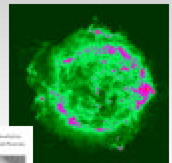
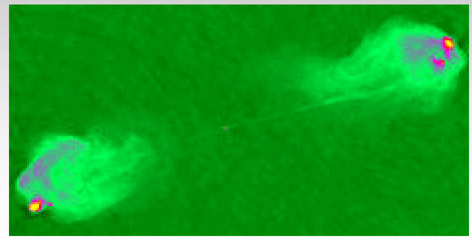
Imaging and deconvolution

Tim Cornwell, NRAO



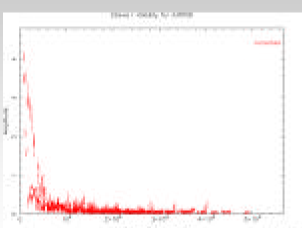
Overview

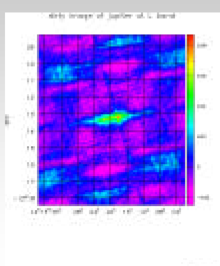
- How do we make images of the radio sky from measurements of the coherence function of the electric field measured by our antennas?



Imaging

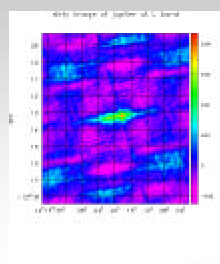
- Go from samples of visibility function to “dirty” image

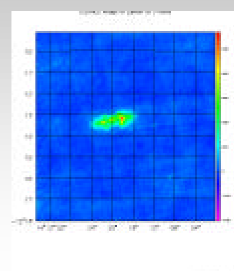




Deconvolution

- Go from dirty image to deconvolved image





Outline

- The relationship between sky brightness and visibility
- Sampling of the Fourier plane
- Fourier inversion
  - Weighting schemes
  - The problem with the dirty image - sidelobes
- Deconvolution
  - CLEAN
  - Maximum Entropy Method
  - Algebraic deconvolution
  - Other methods
- Some examples

Formal description

- For small fields of view, the visibility function is the 2-D Fourier transform of the sky brightness:

$$V(u,v)=\int I(l,m)\cdot e^{j2\pi(ul+vm)}dl\,dm$$

- We sample the Fourier plane at a discrete number of points:

$$S(u,v)=\sum_k \delta(u-u_k)\cdot\delta(v-v_k)$$

- So the inverse transform is:

$$I^D(x,y)=F^{-1}[S(u,v)\cdot V(u,v)]$$

- Applying the Fourier convolution theorem:

$$I^D(x,y)=B(x,y)\otimes I(x,y)$$

  where B is the point spread function:

$$B(x,y)=F^{-1}[S(u,v)]$$

Convolution theorem

- Inverse Fourier transform of sampled visibilities yields the true sky convolved with the point spread function
- Different ways to understand this effect:
  - Incomplete Fourier sampling => missing information about the sky brightness
  - Array ≡ masked aperture => diffraction patterns in image plane

  To find the true sky brightness I, we must "deconvolve" the point spread function B from the dirty image I^D
- What are the properties of the point spread function?
  - "sidelobes" with infinite extent
  - Invisible distributions

A digression: Fast Fourier Transforms

- FFTs are much faster than simple Fourier summation but a regular gridding is required
- Visibility data are irregularly re-sampled so we must resample the data on a regular grid
- Convolutional gridding is used: the discrete visibility samples are notionally smoothed to a continuous function, and then resampled at the regular grid points.
- Time-consuming but generally worthwhile
- Some fraction of the power is applied to the incorrect spatial frequencies: *aliasing* or *spurious sources*, usually at a very low level
- Long description in *Synthesis Imaging II*

Sidelobes

- From the sampling pattern, we can find that:

$$B(x,y)=\frac{1}{K}\sum_k \cos(u_k x+v_k y)$$

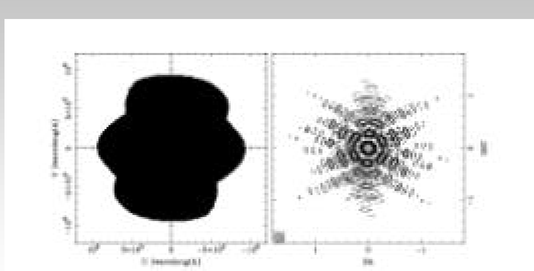
- So the point spread function is always a collection of co-sinusoids, and extends forever in the image plane
- At the center, $B(x,y)\sim 1$
- The PSF has a width $\Delta x\sim 1/u_{\max}$ $\Delta y\sim 1/v_{\max}$
- The RMS level is ~ $1/\sqrt{K}$

More on sidelobes

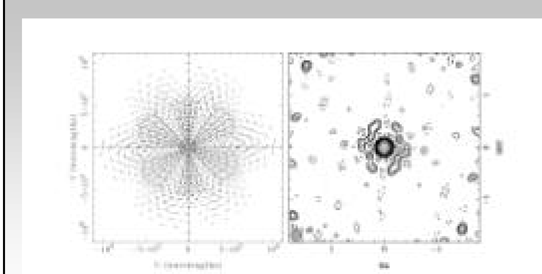
- Far-out sidelobes:
  - From the Fourier derivative theorem, if the sampling pattern has a discontinuous first derivative, the the drops off as the inverse of the radius in the image plane
- Close-in sidelobes:
  - Suppose that the sampling pattern is bounded by a circle, then the PSF close in must resemble the inverse Fourier transform of a circle: first order Bessel function divided by radius: *Jinc* function
- Can apply weighting to ameliorate these two effects:

$$B(x,y)=\frac{\sum_k w_k \cos(u_k x+v_k y)}{\sum_k w_k}$$

Close-in sidelobes



Close-in sidelobes



Weighting

- Choose the weighting function to alter properties of PSF:
- Uniform weighting
  - To minimize RMS sidelobes over entire image requires:

$$w_k = 1/r(u_k, v_k)$$

  - But SNR suffers...

Weighting

- Natural weighting
  - To minimize noise over entire image requires:

$$w_k = 1/s_k^2$$

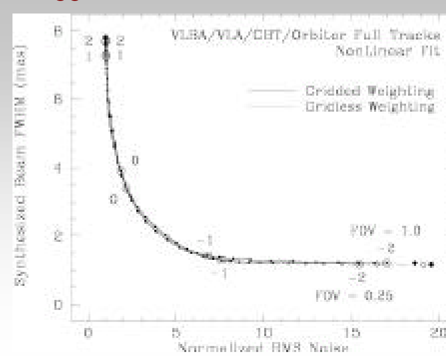
- Briggs (robust) weighting
  - To minimize noise plus sidelobes for point source of strength S requires

$$w_k = 1/\left[S^2 \cdot r(u_k, v_k) + s_k^2\right]$$

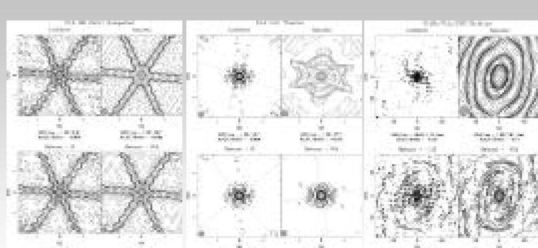
More on weighting

- Super-uniform weighting
  - Can choose to minimize sidelobes over smaller region than entire image
  - Divide out density averaged over large region in Fourier space
- All weighting decreases the sensitivity relative to natural weighting
- Uniform weighting increases the resolution relative to natural weighting
- Briggs' weighting allows a compromise between sensitivity and resolution

Briggs' curves



Examples of weighting



Tapering

- Can go further, and multiply by a desired sampling shape:

$$w_k = T(u,v)/r(u_k, v_k)$$

- For example, the desired shape could be a Gaussian, which transforms to a Gaussian, and therefore falls away rapidly in the image plane
- BUT, the underlying sampling pattern eventually wins...

Weighting and tapering help, but cannot entirely remedy the limitations in the image due to finite Fourier plane sampling

Invisible distributions

- There are sky brightness distributions Z that are invisible:

$$B \otimes Z = 0$$

- This occurs when the spatial frequencies (u,v) in the invisible distribution Z are not sampled
- Some examples:
  - Total integrated brightness (usually but not always)
  - Short spacings below the *minimum* separation of antennas
  - Long spacings beyond the *maximum* separation of antennas
  - Holes in the sampling pattern
  - Any combination of the above!
- No *linear* method can ever recover the invisible distributions

$$D \otimes I^D = D \otimes [B \otimes (I + Z)] = D \otimes B \otimes I$$

How to find invisible distributions?

- Apply *a priori* knowledge about the sky brightness
- What do we know?
  - Sky brightness is positive, sum of co-sinusoids is not necessarily
  - Sky is mostly dark, sum of co-sinusoids is not
  - Sky is collection of point sources, sum of co-sinusoids is not
  - Sky may be smooth, sum of co-sinusoids is probably not
- *Non-linear* deconvolution algorithms solve for an estimate of the true sky brightness I, from the convolution equation, while applying *a priori* constraints on the final solution

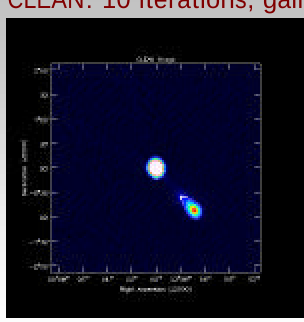
Popular deconvolution algorithms

- CLEAN:
  - sky is composed of point sources on a dark sky
  - sky is composed of resolved sources of known extent on a dark sky
- Multi-scale CLEAN:
  - sky is composed of smooth, limited extent blobs on a dark sky
- Maximum Entropy Method:
  - sky is smooth and positive
- Non-negative least squares:
  - sky is non-negative and compact
- Hybrid algorithms:
  - Some combination of the above...

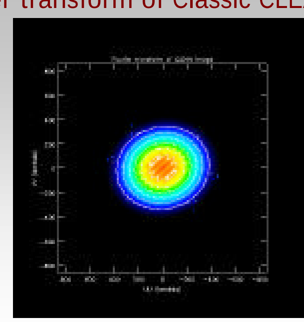
Classic CLEAN

- *A priori* constraint: sky is composed of point sources on a dark sky
- Uses iterative algorithm to find sequence of point sources
  - Find peak in image
  - Subtract a PSF centered and scaled appropriately to remove the effect of the brightness point, store component thus found
  - If any significant points left, return to first step
  - Convolve point components by "Clean" point spread function
    - Same width as dirty psf but no sidelobes
  - Add residuals image to obtain "restored" image
- Classic CLEAN algorithm due to Hogbom (1974)

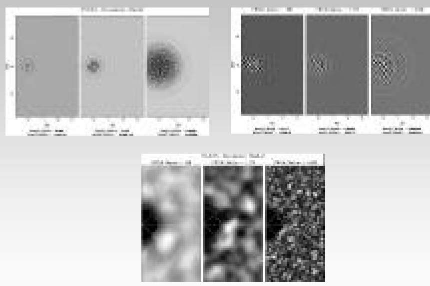
Classic CLEAN: 10 iterations, gain=1.0



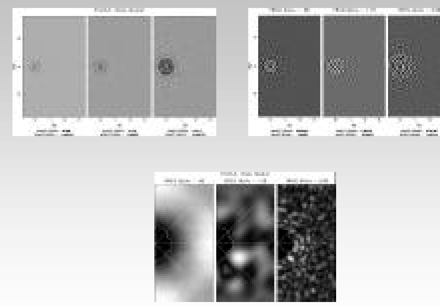
Fourier transform of Classic CLEAN



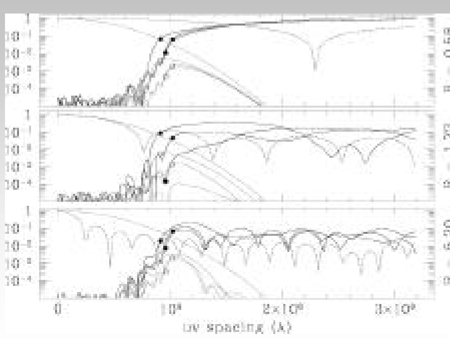
Classic CLEAN: Gaussian model



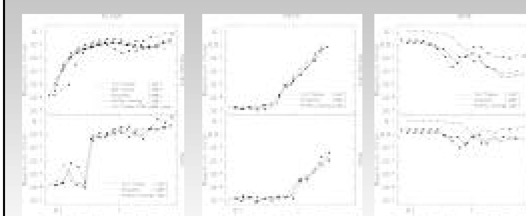
Classic CLEAN: disk model



Classic CLEAN: disk model



Algorithm comparisons



Classic CLEAN details

- Usually stabilize algorithm by subtracting only a fraction (the loop gain ~ 0.1) of the strength of a point source
- Usually stop either after finding a given number of components or when the peak residual is reaches a threshold, such as a multiple of the intrinsic noise level
- Schwarz (1978) showed that
  - Classic CLEAN must converge *i.e.* the peak residual must decrease
  - Classic CLEAN is equivalent to a least square fit of sinusoids to the visibility data
- Excellent at reducing identifying and correcting for point sources, less effective for extended emission in neighboring pixels

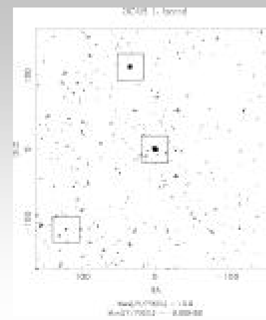
Classic Window CLEAN

- *A priori* constraint: sky brightness extent is known a priori
- Uses Classic CLEAN iterative algorithm to find sequence of point sources in restricted region delimited by CLEAN boxes
- Allows close specification of source support contraints
- Very useful for poor Fourier plane coverage *e.g.* VLBI

CLEAN variants

- Clark CLEAN: faster variant of Hogbom CLEAN
  - Split into two stages
  - Cleans subset of brightness points in minor cycle
  - Subtracts sidelobes completely using Fast Fourier Transform convolution in major cycle
  - 0.1-10 times faster than Hogbom
- Schwab-Cotton CLEAN: another variant of Clark CLEAN
  - Clark minor cycle
  - Major cycle subtracts components directly from visibility data
  - Sometimes faster, always more accurate than Clark CLEAN
  - Can clean multiple fields
- Steer-Dewdney-Ito: variant of Clark CLEAN
  - Minor cycle simply takes scaled version of pixels brighter than some trim level

Schwab-Cotton CLEAN



Multi-scale CLEAN

- *A priori* constraint: sky is composed of smooth blobs on a dark background
- Decompose sky into summation of blobs of various sizes *e.g.* truncated parabolas of width 0, 3, 10, 30 pixels.
- Perform one CLEAN algorithm for each scale size in parallel, and choose blob that gives the greatest reduction in peak residual
- Excellent at identifying large-scale coherent structure
- Residuals are quite noise-like

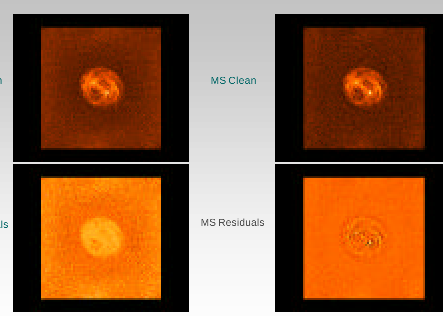
Multi-scale CLEAN

Clark Clean

MS Clean

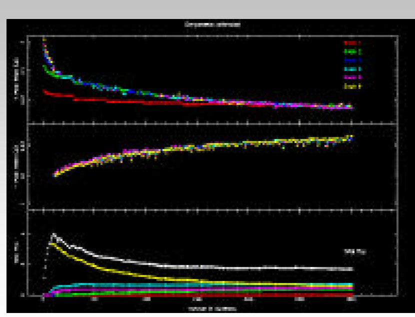
Clark Residuals

MS Residuals



Multi-scale CLEAN

Convergence on various scale sizes. Clean gain=0.5



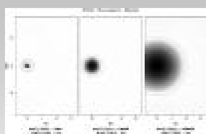
Maximum Entropy Method

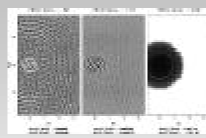
- *A priori* constraint: sky is smooth and positive
- Algorithm maximize a measure of smoothness (entropy) while solving the convolution equation

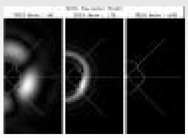
$$H(I) = -\sum_k I_k \cdot \log(I_k / m_k)$$

- where m_k is a "default" image which is the image obtained with no data. Usually a flat default image is used.
- Non-linear optimization problem: AIPS, AIPS++ use Cornwell-Evans (1983) algorithm
- Excellent for large diffuse emission
- Default image is very powerful for incorporating prior images
- Extensible to multiple simultaneous convolution equations
  - *See Mosaicing lecture*

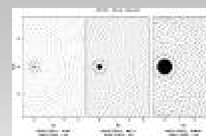
Maximum Entropy Method: Gaussian

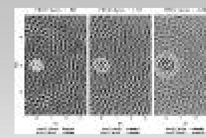


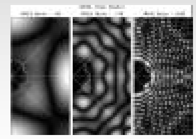




Maximum Entropy Method: Disk







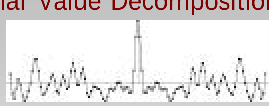
Maximum Entropy Method details

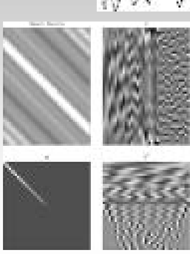
- Fast and efficient for million or more pixels
- Excellent on smooth extended emission with limited dynamic range
- Point sources cause problems
  - Should be removed using CLEAN before applying MEM
- Lots written about philosophy and meaning of MEM

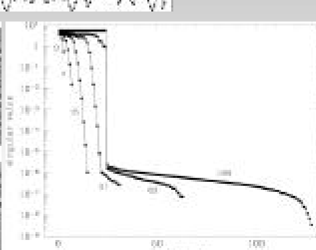
Algebraic deconvolution

- Pixellate convolution equations and represent via linear algebra $A\underline{x}=\underline{b}$ where the matrix A represents the point spread function, $\underline{x}$ is the unknown image as a vector, and $\underline{b}$ is the dirty image as a vector.
- The problem is then to solve this linear equation using various constraints
  - *e.g.* support constraints: we know that the emission is non-zero for only some areas
- Solve equation using *e.g.* Singular Value Decomposition
  - usually inadequate to get reasonable result but useful as indication of conditioning of problem

Singular Value Decomposition







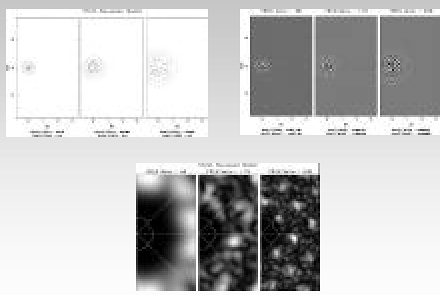
Non-Negative Least Squares

- Impose non-negativity using any of a variety of solvers

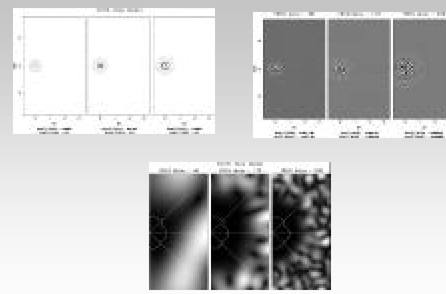
Solve $A\underline{x}=\underline{b}$ subject to $\underline{x}\geq 0$

- Works well for high dynamic range images of moderately resolved sources (Briggs' thesis, 1995)

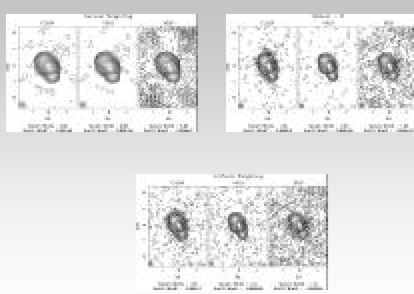
Non-Negative Least Squares: Gaussian



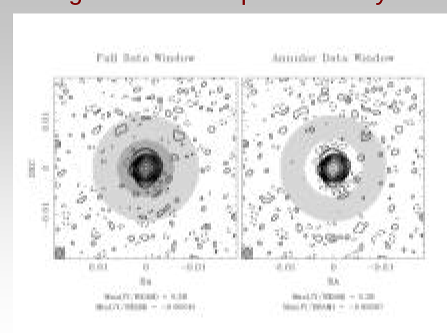
Non-Negative Least Squares: Disk



Non-Negative Least Squares: Source



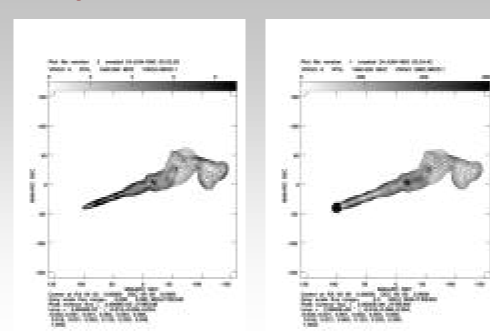
Non-Negative Least Squares: Party trick



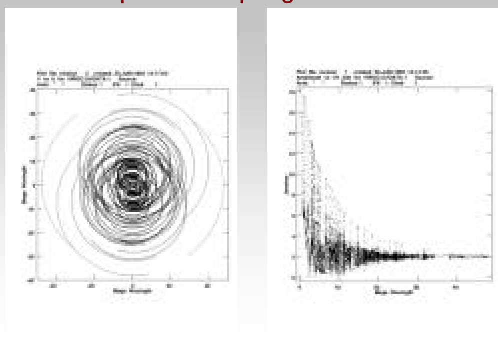
Example

- VLBA simulated observations of M87-like jet source
- Will show
  - UV coverage
  - Visibility function
  - Point Spread Function
  - Dirty image
  - Clean images
  - Maximum Entropy images

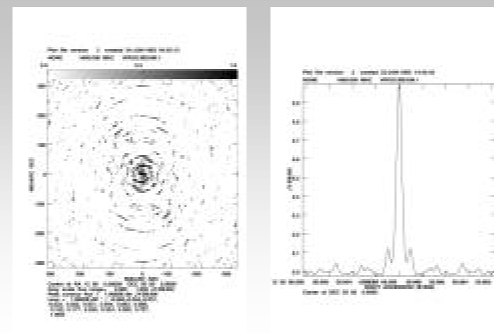
Original and smoothed model



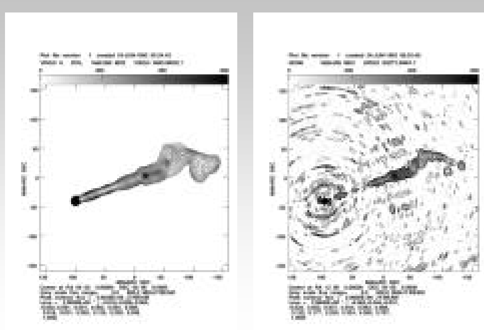
Fourier plane sampling



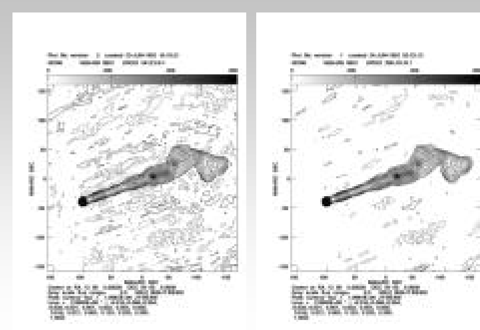
Point Spread Function



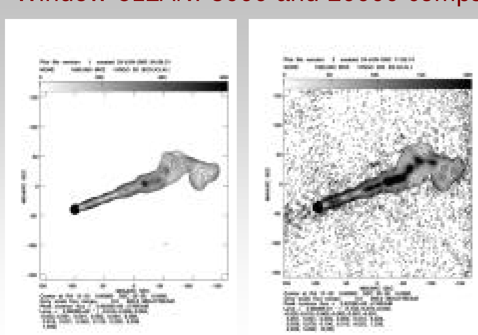
Original model and Dirty image



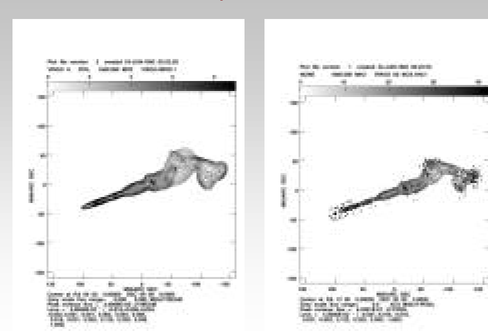
Classic CLEAN: 5000 and 20000 comps



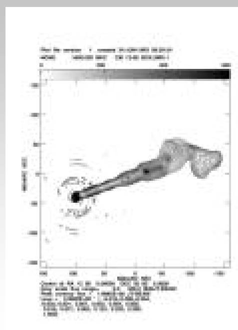
Window CLEAN: 5000 and 20000 comps

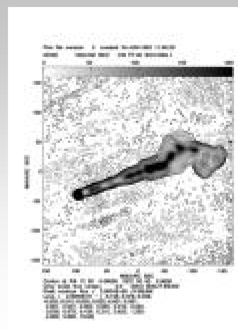


MEM: failure of super-resolution

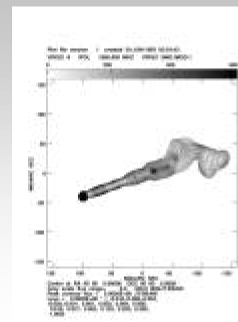


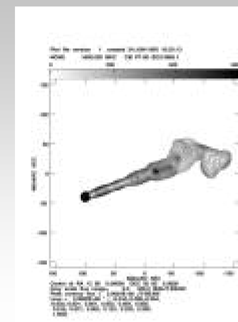
MEM: boxed, with point source removed



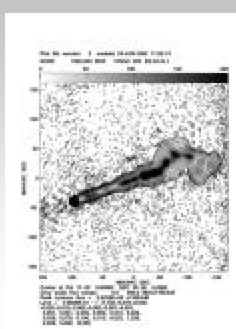


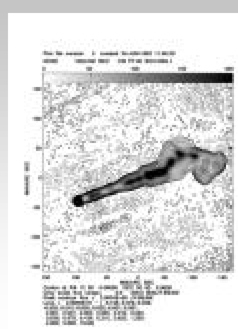
Original model and best image





Best Clean and Best MEM

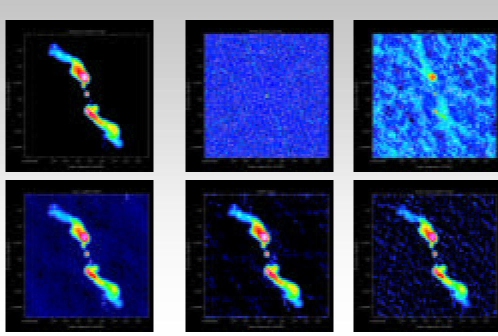




Example

- VLA multi-snapshot observation of Hydra A-like source
- Will show
  - UV coverage
  - Visibility function
  - Point Spread Function
  - Dirty image
  - Clean images
  - Maximum Entropy images

Model, PSF, Dirty image, CLEAN, MEM, Multi-scale CLEAN



Summary

- Incomplete Fourier plane coverage leads to diffraction patterns in images
- Deconvolution algorithms can correct for these patterns
- A number of complementary algorithms exist for image deconvolution

Bibliography

- *"Imaging and deconvolution"* in Synthesis Imaging II
- Dan Briggs' PhD thesis, *"High Fidelity Deconvolution of Mildly Resolved Sources",* New Mexico Tech, 1995
  - ftp://ftp.aoc.nrao.edu/pub/dissertations/dbriggs/diss.html