

Imaging and Deconvolution

Rick Perley, NRAO/Socorro



ATNF Radio School

Narrabri, NSW

29 Sept – 03 Oct 2014



Topics

- **Imaging**
 - **Formal Solution**
 - **Discrete Data**
 - **The Direct Transform**
 - **The Fast Fourier Transform**
 - **Weighting, Gridding, Aliasing, and more**
- **Deconvolution**
 - **'Filling the Holes'**
 - **Clean, MSClean**
 - **Max Ent**
 - **Compressed Sensing Methods**



The Formal Relations

- We learned, from my Lectures #1 and #2, that under some conditions (often true to good accuracy) the sky brightness $I_\nu(l,m)$ is related to the interferometer measurements $V(u,v)$ by:

$$V_\nu(u, v) = \iint I_\nu(l, m) e^{-i2\pi(ul+vm)} du dv$$

- The formal inversion of this equation is:

$$I_\nu(l, m) = \iint V_\nu(u, v) e^{i2\pi(ul+vm)} dl dm$$

- This is very simple (and very pretty), but cannot directly be used, since the functions concerned are analytic – known for all angles and distances.
- What we actually have are (often a large number of) discrete samples of the visibility, which we can write as $V_i(u_i, v_i)$.
- Thus, there is missing information, and this means we cannot guarantee a correct determination of $I(l,m)$.



The Sampling Function, and the Dirty Map

- We define the ‘Dirty Image’ by:

$$I_D(l, m) = \iint S(u, v) V_v(u, v) e^{i2\pi(ul+vm)} dl dm$$

where $S(u, v)$ is the ‘sampling function’, which describes the actual locations in the (u, v) plane where samples of the Visibility are taken.

- We can write this as:

$$S(u, v) = \sum_{k=1}^M \delta(u - u_k, v - v_k)$$

- Then, using a property of the Dirac delta function, we find

$$I_D(l, m) = \frac{1}{M} \sum_{k=1}^M V(u_k, v_k) e^{i2\pi(u_k l + v_k m)} .$$

- This function is often called the ‘Principle Solution’.

The Simplest Inversion: The Direct Transform

- The simplest procedure is to utilize ‘brute force’, and directly invert the equation through direct summation.
- Define a square grid, $N \times N$ cells in angle space (l, m) .
- For each cell, compute directly the sum:

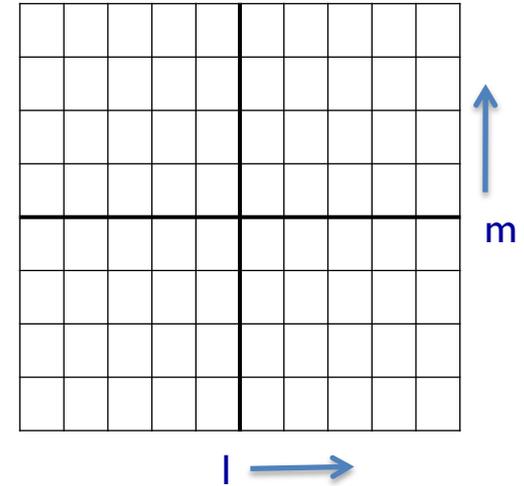
$$I_D(l, m) = \frac{1}{M} \sum_{k=1}^M V(u_k, v_k) e^{i2\pi(u_k l + v_k m)}$$

- In fact, this form is not quite right, as we have not utilized the Hermitian nature of the data.
- Because the sky is real, the visibilities are Hermitian, so that:

$$V(-u, -v) = V^*(u, v)$$

- If we write: $V_k = A_k e^{i\varphi_k}$
- Then we can write, for the ‘principle solution’

$$I_D(l, m) = \frac{1}{M} \sum_{k=1}^M A(u_k, v_k) \cos[2\pi(u_k l + v_k m) + \varphi_k]$$



Some Illustrations ...

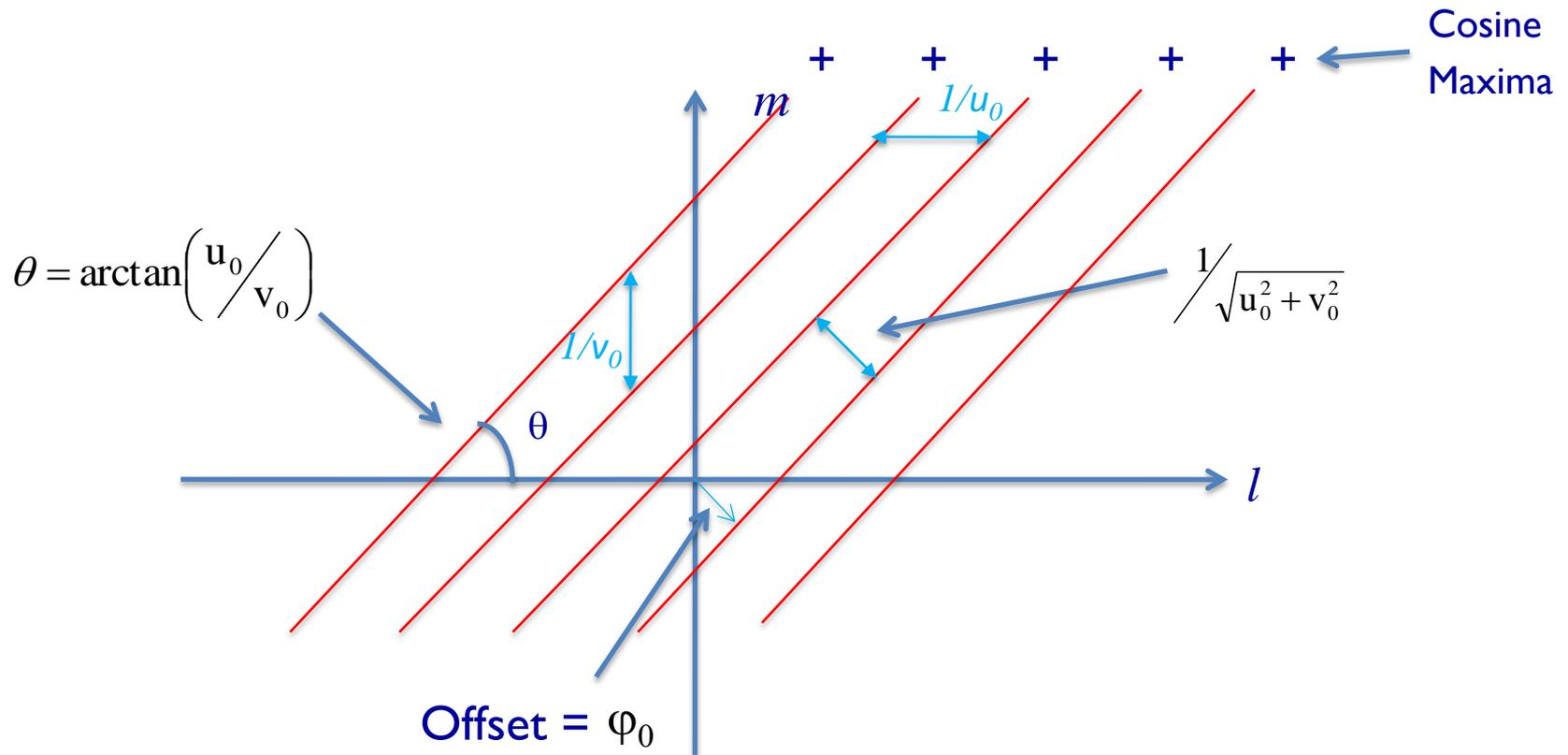
- To gain a better appreciation of what is happening here, consider a single observation with a single pair of antennas – i.e., one baseline.
- Then $S(u, v) = \delta(u - u_0, v - v_0)$
- The baseline formed by this pair has coordinates (u_0, v_0) .
- The visibility amplitude for the observation is A_0 , the phase is φ_0 .
- The ‘map’ from this single visibility is then:

$$I_D(l, m) = A_0 \cos[2\pi(u_0 l + v_0 m) + \varphi_0].$$

- This is a ‘cosinoidal sea’, of amplitude A and spatial period $\frac{1}{\sqrt{u_0^2 + v_0^2}}$ oriented at an angle $\theta = \arctan\left[\frac{u_0}{v_0}\right]$

and offset by angle φ_0 .

The Contribution of a Single Visibility.

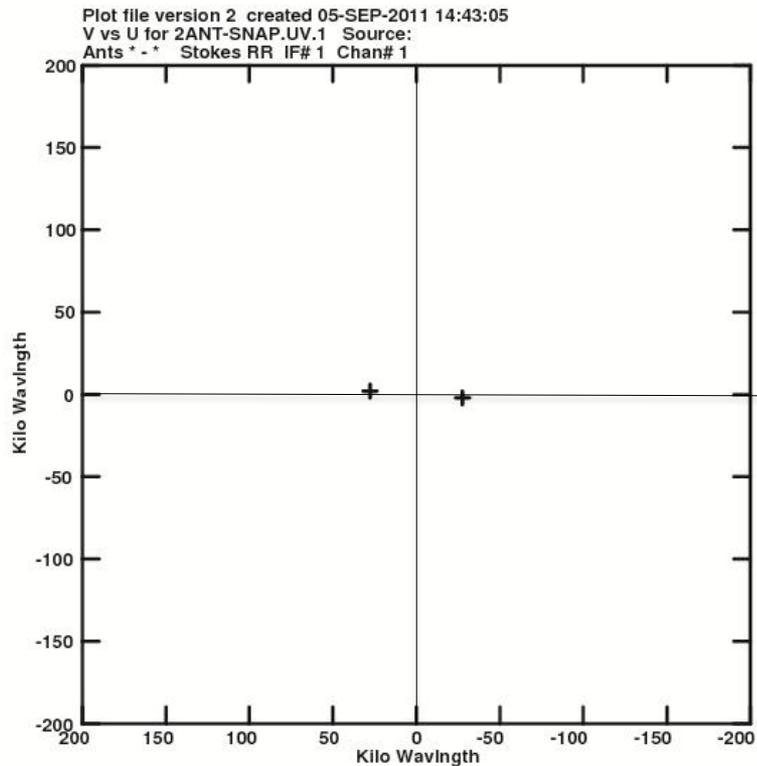


Every single visibility observation adds another cosinuoidal 'sea' with its own amplitude, orientation, spatial period, and offset.

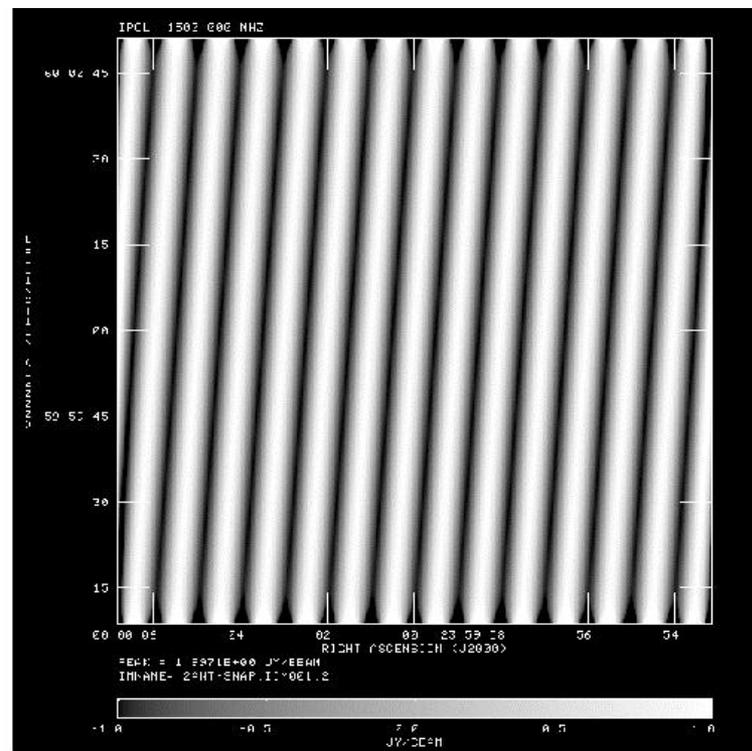
The map is the sum of all these cosinusoids.

Examples, using the VLA ...

Two antennas -- single baseline, single observation, 1 Jy point source.



UV-Coverage



Image

More Antennas = Better Images!

- A single observation of a point source.

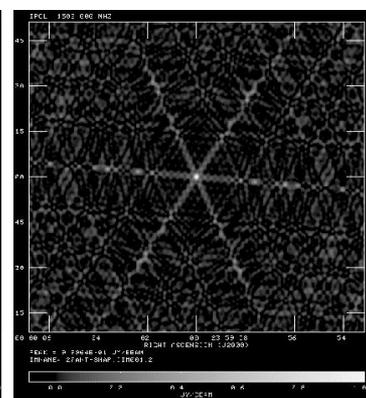
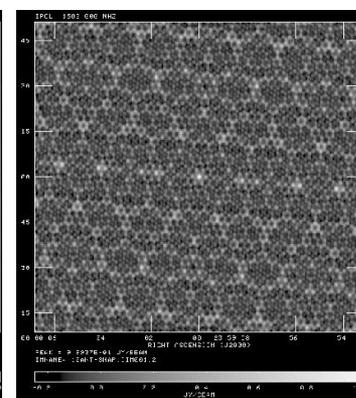
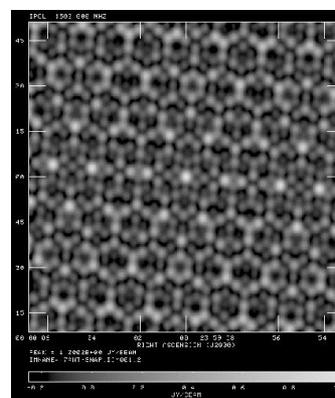
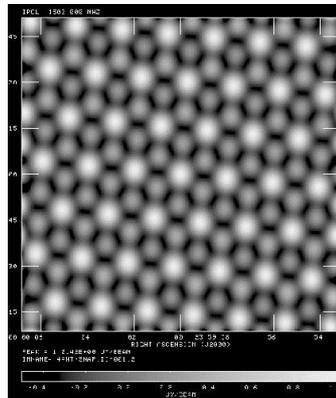
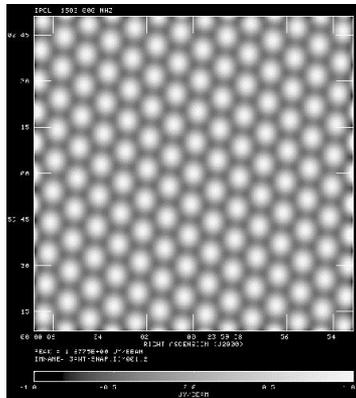
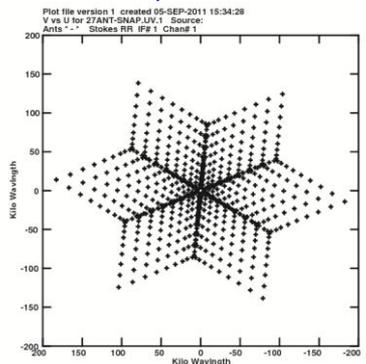
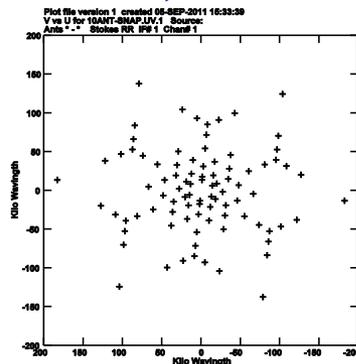
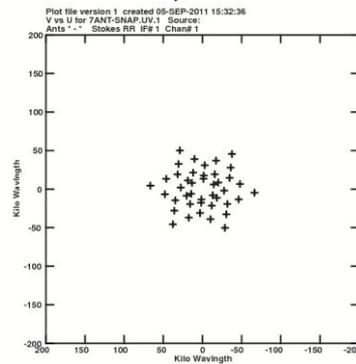
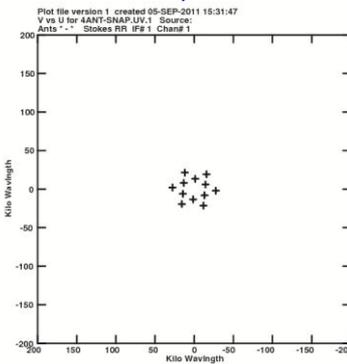
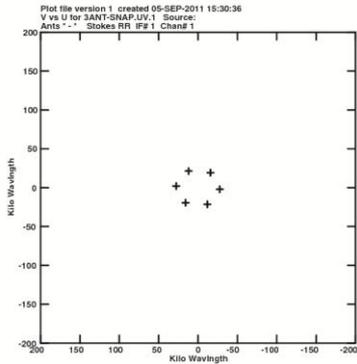
3 ant. 3baselines

4, 6

7, 21

10, 45

27, 351



100%
71%
-280%

100%
41%
-104%

93%
15%
-32%

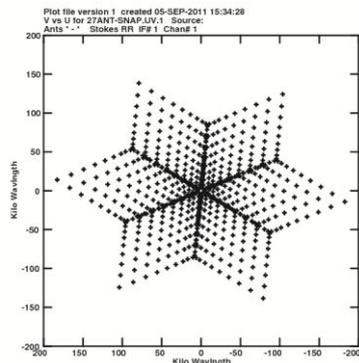
82%
11%
-23%

41% = Pk. sidelobe
4% = rms.
-10% = Pk. neg.

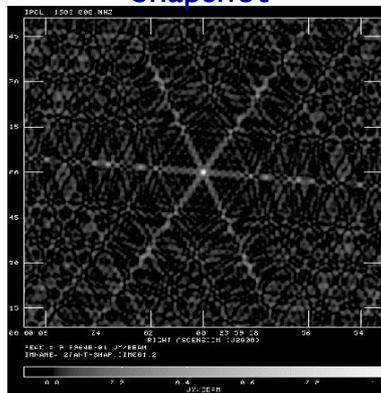


To do better – Observe over Time.

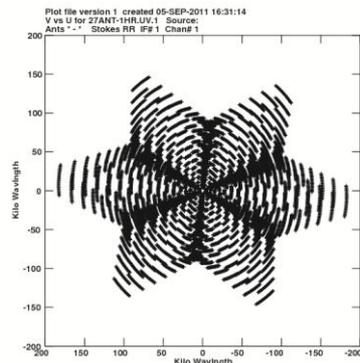
- UV Coverage and Images for a point-source for 0, 1, 6, and 12 hours' observing, with the 27-antenna VLA, at $\delta = 60$.



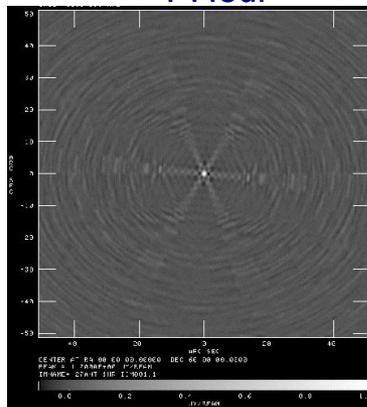
Snapshot



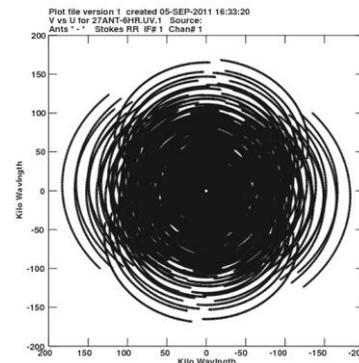
Pk Sidelobe 41%
rms 11%
Max. Neg. -10%



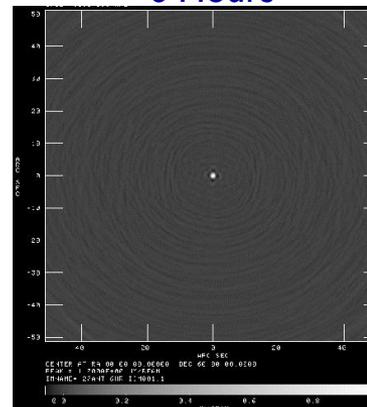
1 Hour



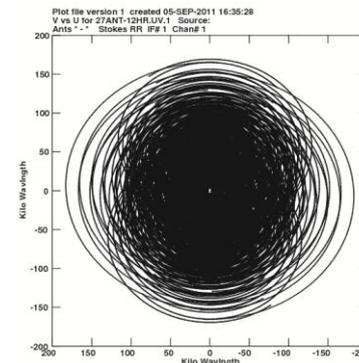
14%
4%
-10%



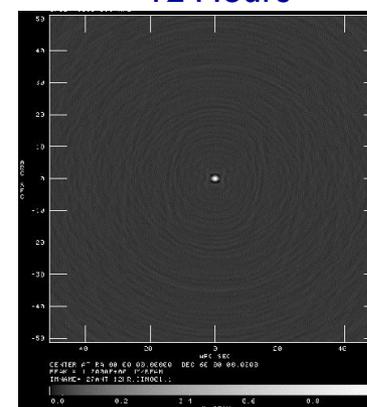
6 Hours



2%
0.5%
-5%



12 Hours

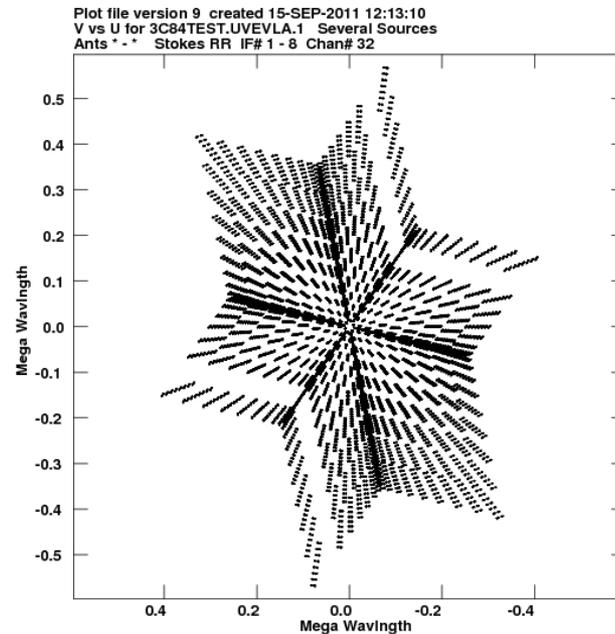
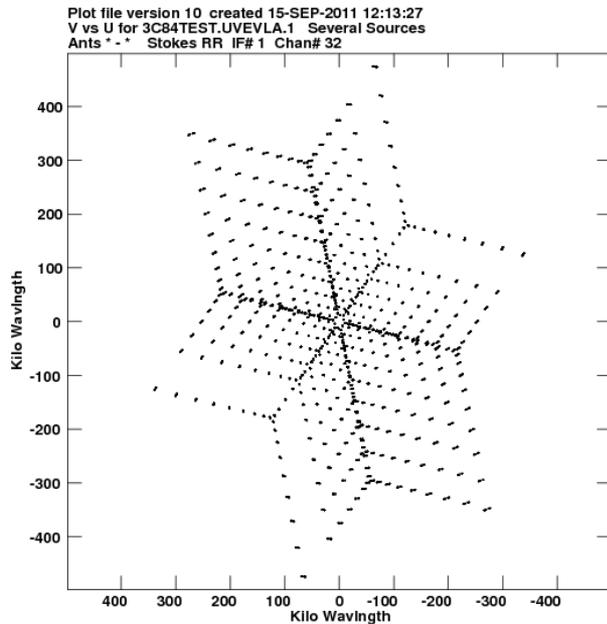


1%
0.3%
-4%



Multi-Frequency Imaging

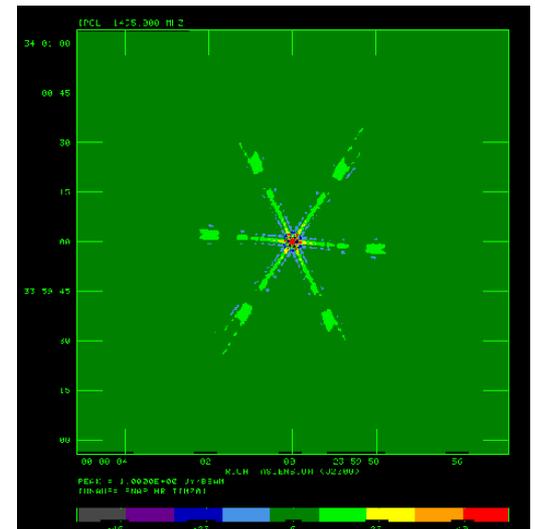
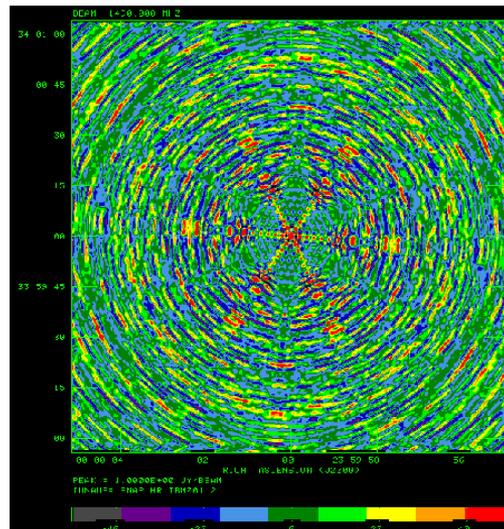
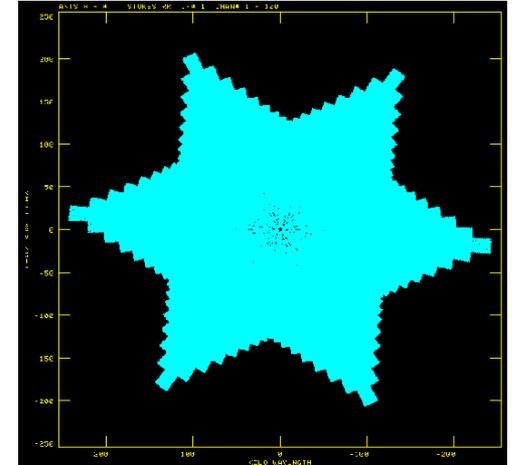
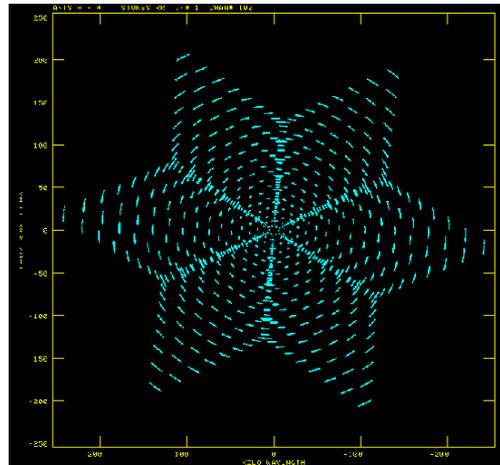
- One can dramatically improve the (u,v) coverage with a wide-band system – and a wide-bandwidth correlator.
- Examples: 1 frequency channel at 5.5 GHz, and 8 frequency channels, spread over 1 GHz (from 5.5 to 6.5 GHz) -- 12% spread.



Multi-Frequency Imaging – 2:1 BWR

- The use of wide-bandwidth, multi-frequency visibilities can dramatically improve the basic imaging characteristics.
- Shown is a 30 minute snapshot – left side is single frequency, right side is 2:1 wideband.
- RMS narrow: 1.4%
- RMS wide: 0.2%
- Peak sidelobes:

12.3% vs. 2.1%



But, Nothing Comes for Free ...

- The advantages in using wide bandwidths to improve (u,v) coverage are obvious.
- But, optimal usage of this extra information is not simple:
 - Extended sources have spectral gradients – often very steep.
 - This means the information in the visibility can change dramatically from one end of the band to the other.
 - You cannot simply invert all the data without accounting for this.
 - In addition – the antenna characteristics (primary beam) change over the frequency range – attenuating source brightnesses.
- More on this subject, later.

Problems with Sampling ...

- You will note that the (u,v) sampling by the interferometer is very non-uniform.
- There are large holes, and a large oversampling in the inner regions (short spacings).
- What we prefer is a generally uniform – but not regular! -- sampling, from the shortest spacings (zero) to the largest (λ/B_{\max}).
- For a single observation, certain array designs (equi-spaced linear, equi-spaced T, equi-spaced X) will provide a uniform (and regular) sampling.
- But even for these, the integration over time still causes a huge overweighting for short spacings.
- So some sort of averaging, or some sort of correction procedure will be needed to reduce the dominance of these short spacings.
- This is necessary to provide us a reasonable estimation of the true brightness.

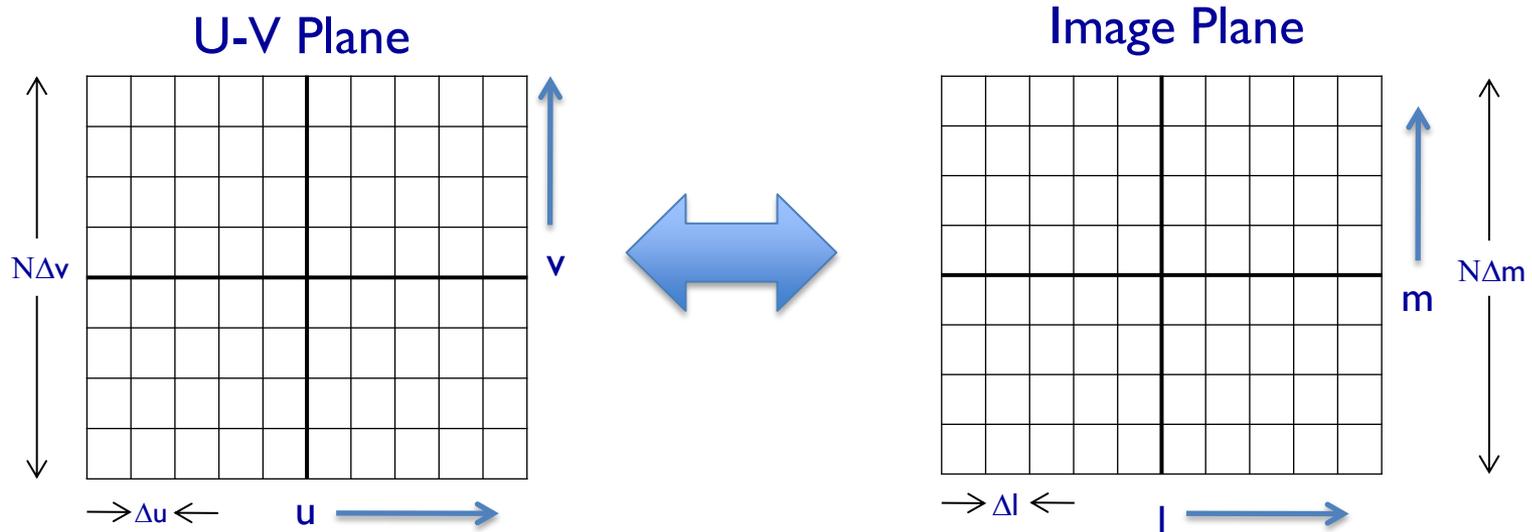


Big Data Sets – Why we need the FFT.

- The simple direct transform approach involves a lot of calculations.
- We have N^2 image cells, each of which requires a determination of the cosine factor, followed by a multiplication by the visibility amplitude, followed by a summation over all M measurements.
- We thus have $\sim 2MN^2$ real multiplications, and since $M \sim N^2$, the number of cosine evaluations, multiplications, and additions is $\sim N^4$!
- This can be managed with modern computers for modest values of N (say, ~ 100). But large-scale imaging requires $N > 10^3$, so another way is needed.
- Better Way: Use the magic of the FFT algorithm. The number of computations now scales as: $N^2 \log_2(N)$ – much better than N^4 !
- But – nothing comes for free -- the speed gained is because the data must be regularly arranged – which means we must ‘grid’ the uv-data onto a regular grid.

Two Grids: the (u,v), and (l,m).

- We now have two $N \times N$ grids – one for the (u,v) data, and the other – its Discrete Fourier Transform -- for the image.



- The cell sizes are set by the maximum baseline and field of view

$$\Delta u = \Delta v = 1/(2l_{\max}) \quad (2l_{\max} \text{ is the full field-of-view, in radians})$$

- $\Delta l = \Delta m = 1/u_{\max}$ (u_{\max} is the longest baseline, in wavelengths)

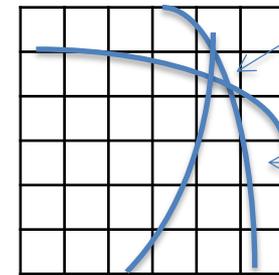
Example: Size of the Grids

- We can use these simple relations to establish some guidelines about image sizes.
- The largest field of view (size of the image) would normally be the angular separation of the antenna's first nulls. $l_{\max} = \lambda / D$
- From this, $\Delta u = 1/2l_{\max} = D/2\lambda$
- If the longest baseline is: $u_{\max} = B_{\max} / \lambda$
- Then, $\Delta l = \lambda / 2B_{\max}$
- The size of the grids comes from the relation: $N = 2u_{\max} / \Delta u = 4B/D$
- For the VLA at 6 GHz ($\lambda=5\text{cm}$) $D = 25$ m, so: $\Delta u = 500$ (wavelengths)
- And in the A-configuration, $B_{\max} = 35$ km, so $\Delta l = 1.4 \times 10^{-6}$ (radians)
- And for the size of the grid: $N = 5600$ (cells).
- Note that the image size (in cells) is independent of wavelength!



Advantages and Disadvantages of Gridding

- Use of the FFT is necessary for large databases
- Thus, we must grid the data onto a regular grid.
- All data within any one cell will be averaged in some way, and the value placed at the cell center.
- There are some consequences:
 - This introduces errors: The visibilities are ‘moved’ to a (u,v) point where they were not measured.
 - The regularity of the gridded data causes aliasing.
- The first issue is not critical if the number of data points is large (lots of points within a cell – the mean will be ‘about right’).
- The second issue is definitely a problem, but can be handled by suitable anti-aliasing convolution functions.
- What is ‘aliasing’, and how can we control it?

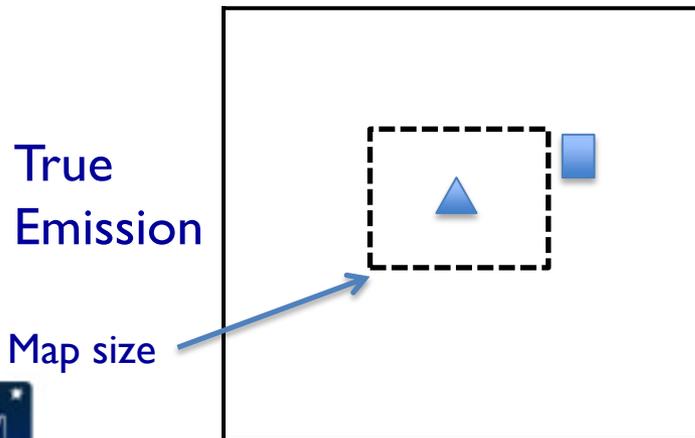


This cell is fairly well sampled

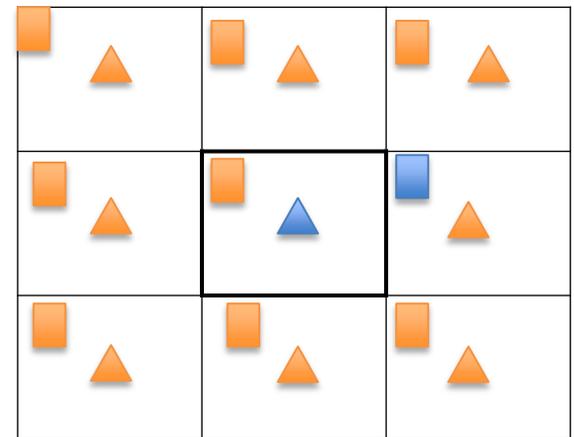
This cell's data are along one side.

Aliasing

- Aliasing is a phenomenon which occurs when a Fourier sum is made over data which are regularly spaced.
- In this case, there is no way to discriminate between a phase change of 0 , 2π , 4π , ... per cell.
- When we grid data, we are generating a regularity which enables emission from outside the image to appear in it!
- From a Fourier series point of view, the transform of a regularly gridded function is itself a regularly gridded function – in the case, the image is itself replicated indefinitely in all directions.



After
FFT

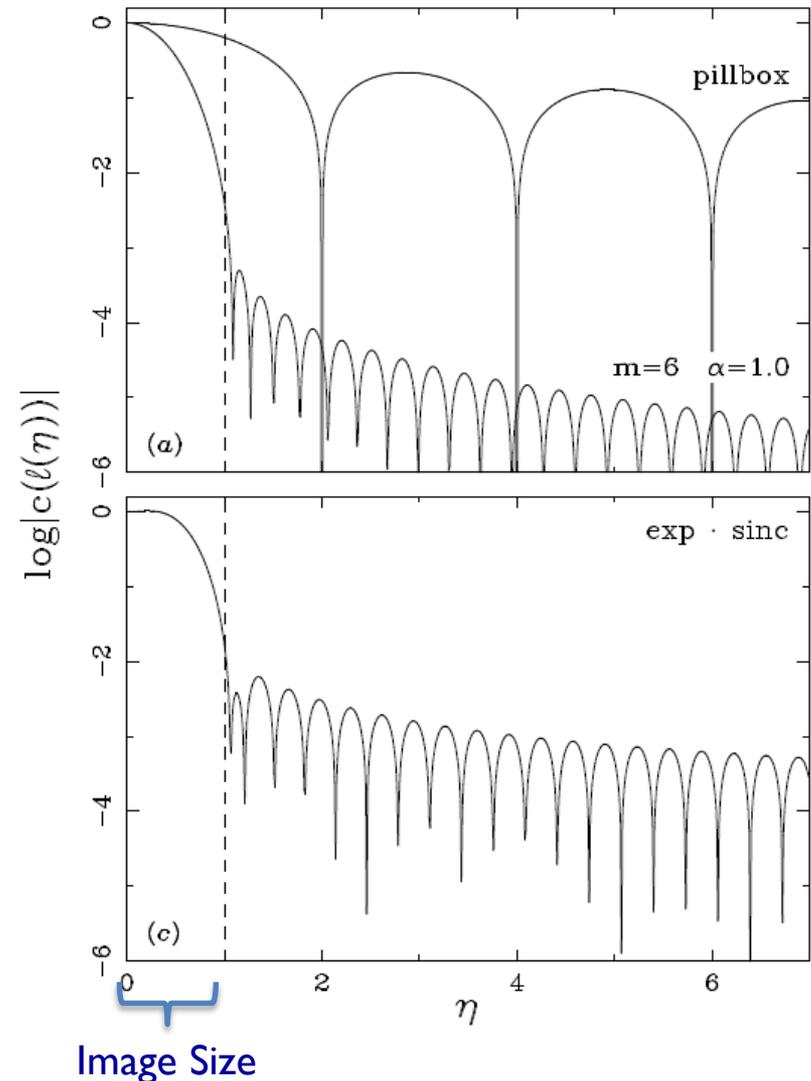


Managing Aliasing

- In general, we have to use the FFT – thus we are vulnerable to aliasing.
- How to manage?
 - Make a bigger image! Big enough that all real emission lies within the transform area.
 - But this rather defeats the purpose of the FFT, especially if the outside emission is far away.
- A generally better method is to utilize the data within the cell and the adjacent cells.
 - The information to discriminate lies within the data within these cells – far away emission has a strong phase gradient within each cell.
 - With enough data, and with a averaging (smoothing) function, out-of-field emission can be very strongly suppressed.
 - Some sophisticated gridding functions are utilized, which suppress the aliasing very well.

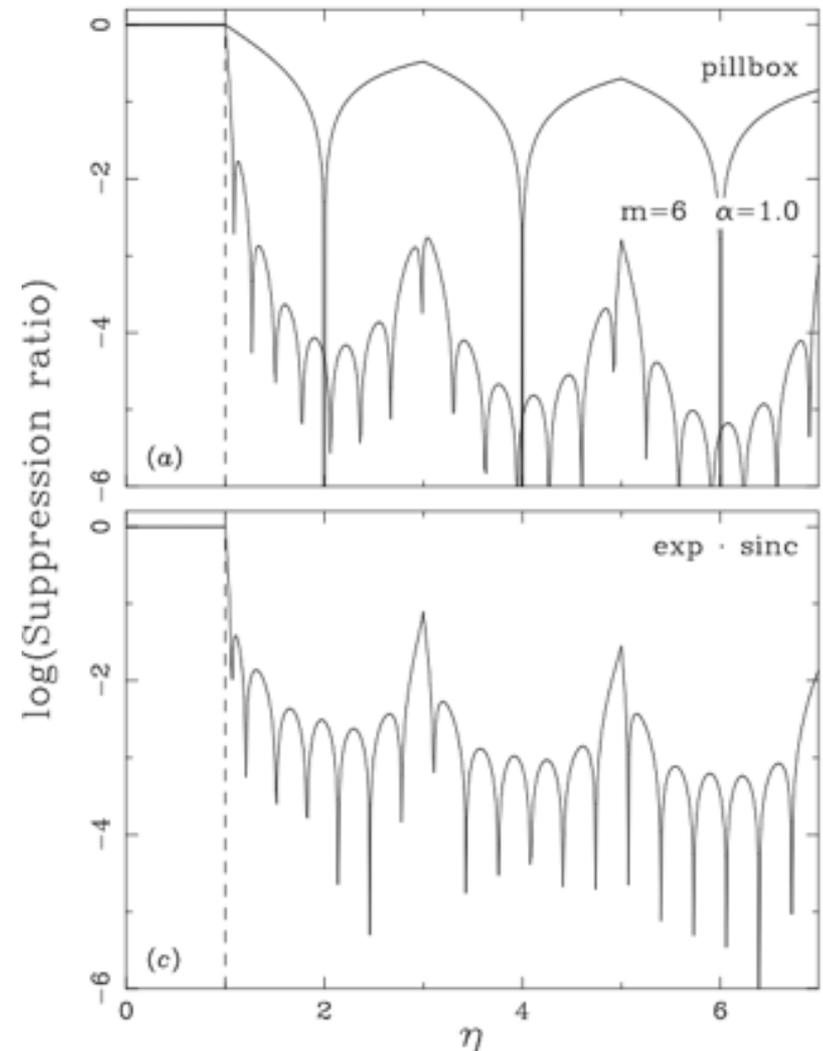
Various Antialiasing Functions

- Standard Imaging packages have numerous convolution functions available.
- Shown here are three which are available in AIPS:
 - Pillbox (1 cell simple average)
 - Spheroidal
 - Exponential * $\sin(x)/x$
- X-axis is in units of the half map width
- Y-axis is the log(suppression).
- Note: these functions are correct only when data are uniformly distributed.



Gridding Correction

- The anti-aliasing functions partially suppress emission within the central field.
- This suppression must be corrected for, using the known functions.
- The result of this is shown to the right.
- Note that the correction factors are valid for a uniform distribution within the convolving window.
- The actual distribution will be different, so there will be errors in the map amplitudes, especially towards the edges.



More on u-v Plane Weighting

- We can do more than apply anti-aliasing functions when preparing the data for interpolation to the regular (u,v) grid:
 - **Tapering:** The sharp cut-off in sampling (due to the longest baseline), and the sparseness of the sampling at the longest baselines both lead to ‘ringing’ in the dirty image.
 - This can be reduced by adjusting the weighting of the data (reducing the amplitude) as a function of the radius. A gaussian weighting, with unity at the origin, and ~30% amplitude at the longest baseline, is common.
 - Tapering reduces the sidelobes, but degrades the resolution.
 - **Natural vs. Uniform Weighting:** Ideally, the amplitude of the visibility gridded in each cell should be the mean of the data within that cell. This is termed ‘Uniform Weight’.
 - But since there are generally far more points in the inner parts of the (u,v) plane than in the outer, doing this simple averaging can considerably reduce sensitivity. Using the sum maximizes sensitivity. This is ‘Natural Weighting’



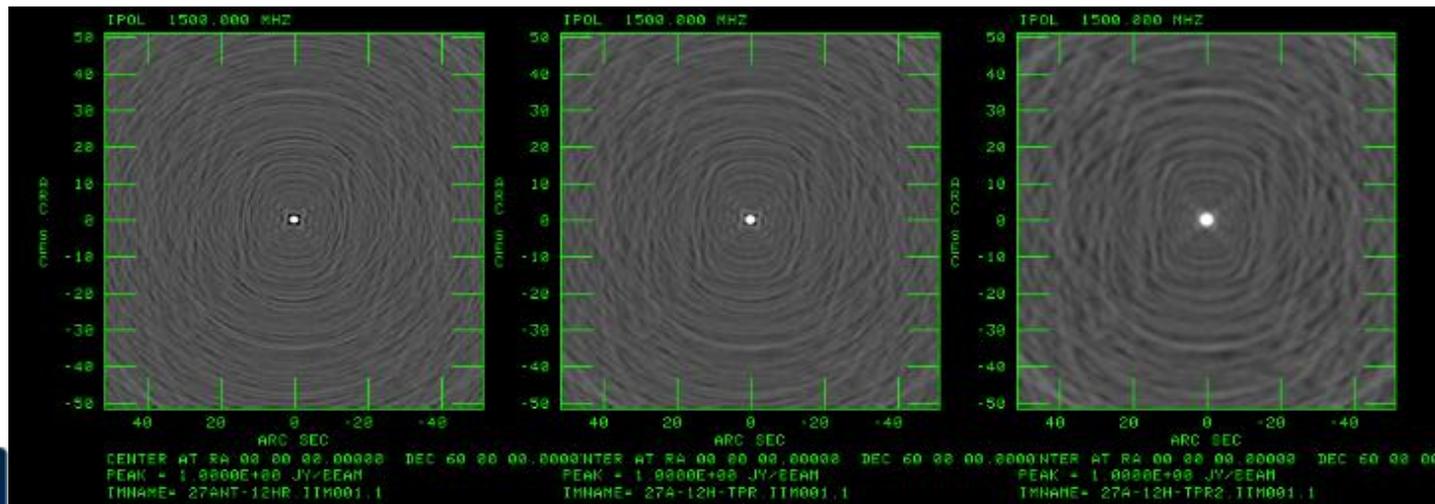
'Natural' vs. 'Uniform' Weighting

- Two extremes can be identified:
 - 'Natural' Weighting: The sum of all values within the cell is utilized. This maximizes sensitivity, but greatly degrades the resolution.
 - 'Uniform' Weighting: The average of all values within the cell is utilized. This is closest to what we really want – but loses sensitivity.
 - Other schemes are possible. Dan Briggs developed a weighting scheme which allows a continuous gradation between fully 'uniform' and fully 'natural'.
 - The 'right' weighting is a matter of experience, goals (and prejudice).
- Some examples should help clarify this.

Tapering

- Besides the convolution/gridding operations, we can also impose an overall taper to the (u,v) plane, in order to generate a Dirty Beam with more desirable characteristics.
- In most imaging packages, tapering is by a circular Gaussian, specified to some standard attenuation value at a given radial distance.

Taper	None	30% @ 160 k λ	30% at 80 k λ
Beam	1.25 x 1.09	1.40 x 1.26	1.85 x 1.74
Pk Neg.	-3.9%	-1.7%	-1.1%
rms	0.29%	0.27%	0.31%



Robust Weighting Functions.

- There are more options than simply ‘natural’ and ‘uniform’.
- Dan Briggs generated a simple, single valued weighting system – applied as part of the gridding operations – which allow a kind of ‘hand tailoring’ to balance the gain of sensitivity against the loss of resolution.
- Available in both AIPS and CASA (and presumably in other imaging packages).
- When should the user use ‘natural’ weighting?
 - Only for detection of faint, small objects, when fidelity of the imaging is not a primary concern.
 - There is a loss of ~20% in sensitivity when using ‘uniform’ weighting for VLA data.
 - The more centrally concentrated the array (as in some SKA designs), the more severe is the loss in sensitivity when uniform weighting is used.



Deconvolution

- ‘Deconvolution’ is the process of removing artifacts caused by the irregular and/or incomplete sampling in the (u,v) plane.
- The ‘dirty’ image provided by the Fourier inversion, $I_D(l,m)$ is related to the correct image $I(l,m)$ by a convolution relation:

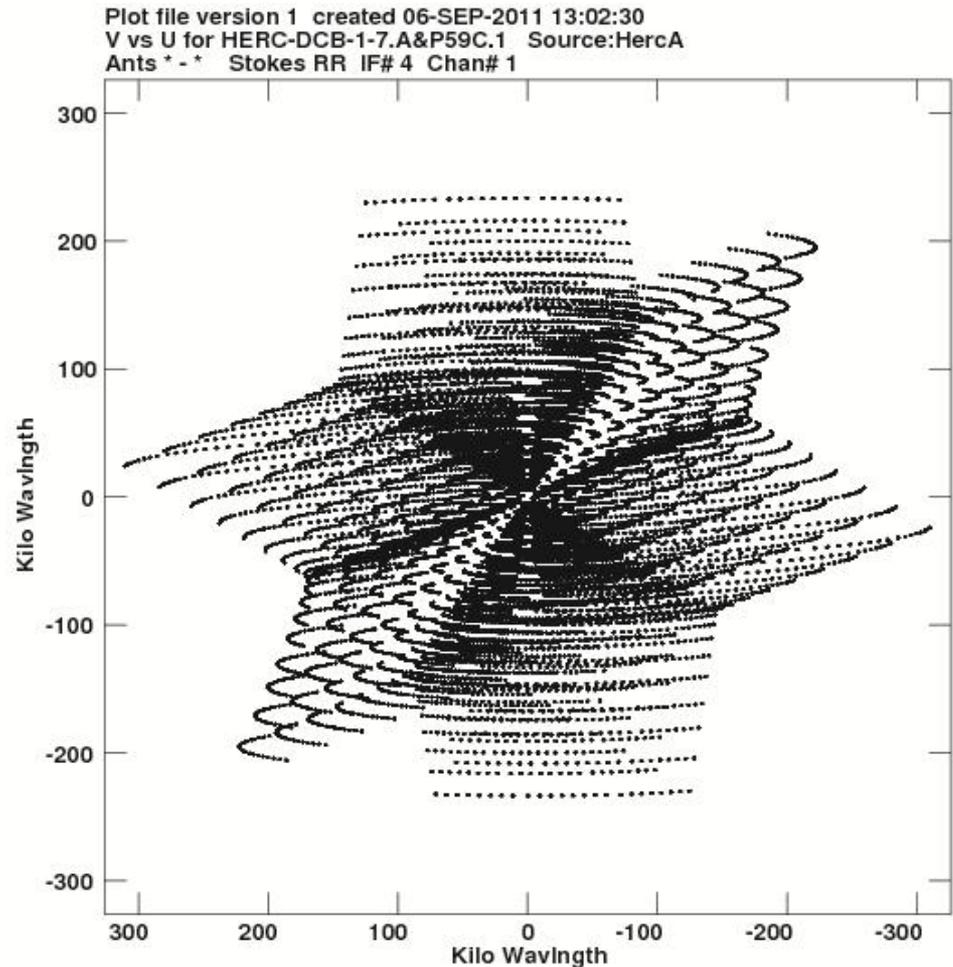
$$I_D(l,m) = I(l,m) * B_D(l,m)$$

where $B_D(l,m)$ is the ‘dirty beam’ -- the F.T. of $S(u,v)$.

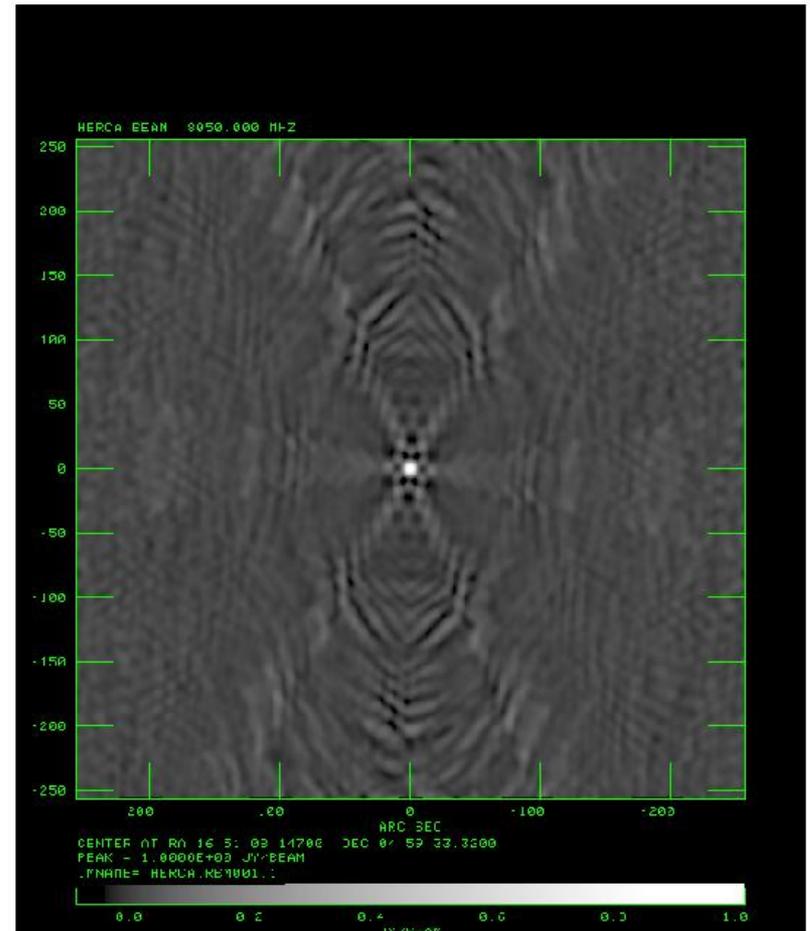
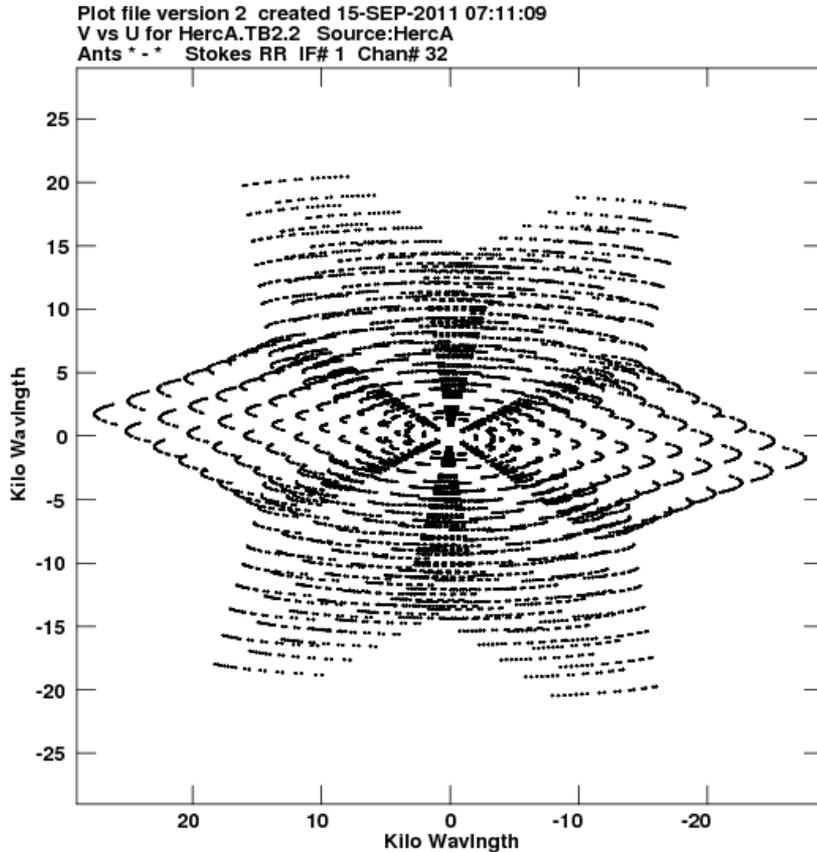
- The ‘dirty’ map and ‘dirty’ beam are called this because they look terrible.
- This is caused by a critical lack of information – there are many ‘cells’ in the (u,v) plane for which we have no measurements!
- In principle, these ‘holes’ could have any values at all.
- However, the true sky has characteristics which we can use to assist us in ‘guessing’ the values in the missing cells.
- Thus, deconvolution is the process of estimating the visibilities for the locations where we have no observations.

Typical U-V Coverage, and the Gaps.

- Shown is a short (4 hours') uv-coverage with the VLA of a low-declination source (Hercules A) at 3.6cm.
- Key features:
 - No data beyond 300 $K\lambda$
 - A (small) central hole.
 - Many large gaps!
- The gaps are much larger than the inverse source size.
- Without additional information, there is no way to fill in these 'gaps'.
- What information can help?



A Typical Dirty Beam and its Coverage



Deconvolution: Constraints we can Use...

- Because the ‘holes’ in the UV coverage are generally much larger than the inverse image size, the missing information cannot – in general -- be interpolated from nearby data without extra information.
- The deconvolution programs can use all the help we can provide.
- Some useful constraints include:
 1. Finite source sizes: If we know (or think we know) the regions from which emission arises, deconvolution proceeds much better.
 2. Knowledge of the shape/size/pattern of the dirty beam helps in identifying likely regions of emission.
 3. Mostly empty field: In many cases, the fraction of the pixels with emission is small.
 4. All positive emission: Useful for Stokes I. (Cannot be used for Q, U, or V).

Major Forms of Deconvolution

- There are two well-known algorithms used in radio interferometry, and a number of lesser known ones under development.
- There are two major classes of deconvolution algorithms (that I know of).
 1. CLEAN: An iterative procedure which subtracts the PSF (Dirty Beam) from the image.
 2. MEM: The 'Maximum Entropy Method' algorithm which finds an all-positive image with maximum smoothness whose transform matches the original data.

CLEAN

- The 'CLEAN' algorithm was developed by Hogbom in 1974. Various improvements (mostly for speed) have been developed since then.
- CLEAN remains the most commonly used procedure.
- The procedure is:
 1. Locate the highest point of emission in the image, and subtract from the entire image the dirty beam (psf), scaled by some 'loop gain' γ . Remember the location and peak brightness removed.
 2. Repeat step 1 until some level of noise is reached. This image is the 'residual'
 3. Rebuild the image, using the list of locations and peaks, using a 'clean beam' in place of the psf.
 4. Add the residuals back.
- In essence, CLEAN imagines the emission to be a set of unresolved or partially resolved, well separated objects.

Despite its simplicity, CLEAN works remarkably well.

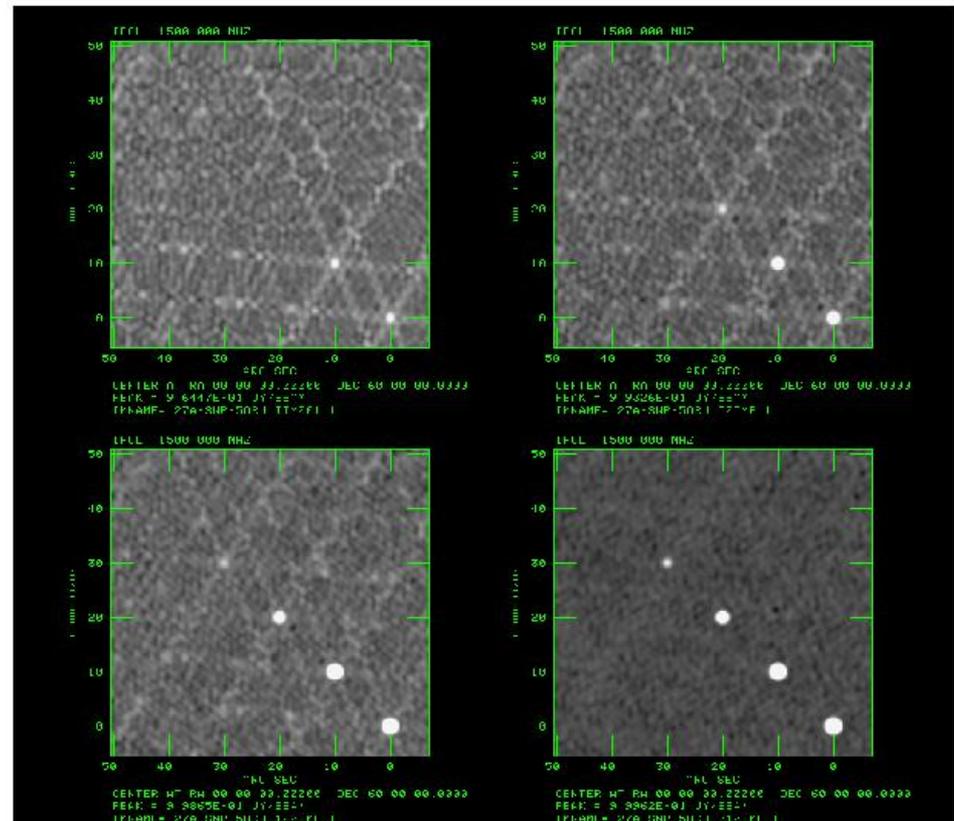


CLEAN – some additional points.

- There are a number of variants of 'CLEAN': the 'Clark' and 'Cotton-Schwab' algorithms are variants which permit much faster searching and subtracting. Multi-Resolution Clean helps on extended objects.
- CLEAN is naturally best with compact emission.
- CLEAN often does a poor job on low-brightness sources – the typical signature being 'rippling' in the restored image. The procedure can diverge in these situations, especially when the (u,v) coverage is poor.
- CLEAN is very slow when deconvolving large extended sources.
- CLEAN is not effective at recovering the total (zero-spacing) flux, nor at reliably providing 'super-resolution'.
- There is a poor theoretical understanding of CLEAN.
- The use of 'Clean boxes' is highly recommended.
- The interaction between the key parameters: loop gain, the number of iterations, and the location of boxes, to provide the 'best' image, is a matter of experience.

Examples of CLEAN at work.

- Shown are partial deconvolutions of a field with 5 sources, of strengths 1.0, 1.0, 0.1, 0.01, and 0.001 Jy. Loop factor = 20%.
- Simulated snapshot with VLA, with suitable noise added.
- Top Left: Dirty Image
 - $\sigma = 53$ mJy
- Top Right: After 50 comp.
 - $\sigma = 5.7$ mJy
- Bottom Left: After 150 comp.
 - $\sigma = 0.74$ mJy
- Bottom Right: Done (310 comp).
 - $\sigma = 0.19$ mJy
- Note: the weakest source is not visible – below the noise.

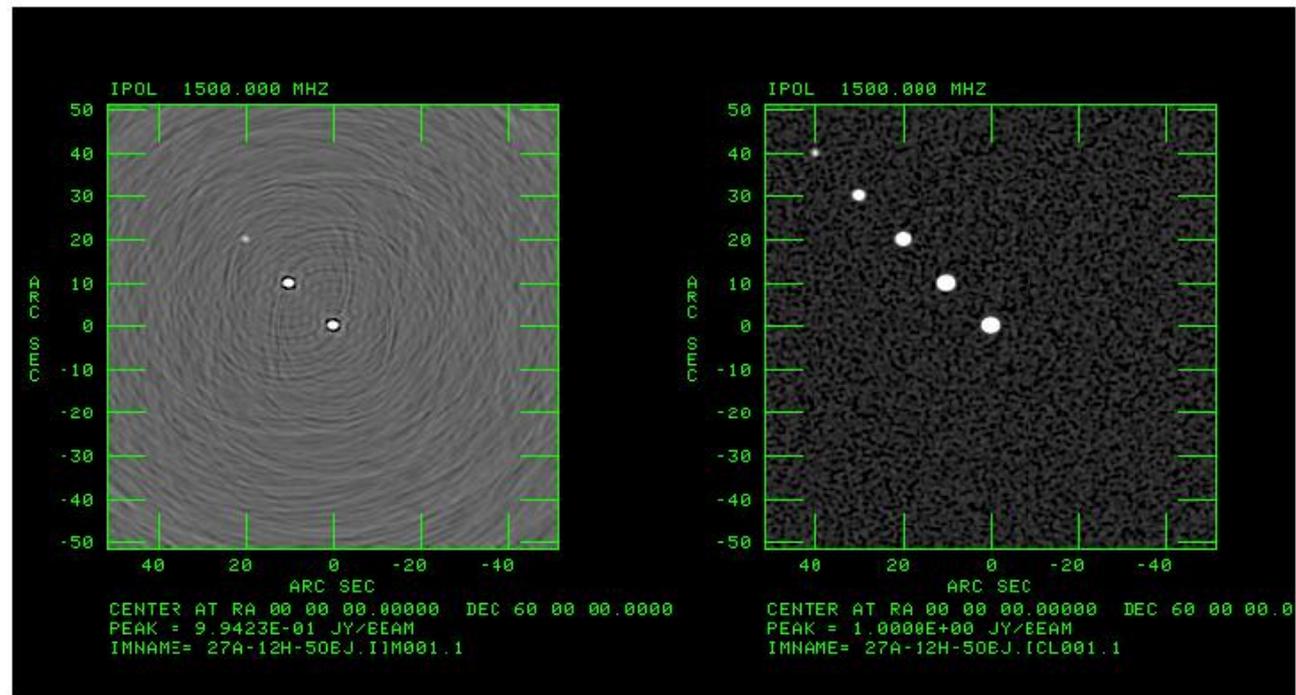


With a Full Synthesis:

- Unsurprisingly, the quality and sensitivity of the image improves with more data.
- Here are the dirty and CLEANed images for a full 12-hour observation.
- Note that the weakest source is easily visible now.

- Dirty Image rms:
4.1 mJy

- Clean Image rms:
0.012 mJy



Multi-Scale Clean

- Two major issues with basic Clean were mentioned earlier:
 - Smooth emission is usually broken into beam-sized ‘lumps’.
 - Extended regions take a very long time to deconvolve.
- These two issues are addressed by ‘multi-scale’ Clean.
 - Multiple images, at a range of resolutions, are generated.
 - Deconvolution proceeds in parallel amongst these, with some internal watchdog to decide which resolution image is accessed.
 - In the major loop, the visibilities corresponding to the parallel smoothed components are subtracted from the visibility data, and the multiple multi-scaled images are regenerated.
 - Upon completion, the multiple multi-scaled CLEAN components are combined, and the final image generated.
- This implementation provides smoother images in less time.



Compressed Sensing Methods – the ‘cutting edge’.

- Deconvolution is the most uncertain – and critical – step in the imaging process.
- It is also very time-consuming, and compute-intensive.
- Clean-based, and MEM-based methods have well known drawbacks.
- A ‘new generation’ of imaging scientists, driving by these new arrays and the data they produce, is experimenting with new methods.
- An example is Urvashi Rao’s ‘ASP’ algorithm – similar to MSClean, but automatically adjusts the size of the best-fit gaussians during the subtraction process.

MEM Methods -- VTESS

- Another algorithm less commonly used is ‘VTESS’, which uses a ‘MEM’ method. (‘MEM’ = ‘Maximum Entropy Method’).
- There was (and remains) a vigorous debate amongst the experts (and in the literature) concerning this method.
- This algorithm utilizes a ‘dirty’ image and psf to provide an output image which:
 - fits the data, and
 - maximizes a function chosen to emphasize output smoothness and positivity.
- A function commonly employed is:

$$H = -\sum I_k \ln\left(\frac{I_k}{M_k}\right)$$

where M_k is a ‘default’ image.

- Various other functions have been proposed, and there has been much discussion on the informatic content of such functions.



MEM, (cont.)

- The procedure to find the new image must be constrained to match the original visibility data, within the noise.

- The algorithms, at each step, do this by minimizing the χ^2 (to within the noise):

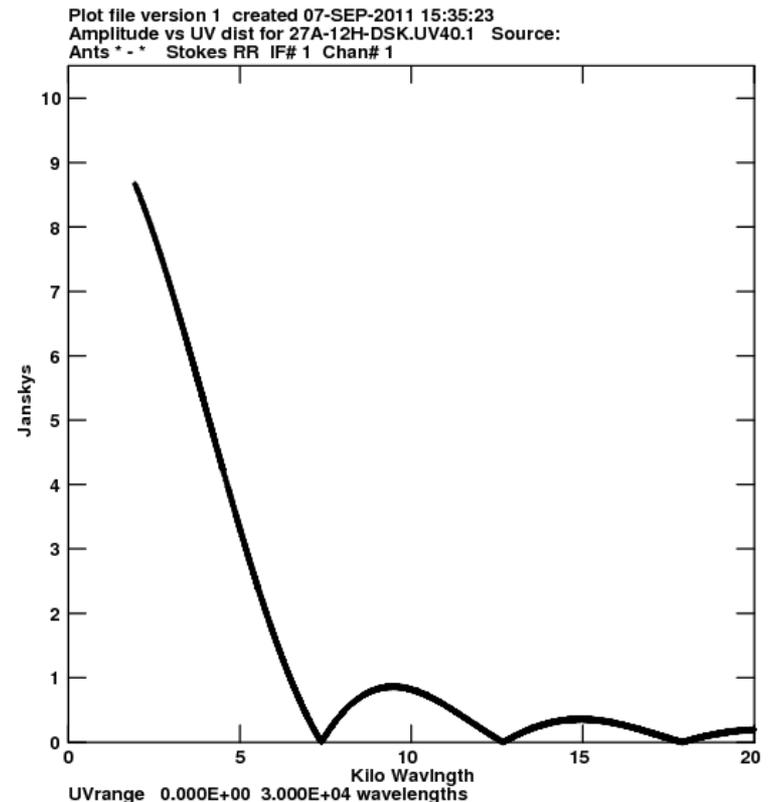
$$\chi^2 = \sum_i \left| (V_{obs,i} - V_{mod,i}) \right|^2 / \sigma_i^2$$

- As written, the resulting image fits long spacings too well, and the short spacings (sensitive to total flux) too poorly.
- This problem is avoided by adding another constraint: That the solution match the known total flux.
- The ‘default image’ is important to assist the algorithm to find an acceptable solution.
- Most common ‘default’ is a blank image whose sum equals the total flux.
- Better would be a low resolution image from (say) a single dish, or a lower resolution array.



Simulation Observations of a Big Disk.

- To allow comparison of VTESS with CLEAN, I generated a 12-hour observation of a 10 Jy, smooth 40 arcsecond wide disk.
- The radial visibility function is shown on the right.
- It is a 'jinc' function – $J_1(x)/x$.
- Note that there are no data between 0 and 2 $k\lambda$ – the complete visibility function should be 10 Jy at 0 baseline.
- In essence, the imaging programs use '0' when there are no data.

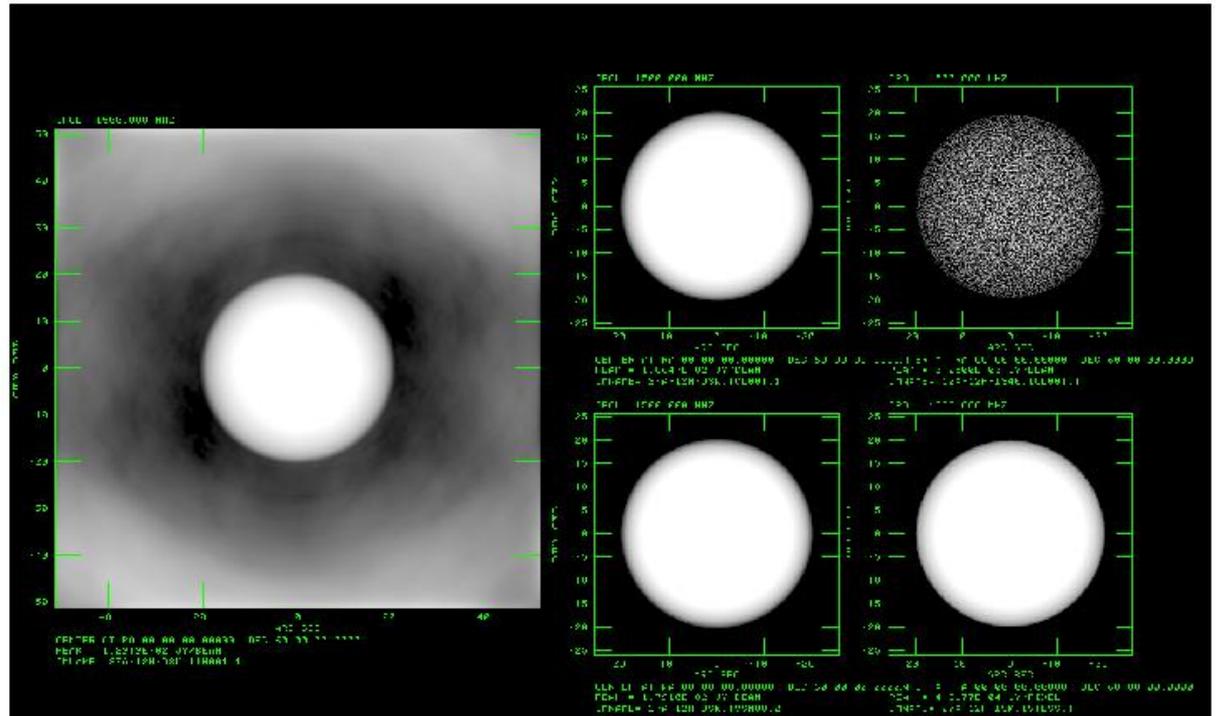


VTESS vs. CLEAN – Big Disk

- Shown below are test deconvolutions (using simulated data) for an image of a 40 arcsecond-wide disk.
- The large panel shows the ‘dirty’ image – note the negative bowl, caused by the absence of data at short spacings.
- Top Panels: CLEAN. Bottom Panels: VTESS

Small Panels:

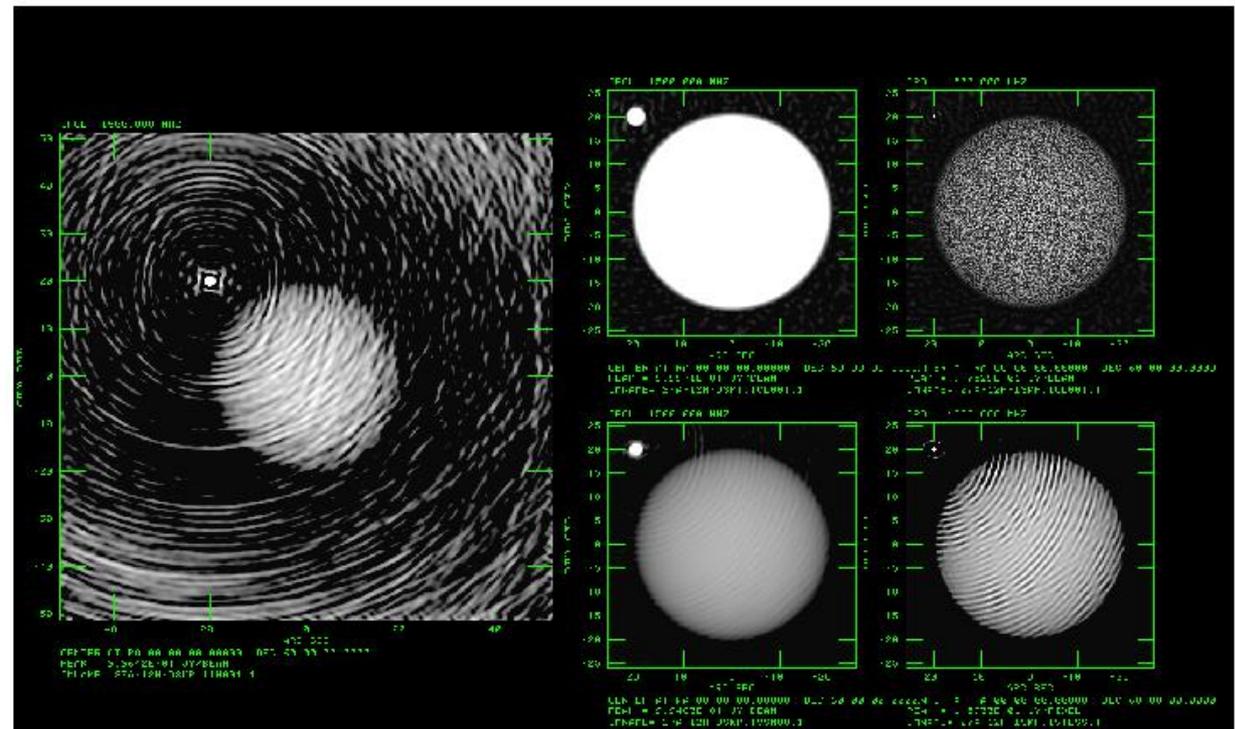
- Top Left: 100,000 component CLEAN, with restore.
- Top Right: Same CLEAN, but no clean beam restore.
- Bottom Left: 40-cycle VTESS, with gaussian smoothing.
- Bottom Right: Same VTESS, but no smoothing.
- Disk brightness: ~ 20 mJy/bm.



VTESS is not suitable for Point Sources.

- VTESS seeks maximum smoothness. It is thus not expected to handle point objects very well.
- The simulation below demonstrates this.
- I've added a 1 Jy point source to the 10 Jy large disk.

- CLEAN handles the additional point source well.
- VTESS has done rather badly.
- A hybrid method would be optimal.
- VTESS has provided super-resolution.



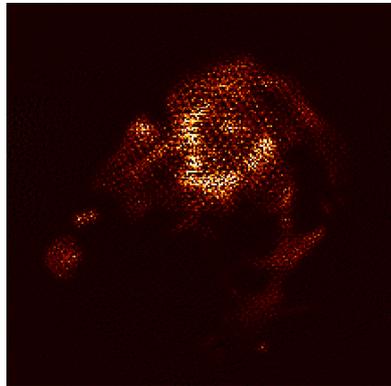
Some Final Points: VTESS and CLEAN

- VTESS produces images which have ‘variable resolution’ – point sources will deconvolve to different shapes, depending on SNR.
- If a uniform ‘flux/beam’ is wanted, one must smooth the output image with a suitable Gaussian.
- But since VTESS provides good ‘super-resolution’ capabilities (much better than CLEAN), and provides output images in ‘Jy/pixel’, smoothing the output is throwing away information which is likely good.
- CLEAN *must* smooth the output images, since its ‘super-resolution’ capabilities are obviously poor.
- The ‘MEM’ methods do well in some situations, but CLEAN still dominates – for two important reasons (IMHO):
 - Most fields are indeed largely dominated by small objects.
 - It’s easier to use.

Deconvolution – Comparison of Algorithms

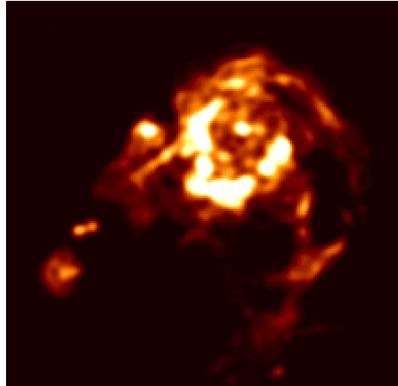
CLEAN

Minimize L2
(assume sparsity
in the image)



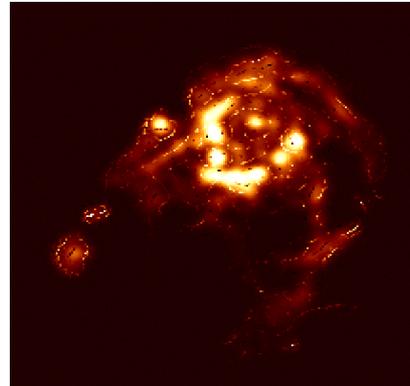
MEM

Minimize L2 subject to
an entropy-based prior
(e.g. smoothness)



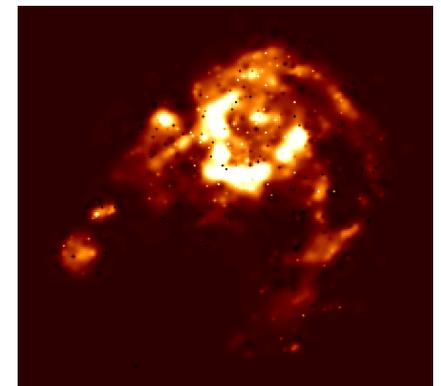
MS-CLEAN

Minimize L2
(assume a set of
spatial scales)

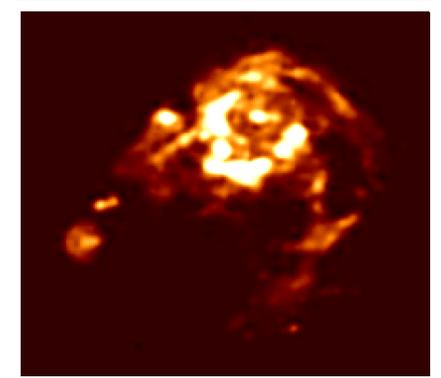
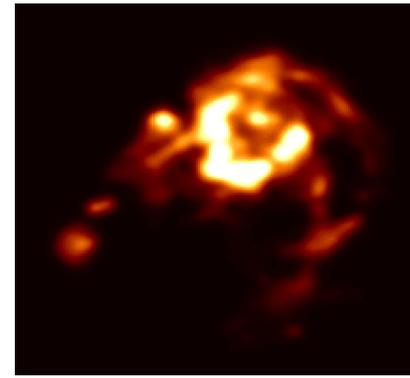
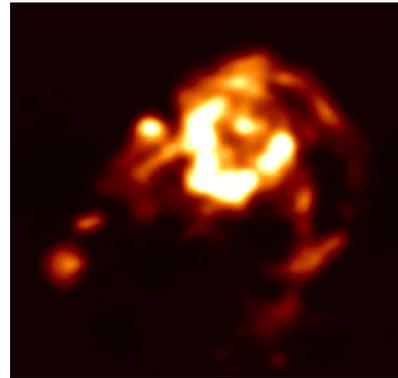
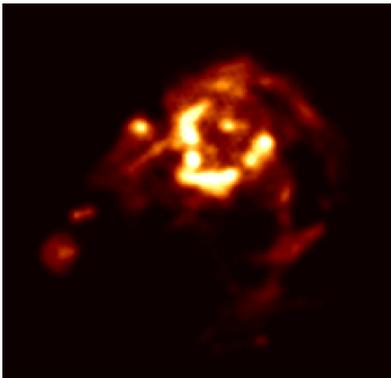


ASP

Minimize L2 with
TV-based subspace
searches



I^m

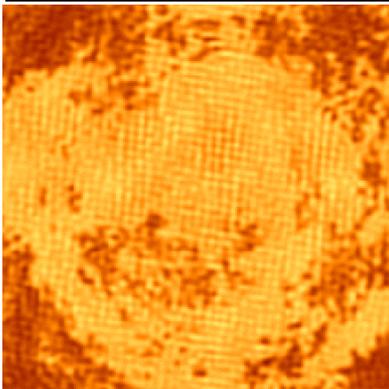
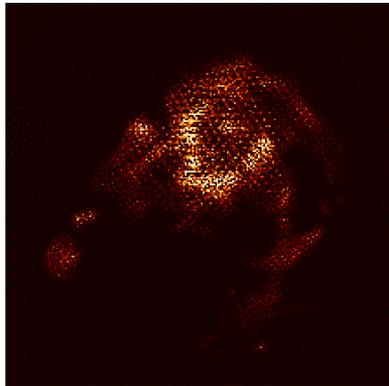


I^{out}

Deconvolution – Comparison of Algorithms

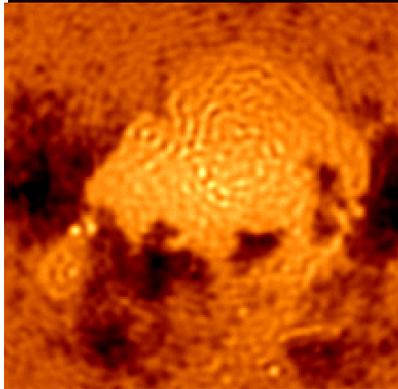
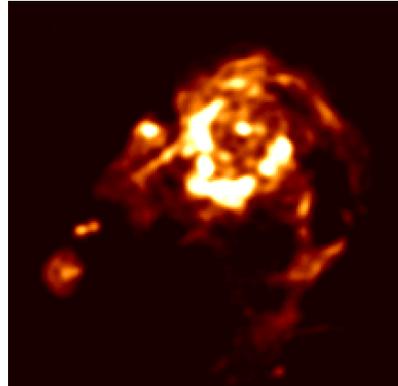
CLEAN

Minimize L2
(assume sparsity
in the image)



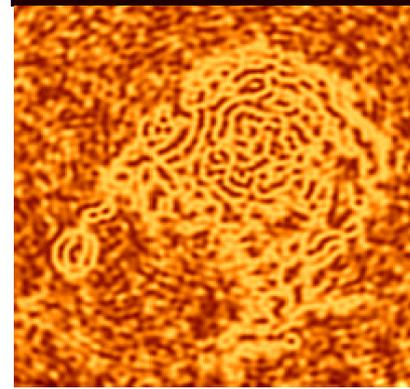
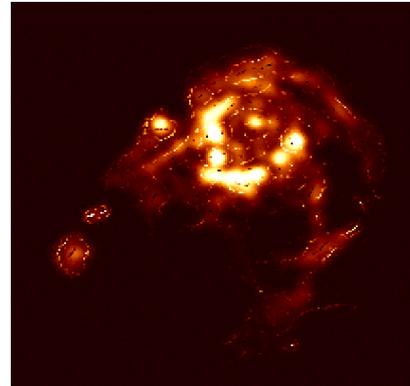
MEM

Minimize L2 subject to
an entropy-based prior
(e.g. smoothness)



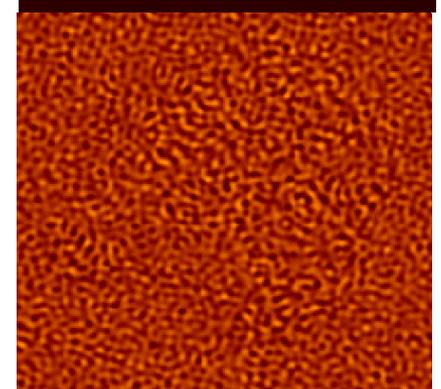
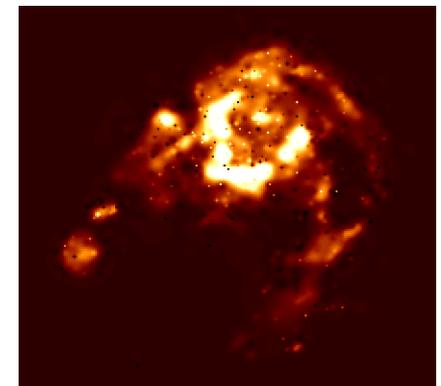
MS-CLEAN

Minimize L2
(assume a set of
spatial scales)



ASP

Minimize L2 with
TV-based subspace
searches



I^m

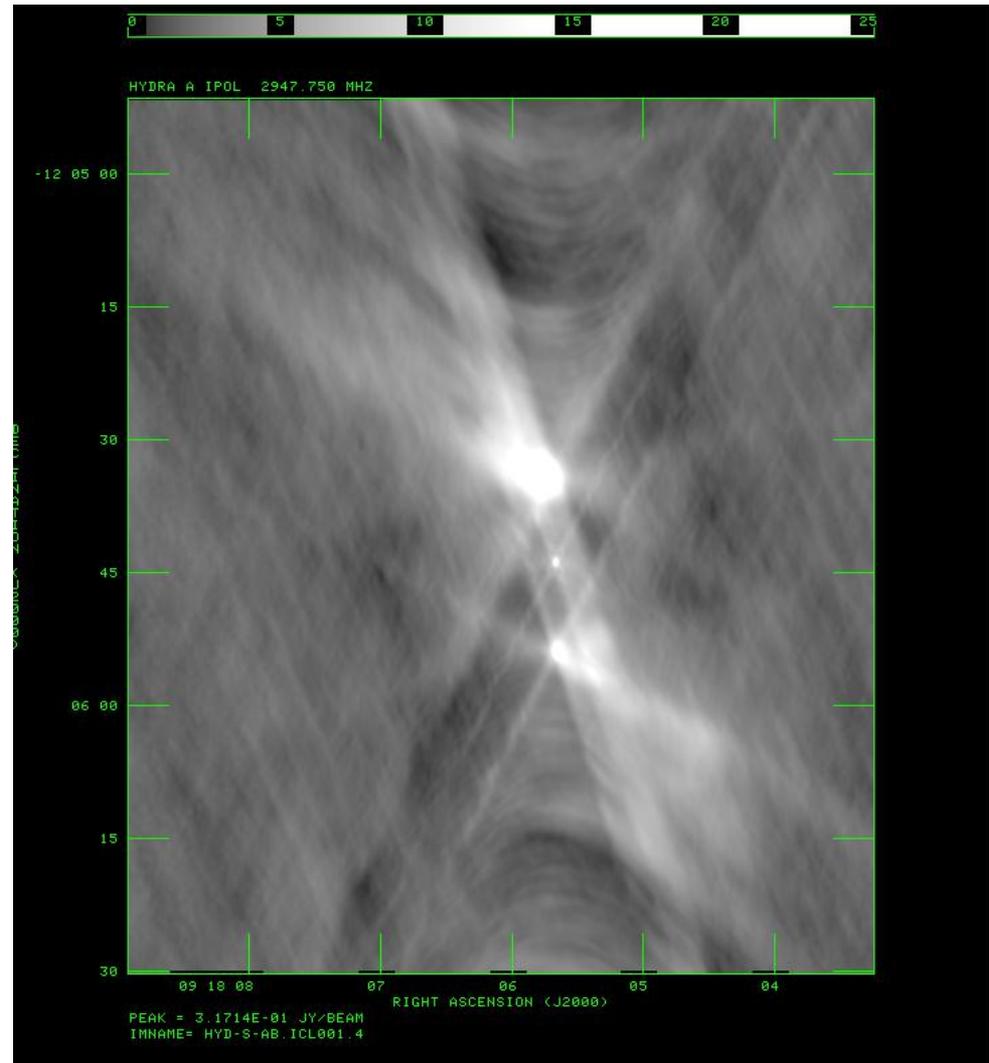
I^{res}

A comparison, using Real Data

- Source is Hydra A (3C218), a nearby FRI-type radio galaxy.
- Recent VLA data: L, S, C, and X bands (1 through 12 GHz, no gaps), all four configurations, full polarimetry. A major effort.
- Shown will be a narrowband (10 MHz BW) S-band (3 GHz) image, utilizing three of the four configurations (A, B, D).
- The source is very extended with compact structure near the core, and large, billowy smooth emission extending to great distance.
- A real challenge for 'Clean'.

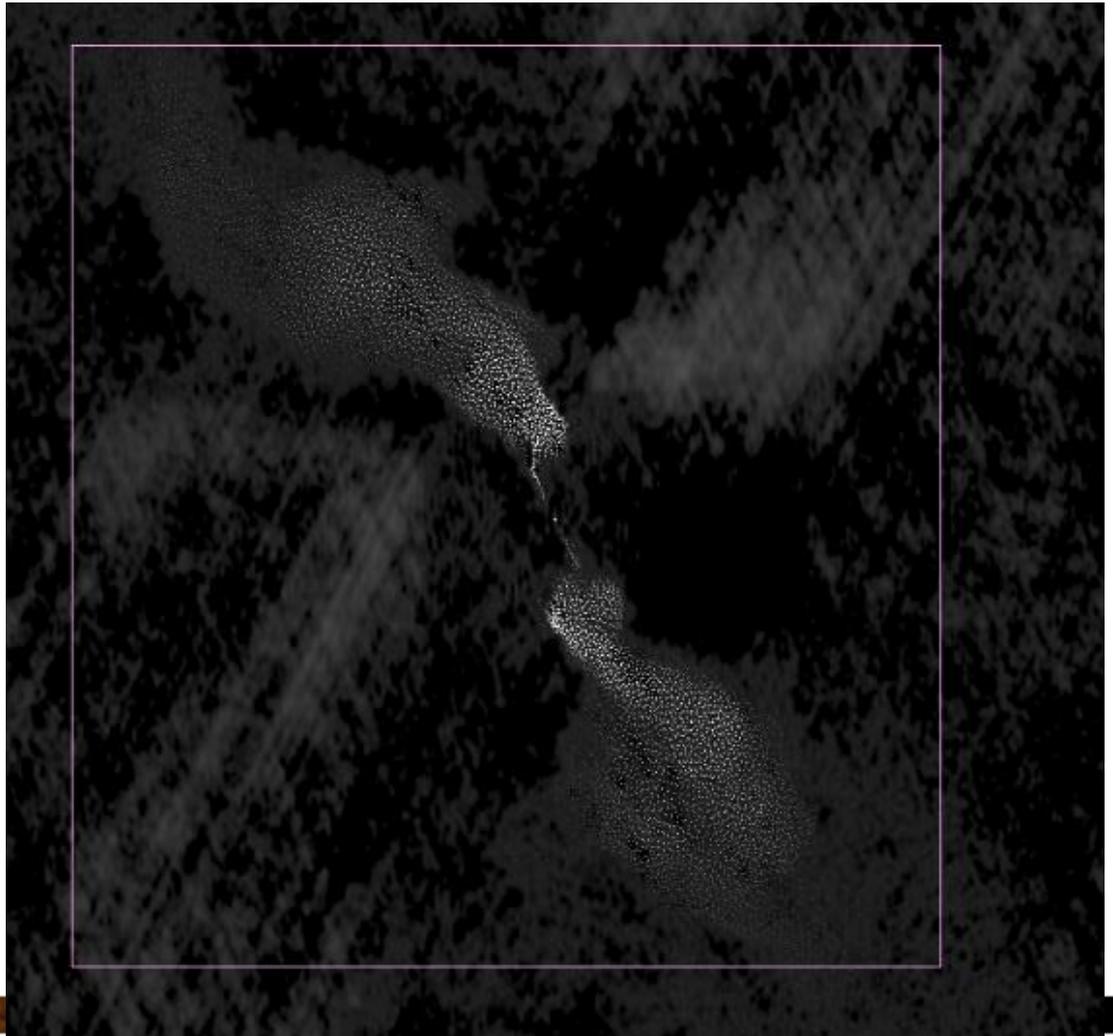
The 'Dirty' Image

- About 3 hours total integration, taken over a 12 hour span.
- The X-shaped sidelobes are a manifestation of the low, near-equatorial, declination (-12).
- The lobes and nucleus are visible, but little else.
- No credible science possible with this.



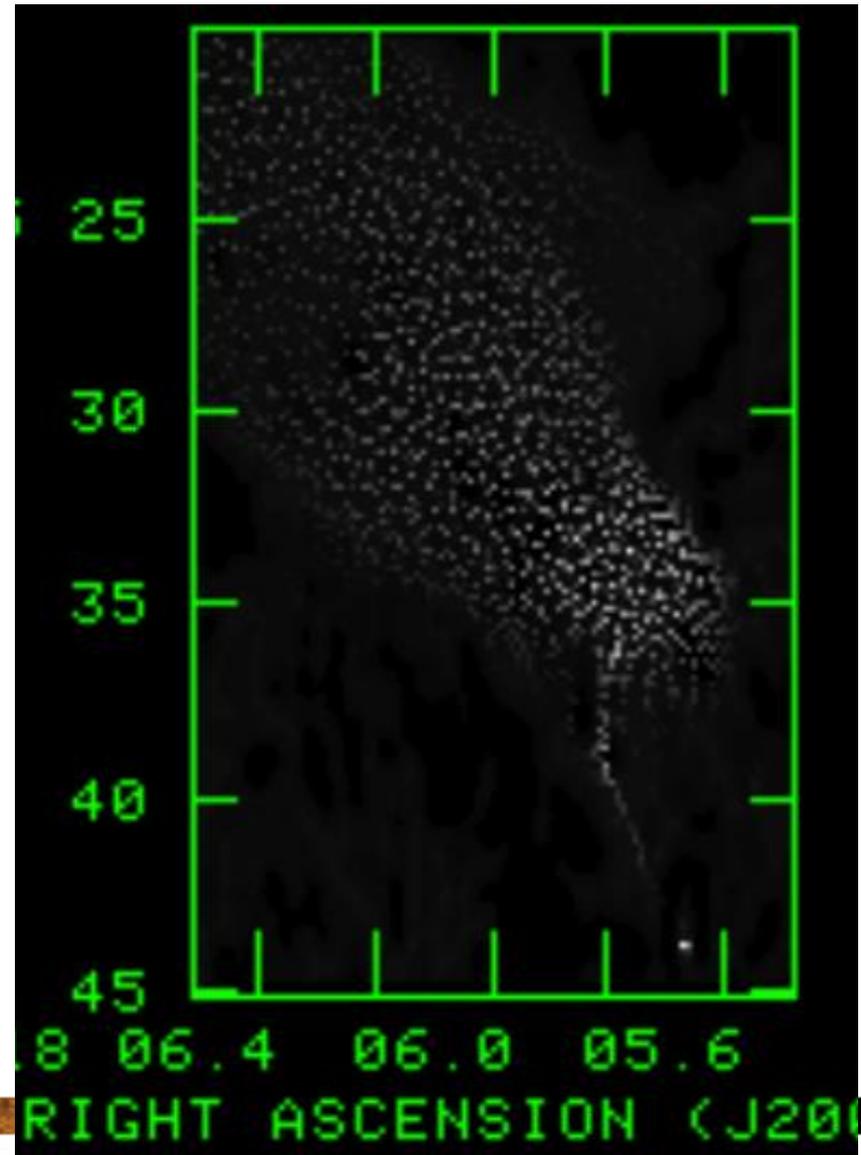
The Raw Clean Solution

- Shown here are the 50,000 Clean components, with the residuals added.
- This mass of point components is actually the *correct* solution, in the sense that its Fourier Transform will match the original data.
- But nobody would ever publish such an image, since nobody believes the lobes of this source are actually like this.
- So we blur the result to make it 'look better'.



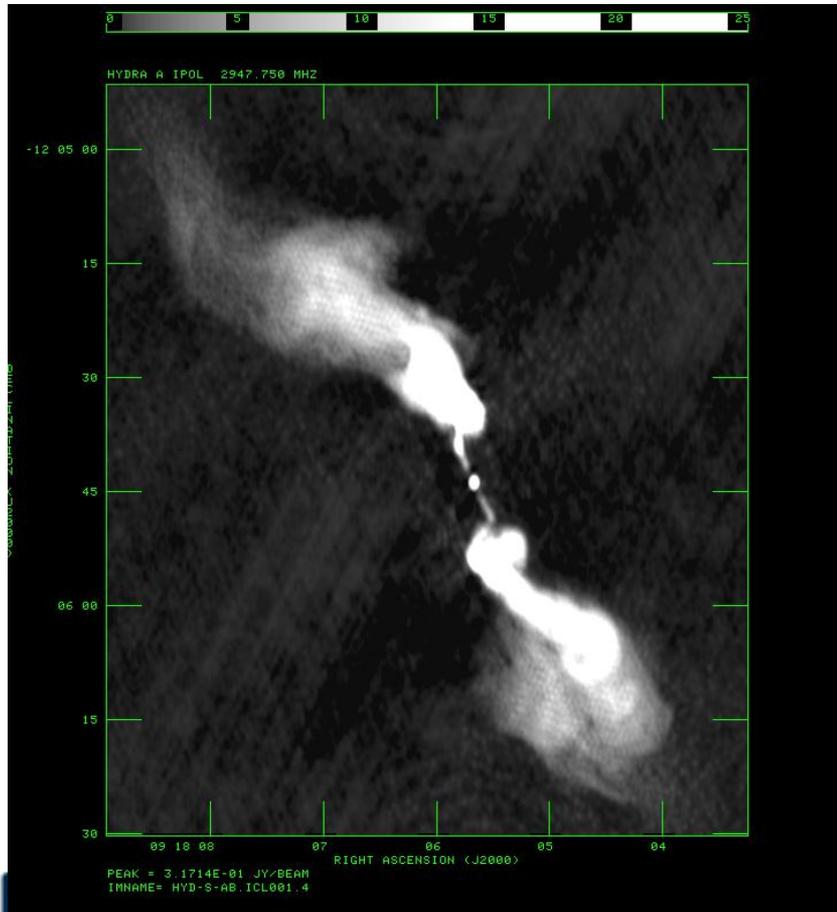
Zoom-In to the Northern Lobe

- Can see here how what we are sure is smooth and continuous emission has been broken into discrete ‘delta-functions’.
- The fundamental problem is the unsmooth sampling in the u-v plane.
- But this is seriously aggravated by the use of the point-spread function in areas of broad and smooth emission.

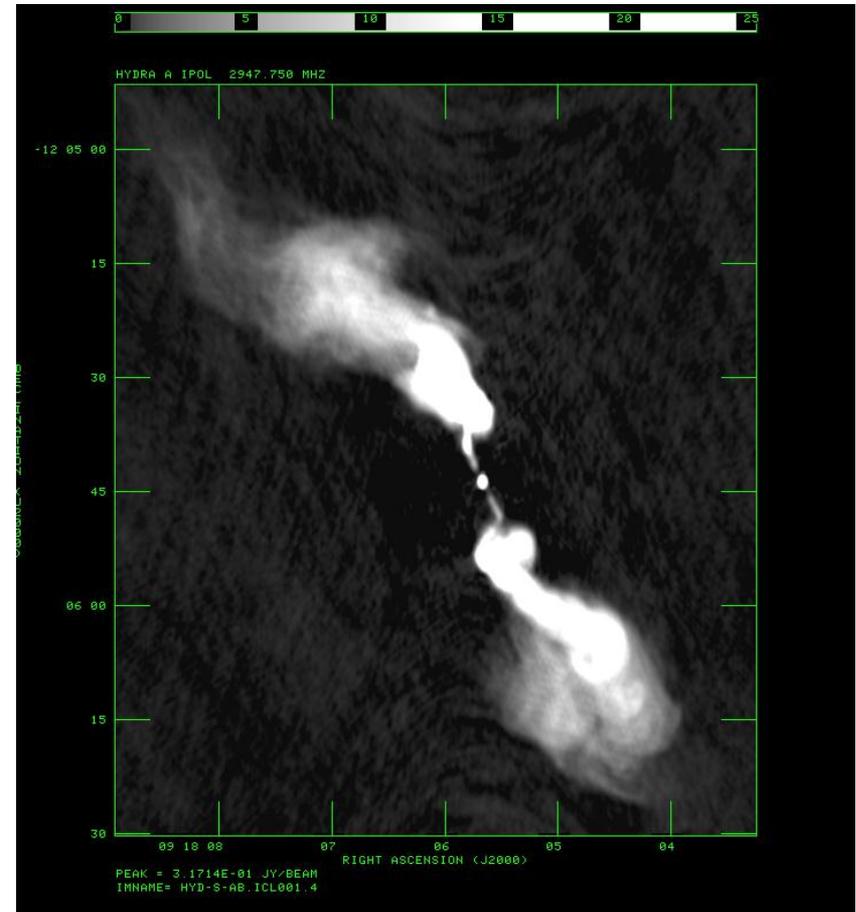


Standard Clean, and Multi-Scale Clean

- Standard CLEAN

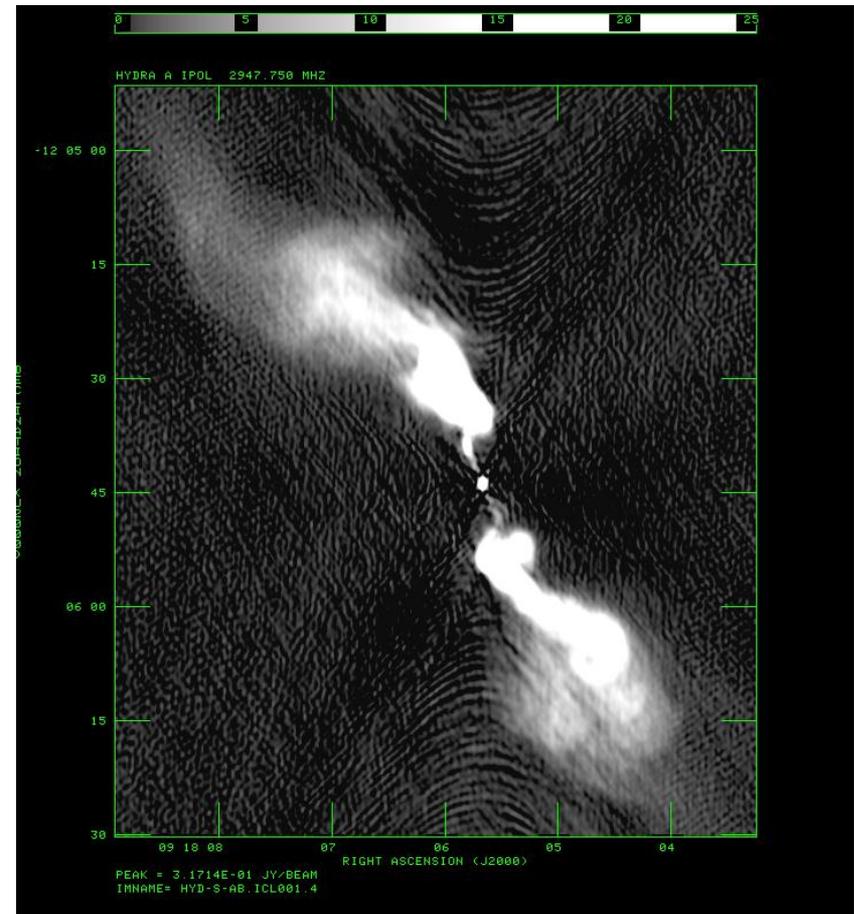


- Multi-Scale CLEAN



Max-Entropy (VTESS)

- The maximum Entropy image is definitely smoother, but clearly has made significant errors near the nucleus, and has failed to remove the fine-structure sidelobes.



MFS (Multi-Frequency Synthesis)

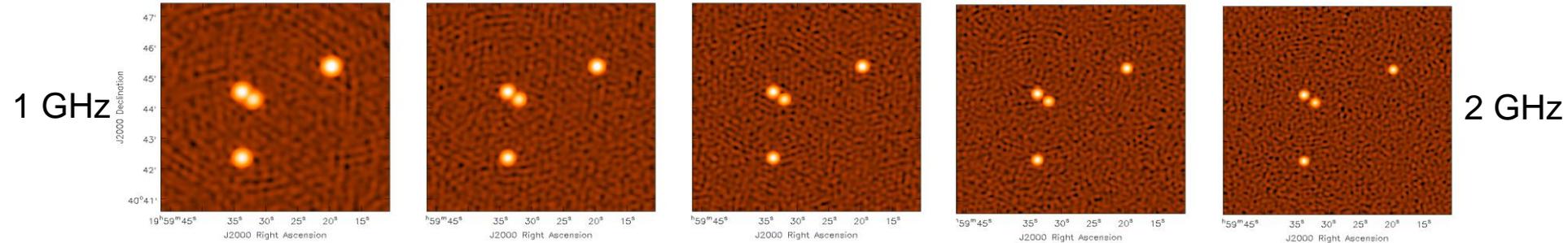
- Modern Telescopes provide huge fractional bandwidth.
- Properties of the emission can change dramatically over the BW.
- How best to image such data?
 - Break into subbands, image each separately, form a cube, and fit for median, slope, curvature? (Conceptually simple)
 - Develop a joint deconvolution, to form a single image of the median, spectrum, and curvature?
- Not an easy problem. Note that the resolution changes over the BW => opposite ends of a 2:1 BW band gives very different information.
- Urvashi Rao has produced MFS Clean in CASA.



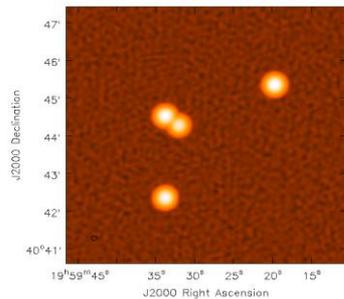
Spectral Cube (vs) MFS imaging

Simulation : 3 flat-spectrum sources + 1 steep-spectrum source (1-2 GHz VLA observation)

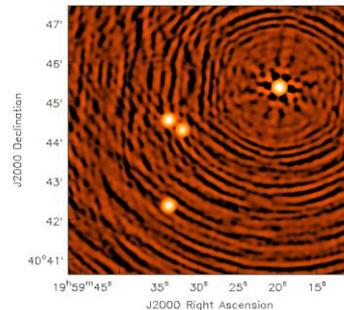
Images made at different frequencies between 1 and 2 GHz (limited to narrow-band sensitivity)



Add all single-frequency images (after smoothing to a low resolution)

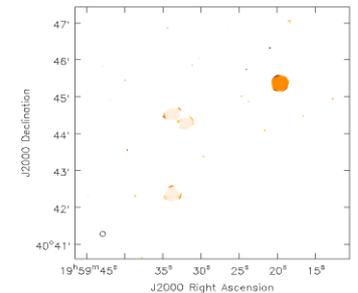
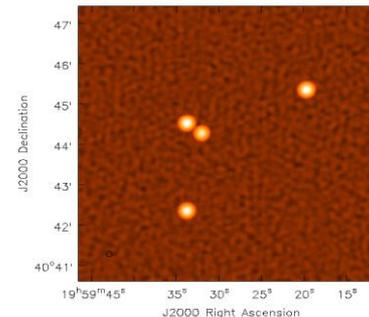


Use wideband UV-coverage, but ignore spectrum (MFS, nterms=1)



Use wideband UV-coverage + Model and fit for spectra too (MT-MFS, nterms > 1)

Output : Intensity and Spectral-Index



A Short Summary

- Although the Direct Transform will give the most accurate result, data volumes require the FFT, and hence gridding.
- The negative effects of gridding (aliasing, and attenuation over the field) can largely be negated through use of good convolution functions, and by utilizing a lot of data. (!)
- The mathematics of imaging is well established.
- Deconvolution is the process of removing sidelobes from images – equivalent to interpolating data into cells where none were taken.
- Deconvolution is as much an art as science – the CLEAN algorithm still has no good theoretical understanding.
- CLEAN methods do best on point sources, MEM methods are better for extended emission. MS Clean probably the best compromise.
- New methods – many based on Clean concepts -- will allow utilization of modern instruments to general sources.

