mm Interferometry: Imaging

S.Guilloteau (IRAM/ESO)

ALMA European Project Scientist

Bibliography:

- "Synthesis Imaging". Proceedings of the lectures from NRAO summer school. Eds R.Perley, F.Schwab & A.Bridle.
- " Proceedings of the IMISS2". Ed. A.Dutrey
- "Interferometry and Synthesis in Radio Astronomy". R.Thompson, J.Moran & G.W.Swenson, Jr.

Fourier Transform and Related Approximations

- The *Complex Visibility* is

$$V = |V|e^{i\phi_V} = \int_{Sky} A(\sigma)I(\sigma)e^{-2i\pi\nu\mathbf{b}.\sigma/c}d\Omega \quad (1)$$

- Let (u,v,w) be the coordinate of the baseline vector, in units of the observing wavelength ν, in a frame of the phase tracking vector $\mathbf{s}_o$, with w along $\mathbf{s}_o$. (x,y,z) are the coordinates of the source vector $\mathbf{s}$ in this frame. Then

$$\begin{aligned} \nu\mathbf{b}.\mathbf{s}/c &= ux + vy + wz & \nu\mathbf{b}.\mathbf{s}_o/c &= w \\ z &= \sqrt{1-x^2-y^2} & d\Omega &= \frac{dxdy}{z} = \frac{dxdy}{\sqrt{1-x^2-y^2}} \end{aligned} \quad (2)$$

$$V(u,v,w) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} A(x,y)I(x,y)\frac{e^{-2i\pi(ux+vy+w(\sqrt{1-x^2-y^2}-1))}}{\sqrt{1-x^2-y^2}}dxdy \quad (3)$$

with $I(x,y) = 0$ when $x^2 + y^2 \geq 1$

Field of View: 2-D Fourier Transform

- If (x, y) are sufficiently small,

$$(\sqrt{1 - x^2 - y^2} - 1)w \simeq \frac{1}{2}(x^2 + y^2)w \simeq 0 \tag{4}$$

and Eq.3 becomes

$$\begin{aligned} V(u,v) &= \int\!\!\int B(x,y)I(x,y)e^{-2i\pi(ux+vy)}e^{-i\pi(x^2+y^2)w}dxdy \\ \text{with} \qquad B(x,y) &= \frac{A(x,y)}{\sqrt{1-x^2-y^2}} \end{aligned} \tag{5}$$

i.e. basically a 2-D Fourier Transform of BI, but with a phase error term $\pi(x^2 + y^2)w$.

- We want $|\pi(x^2 + y^2)w| \ll 1$ in all circumstances. Noting that $w < w_{\mathrm{max}} = B_{\mathrm{max}}/\lambda \simeq 1/\theta_{\mathrm{s}}$, we can show that for a maximum phase error of 6°, this limits the field of view to $\theta_{\mathrm{f}} < 0.35\sqrt{\theta_{\mathrm{s}}}$ (in radians...).

Field of View: Bandwidth Smearing

- Assume u, v are computed for the center frequency ν_0. At frequency ν_0, we have ($\rightleftharpoons$ being the Fourier transform operator)

$$V(u, v) \rightleftharpoons BI(x, y) \tag{6}$$

The similarity theorem on Fourier Transform pairs give

$$V(\frac{\nu_0}{\nu}u, \frac{\nu_0}{\nu}v) = (\frac{\nu}{\nu_0})^2 I(\frac{\nu}{\nu_0}x, \frac{\nu}{\nu_0}y) \tag{7}$$

- Averaging of bandwidth $\Delta\nu$, there is a *radial smearing* equal to

$$\sim \frac{\Delta\nu}{\nu_0}\sqrt{x^2 + y^2} \tag{8}$$

and hence the constraint

$$\sqrt{x^2 + y^2} \leq 0.1\frac{\theta_s \nu_0}{\Delta\nu} \tag{9}$$

Field of View: Time Averaging

- Assume observations of the Celestial Pole. The baselines cover a sector of angular width $\Omega_{earth}\Delta t$, where Ω_{earth} is the Earth rotation speed, and Δt the integration time.
- The smearing is *circumferential* and of magnitude $\Omega_{earth}\Delta t\sqrt{x^2+y^2}$, hence the constraint

$$\sqrt{x^2+y^2} \leq 0.1\frac{\theta_{\rm s}}{\Omega_{earth}\Delta t} \tag{10}$$

- For other declinations, the smearing is no longer rotational, but of similar magnitude.
- Values for Plateau de Bure

$\theta_{\rm s}$	ν (GHz)	2-D Field	0.5 GHz Bandwidth	1 Min Averaging	Primary Beam
$5''$	80	$5'$	$80''$	$2'$	$60''$
$2''$	80	$3.5'$	$30''$	$45''$	$60''$
$2''$	220	$3.5'$	$1.5'$	$45''$	$24''$
$0.5''$	230	$1.7'$	$22''$	$12''$	$24''$

Array Geometry & Baseline Measurements

- Using a Cartesian coordinate system (X, Y, Z) with Z towards the pole, X towards the meridian, and Y towards East, the conversion matrix to u, v, w is

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} \sin(h) & \cos(h) & 0 \\ -\sin(\delta)\cos(h) & \sin(\delta)\cos(h) & \cos(\delta) \\ \cos(\delta)\cos(h) & -\cos(\delta)\sin(h) & \sin(\delta) \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{11}$$

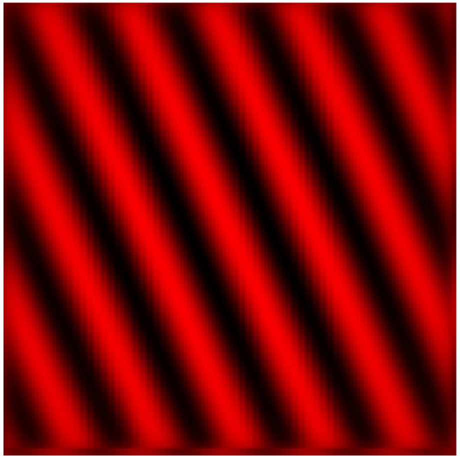
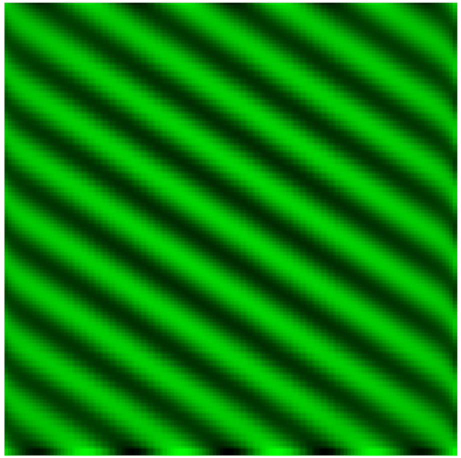
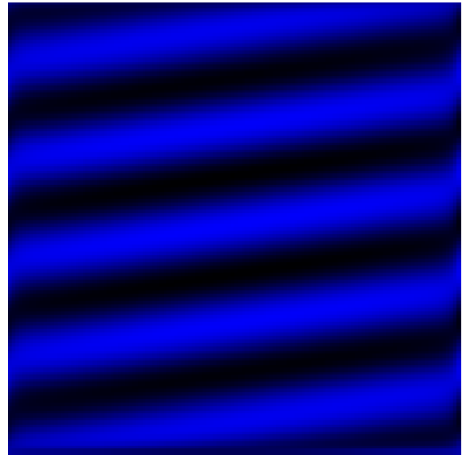
  where h, δ are the hour angle and declination of the phase tracking center.

- Eliminating h from Eq.11 gives the equation of an ellipse:

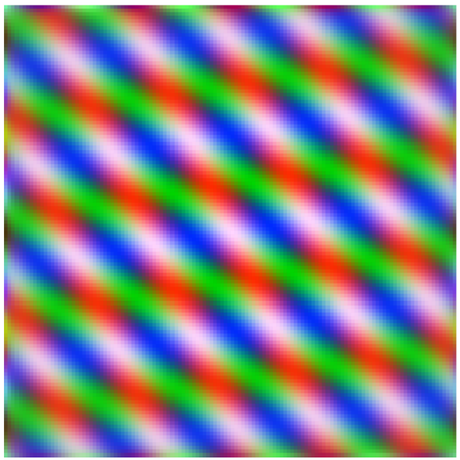
$$u^2 + \left(\frac{v - (Z/\lambda)\cos(\delta)}{\sin(\delta)} \right)^2 = \frac{X^2 + Y^2}{\lambda^2} \tag{12}$$

- the UV coverage is an ensemble of such ellipses. Choice of antenna configurations is made to cover the UV plane as much as possible. Array configuration design takes this constraint into consideration, with the restriction due to the site ...

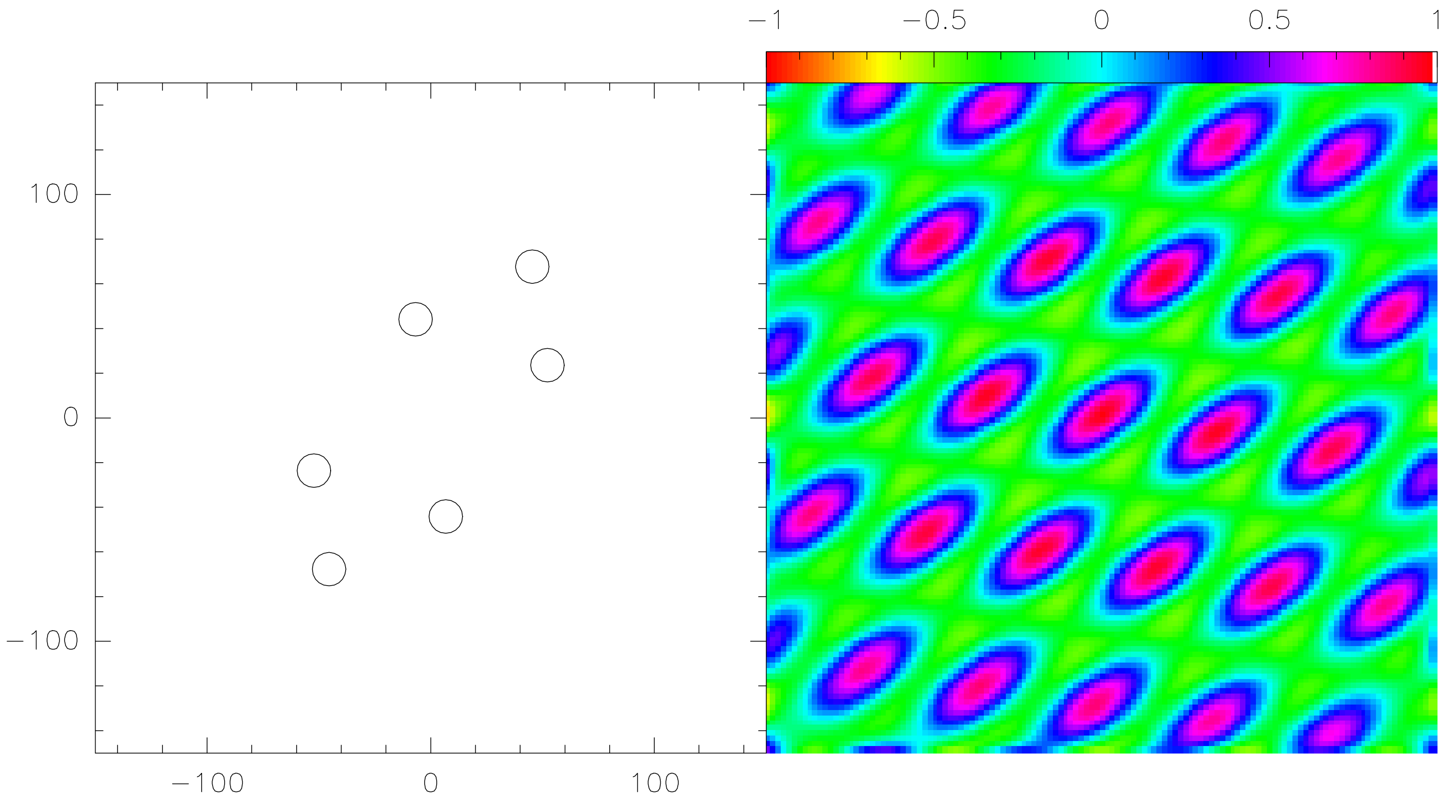
Fringes: 2 —> 3 antennas

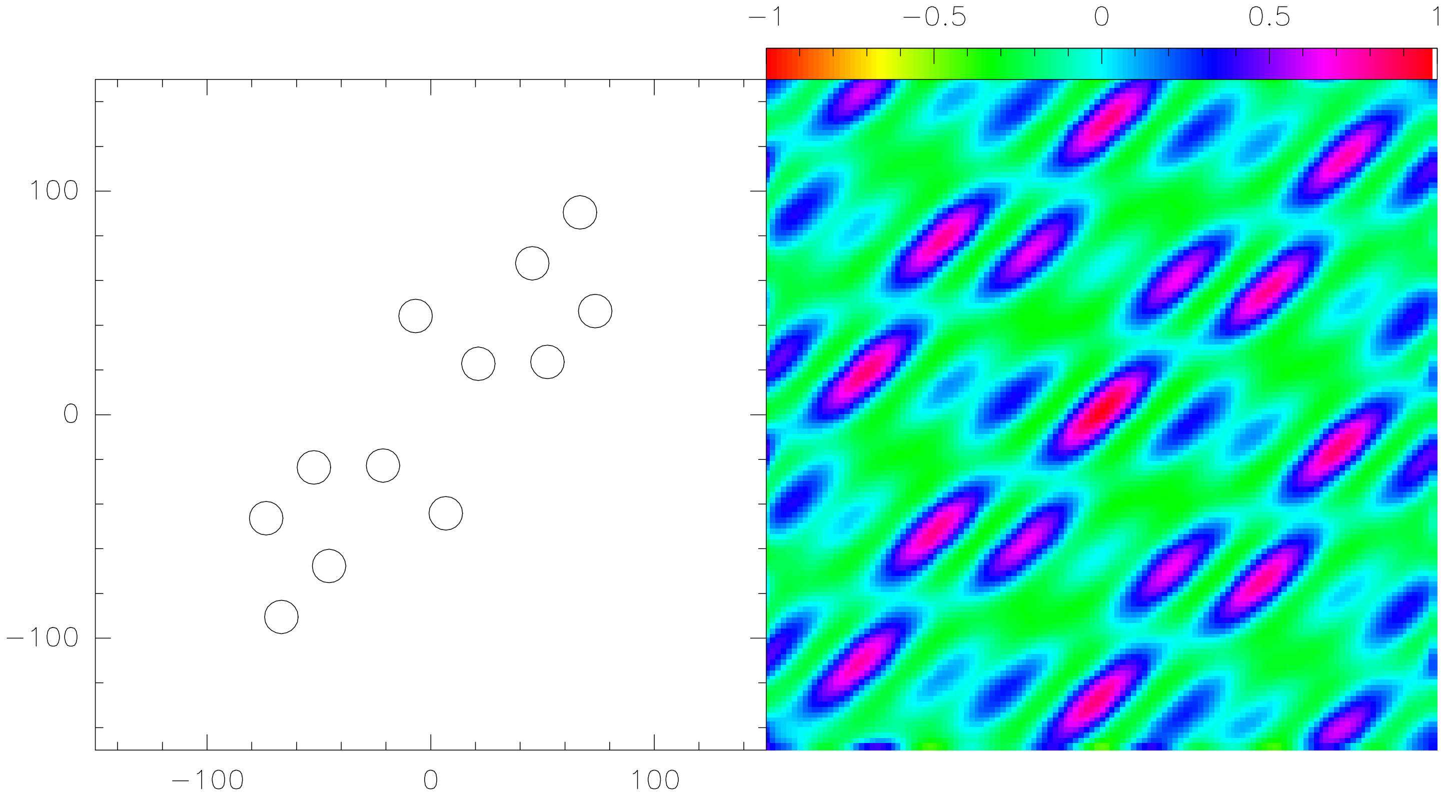
With 3 antennas:



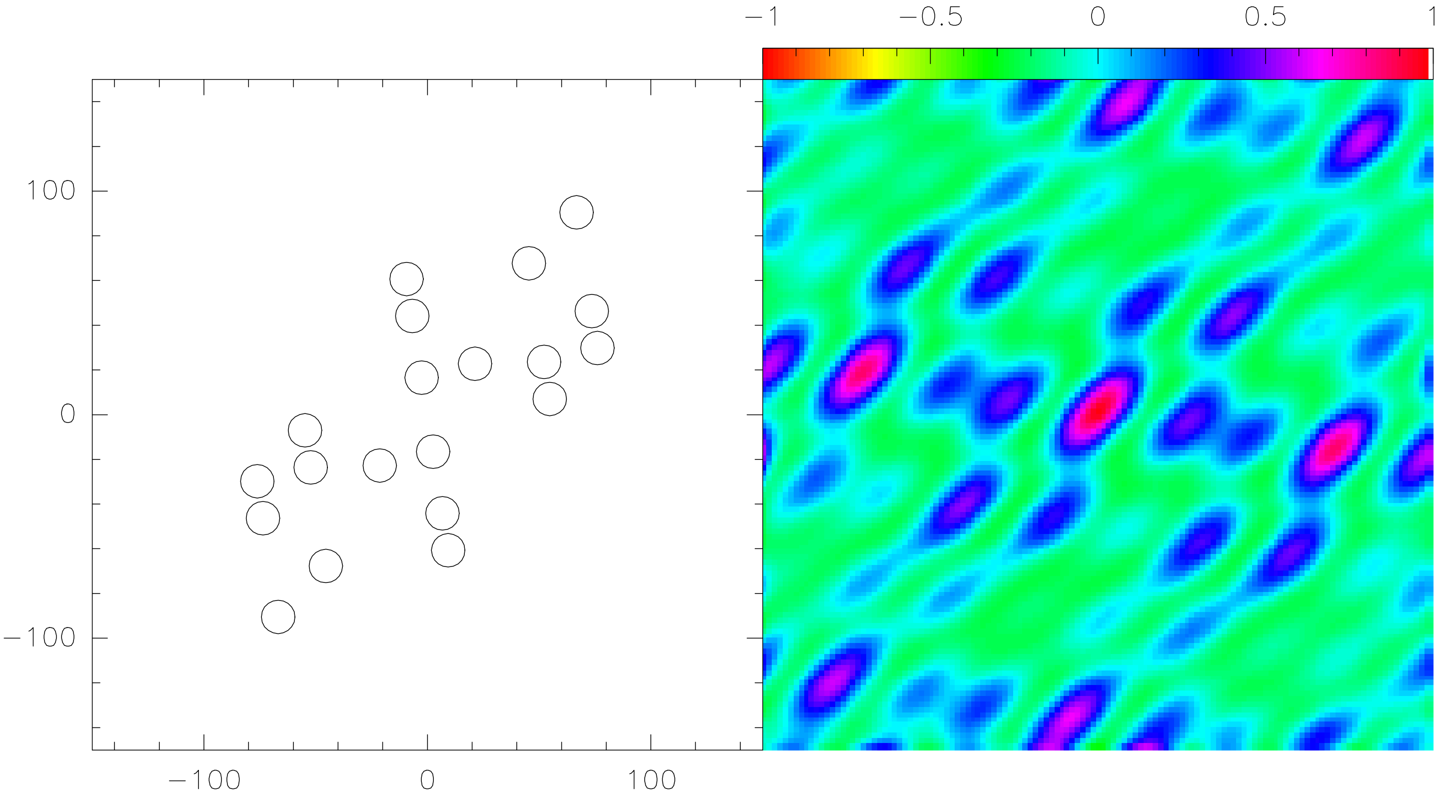
Fringes: 3 antennas



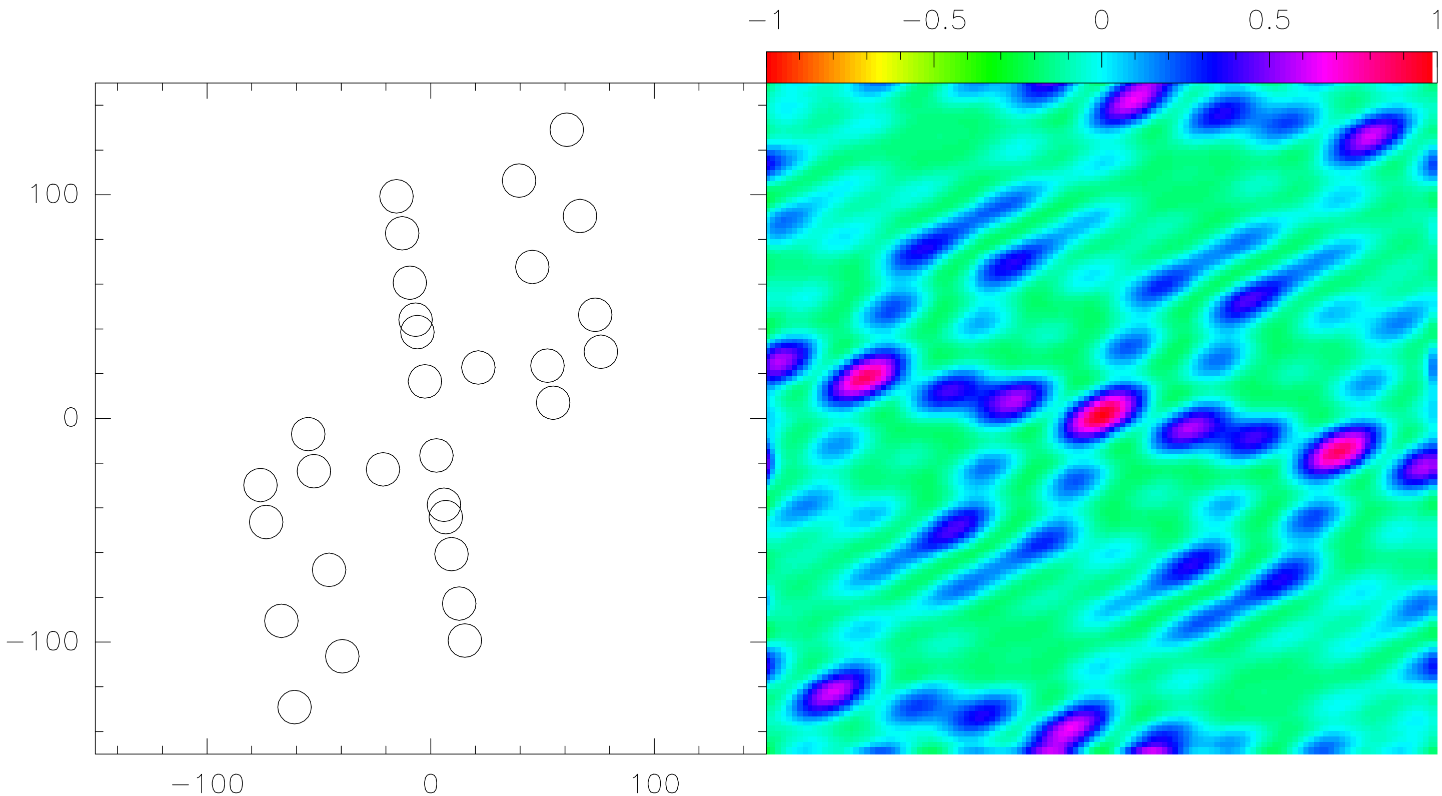
Fringes: 4 antennas



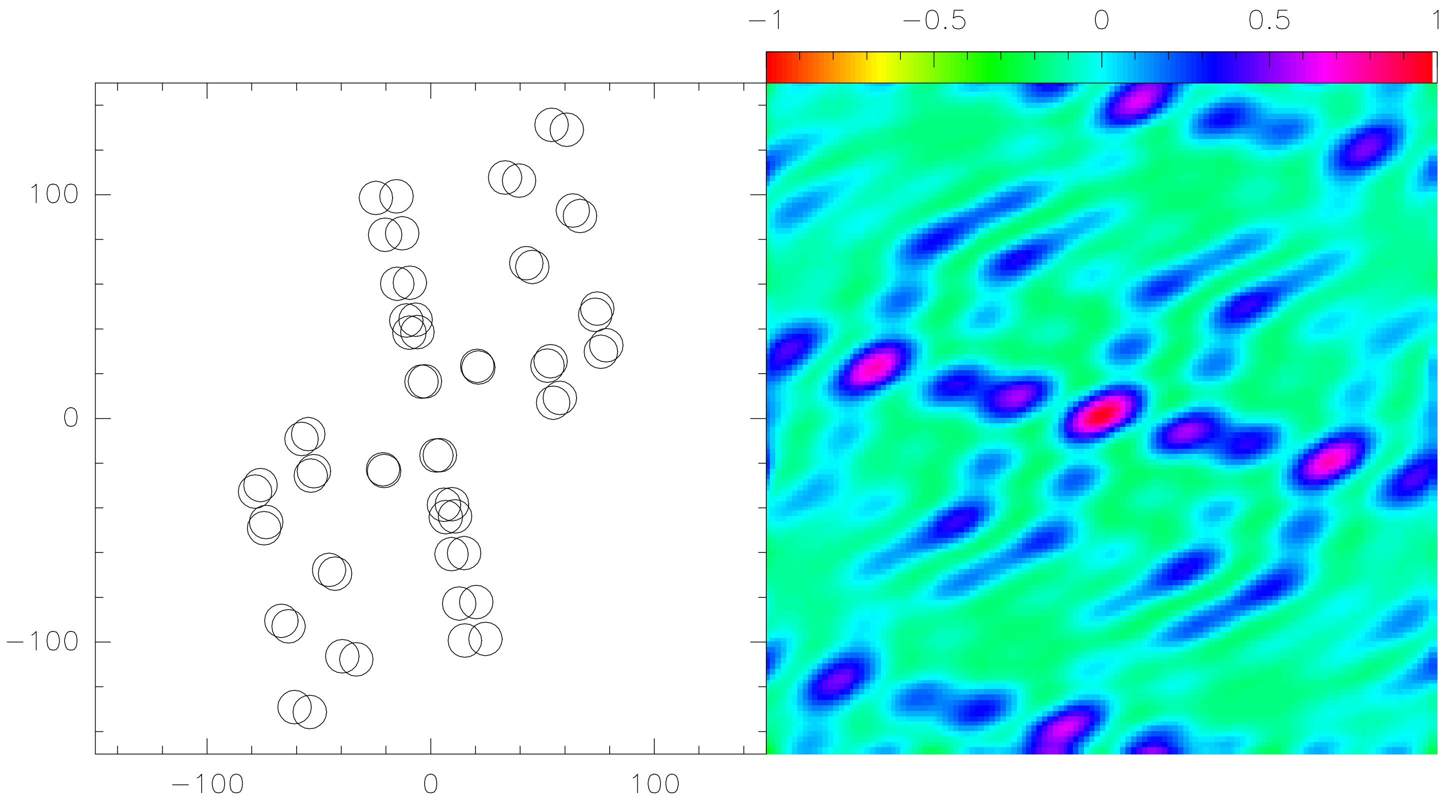
Fringes: 5 antennas



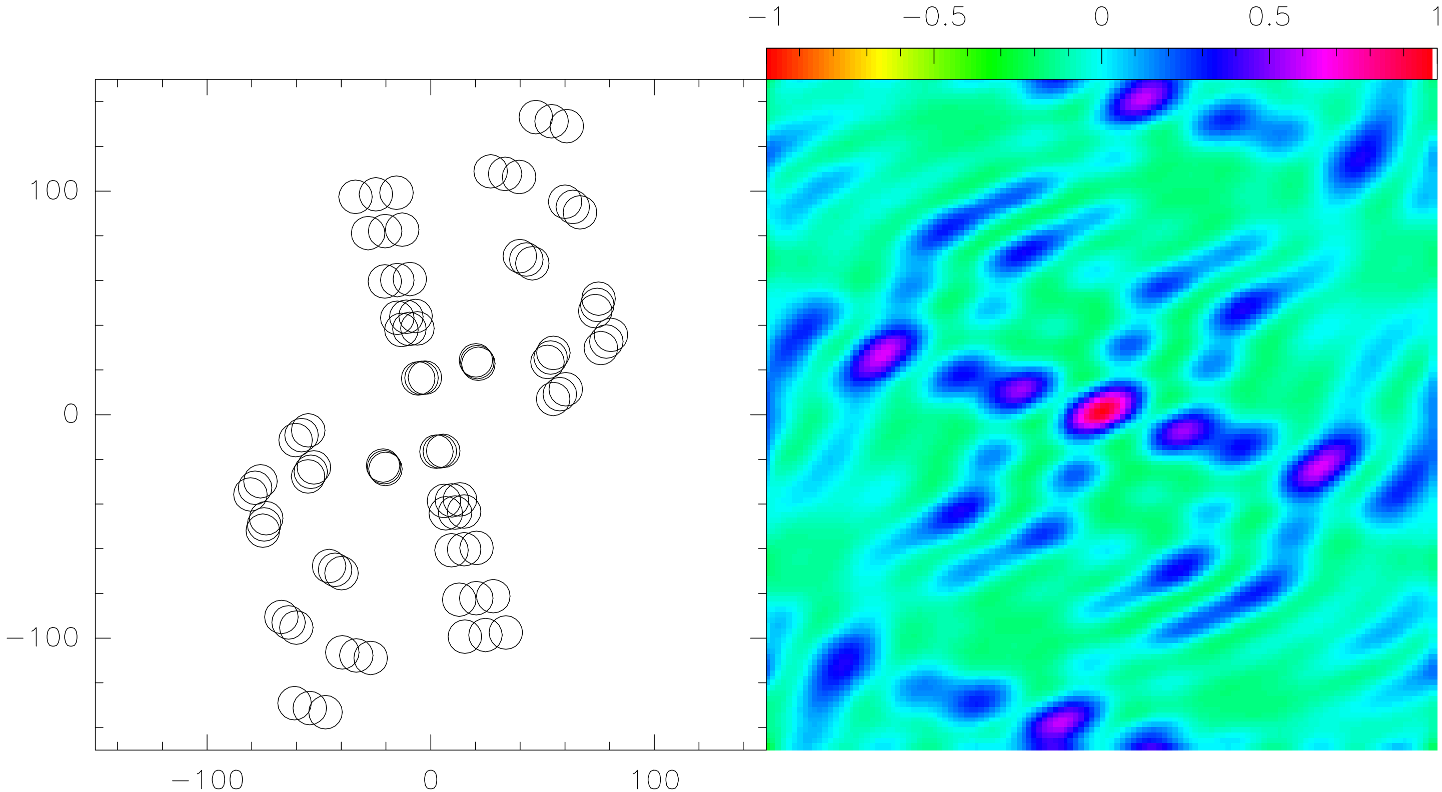
Fringes: 6 antennas



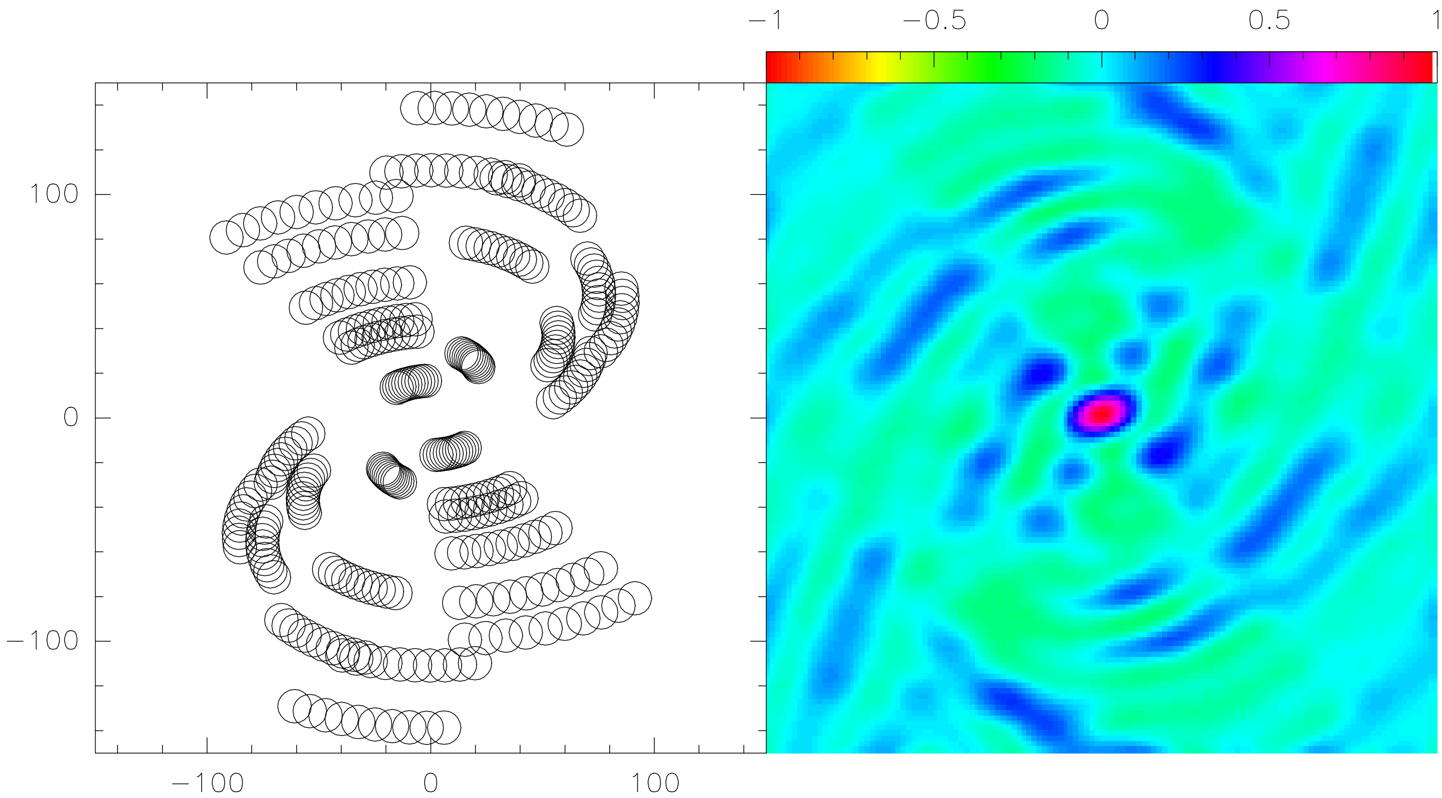
Aperture Synthesis



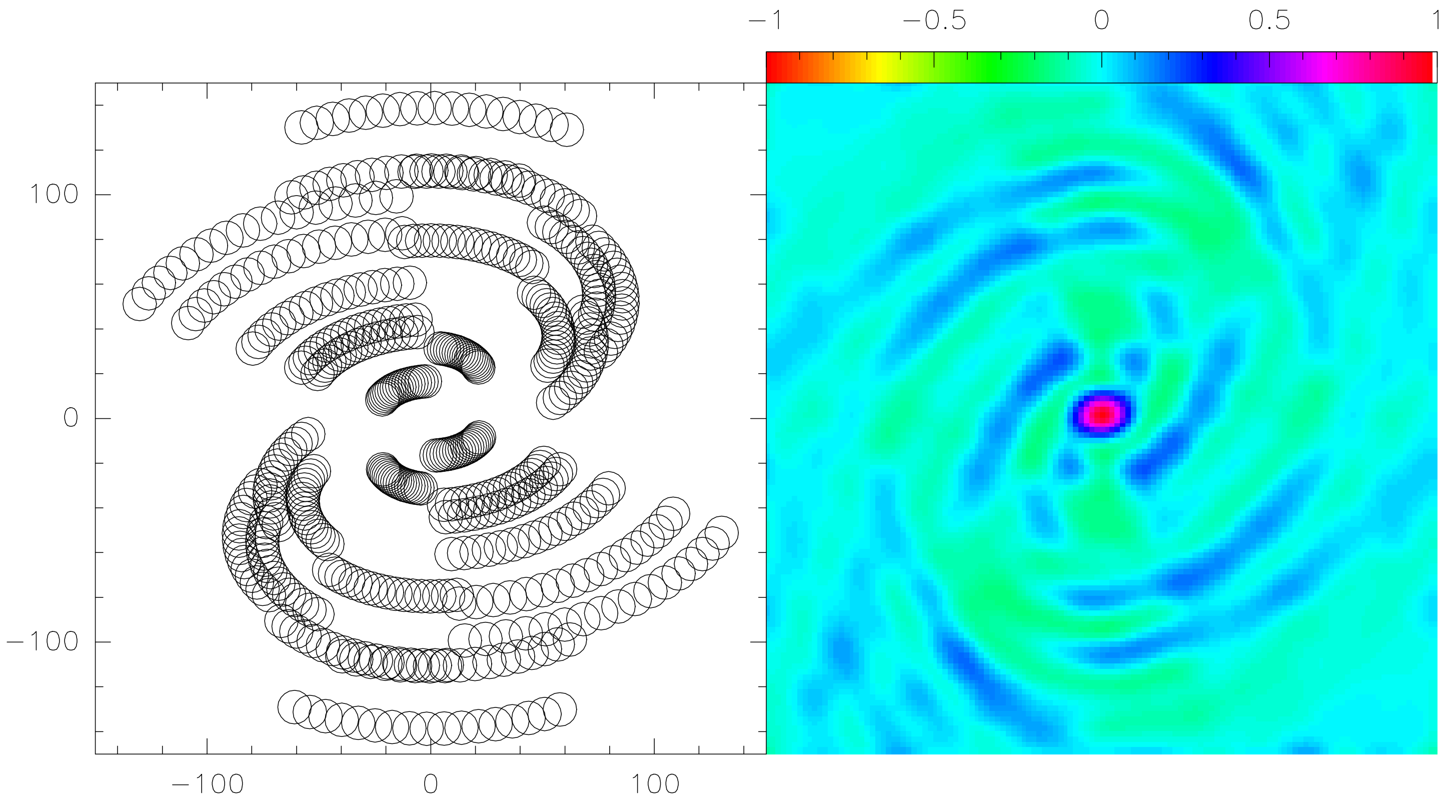
Aperture Synthesis: Earth Rotation



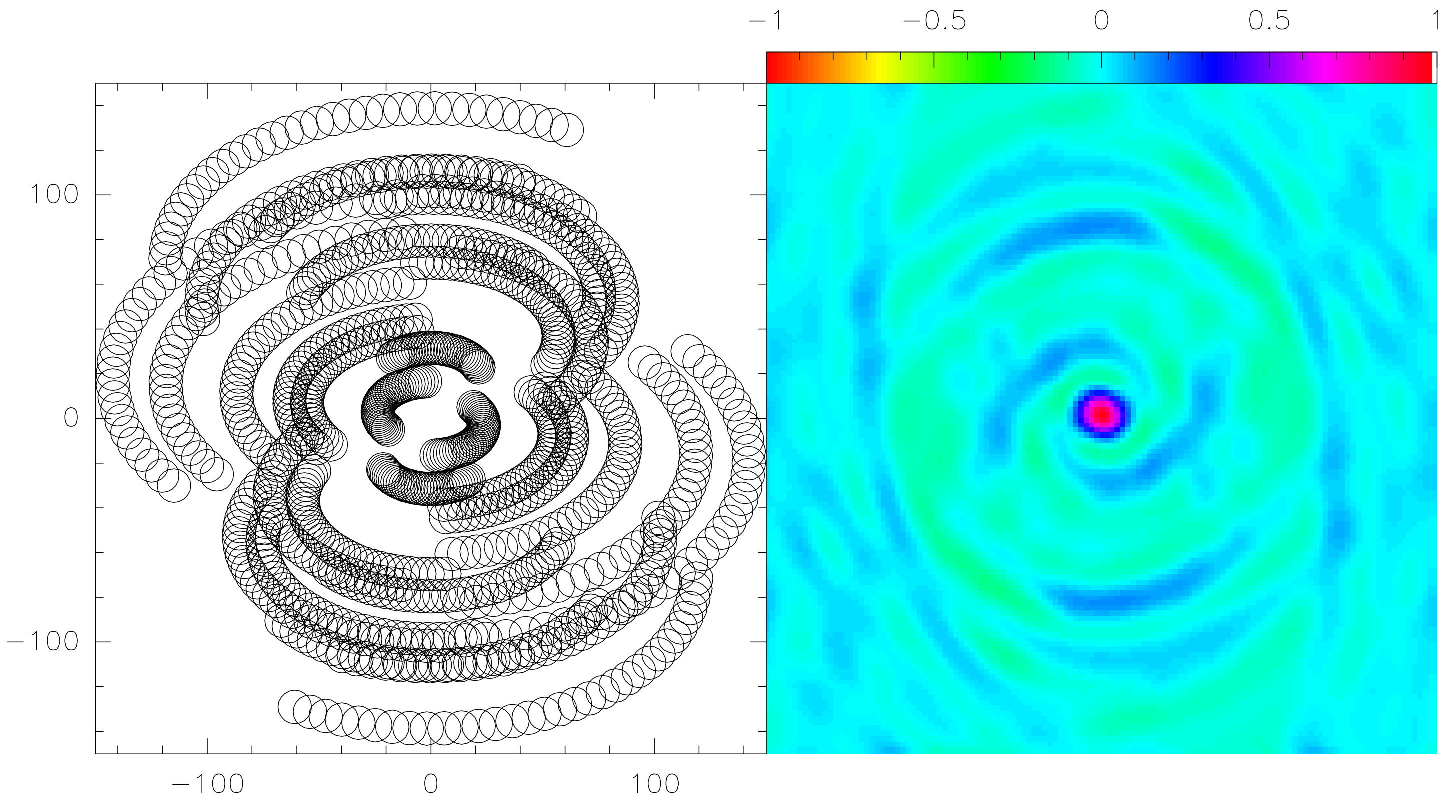
Aperture Synthesis: Earth Rotation



Aperture Synthesis: Earth Rotation



Aperture Synthesis: Earth Rotation



From Visibilities to Images

- An interferometer measures the *visibility function*

$$V(u,v) = \iint B(x,y)\, I(x,y)\, e^{-2i\pi(ux+vy)}\, dx\, dy \tag{13}$$

over an ensemble of points $(u_i, v_i), i = 1, n$, where $B(x,y)$ is the (slightly modified) power pattern of the antennas and $I(x,y)$ the sky brightness distribution.

- The imaging process consists in determining as best as possible the sky brightness $I(x,y)$.
- Let $S(u,v)$ be the sampling function (also called **spectral sensitivity** function).

$$\begin{aligned} S(u,v) \neq 0 &\Longleftrightarrow \exists i \in 1,n\, (u_i,v_i) = (u,v) \\ S(u,v) = 0 &\Longleftrightarrow \forall i \in 1,n\, (u_i,v_i) \neq (u,v) \end{aligned} \tag{14}$$

S contains information on the relative weights of each visibility, derived e.g. from theoretical noise.

- Let us define

$$I_w(x,y) = \iint S(u,v)\, W(u,v)\, V(u,v)\, e^{2i\pi(ux+vy)}\, du\, dv \tag{15}$$

where $W(u,v)$ is an arbitrary *weighting function*.

- Since the Fourier Transform of a product of two functions is the convolution of the Fourier Transforms of the functions, $I_w(x,y)$ can be identified with

$$I_w(x,y) = (B(x,y)I(x,y)) ** (D_w(x,y)) \tag{16}$$

where

$$\begin{aligned} D_w(x,y) &= \iint S(u,v)\, W(u,v)\, e^{2i\pi(ux+vy)}\, du\, dv \\ &= \widehat{SW} \end{aligned} \tag{17}$$

$D_w(x,y)$ is called the *dirty beam*, and is directly dependent on the choice of the *weighting function* W.

Fourier Transform and **Deconvolution**
are thus two key issues in imaging (see Eq.16).

Back to Eq.18

Fourier Transform

- Direct Fourier Transform
  - Computes $\sin$ and $\cos$ functions directly in Eq.16 for all combinations of visibilities and pixels in the image.
  - Straightforward, but slow
- Fast Fourier Transform (FFT)
  - Use the speed of FFT
  - Gridding process for the visibilities
  - Gridding correction for the images and beam
  - Aliasing effects due to periodic function
    - images must not be too small
    - hopefully, $B \times I$ in Eq.16 has a finite support.

Gridding Process

- Definition of gridding
  - Goal: Re-sample the visibilities on a regular grid for subsequent use of the FFT
  - General Method: Resample on the grid by some interpolation from the values of nearest visibilities around each grid point.
  - Specific Method: Use a *convolution* technique to perform the interpolation.
- Why a convolution ?
- From Eq.13, $V = \widehat{BI} = \widehat{B} ** \widehat{I}$
  $\Longrightarrow V$ is already a convolution of a nearly Gaussian function $(\widehat{B})$ with the FT I, $\widehat{I}$.
  $\Longrightarrow$ nearby visibilities are not independent.
- *Exact interpolation* not desirable, since original data points are *noisy samples of a smooth function*. *Some smoothing* is desirable.
- If the width of the convolution kernel used in gridding is small compared to $\widehat{B}$, the convolution added in the gridding process will not significantly degrade the information

Gridding Process

- By definition

$$I_w = I ** D_w = \widehat{V} ** \widehat{SW}$$

- Let G the gridding convolution kernel. Eq.15 becomes

$$I_w^g \rightleftharpoons G ** (S \times W \times V) \tag{18}$$

$$I_w^g = \widehat{G} \times (\widehat{SW} ** \widehat{V}) \;=\; \widehat{G} \times I_w \tag{19}$$

$$D_w^g \rightleftharpoons G ** (S \times W) \quad D_w^g \;=\; \widehat{G} \times \widehat{SW} \tag{20}$$

$$\frac{I_w^g}{\widehat{G}} = \frac{D_w^g}{\widehat{G}} ** (BI) \tag{21}$$

- Thus the dirty image and dirty beams are obtained by dividing the Fourier Transform of the gridded data by the Fourier Transform of the gridding function.

Back to Eq.26

Sampling & Aliasing

- Sampling is equivalent to multiplying by a series of delta functions, or the *shah* function:

$$[\frac{1}{\Delta u}]\mathrm{III}(\frac{u}{\Delta u}) = \sum_{k=-\infty}^{\infty} \delta(u - k\Delta u) \tag{22}$$

- The Fourier Transform of the *shah* function above is the *shah* function

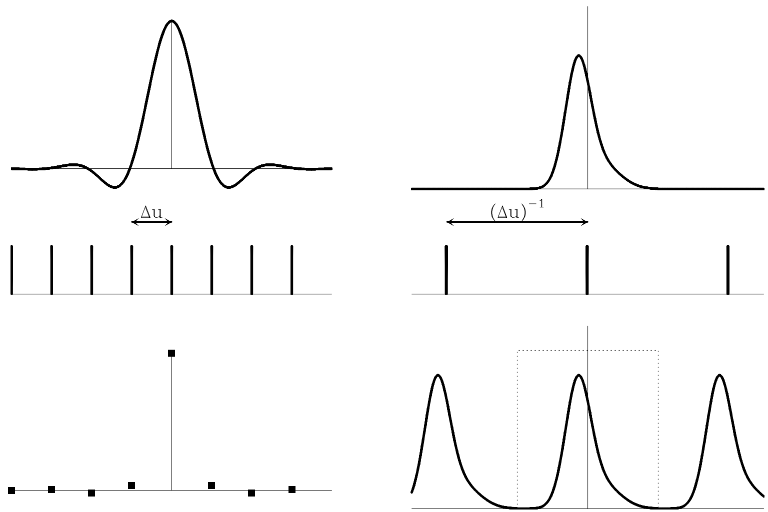
$$\mathrm{III}(x\Delta u) = \frac{1}{\Delta u} \sum_{m=-\infty}^{\infty} \delta(x - \frac{m}{\Delta u}) \tag{23}$$

- Hence, sampling the visibilities V results in convolving its Fourier Transform $\widehat{V}$ by a periodic *shah* function. This convolution reproduces in a periodic way the Fourier Transform of the visibilities $\widehat{V}$.
- If the Fourier Transform of the visibilities $\widehat{V}$, i.e. the brightness distribution BI, has finite support ΔX, the replication poses no problem provided

$$(\Delta u)^{-1} \geq (\Delta X) \qquad \Delta u \leq (\Delta X)^{-1} \tag{24}$$

Sampling & Aliasing (...)

- If not, data outside $(\Delta u)^{-1}$ are *aliased* in the *imaged area* $(\Delta u)^{-1}$. (see Fig)
- Finite support is ensured to first order by the finite width of the antenna primary beam B. However, strong sources in antenna sidelobes may be aliased if the imaged area is too small.
- **Noise** does not have finite support, because white noise in the UV plane result in white noise in the map plane. Noise aliasing produces an increased noise level at the map edges.



Convolution and Aliasing

- Combination of Gridding and Sampling produces the UV data set

$$V_m = \frac{1}{\Delta u \Delta v} \mathrm{III}(\frac{u}{\Delta u}, \frac{v}{\Delta v}) \times (G ** (S \times W \times V))(u,v) \quad (25)$$

$$= \mathrm{III} \times (G ** (S \times W \times V))/(\Delta u \Delta v) \quad (26)$$

which analogous with Eq.18

- The Fourier Transform of this UV data set is

$$\widehat{V}_m = \mathrm{III}(x\Delta u, y\Delta v) ** (\widehat{G} \times (\widehat{SW} ** \widehat{V}))$$

$$\widehat{V}_m = \mathrm{III} ** (\widehat{G} \times (\widehat{SW} ** (BI))) \quad (27)$$

- V_m is thus the sky brightness multiplied by the primary beam (B I), convolved by the the dirty beam $\widehat{SW}$, then multiplied by the Fourier transform of the gridding function $\widehat{G}$ and periodically replicated (by the convolution with last Shah function).
- Aliasing of $\widehat{G}$ in the map domain will thus occur.

Back to Eq.29

Convolution and Aliasing (...)

- Note 1: at this stage, providing that aliasing of $\widehat{G}$ remains negligible, an exact convolution equation is preserved

$$\frac{\widehat{V_m}}{\widehat{G}} = \widehat{SW} ** BI \tag{28}$$

- Note 2: aliasing of $\widehat{G}$ is not completely negligible. Choice of the convolving function must be made to minimize it.
- Note 3: The weighting function W is usually smooth, while the gridding function G is a relatively sharp function (since it ensures the re-gridding by convolution from *nearby* data points). Thus, to first order $G ** W = W$, and we can rewrite Eq.26 as

$$V_m = \mathrm{III} \times W \times (G ** (S \times V))/(\Delta u \Delta v) \tag{29}$$

i.e. the weighting can be performed *after* the gridding.

Choice of the Gridding function

- The gridding function will be selected according to the following principles:

  1. small support, typically one or two cells wide (Δu).
  2. small aliasing.
  3. fast computation.

- 1 and 2 are contradictory, since a small support for G implies a large extent of $\widehat{G}$. Some compromise is required. Gridding functions are usually selected among those with separable variables:

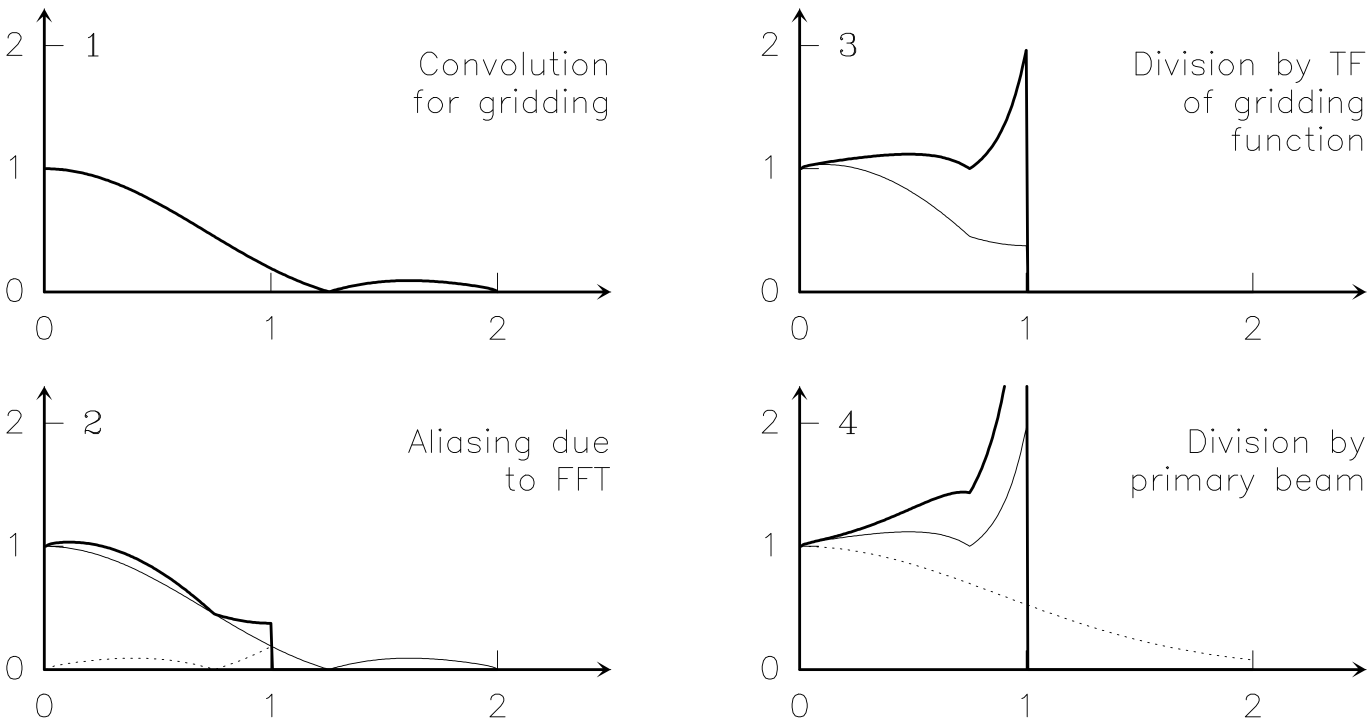
$$G(u, v) = G_1(u) G_1(v)$$

- **Gaussian-Sinc function**
  $\widehat{G}$ should ideally be a rectangular function (1 inside the map, 0 outside). G would be a sinc function, but this falls off too slowly, and would require a lot of computations in the gridding; and its truncation would destroy the sharp edges of $\widehat{G}$.
- **Spheroidal functions**
  We actually want $\widehat{G}$ to fall off as quickly as possible, but G to be support limited. Mathematically, this defines a class of functions known as *spheroidal* functions. They are solutions of differential equations, and cannot be expressed analytically.

  Tabulated values are used in the task `UV_MAP`.

The noise problem

- **Noise**
  Aliasing increases noise at the map edges (by aliasing and then by the gridding correction). This amounts to $(\pi/2)^2$ at map corners for the Gaussian-Sinc function. Near the map center, the effect is negligible.



Weighting and Tapering

- At UV table creation, the sampling function is defined as

$$S(u,v) = \frac{1}{\sigma^2(u,v)} \tag{30}$$

where the noise σ is computed from system temperature, bandwidth, integration time, and effective antenna efficiency (including decorrelation).

$$\sigma(u,v) = \frac{\mathcal{J}_I T_{\rm sys}}{\eta_{\rm Q}\sqrt{2\Delta\nu t_{\rm int}}} \tag{31}$$

- The weights $W(u,v)$ can be freely chosen.
- **Natural Weighting, no taper**
corresponds to $W(u,v) = 1$. This maximizes sensitivity to *point sources*.
- **Tapering** consists in apodizing the UV coverage by $T(u,v) = \exp(-(u^2+v^2)/t^2))$ where t is a tapering distance. This corresponds to smoothing the data in the map plane (by convolution with a gaussian), and thus, to some extent increases sensitivity to "medium size" structure.

Weighting and Tapering

- **Tapering** is always throwing out some information. To increase sensitivity, use compact arrays, not tapering. (You usually don't want to throw away 50% of the information...).
- **Uniform Weighting** consists in selecting the weights $W(u, v)$ so that the sum of weights $\sum W \times S$ over a UV cell is a constant function (or zero if no UV data exists in that cell). The size (radius) of the UV cell is an arbitrary parameter. In practice, it should be compatible with sampling considerations (e.g. equal to half the dish diameter).
- **Uniform Weighting** gives more weight to long baselines than natural weighting (because you spend less time per UV cell on long baselines than on short baselines for earth synthesis). **Uniform Weighting** produces smaller beam and sometimes (but not always) lower sidelobes.
- Weighting and Tapering reduce point source sensitivity by

$$\sqrt{\sum T^2 W^2} / \sum TW \tag{32}$$

Robust Weighting

- This is a variant of uniform weighting which avoids to give too much weight to a UV cell with low natural weight.
- Roughly speaking, if the sum of natural weights in a cell is less than a threshold, the weighting is unchanged, if it is more, the weight is set to this threshold.
- **Robust** weighting combines the advantages of **Natural** and **Uniform** weighting, by increasing the resolution and lowering the sidelobes without degrading too much the sensitivity.
- By adjusting the threshold, it approaches either case (large threshold $\longleftrightarrow$ Natural, small threshold $\longleftrightarrow$ Uniform).

Imaging Parameters

- the **map size**: The imageable area is normally limited by the **field of view**, which is determined by the *primary beam*, but also by the *2-D approximation*, the *bandwidth smearing*, and the *integration time* smearing. For some deconvolution methods, the **map size** should be twice larger than the imageable area.

  Also, an interferometer is a *spatial filter*: large scale structures are not properly measured (see *Short spacing* and *Mosaics*).

- the **map cell size**
  The map pixel size used in gridding is a free parameter. It Should respect proper sampling compared to the synthesized beamwidth. In practice, 3 – 4 pixels per beamwidth are required.

- the **Weight Mode**
  Natural weighting preserves best sensitivity. However, *Robust* weighting may be a better compromise: it has two free parameters (cell size and robustness threshold).

- the **Taper**
  No taper should be used in general, but it is often important to look at the same data with different angular scales.

- Note that except for the **weight mode** and **taper**, the UV coverage contains enough information to allow appropriate values to be derived.

Good Use & Common Mistakes

- Image too small in angular size: aliasing will occur.
- Pixel size too large: synthesized beam is undersampled, leading to deconvolution problems
- Image too big
  - Waste of disk space
  - Waste of computer time
  - Increase numerical errors (more deconvolution iterations required)
  - Increase risk of "Insufficient memory" error message
- Poor choice of weighting and/or tapering function. It is recommended to look at your image with two sufficiently different angular resolutions (e.g. a Robust weighting with no taper, and a significantly tapered image).
- For spectral line observations: too many channels treated at once: data cubes get too big. Select only the ones you need... Remember: a 256x256 (image) x 256 channels is 64 MBytes. Most image processing tasks require 3 to 8 times that in (virtual) memory to operate

Deconvolution: example of CLEANing

- Principle: assume the sky is dark, but full of stars.
- Method: "matching pursuit"

1. Initialize a *Residual* map to the *Dirty* map
2. Initialize a *Clean component list* to zero.
3. Assume strongest feature in *Residual* map originates from a point source
4. Add a fraction γ (the *Loop Gain*) of this point source to the *Clean component list*, remove the same fraction, convolved with the dirty beam, from the *Residual* map.
5. If strongest feature larger than some threshold, go back to point 3. If not, go to 7. Each such step is called an iteration.
6. If number of iterations N_{iter} is too large, go to point 7. If not loop to point 3.
7. Convolve the *Clean component list* by a properly chosen *Clean Beam*, and add the *Residual* map to obtain the *Clean Map*.

CLEANing

- Loop Gain and Number of iterations
  - $0 < \gamma < 2$ in theory
  - $\gamma \simeq 0.2$ in practice, depending on sidelobe levels, source structure and dynamic range
  - In theory, remaining flux $\propto (1-\gamma)^{N_{\rm iter}}$ if the object is made of a single point source...
  - about $\propto (1-\gamma)^{(N_{\rm iter}/N_{\rm beams})}$ if the source extends over $N_{\rm beams}$ (to first order...)
  - Example: 500 iterations, 25 beams (i.e. only a 5×5 beams patch): $0.8^{500/25} = 0.8^{20} = 1.1\%$ residual. You may need a large number of iterations...
- Clean Beam
  - Usually Gaussian: simple to compute...
  - Size should be matched to the synthesized beam
  - If not, flux density estimates will be incorrect
  - Restoring with a round beam is allowed
  - Sometimes no proper choice (e.g. synthesized beam with narrow central peak on top of broad "shoulder")

CLEAN: Interpretation

- The *Clean component list* is a plausible solution of the measurement equation, but it is not unique...
- Clean image is **not** a solution !
- But convolution by the Clean Beam smears out artifacts due to extrapolation beyond the measured area of the UV plane (**a posteriori regularization**).
- Clean Box and Support
  - Only the inner quarter of a map can be properly Cleaned
  - The Clean Box provides a way of minimizing the number of solutions
  - A more flexible support can be used (e.g. in the MAPPING program)
  - Clean Box or support should not be too limited: Cleaning the noise is necessary too...
- Negative Clean components are necessary.

CLEAN variants

- Hogbom or SIMPLE: from the inventor himself. Only (a little bit less than) the inner quarter of the image is properly restored.
- CLARK from B.Clark: A Major-Minor cycle approach.
  - Minor cycles: Hogbom Clean on a subset of the image (pixels with high enough absolute values), with a subset of the beam (which only includes the larger sidelobes).
  - Major cycles: Remove the cumulative list of point sources by using an FFT
  - Fast, but could diverge for improper choices of the major/minor decision.
- MX (from W.Cotton and F.Schwab): As the Clark method, but the major cycle subtraction is done in the ungridded u, v data, which is then gridded again. Can almost CLEAN the whole image...
- Steer-Dewdney-Ito (SDI): As the Clark method, but no deconvolution is made in the minor cycle. A simple selection and scaling is done.

CLEAN variants ...

- Multi-Resolution Clean (`MRC`):
  - Separate the problem into a smoothed map and a difference map
  - Since the measurement equation is linear, both maps can be cleaned independently
  - The method is faster than standard CLEAN because the smooth map can be compressed, and the smooth and difference map both contains a smaller number of components
  - the large sidelobes in the difference beam can prevent proper convergence.
  - It is not applicable to **Mosaics**, where the measurement equation is non linear.
- Multi-Scale Clean (`MULTI`)
  - Produce N_{scale} smoothed maps, with different smoothing kernels.
  - At each iteration, select the component from the smoothed map which has the highest Signal to noise ratio.
  - Remove the corresponding source list for all N_{scale} smoothed maps.
  - Very stable, but relatively slow.

Maximum Entropy Methods

- Principle: assume you now nothing about the source, and find the *least biased* image which agrees with the data
- In math form, minimize

$$J = H - \alpha\chi^2 \tag{33}$$

where H is an image *entropy*

$$H = -\sum_i I_i \log\left(\frac{I_i}{eM_i}\right) \quad \texttt{maximum entropy, requires positivity}$$

$$H = -\sum_i I_i \log\left(\cosh\left(\frac{I_i}{M_i}\right)\right) \quad \texttt{maximum emptiness, does not}$$

α a Lagrange multiplier, and χ^2 is a measurement of the agreement with the data, which can be expressed in the visibility plane as:

$$\chi^2 = \sum_j \frac{(\widehat{I_j} - V_j)^2}{\sigma_j^2} \tag{34}$$

(but the analogous form in the image plane could be used too)

- Numerical techniques, based on conjugate gradients, exist to solve this problem.

MEM

- Advantages:
  - MEM is relatively fast on large images.
  - With the classical entropy form, it has some theoretical background in information theory.
  - In the absence of information, just gives the *prior* image M.
- Drawbacks:
  - MEM uses no direct information about the dirty beam shape, and the deconvolution only proceeds through a global minimum of Eq.33. MEM images can be locally poor.
  - Long baselines are usually better fitted than short baselines (they are more numerous). A kludge is to **fix the zero spacing**.
  - MEM will perform badly with limited u, v coverage
  - The MEM image is biased toward the prior.
  - The MEM image has varying angular resolution.
  - MEM resolution is signal to noise dependent.
  - No simple noise statistics is possible on a MEM deconvolved image: the noise in MEM images has a non zero average.

WIPE

- Principle: only a **regularized** image of the sky can be obtained
- Method: minimize

$$\chi^2 = \sum_j \frac{(\widehat{I_j} - V_j)^2}{\sigma_j^2} + \alpha \sum_l (\widehat{I_l})^2 \tag{35}$$

  where j runs over the ensemble of measured visibilities and l over an ensemble of visibilities for long baselines. α is a weight factor which can be computed from the number of visibilities.
- To minimize the χ^2, WIPE will tend to find solutions with zero visibilities on the long baseline ensemble l.
- WIPE will thus produce a **resolution limited** image.
- I is constrained by a support constraint (like in CLEAN). Positivity can be enforced.
- WIPE is closer to CLEAN than to MEM: it provides a well defined angular resolution.
- WIPE is slow, and somewhat conservative regarding the level of deconvolution performed.
- WIPE allows reconstruction errors to be quantified (in the upper limit sense).

From Flux density to Brightness temperature

- Units of dirty map are ill defined:
  - A single point source of 1 Jy appears with peak intensity of 1.
  - But if more than 1 point source, combination of positive or negative sidelobes from the other source modify this result.
- Deconvolution is thus required
  - However, it is impossible to deconvolve weak structures near the noise level.
  - If extended, these structures *may contribute to a significant flux*.
  - But they in general **do not contribute to a significant brightness**.
- After deconvolution, the beam area is well defined
- The clean map unit is Jy per beam area
- Conversion to brightness is done using the standard equation

$$S_\nu = \frac{2k\Omega_s}{\lambda^2} T_B = \frac{2k\pi\theta_s^2}{4\log 2\lambda^2} T_B \tag{36}$$

Accuracy of Flux density estimates

- Seeing effects:
  - Point source flux underestimated
  - Flux spread over the "seeing disk" $\Longrightarrow$ Total flux preserved
- Noise Estimate:
  - Statistics on the (deconvolved) image gives the rms noise for point sources
  - Avoid including map edges, since noise increases at map edges due to aliasing and gridding
  - However, for total flux, it is not sufficient to count the number of beams ...
  - UV plane analysis is preferred for simple sources.
- **Do not forget primary beam correction**
- Do not forget amplitude calibration uncertainty...

Short Spacings

- Extended structure are missed, attenuated or distorted in interferometric maps by lack of short spacing information.
- Deconvolution recovers some of them, but under-estimate the total flux because the integral of the dirty beam is zero (no short spacing)
- Constructing a beam with a non zero integral can help deconvolution $\Rightarrow$ **Zero spacing** flux or spectrum.
- **Short spacings** provides even more information.
- Can be provided by
  - A larger single dish ($\mathbf{D} > D$): it measures $V(u,v)$ up to $\sqrt{u^2+v^2} = \mathbf{D} - D$
  - A smaller interferometer ($d < D$), combined with zero spacing from the large antennas. In the field of view of the small antennas, the large antennas measure visibilities up to $\sqrt{u^2+v^2} < D - d$, the small interferometer from $\sqrt{u^2+v^2} > d$, so if $d \simeq D/2$, full sampling of the u, v plane is guaranteed.

Short Spacing and Mosaics

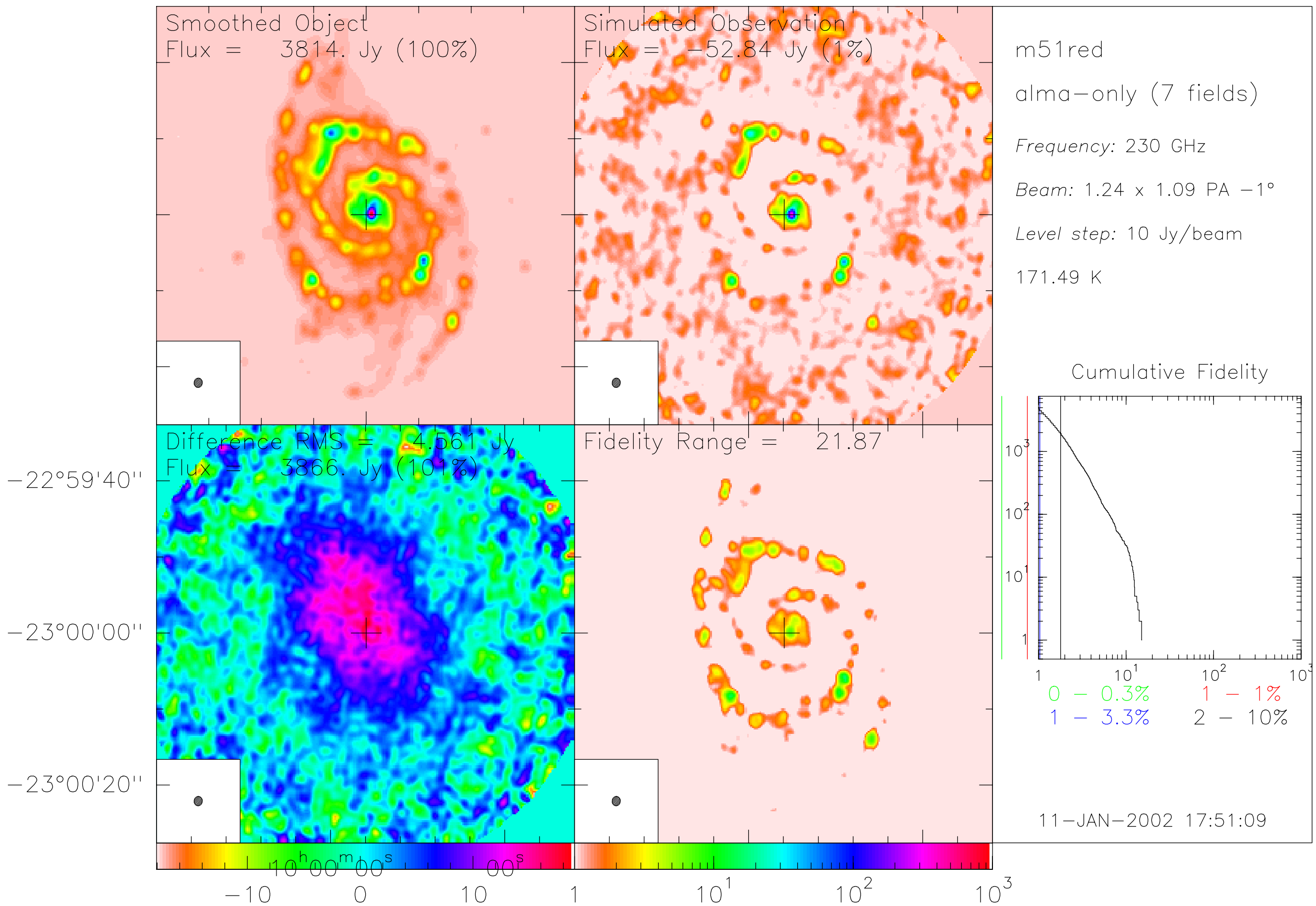
- an interferometer is a *spatial filter*: it is insensitive to large structures
- an interferometer has limited *field of view*
- To image areas of the sky larger than the field of view, **Mosaicing** with inclusion of **short spacing** is required
- **Mosaicing** principle
  - Make overlapping pointings with the interferometer
  - Each pointing measures (after deconvolution) $J_i = B_i I$ where I is the sky brightness
  - The linear combination

$$J = \frac{\sum_i B_i J_i}{\sum_i B_i^2} \equiv I \tag{37}$$

    is an optimal (vis-a-vis noise) estimate of I.
  - J can be corrupted by improper deconvolution of one field. Taking advantage of the overlapping nature of the B_i would certainly help minimizing deconvolution errors.

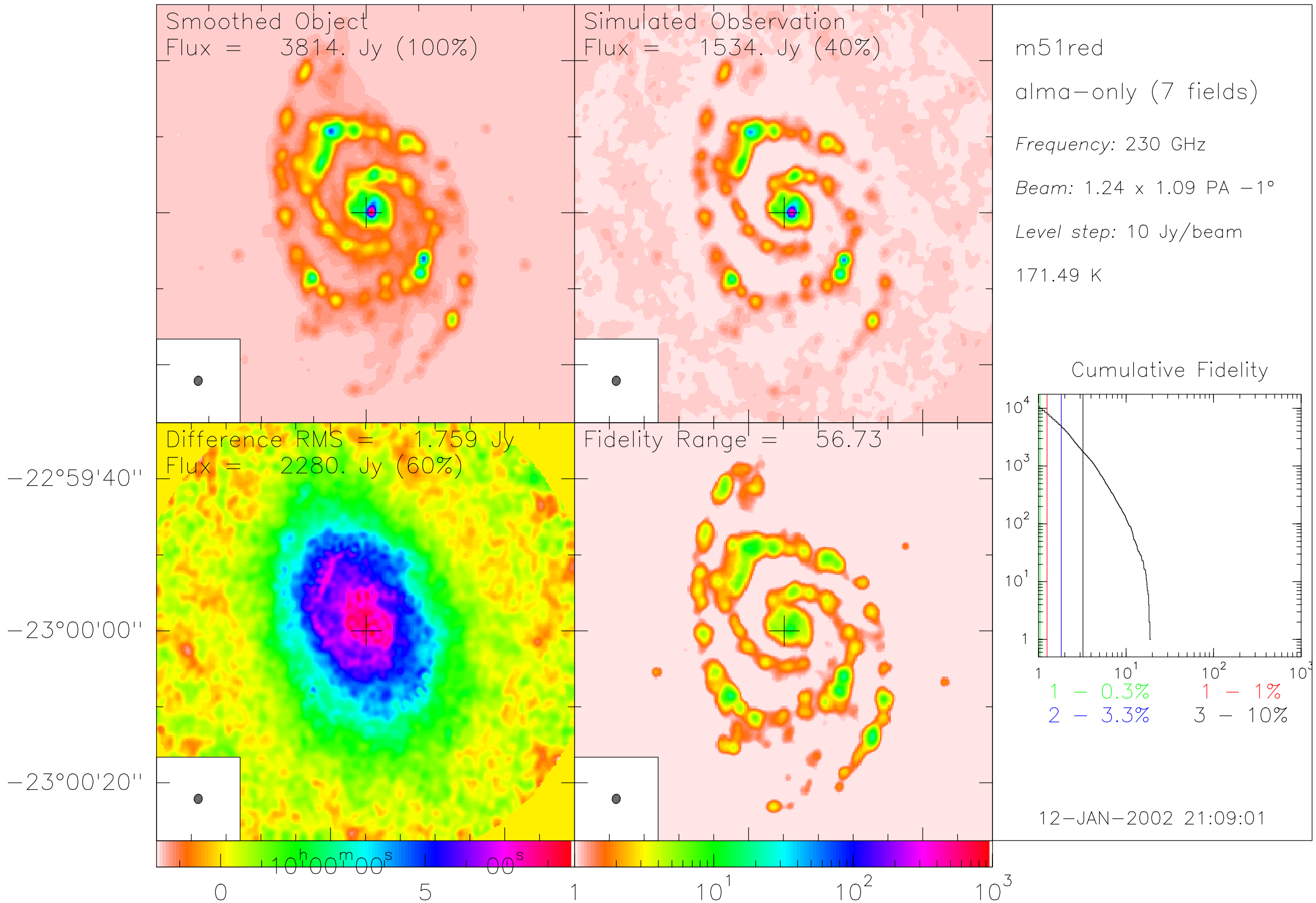
Mosaic Deconvolution: i) raw image

- Linear combination, with no prior deconvolution



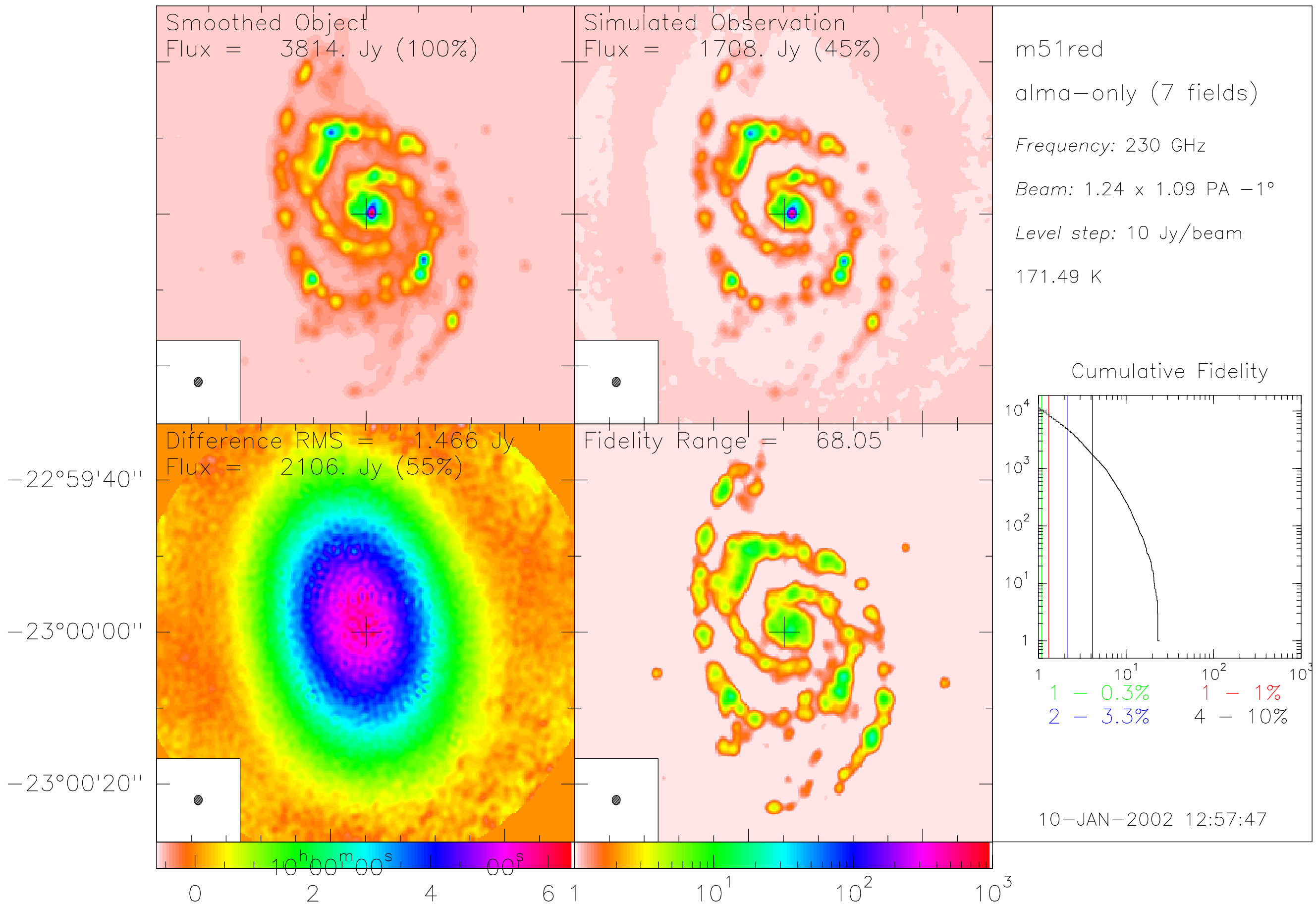
Mosaic Deconvolution: Separate Deconvolution

- Linear combination after separate field deconvolution



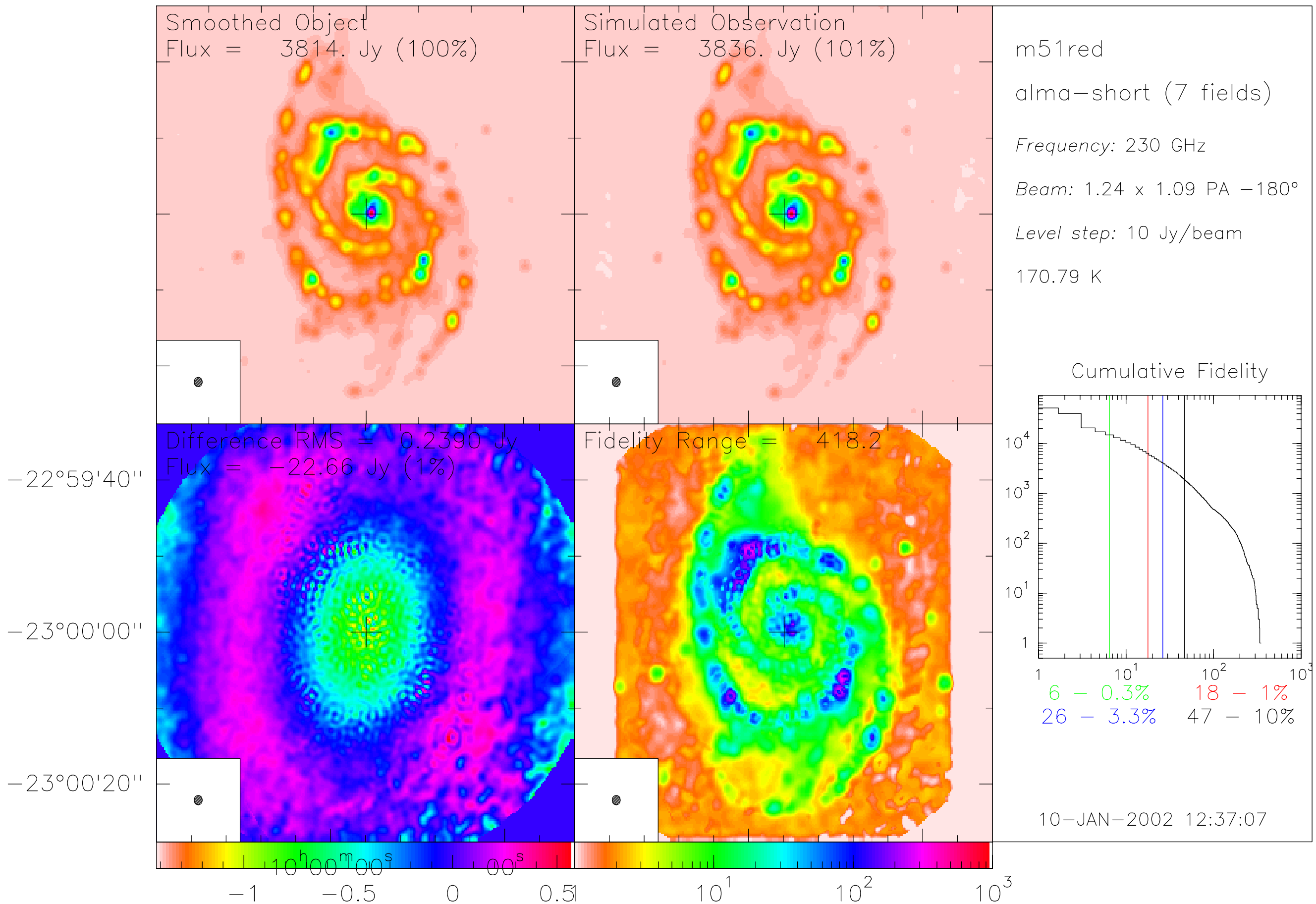
Mosaic Deconvolution: Joint Deconvolution

- Joint deconvolution by CLEAN adapted to mosaics (Gueth & Guilloteau 1999)



Mosaic Deconvolution: Zero Spacing

- Inclusion of single-dish data (zero spacing) does help



Mosaic Deconvolution: Short Spacings

- Short spacings form an array of smaller dishes ($d < D$) provides better information