Advanced radio interferometric imaging

European Radio Interferometry School (ERIS) 2017

2017-10-17

André Offringa
Netherlands Institute for Radio Astronomy (ASTRON)

RadioNet

ASTRON

erc European Research Council
Established by the European Commission
Supporting top researchers from anywhere in the world

- Output of an interferometer after calibration:

$$V(u,v,w) = \iint \frac{I(l,m)}{\sqrt{1-l^2-m^2}} e^{-2\pi i\left(ul+vm+w(\sqrt{1-l^2-m^2}-1)\right)} dl dm$$

- (u,v,w) : interferometer's geometrical vector
- (l,m) : position on the sky
- I : sky brightness (“image”)

Imaging : Calculating I(l,m) from V(u,v,w)

Visibility function

- Full visibility function:

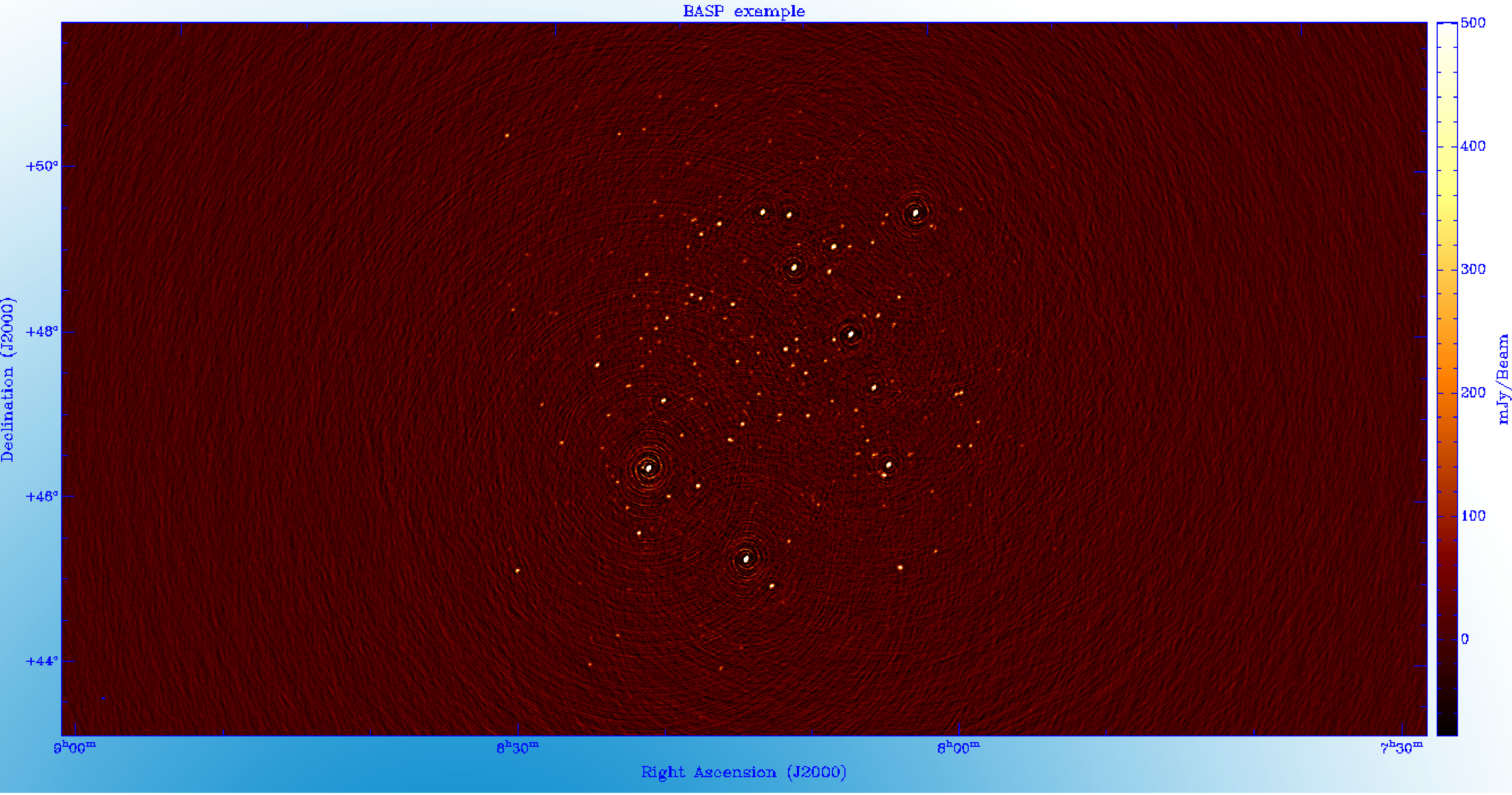
$$V(u,v,w) = \iint \frac{I(l,m)}{\sqrt{1-l^2-m^2}} e^{-2\pi i\left(ul+vm+w(\sqrt{1-l^2-m^2}-1)\right)} dl dm$$

- For small field of view (l~0, m~0) or w~0 :

$$V(u,v,\cancel{w}) \approx \iint I(l,m) e^{-2\pi i(ul+vm)} dl dm$$

- (u,v,w) : interferometer's geometrical vector
- (l,m) : position on the sky
- $\boldsymbol{I}$: sky brightness (“image”)

Fourier relation

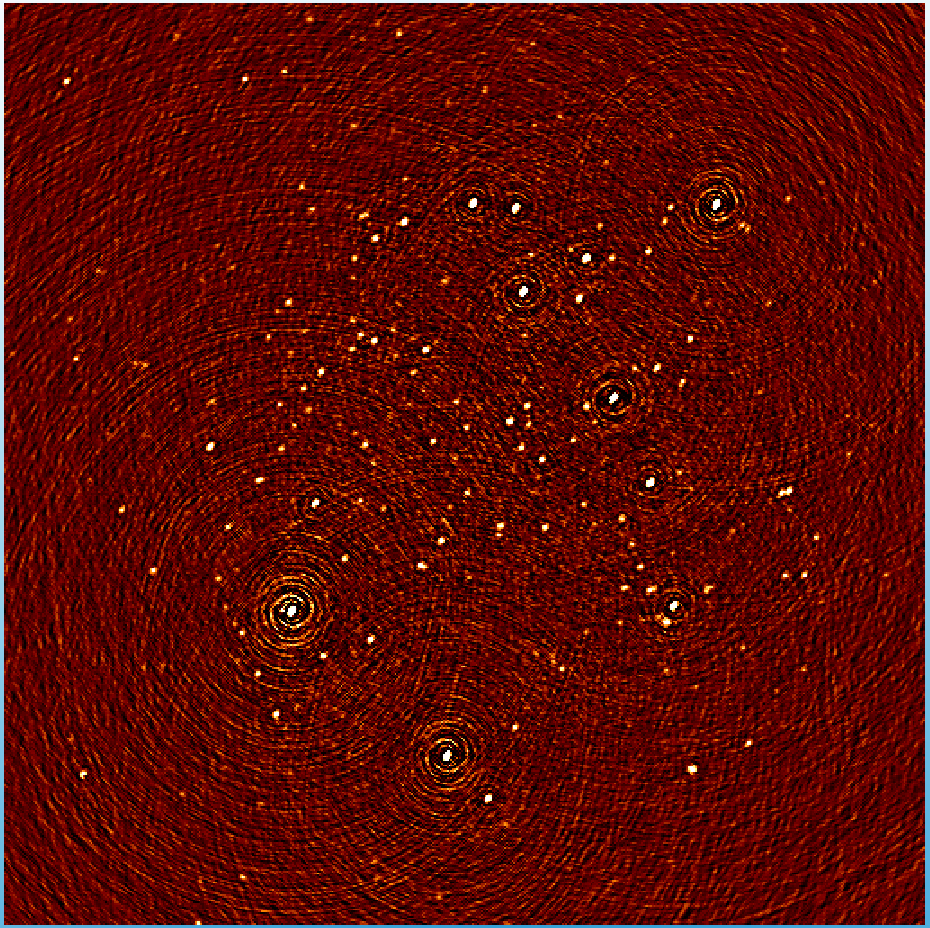


LOFAR dirty image (3c196)

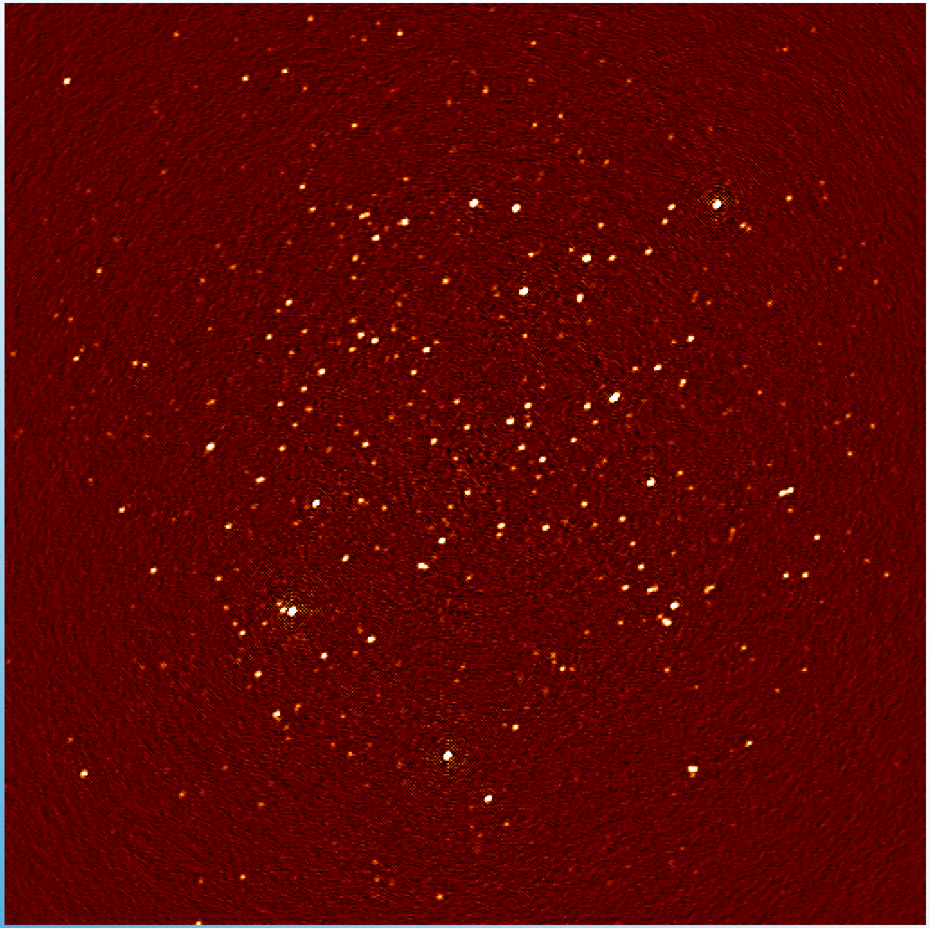
The dirty image

Högbom CLEAN

- Högbom CLEAN algorithm (1974):
  - Find largest peak in image
  - Scale PSF to fraction of peak and subtract
  - Repeat until peak < threshold or nIter > limit
  - Finally: restore subtracted components



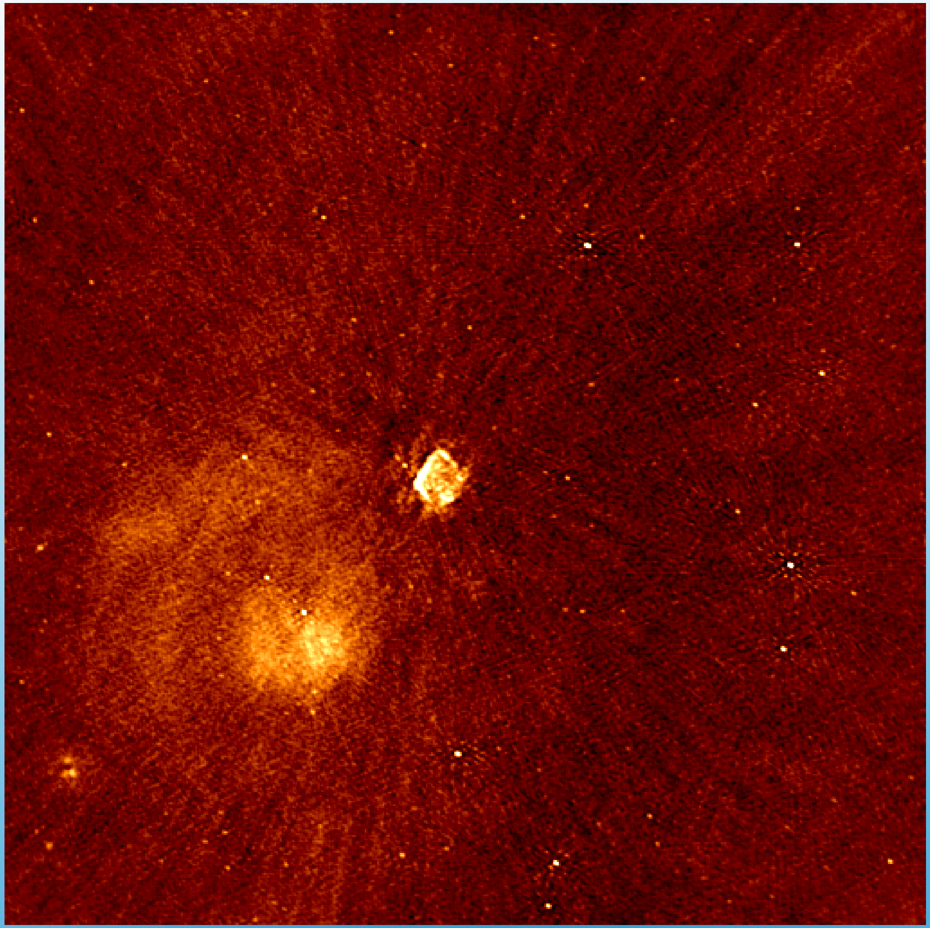
LOFAR undeconvolved (“dirty”) image



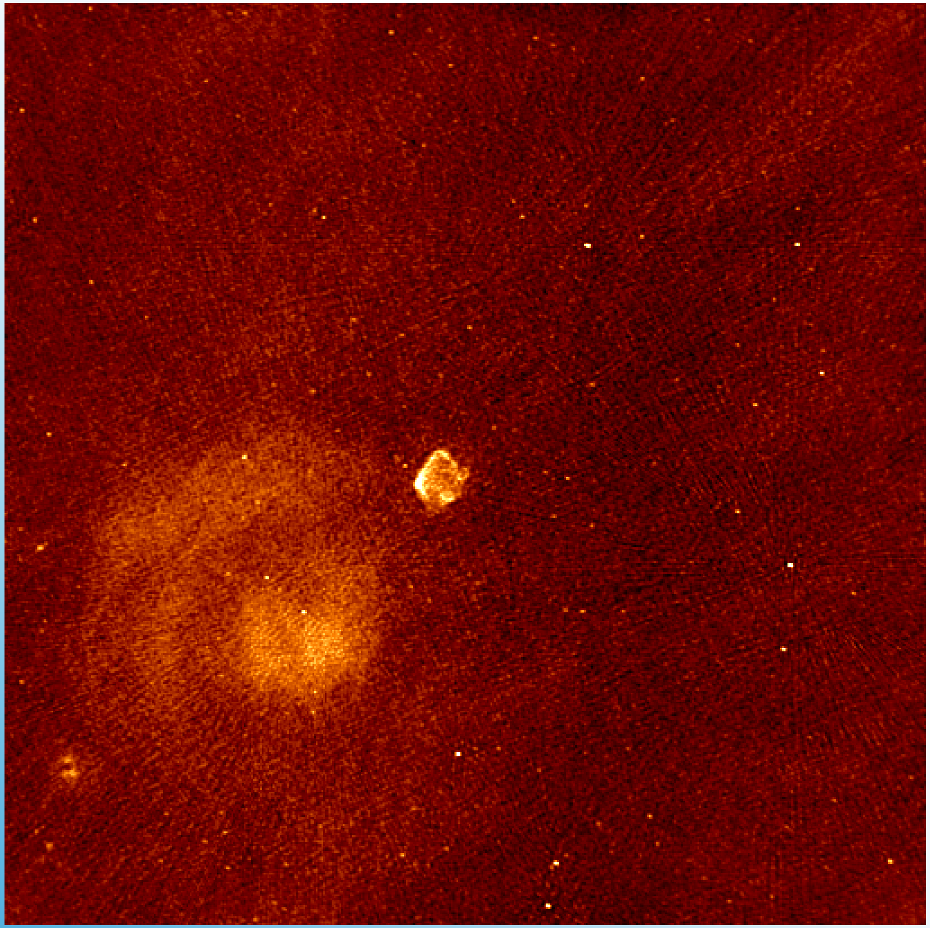
Deconvolved with Högbom CLEAN

Högbom CLEAN

Deconvolving diffuse structures

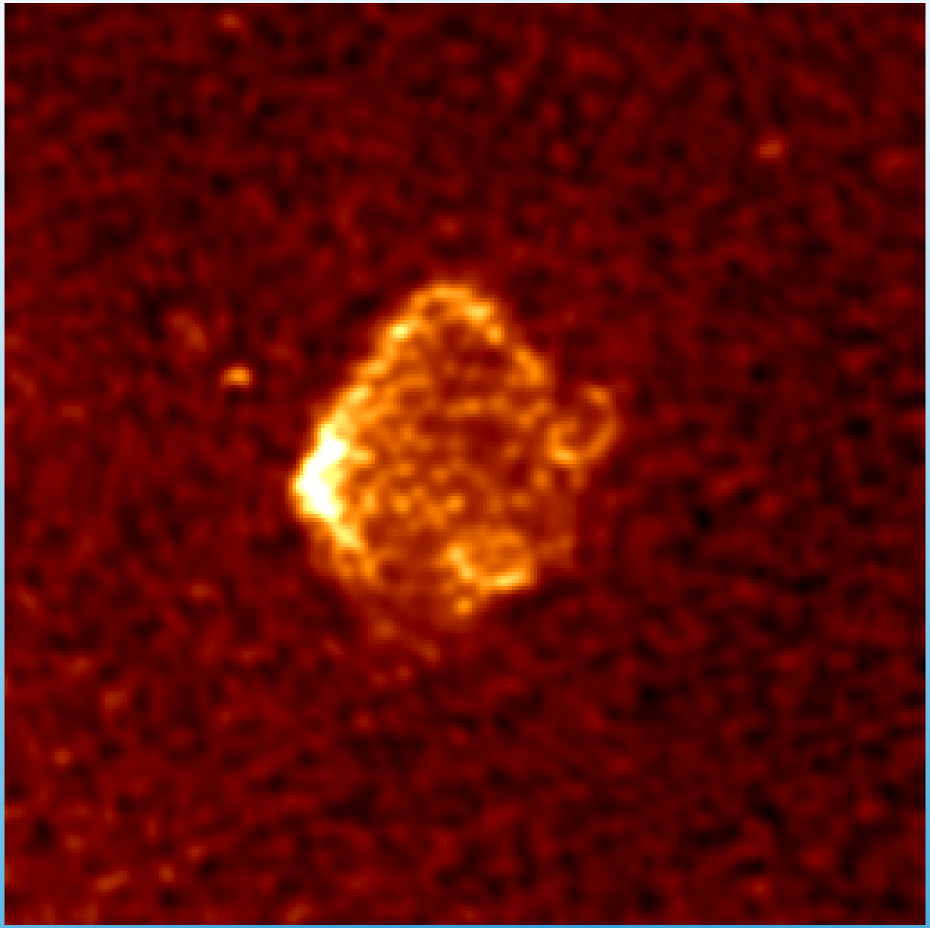


Undeconvolved “dirty” image

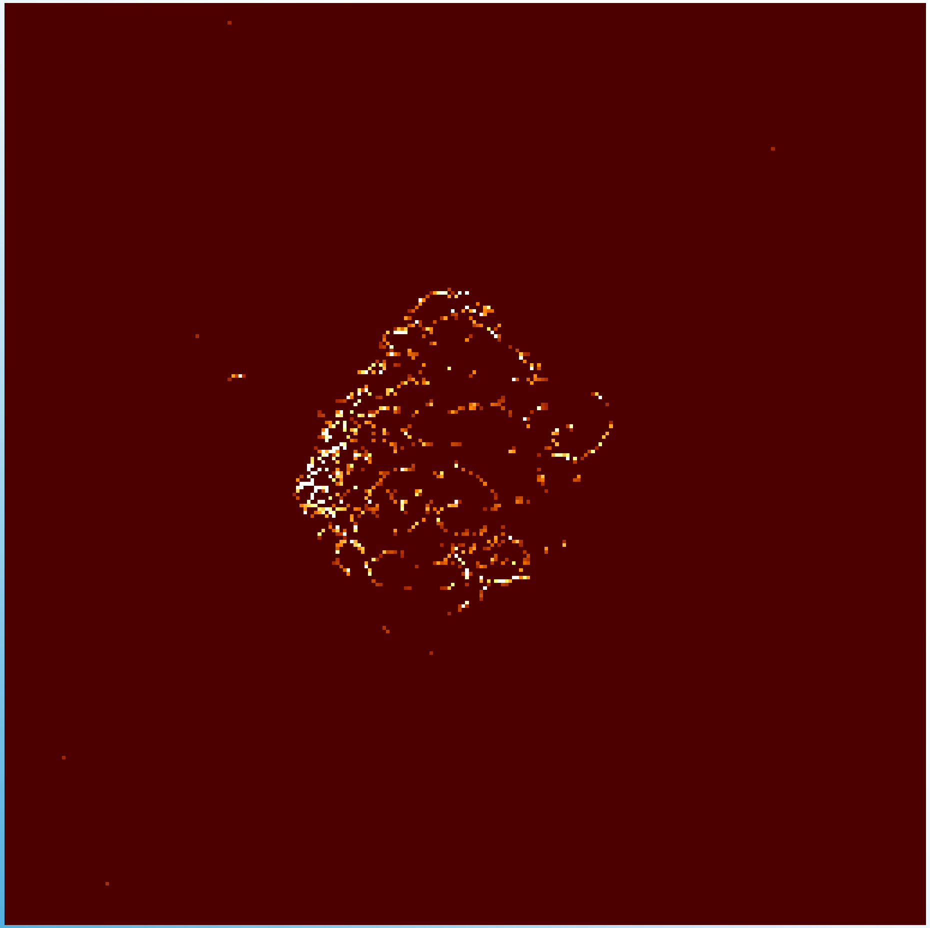


Deconvolved image with Högbom CLEAN

Deconvolving diffuse structures



Deconvolved image (Högbom CLEAN)

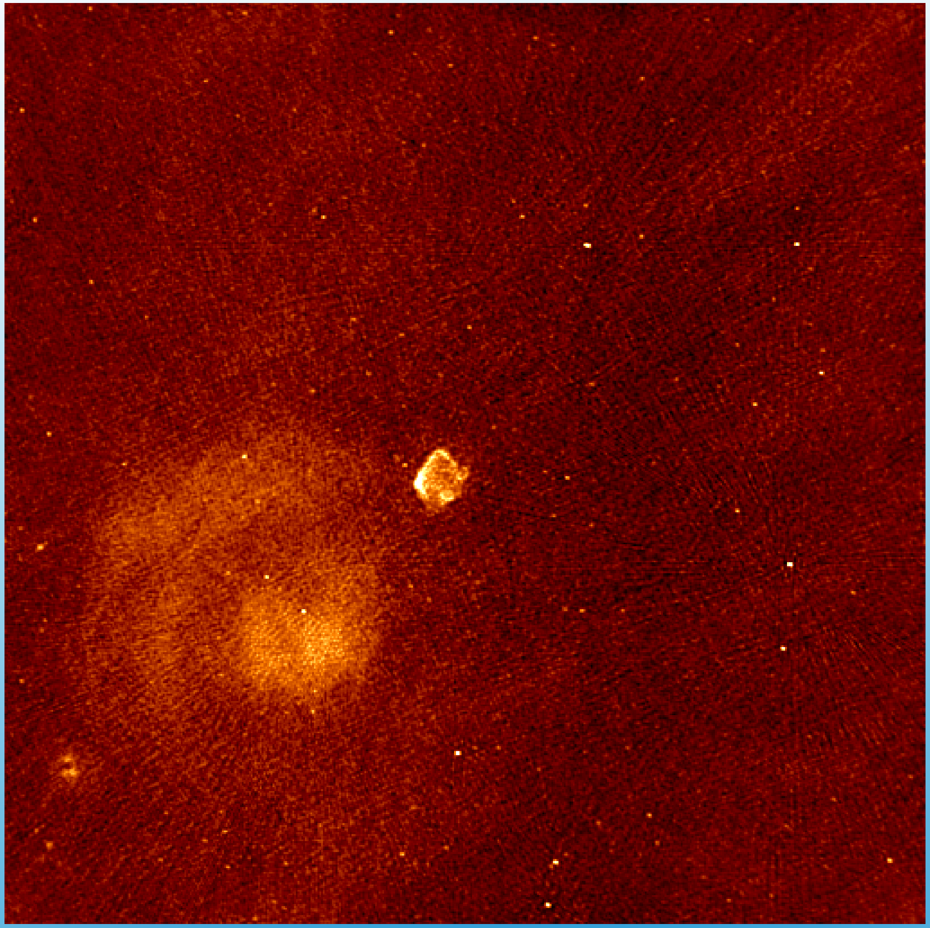


Actual model

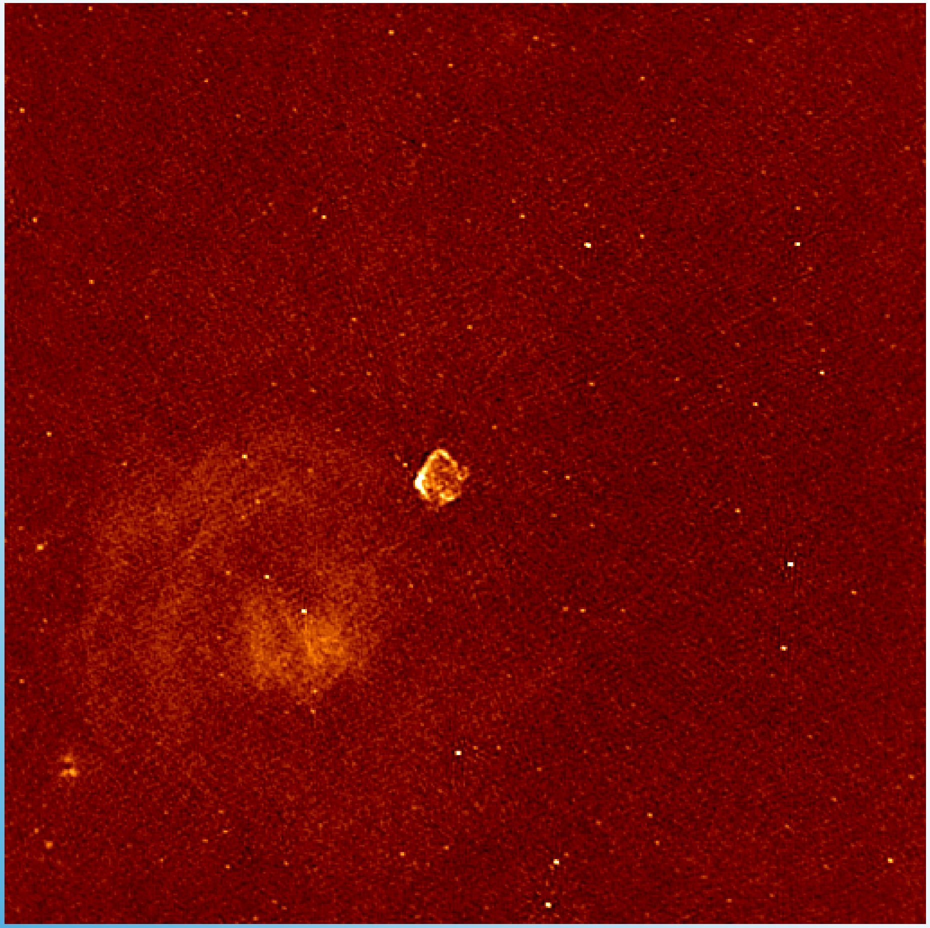
Improved algorithm by Cornwell (2008) :

- “Multi-scale clean”
- Fits small smooth kernels (and delta functions) during a Högbom CLEAN iteration

Multi-scale CLEAN

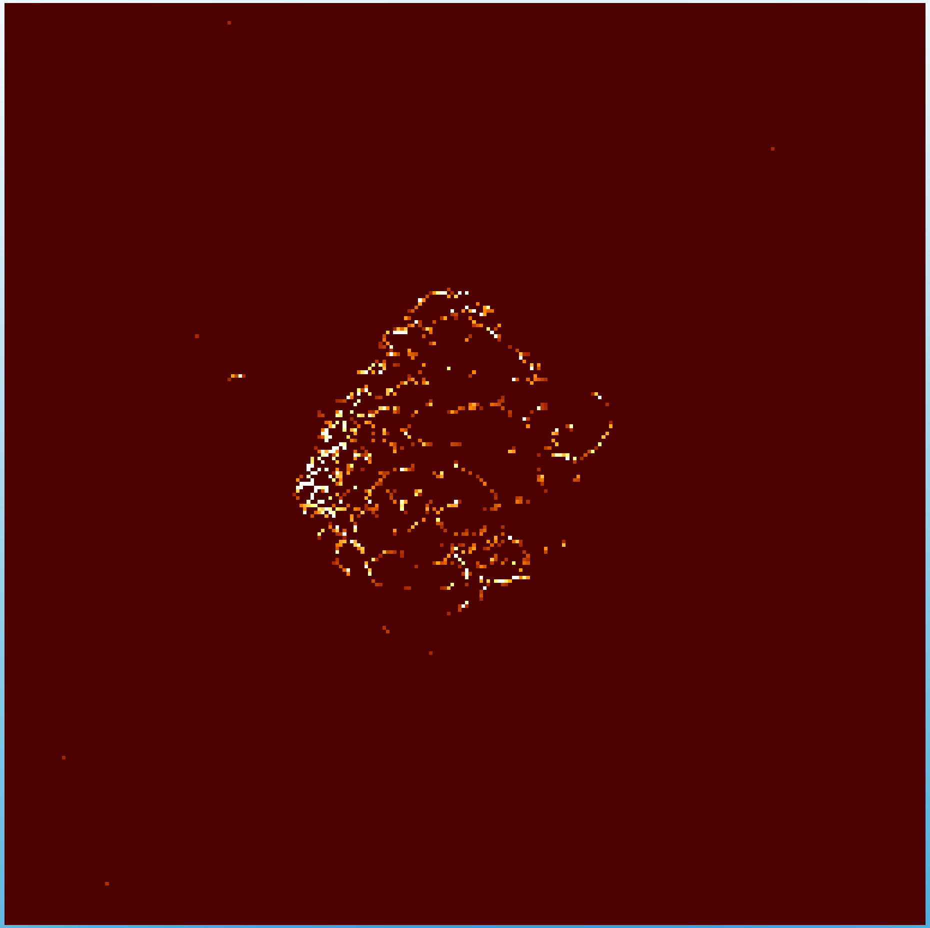


Normal Högbom CLEAN

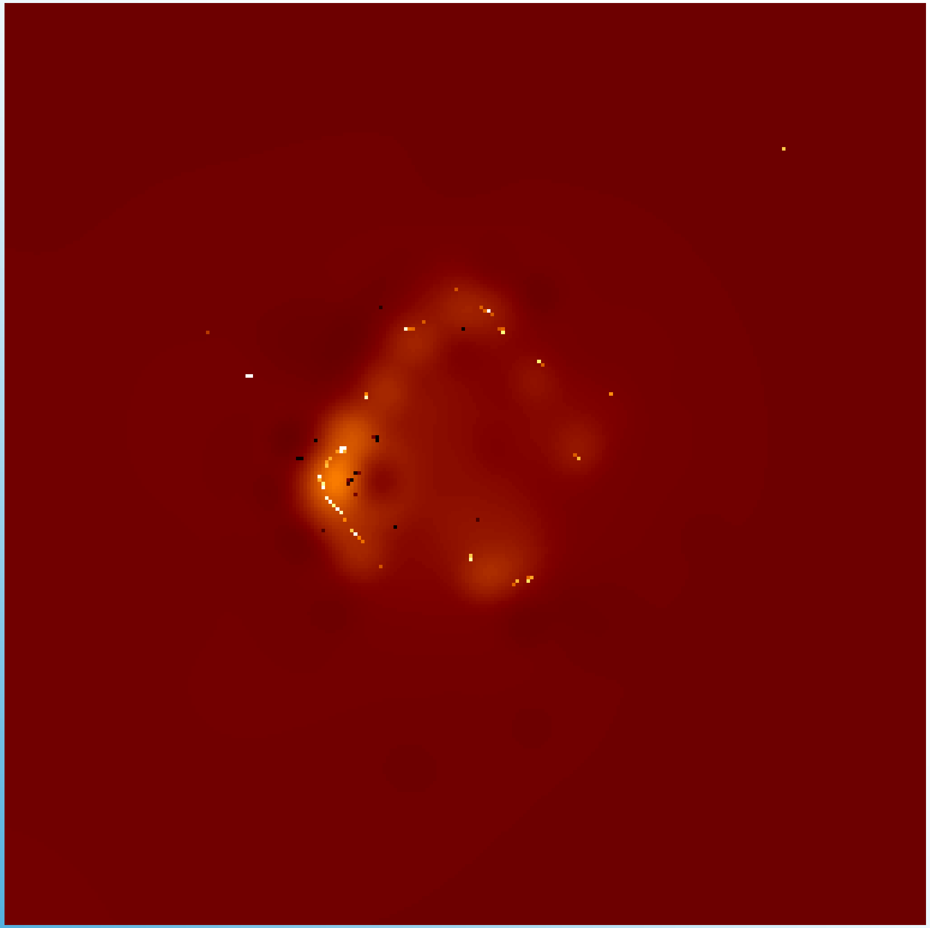


Multi-scale CLEAN
(implementation in WSClean)

Multi-scale CLEAN



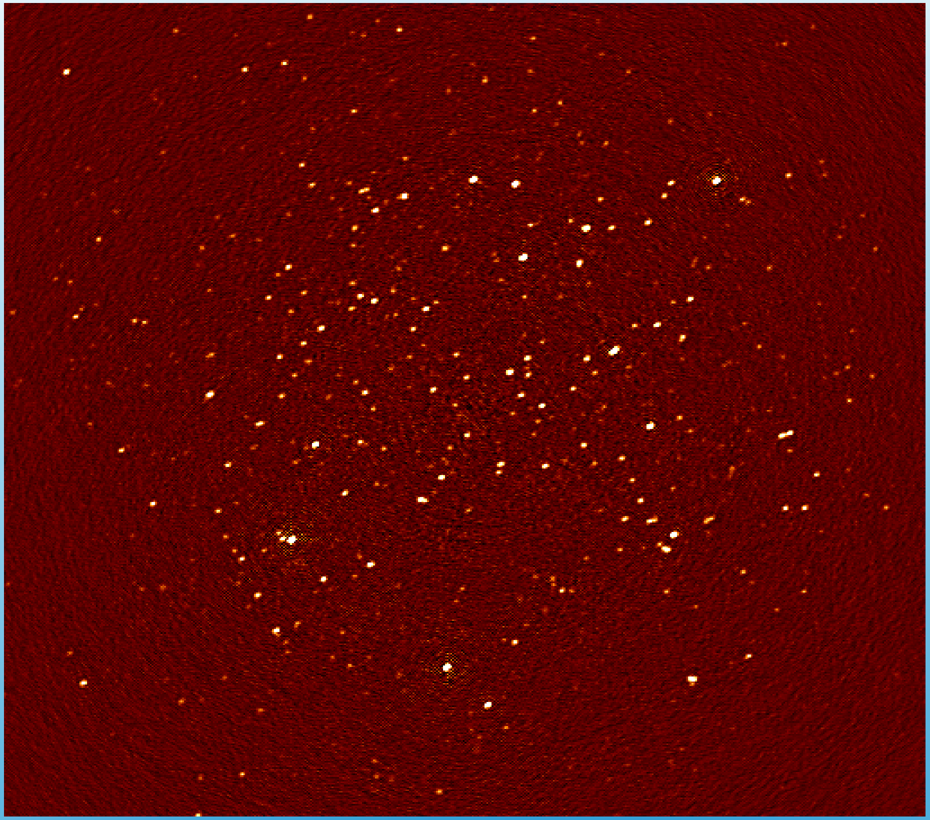
Normal Högbom CLEAN
Output model



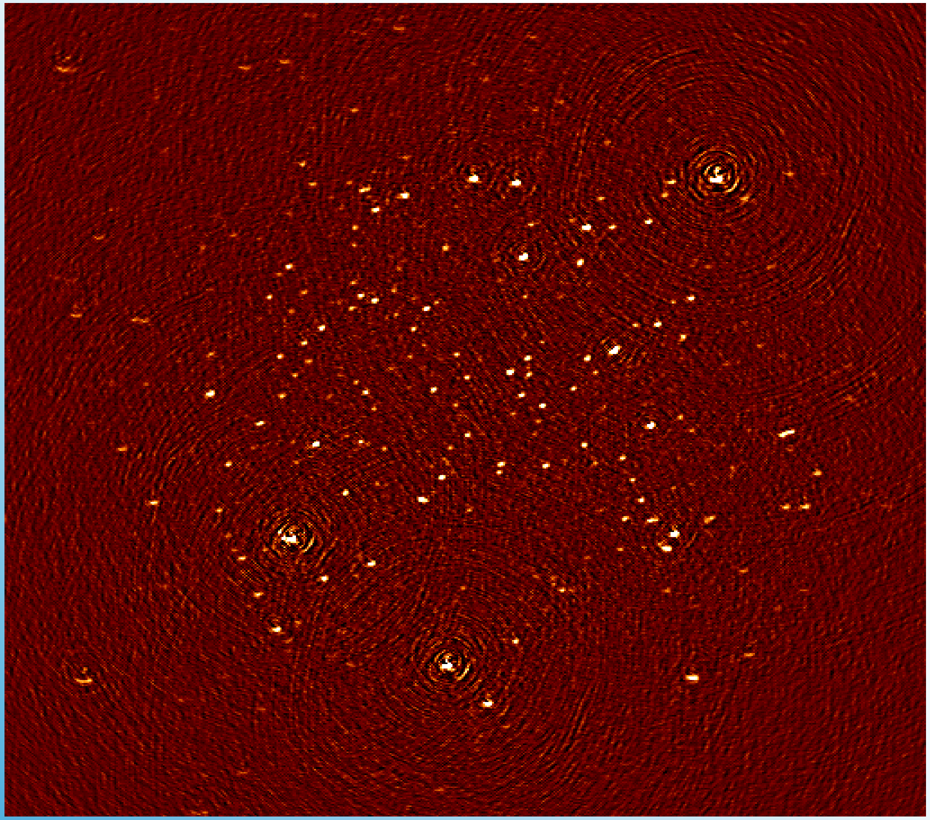
Multi-scale CLEAN
(as implementation in WSClean)

Multi-scale CLEAN

- 2D FT does not hold for new arrays: l,m,w >> 0



Correcting w-terms



Without correcting w-terms

The w-term

- 2D FT relationship does not hold for new arrays: l,m,w >> 0
- Have to use full function:

$$V(u,v,w) = \iint \frac{I(l,m)}{\sqrt{1-l^2-m^2}} e^{-2\pi i\left(ul+vm+w(\sqrt{1-l^2-m^2}-1)\right)} dldm$$

- Easy solution: facetting
    - But: slow, stitching artefacts
- Better & most used solution: 'w-projection'

The w-term

- Visibility function:

$$V(u,v,w) = \iint \frac{I(l,m)}{\sqrt{1-l^2-m^2}} e^{-2\pi i\left(ul+vm+w(\sqrt{1-l^2-m^2}-1)\right)} dldm$$

- W-projection: *(Cornwell et al, 2008)*

$$V(u,v,w) * \mathcal{F}(e^{-2\pi i w(\sqrt{1-l^2-m^2}-1)}) = \iint \frac{I(l,m)}{\sqrt{1-l^2-m^2}} e^{-2\pi i(ul+vm)} dldm$$

This convolution turns out
to have a “limited” support

- Performance very dependent on zenith angle, coplanarity of array, field of view and resolution.

w-projection

- Another problem; convolution theorem no longer works when w-terms present in

$$V(u,v,w) = \iint \frac{I(l,m)}{\sqrt{1-l^2-m^2}} e^{-2\pi i\left(ul+vm+w(\sqrt{1-l^2-m^2}-1)\right)} dl dm$$

- Högbom CLEAN assumes constant PSF
- But PSF changes (slightly) over the image
- Solved with Cotton-Schwab algorithm (schwab 1984)
- Normal CASA imaging mode will automatically use CS

w-projection

- Another problem; convolution theorem no longer works when w-terms present in

$$V(u,v,w) = \iint \frac{I(l,m)}{\sqrt{1-l^2-m^2}} e^{-2\pi i \left(ul+vm+w(\sqrt{1-l^2-m^2}-1)\right)} dl dm$$

- Högbom CLEAN assumes constant PSF
- But PSF changes (slightly) over the image
- Solved with Cotton-Schwab algorithm (schwab 1984)
- Normal CASA imaging mode will automatically use CS (i.e., **C**otton-**S**chwab, *not* **C**ompressed **S**ensing)

w-projection

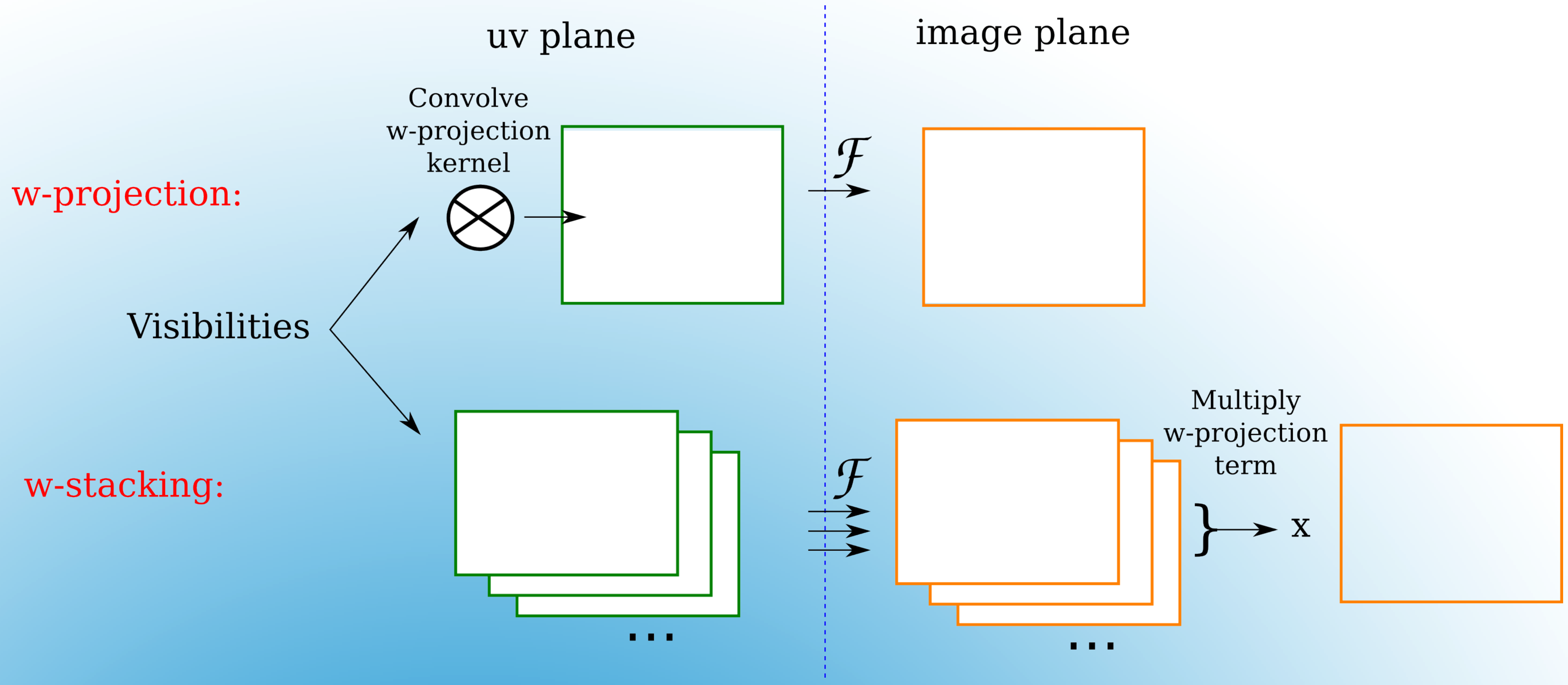
- The Cotton-Schwab + w-projection algorithm:
  - Make initial dirty image & central PSF
  - Perform minor iterations:
    - Find peak
    - Subtract scaled PSF at peak with small gain
    - Repeat until highest peak ~ 80-90% decreased
  - Major iteration: “Correct” residual
    - Predict visibility for current model
    - Subtract predicted contribution and re-image

w-projection

- W-projection is the standard way to solve w-terms in radio astronomy
- W-term convolution can be slow
  - Imaging 2 minutes of data of the MWA telescope (30 degree FOV) costs hours
- New imager with new algorithm implemented: WSClean[1] ("w-stacking clean").
  - Offringa et al, 2014

[1]see http://wsclean.sourceforge.net/

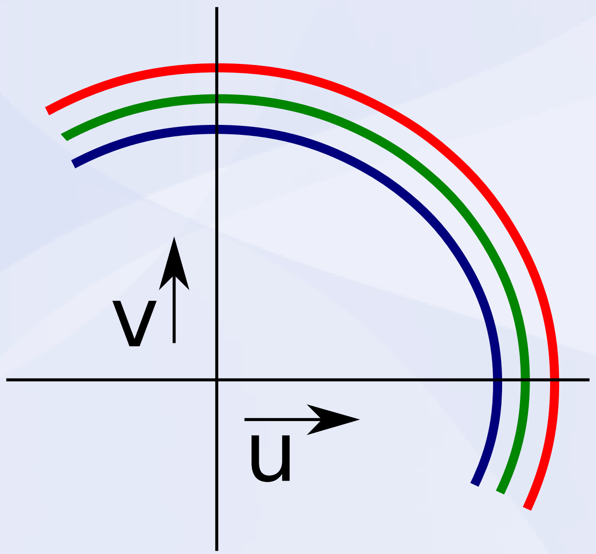
w-projection


uv plane
image plane
Convolve
w-projection
kernel
w-projection:
$\mathcal{F}$
Visibilities
w-stacking:
$\mathcal{F}$
...
Multiply
w-projection
term
x
...

w-stacking

Multi-frequency synthesis

- Multi-frequency synthesis (MFS) means gridding different frequencies on the same uv grid:



- This is the standard for modern telescopes
- Appropriate citation: **Sault & Conway (1999)**

Synthesis Imaging in Radio Astronomy II
ASP Conference Series, Vol. 180, 1999
G. B. Taylor, C. L. Carilli, and R. A. Perley (eds.)

21. Multi-Frequency Synthesis

R.J. Sault
Australia Telescope National Facility, Epping, NSW 2121, Australia

J.E. Conway
Onsala Space Observatory, S-439 00 Onsala, Sweden

Abstract. Multi-frequency synthesis is the practice of using visibility data measured over a range of frequencies when forming a continuum image. Because observing frequency is easier to vary than antenna location, it is an effective way of filling the (u, v) plane for an observation. Here we consider the artifacts in MFS images caused by source spectral variation. For frequency ranges of about 30%, for observations where only modest dynamic range is required, the artifacts of MFS can be completely ignored. For higher dynamic range observations, some calibration techniques and deconvolution algorithms are described which minimize the artifacts.

Definition from the bible:

'mfs' : Images made by combining data from multiple channels use multi-frequency-synthesis to grid visibilities according to their respective observing frequencies. The frequency at which the output image is created is the middle of the sampled frequency range.

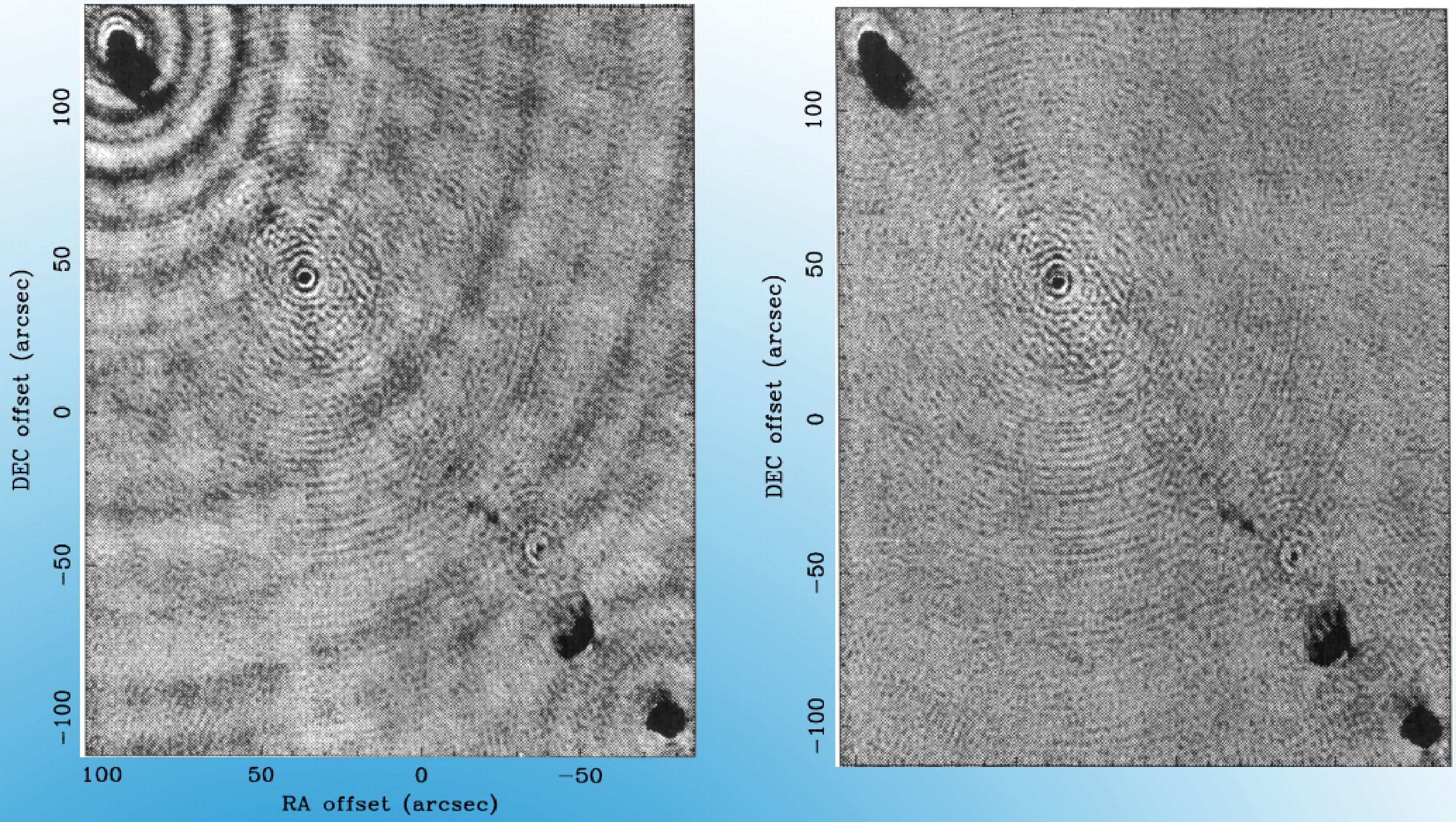
Related, but not the same:

- **Multi-frequency deconvolution** sometimes called **multi-term deconvolution**

  Selected by setting **`nterms`** *in CASA's clean task*

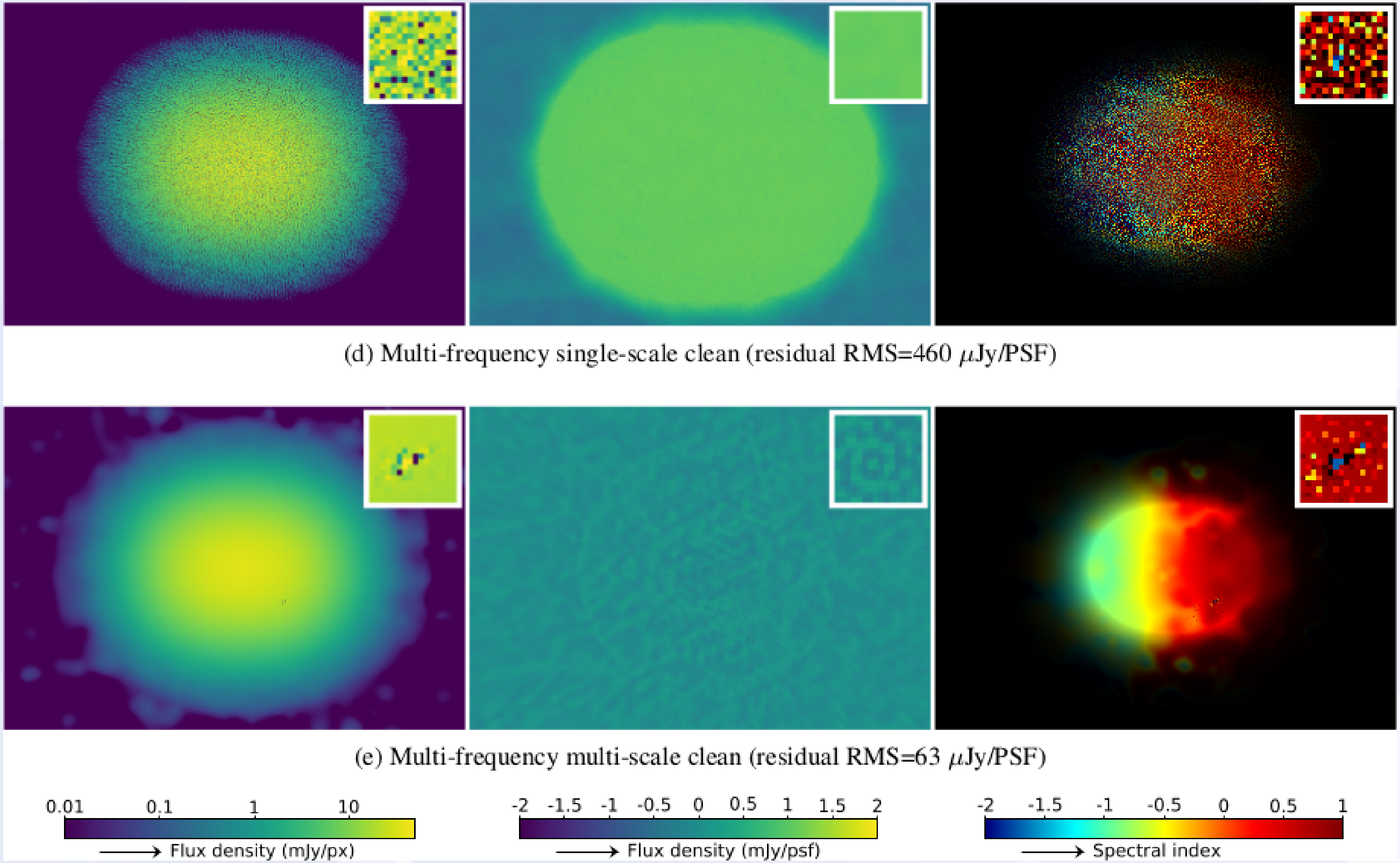
- Takes spectral variation into account during deconvolution
- Useful for wideband, sensitive imaging
- MSMFS in CASA (Rau and Cornwell 2011),
  Joined channel cleaning in WSClean (Offringa and Smirnov 2017)

Multi-frequency deconvolution



- Right image: fit for flux over frequency to improve deconvolution (Sault & Wieringa, 1994)

Frequency-dependent deconvolution



(d) Multi-frequency single-scale clean (residual RMS=460 μJy/PSF)

(e) Multi-frequency multi-scale clean (residual RMS=63 μJy/PSF)

- Comparison of WSClean MF single scale and multi-scale cleaning
- Simulated bandwidth of 30 MHz at 150 MHz.
- MWA layout, 2 min snapshot

Offringa and Smirnov (2017)

Nomenclature summary & references

- MFS: Multiple channels gridded

  Sault & Conway (1999)

- Multi-term/multi-frequency deconvolution: nterms>1

  Sault & Wieringa (1994)

- MSMFS: Deconvolution with `nterms`>1 **and multi-scale**

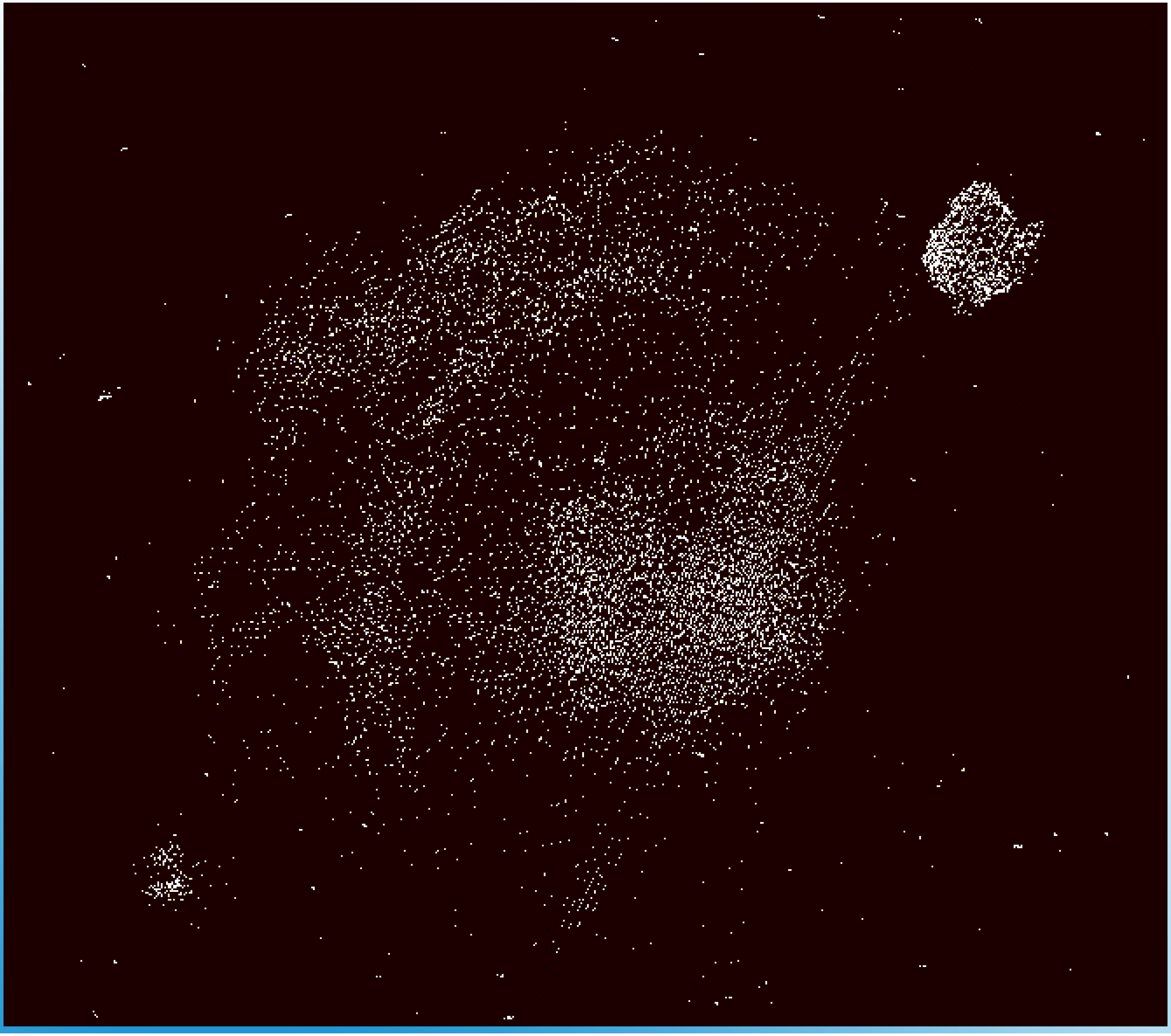
  Rau & Cornwell (2011)

- Joined-frequency cleaning: channelsout>1

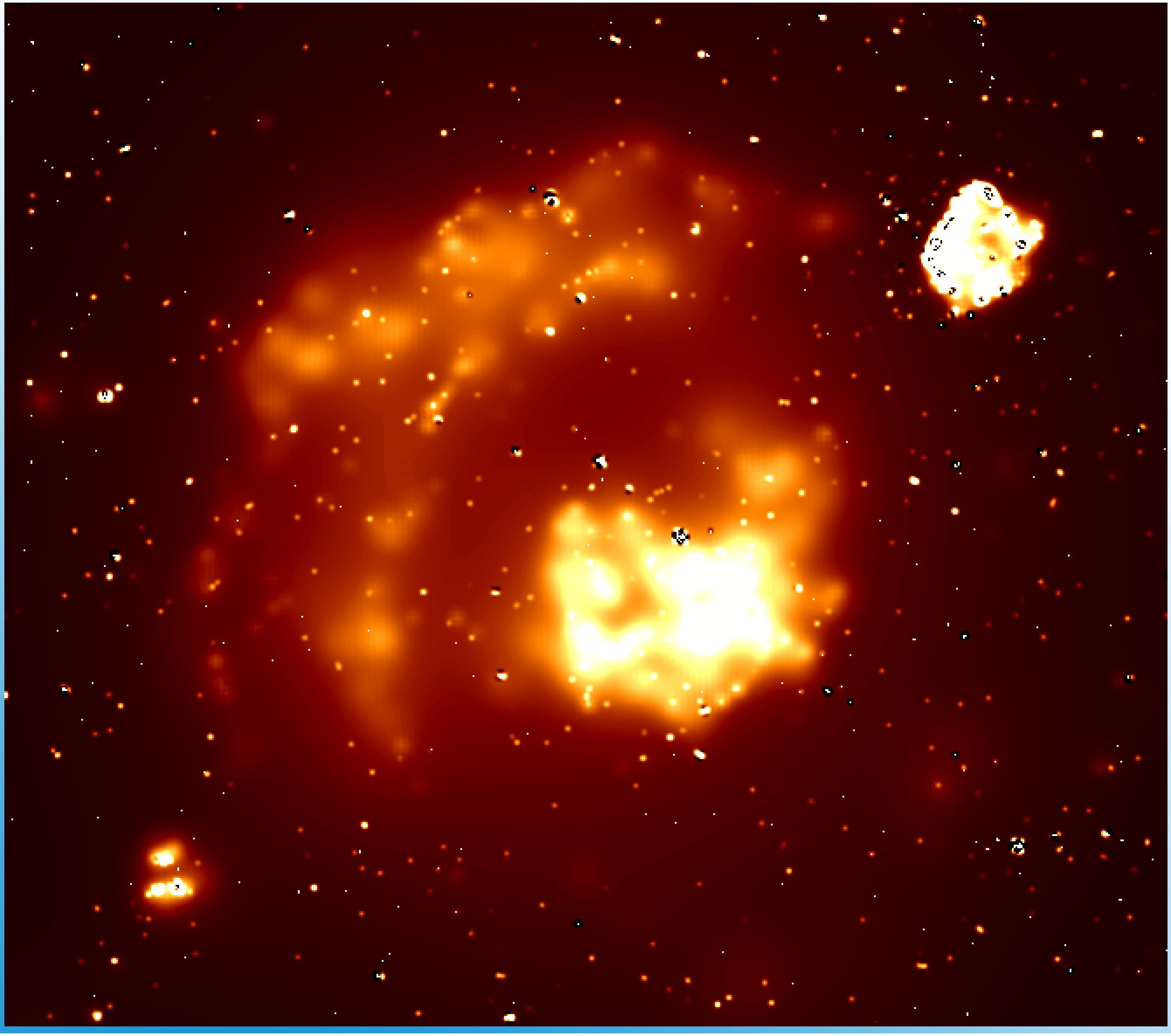
  In WSClean: Offringa & Smirnov (2017)
  In OBIT: Cotton (2008)

- Recent focus on deconvolution using '**c**ompressed **s**ensing' (abbrev. CS – but CS can mean "Cotton-Schwab" too)
- CS methods assume the sky is 'sparse' ("solution matrix is sparse in some basis")
- Minimizes "L1-norm" (= abs sum of CLEAN components)
- Högbom clean is actually (almost) a compressed sensing method called "Matching Pursuit"
- CS considers MP to be non-ideal… but radio data is not the perfect CS case: Calibration errors, w-terms

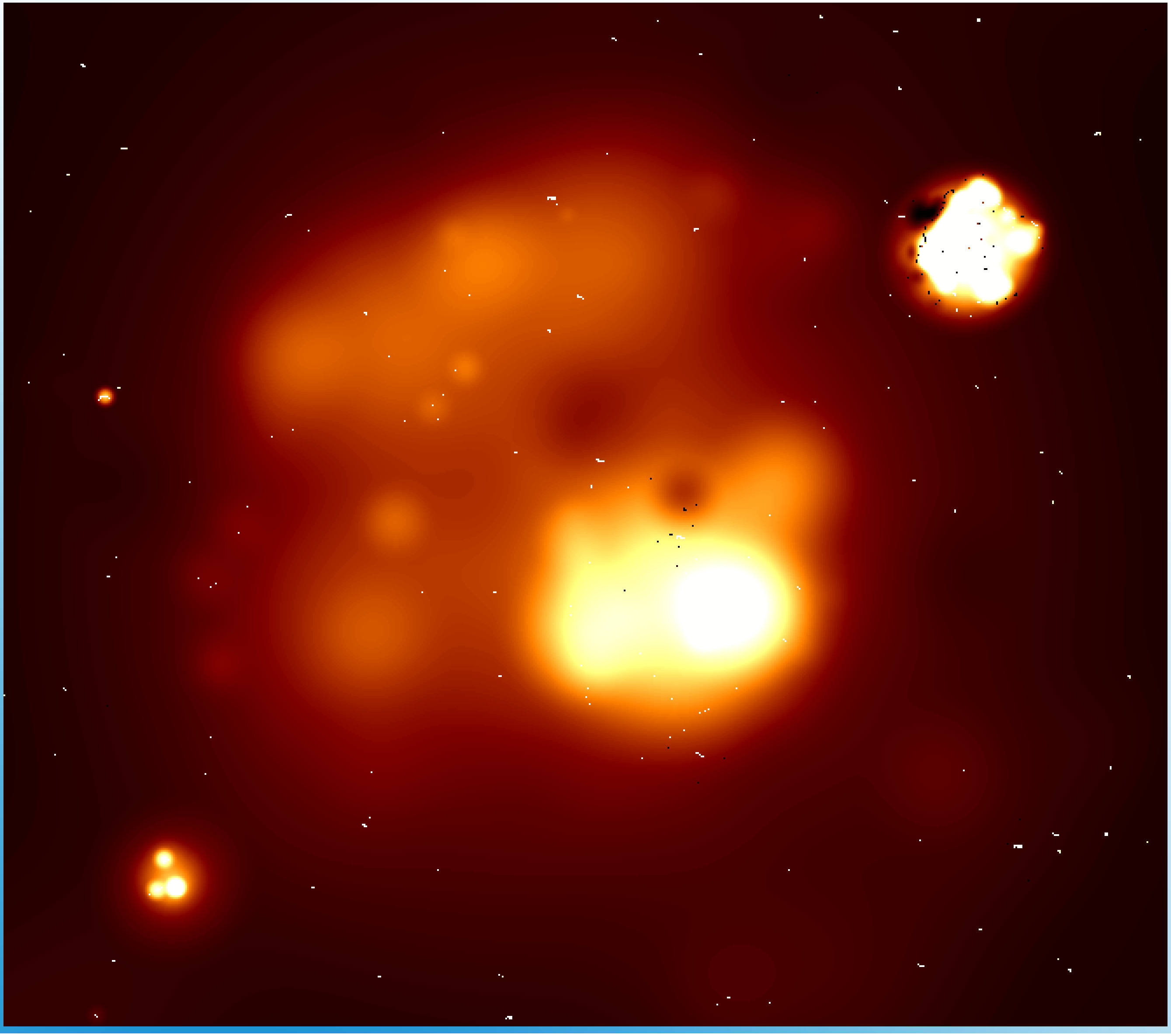
Compressed sensing



Model created by Högbom clean



Model created by a CS method
("non-linear conjugate gradient using IUWT")

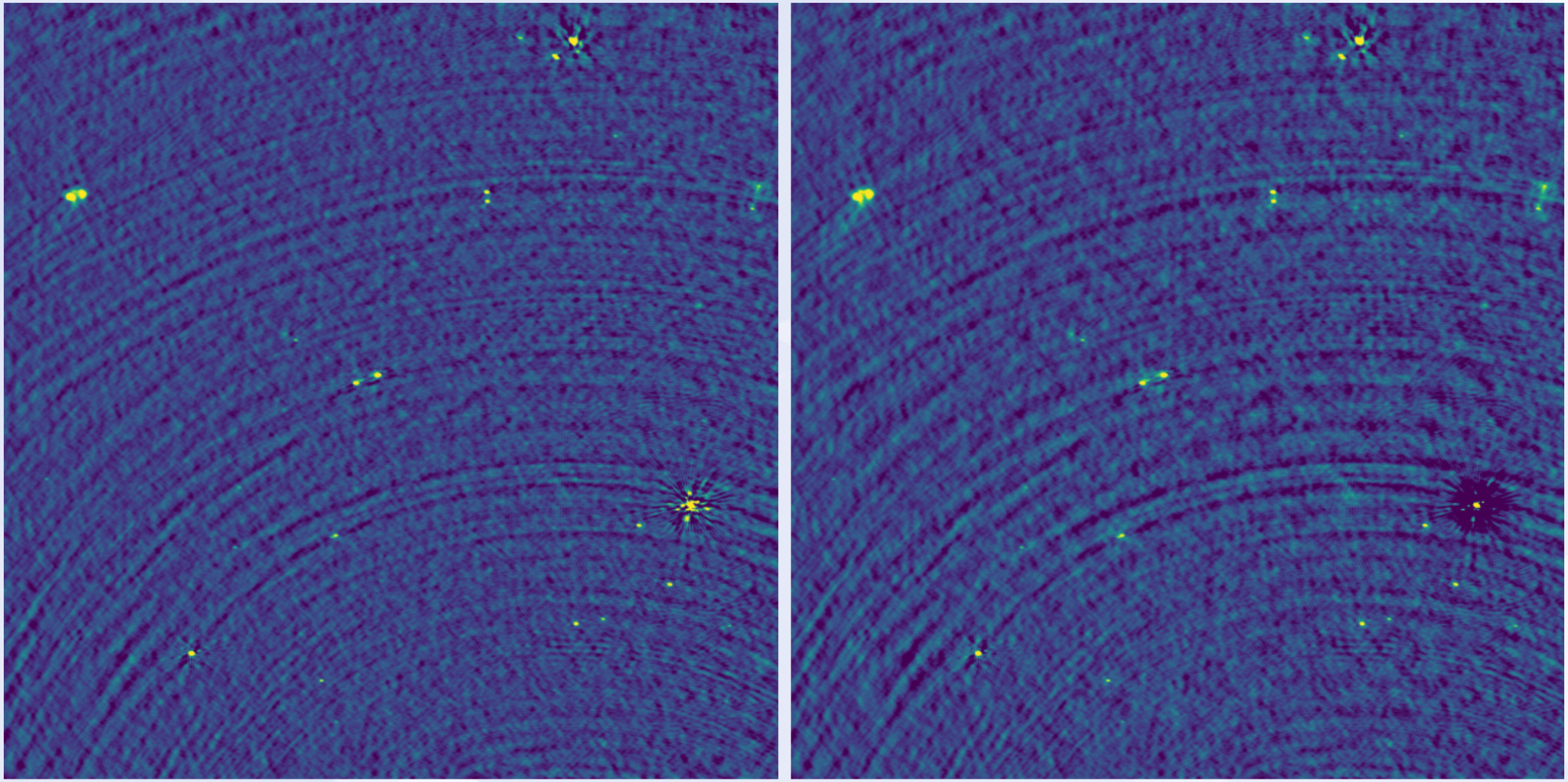


Model created by multi-scale clean

- Compressed sensing does not work well with calibration artefacts
- Multi-scale is more robust
- On well-calibrated data:
  - CS gives more accurate model
  - But residuals don't improve much
- Compressive sensing methods are still in development, and only in very specific cases practically useful.

Compressed sensing

Issue with stability of compressive sensing methods



(a) WSCLEAN multi-frequency, multi-scale with β=0.6 (rms=1.4 mJy)

(b) WSCLEAN multi-frequency, iuwt (rms=2,7 mJy)

Offringa and Smirnov (2017)

Self-cal & CLEAN

- Clean components can be used as calibration model
- Often applied as:

Phase cal
↓
Shallow clean
↓
Phase cal
↓
Deep clean
↓
Phase & amplitude cal
↓
Deep clean

Self-cal & CLEAN

- Clean components can be used as calibration model
- Often applied as:

Phase cal
↓
Shallow clean
↓
Phase cal — **Why only phase?**
↓
Deep clean
↓
Phase & amplitude cal
↓
Deep clean

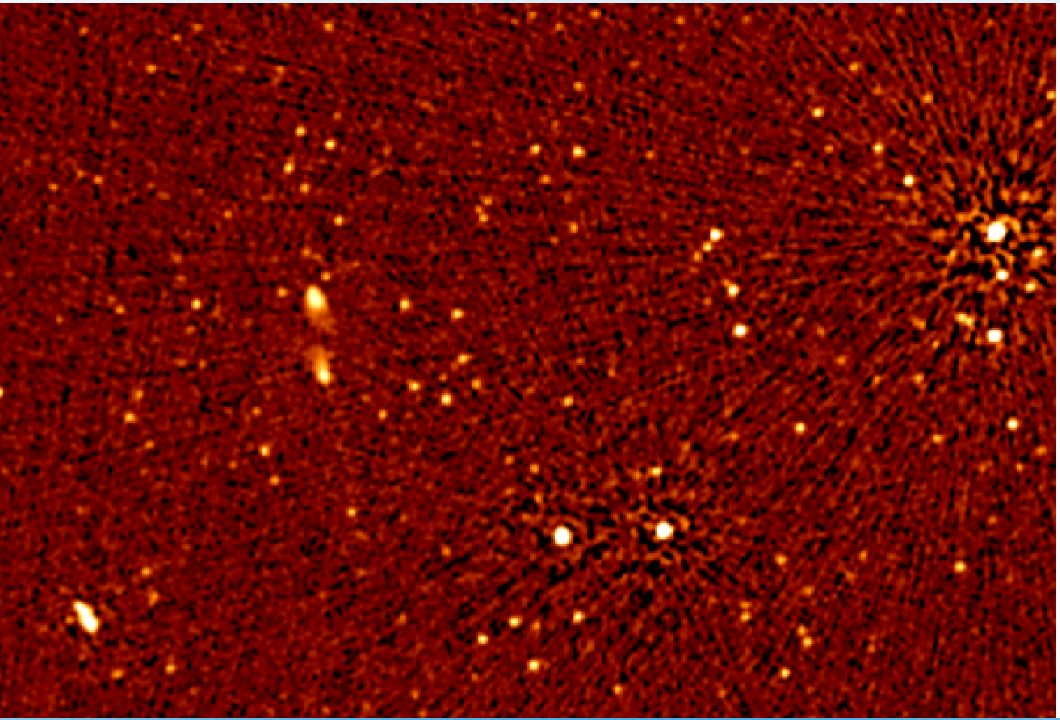
Self-cal & CLEAN

- Clean components can be used as calibration model
- Often applied as:

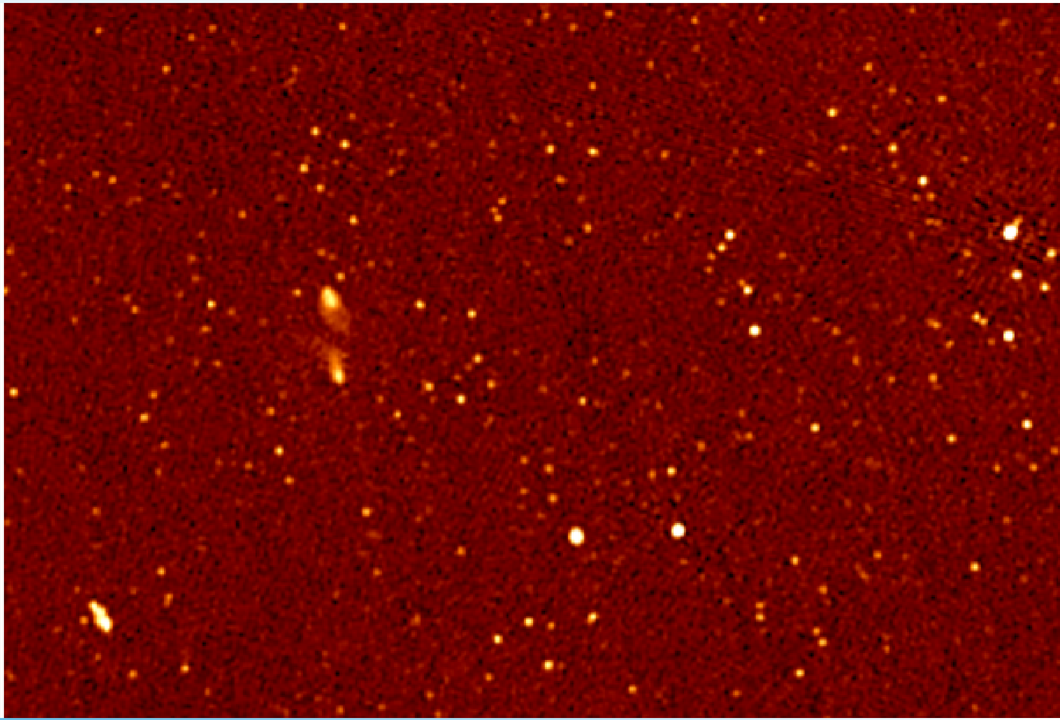
Phase cal
↓
Shallow clean
↓
Phase cal
↓
Deep clean
↓
Phase & amplitude cal
↓
Deep clean

Why only phase?

→ Avoid self-cal bias



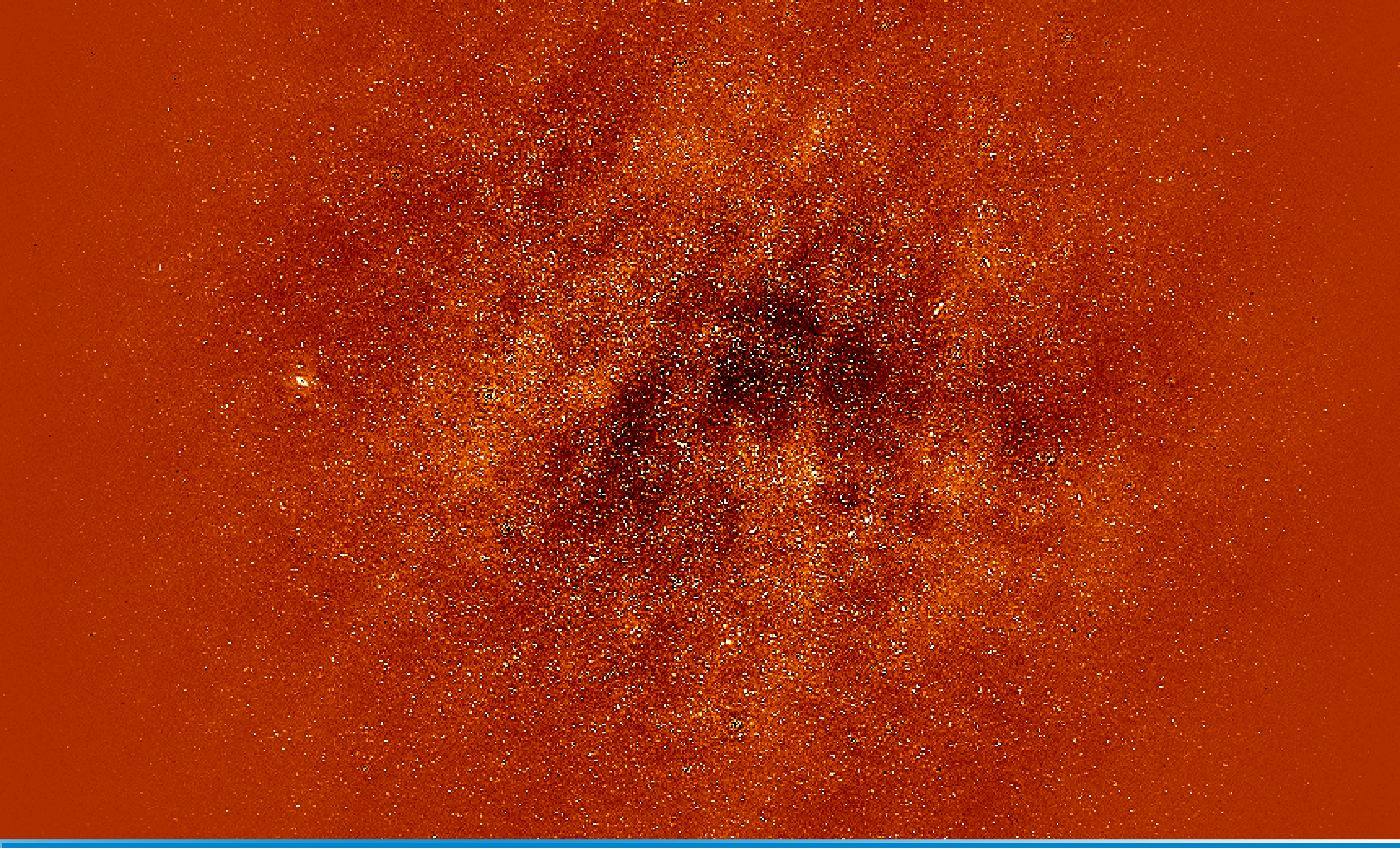
After initial calibration



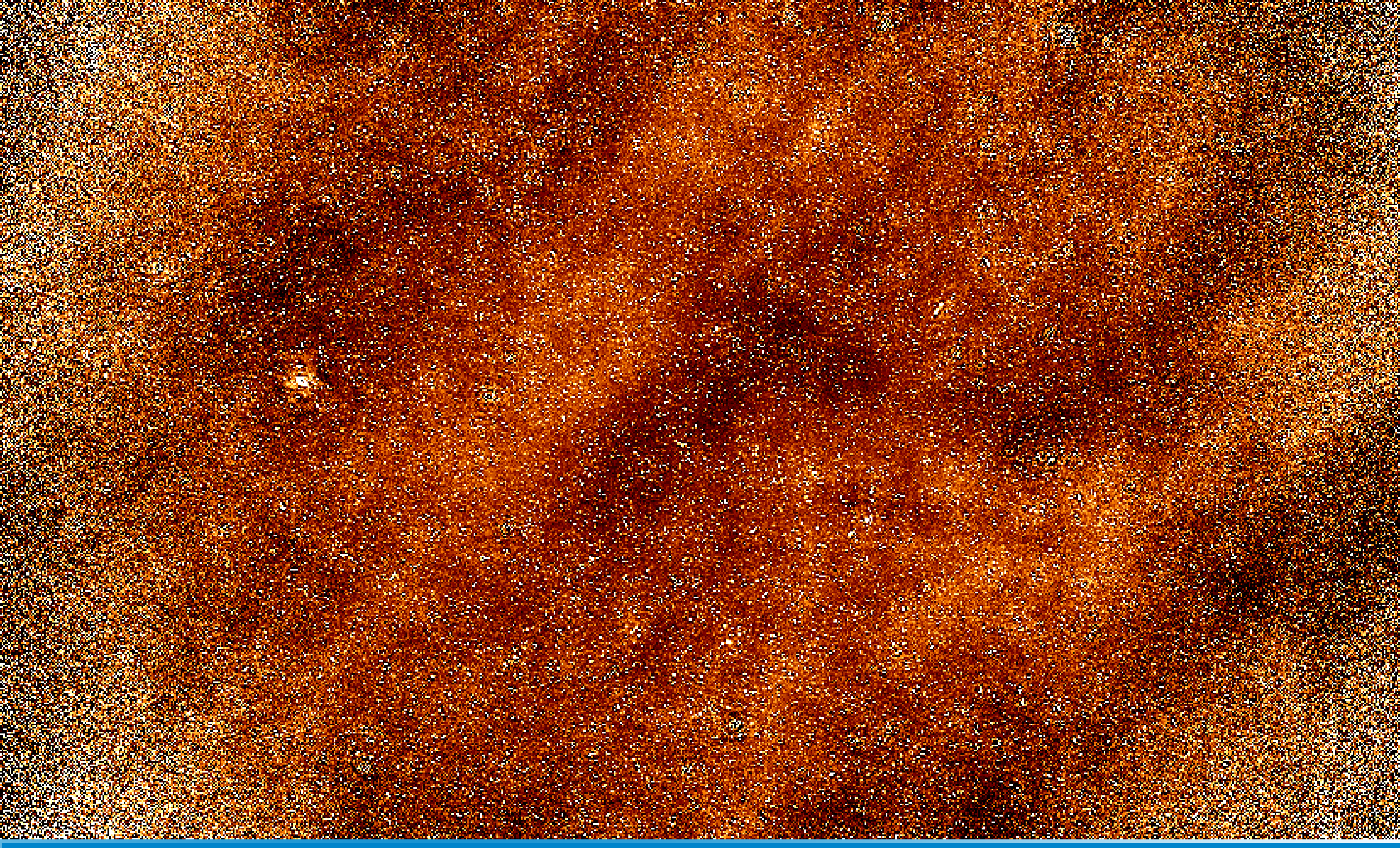
After self-cal on clean components

Image credit: N. Hurley walker (using the MWA)

Self-calibration using CLEAN



- **Result of imaging – what is still missing?**



- **Result of imaging – now with beam correction**

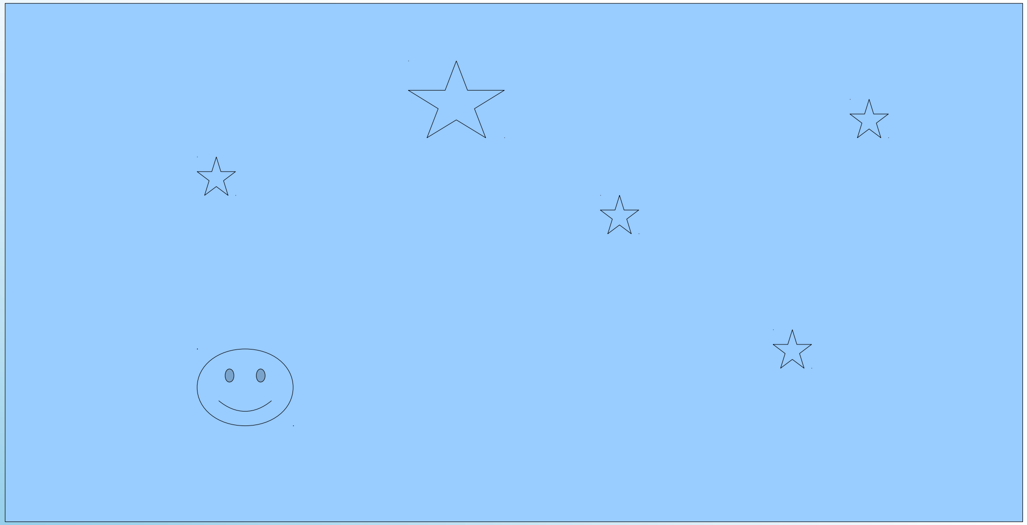
- Correction is required for the antenna response
- This is called “primary beam” correction
(as opposed to the synthesized beam / psf)

- For dishes, the primary beam is ~constant
- To correct for: multiply final image with the inverse beam
- **Scalar** for total brightness, **matrix** for polarized

Primary beam correction

What if…

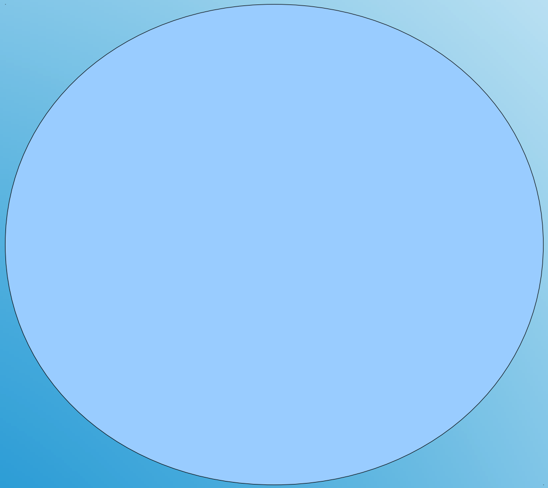
This is our field of interest →



(In practice, actual galaxies look different)

… and

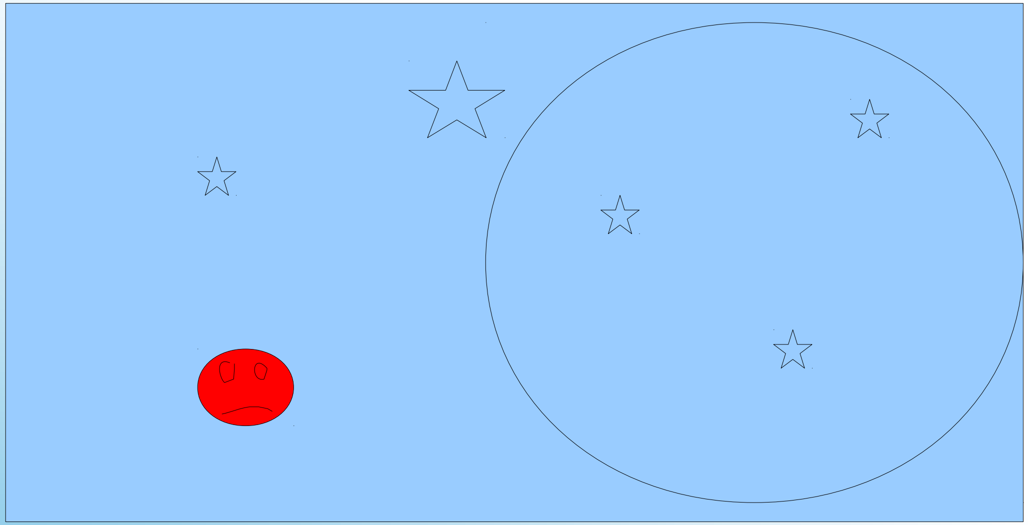
this is our primary beam →



Mosaicing

What if…

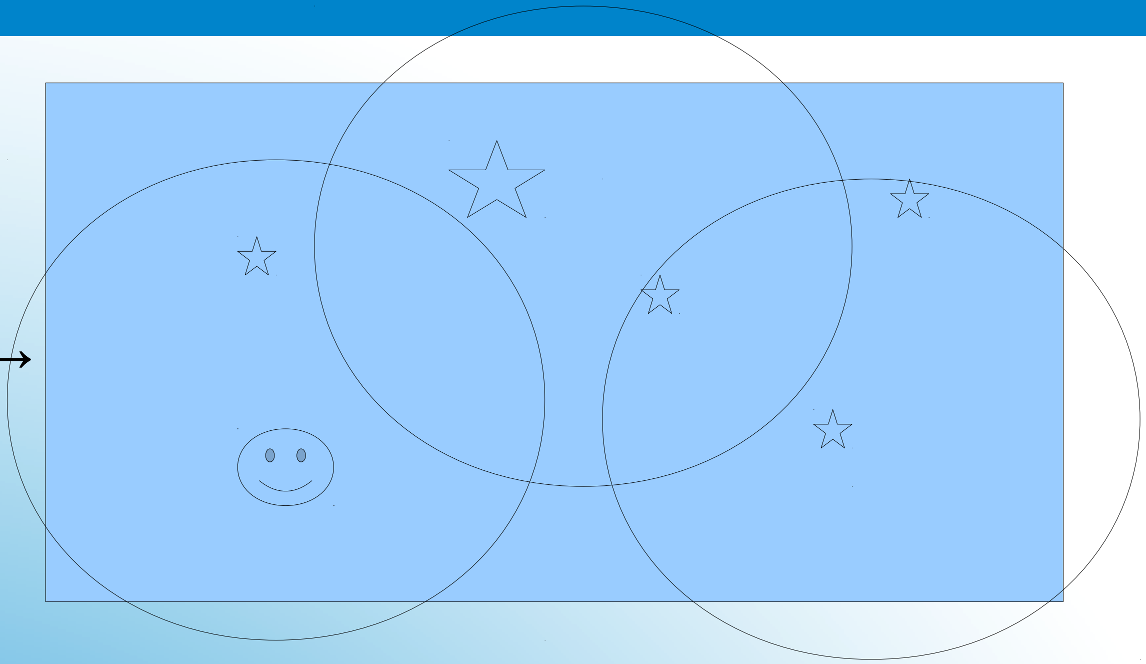
This is our field of interest→



(In practice, actual galaxies look different)

What if…

This is our field of interest →



- This is called mosaicking
- Should we average the 3 primary-beam-corrected images together?

Mosaicking

Inverse-variance weighting

Primary-beam-corrected image

$$M(l,m) = \frac{\sum_i B_i^2(l,m)\left(I_i(l,m)/B_i(l,m)\right)}{\sum_i B_i^2(l,m)} = \frac{\sum_i B_i(l,m) I_i(l,m)}{\sum_i B_i^2(l,m)}$$

- This is called mosaicking
- Should we average the 3 primary-beam-corrected images together?

No → Weight with $1/\sigma^2$ = (primary beam)^2

Mosaicing

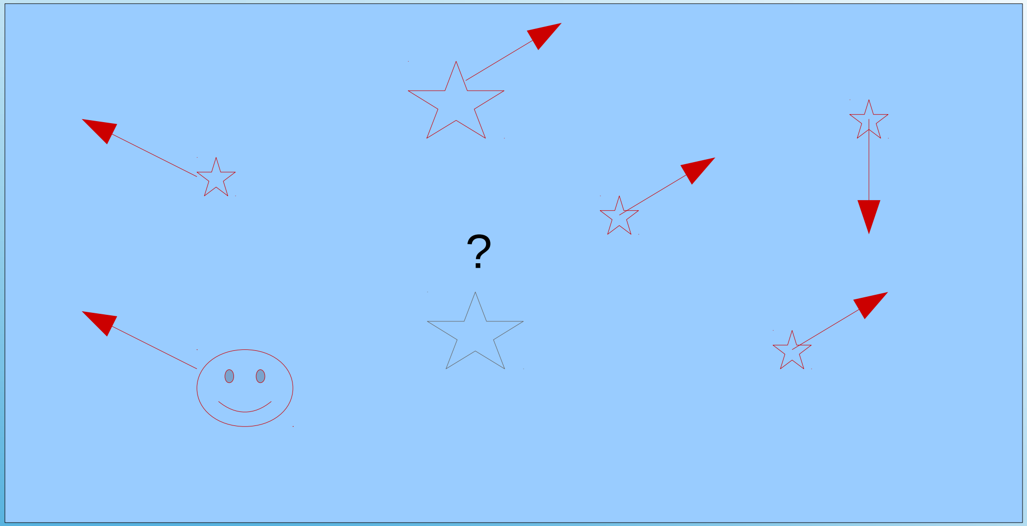
- Primary beam of tiled arrays varies in time, per station

Variable primary beam

Variable primary beam

- Primary beam of tiled arrays varies in time
  - Or even per station
- Has to be accounted for during cleaning
- Algorithm to do this is "aw-projection"
  - similar to w-projection
  - Specialized software package for LOFAR ("**AWImager**")
- Homogeneous arrays can also use snapshot imaging

- **Direction-dependent effects** might require further correction during imaging:

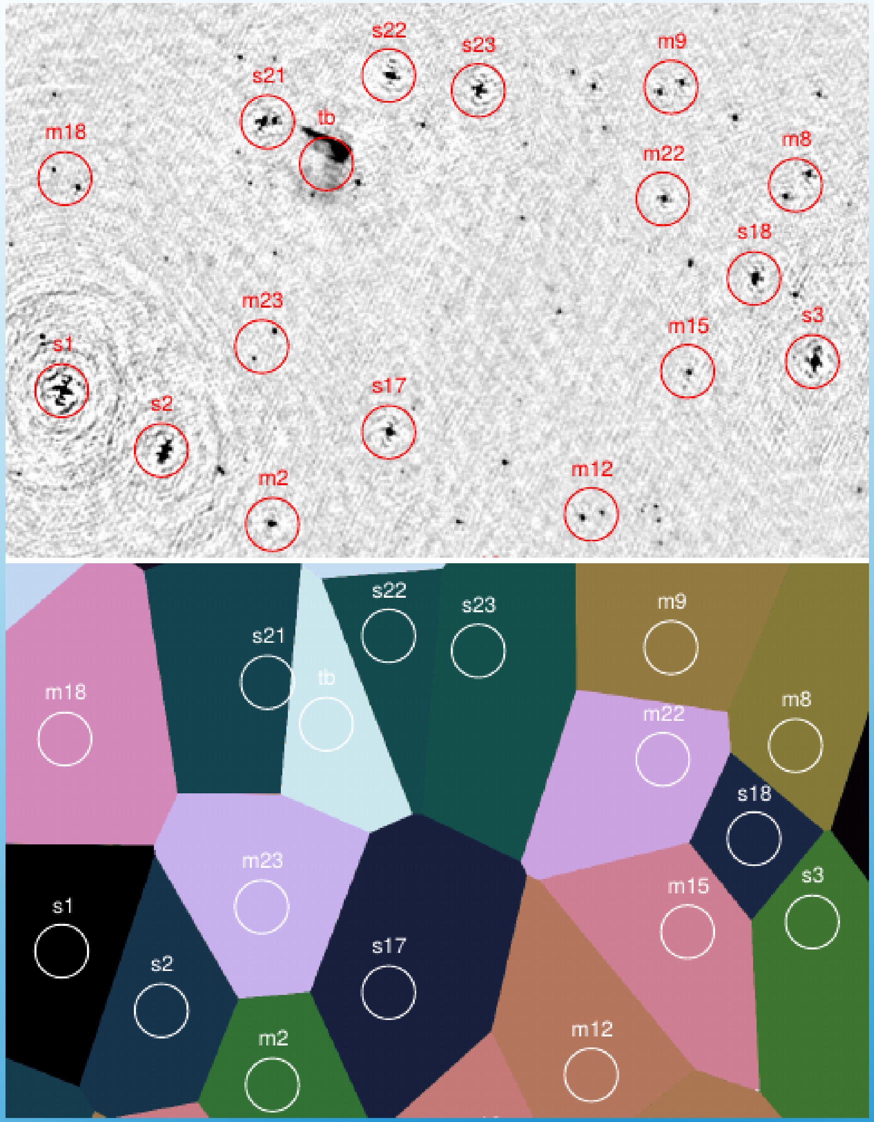


- Positions of 'calibrators' (red) are known
- Apparent position has moved due to ionosphere

More variable effects...

Direction-dependent effects

- **Direction-dependent effects** might be time-variable (e.g. ionosphere)
- Besides position, DD effects can also affect polarization angle and brightness
- Not a fully solved problem, but possible solutions:
  - image in small “facets” where DDE's are constant
    Popular software: Factor (van Weeren), DDFacet (Tasse)
  - or interpolation – AWImager can do this.
  - Peeling



Factor is one of the pipelines to produce high-resolution high-dynamic-range images.

- Works by facetting the sky
- Each facet is independently self-calibrated

Factor

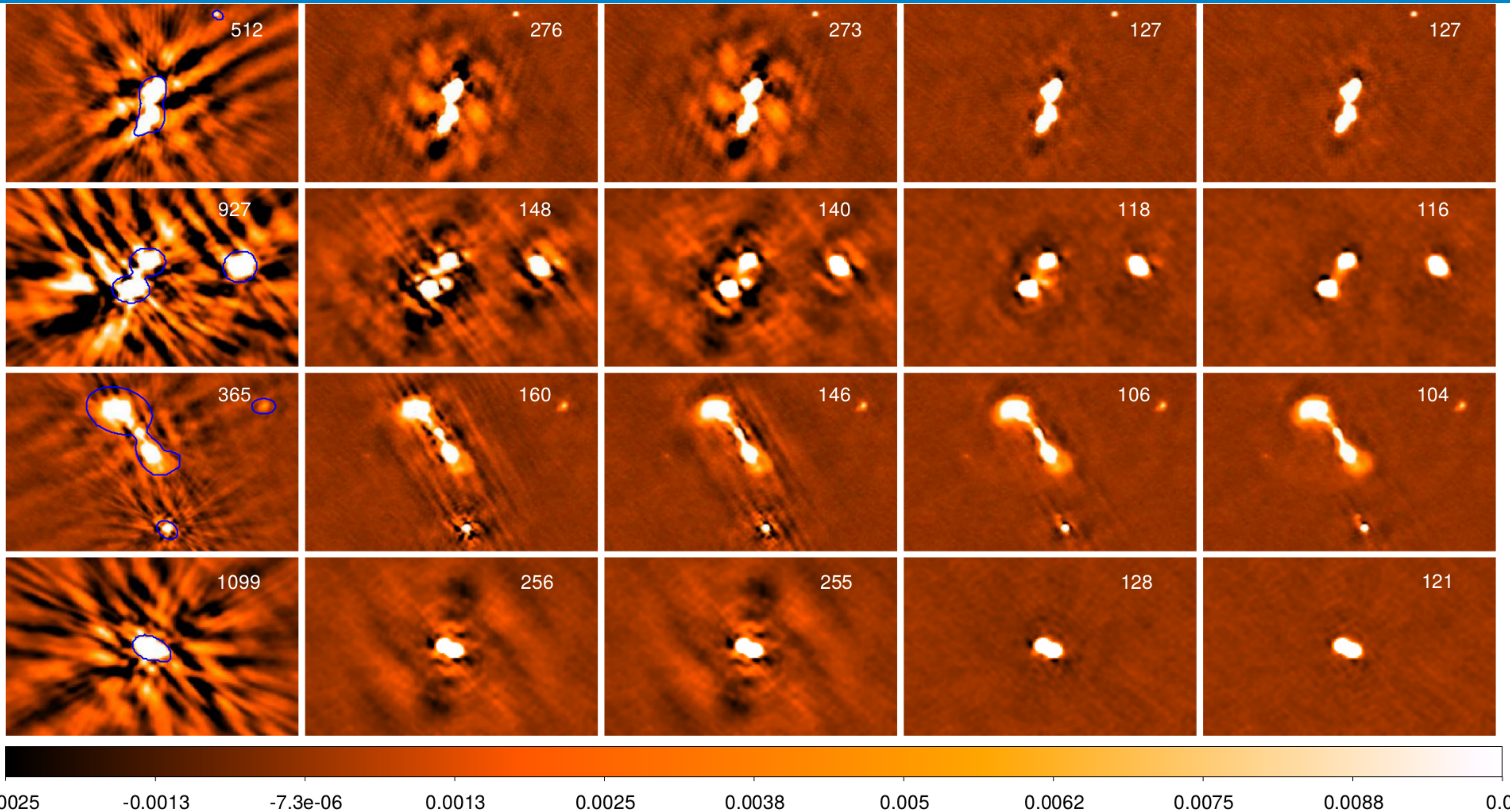
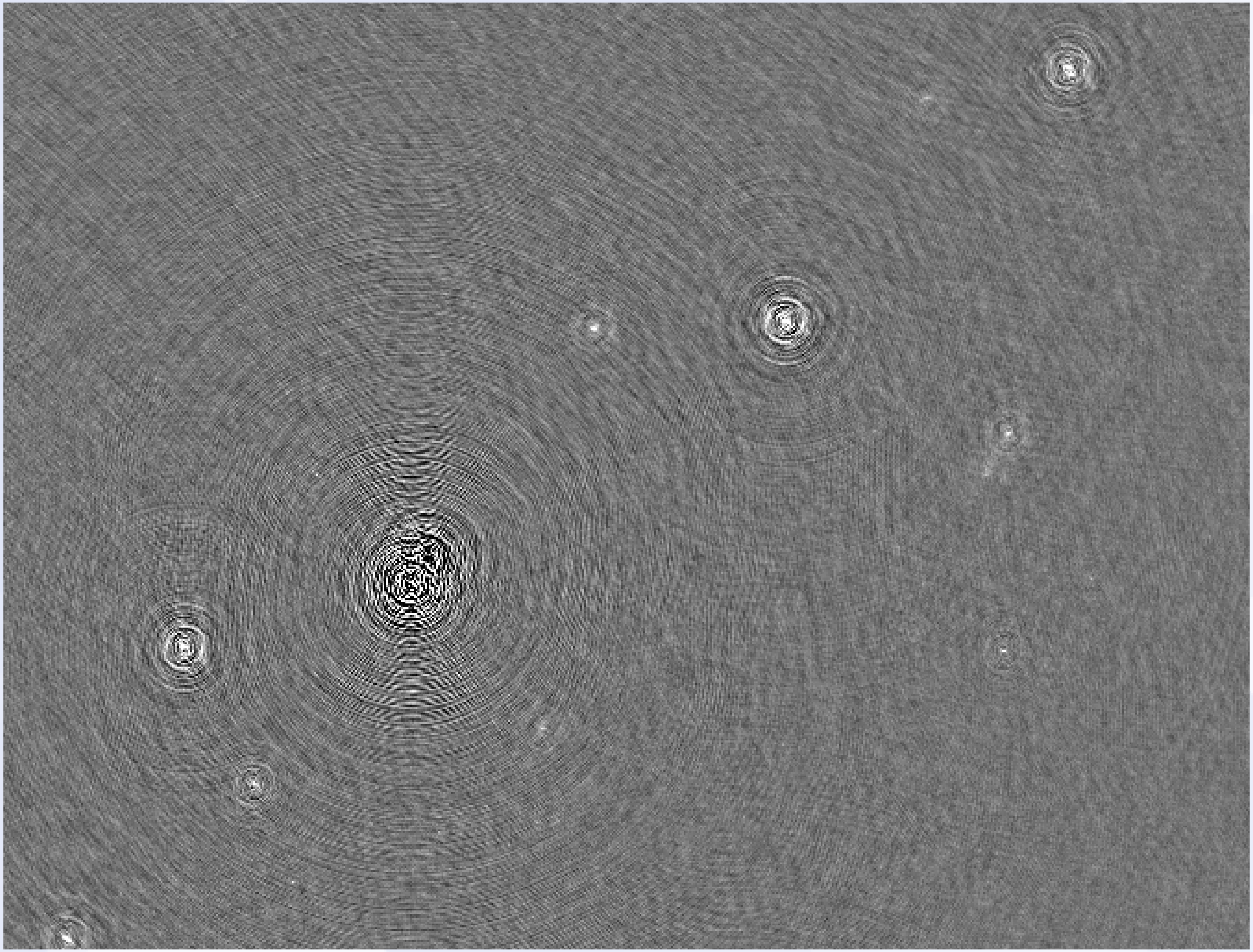
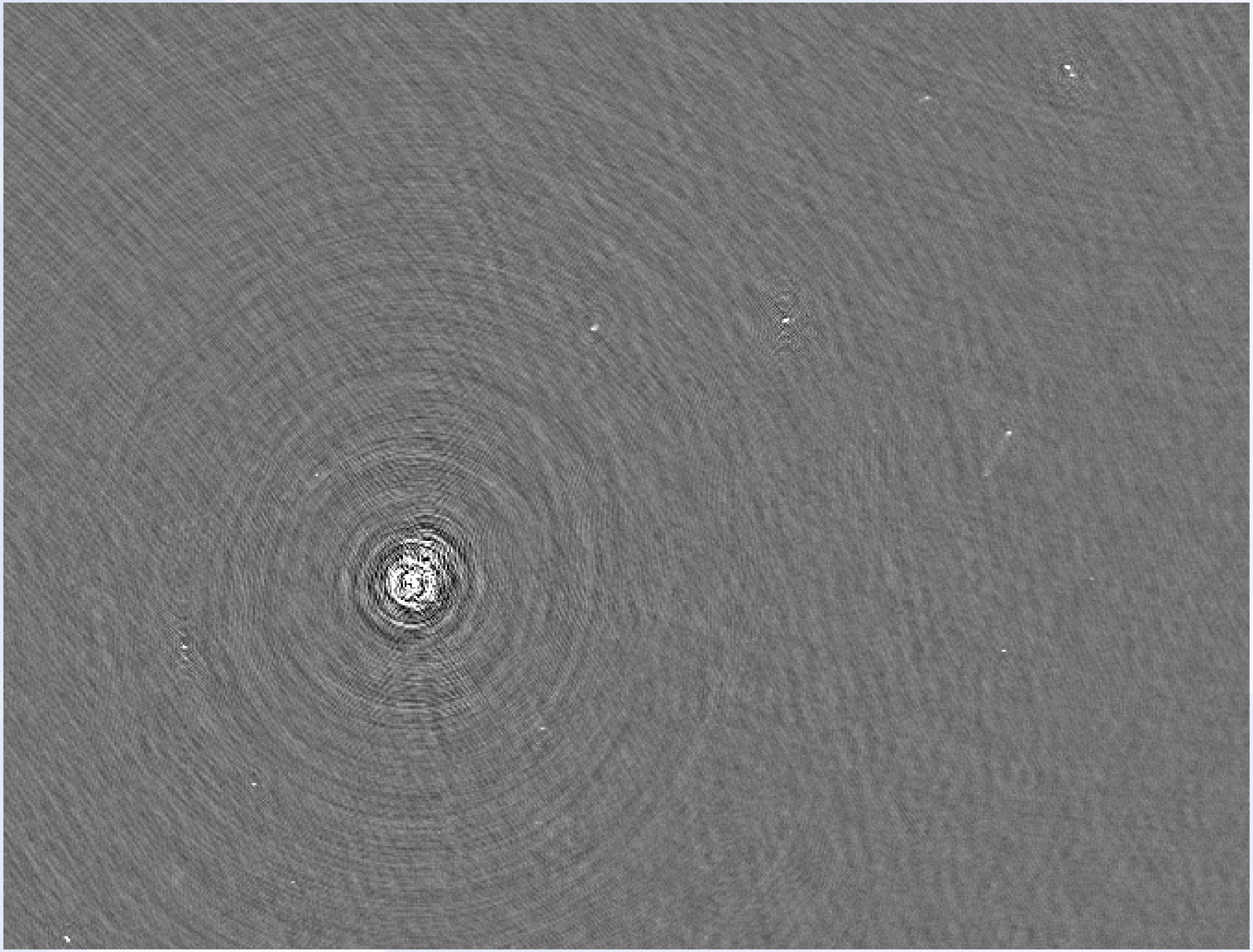


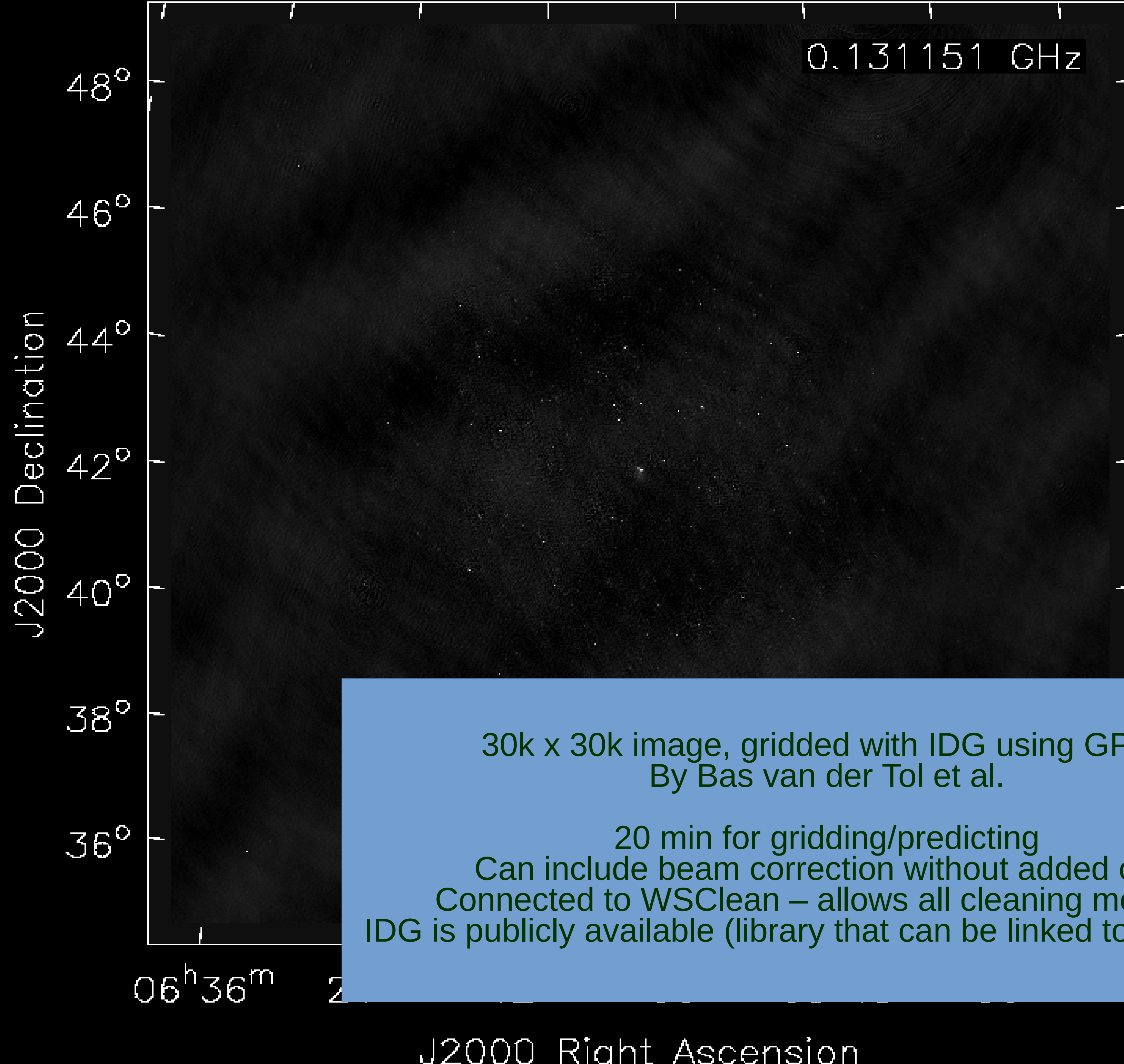
FIG. 10.— Images showing the incremental improvements during the DDE calibration, see Sect. 5.3. For reference, the first and second row of images show direction s2 and s21, respectively (Figure 6). All images are made using the full dataset (120–181 MHz, `robust=-0.25`) and have a resolution of $8'' \times 6.5''$. Note that at this resolution many of the bright DDE calibrator sources are resolved. The first column displays the images made with the (direction independent) self-calibration solutions, see Sect 4.6. The blue contours show the clean mask that was created with `PyBDSM` for the imaging. The clean mask is updated at each imaging step during the DDE calibration (not shown). The next columns display improvements during the DDE calibration step (see also Figure 9). Second column: first DDE TEC+phase iteration. Third column: second DDE TEC+phase iteration. Fourth column: third DDE TEC+phase iteration and first DDE XX and YY gain (amplitude and phase) iteration. Fifth column: fourth DDE TEC+phase iteration and second DDE XX and YY gain (amplitude and phase) iteration. For all four directions the TEC+phases were solved for on a 10 s timescale. The XX and YY gains were solved for on a 10 min timescale, except for the source in the top row for which this was 5 min. The scale bar at the bottom is in units of Jy beam^{-1}. The images in the first and third row were cleaned with multi-scale clean because of extended emission. The r.m.s. noise level in each of the images is indicated in the top right corner in units of μJy beam^{-1}.

Local RMS cleaning



Local RMS cleaning





30k x 30k image, gridded with IDG using GPUs
By Bas van der Tol et al.

20 min for gridding/predicting
Can include beam correction without added cost
Connected to WSClean – allows all cleaning methods
IDG is publicly available (library that can be linked to WSClean)

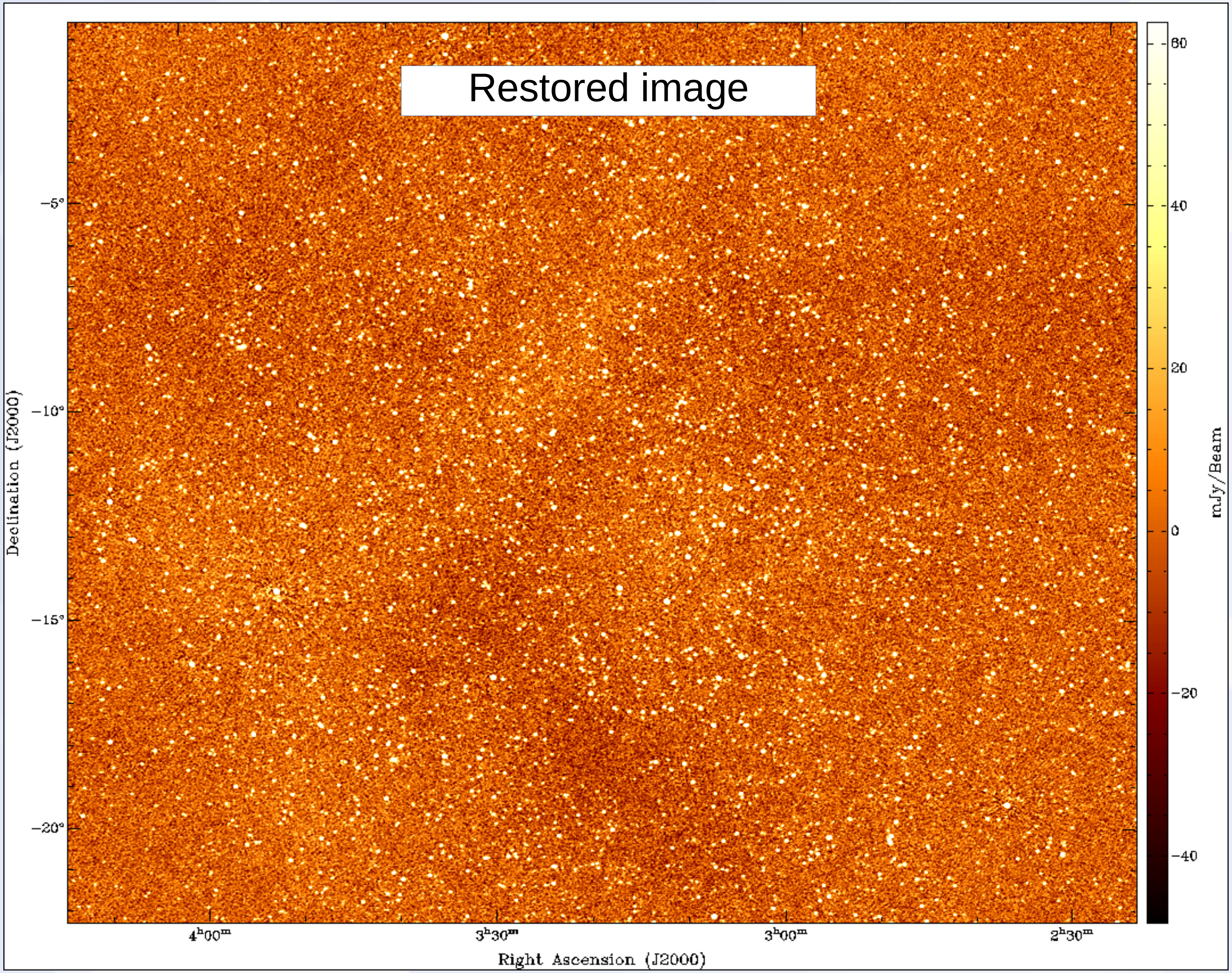
Automatic scale-dependent masking

- Normal cleaning requires manual threshold tweaking, manual masking, etc…
- Masking is hard when structures are diffuse
- Move towards non-interactive, fully automatic cleaning
- “Automatic scale-dependent masking” :
  - For each scale, a mask is accumulated
  - Clean normal to 3-5σ, continue to 0.5σ with a scale-dependent mask. In one run.

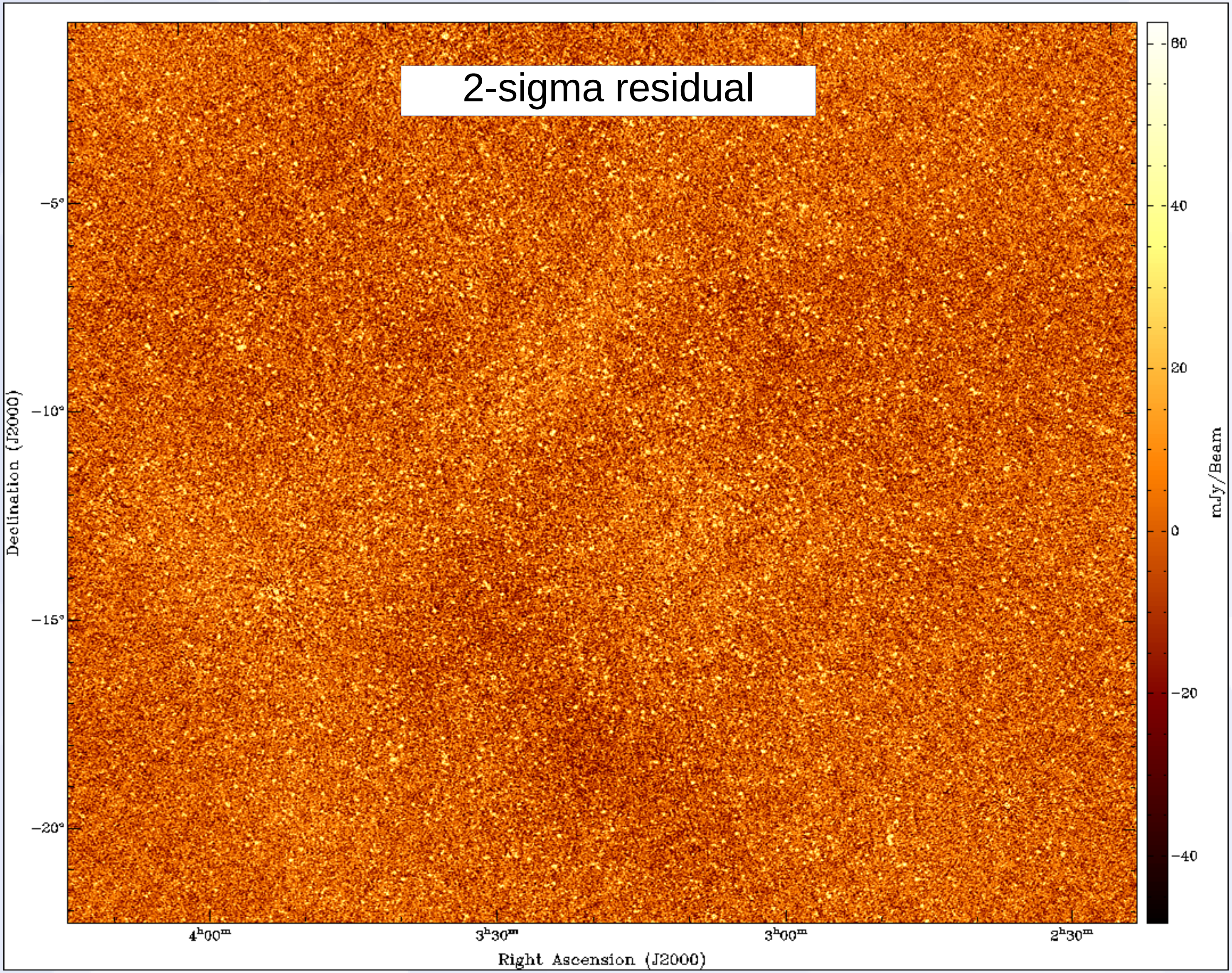
Automatic masking

- Threshold is relative to RMS estimate
- RMS estimate can be “local” when RMS is expected to change over the image
  (avoids picking up calibration errors)
- Avoids interaction & somewhat-arbitrary selection of features, etc.
- Allows deeper & more stable cleaning of complex structures. Limits clean bias.
- Can be done in multi-frequency mode

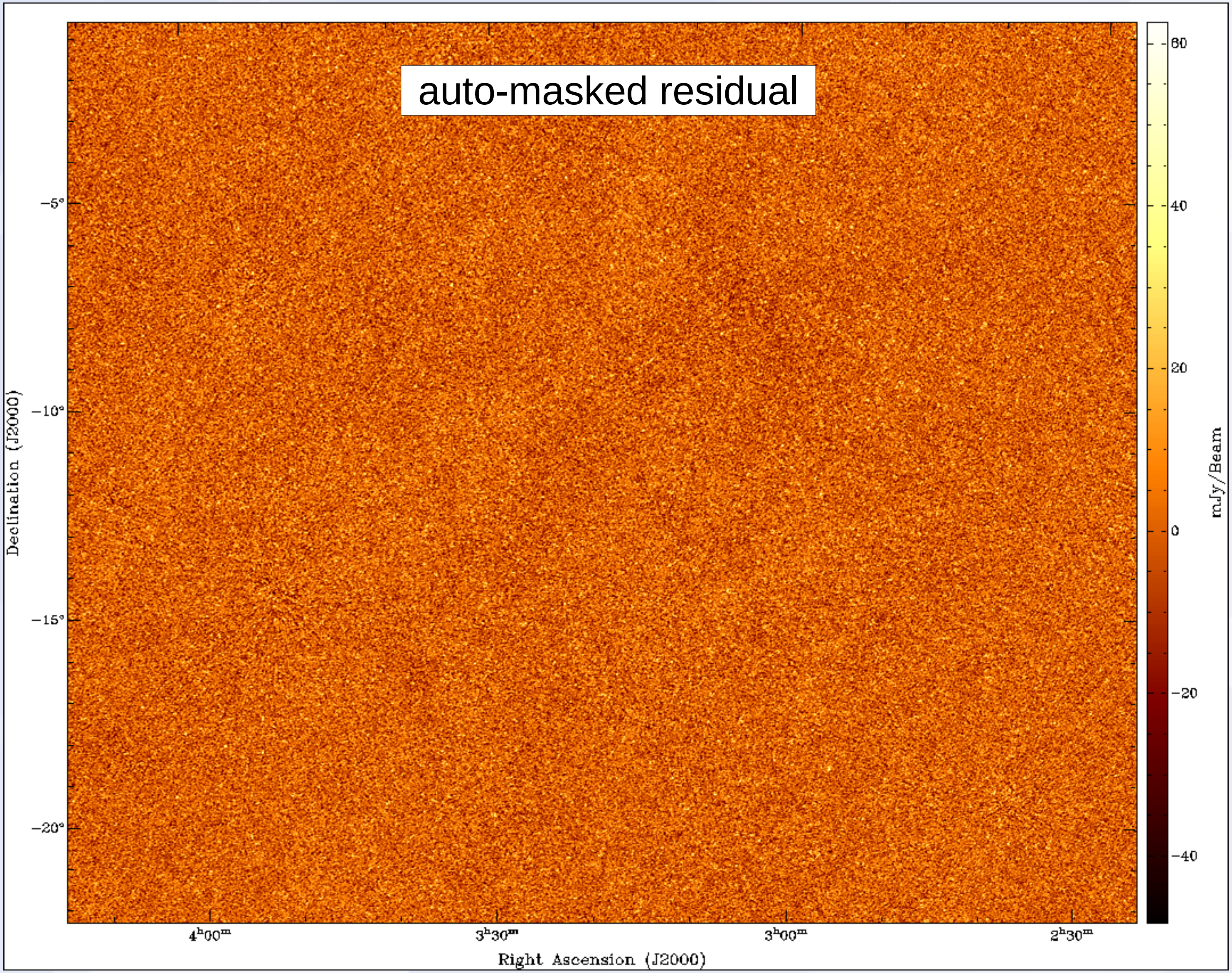
Auto-masking on point sources



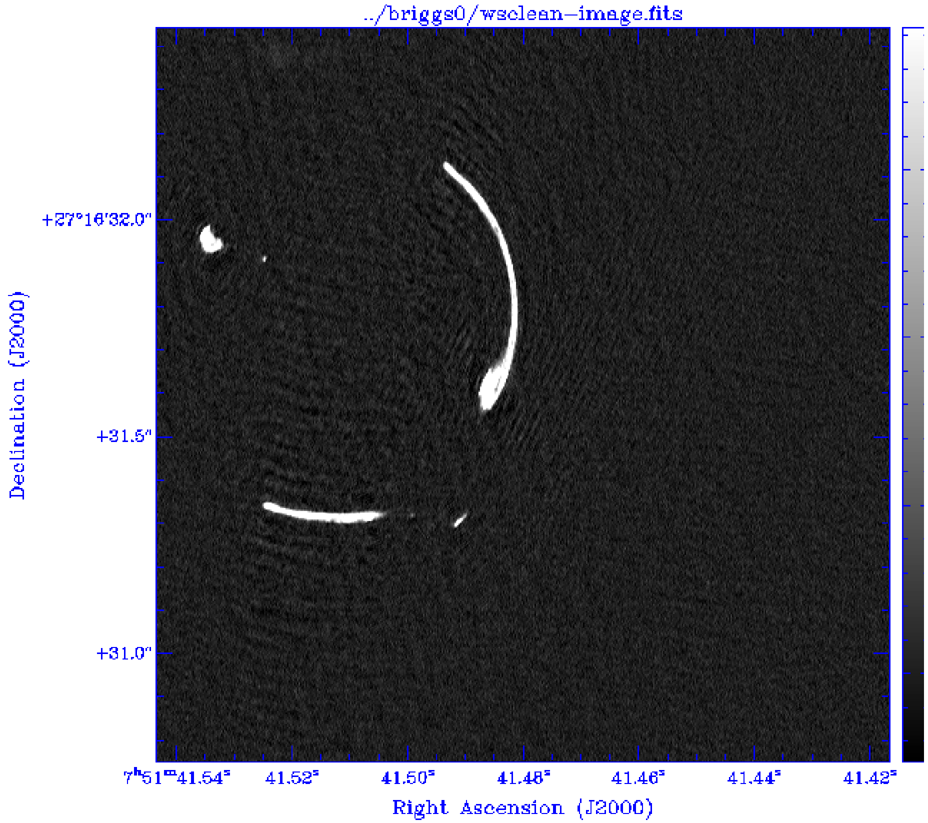
Auto-masking on point sources



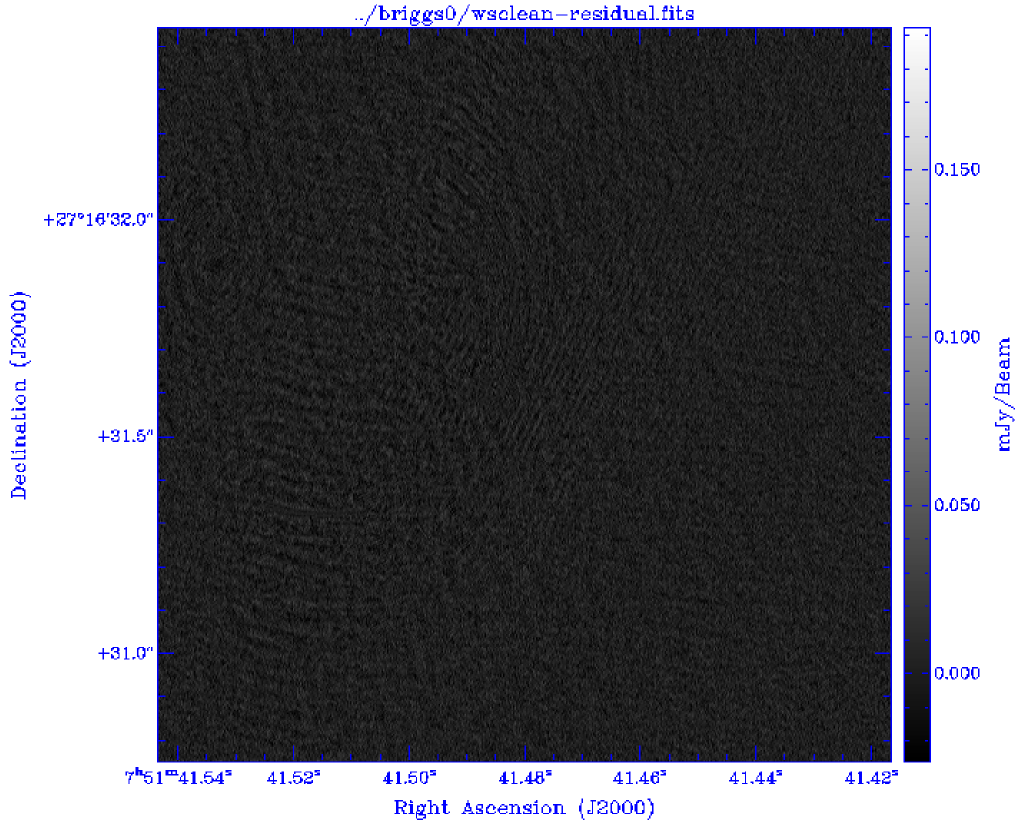
Auto-masking on point sources



Automasking VLBI example

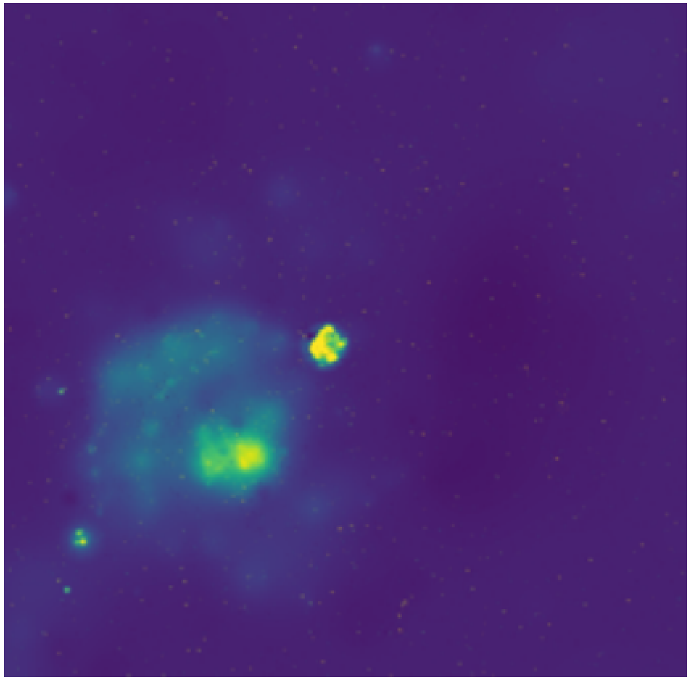


Restored

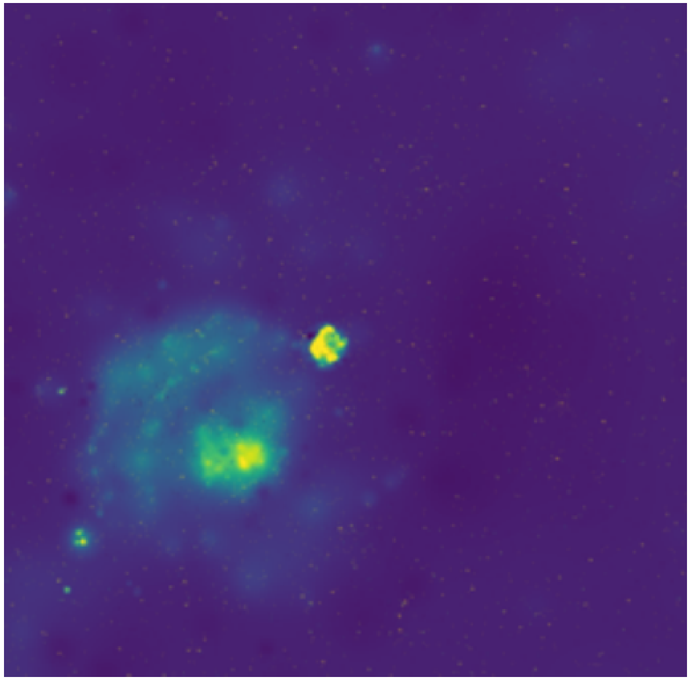


Residual

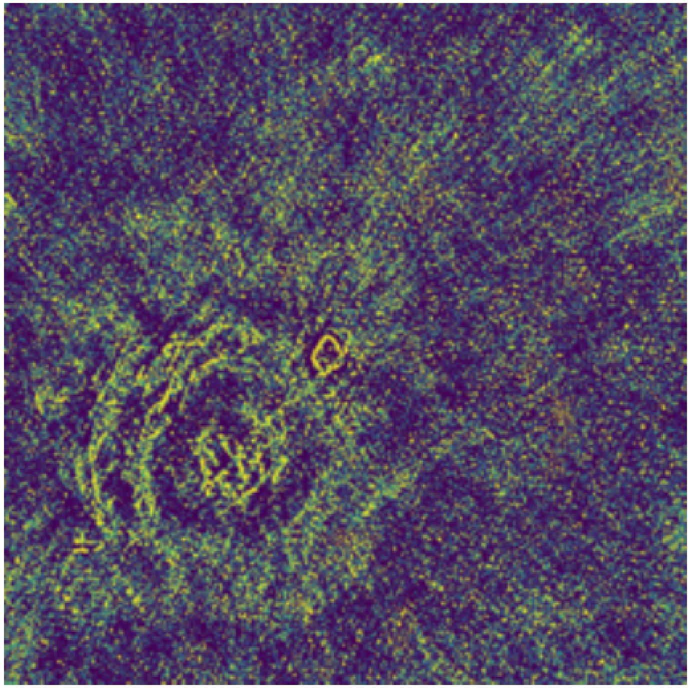
Data by J. P. McKean
and C. Spingola



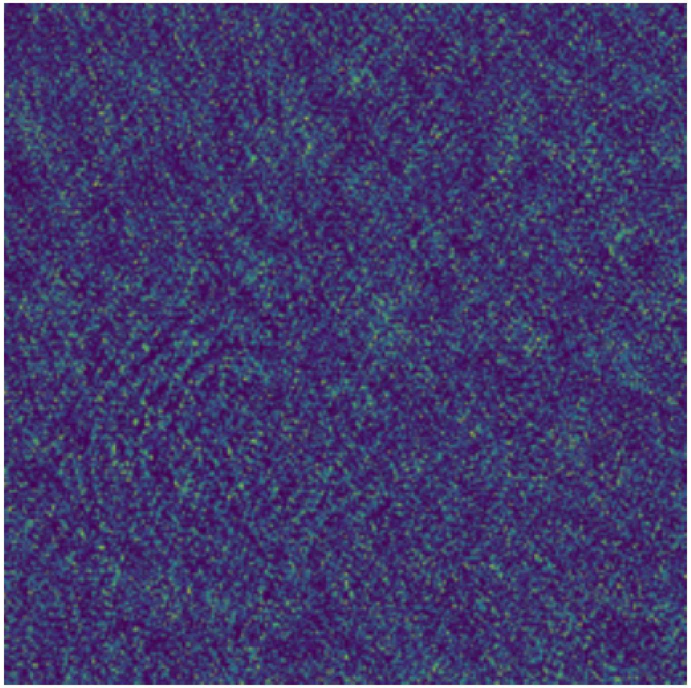
(a) Multi-scale model image without masking



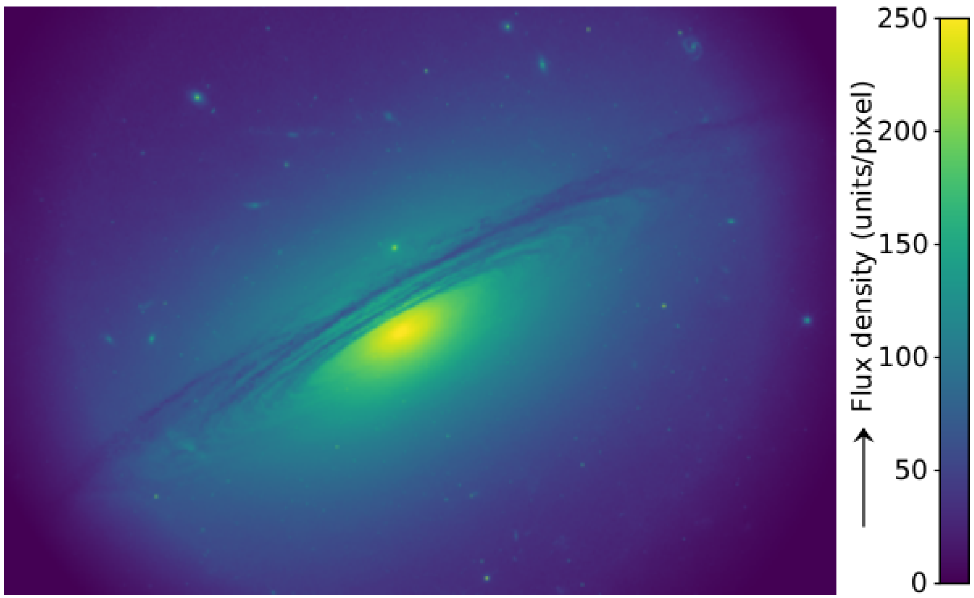
(b) Multi-scale model image with automatic masking



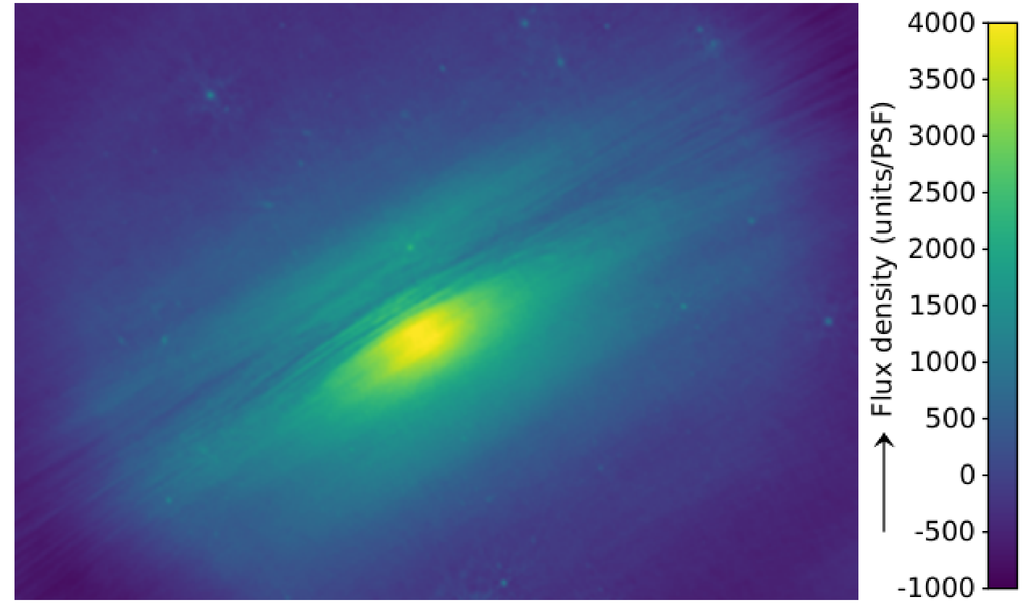
(c) Multi-scale residual without masking (rms=50 mJy/B)



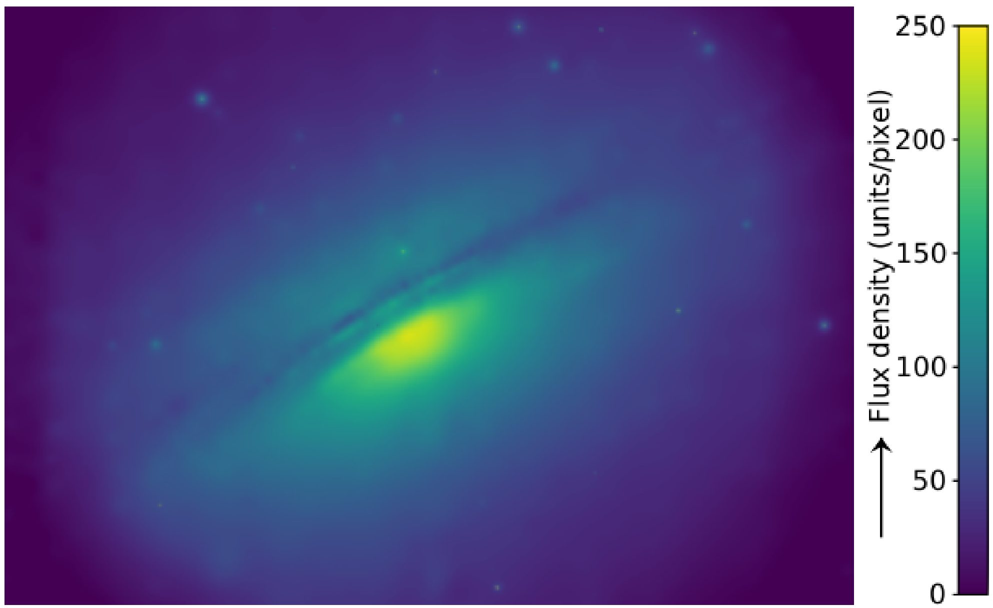
(d) Multi-scale residual with automatic masking (rms=38 mJy/B)



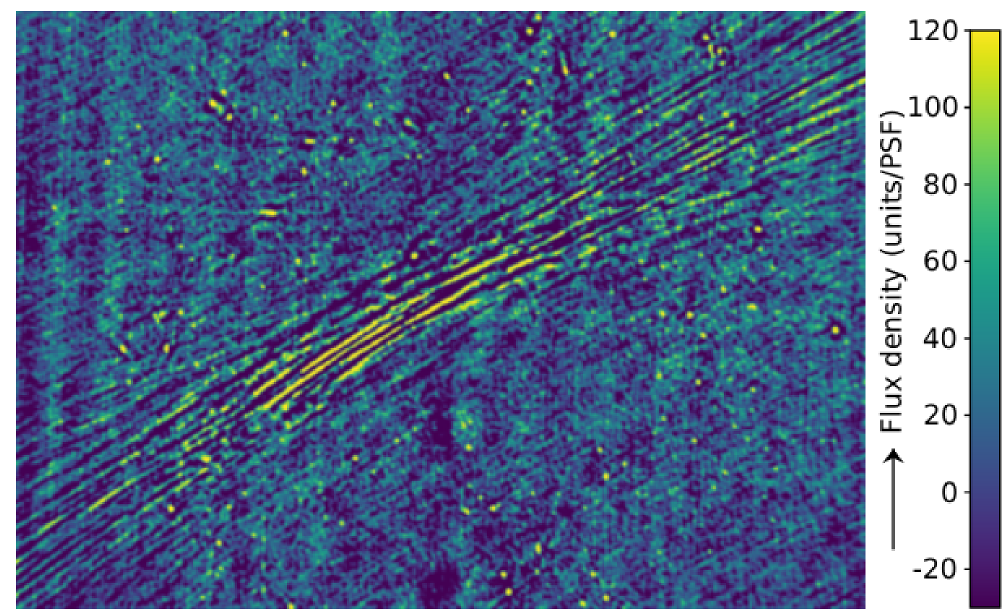
(a) Original



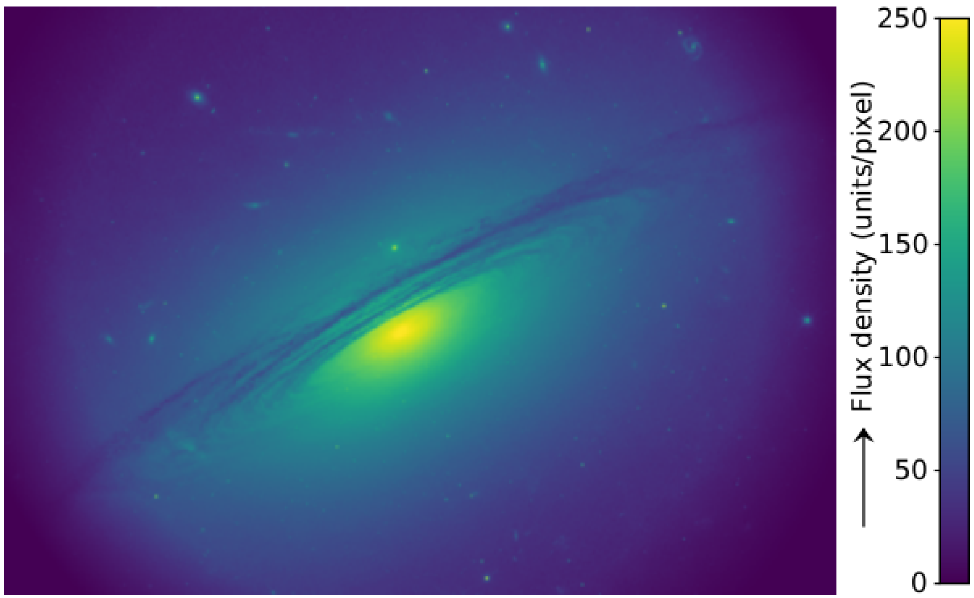
(b) Convolved image (σ=640,000 units/PSF)



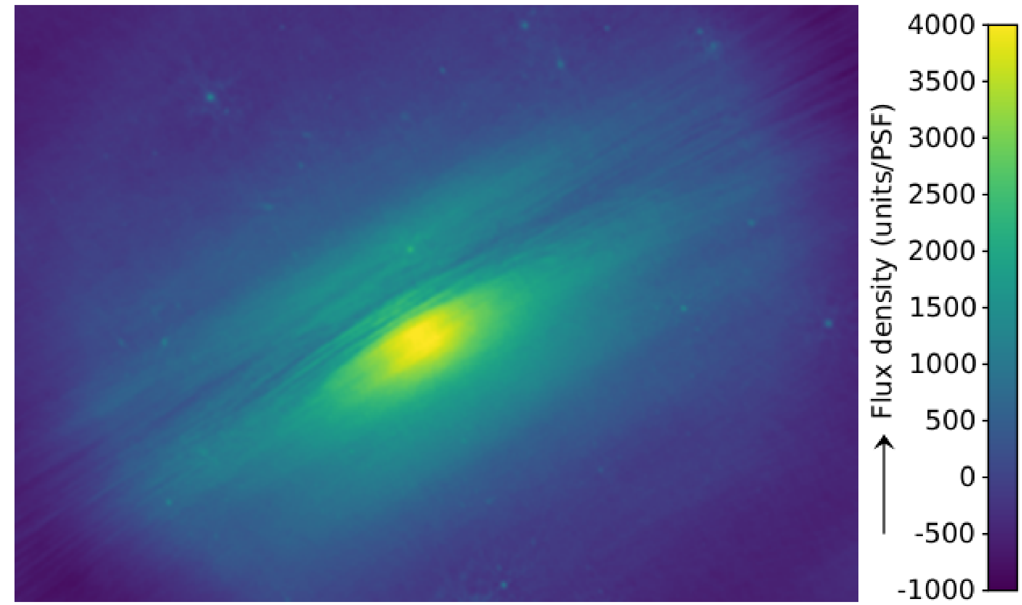
(c) CASA model



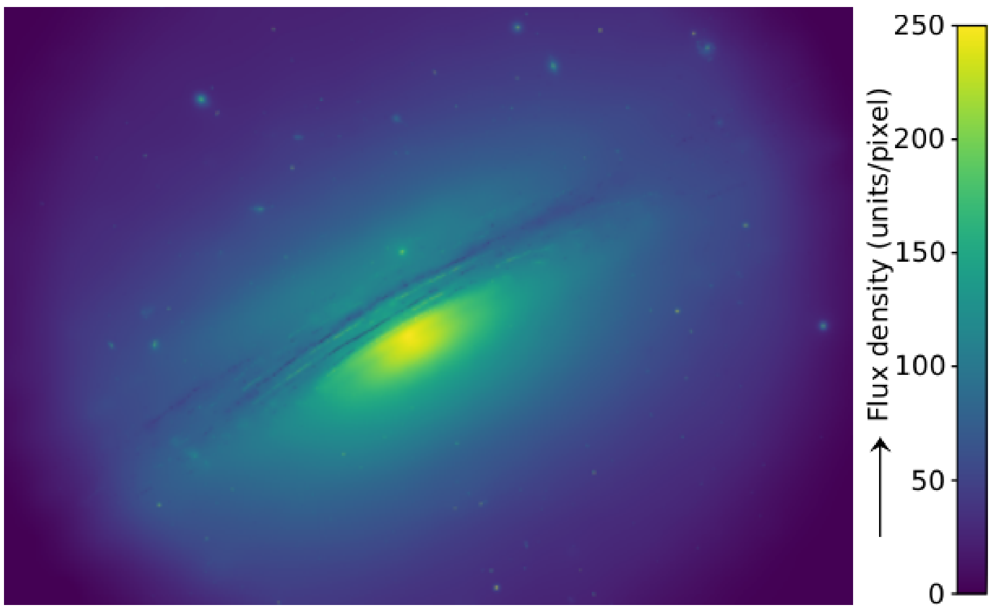
(d) CASA residual (σ=37 units/PSF)



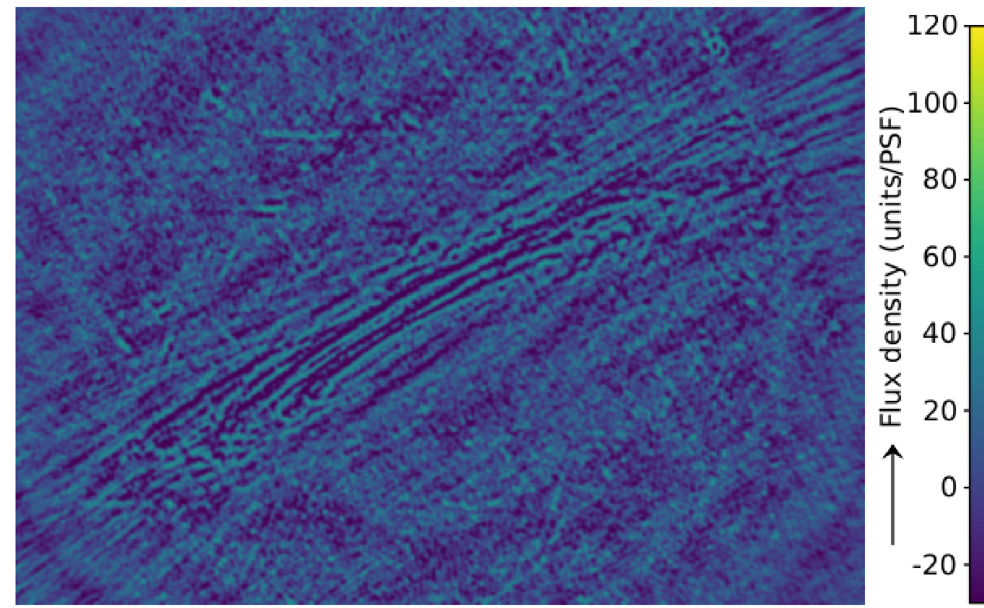
(a) Original



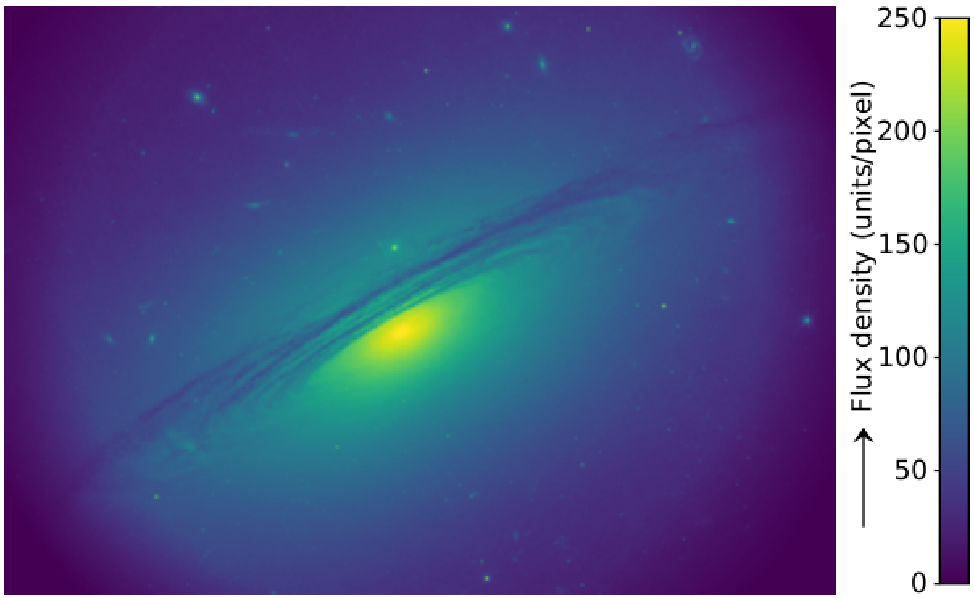
(b) Convolved image (σ=640,000 units/PSF)



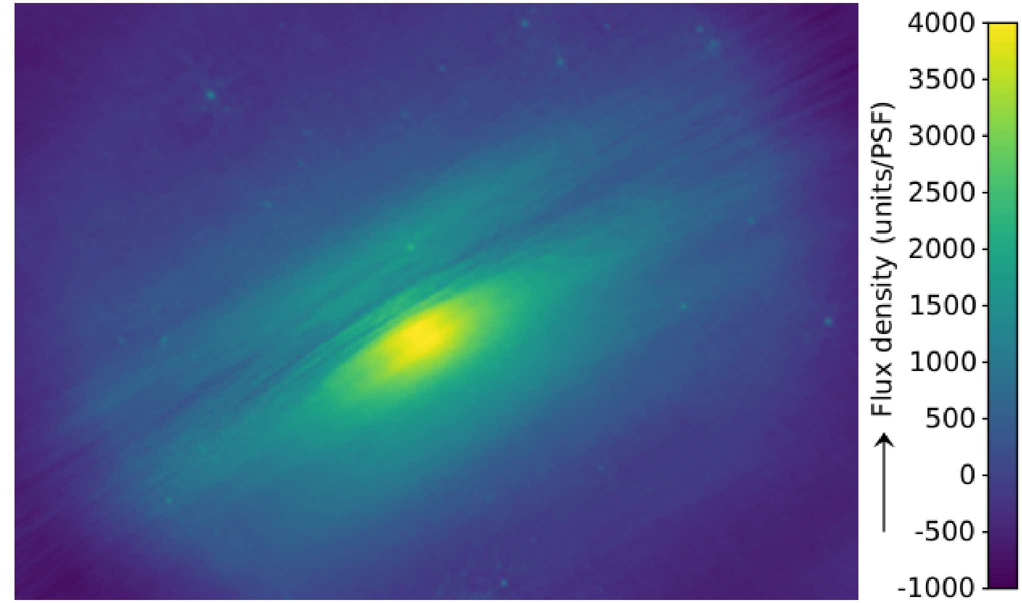
(e) WSCLEAN model



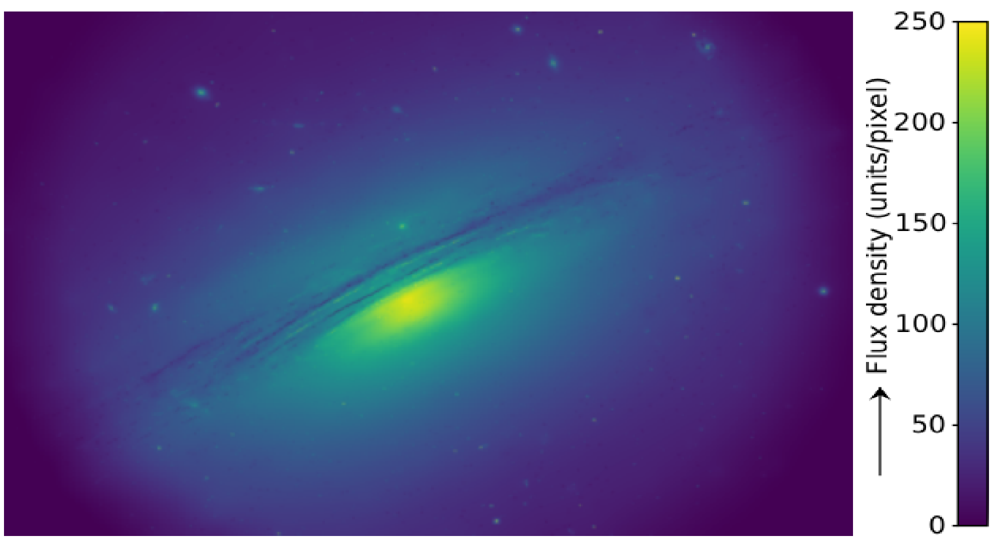
(f) WSCLEAN residual (σ=15 units/PSF)



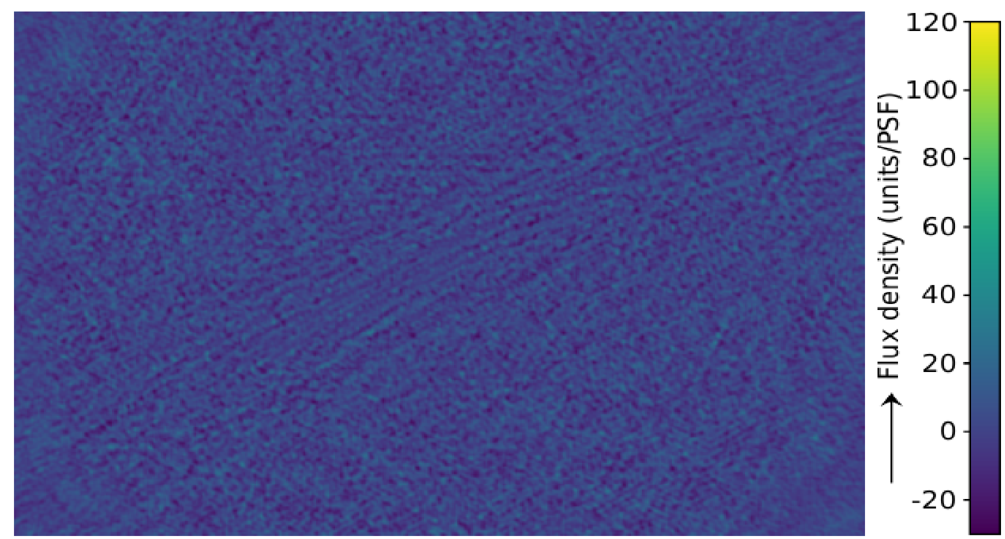
(a) Original



(b) Convolved image (σ=640,000 units/PSF)



(a) Reconstructed model image



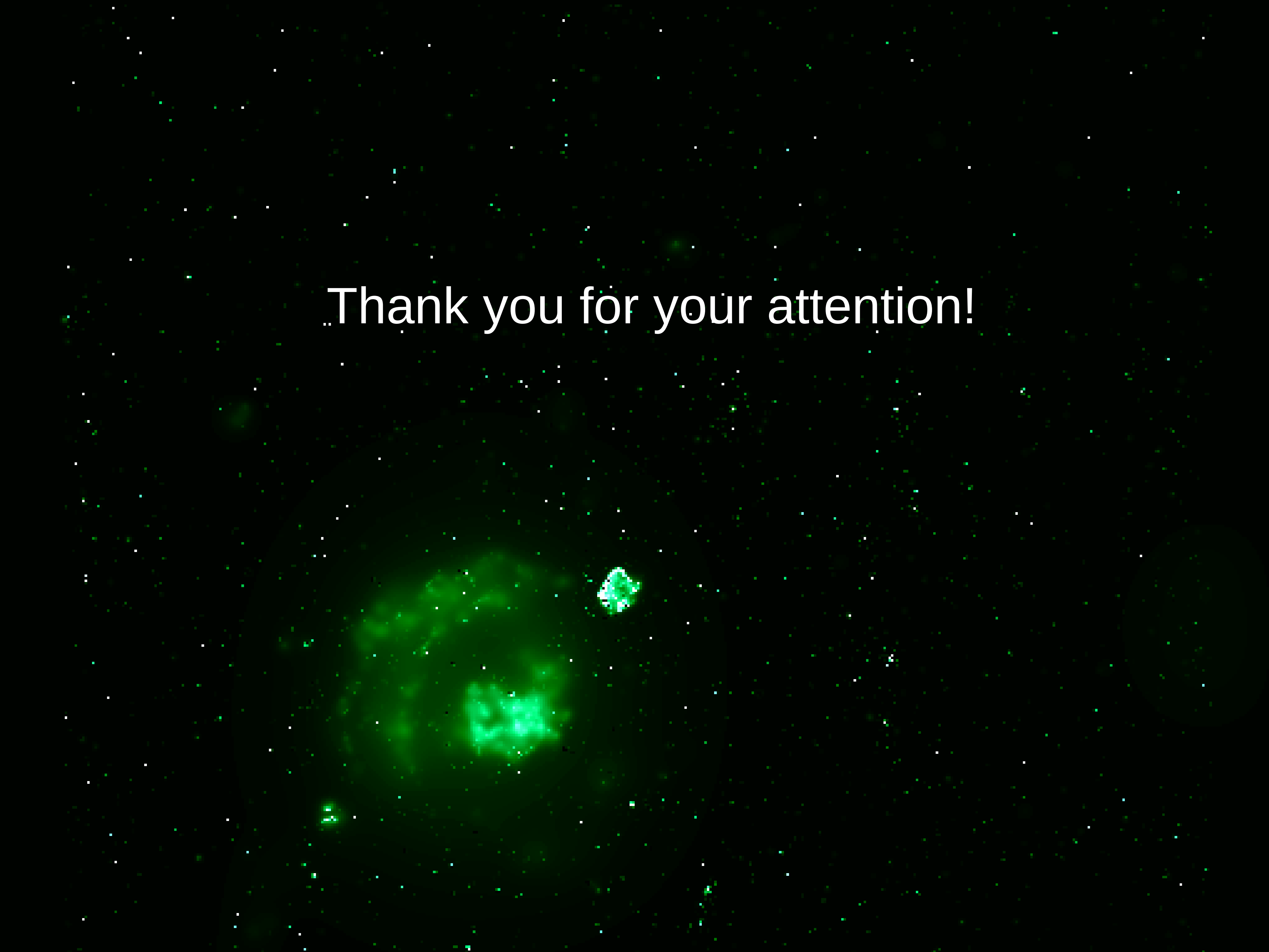
(b) Residual image (σ=8.6 units/PSF)

Figure 9. Automatic scale-dependent masking applied on the UGC12591 test-set.

Discussed topics:

- CLEAN
- When to use Multi-scale or other deconvolution methods
- The effect of and solution to w-terms
- Multi-term deconvolution
- Self-cal using CLEAN components
- Primary beam correction
- Mosaicing
- Direction-dependent effects during imaging

Summary


Thank you for your attention!