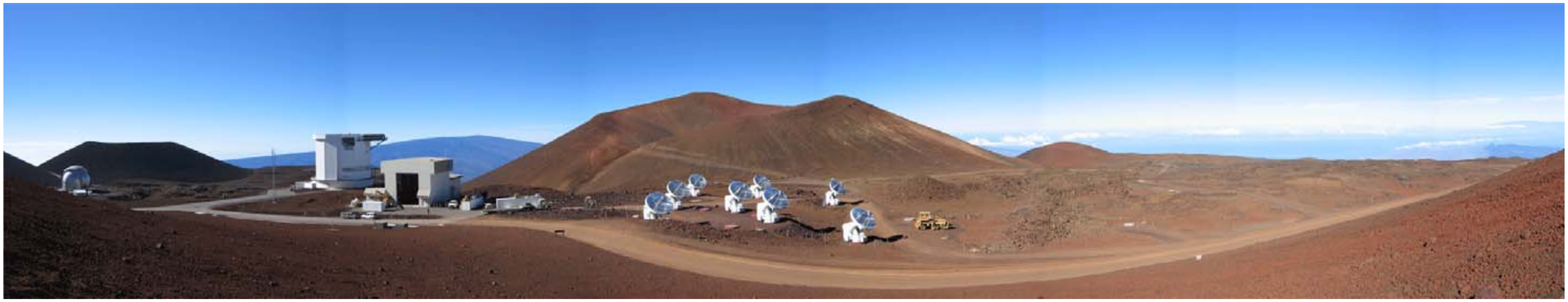
Imaging and Deconvolution

David J. Wilner
(Harvard-Smithsonian CfA)

Eleventh Synthesis Imaging Workshop
June 11, 2008



References

- Thompson, A.R., Moran, J.M. & Swensen, G.W. 1986, “Interferometry and Synthesis in Radio Astronomy” (New York: Wiley); 2nd edition 2001

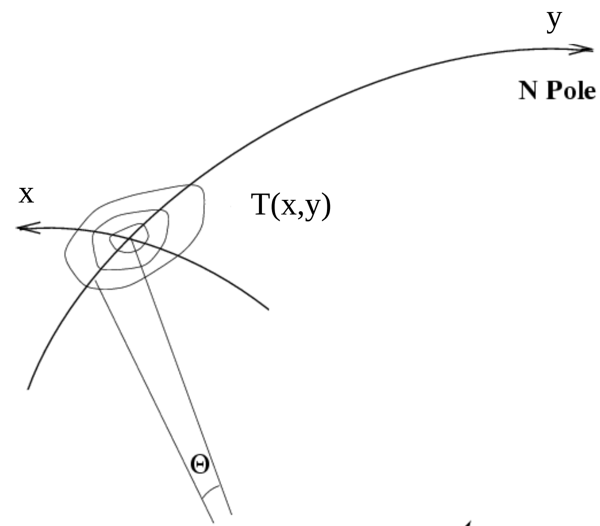
- NRAO Summer School proceedings
  - http://www.aoc.nrao.edu/events/synthesis/
  - Perley, R.A., Schwab, F.R. & Bridle, A.H., eds. 1989, ASP Conf. Series 6, Synthesis Imaging in Radio Astronomy (San Francisco: ASP)
    - Chapter 6: Imaging (Sramek & Schwab), Chapter 8: Deconvolution (Cornwell)
  - Lecture by T. Cornwell 2002, Imaging and Deconvolution
  - Lectures by S. Bhatnagar 2004, 2006 Imaging and Deconvolution
- IRAM Summer School proceedings
  - http://www.iram.fr/IRAMFR/IS/archive.html
  - Guilloteau, S., ed. 2000, “IRAM Millimeter Interferometry Summer School”
    - Chapter 13: Imaging Principles, Chapter 16: Imaging in Practice (Guilloteau)
  - Lecture by J. Pety 2004 “Imaging, Deconvolution & Image Analysis”

- Talks by many practitioners (Chandler, Wright, Welch, Downes, …)

Visibility and Sky Brightness

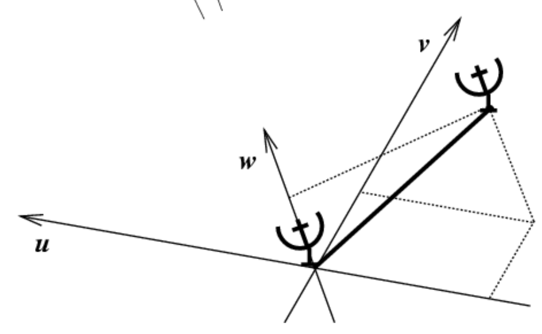
- from the van Citttert-Zernike theorem (TMS Appendix 3.1)

  – for small fields of view:
  the complex visibility,V(u,v)
  is the 2D Fourier transform of
  the brightness on the sky,T(x,y)



$$V(u,v) = \int\int T(x,y) e^{2\pi i (ux+vy)} dxdy$$

$$T(x,y) = \int\int V(u,v) e^{-2\pi i (ux+vy)} dudv$$



  – u,v (wavelengths) are spatial frequencies in
  E-W and N-W directions, i.e. the baseline lengths
  – x,y (rad) are angles in tangent plane relative to
  a reference position in the E-W and N-S directions

$$V(u,v) \rightleftharpoons T(x,y)$$

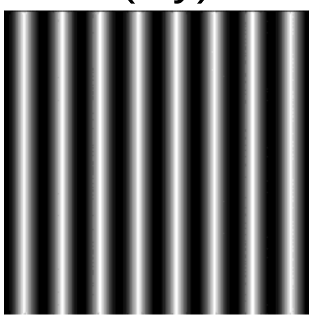
The Fourier Transform

Jean Baptiste

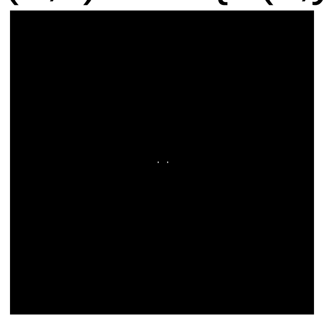
Joseph Fourier
1768-1830

- Fourier theory states that any signal (here images) can be expressed as a sum of sinusoids
- (x,y) plane and (u,v) plane are conjugate

T(x,y)



V(u,v) = FT{T(x,y)}



- in this example a single Fourier component encodes all
  - the spatial frequency = period of the wave
  - the magnitude = contrast
  - the phase (not shown) = shift of wave with respect to origin
- Fourier Transform image contains all information of original image

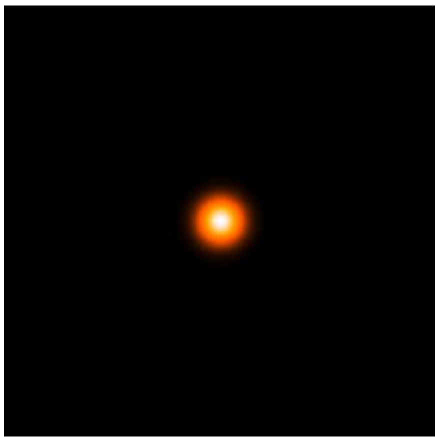
The Fourier Domain

- acquire comfort with the Fourier domain…
  - in older texts, functions and their Fourier transforms occupy *upper* and *lower* domains, as if "functions circulated at ground level and their transforms in the underworld'' (Bracewell 1965)

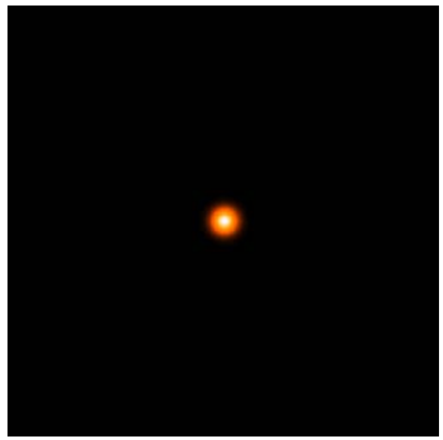
- a few properties of the Fourier transform: $f(x) \rightleftharpoons F(s)$
  - scaling: $f(\alpha x) = \alpha^{-1} F(s/\alpha)$
  - shifting: $f(x - x_0) = F(s) e^{i 2\pi x_0 s}$
  - convolution/multiplication: $g(x) = f(x) \otimes h(x); \quad G(s) = F(s) H(s)$
  - sampling theorem: $f(x) \subset \Theta$ completely determined if $F(s)$ sampled at intervals $\leq 1/\Theta$

Some 2D Fourier Transform Pairs

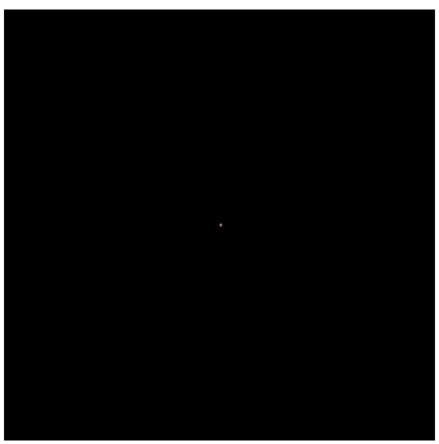
T(x,y) Amp{V(u,v)}



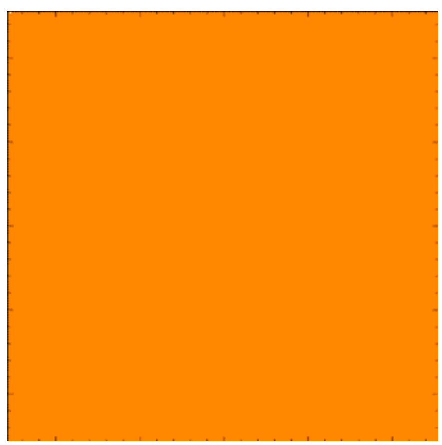
$\rightleftharpoons$



Gaussian Gaussian



$\rightleftharpoons$

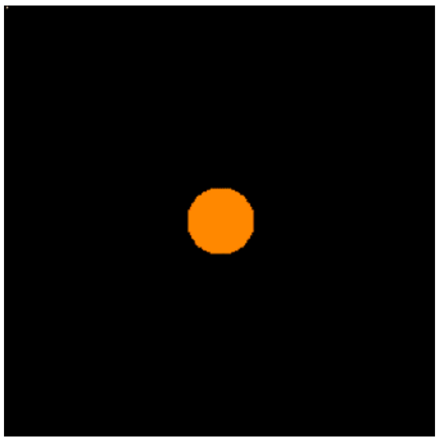


δ Function Constant

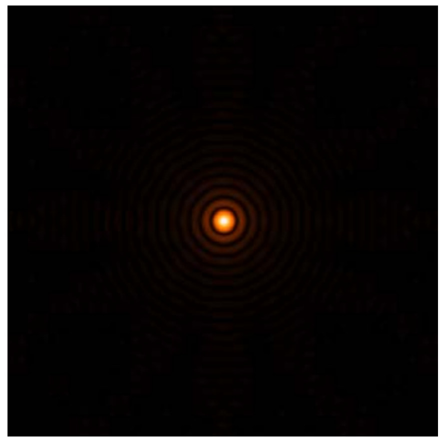
narrow features transform to
wide features (and vice-versa)

More 2D Fourier Transform Pairs

T(x,y) Amp{V(u,v)}

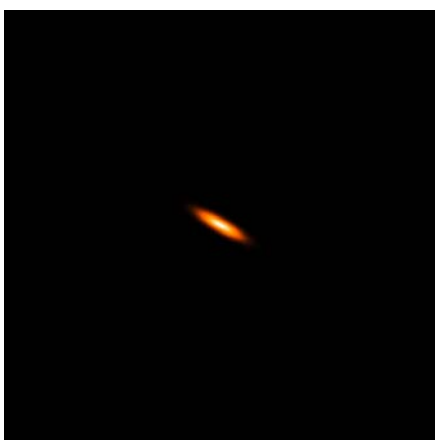


$\rightleftharpoons$

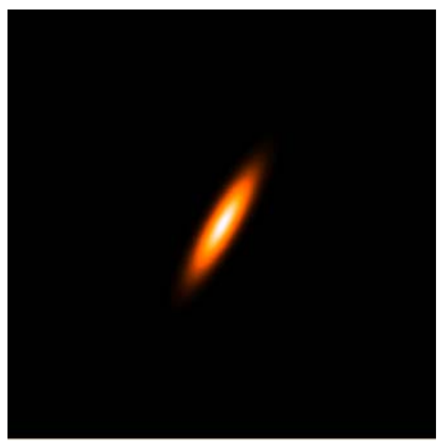


Disk Bessel

sharp edges result in many high spatial frequencies



$\rightleftharpoons$

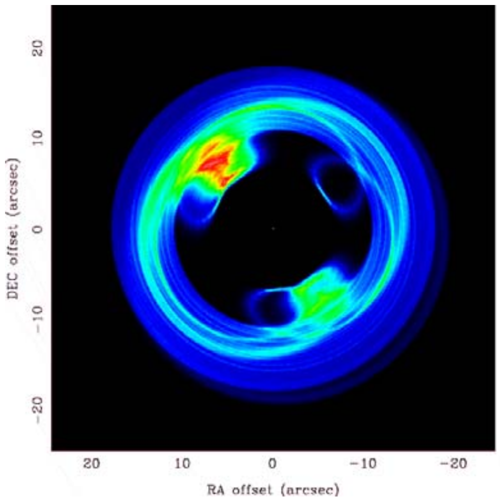


Ell. Gaussian Ell. Gaussian

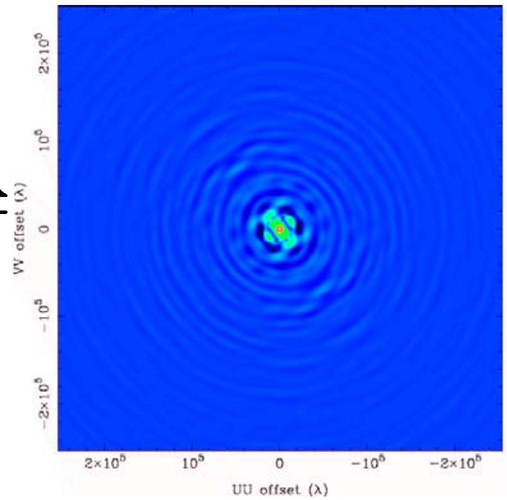
orientations are orthogonal in the (x,y) and (u,v) planes

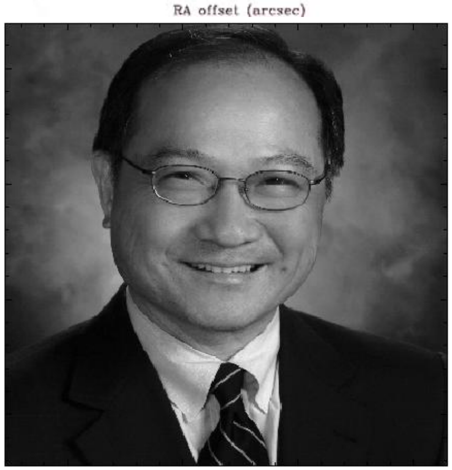
2D Fourier Transform Pairs

T(x,y) Amp{V(u,v)}

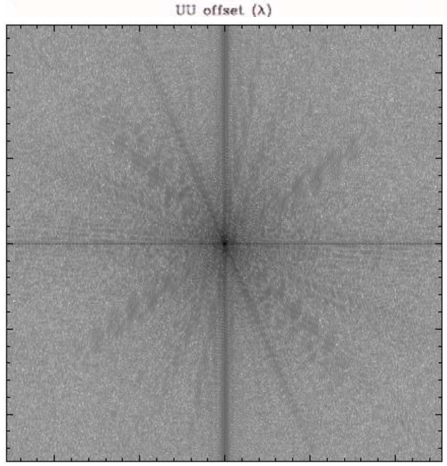


⇌





⇌

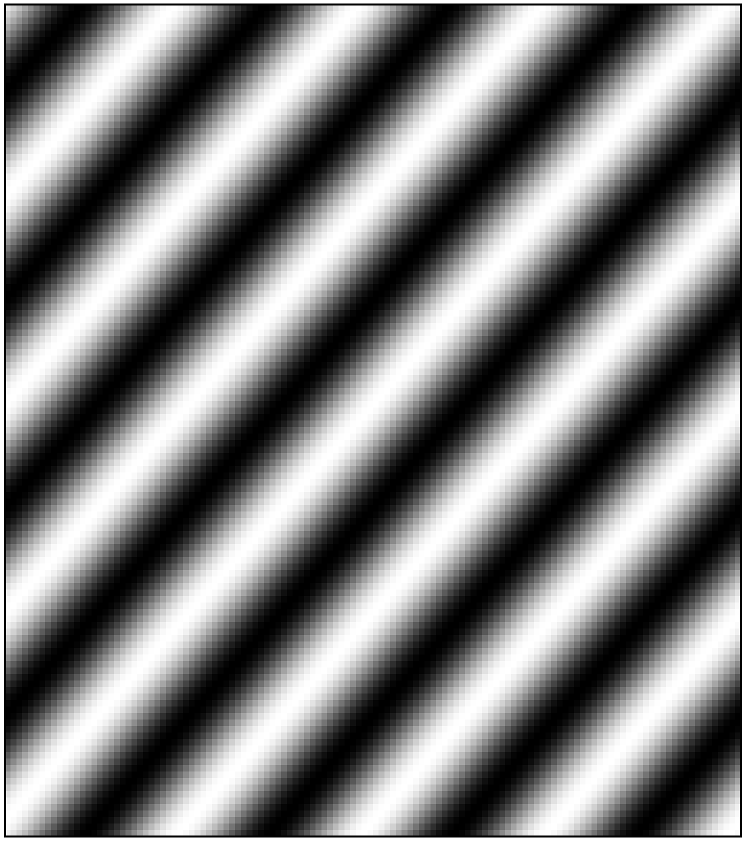


structure on many scales

- T(x,y) is real, but V(u,v) is complex (in general)
  - Real and Imaginary
  - Amplitude and Phase
  - Amplitude tells "how much" of a certain frequency component, Phase tells "where"
  - V(-u,-v) = V*(u,v) where * is complex conjugation (Hermitian)
- V(u=0,v=0) $\rightarrow$ integral of T(x,y)dxdy = total flux

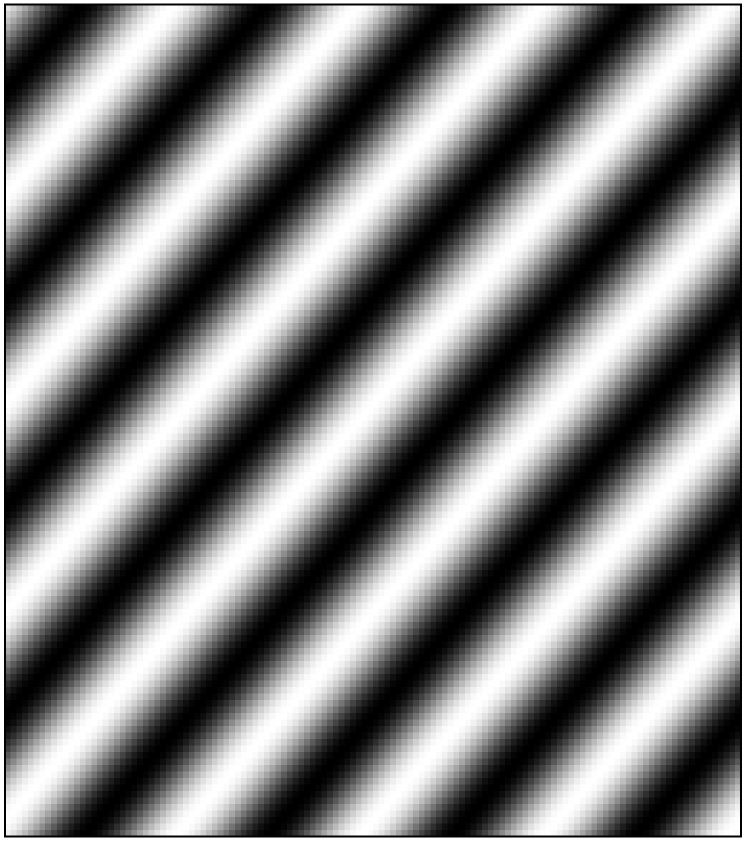
Visibility and Sky Brightness

$$V(u,v) = \int\int T(x,y) e^{2\pi i(ux+vy)} dx dy$$



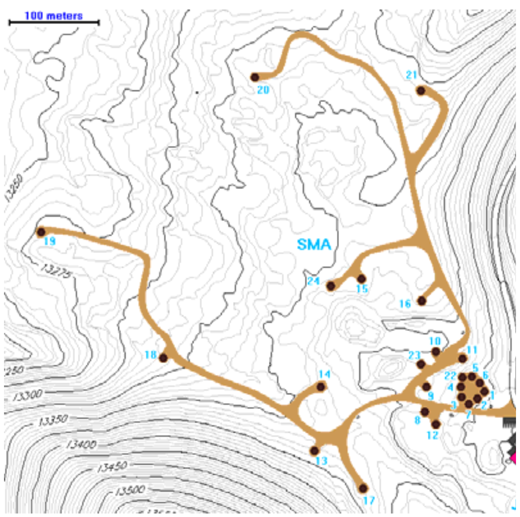
Visibility and Sky Brightness

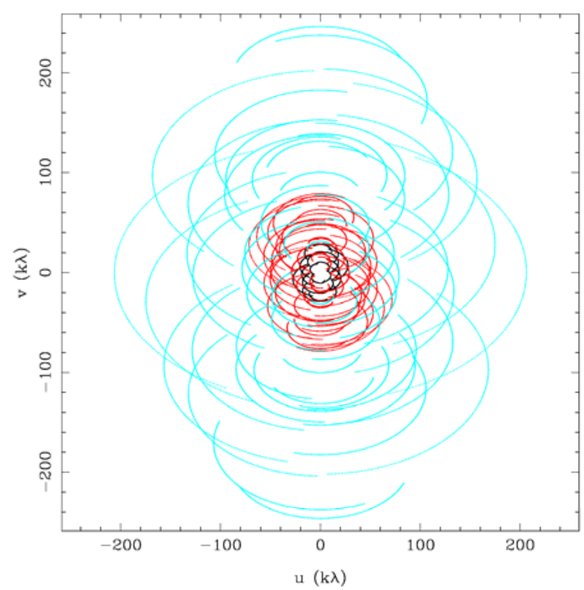
$$V(u, v) = \int \int T(x, y) e^{2\pi i (ux + vy)} dx dy$$



Aperture Synthesis

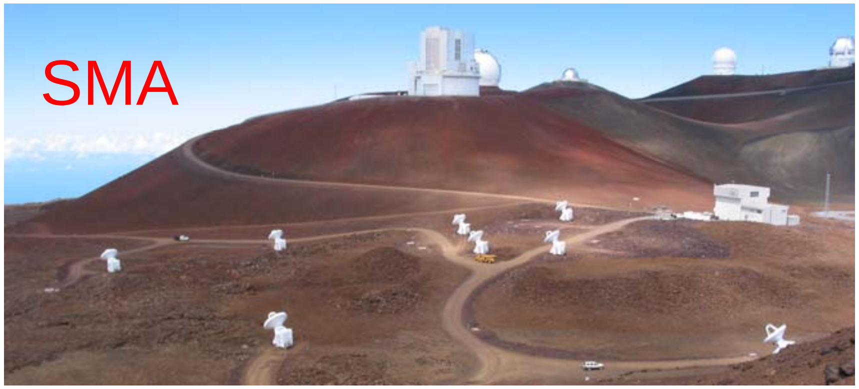
- sample V(u,v) at enough points to synthesis the equivalent large aperture of size (u_{max},v_{max})
  - 1 pair of telescopes → 1 (u,v) sample at a time
  - N telescopes → number of samples = N(N-1)/2
  - reconfigure physical layout of N telescopes for more
  - fill in (u,v) plane by making use of Earth rotation (Sir Martin Ryle, 1974 Nobel Prize in Physics)





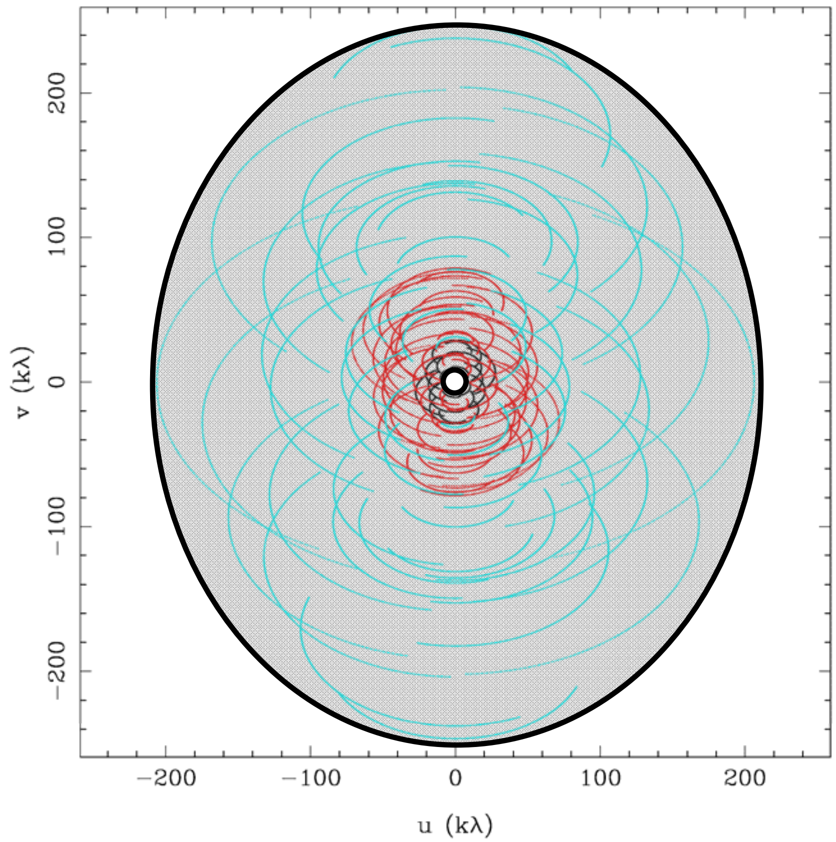
3 configurations
of 8 SMA antennas
at 345 GHz

Aperture Synthesis Telescopes



Imaging: (u,v) plane Sampling

- in aperture synthesis, V(u,v) samples are limited by number of telescopes, and Earth-sky geometry



- high spatial frequencies
  - maximum angular resolution
- low spatial frequencies
  - extended structures invisible
- irregular within high/low limits
  - sampling theorem violated
  - information lost

Formal Description

- sample Fourier domain at discrete points

  $B(u, v) = \sum_k (u_k, v_k)$

- the inverse Fourier transform is

  $T^D(x, y) = FT^{-1}\{B(u, v) \times V(u, v)\}$

- the convolution theorem tells us

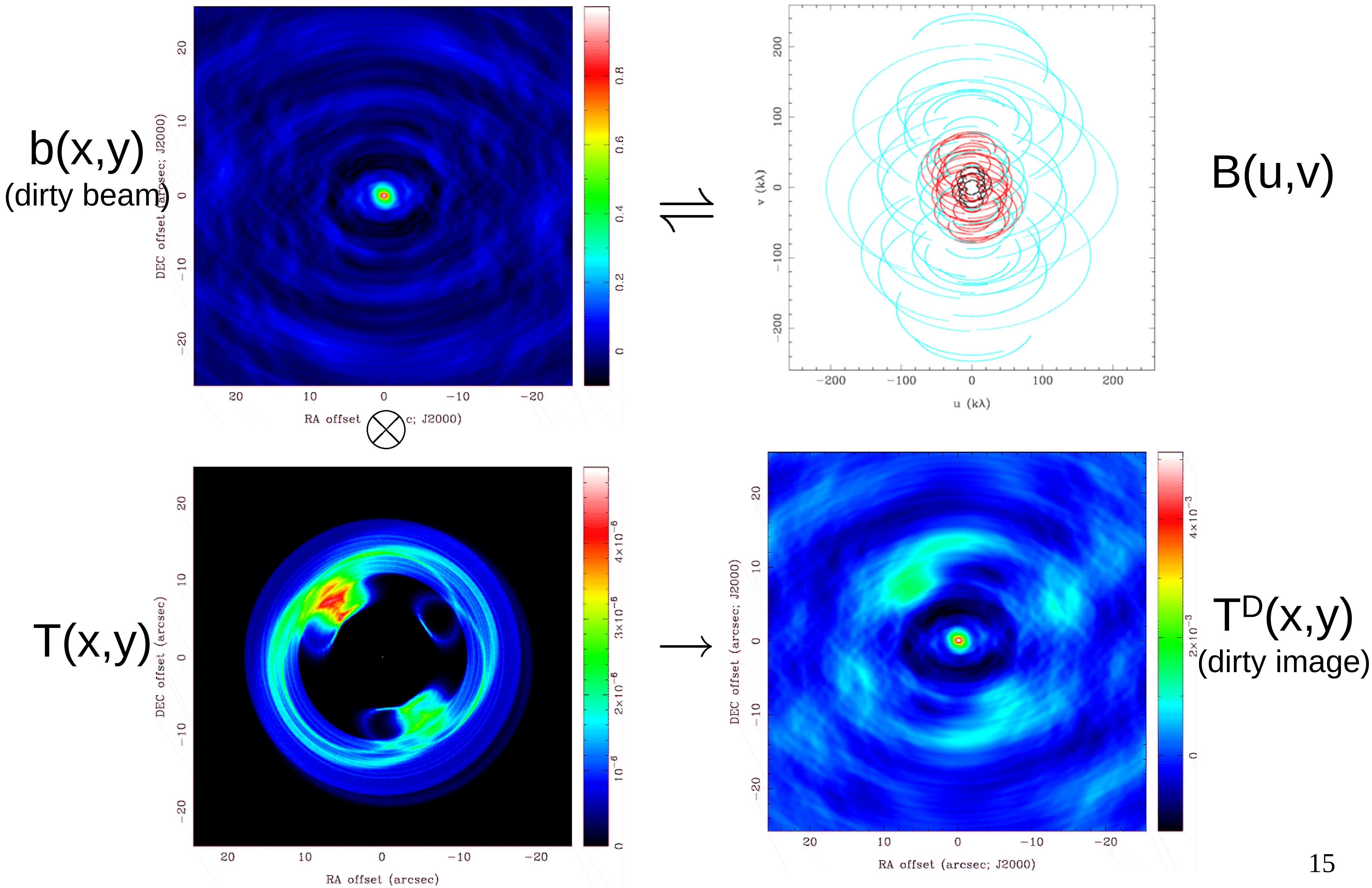
  $T^D(x, y) = b(x, y) \otimes T(x, y)$

  where $b(x, y) = FT^{-1}\{B(u, v)\}$ (the point spread function)

Fourier transform of sampled visibilities yields the true sky brightness convolved with the point spread function

(the “dirty image” is the true image convolved with the “dirty beam”)

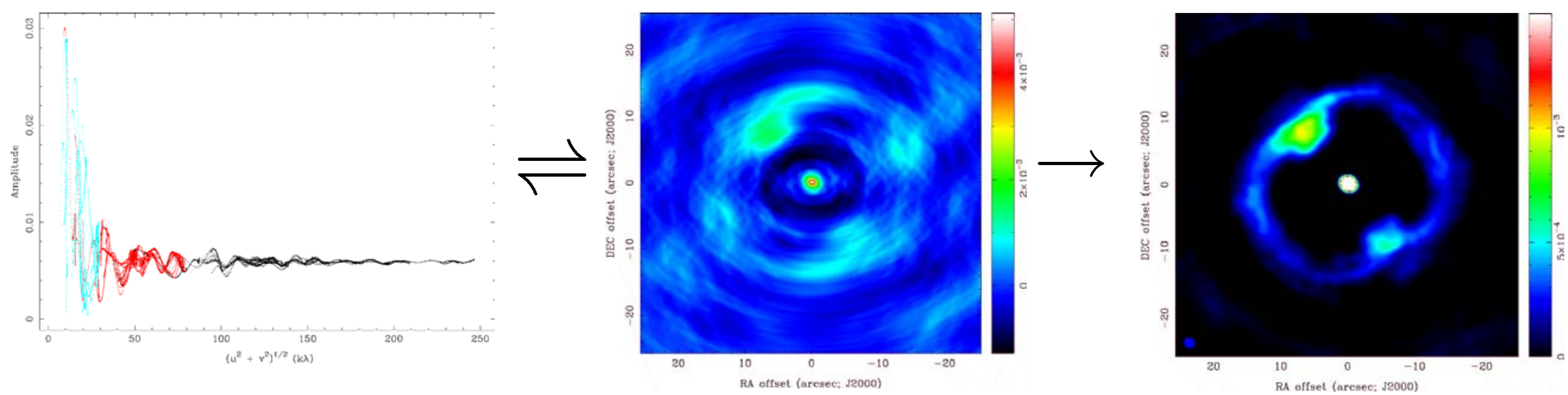
Dirty Beam and Dirty Image



How to analyze interferometer data?

- uv plane analysis
  - best for “simple” sources, e.g. point sources, disks
- image plane analysis
  - Fourier transform V(u,v) samples to image plane, get $T^D(x,y)$
  - but difficult to do science on dirty image
  - deconvolve b(x,y) from $T^D(x,y)$ to determine (model of) T(x,y)

visibilities → dirty image → sky brightness



Fourier transform deconvolve

Details of the Dirty Image

- Fourier Transform
  - Fast Fourier Transform (FFT) much faster than simple Fourier summation, O(NlogN) for 2^N x 2^N image
  - FFT requires data on regularly spaced grid
  - aperture synthesis observations not on a regular grid…

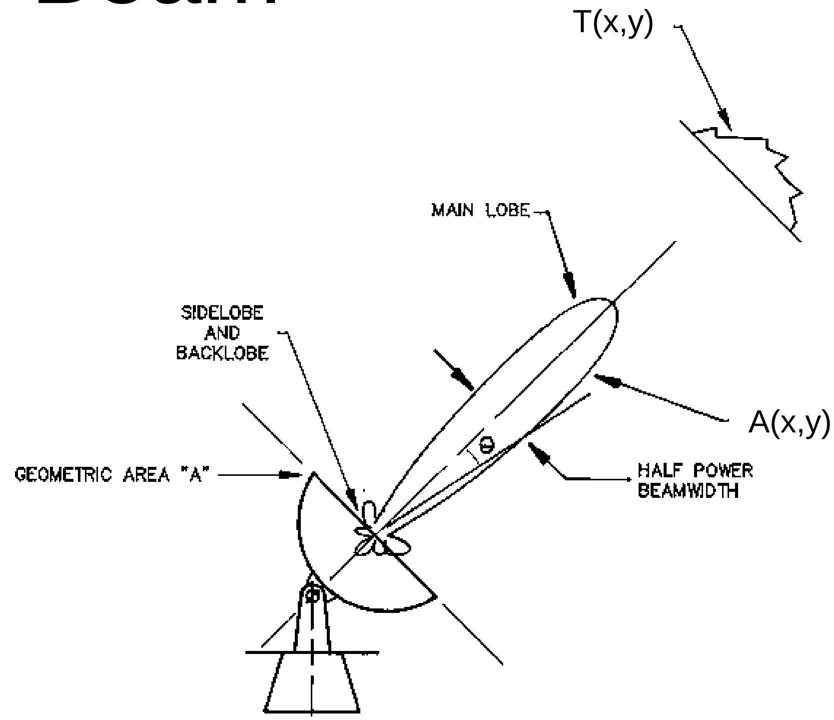
- "Gridding" is used to resample V(u,v) for FFT
  - customary to use a convolution technique
    - visibilities are noisy samples of a smooth function
    - nearby visibilities not independent
  - use special ("Spheroidal") functions with nice properties
    - fall off quickly in (u,v) plane (not too much smoothing)
    - fall off quickly in image plane (avoid aliasing)

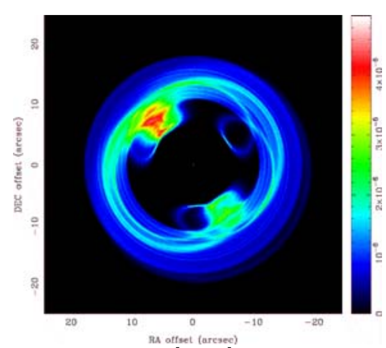
$$V^G(u,v) = V(u,v)B(u,v) \otimes G(u,v) \rightleftharpoons T^D(x,y)g(x,y)$$

Primary Beam

- A telescope does not have uniform response across the entire sky
  - main lobe approximately Gaussian, fwhm ~1.2λ/D (where D is ant diameter) = “primary beam”
  - limited field of view
  - sidelobes, error beam (sometimes important)

- primary beam response modifies sky brightness: $T(x,y) \rightarrow A(x,y)T(x,y)$
  - correct with division by A(x,y) in image plane





T(x,y)

large A(x,y)

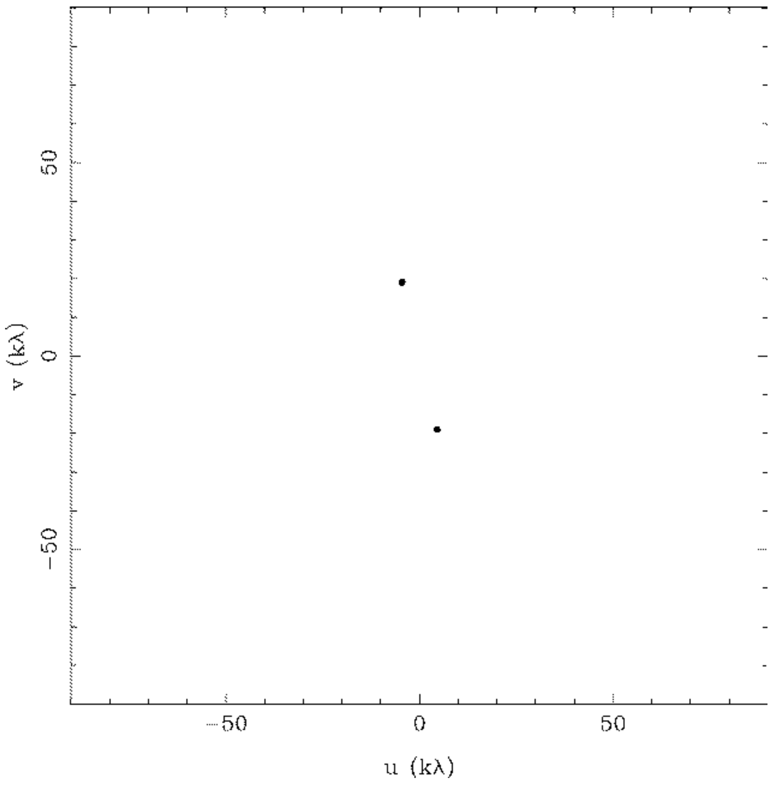
small A(x,y)

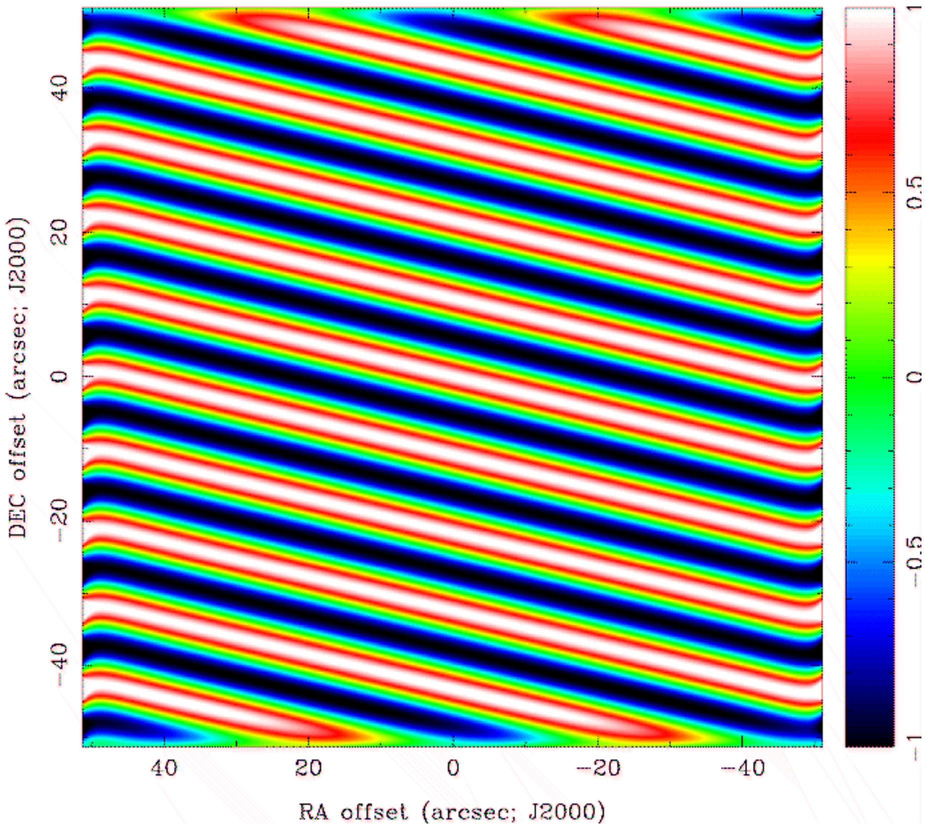
Pixel Size and Image Size

- pixel size
  - should satisfy sampling theorem for the longest baselines, $\Delta x < 1/2\ u_{max}$, $\Delta y < 1/2\ v_{max}$
  - in practice, 3 to 5 pixels across the main lobe of the dirty beam (to aid deconvolution)
- image size
  - natural resolution in (u,v) plane samples FT{A(x,y)}, implies image size 2x primary beam
  - if there are bright sources in the sidelobes of A(x,y), then they will be aliased into the image (need to make a larger image)

Dirty Beam Shape and N Antennas

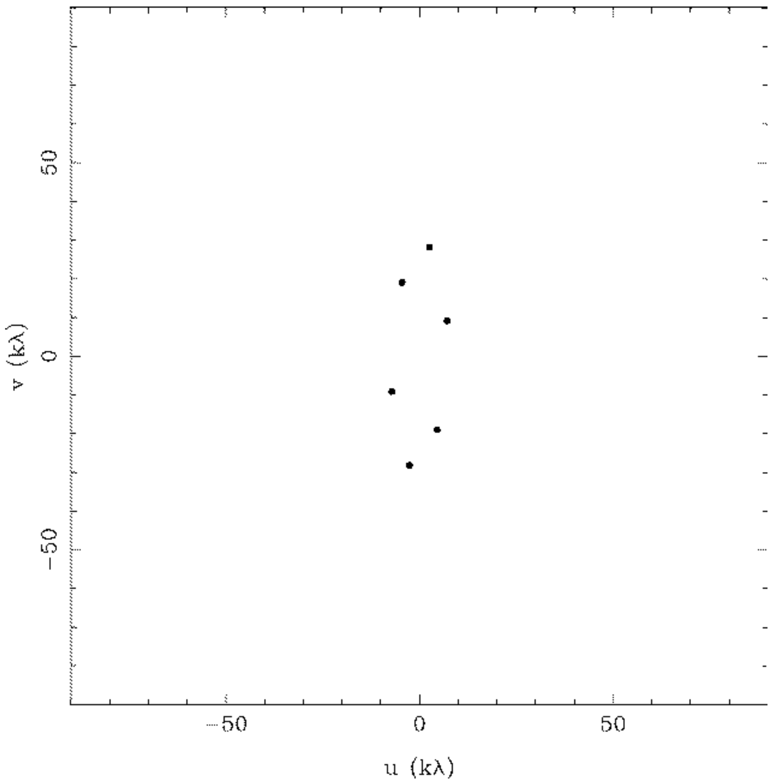
2 Antennas

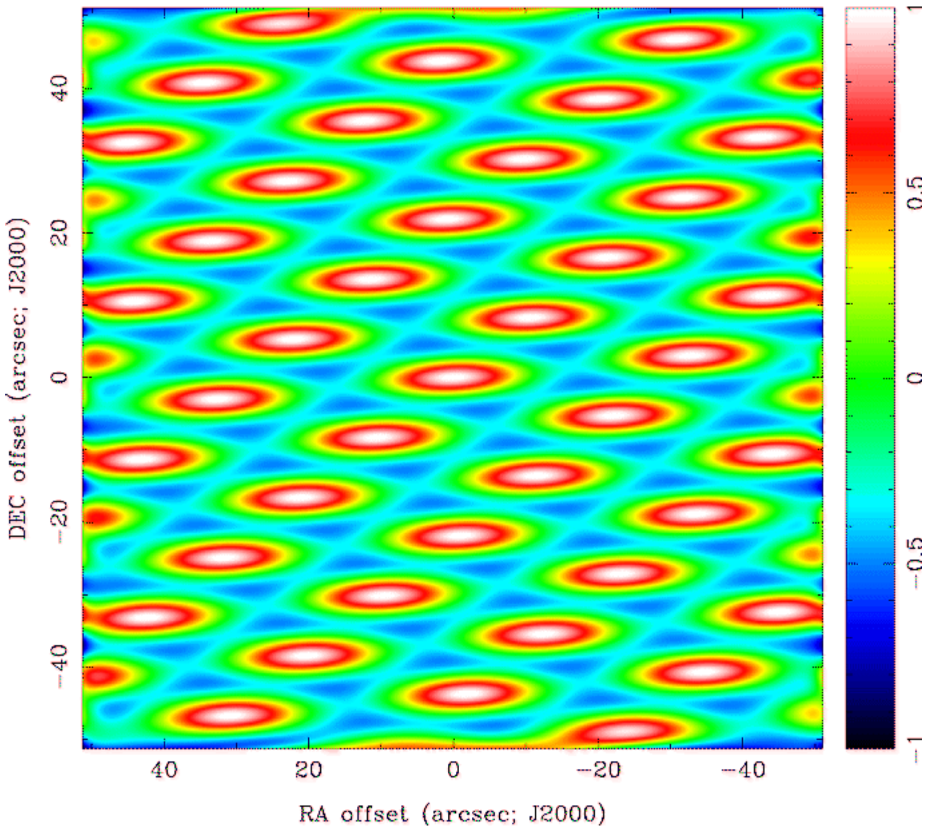




Dirty Beam Shape and N Antennas

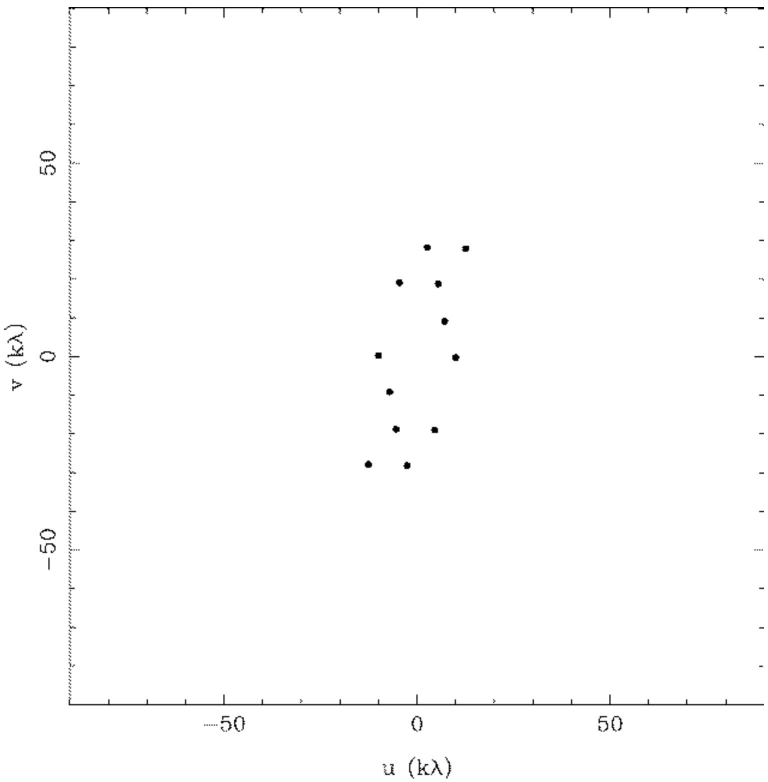
3 Antennas

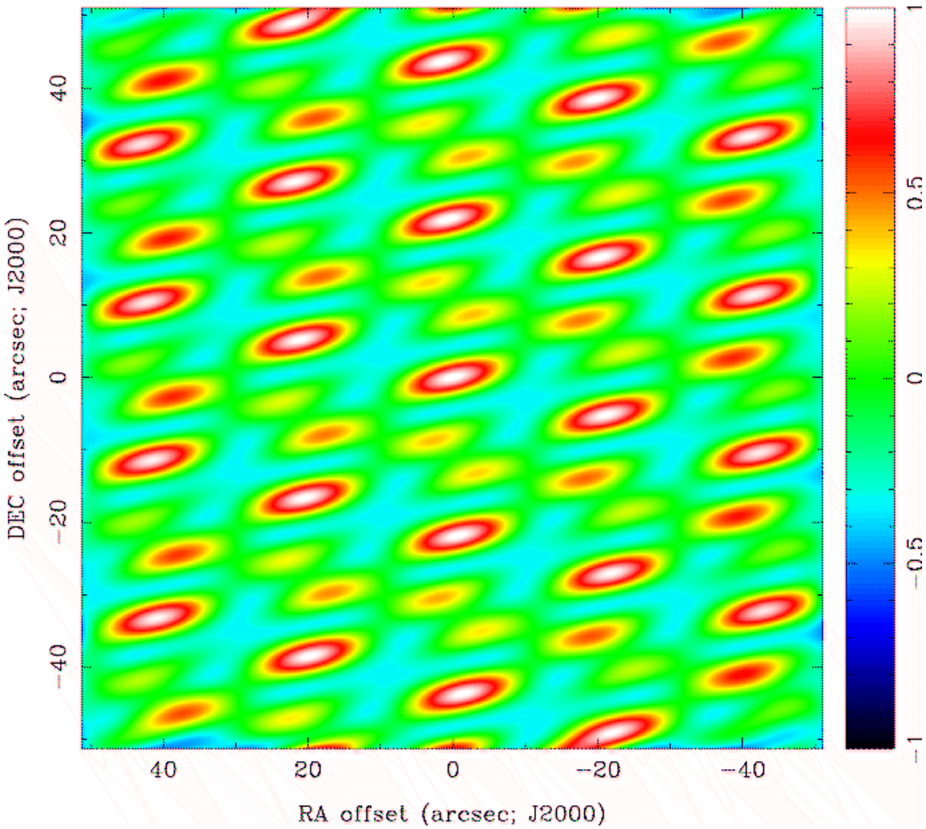




Dirty Beam Shape and N Antennas

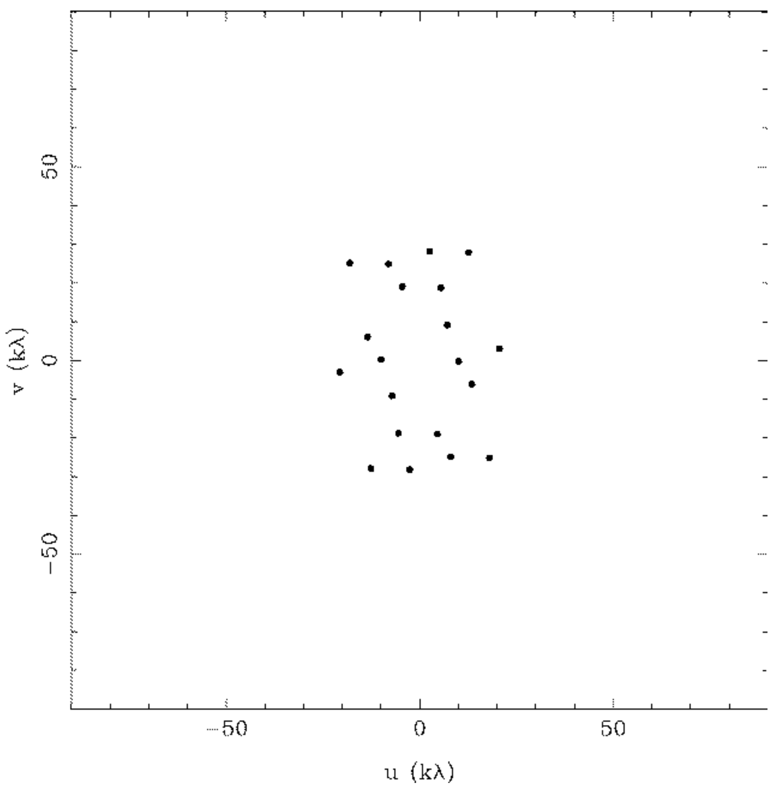
4 Antennas

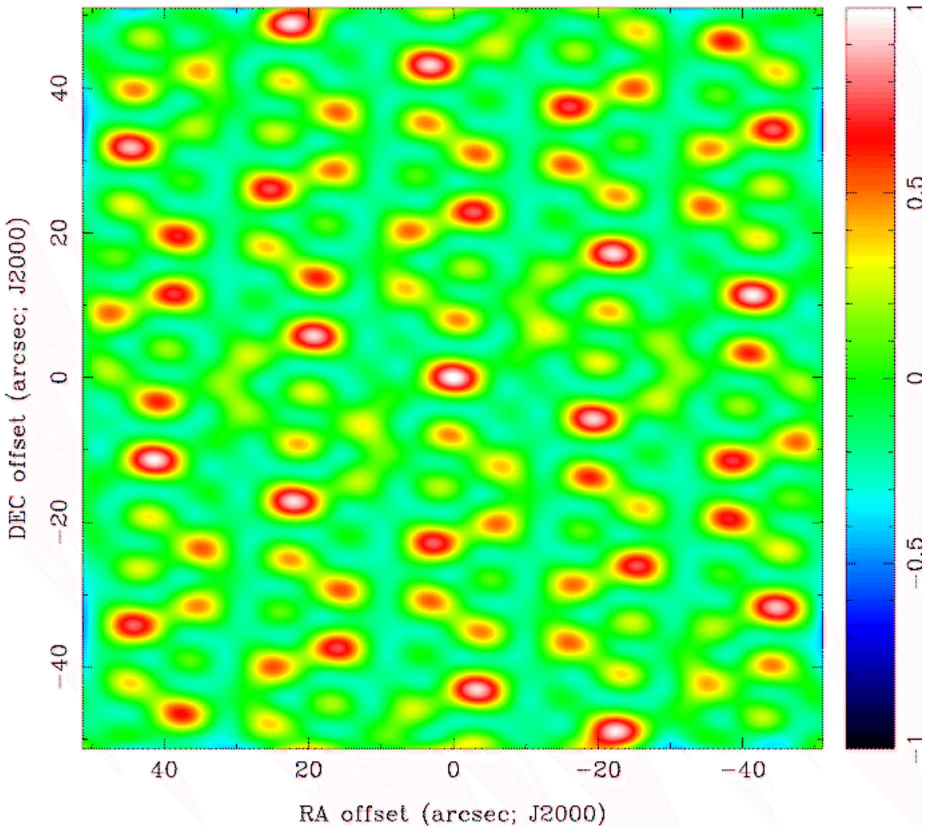




Dirty Beam Shape and N Antennas

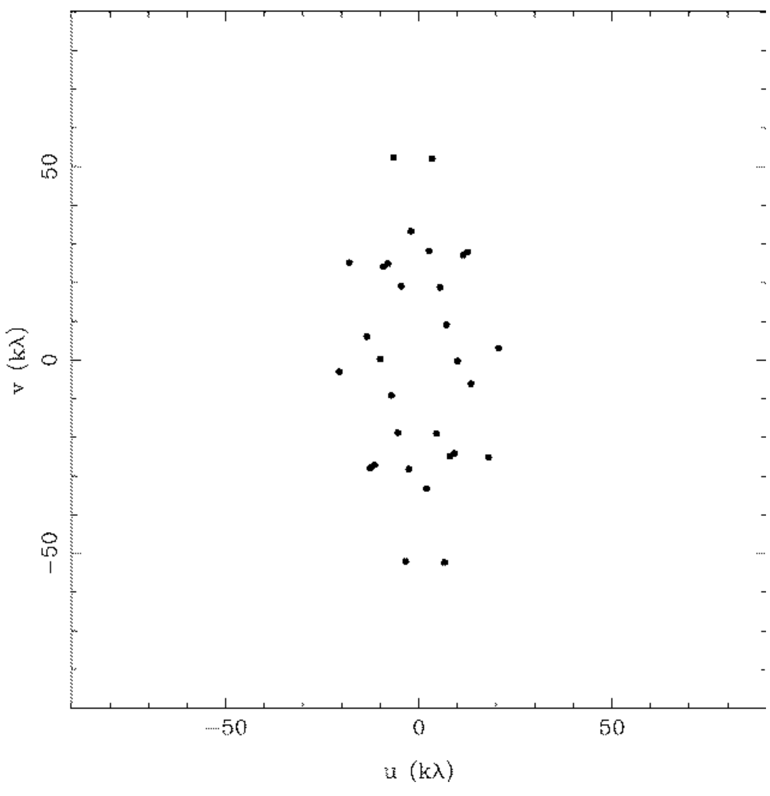
5 Antennas

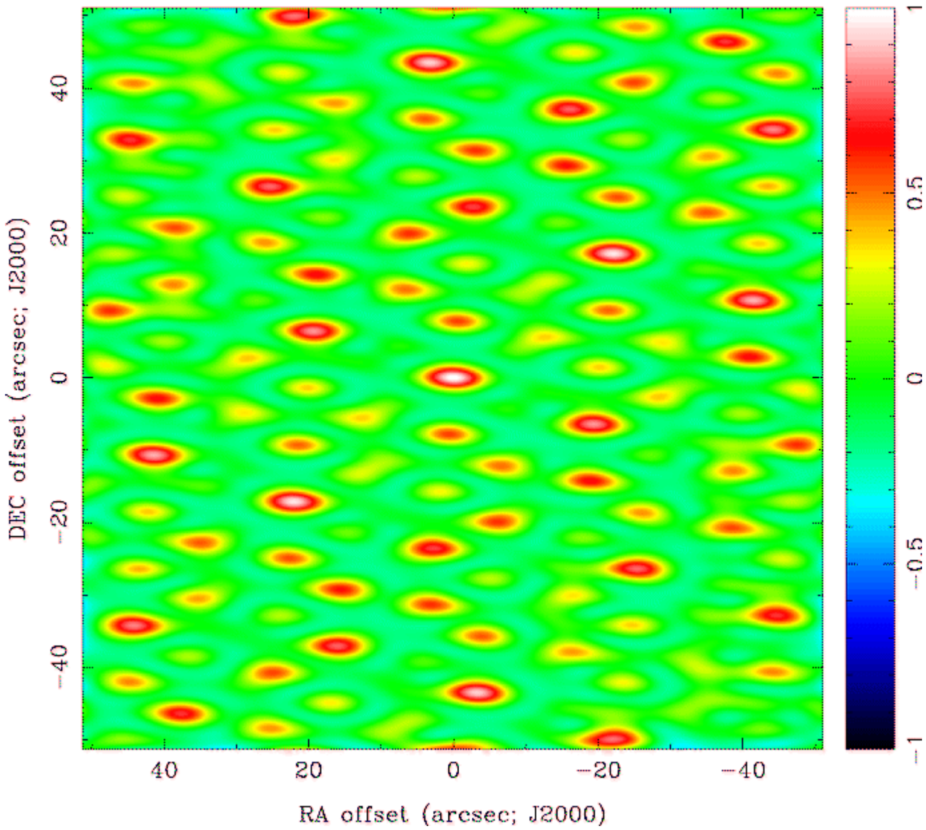




Dirty Beam Shape and N Antennas

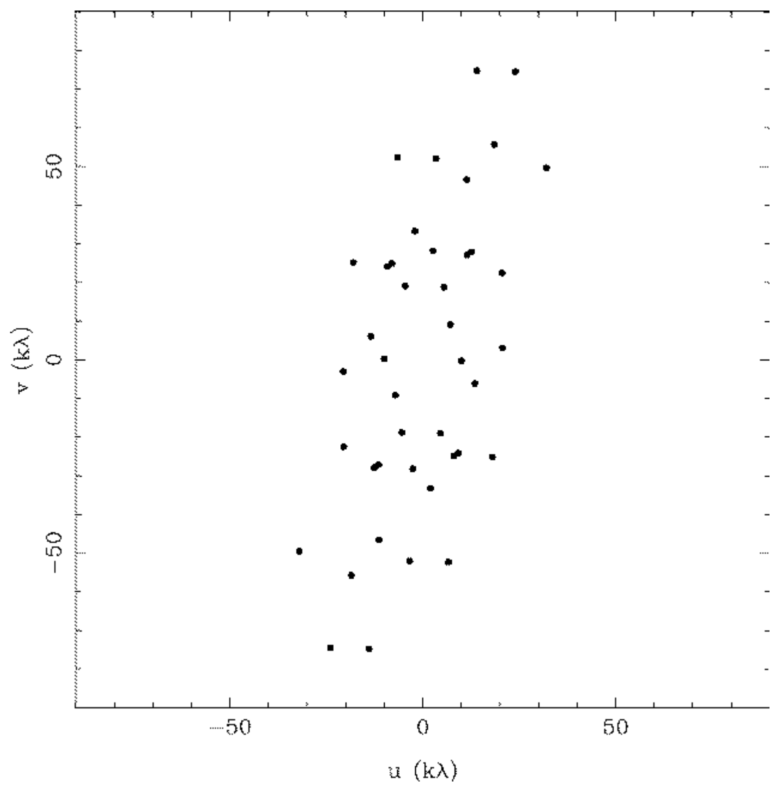
6 Antennas

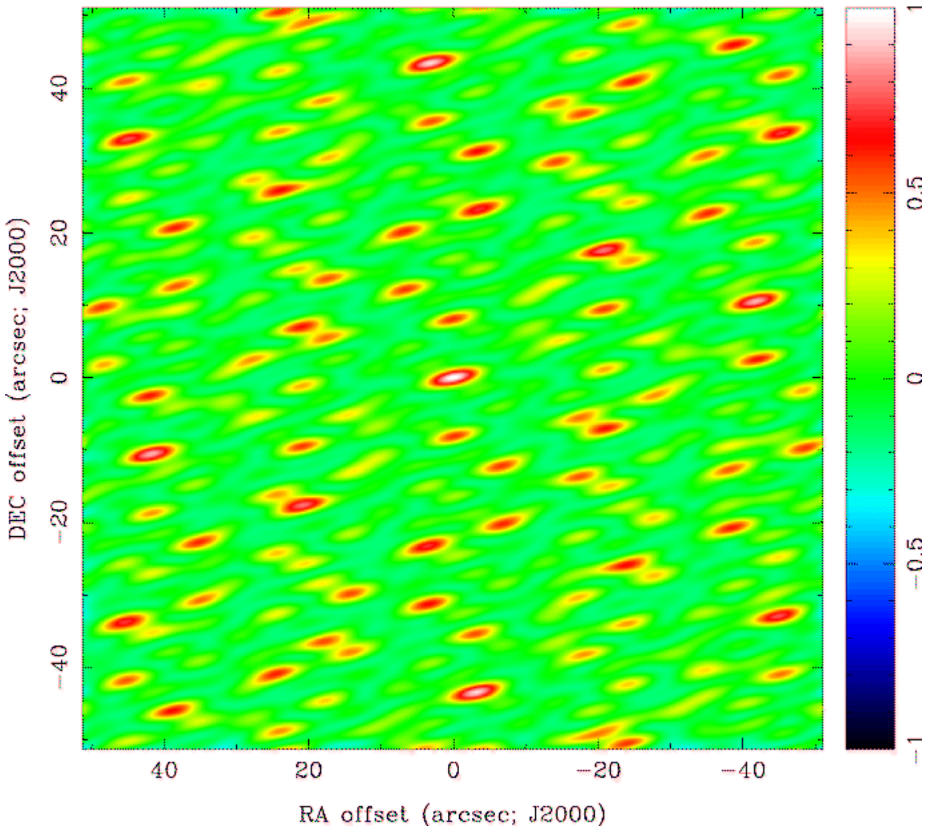




Dirty Beam Shape and N Antennas

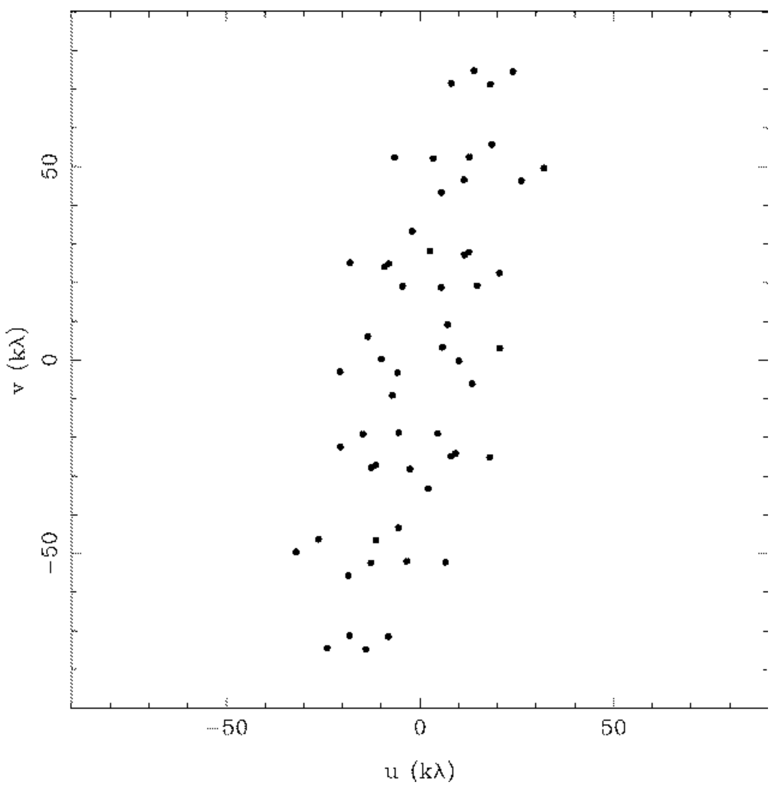
7 Antennas

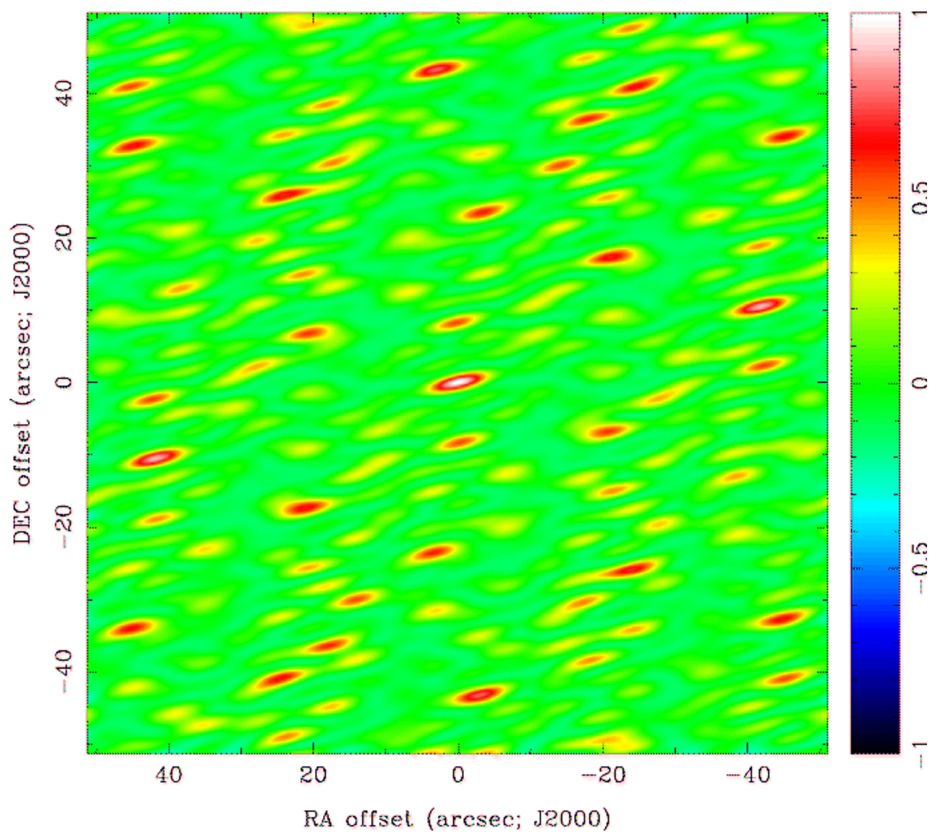




Dirty Beam Shape and N Antennas

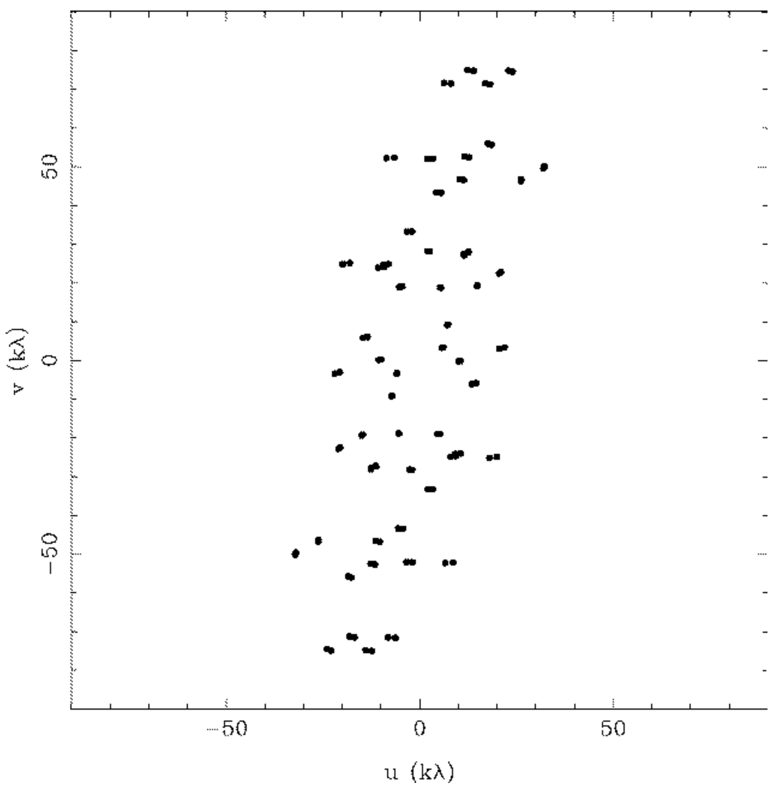
8 Antennas

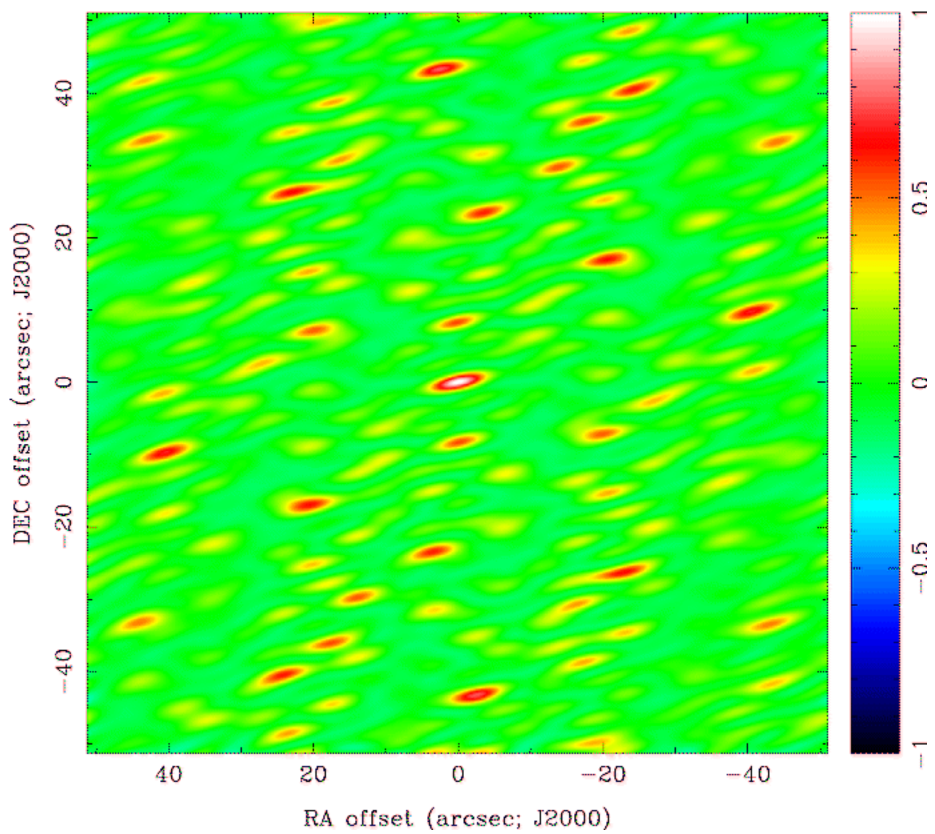




Dirty Beam Shape and Super Synthesis

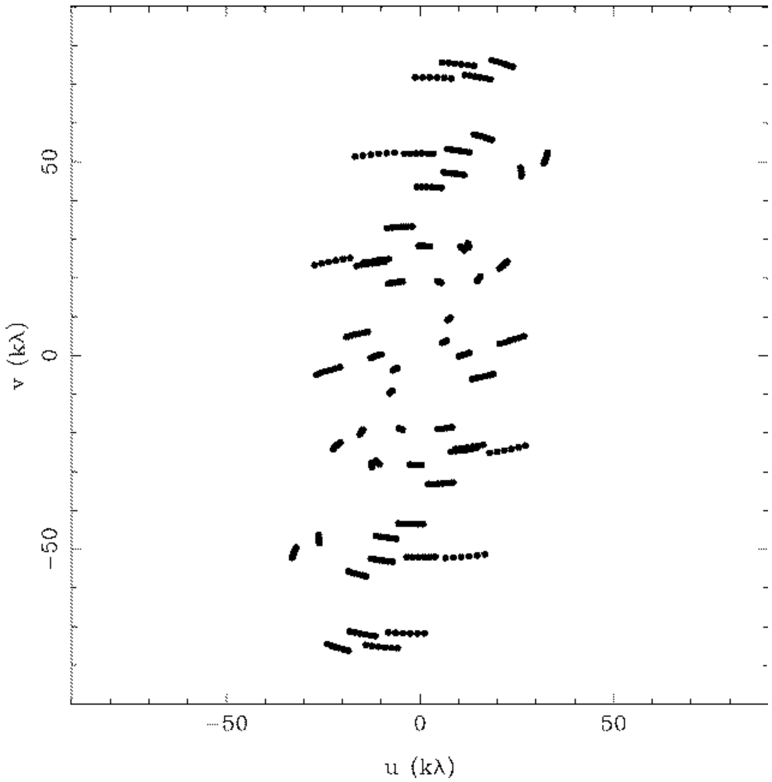
8 Antennas x 2 samples

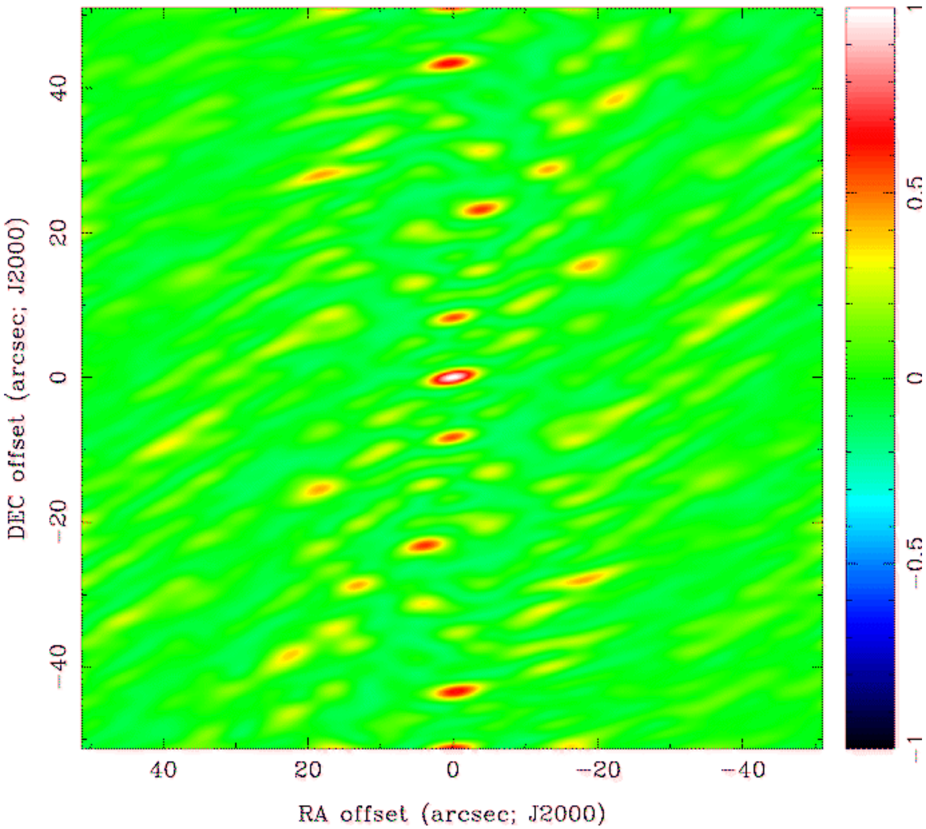




Dirty Beam Shape and Super Synthesis

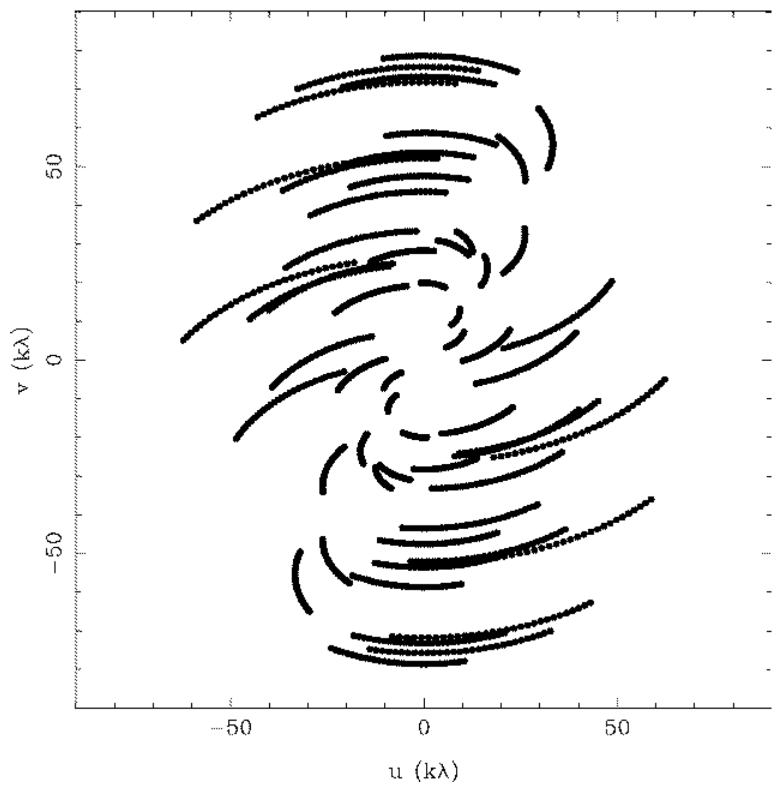
8 Antennas x 6 samples

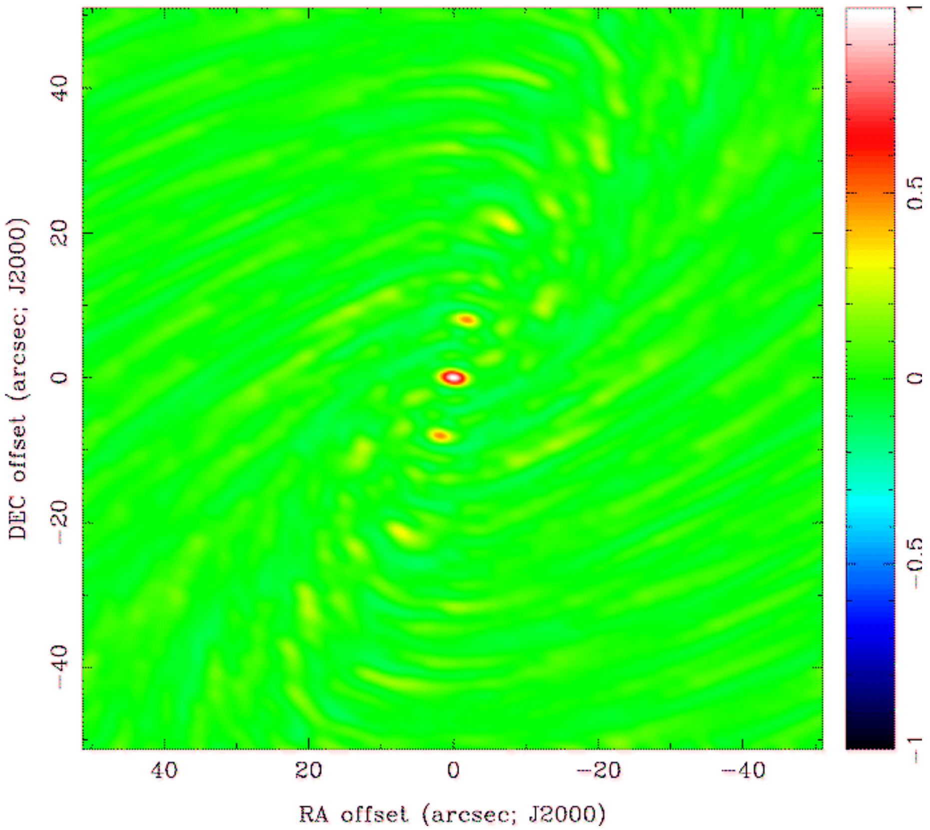




Dirty Beam Shape and Super Synthesis

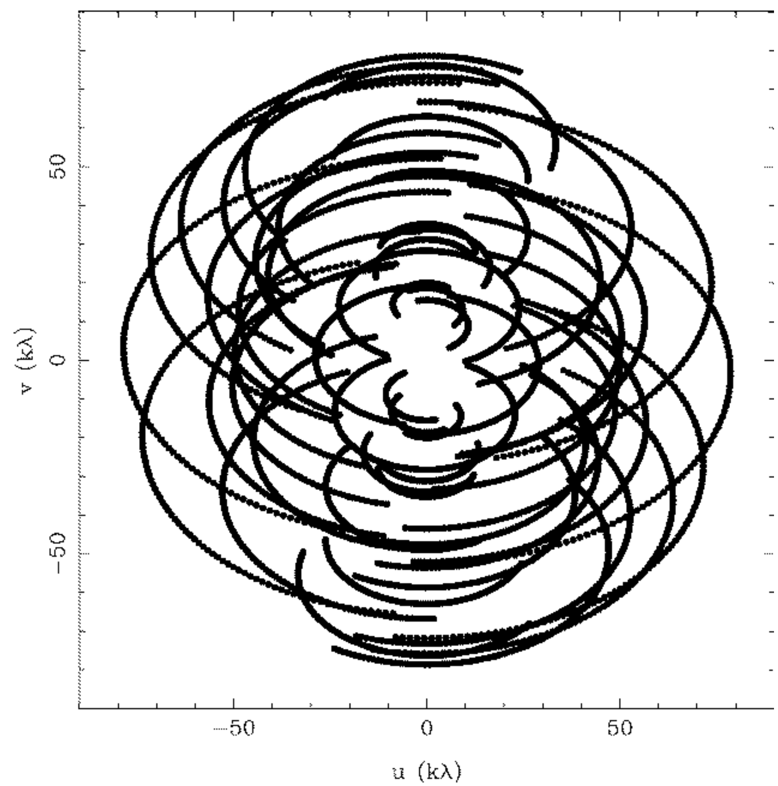
8 Antennas x 30 samples

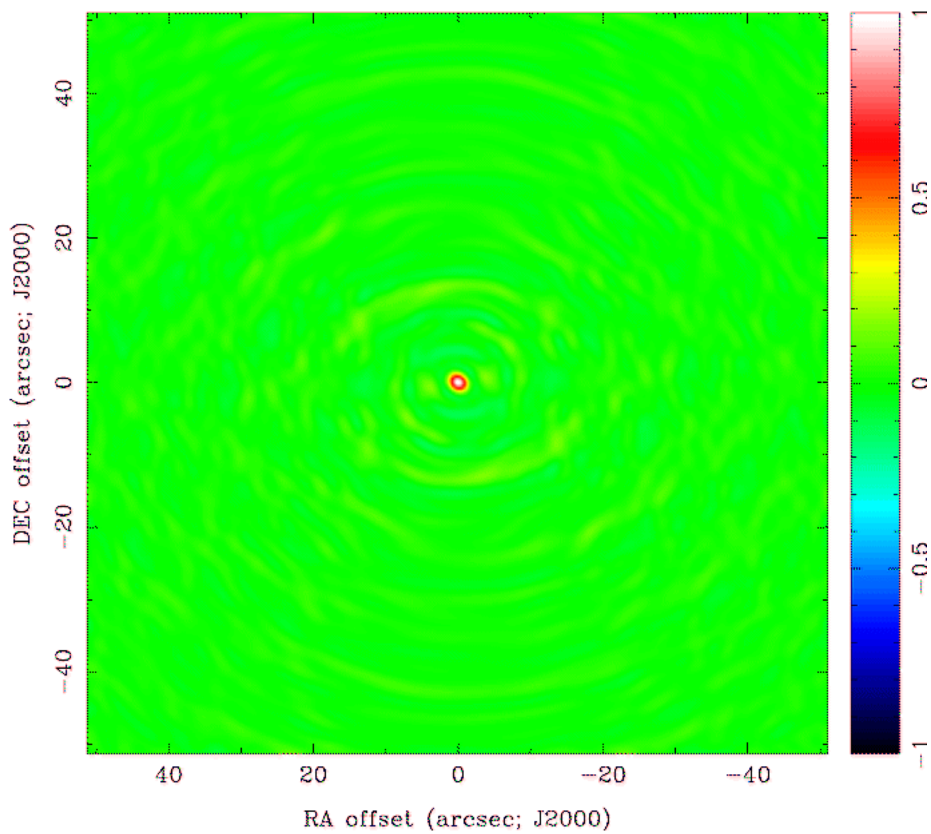




Dirty Beam Shape and Super Synthesis

8 Antennas x 107 samples



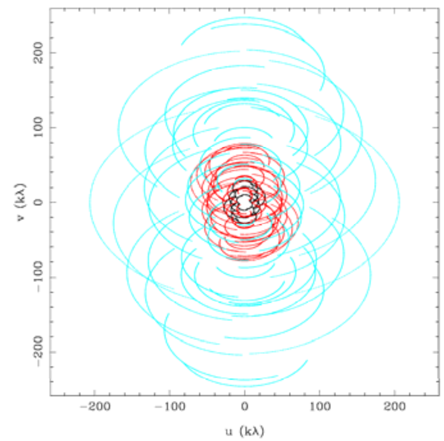


Dirty Beam Shape and Weighting

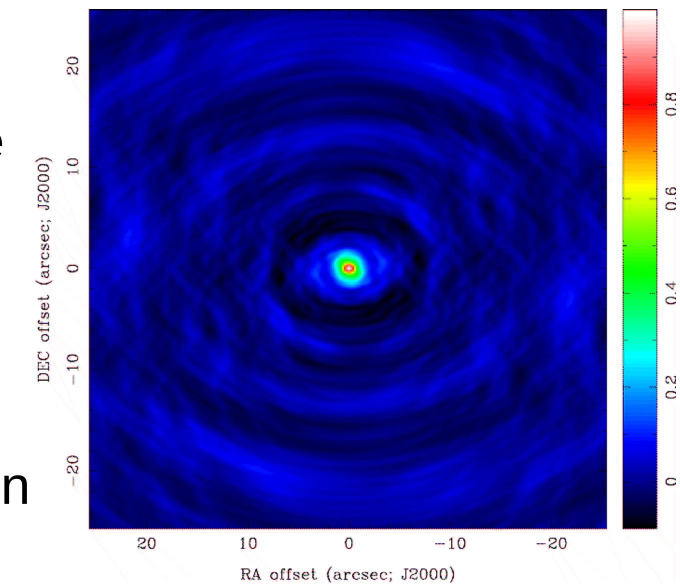
- introduce weighting function W(u,v)

  $$b(x,y) = FT^{-1}\{W(u,v)B(u,v)\}$$

  – W modifies sidelobes of dirty beam
  – W is also gridded for FFT

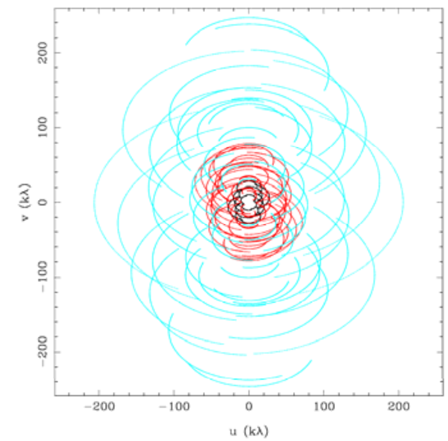


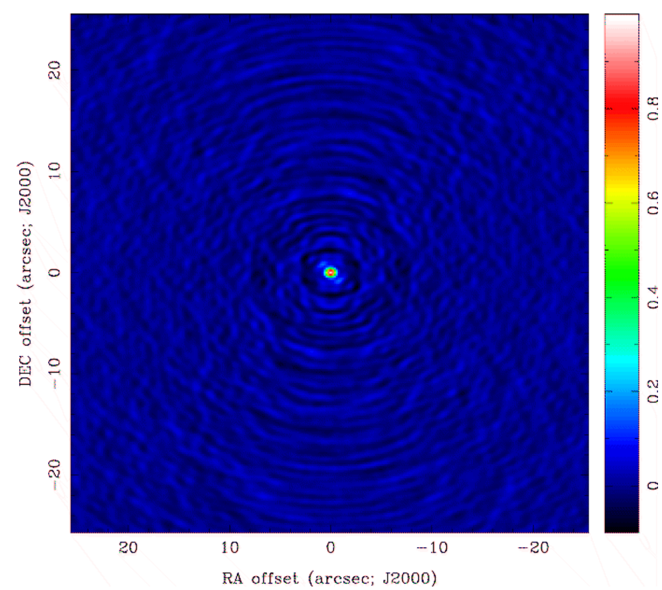
- "Natural" weighting
  – $W(u,v) = 1/\sigma^2(u,v)$ at points with data and zero elsewhere where $\sigma^2(u,v)$ is the noise variance of the (u,v) sample
  – maximizes point source sensitivity (lowest rms in image)
  – gives more weight to shorter baselines (larger spatial scales), degrades resolution



Dirty Beam Shape and Weighting

- "Uniform" weighting
  - W(u,v) is inversely proportional to local density of (u,v) points, so sum of weights in a (u,v) cell is a constant (or zero)
  - fills (u,v) plane more uniformly, so (outer) sidelobes are lower
  - gives more weight to long baselines and therefore higher angular resolution
  - degrades point source sensitivity (higher rms in image)
  - can be trouble with sparse sampling (cells with few data points have same weight as cells with many data points)



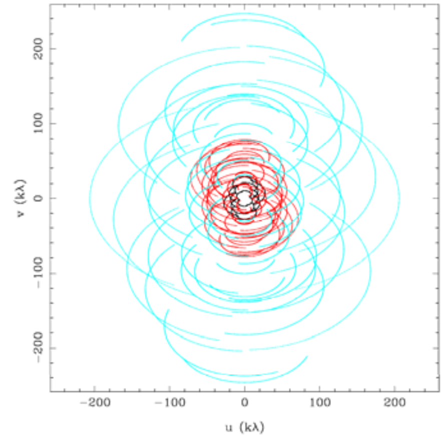


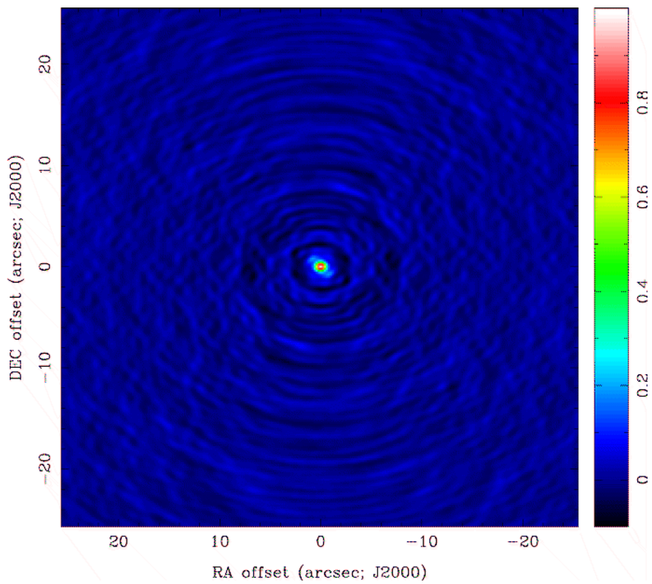
Dirty Beam Shape and Weighting

- "Robust" (Briggs) weighting
  - variant of "uniform" that avoids giving too much weight to cell with low natural weight
  - implementations differ, e.g. S_N is natural weight of a cell, S_t is a threshold

$$W(u,v) = \frac{1}{\sqrt{1+S_N^2/S_{thresh}^2}}$$

  - large threshold$\rightarrow$ natural weighting
  - small threshold $\rightarrow$ uniform weighting
  - parameter allows continuous variation between optimal angular resolution and optimal point source sensitivity



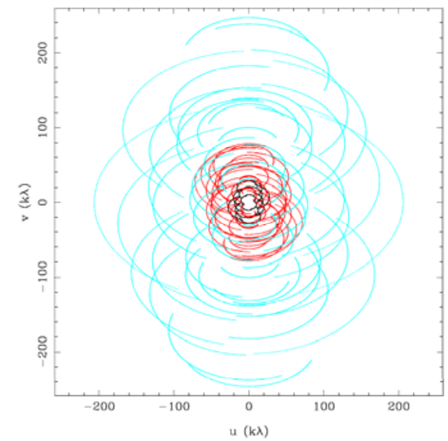


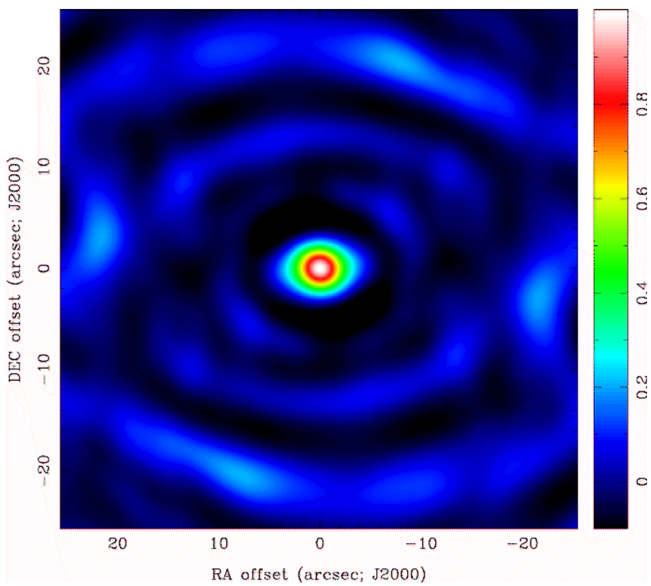
Dirty Beam Shape and Weighting

- "Tapering"
  - apodize the (u,v) sampling by a Gaussian

    $$W(u,v) = exp\left\{-\frac{(u^2+v^2)}{t^2}\right\}$$

    t = tapering parameter (in kλ; arcsec)
  - like smoothing in the image plane (convolution by a Gaussian)
  - gives more weight to shorter baselines, degrades angular resolution
  - degrades point source sensitivity but can improve sensitivity to extended structure
  - could use an elliptical Gaussian
  - limits to usefulness





Weighting and Tapering: Noise

Natural
1.7x1.4

σ=1.0

Robust
1.0x0.8

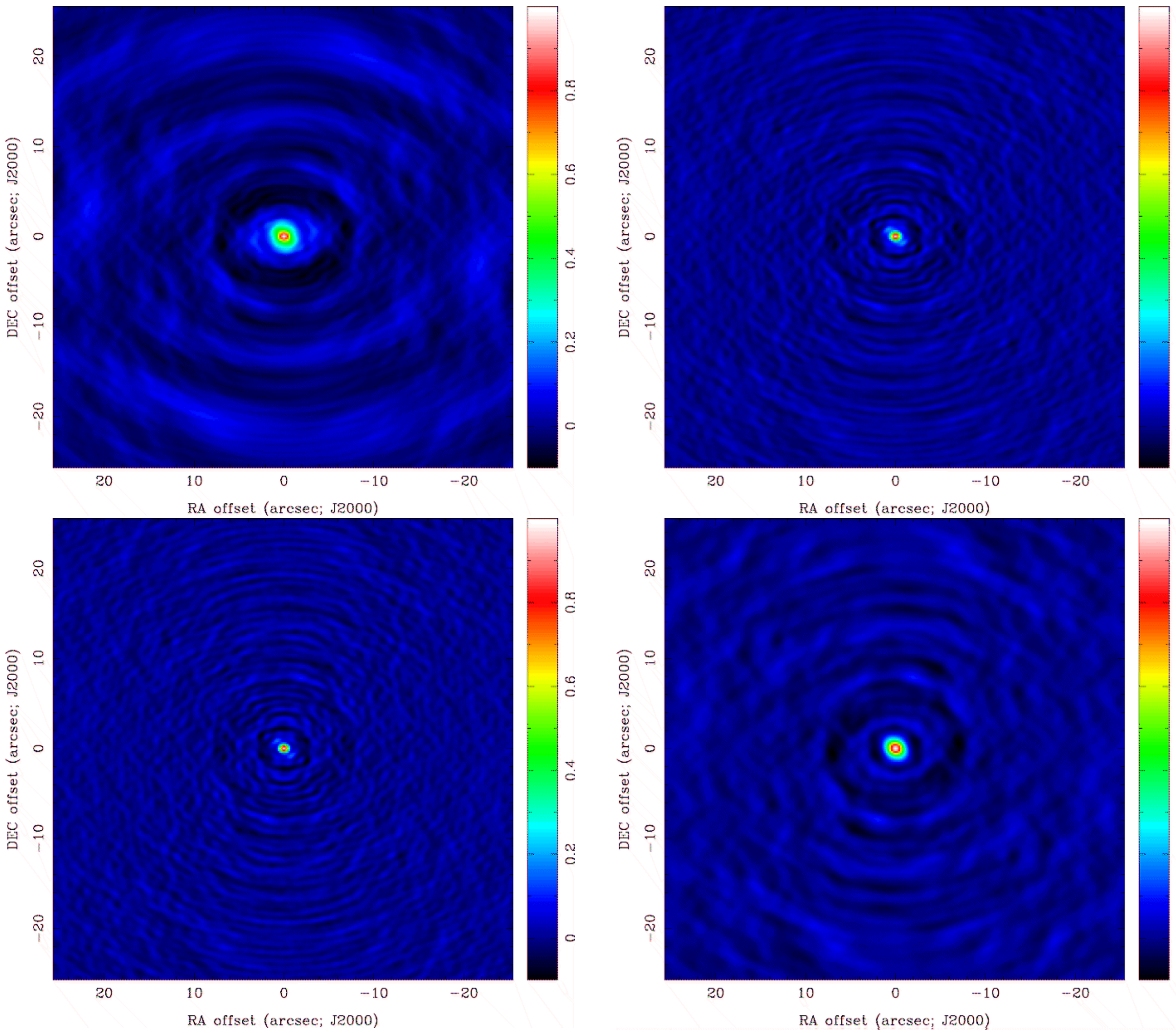
σ=1.28

Uniform
0.9x0.7

σ=1.58

Robust
+ Taper
1.7x1.4

σ=1.30



Weighting and Tapering: Summary

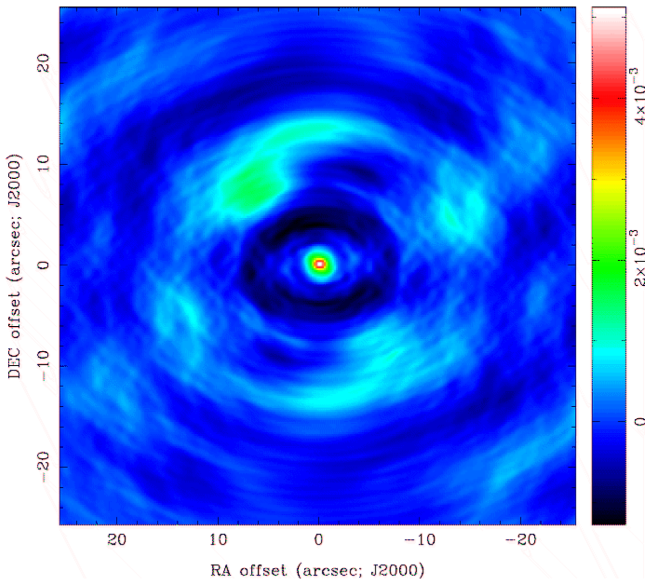
- imaging parameters provide a lot of freedom

	Robust/Uniform	Natural	Taper
Resolution	high	medium	low
Sidelobes	lower	higher	depends
Point Source Sensitivity	lower	maximum	lower
Extended Source Sensitivity	lower	medium	higher

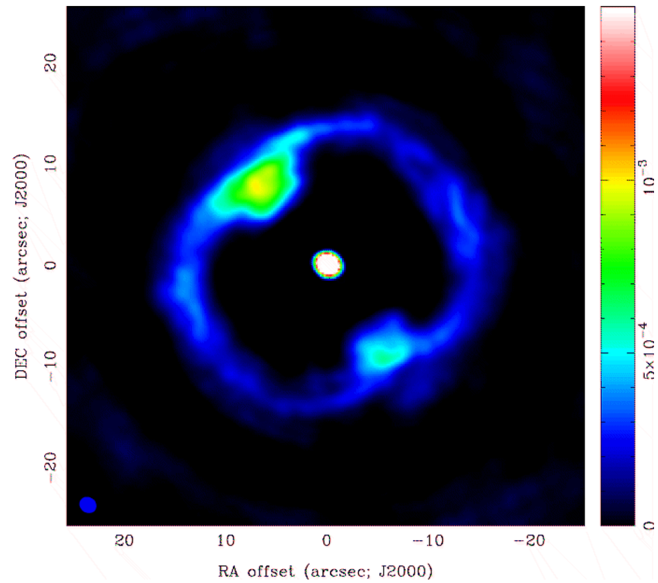
Deconvolution

- difficult to do science on dirty image
- deconvolve $b(x,y)$ from $T^D(x,y)$ to recover $T(x,y)$
- information is missing, so be careful!

dirty image



“CLEAN” image



Deconvolution Philosophy

- to keep you awake at night
  - $\exists$ an infinite number of T(x,y) compatible with sampled V(u,v), i.e. “invisible” distributions R(x,y) where $b(x,y) \otimes R(x,y) = 0$
    - no data beyond $u_{max}, v_{max} \rightarrow$ unresolved structure
    - no data within $u_{min}, v_{min} \rightarrow$ limit on largest size scale
    - holes between u_{min}, v_{min} and $u_{max}, v_{max} \rightarrow$ sidelobes
  - noise $\rightarrow$ undetected/corrupted structure in T(x,y)
  - no unique prescription for extracting optimum estimate of true sky brightness from visibility data
- deconvolution
  - uses non-linear techniques effectively interpolate/extrapolate samples of V(u,v) into unsampled regions of the (u,v) plane
  - aims to find a **sensible** model of T(x,y) compatible with data
  - requires *a priori* assumptions about T(x,y)

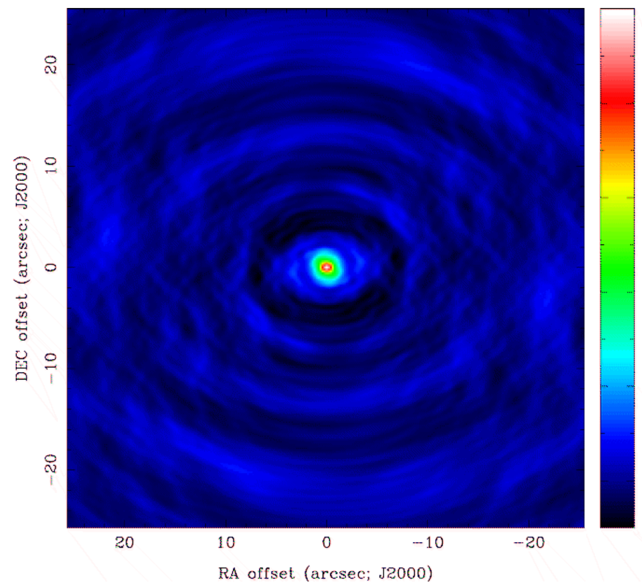
Deconvolution Algorithms

- most common algorithms in radio astronomy
  - CLEAN (Högbom 1974)
    - *a priori* assumption: T(x,y) is a collection of point sources
    - variants for computational efficiency, extended structure
  - Maximum Entropy (Gull and Skilling 1983)
    - *a priori* assumption: T(x,y) is smooth and positive
    - vast literature about the deep meaning of entropy (Bayesian)
  - hybrid approaches of these can be effective

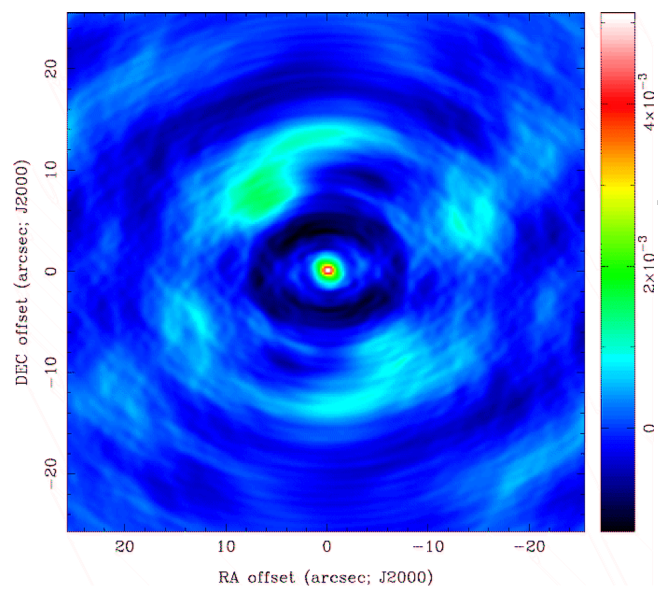
- deconvolution requires knowledge of beam shape and image noise properties (usually OK for aperture synthesis)
  - atmospheric seeing can modify effective beam shape
  - deconvolution process can modify image noise properties

Basic CLEAN Algorithm

1. Initialize
   - a *residual* map to the dirty map
   - a *Clean component* list to empty
2. Identify strongest feature in *residual* map as a point source
3. Add a fraction g (the loop gain) of this point source to the clean component list
4. Subtract the fraction g times b(x,y) from *residual* map
5. If stopping criteria not reached, goto step 2 (an interation)
6. Convolve *Clean component* (cc) list by an estimate of the main lobe of the dirty beam (the “Clean beam”) and add *residual* map to make the final “restored” image



b(x,y)



$T^D(x,y)$

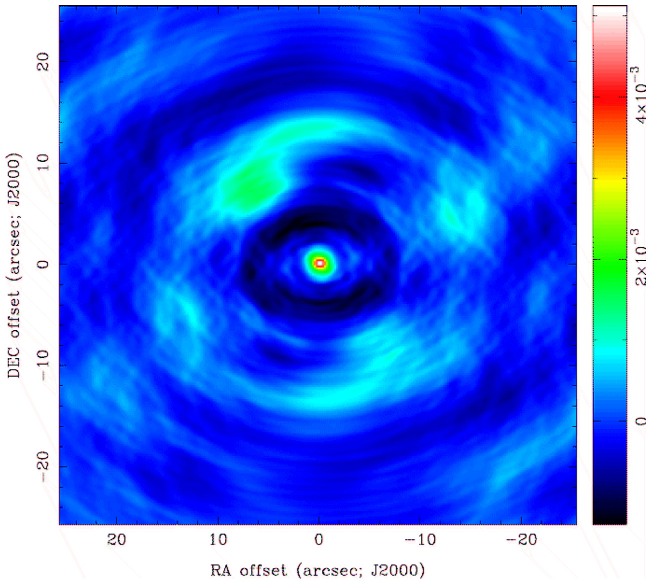
Basic CLEAN Algorithm (cont)

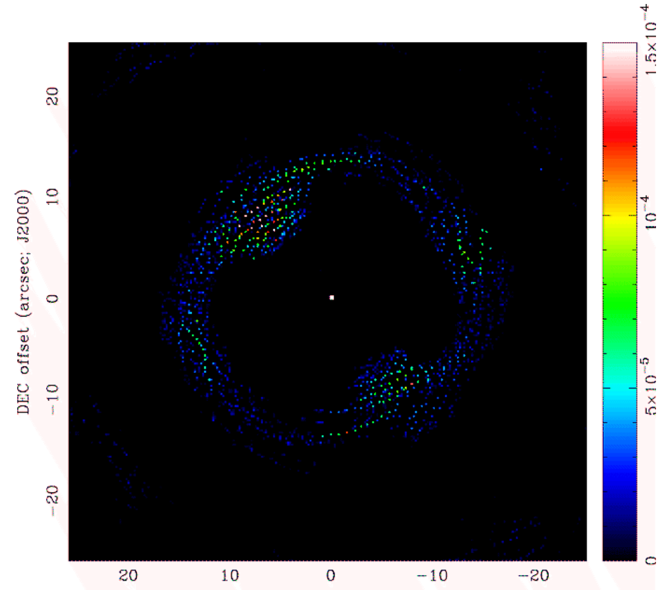
- stopping criteria
  - r*esidual* map max < multiple of rms (when noise limited)
  - r*esidual* map max < fraction of dirty map max (dynamic range limited)
  - max number of clean components reached
- loop gain: good results for g ~ 0.1 to 0.3
- easy to include *a priori* information about where to search for clean components (“clean boxes”)
  - very useful but potentially dangerous!

- Schwarz (1978): CLEAN is equivalent to a least squares fit of sinusoids (in the absense of noise)

CLEAN

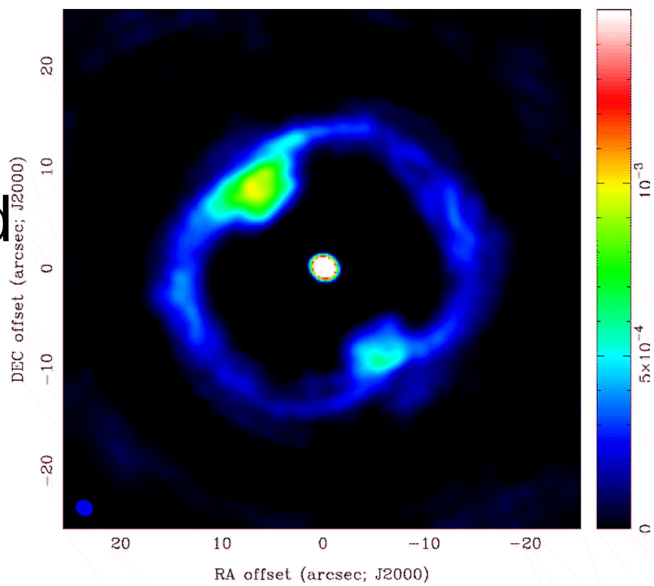
$T^D(x,y)$

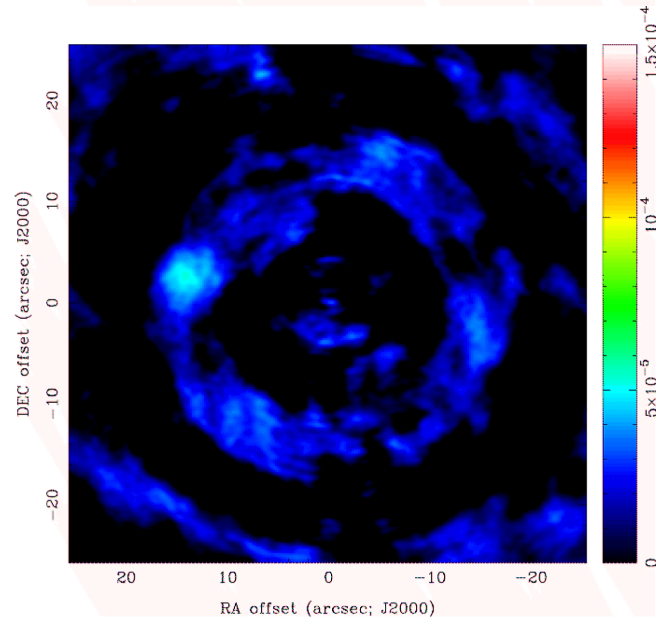




CLEAN model

restored image

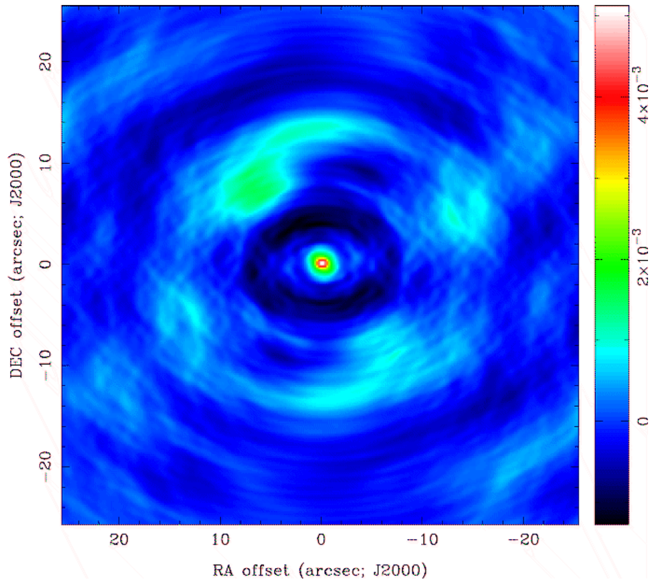


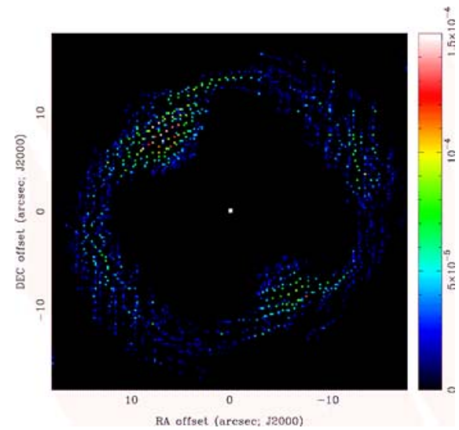


residual map

CLEAN with Box

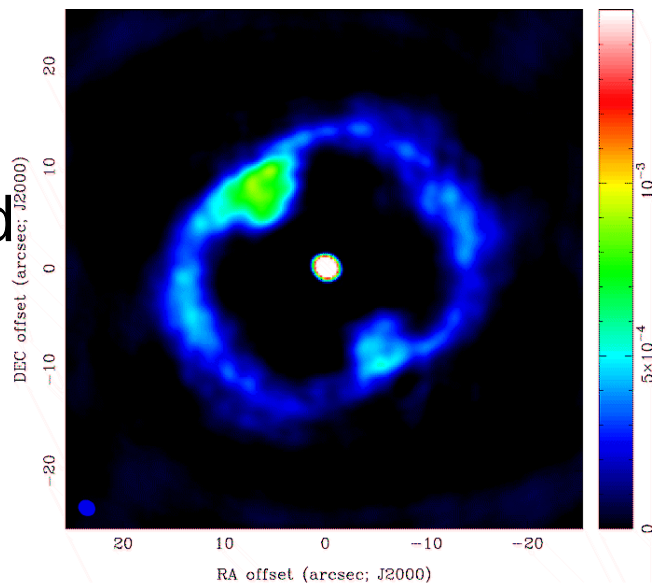
$T^D(x,y)$

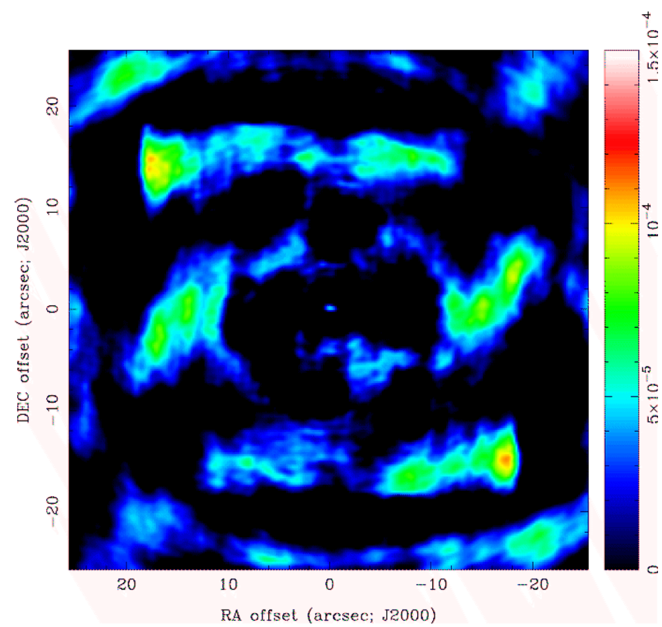




CLEAN model

restored image

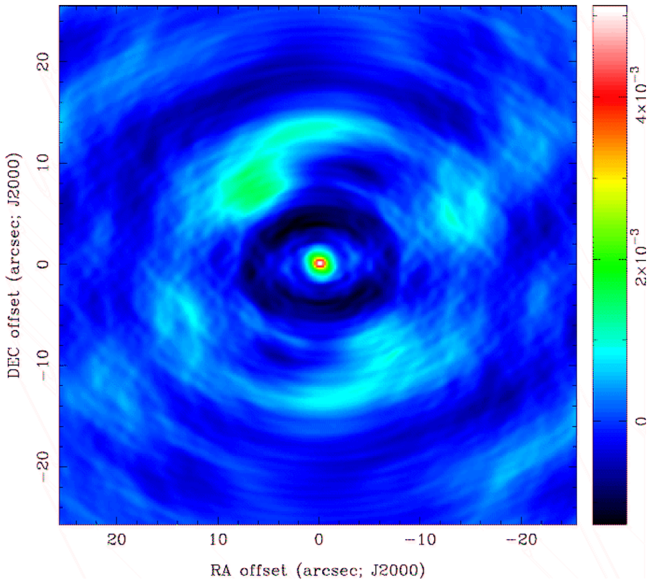


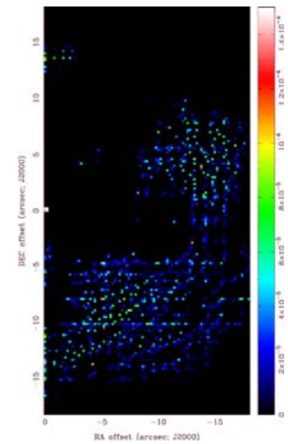


residual map

CLEAN with Poor Choice of Box

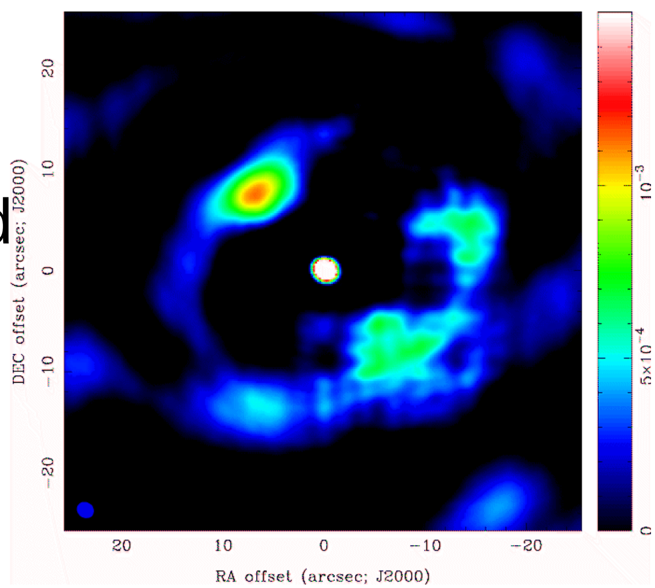
$T^D(x,y)$

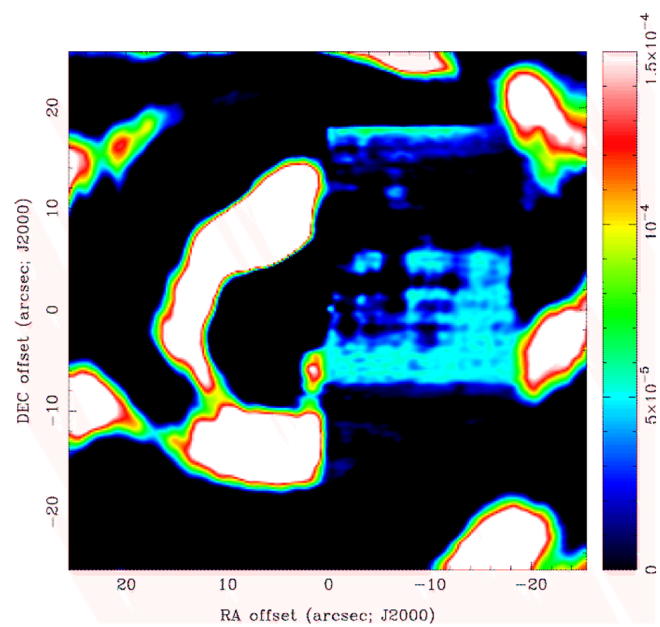




CLEAN model

restored image





residual map

CLEAN Variants

- Clark CLEAN
  - aims at faster speed for large images
  - Högbom-like “minor” cycle w/ truncated dirty beam, subset of largest residuals
  - in “major” cycle, cc’s are FFT’d and subtracted from the FFT of the residual image from the previous “major” cycle
- Cotton-Schwab CLEAN (MX)
  - in “major” cycle, cc’s are FFT’d and subtracted from ungridded visibilities
  - more accurate but slower (gridding steps repeated)
- Steer, Dewdny, Ito (SDI) CLEAN
  - aims to supress CLEAN “stripes” in smooth, extended emission
  - in “minor” cycles, any point in the residual map greater than a fraction (<1) of the maximum is taken as a cc
- Multi-Resolution CLEAN
  - aims to account for coupling between pixels by extended structure
  - independently CLEAN a smooth map and a difference map, fewer cc’s

“Restored” Images

- CLEAN beam size:
  - natural choice is to fit the central peak of the dirty beam with elliptical Gaussian
  - unit of deconvolved map is Jy per CLEAN beam area (= intensity, can convert to brightness temperature)
  - minimize unit problems when adding dirty map residuals
  - modest super resolution often OK, but be careful
- “restored” image does not fit the visibility data
- photometry should be done with caution
  - CLEAN does not conserve flux (extrapolates)
  - extended structure missed, attenuated, distorted
  - phase errors (e.g. seeing) can spread signal around

Noise in Images

- point source sensitivity: straightforward
  - telescope area, bandwidth, integration time, weighting
  - in image, modify noise by primary beam response
- extended source sensitivity: problematic
  - not quite right to divide noise by $\sqrt{n}$ beams covered by source: smoothing = tapering, omitting data $\rightarrow$ lower limit
  - always missing flux at some spatial scale
- be careful with low signal-to-noise images
  - if position known, 3σ OK for point source detection
  - if position unknown, then 5σ required (flux biased by $\sim 1\sigma$)
  - if $< 6\sigma$, cannot measure the source size (require $\sim 3\sigma$ difference between "long" and "short" baselines)
  - spectral lines may have unknown position, velocity, width

Maximum Entropy Algorithm

– Maximize a measure of smoothness (the entropy)

$$H = -\sum_k T_k \log\left(\frac{T_k}{M_k}\right)$$

subject to the constraints

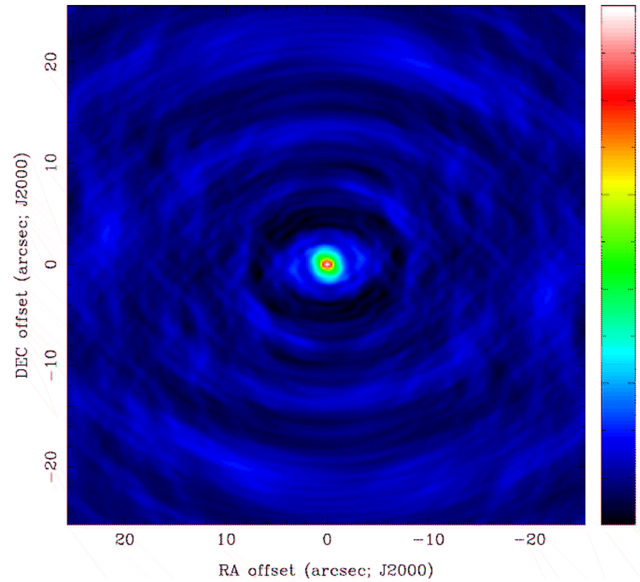
$$\chi^2 = \sum_k \frac{|V(u_k,v_k) - \mathrm{FT}\{T\}|^2}{\sigma_k^2}$$

$$F = \sum_k T_k$$

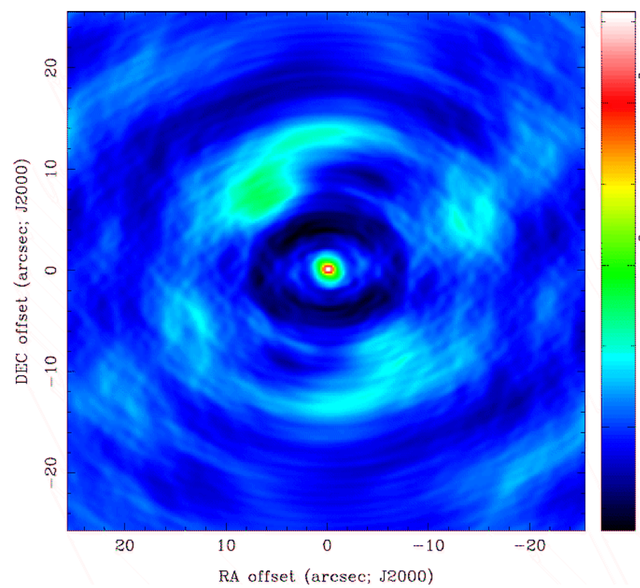
– M is the “default image”

– fast (NlogN) non-linear optimization solver due to Cornwell and Evans (1983)

– optional: convolve with Gaussian beam and add residual map to make map



b(x,y)



$T^D(x,y)$

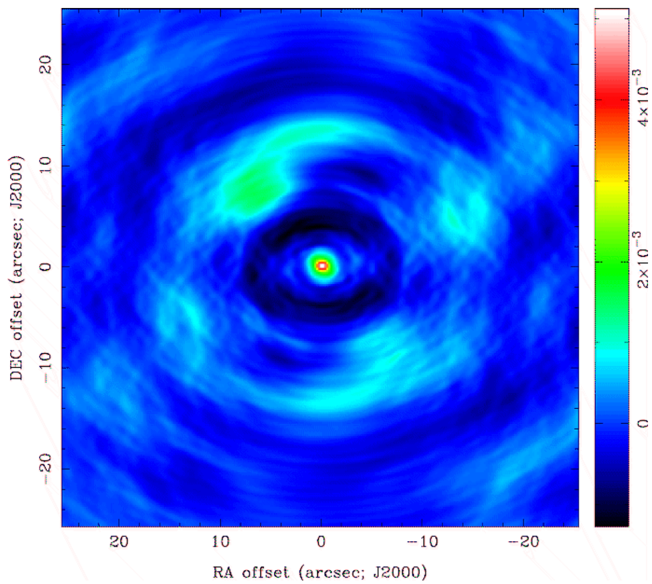
Maximum Entropy Algorithm (cont)

- easy to include *a priori* information with default image
  - flat default best only if nothing known (or nothing observed!)
- straightforward to generalize χ^2 to combine different observations/telescopes and obtain optimal image
- many measures of "entropy" available
  - replace log with cosh $\rightarrow$ "emptiness" (does not enforce positivity)

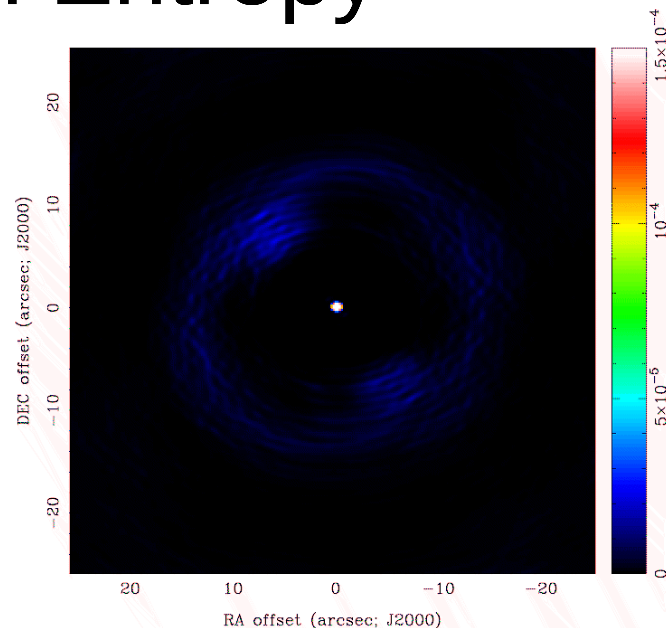
- less robust and harder to drive than CLEAN
- works well on smooth, extended emission
- trouble with point source sidelobes
- no noise estimate possible from image

Maximum Entropy

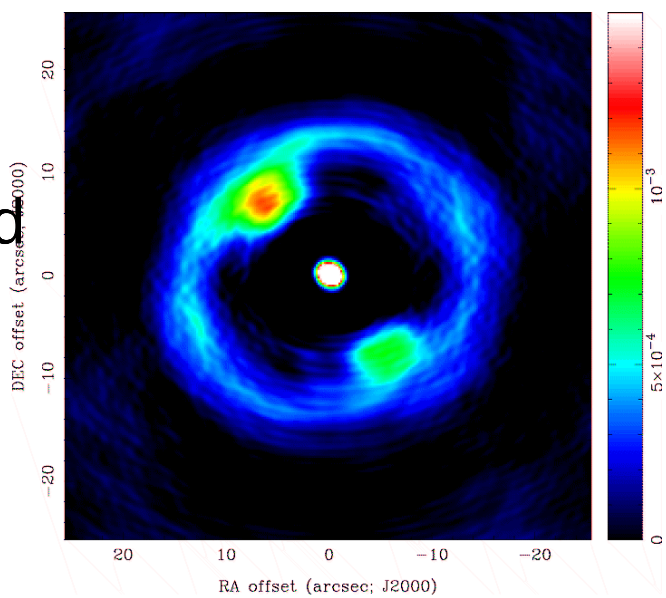
$T^D(x,y)$



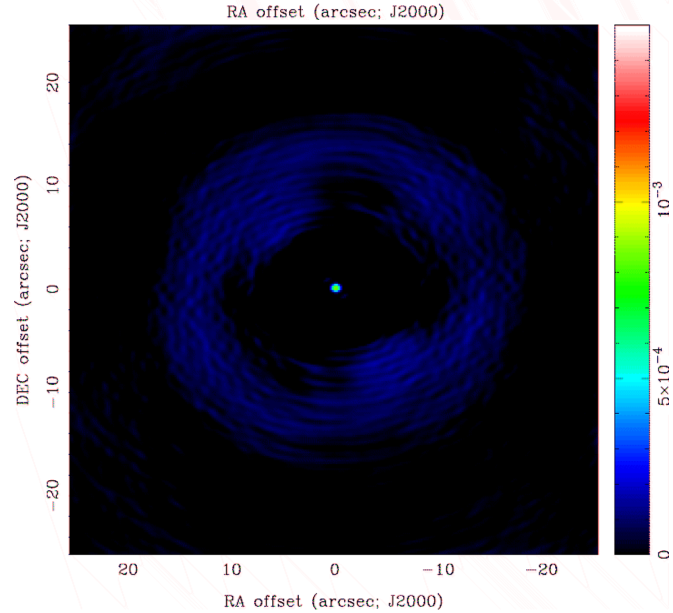
MAXEN model



restored image



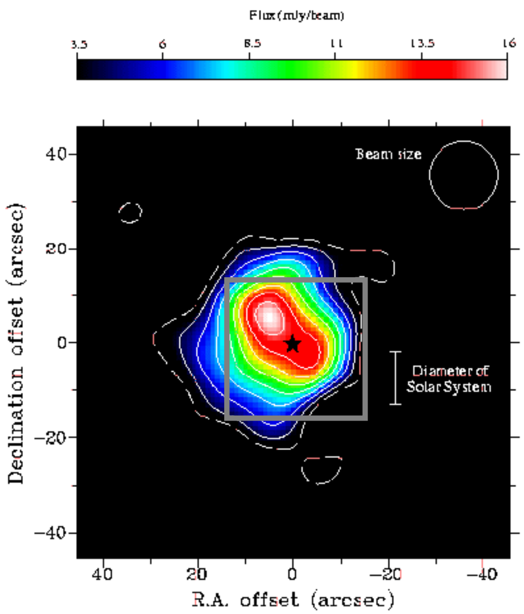
residual map



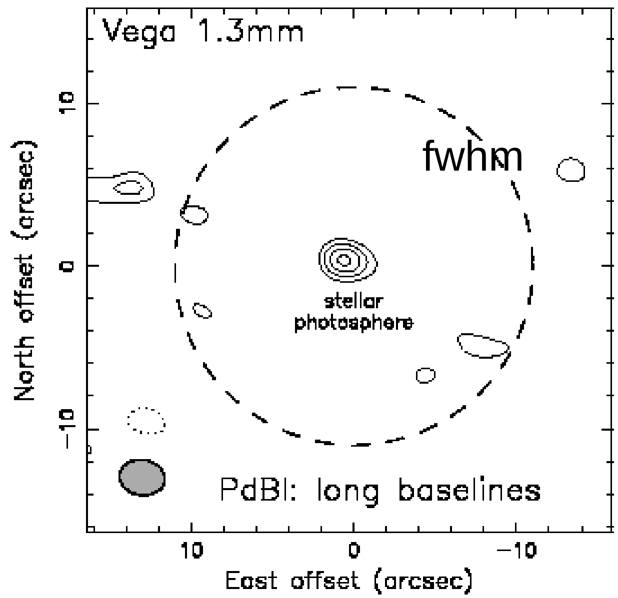
Example: Dust around Vega

- tune resolution and sensitivity to suit science
- Wilner et al. 2002, ApJ, 569, L115:

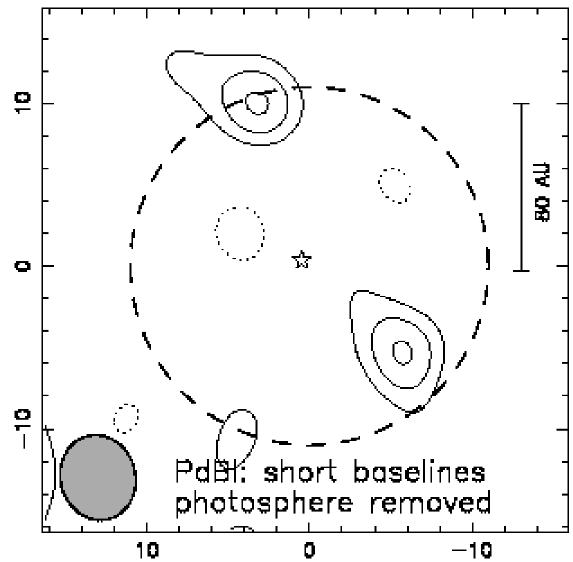
Vega 850 μm
Holland et al. 1998



SCUBA: 14 arcsec



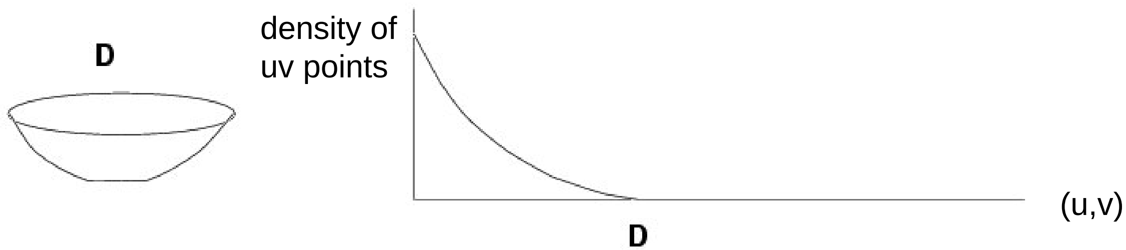
2.8 x 2.1 arcsec
stellar photosphere



5.3x4.6 arcsec
and dust blobs

Missing Low Spatial Frequencies (I)

- Large Single Telescope
  - make an image by scanning across the sky
  - all Fourier components from 0 to D sampled, where D is the telescope diameter (weighting depends on illumination)



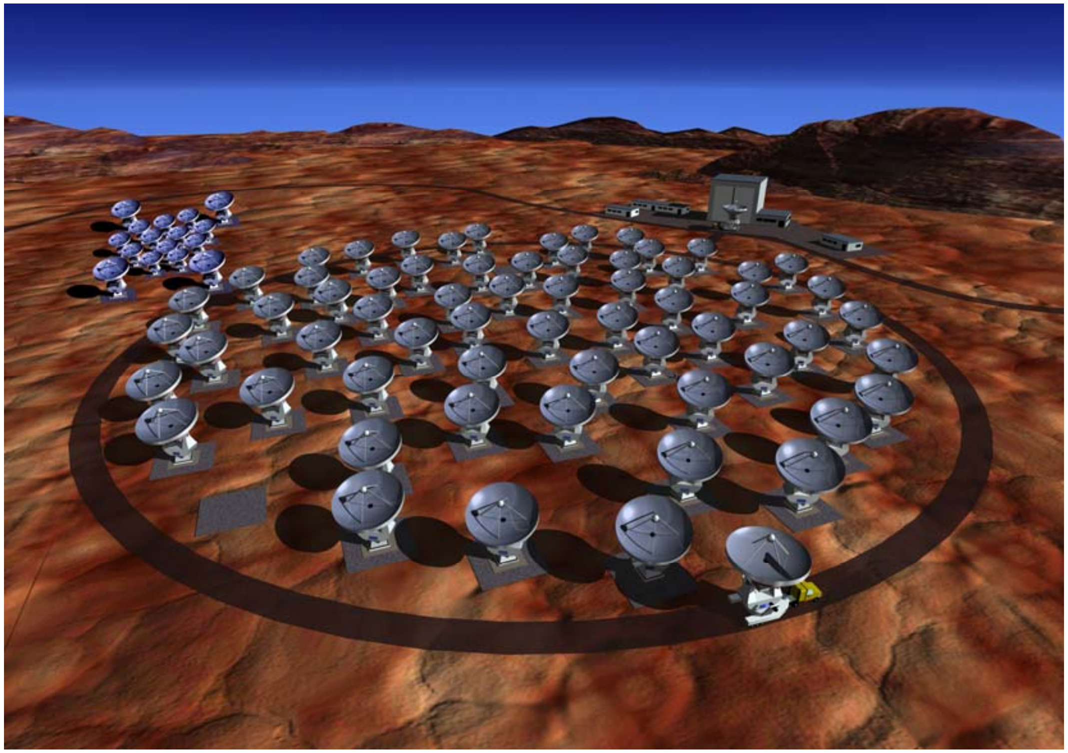
  - Fourier transform single dish map = T(x,y) ⊗ A(x,y), then divide by a(x,y) = FT{A(x,y)}, to estimate V(u,v)

$$\hat{V}(u,v) = \frac{[V(u,v)a(u,v)]}{\hat{a}(u,v)}$$

  - choose D large enough to overlap interferometer samples of V(u,v) and avoid using data where a(x,y) becomes small

Missing Low Spatial Frequencies (II)

- separate array of smaller telescopes
  - use smaller telescopes observe short baselines not accessible to larger telescopes
  - shortest baselines from larger telescopes total power maps

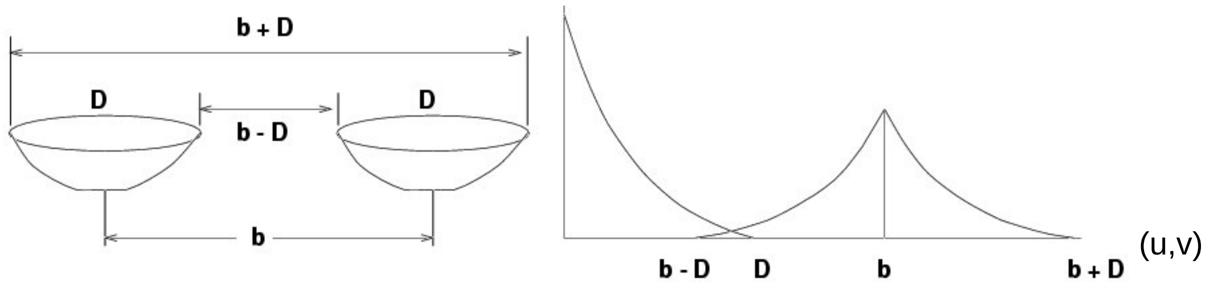


ALMA with ACA

64 x 12 m:	12 to 14000 m
+12 x 7 m:	fills ~7 to 12 m
+ 4 x 12 m:	fills 0 to ~7 m

Missing Low Spatial Frequencies (III)

- mosaic with a homogeneous array
  - recover a range of spatial frequencies around the nominal baseline b using knowledge of A(x,y) (Ekers and Rots 1979) (and get shortest baselines from total power maps)



  - V(u,v) is linear combination of baselines from b-D to b+D
  - depends on pointing direction (x_o,y_o) as well as (u,v)

$$V(u,v;x_o,y_o) = \int\int T(x,y)A(x-x_o,y-y_o)e^{2\pi i(ux+vy)}dxdy$$

  - Fourier transform with respect to pointing direction (x_o,y_o)

$$V(u-u_o,v-v_o) = \frac{\int\int V(u,v;x_o,y_o)e^{2\pi i(u_o x_o+v_o y_o)}dx_o dy_o}{a(u_o,v_o)}$$

Self Calibration

- a priori calibration not perfect
  - interpolated from different time, different sky direction from source
- basic idea of self calibration
  - correct for antenna-based errors *together with imaging*
- works because
  - at each time, measure N complex gains and N(N-1)/2 visibilities
  - source structure represented by small number of parameters
  - highly overconstrained problem if N large and source simple
- in practice, an iterative, non-linear relaxation process
  - assume initial model $\rightarrow$ solve for time dependent gains $\rightarrow$ form new sky model from corrected data using e.g. CLEAN $\rightarrow$ solve for new gains…
  - requires sufficient signal-to-noise ratio for each solution interval
- loses absolute position information
- dangerous with small N, complex source, low signal-to-noise

Summary

Calibrated Visibilities

$\Downarrow$ Fourier Transform

Dirty Beam and Dirty Image

$\Downarrow$ Deconvolution

Clean Beam, Sky Model, “restored” Image

$\Downarrow$ Analysis

Physical Information on Source

	AIPS	Miriad	GILDAS
Fourier Transform	imagr	invert	UV_MAP
CLEAN deconvolution	imagr	clean, restor	CLEAN

Concluding Remarks

- interferometry samples visibilities that are related to a sky brightness image by the Fourier transform
- deconvolution corrects for incomplete sampling

- remember… there are usually an infinite number of images compatible with the sampled visibilities
- astronomer must use judgement in imaging process
- imaging is generally fun (compared to calibration)

- many, many issues not covered today (see References)