Juan Lorente Guarnieri 816020

# Final project: Very-long-baseline Interferometry

## Understanding, running, and modifying the code

In this section we address in turn each of the key conceptual questions, and then lay out how we will organize our presentation of the experiments prescribed in (a), (b) and (c).

### The input image and input telescope array

- **How is the input image defined?**
  The notebook reads a plain-text file where each line gives the x, y coordinate on a regular grid and four values I, Q, U, V. These are parsed by eh.image.load_image into a 2D array of Stokes parameters.

- **What do I, Q, U, V represent?**

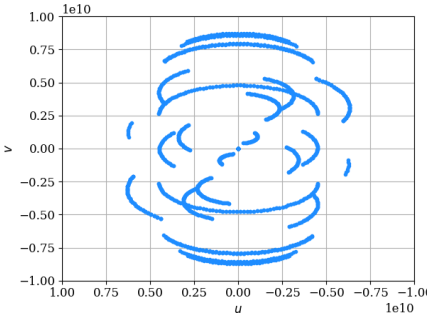
  - **I** = total intensity (brightness)

  - **Q, U** = linear polarization components

  - **V** = circular polarization
    In the tutorials it is used only Stokes I (total intensity) for imaging; Q/U/V are set to zero by default.

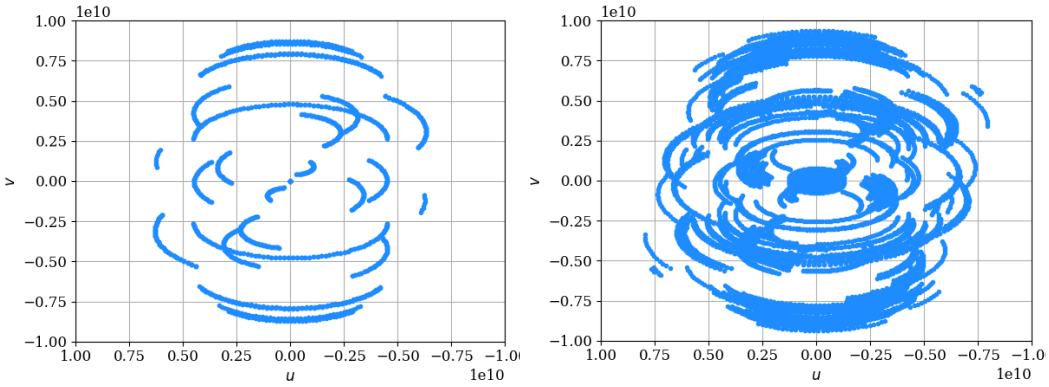
### How telescope arrays sample the sky (U–V coverage)

- **U–V coverage plots**
  Each baseline (pair of telescopes) as the Earth rotates sweeps out a “track” in the 2D spatial-frequency plane (the *u*–*v* plane).




In the plot, each colored line is one baseline’s samples over time. When a telescope dips below the horizon, its baselines disappear, creating gaps.

- **Switching arrays**
  By loading different station lists (e.g. `arrays/EHT2017_m87.txt` vs. `arrays/EHTII.txt`), we see immediately how adding or removing sites changes the density and orientation of these tracks—denser, more isotropic coverage yields better imaging potential.

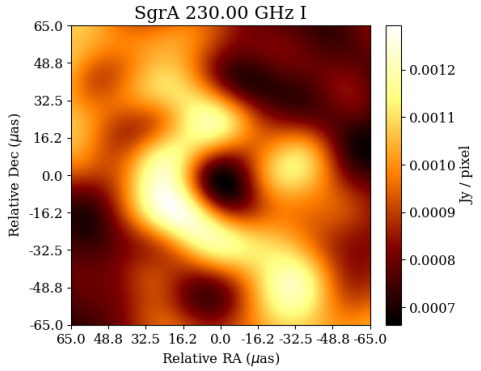



EHT2017_m87 EHTII

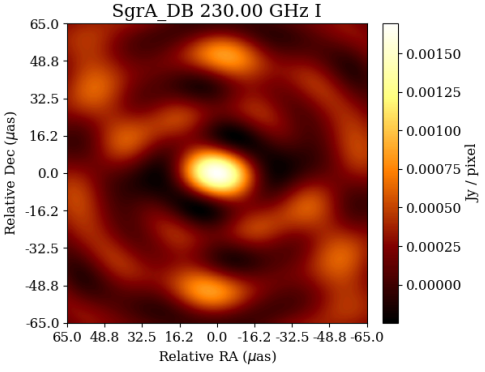
## Dirty image and dirty beam

The results of the dirty image, dirty beam & clean beam are from using the jason_mad_eofn dataset, to see better the results.

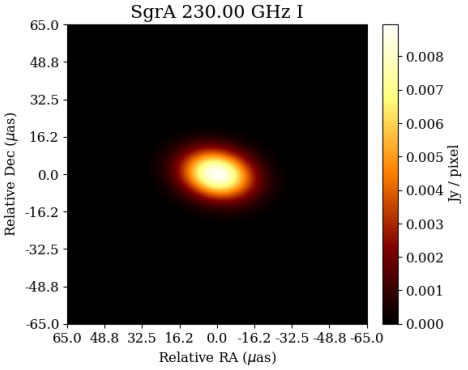
- **Dirty image**: inverse FFT of the sampled visibilities (with zeros where no data exist). It looks blurred and full of sidelobe artifacts.




- **Dirty beam**: point spread function (PSF) of the array, i.e. the IFFT of the sampling mask. Its many sidelobes are what convolve the true sky to produce the dirty image.



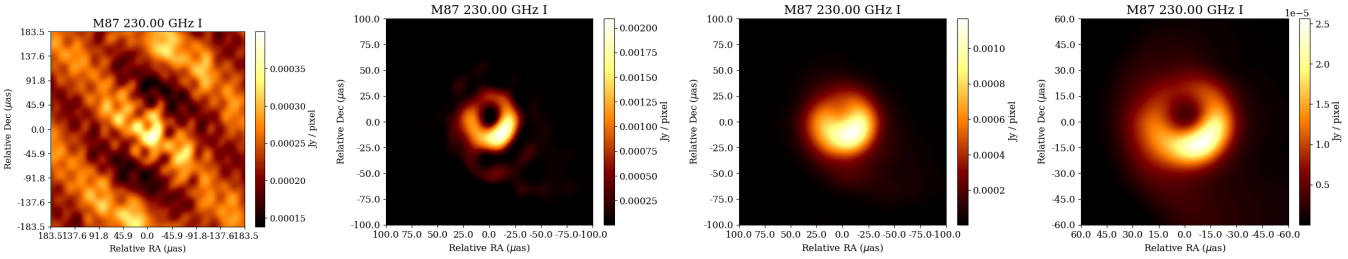
- **Clean beam:** a simplified, idealized PSF—typically a Gaussian fitted to the main lobe of the dirty beam—used to restore the cleaned components. By convolving the CLEAN model with this smooth beam, one creates a final image that suppresses sidelobe artifacts while preserving resolution.




# Experiments to improve reconstruction

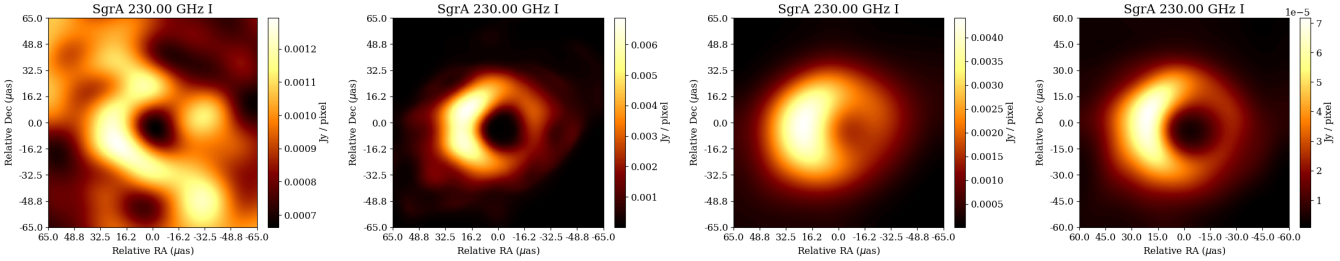
### (a) Easier celestial object

#### rowan_m87 (default)




dirty image | clean image | clean image (blurred) | GT image

#### jason_mad_eofn

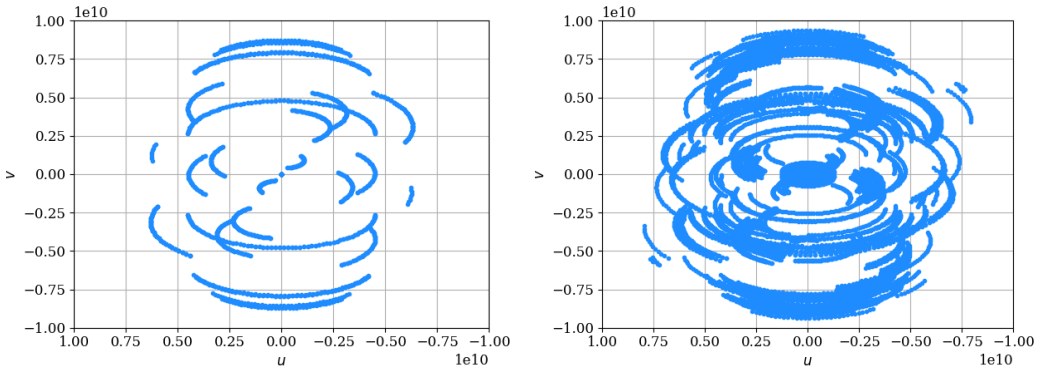



dirty image | clean image | clean image (blurred) | GT image

Using the jason_mad_eofn model instead of the default rowan_m87 yields better reconstruction results because it represents a more extended source that covers a larger portion of the sky. This leads to denser sampling of the u–v plane, especially at lower spatial frequencies, which reduces coverage gaps and sidelobe artifacts in the dirty beam.

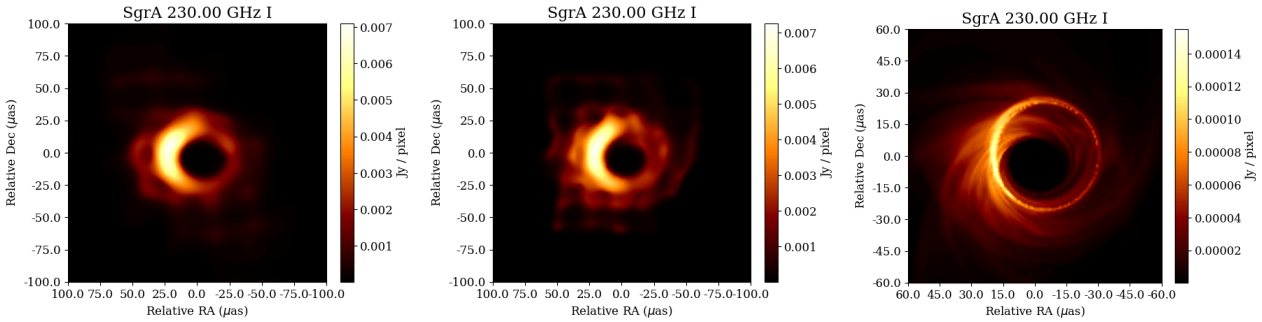
## (b) Larger telescope array

### u-v coverage



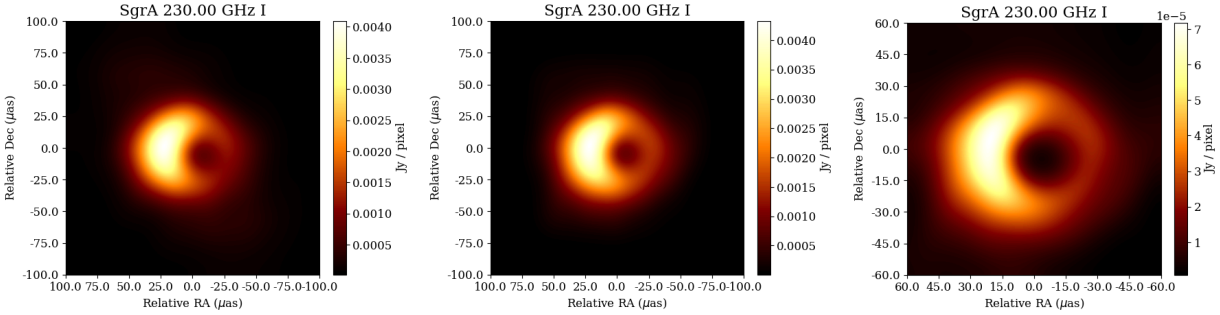

EHT2017_m87 EHTII

### clean image (not blurred)




EHT2017_m87 EHTII GT

### clean image (blurred)

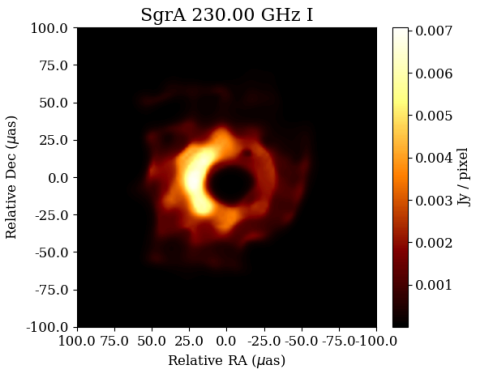



EHT2017_m87 EHTII GT

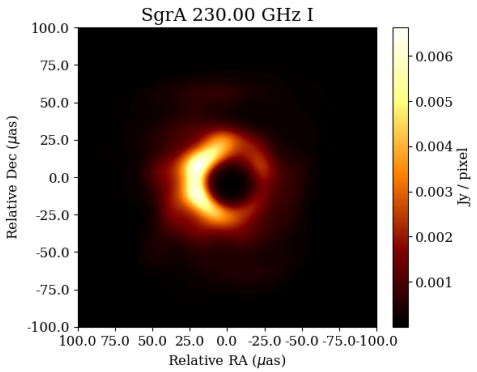
A larger telescope array improves the final reconstructed image primarily by filling more of the u–v plane—the Fourier domain in which VLBI samples the sky's spatial frequencies. Each telescope pair measures one point in u–v space at a time; adding more dishes (especially at diverse locations) creates many additional baselines of different lengths and orientations. This denser, more isotropic sampling reduces gaps in the u–v coverage, yielding a point spread function (dirty beam) with lower sidelobe levels and a sharper central lobe. Consequently, the inverse Fourier transform (dirty image) contains fewer artifacts, and deconvolution algorithms like CLEAN (see in last chapter) can more accurately recover the true sky brightness distribution, resulting in higher-fidelity, higher-resolution images.

**(c) Computational priors and regularizers**

The **'simple'** term applies a relative entropy regularization, encouraging the reconstructed image to resemble a prior map and to remain smooth and diffuse in the absence of strong data constraints.

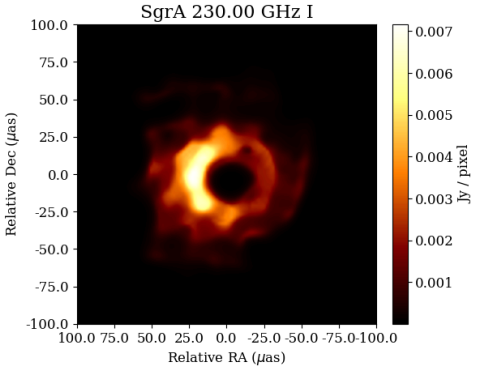



'simple': 0

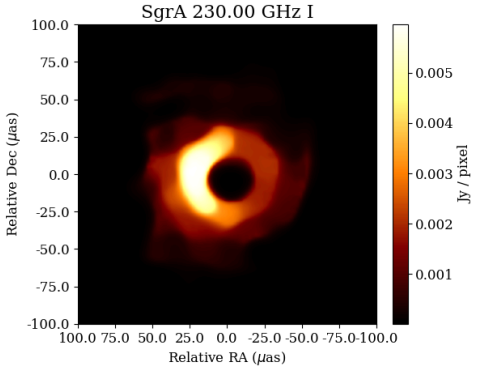



'simple': 1000

The term **'tv'** (total variation) penalizes abrupt pixel-to-pixel changes, preserving real edges while suppressing high-frequency noise.

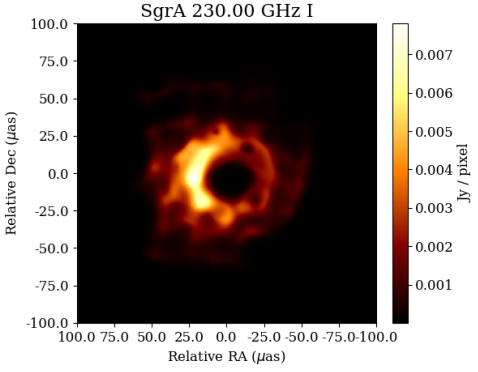



'tv': 1

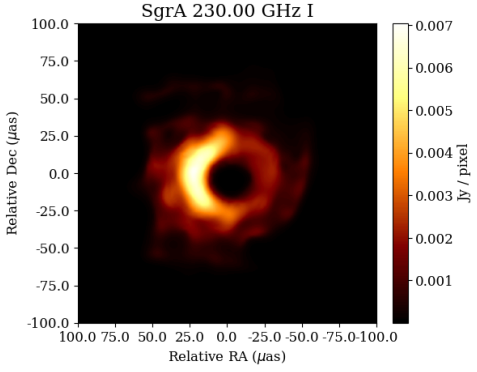



'tv': 10

Whereas **'tv2'** (squared total variation) further enforces smoothness by blurring transitions, even at edges.

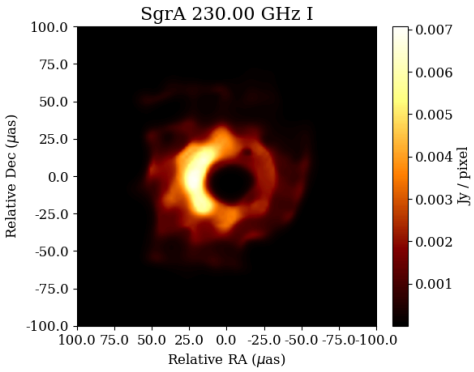



'tv2': 1

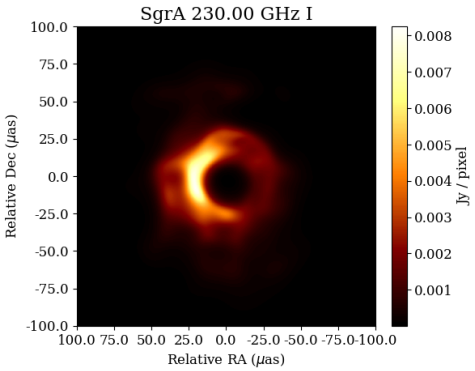



'tv2': 10

The **'flux'** term ensures that the total brightness (integrated flux) of the image matches a known target value, correcting for global scaling errors.

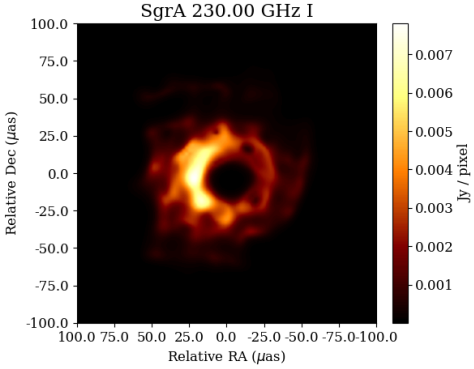



'flux': 1

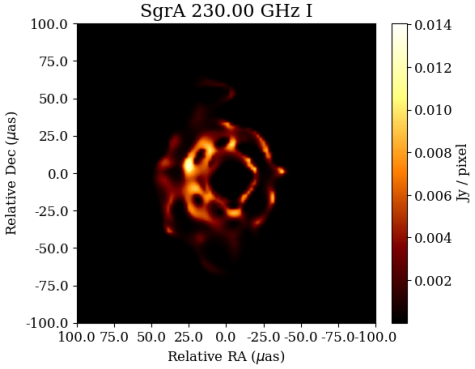



'flux': 1e8

Finally, the **'l1'** term promotes sparsity, favoring reconstructions with a few bright regions and many dark pixels, which helps eliminate weak noise but may also suppress faint extended emission.




'l1': 0.1

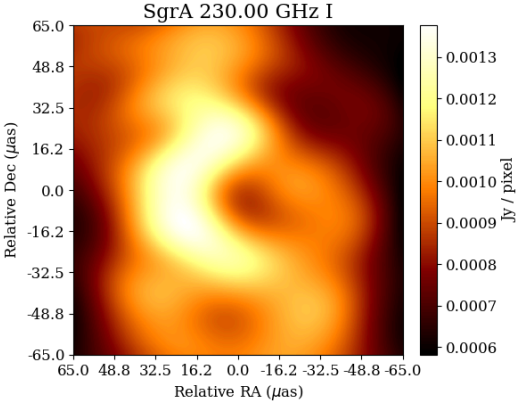



'l1': 100000

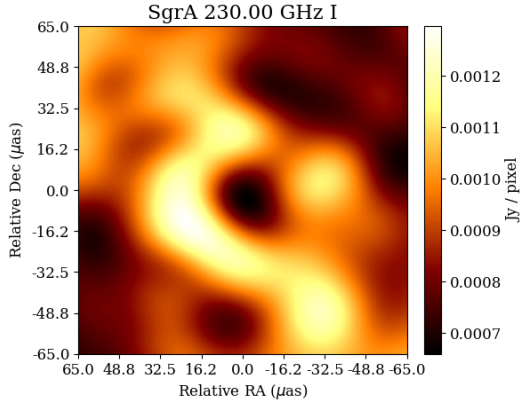
# Forward model: Telescope observations

## Add different amounts of noise

Thermal noise was introduced simply by enabling the add_th_noise=True option in the im.observe() calls, allowing the simulation to include Gaussian noise at the backend automatically. Two datasets (one noise-free and one with thermal noise enabled) were produce with identical parameters to produce dirty images:




dirty image without thermal noise



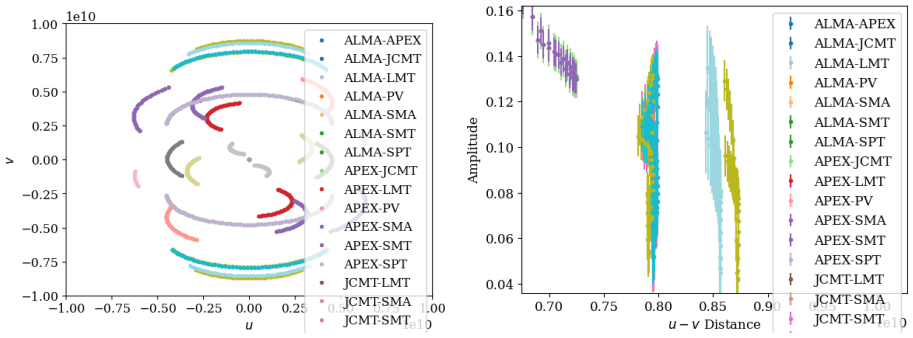

dirty image with thermal noise

The comparison shows that, once noise is included, the dirty image's background becomes noticeably grainy, low-contrast features are washed out, and sidelobe artifacts grow more apparent against the raised noise floor.

## Compute frequencies by each pair of telescopes at a time

The observation dataset is decomposed into separate visibility sets for each antenna pair (baseline) to isolate their individual contributions in the u–v plane. For every unique pair of stations, a two-element subarray is constructed and used to simulate visibilities over the same time and bandwidth parameters as the full array. Each subarray's visibilities yield the (u, v) coordinates sampled by that specific baseline at each integration step. These per-baseline samples are then collected into a list alongside identifiers (e.g., "ALMA–SMA") for subsequent visualization.

By plotting each baseline's u–v track in a distinct color—along with its corresponding u–v distance versus time curve—one can clearly see how different antenna separations and orientations contribute to overall Fourier coverage. This approach highlights which spatial frequencies are uniquely probed by each pair, identifies redundant or sparse regions in the u–v plane, and provides direct insight into how individual baselines shape the array's point spread function. Below are example plots showing colored u–v coverage and u–v distance traces for a subset of baselines.

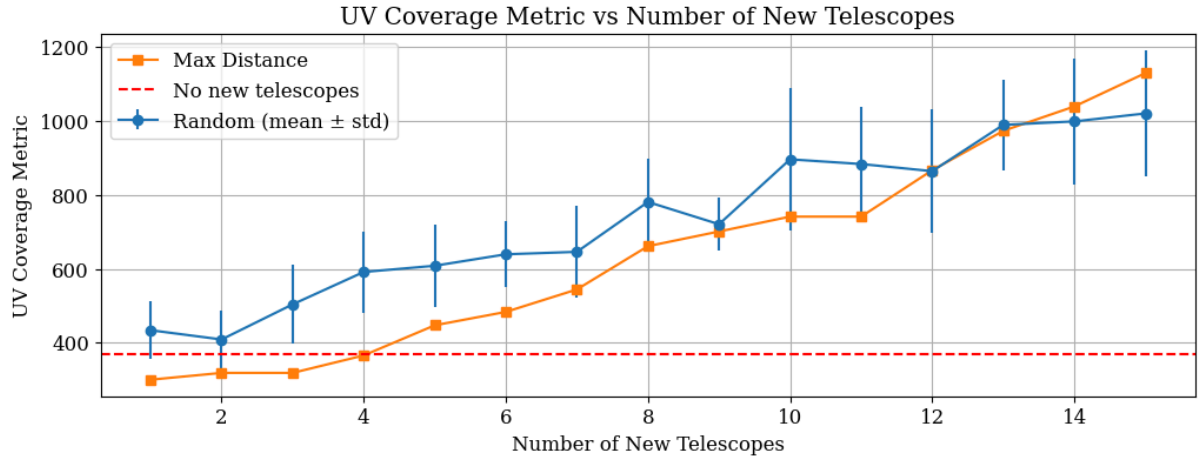



## Add telescopes to maximize frequency coverage

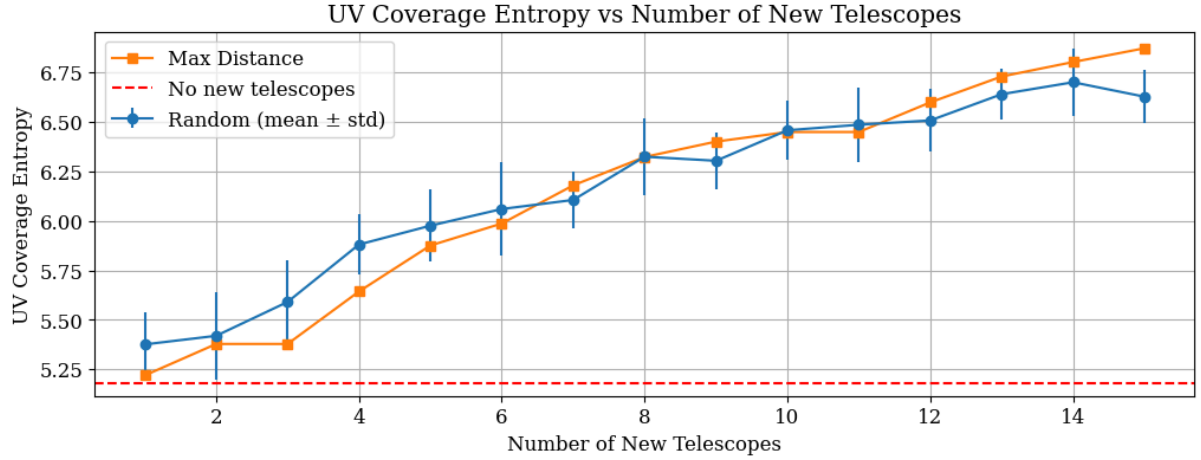
A distance-maximization algorithm was implemented to place new telescopes at the most geographically isolated sites, and its performance was compared against a purely random placement strategy. In the maximization approach, candidate locations on a coarse 3D grid constrained to Earth's radius are evaluated by computing each point's minimum distance to all existing and previously added telescopes; the candidate with the largest minimum separation is selected, ensuring that each new station systematically fills the largest gap in baseline orientations and lengths. The random strategy instead selects valid Earth-surface coordinates uniformly in latitude and longitude, and is repeated ten times to produce an average coverage metric that mitigates run-to-run variability.

After merging the newly added sites with the original EHT array, visibilities are simulated and u–v coverage is computed for each method. With the following six coverage metrics confirm that the distance-maximization algorithm performs equally or better than random telescope placement across multiple dimensions of u–v sampling.

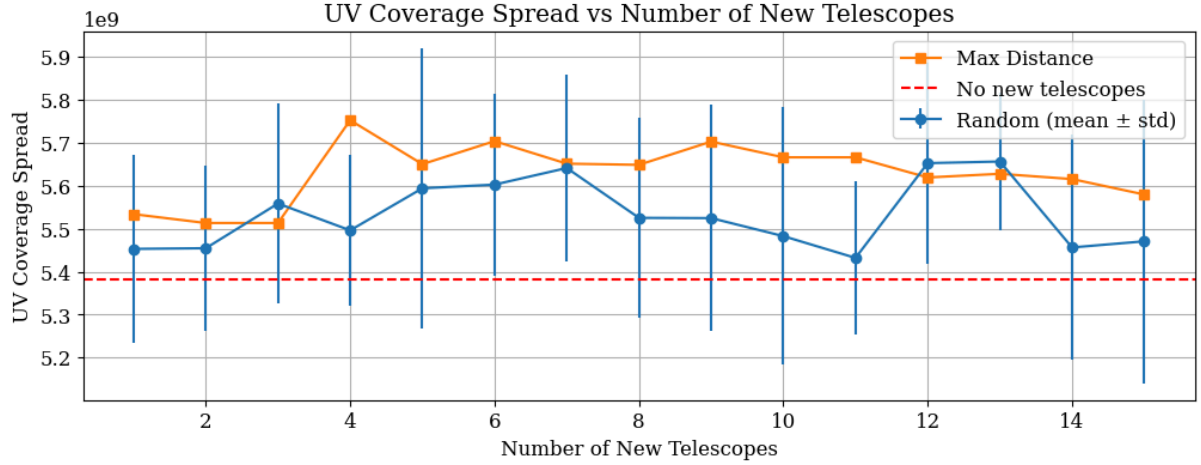
The combined density-uniformity metric rises significantly when using the targeted method, reflecting both more points and a smoother distribution.

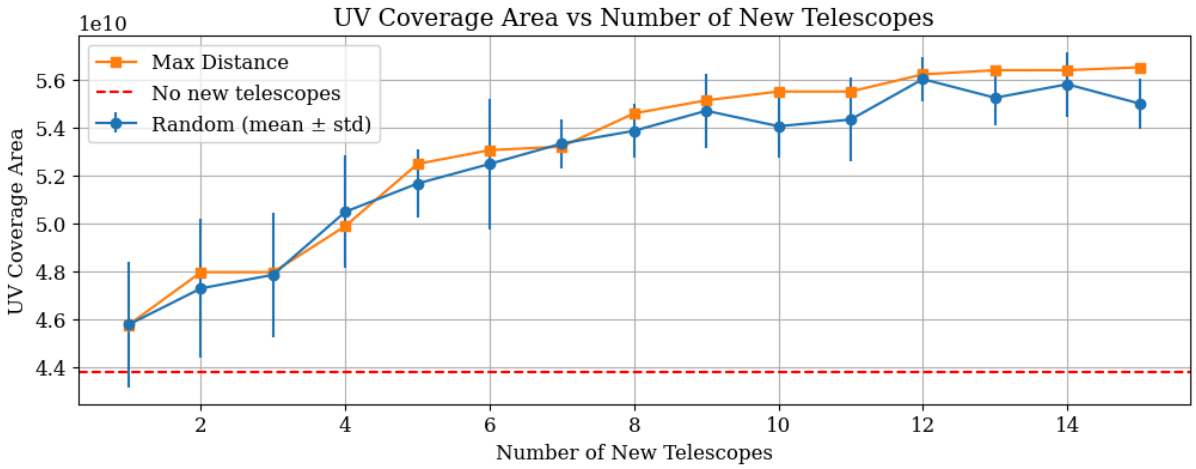



Sample-entropy likewise increases, indicating a more uniform fill of the plane rather than clumped tracks.

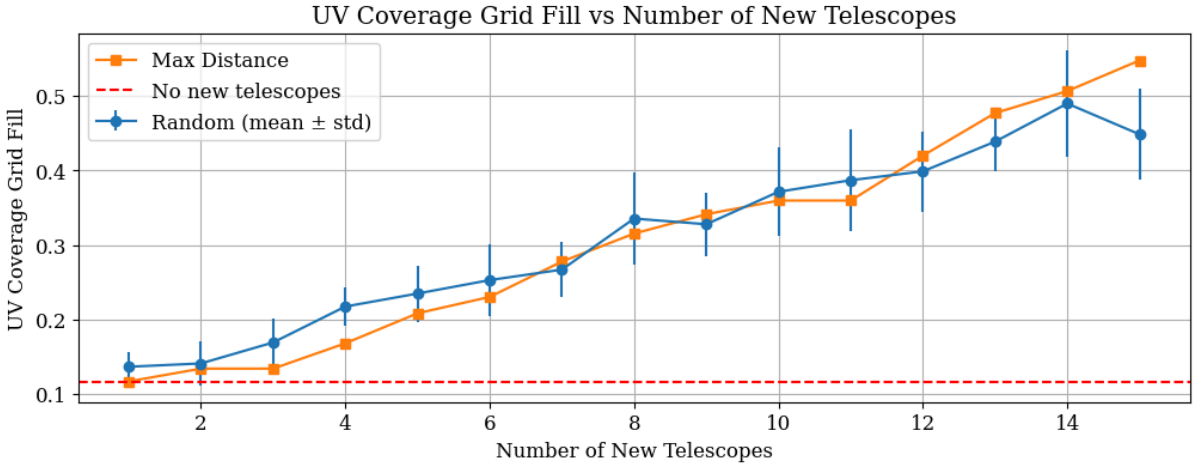


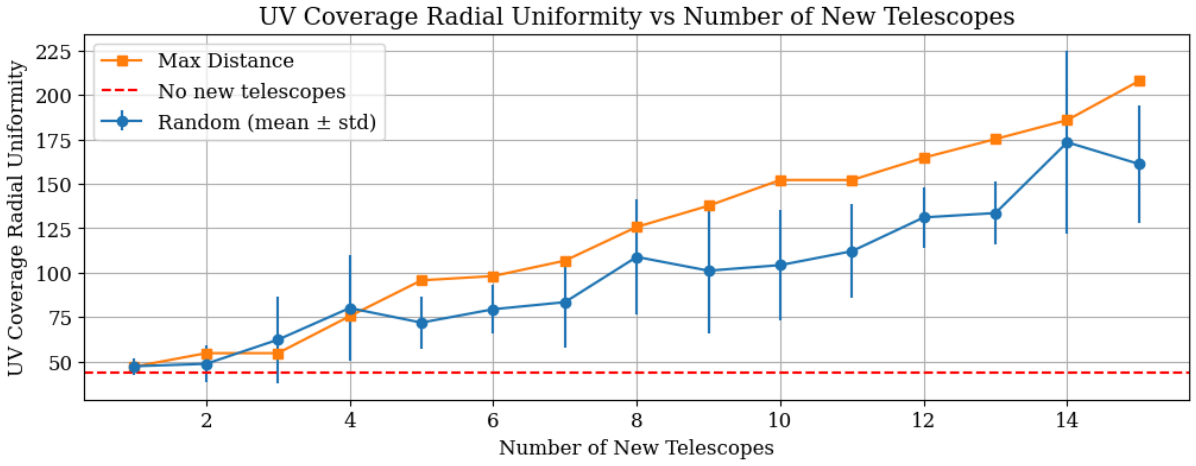

The spread and convex-hull area metrics grow larger under maximized placement, showing that baselines extend over a wider range of spatial frequencies and cover a broader footprint.





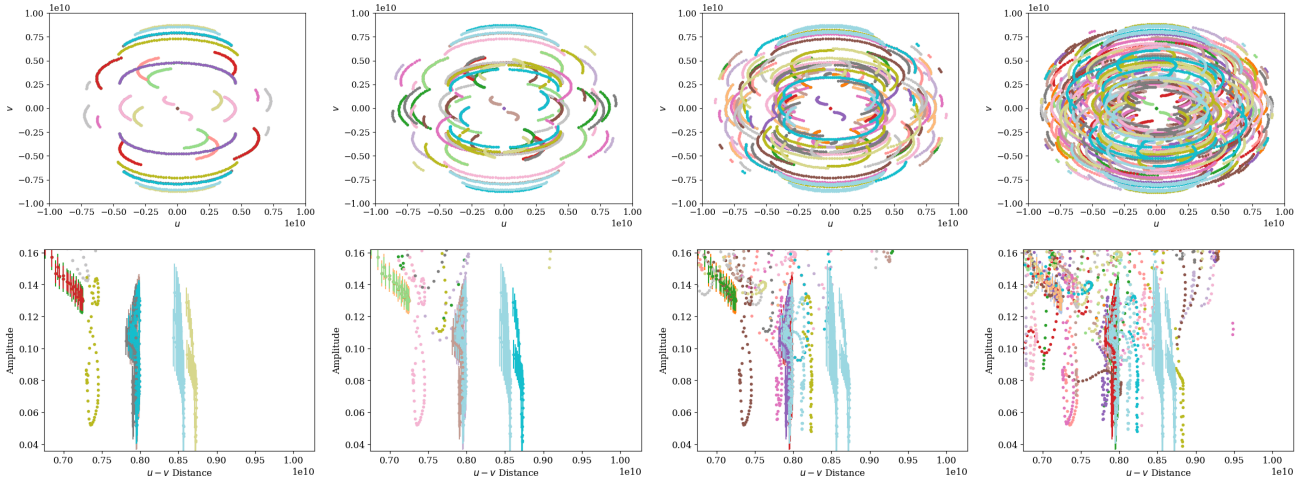

The grid-filling factor jumps higher, demonstrating that more grid cells receive at least one sample, and the radial-uniformity standard deviation drops, signifying a more balanced sampling across baseline lengths.




In contrast, random additions improve these metrics only modestly on average and with far greater run-to-run scatter. Taken together, the six quantitative measures reveal that the maximization algorithm yields consistently denser, more uniform, and more complete u–v coverage with similar added stations compared to naïve random placement.

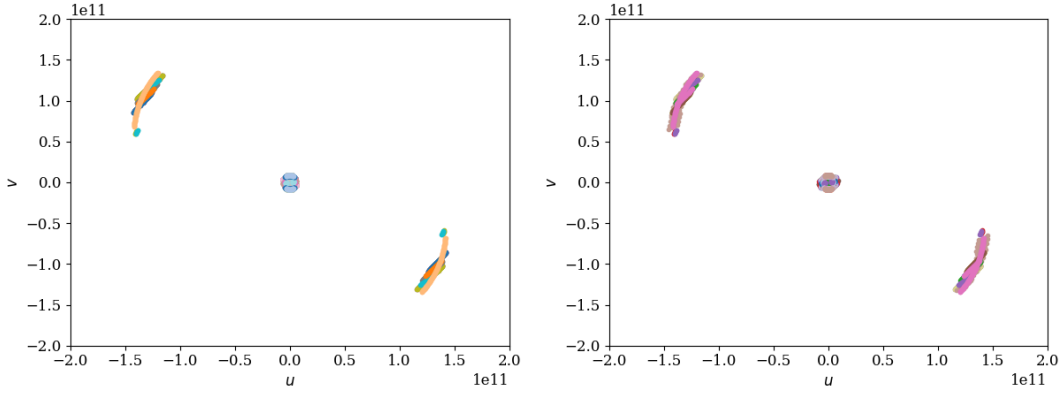
Examples plots of adding new telescopes with the maximum algorithm:



1 new telescope | 5 news telescopes | 10 news telescopes | 20 news telescopes

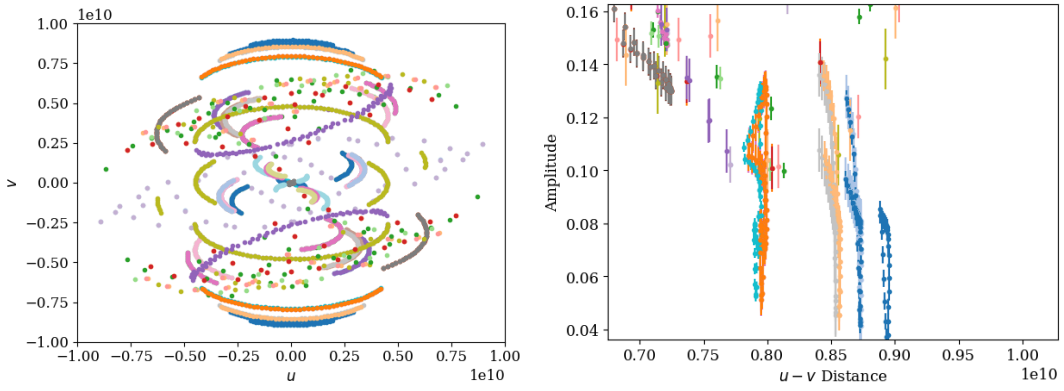
# Evaluate the effect of satellite telescopes

The impact of adding space-based elements (TESS and ISS) was assessed by loading the arrays EHT2025wTESS and EHT2025wTESS&ISS and comparing their u–v coverage to the ground-only configuration.



EHT2025wTESS | EHT2025wTESS&ISS

The full u–v tracks for TESS alone versus TESS + ISS reveals almost no visually discernible improvement when both satellites are included, apart from a few isolated tracks that lie at much larger u–v distances than any ground baselines.

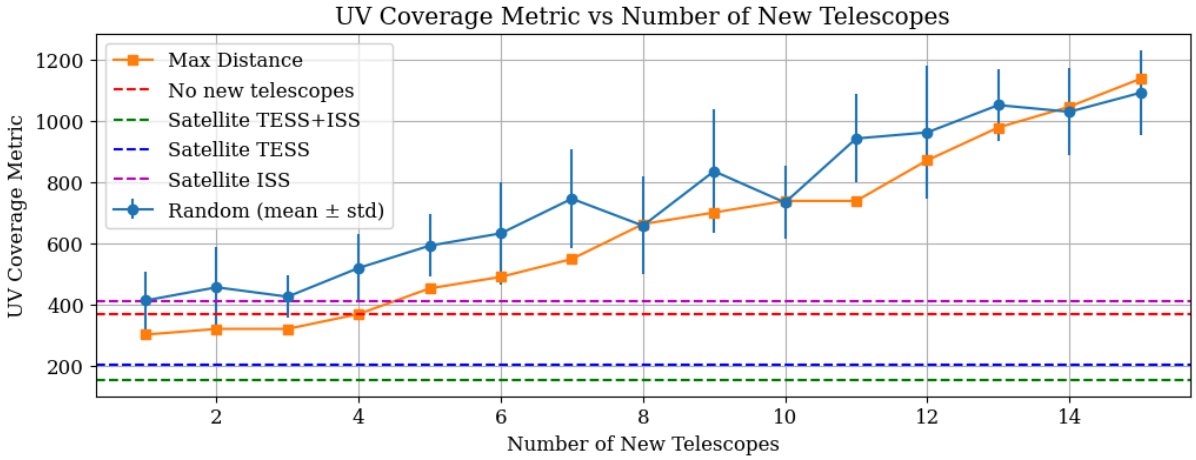


EHT2025wISS from close view

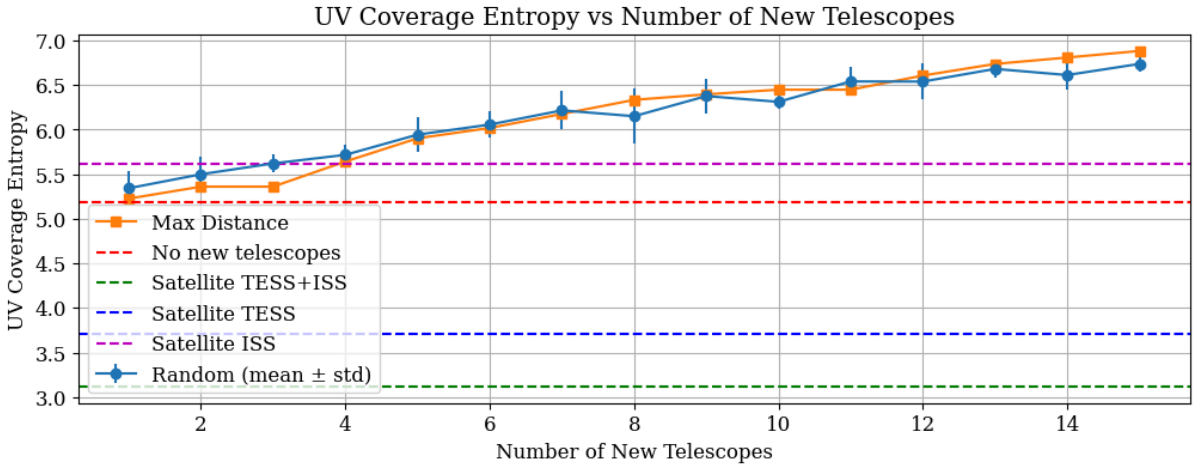
A closer view of the ISS-only coverage against the ground array further highlights a cloud of scattered points filling intermediate gaps between the main tracks, which is also reflected in the u–v distance.

When these configurations are quantified using the coverage metrics, several trends emerge:

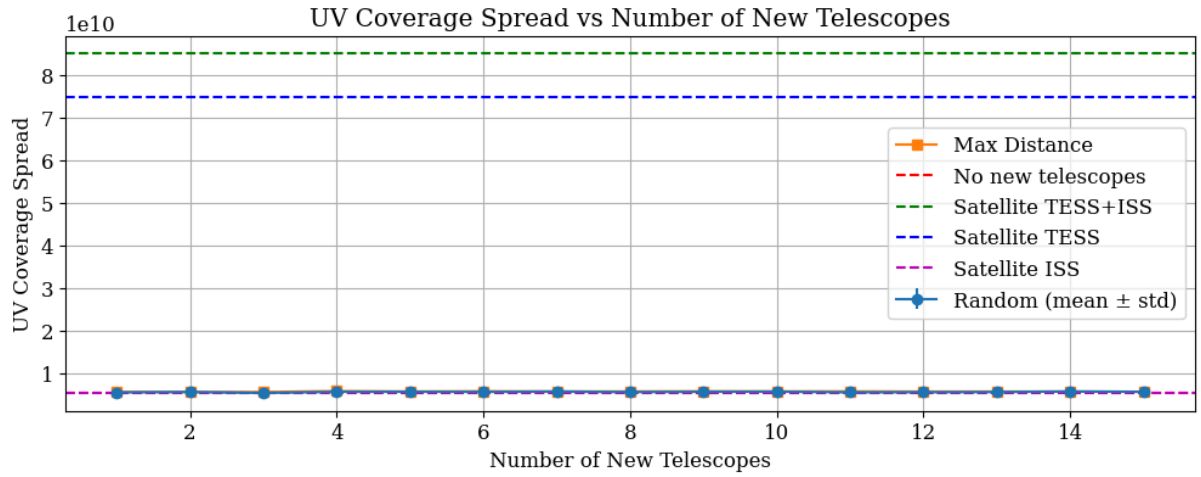
For the combined density–uniformity metric, neither satellite scenario outperforms the single-telescope maximization algorithm: adding one randomized ground station still yields denser and more uniform sampling than either space array.

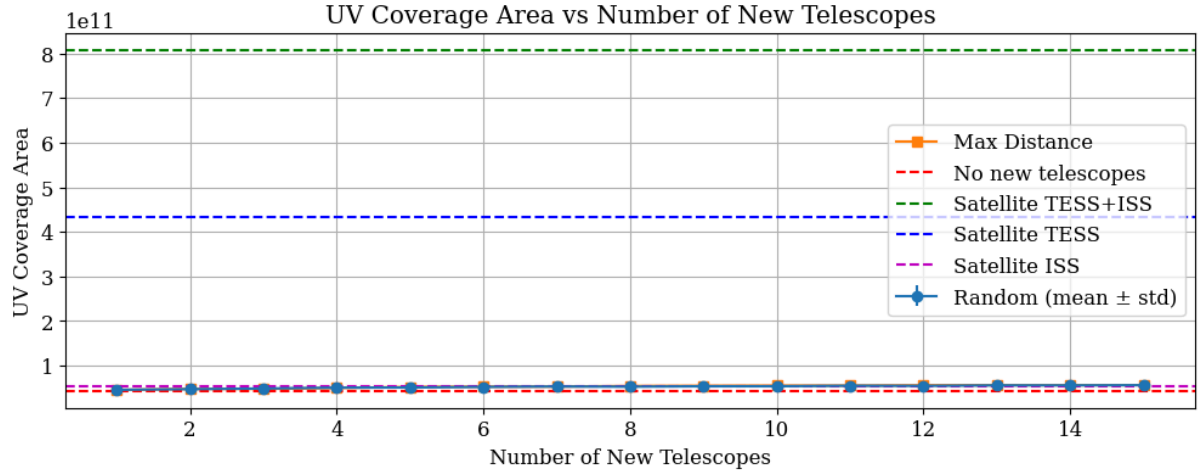


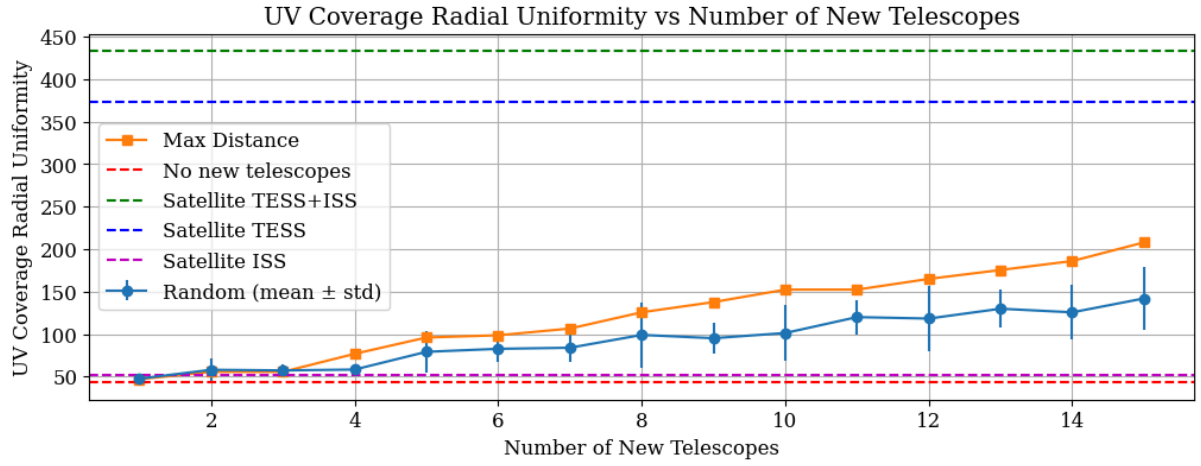

In entropy, only the ISS array matches the effect of adding four ground telescopes via maximization, while TESS alone or TESS + ISS remain below that benchmark.



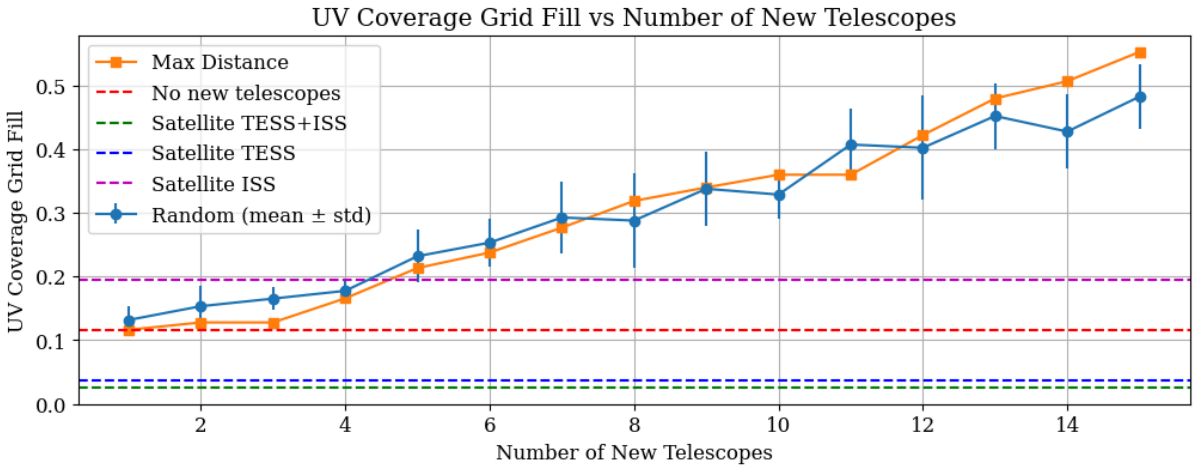

Spread, convex-hull area, and radial-uniformity all show TESS and TESS + ISS delivering dramatically larger values than any other configuration—indicating that the satellites introduce very long baselines that extend the furthest spatial frequencies—whereas ISS alone contributes only modest gains.




Finally, the grid-filling factor behaves like entropy: the satellites barely improve uniform cell coverage compared to a single well-placed ground telescope.




In summary, space-based elements can greatly expand maximum baseline length (boosting spread, area, and radial uniformity), but they do not substitute for dense, uniform sampling in the core u–v plane (as measured by density–uniformity, entropy, and grid fill).

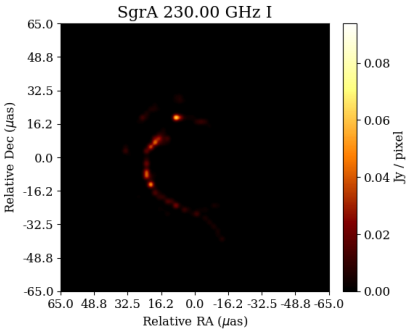
# VLBI Imaging: Estimating an image of a celestial object

The final imaging stage starts with the dirty image—an inverse Fourier transform of incomplete visibilities—and its dirty beam, which together produce blurred features and sidelobe artifacts.

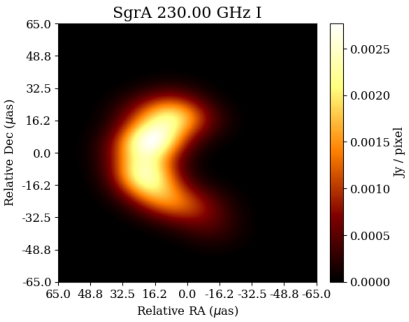
A **Högbom CLEAN** algorithm was implemented to deconvolve the dirty beam from the dirty image. Conceptually, CLEAN proceeds by (1) locating the brightest pixel in the current residual image, (2) subtracting a scaled copy of the dirty beam centered at that location—where the scale factor (“gain”) is a fraction of the peak amplitude—and (3) accumulating those subtracted peaks into a “model” image. This process repeats until the peak residual falls below a predefined threshold or a maximum iteration count is reached. The accumulated model thus represents a sparse reconstruction of the true sky. Finally, the model is convolved with a **clean beam**—an idealized Gaussian fit to the main lobe of the dirty beam—and the remaining residuals are added back, yielding the **final CLEANed image**.

Applying this CLEAN implementation to the simulated visibilities produces a reconstruction in which the key morphological features of the target (e.g. ring structures, extended emission) become sharply defined and sidelobe artifacts are greatly suppressed. When compared side-by-side with the original model image, the CLEANed result achieves closer agreement in both overall flux distribution and spatial resolution. Below are example figures showing the model image, the final CLEANed image obtained using the described algorithm, and the dirty image.

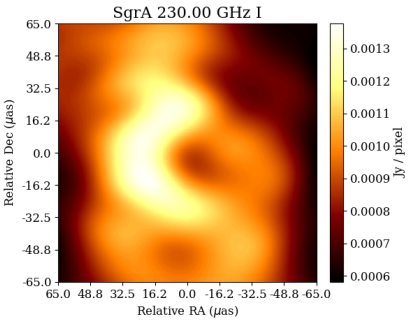
gain = 0.001 threshold = 4e-4




model image

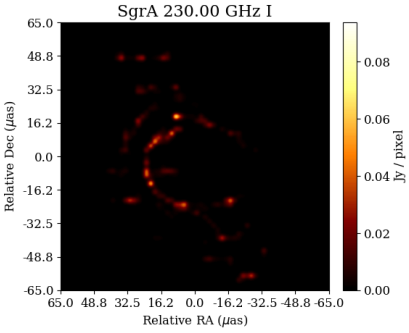



CLEANed image

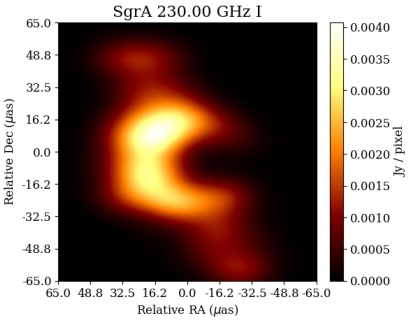



dirty image

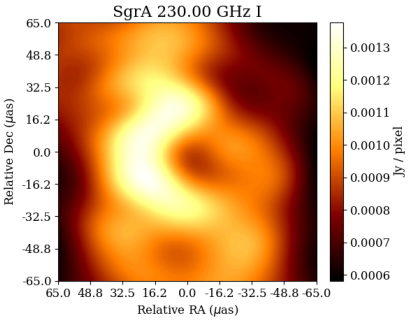
gain = 0.1 threshold = 4e-4




model image




CLEANed image




dirty image