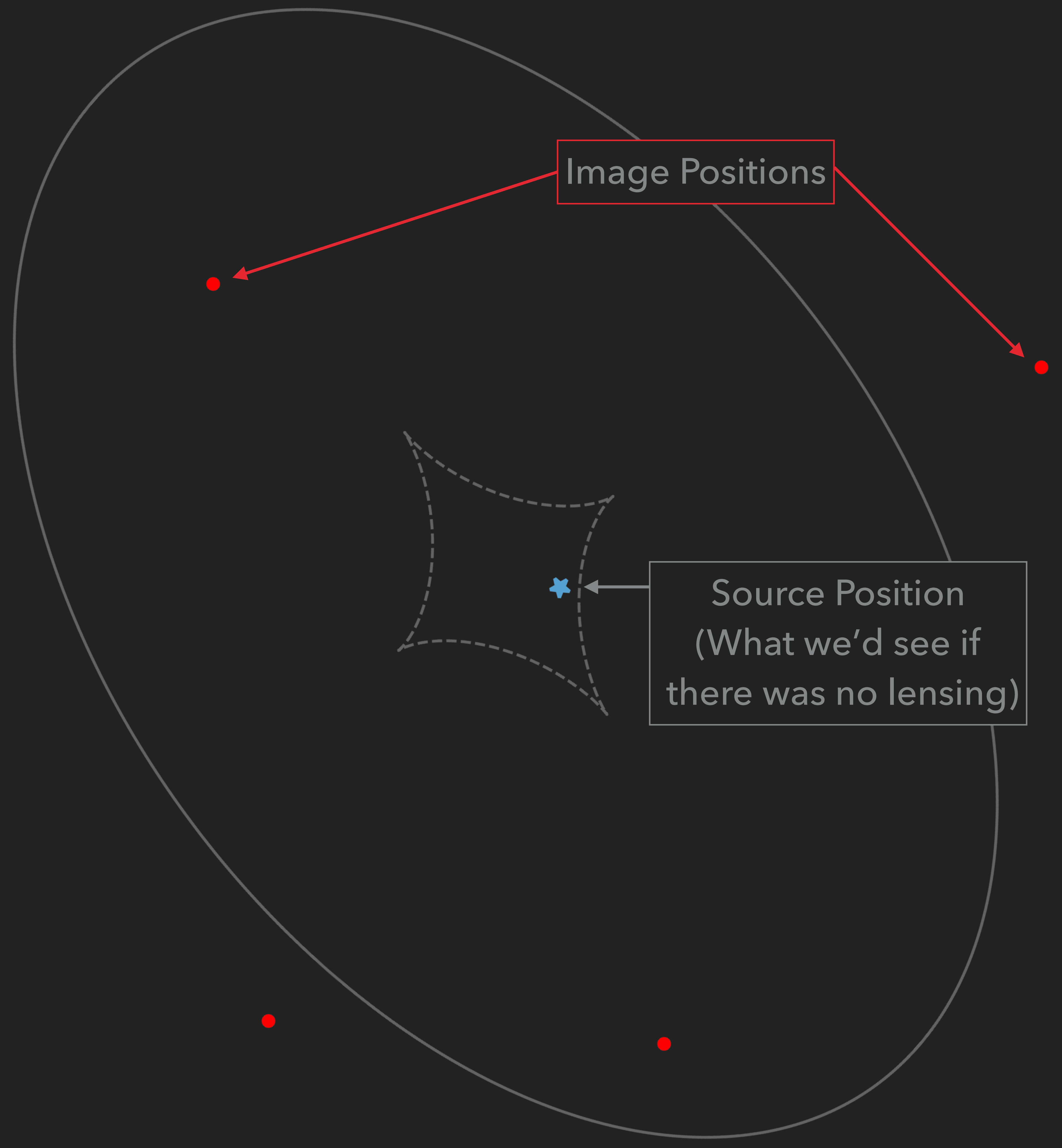




PROJECT 1: DARK MATTER SIMULATIONS

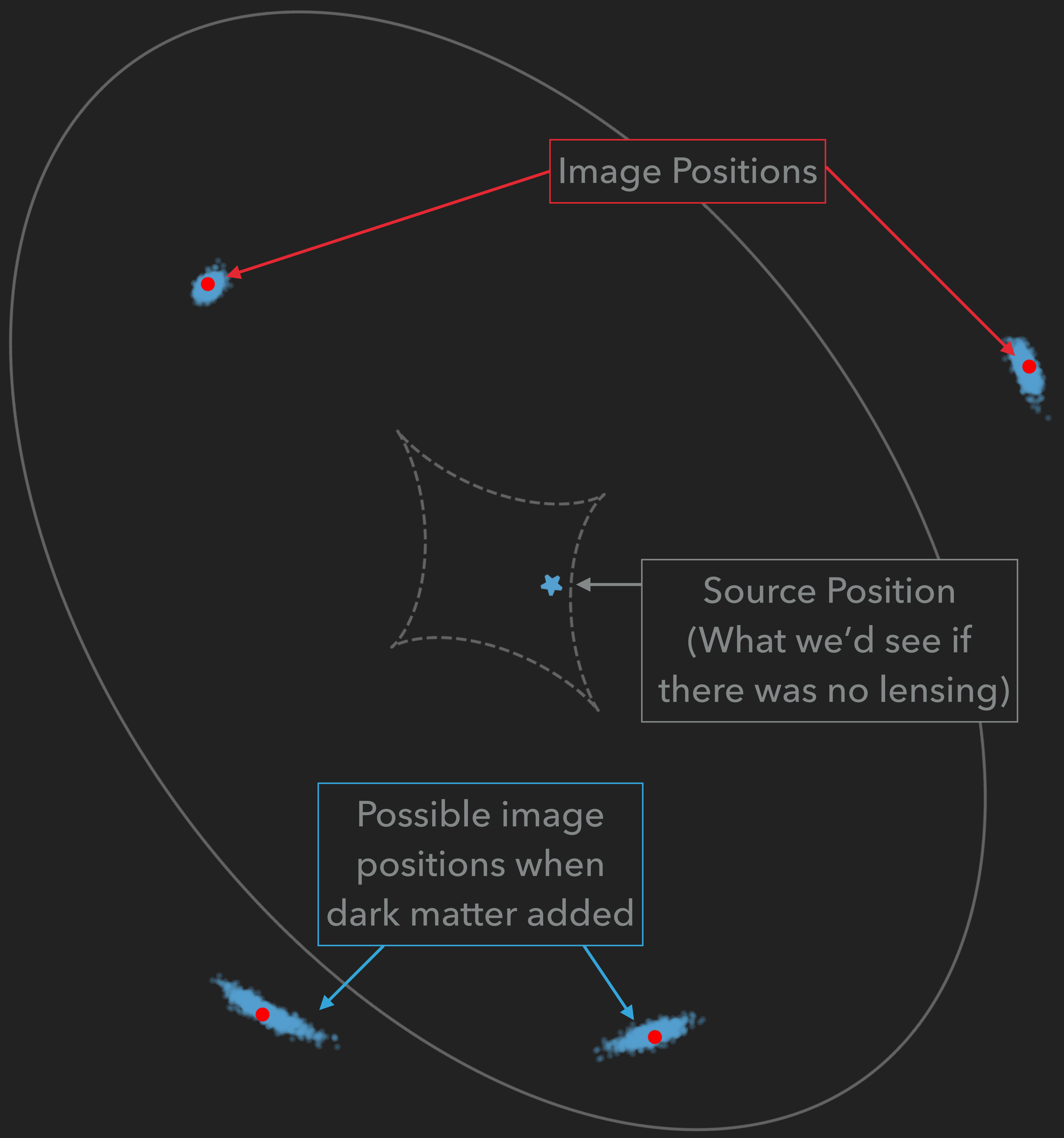
SIMULATING DARK SUBSTRUCTURE ON CHTC

- Goal: Study how different models of dark matter affect image positions
- Many dark matter halos –> simulating lensing takes a long time!
- Enter CHTC: Many simulations at once, with different model parameters
- Red points: image positions for a lens with no dark matter substructure



SIMULATING DARK SUBSTRUCTURE ON CHTC

- Goal: Study how different models of dark matter affect image positions
- Many dark matter halos –> simulating lensing takes a long time!
- Enter CHTC: Many simulations at once, with different model parameters
- Blue cloud: image positions for 1000 full substructure simulations



MOVING TO CHTC

- Very natural task for high-throughput computing
- Python Apptainer image with needed packages
- Each job does its own set of simulations, I collect them after
- ~2 days start to finish, computing took under an hour

```
#Apptainer image
Bootstrap: docker
From: python:3.11

%post
    python3 -m pip install numpy scipy==1.13.1 astropy
            matplotlib lenstronomy==1.11.10 colossus
            mcfit pyhalo==1.2
```

```
#Condor Submission File
container_image = file:///staging/mnmartinez/image.sif
executable = script.py
arguments = $(Cluster)_$(Process)
should_transfer_files = YES
when_to_transfer_output = ON_EXIT_OR_EVICT

log = job_$(Cluster)_$(Process).log
error = job_$(Cluster)_$(Process).err
output = job_$(Cluster)_$(Process).out

requirements = (HasCHTCStaging == true)
request_cpus = 1
request_memory = 2GB
request_disk = 1GB

queue 50
```

PROJECT 2: RADIO INTERFEROMETRY

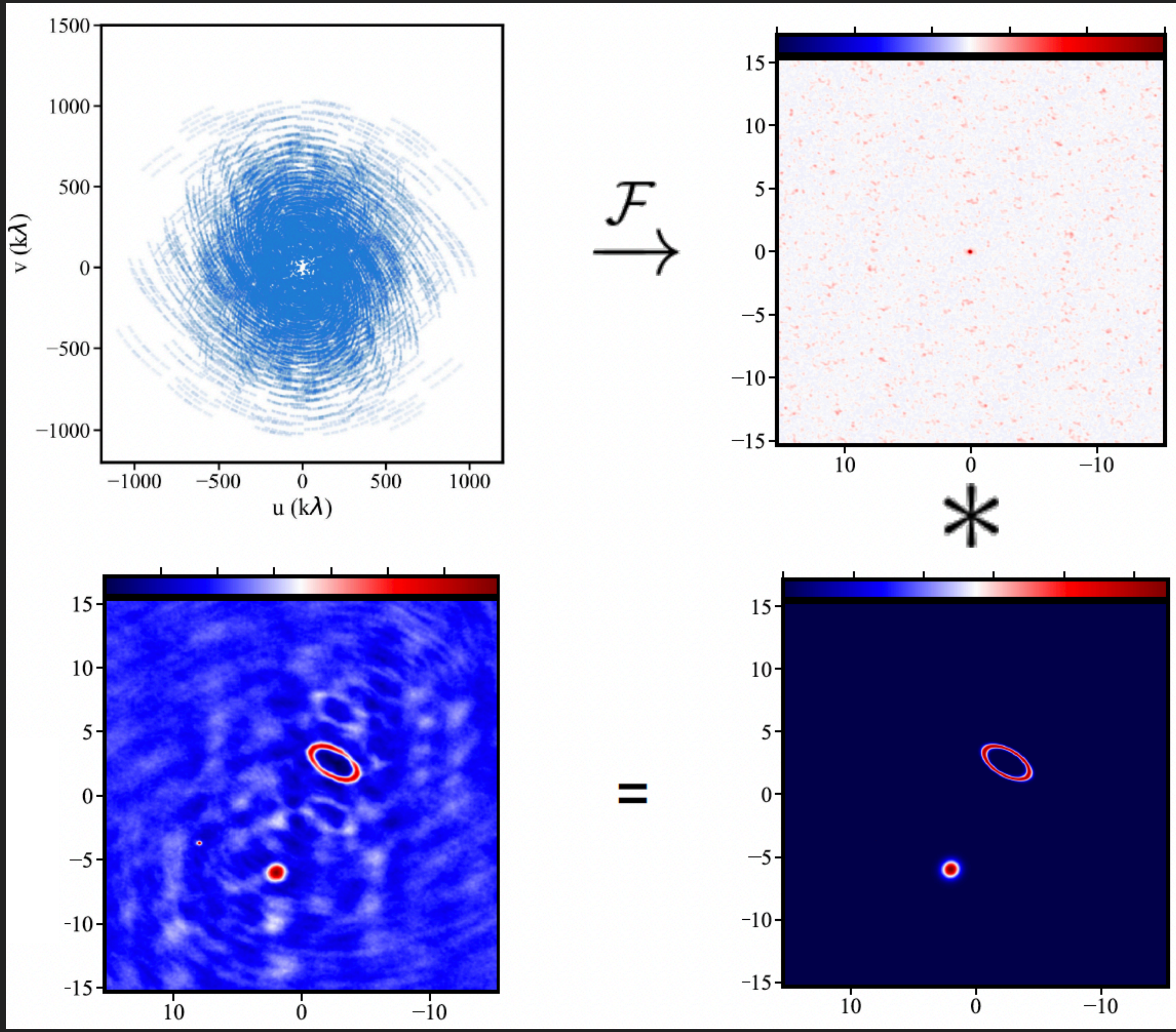
RADIO INTERFEROMETRY

- Radio waves are too long to create high-resolution images with one telescope
- Signals from multiple dishes are combined to create sky image
- Creates effective "telescope" as large as the distance between antennas
- Rather than measuring the sky brightness itself, measures its Fourier transform

NRAO/AUI/NSF

RADIO DATA CHALLENGES

- Radio datasets are massive (up to 100s GB)
- Creating an image requires a lot of computing power
  - FT, convolutions, and more
- Limited opportunity for parallelization (more cores = good, but everything must be on one node), and memory intensive
- Can still use HTC (super-powered laptop)



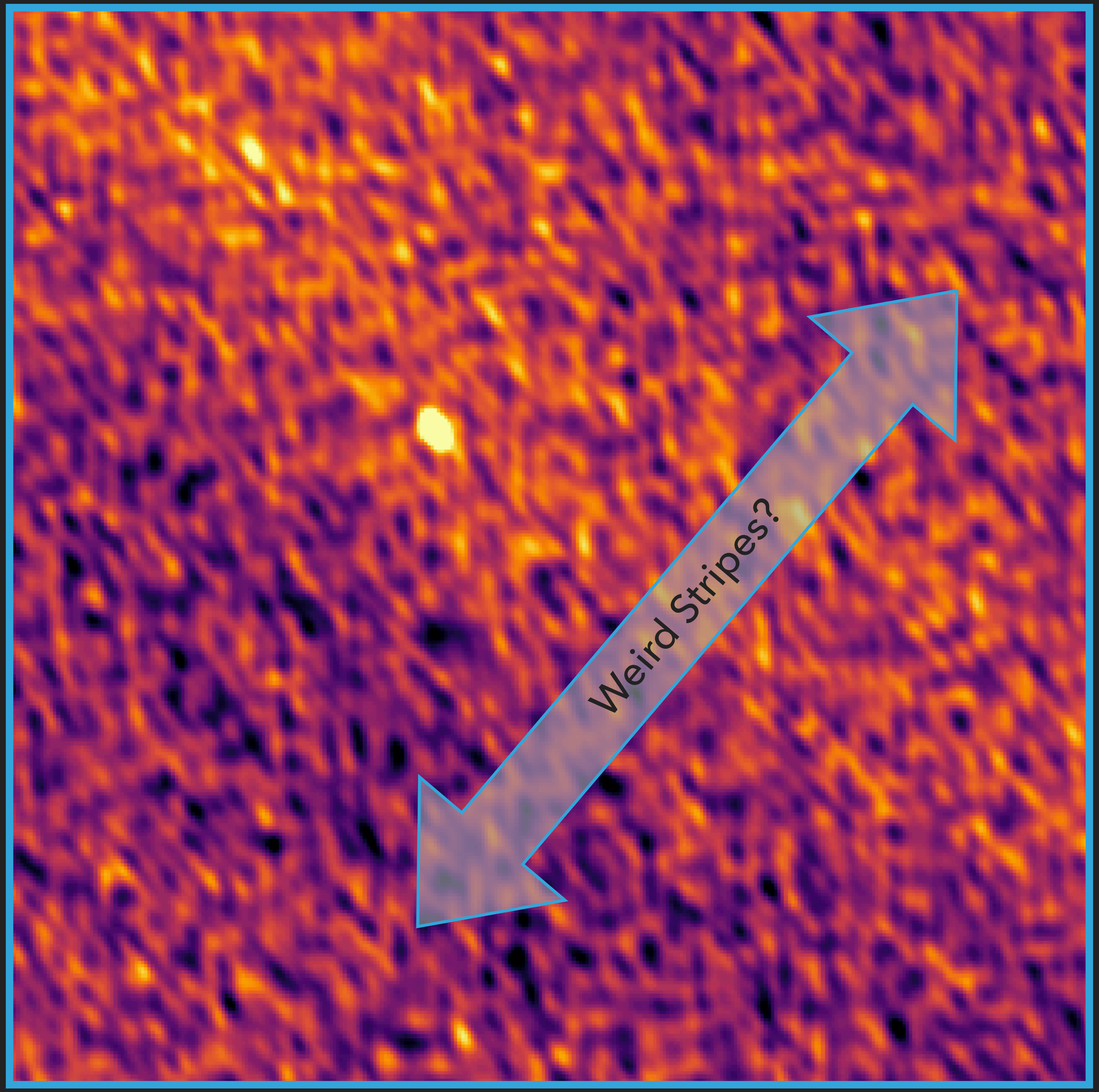
D. Wilner 2023

MOVING TO CHTC

- >1 month to get everything set up
- Custom software in a container, MPI
- Normally interactive process moved to scripts
- Huge datafiles required special transfer procedures
- UW CHTC staff have been extremely helpful!

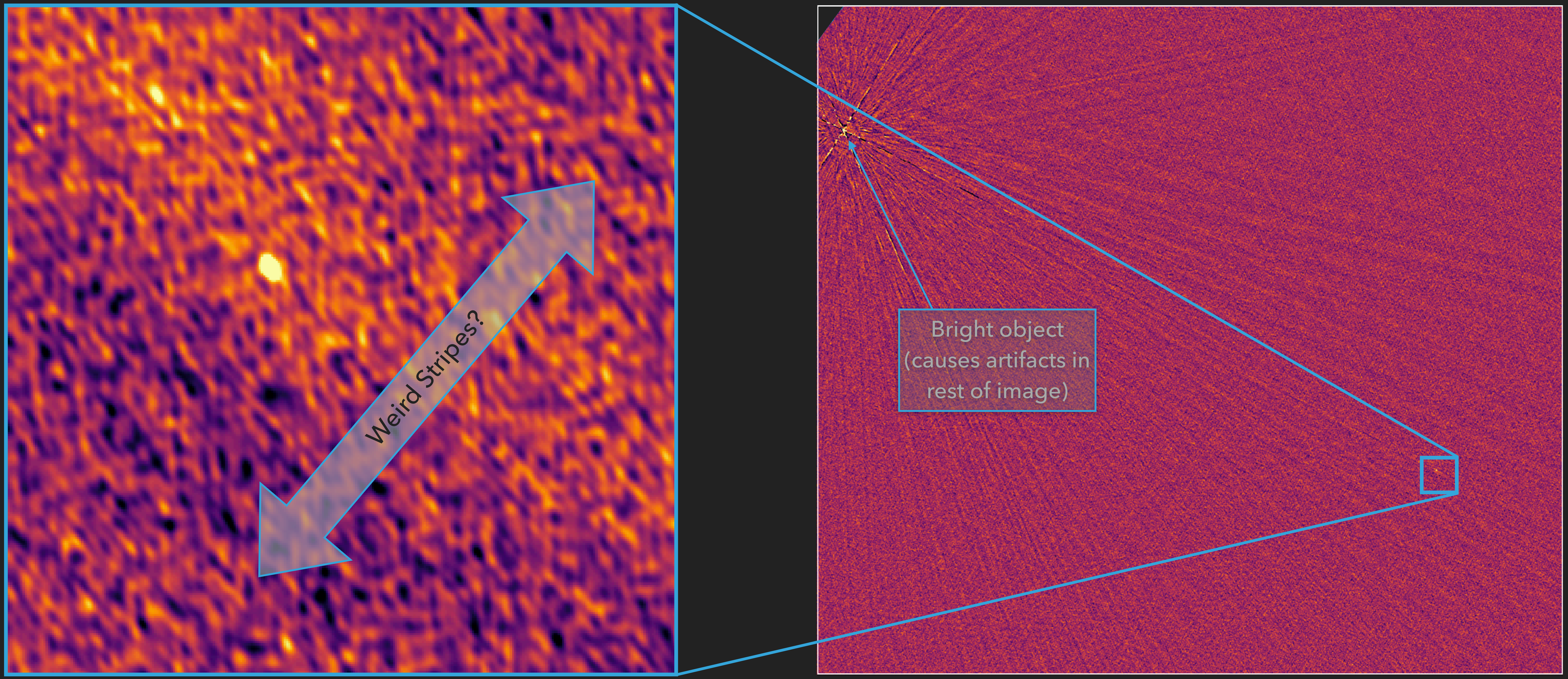
```
Bootstrap: docker
From: ubuntu:24.04

%post

    apt-get update -y --allow-insecure-repositories
    apt-get install python3.12-venv -y
    alias python=python3.12
    python -m venv env --upgrade-deps
    . env/bin/activate
    python -m pip install --upgrade pip
    python -m pip install numpy==1.26.4
    pip install matplotlib
    pip install scipy
    apt-get install ImageMagick*
    apt-get install gfortran -y
    apt-get -y install libnsl2 -y
    apt-get install libgslcblas0 -y
    apt-get install libcanberra-gtk-module -y
    apt-get install build-essential -y
    apt-get install xvfb -y --allow-unauthenticated
    apt-get install zstd -y --allow-unauthenticated
    apt-get install parallel -y
    apt-get install git -y --allow-unauthenticated
    apt-get install wget -y --allow-unauthenticated
    apt-get install openmpi-bin openmpi-doc
    apt-get install libopenmpi-dev
    apt-get install libgsl-dev -y
```

EXAMPLE 1: A LARGE IMAGE

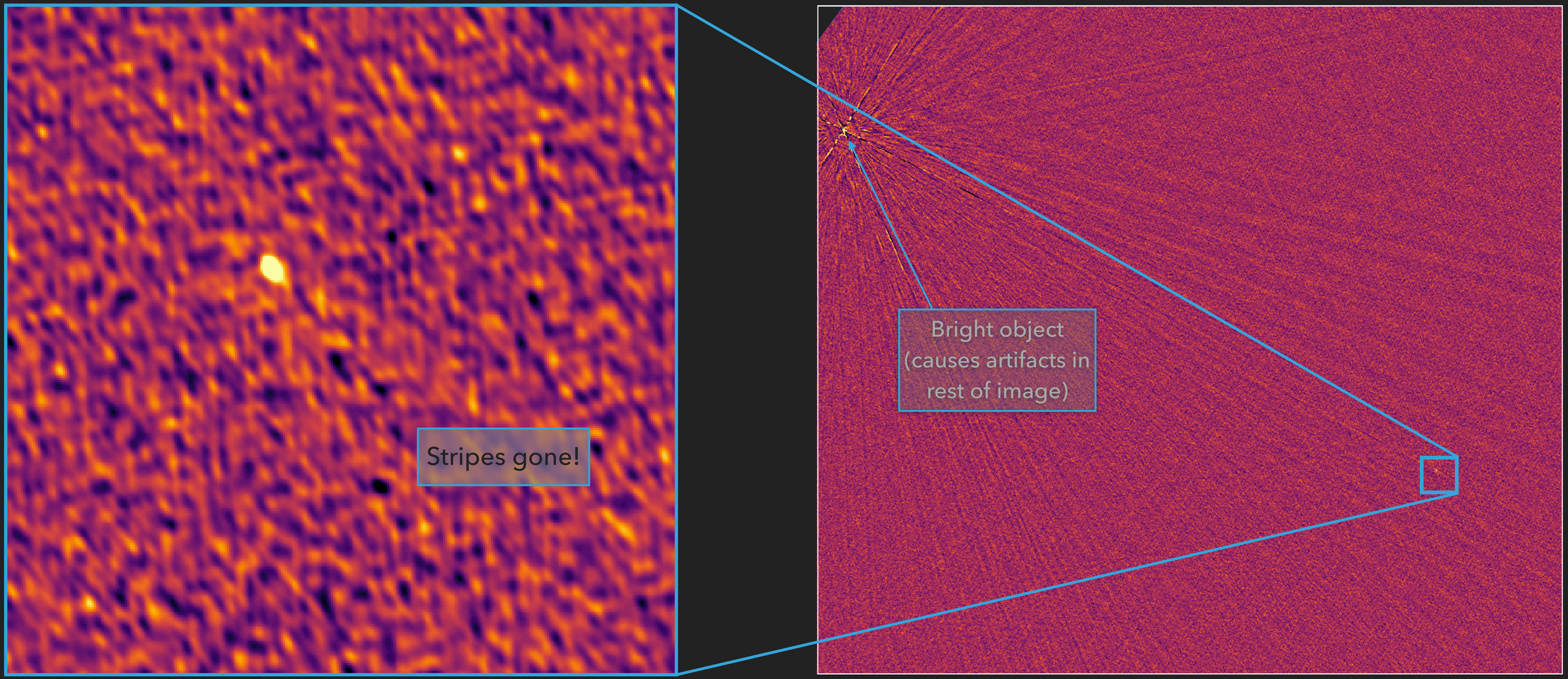


- Laptop image: 320x320 pixels
- Structured noise in image
- Probably due to bright object somewhere else in the image
  - If so, noise can be eliminated
- Created a 4320x4320 pixel image on CHTC to find out

EXAMPLE 1: A LARGE IMAGE



EXAMPLE 1: A LARGE IMAGE



EXAMPLE 2: MANY SOURCES IN ONE FIELD

- Project: look for gravitationally lensed sources that emit radio waves
- Low-resolution survey image suggests target near many other sources
- Outlying sources can mess with image quality
- Can analyze simultaneously given enough computing power

Optical survey image from SDSS9

Optical survey image from SDSS9, close-up from DECaLS DR10

Optical survey image from SDSS9

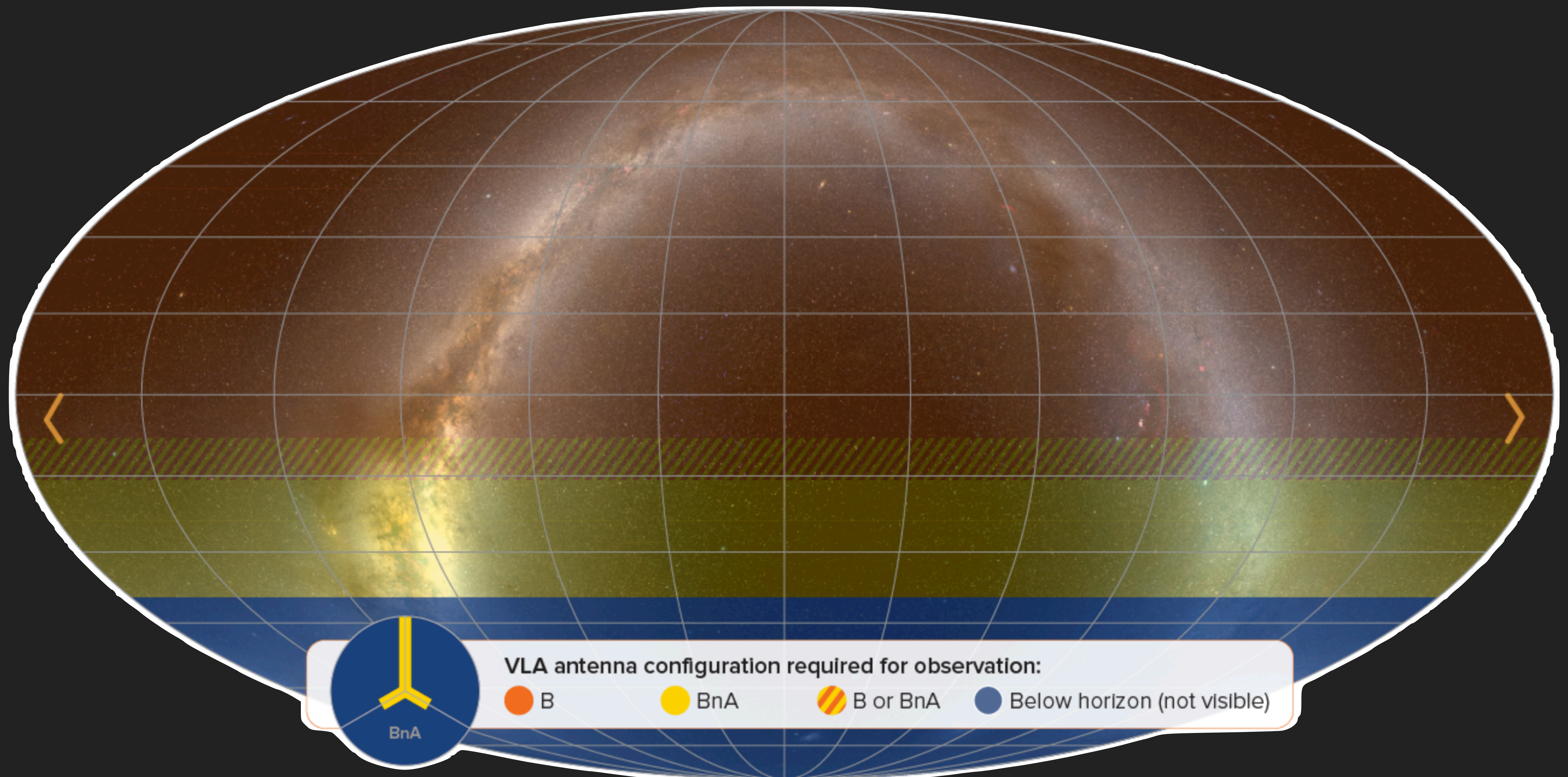

Image Edge
Wide-field radio image from RACS (ATNF, CSIRO)


Wide-field radio image from RACS (ATNF, CSIRO)

PROJECT 3: IMAGE STACKING

THE VERY LARGE ARRAY SKY SURVEY

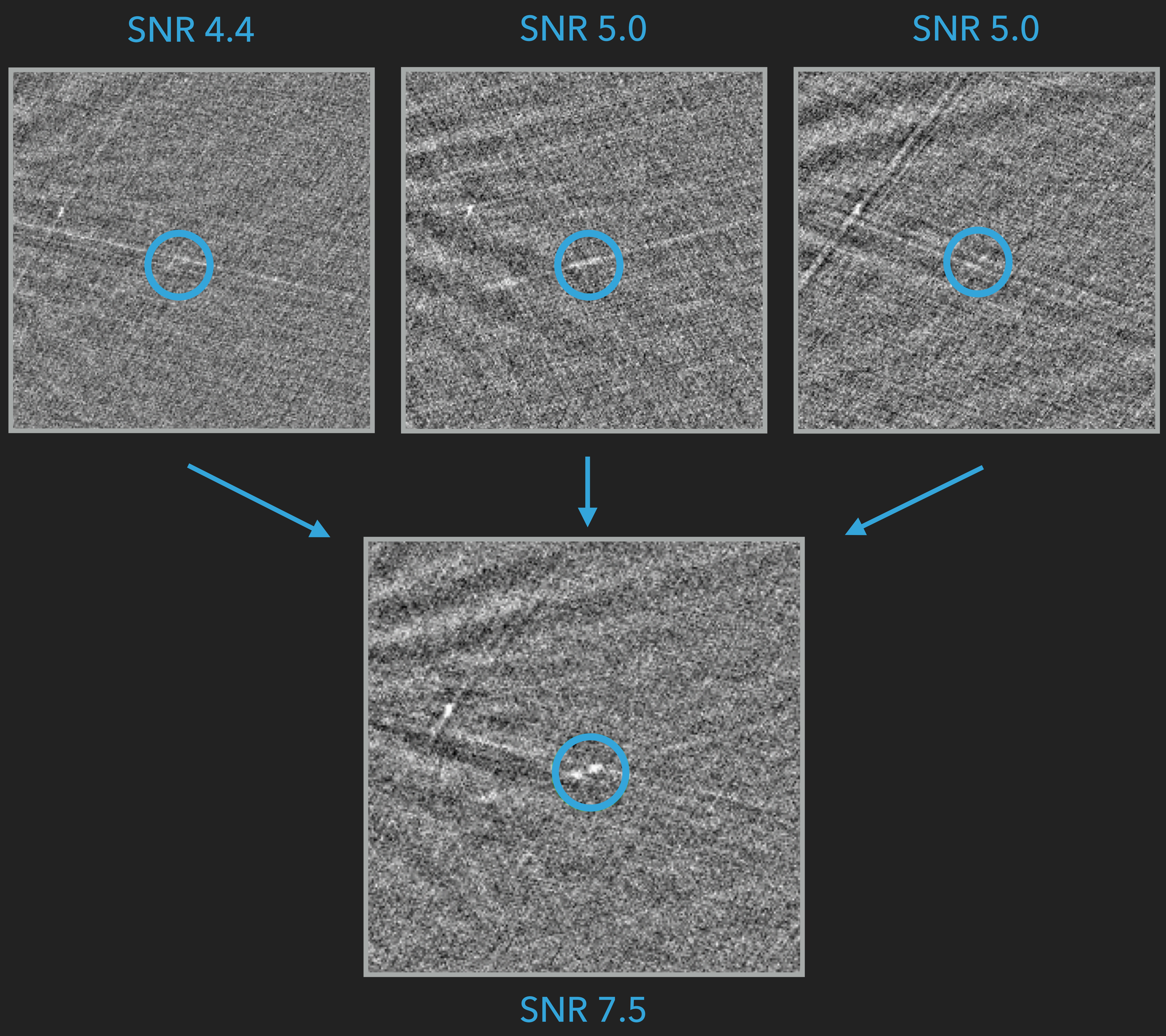
- Radio survey of 3/4ths of the sky
- Time variability survey: 3 epochs between 2017 and 2024
- Used to identify gravitational lens targets for high-resolution followup



NRAO

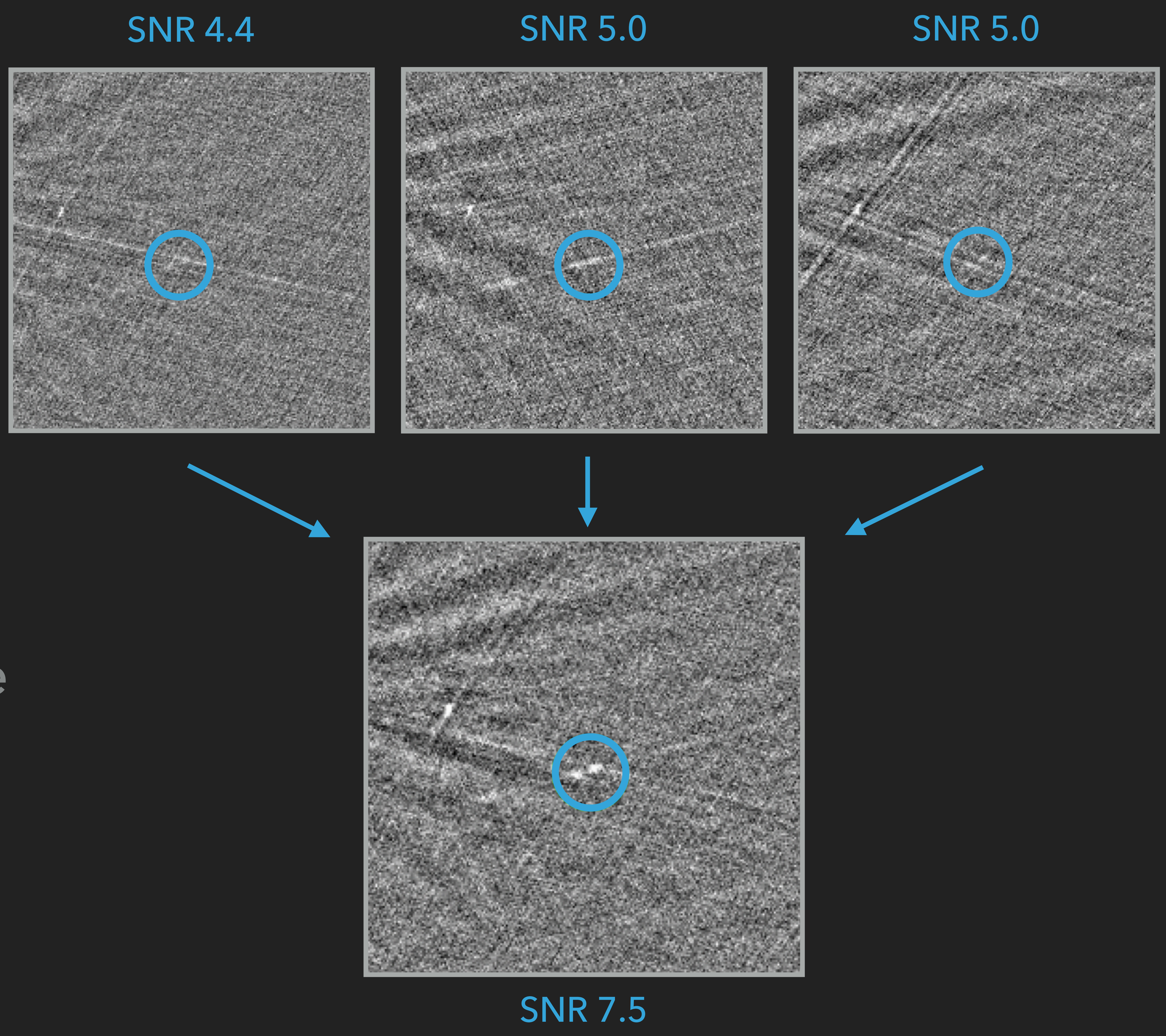
IMAGE-PLANE VLASS STACKING

- Can achieve greater sensitivity by combining the three observational epochs
  - More lenses to observe
- Laptop: about 5 minutes per square degree
- Full survey is 30,000 square degrees
- ~100 days on a laptop



VLASS STACKING ON CHTC

- The perfect HTC use case?
- Each job runs one square degree
    - Stacks images, creates catalogs
    - ~2 hours per job
- Entire survey theoretically stackable in 2 hours (given 30000 jobs)
- More realistically: 1 week?



MOVING TO CHTC

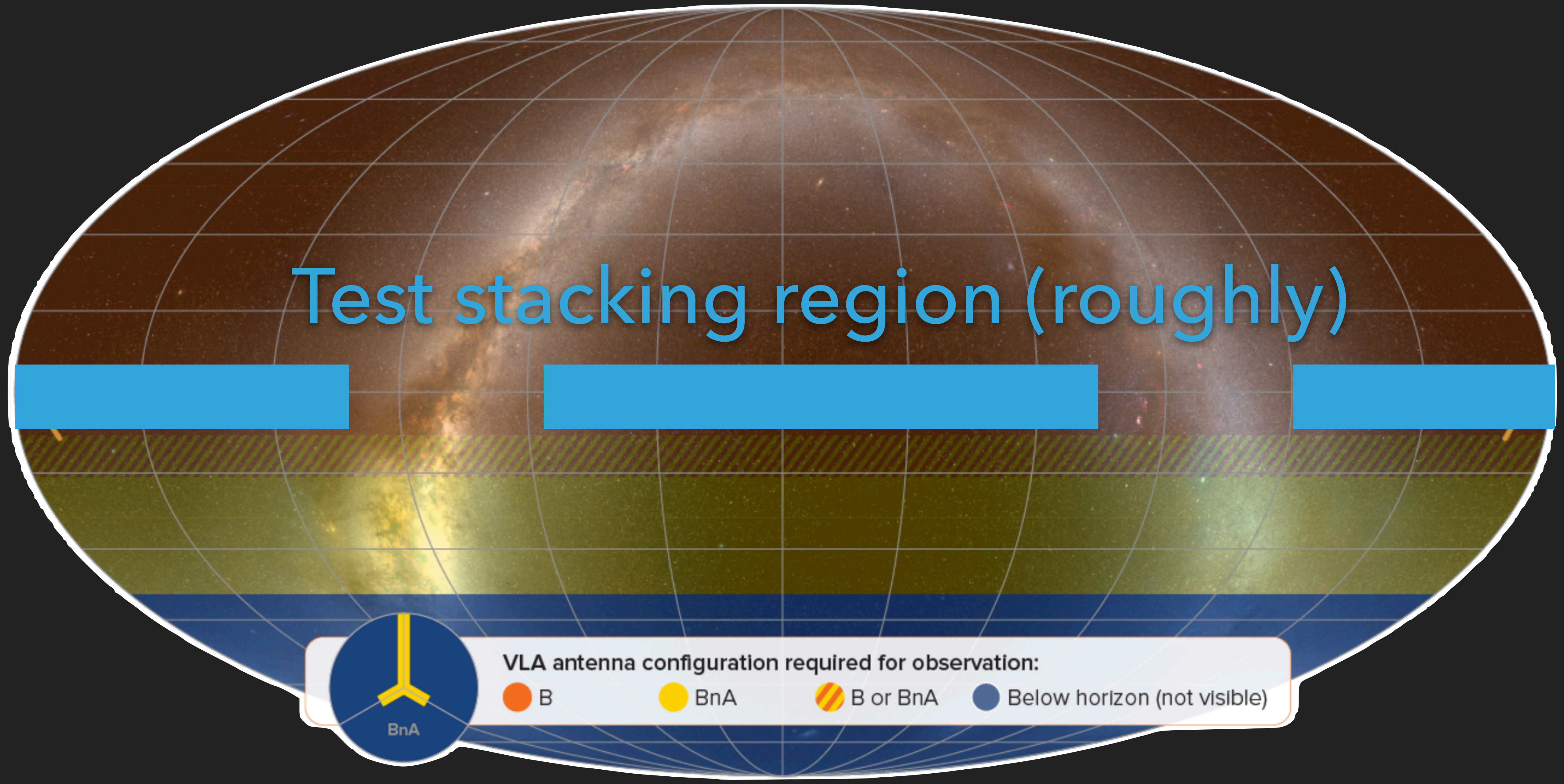
- Image transfer from a server rather than /staging/
- Each image needs different files
- Used a wrapper script to select a different line of a file per job, then run the stacking program
- Future challenge: getting image data off of CHTC for analysis

```
#!/bin/bash
#Wrapper script
head -n $1 tiles.txt | tail -n 1 > foo
sed 's/,/\n/g' foo > bar
wget -i bar
rm foo
rm bar
python3 stack_vlass.py
rm *.ql.*subim.fits
```

```
#Submission file
container_image = file:///staging/mnmartinez/stack.sif
plusone = $(Process) + 1
NewProcess = $INT(plusone,%d)
executable = stack_wrapper.sh
arguments = $(NewProcess)
should_transfer_files = YES
when_to_transfer_output = ON_EXIT_OR_EVICT
transfer_input_files = stack_vlass.py, tiles.txt
log = job_$(Cluster)_$(NewProcess).log
error = job_$(Cluster)_$(NewProcess).err
output = job_$(Cluster)_$(NewProcess).out
requirements = (HasCHTCStaging == true)
request_cpus = 1
request_memory = 4GB
request_disk = 10GB
queue 40
```

VLASS STACKING ON CHTC: CURRENT PROGRESS

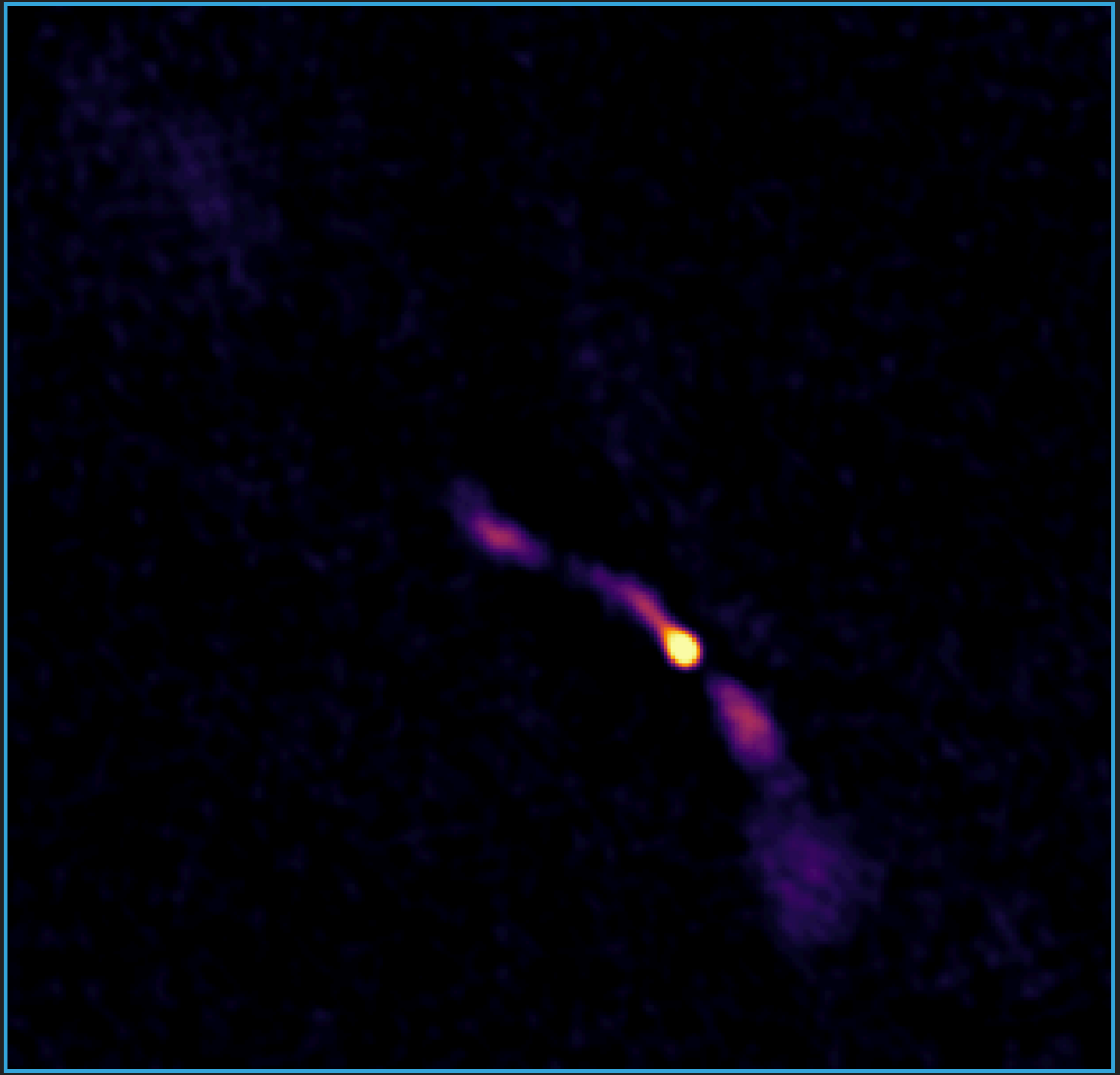
- Preliminary tests with ~40 square degrees have been promising
- Refine procedure/ensure image quality with smaller subset of sky (~1000 square degrees)
- Once details are ironed out, move to whole survey
- Undergraduate research at UW



NRAO

CONCLUSION

- A research project has many different parts
  - Each part can have its own computing requirements
- Computing resources like CHTC or OSG can help with more than you might initially think!



THE END

Questions welcome!