Aperture Synthesis Imaging

(adopted slides from Rick Perley, various radio schools)

Huib Intema

VLA research associate

NRAO Socorro, NM

Associated Universities, Inc.

Atacama Large Millimeter/submillimeter Array
Karl G. Jansky Very Large Array
Robert C. Byrd Green Bank Telescope
Very Long Baseline Array

NRAO

Overview

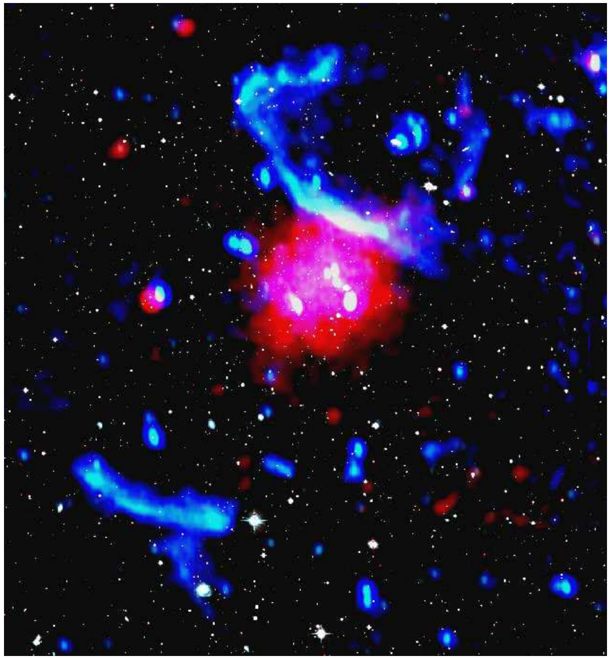
- Part 1: Review
- Part 2: Towards imaging
- Part 3: Imaging
- Start data analysis (Nadeem)
- Afternoon: Lab discussion(s) on calibration and/or other topics

Each part about 30-40 minutes

Additional information

- Synthesis Imaging in Radio Astronomy II (1999)
  editors Taylor, Carilli & Perley
  (pdf version available)
- Interferometry and Synthesis in Radio Astronomy, 2nd edition (2001)
  Thompson, Moran & Swenson
  (bit expensive)
- Tools of Radio Astronomy, 5th edition (2009)
  Wilson, Rohlfs & Huttemeister
  (pdf version available?)
- NRAO Synthesis Imaging Workshop lectures
  https://science.nrao.edu/science/meetings/2014/14th-synthesis-imaging-workshop/lectures
  http://www.aoc.nrao.edu/events/synthesis/2012/lectures.shtml
  …
  http://www.aoc.nrao.edu/events/synthesis/2004/presentations.html

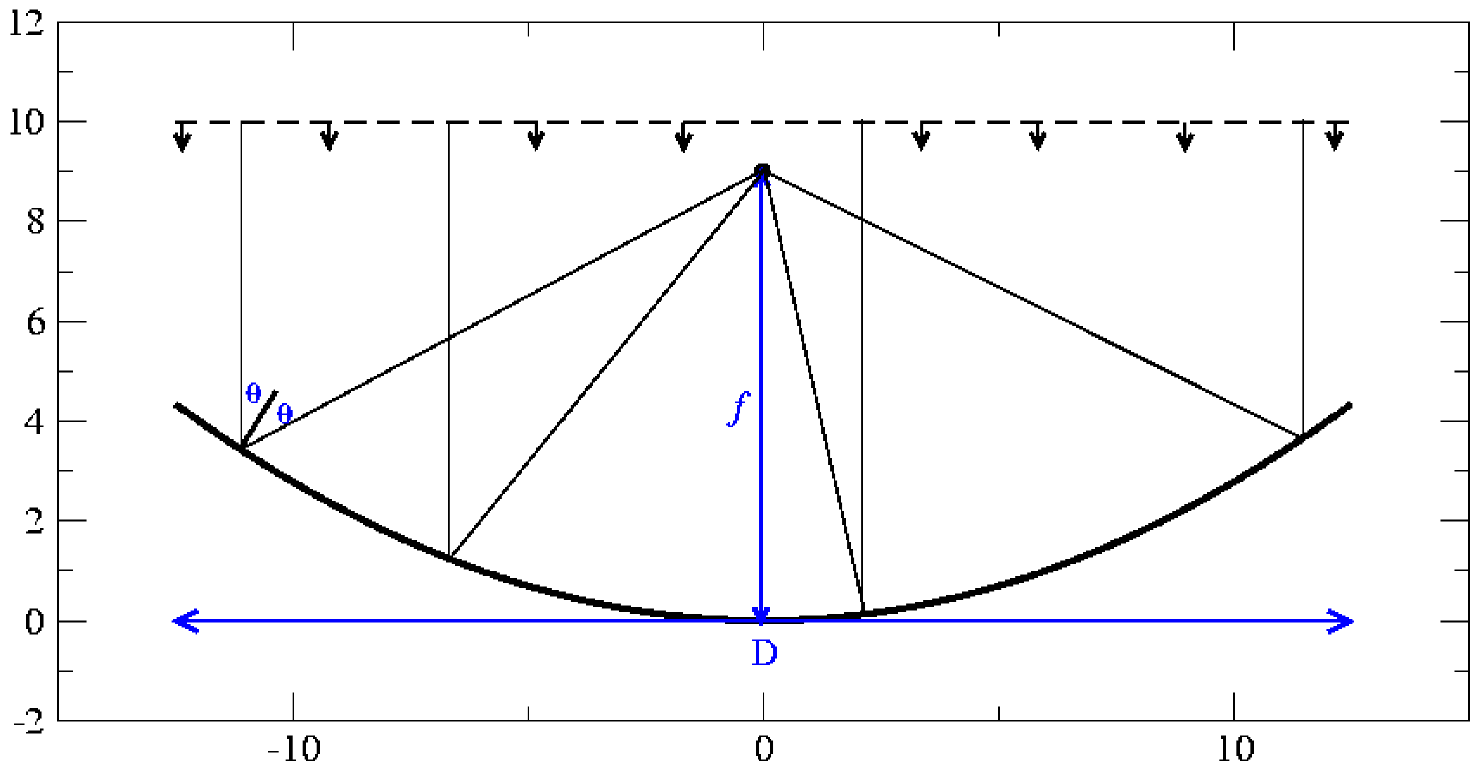
Part 1: Review



- Principles of aperture synthesis
- Sky response on a single baseline
- Definition and behaviour of the visibility
- Tracking the sky

The parabolic reflector

- Along the 'optical axis', the distance from incoming phase front to the focal point is the same for all rays.
- The E-fields will thus all be in phase at the focus – the place for the receiver. They will coherently add together.



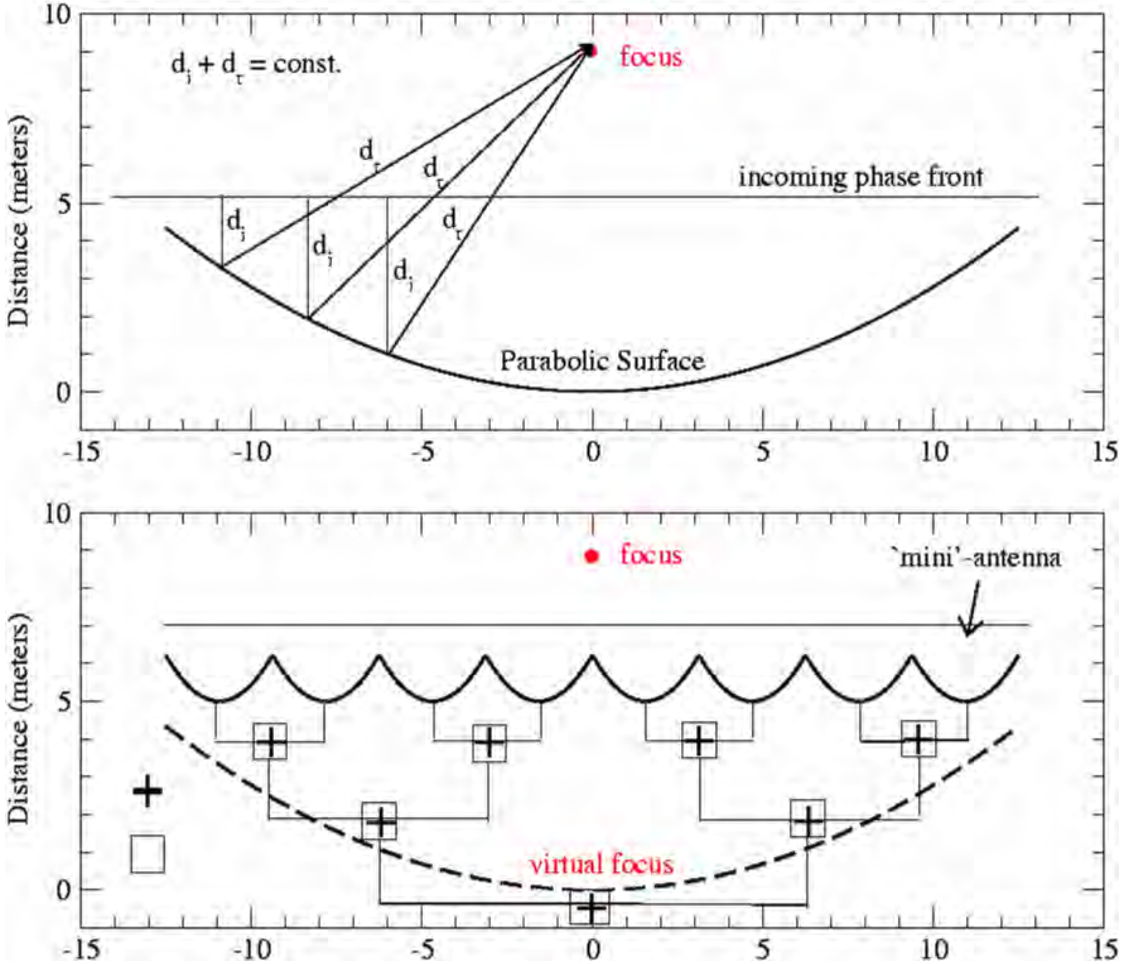
Why aperture synthesis?

- For an aperture of diameter D, and at wavelength λ, the image resolution is $$\theta_{rad} \approx \lambda / D$$
- In 'practical' units: $$\theta_{\mathrm{arcsec}} \approx 2\,\lambda_{\mathrm{cm}} / D_{\mathrm{km}}$$
- To obtain 1 arcsecond resolution at a wavelength of 21 cm (1.4 GHz), we require an aperture of ~42 km!
- The (currently) largest single, fully-steerable apertures are the 100 meter antennas near Bonn (GE), and at Green Bank (USA).

- So we must develop a method of synthesizing an equivalent aperture.
- The methodology of synthesizing a quasi-continuous aperture through summations of separated pairs of antennas is called aperture synthesis.

Aperture synthesis - concept

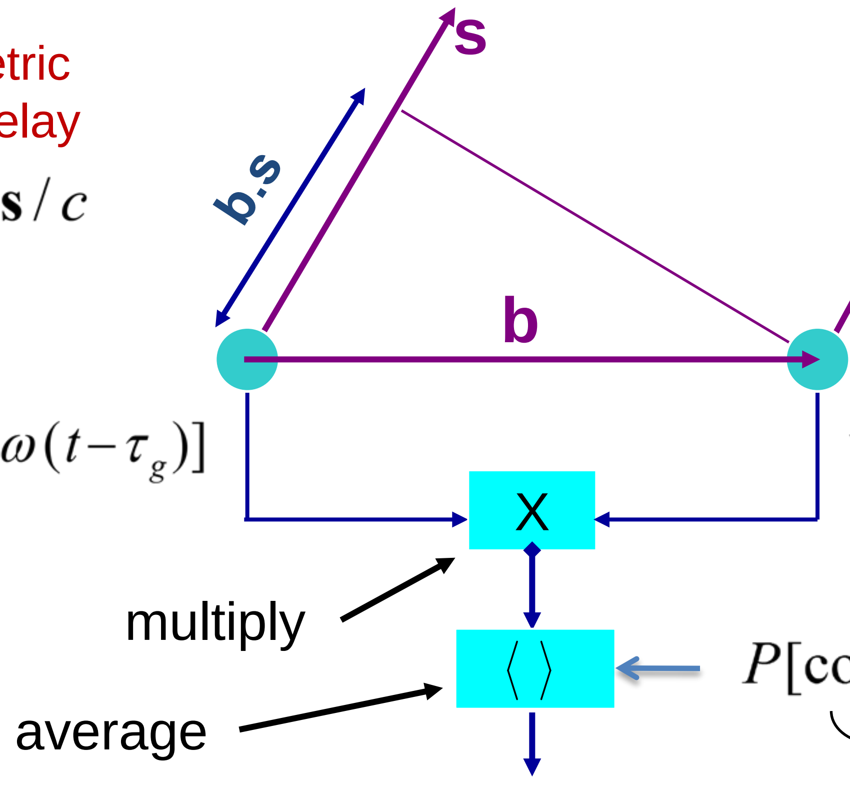
- We don't need a single parabolic structure.
- We can consider a series of small antennas, whose individual signals are summed in a network.
- This is the basic concept of interferometry.
- Aperture Synthesis is an extension of this concept.



Single baseline interferometer

Geometric Time Delay

$$\tau_g = \mathbf{b} \cdot \mathbf{s} / c$$



$$V_1 = E\cos[\omega(t-\tau_g)]$$

$$V_2 = E\cos(\omega t)$$

The path lengths from sensors to multiplier are assumed equal!

$$P[\cos(\omega\tau_g) + \cos(2\omega t - \omega\tau_g)]$$

Unchanging

Rapidly varying,

$$R_C = P\cos(\omega\tau_g)$$

Some general comments

- The averaged product R_C is dependent on the received power, $P = E^2/2$ and geometric delay, τ_g = **b**.**s**/c, and hence on the baseline orientation and source direction:

$$R_C = P\cos(\omega\tau_g) = P\cos\left(2\pi\frac{\mathbf{b}\cdot\mathbf{s}}{\lambda}\right)$$

- Note that R_C is not a a function of:
  - The time of the observation -- provided the source itself is not variable.
  - The location of the baseline -- provided the emission is in the far-field.
  - The actual phase of the incoming signal – the distance of the source does not matter, provided it is in the far-field.

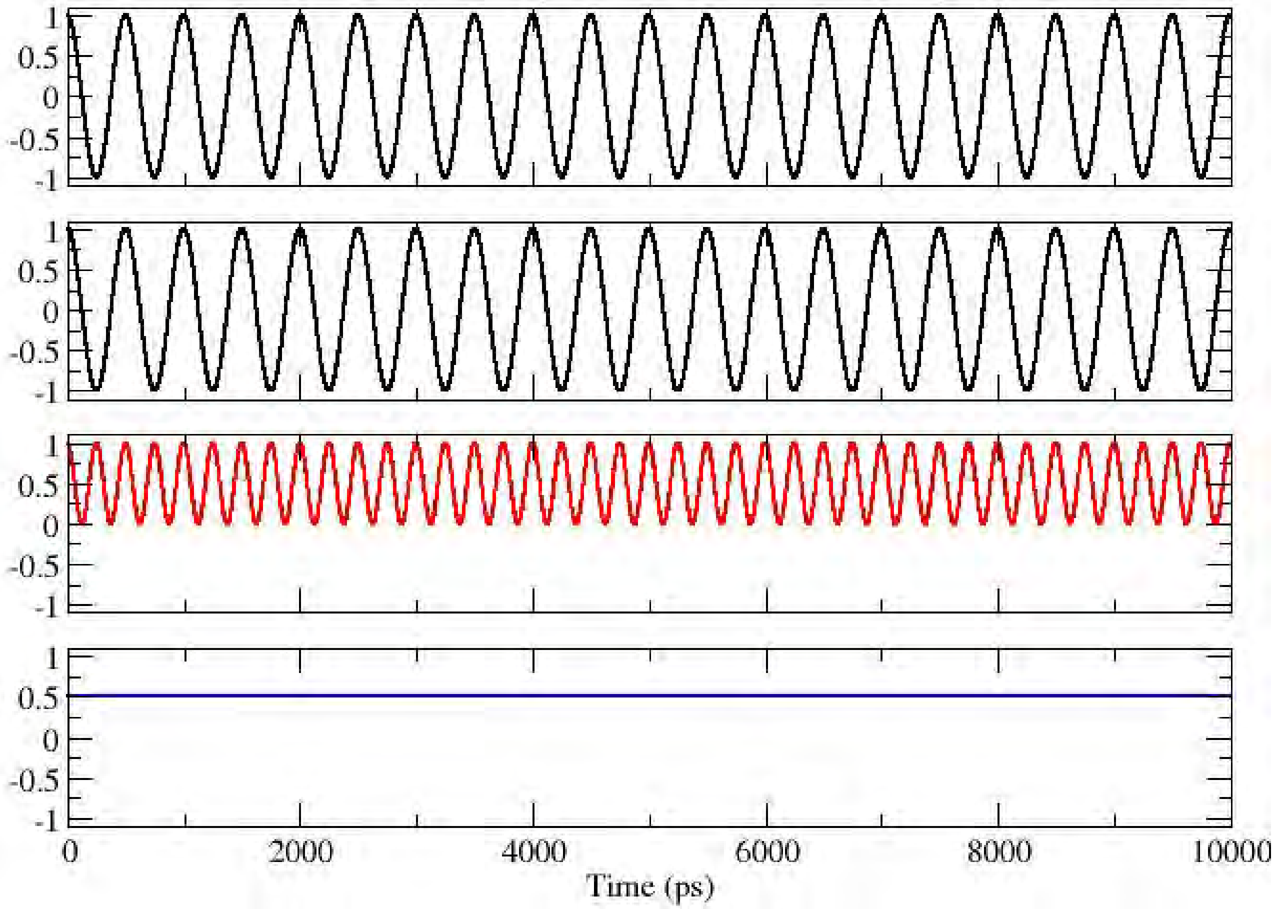
- The strength of the product is dependent on the antenna collecting areas and electronic gains – but these factors can be calibrated for.

Pictorial example: signals in phase

2 GHz Frequency, with voltages in phase:

b.s = nλ, or $\tau_g = n/\nu$

- Antenna 1 Voltage
- Antenna 2 Voltage
- Product Voltage
- Average

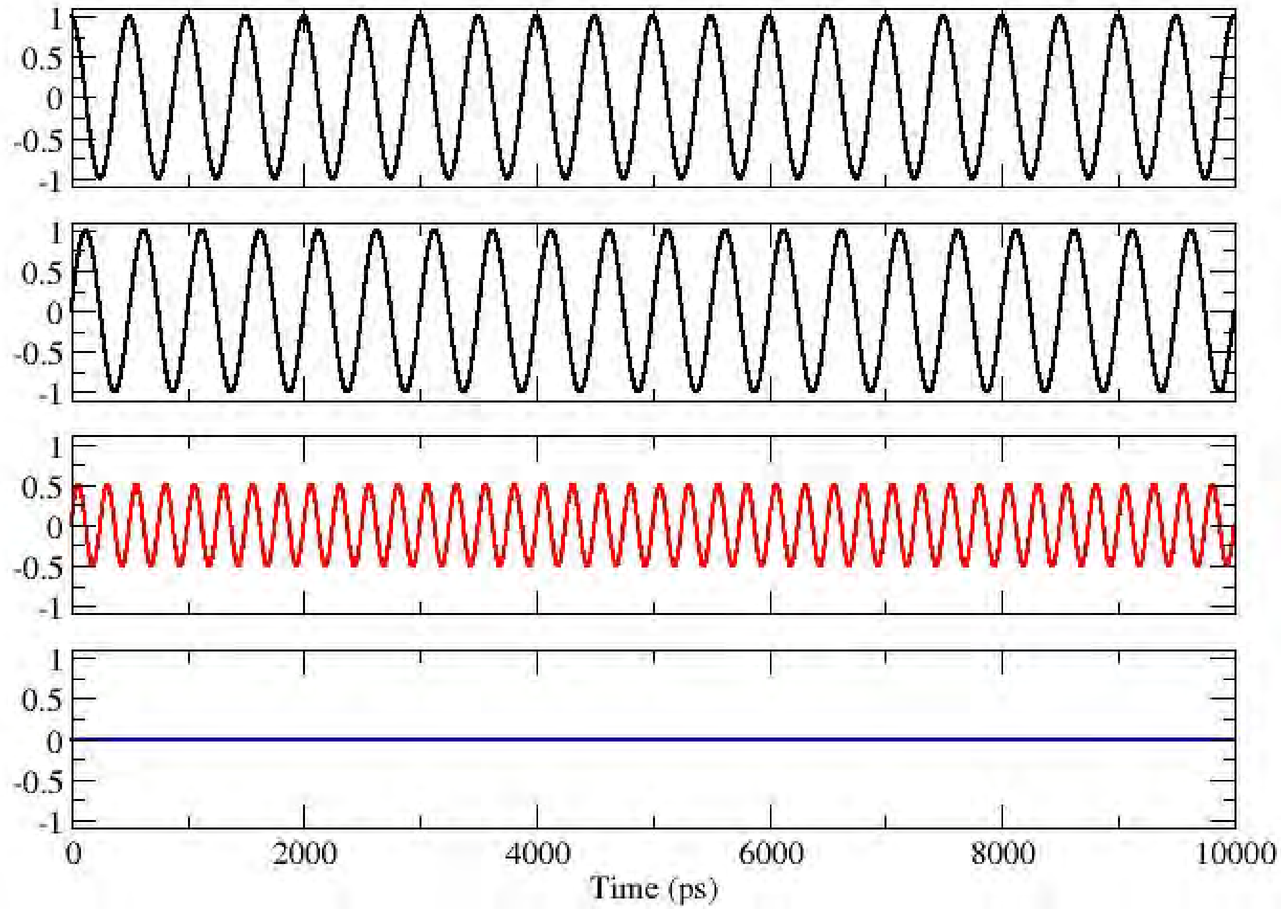


Pictorial example: signals in quad phase

2 GHz Frequency, with voltages in quadrature phase:

$b.s=(n \pm 1/4)\lambda$, $\tau_g = (4n \pm 1)/4\nu$

- Antenna 1 Voltage
- Antenna 2 Voltage
- Product Voltage
- Average

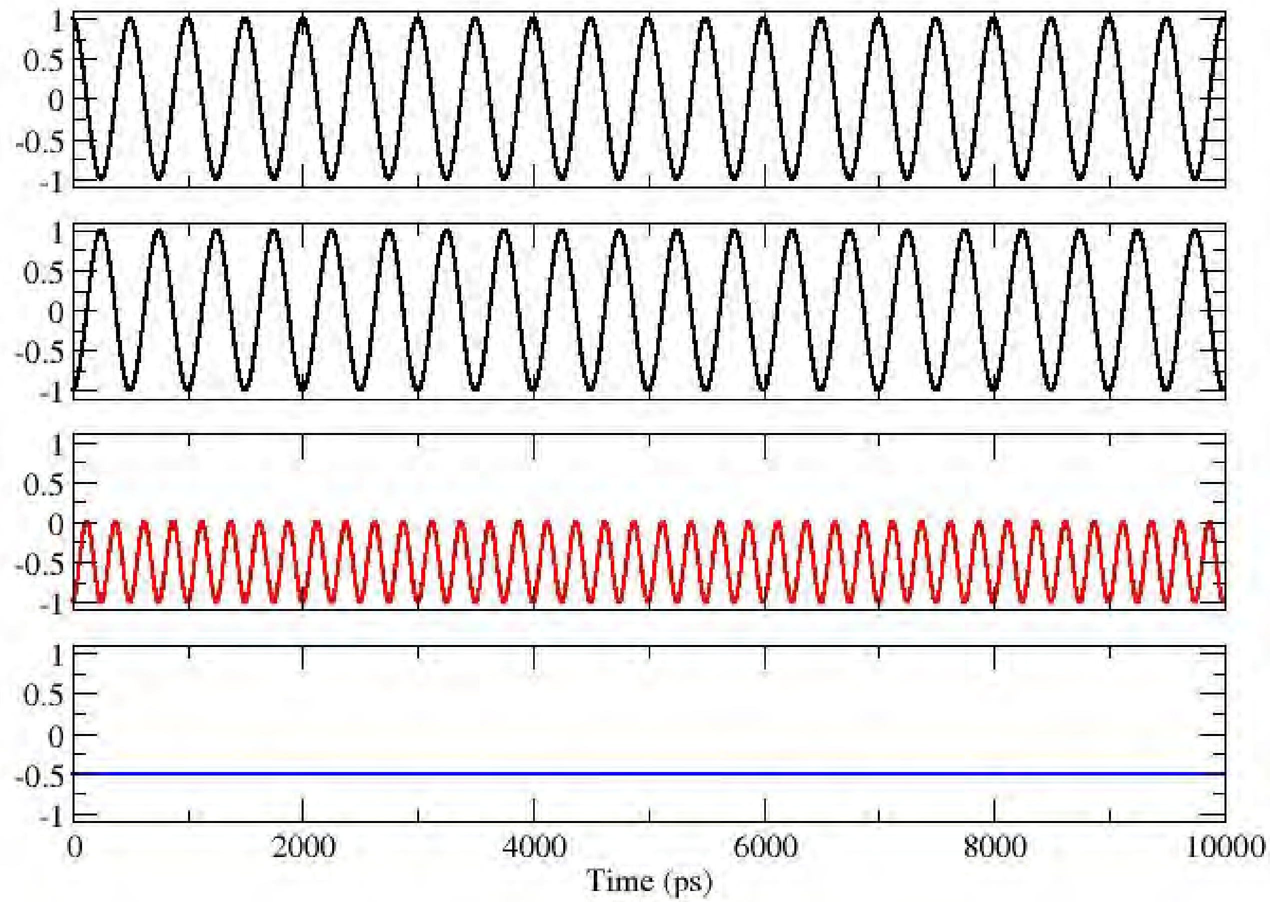


Pictorial example: signals out of phase

2 GHz Frequency, with voltages out of phase:

$b.s=(n \pm \frac{1}{2})\lambda \qquad \tau_g = (2n \pm 1)/2\nu$

- Antenna 1 Voltage
- Antenna 2 Voltage
- Product Voltage
- Average

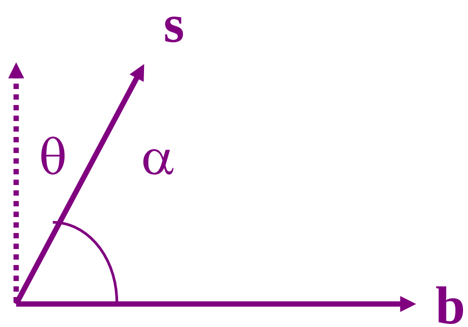


Pictorial illustrations

- To illustrate the response, expand the dot product in one dimension:

$$\frac{\mathbf{b} \bullet \mathbf{s}}{\lambda} = u \cos\alpha = u \sin\theta = ul$$

- Here, **u = b/λ** is the **baseline length in wavelengths**, and θ is the angle w.r.t. the plane perpendicular to the baseline.
- $l = \cos\alpha = \sin\theta$ is the **direction cosine**



- Consider the response R_c, as a function of angle, for two different baselines with u = 10, and u = 25 wavelengths:

$$R_C = \cos(20\pi l)$$

Whole-Sky Response

- Top: u = 10

$$R_C = \cos(20\pi l)$$

There are 20 whole fringes over the hemisphere.

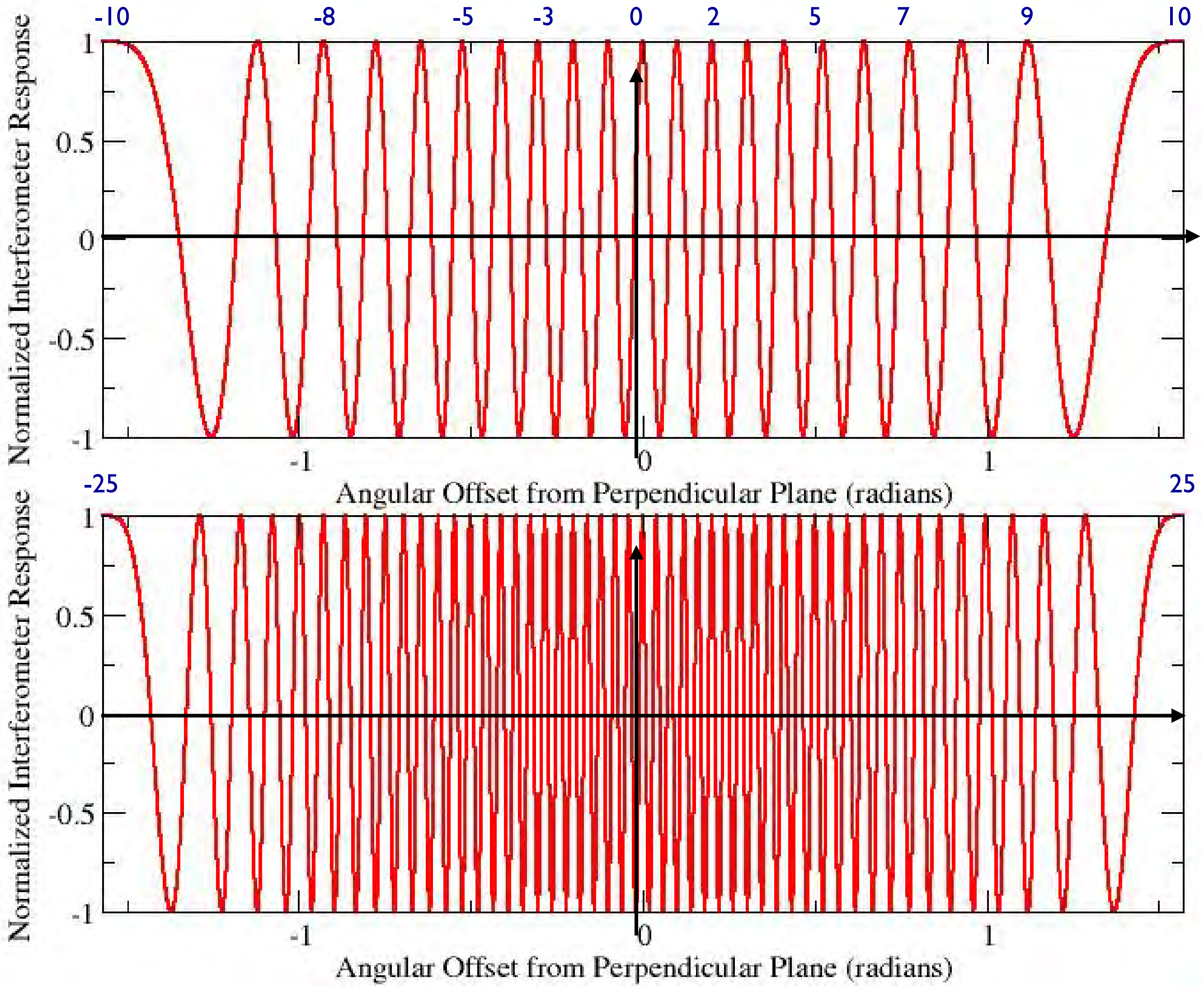
Peak separation 1/10 radians

- Bottom: u = 25

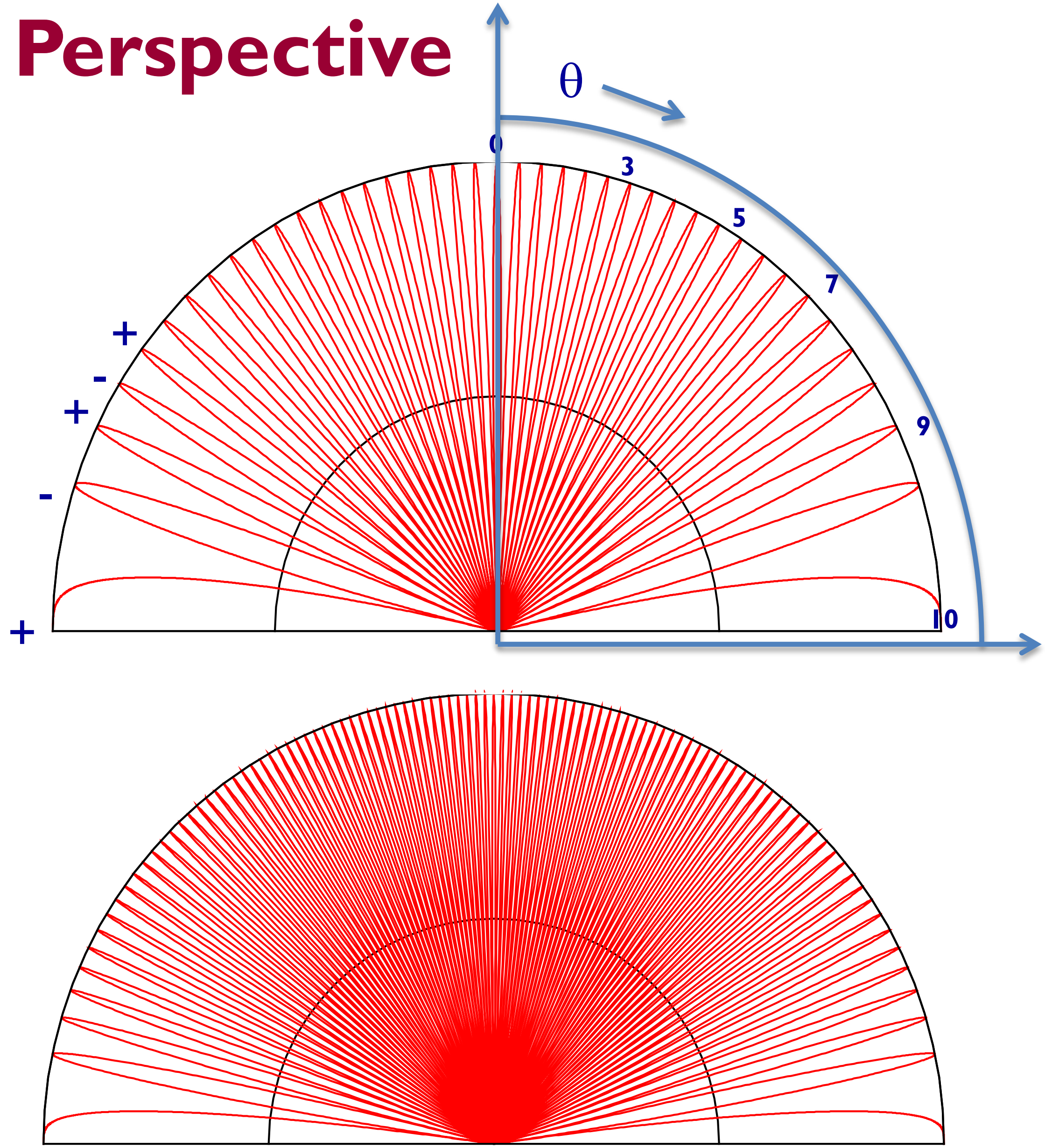
$$R_C = \cos(50\pi l)$$

There are 50 whole fringes over the hemisphere.

Peak separation 1/25 radians.



From an Angular Perspective



Top Panel:

The absolute value of the response for u = 10, as a function of angle.

The 'lobes' of the response pattern alternate in sign.

Bottom Panel:

The same, but for u = 25.

Angular separation between lobes (of the same sign) is

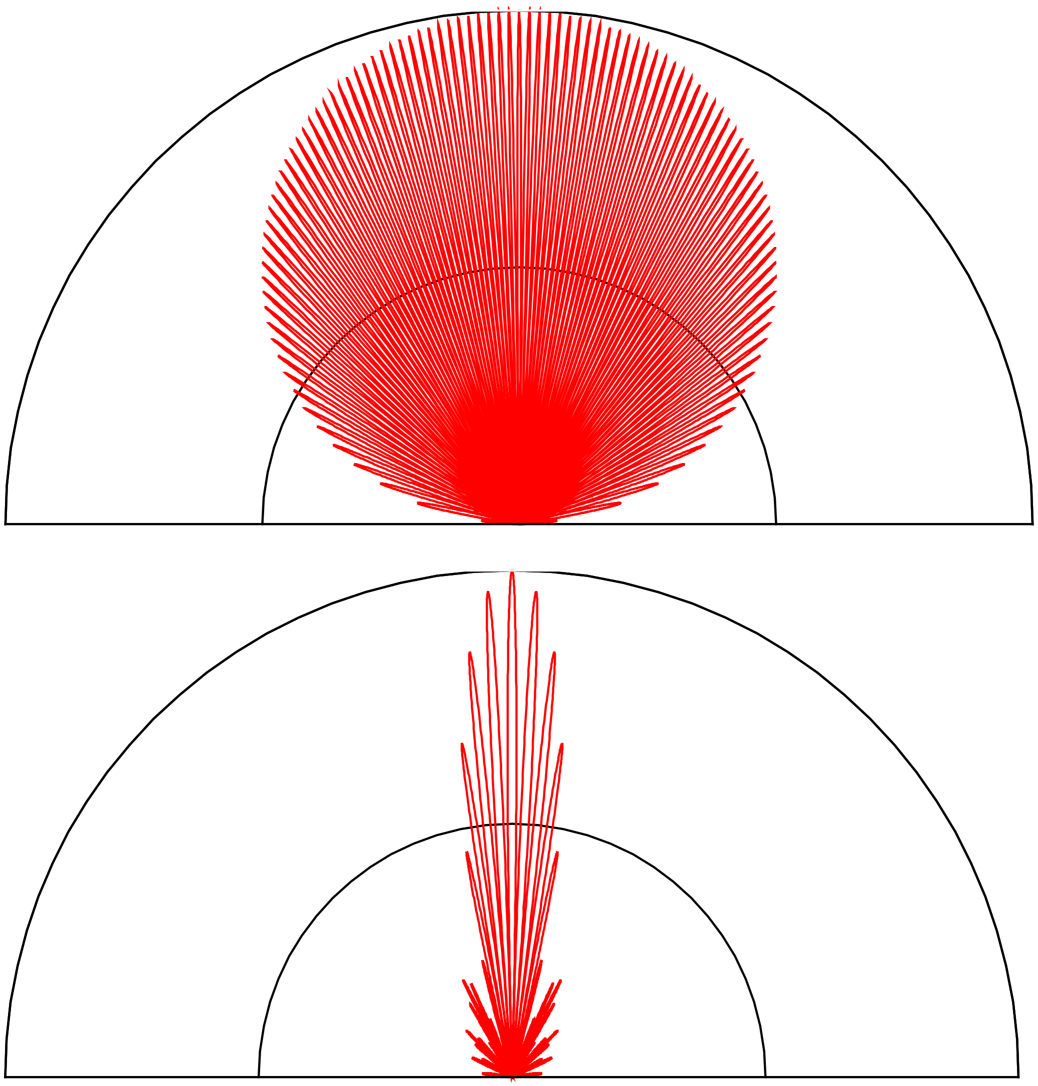
$\delta\theta \sim 1/u = \lambda/b$ radians.

Hemispheric pattern

- The preceding plot is a meridional cut through the hemisphere, oriented along the baseline vector.
- In the two-dimensional space, the fringe pattern consists of a series of coaxial cones, oriented along the baseline vector.
- The figure is a two-dimensional representation when u = 4.
- As viewed along the baseline vector, the fringes show a 'bulls-eye' pattern – concentric circles.

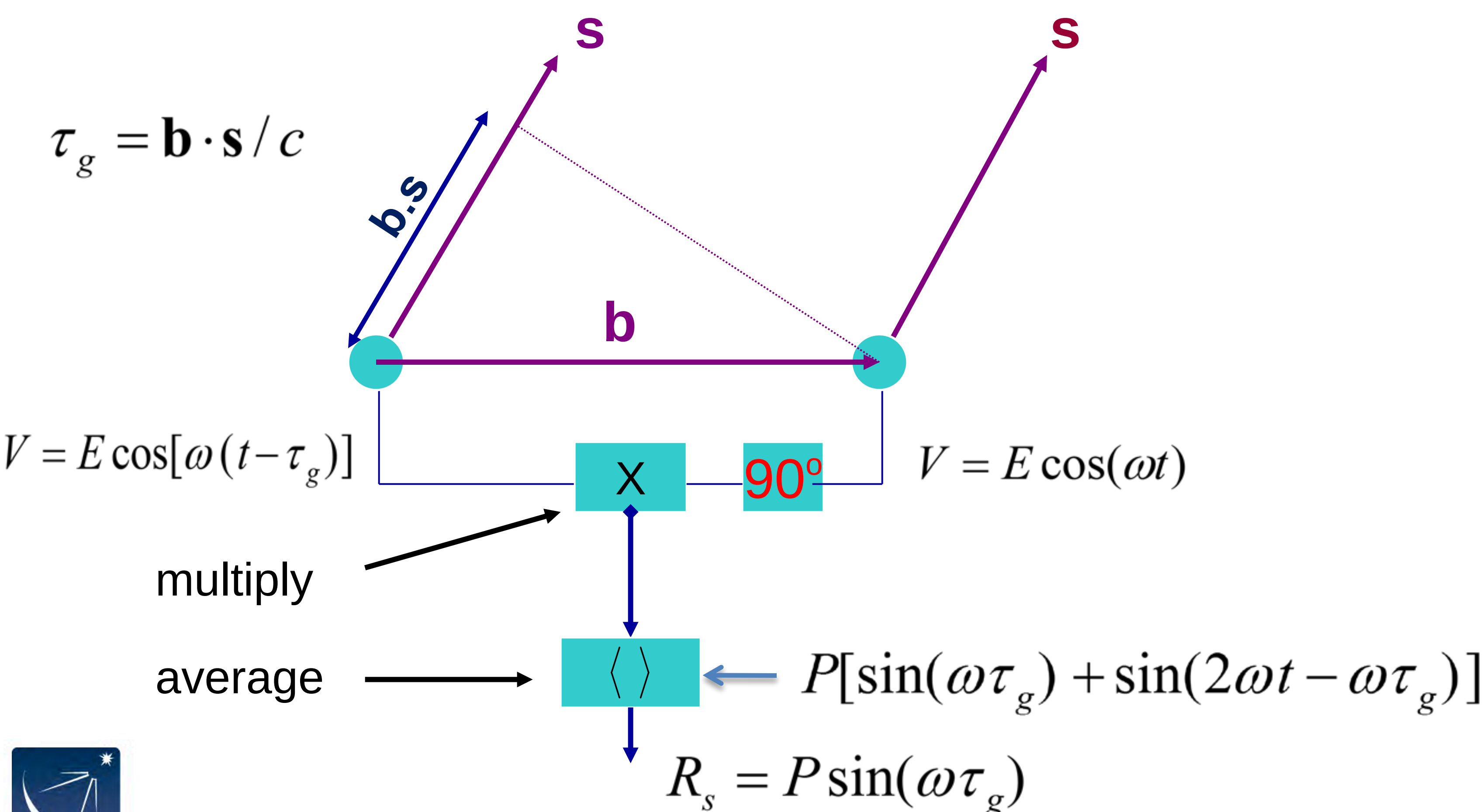
The effect of antenna patterns

- Real sensors (or antennas) are not isotropic, but have their own responses (beam pattern).
- **Top Panel:** The interferometer pattern with a cos(θ)-like sensor response.
- **Bottom Panel:** A multiple-wavelength aperture antenna has a narrow beam, but also sidelobes.



Adding a complementary correlator

- We generate the ‘sine’ pattern by inserting a 90 degree phase shift in one of the signal paths.



NRAO

Define the complex visibility

- We now DEFINE a complex function, the complex visibility, V, from the two independent (real) correlator outputs R_C and R_S:

$$V = R_C - iR_S \equiv Ie^{-i\phi}$$

where

$$I = \sqrt{R_C^2 + R_S^2}, \quad \phi = \tan^{-1}\left(\frac{R_S}{R_C}\right)$$

- This leads to the Fourier relationship between the sky brightness I, and the response V of an interferometer:

$$V_\nu(\mathbf{b}) = R_C - iR_S = \iint I_\nu(\mathbf{s}) e^{-2\pi i \nu \mathbf{b}\cdot\mathbf{s}/c} d\Omega$$

- Rewritten in more familiar terms of **u** and **l**:

$$V_\nu(\mathbf{u}) = \iint I_\nu(\mathbf{l}) e^{-2\pi i \mathbf{u}\cdot\mathbf{l}} d\mathbf{l} \Leftrightarrow I_\nu(\mathbf{l}) = \iint V_\nu(\mathbf{u}) e^{+2\pi i \mathbf{u}\cdot\mathbf{l}} d\mathbf{u}$$

- With the right geometry, this is a 2-D Fourier transform, giving us a well established way to recover $I(\mathbf{l})$ from $V(\mathbf{u})$.

Some comments on visibilities

- The visibility function is a unique function of the sky brightness.
- The two functions are related through a Fourier transform.
- A single-baseline interferometer, at any one time, makes one measure of the visibility function.
- `Sufficient knowledge' of the visibility function (as derived from an interferometer) will provide us a `reasonable estimate' of the sky brightness.
- How many is 'sufficient', and how good is 'reasonable'?
- These simple questions do not have easy answers…

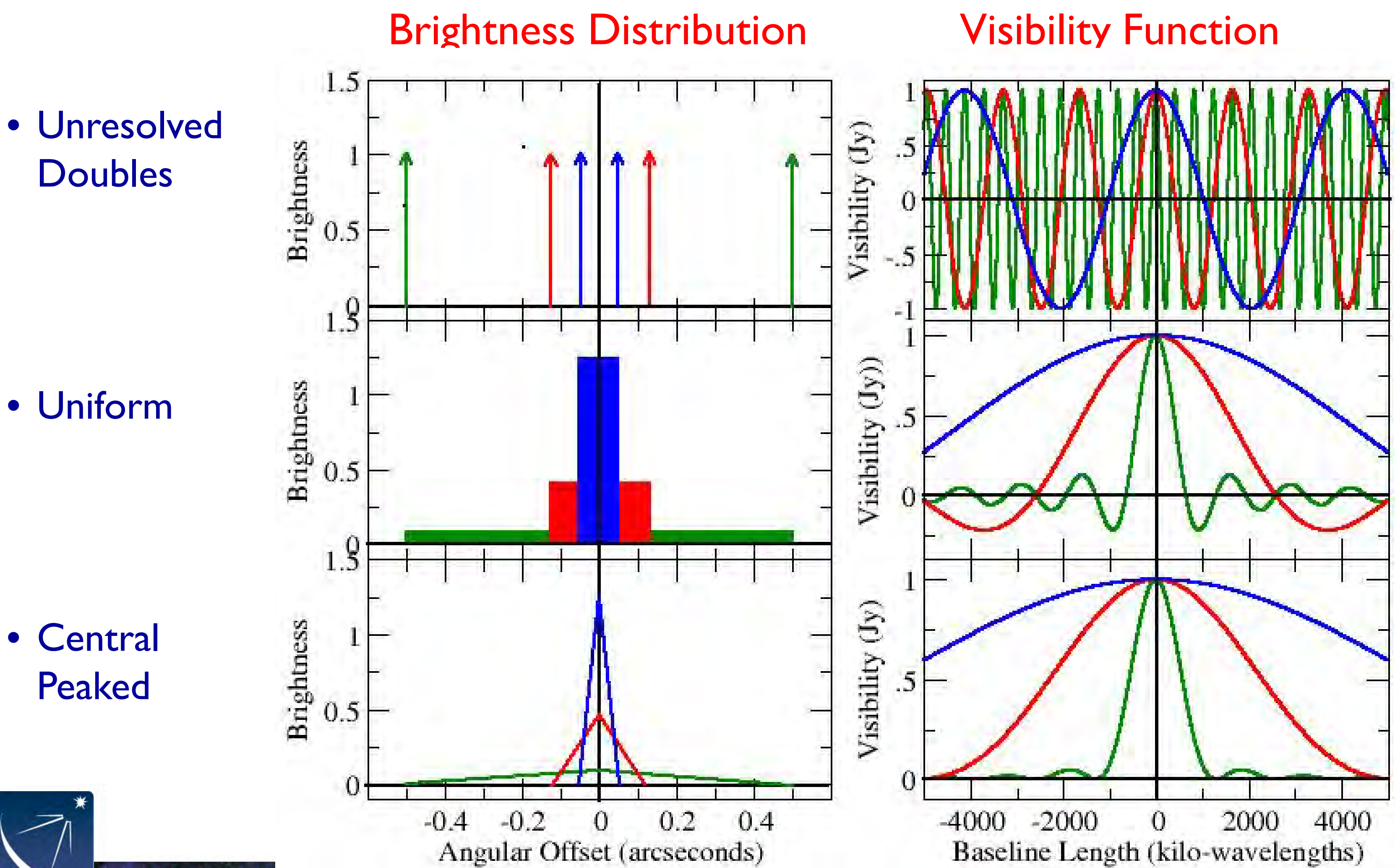
Basic Characteristics of the Visibility

- For a zero-spacing interferometer, we get the 'single-dish' (total-power) response.
- As the baseline gets longer, the visibility amplitude will in general decline.
- When the visibility is close to zero, the source is said to be 'resolved out'.
- Interchanging antennas in a baseline causes the phase to be negated – the visibility of the 'reversed baseline' is the complex conjugate of the original.
- Mathematically, the visibility is Hermitian.

$$V_\nu(\mathrm{u}) = V_\nu^*(-\mathrm{u})$$

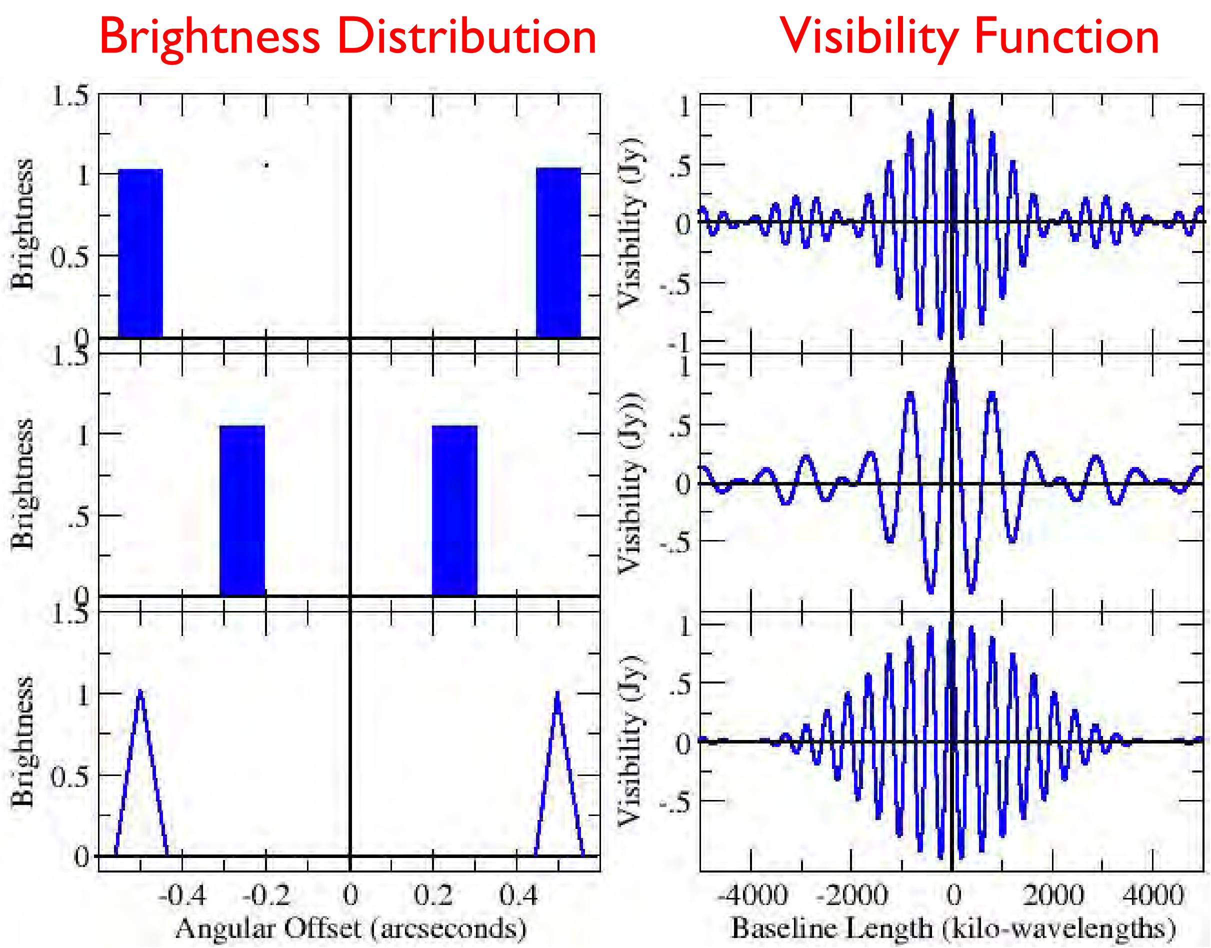
Examples of 1-dimensional visibilities

- Simple pictures are easy to make illustrating 1-dimensional visibilities.



More examples

- Simple pictures are easy to make illustrating 1-dimensional visibilities.



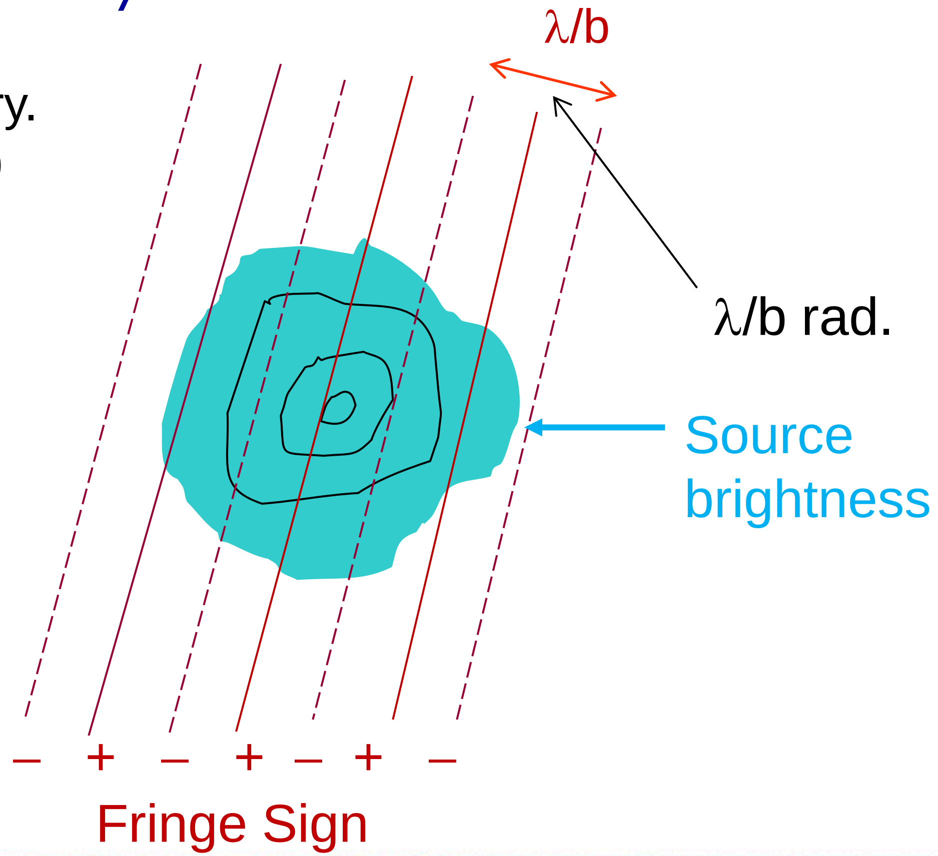
- Resolved Double
- Resolved Double
- Central Peaked Double

The response from an extended source

- The correlator can be thought of 'casting' sinusoidal coherence patterns, of angular scale ~λ/b radians, onto the sky.

- The correlator multiplies the source brightness by its coherence pattern, and integrates (sums) the result over the sky.

- Orientation set by baseline geometry.
- Fringe separation set by (projected) baseline length and wavelength.
  - Long baseline gives close-packed fringes
  - Short baseline gives widely-separated fringes
- Physical location of baseline unimportant, provided source is in the far field.



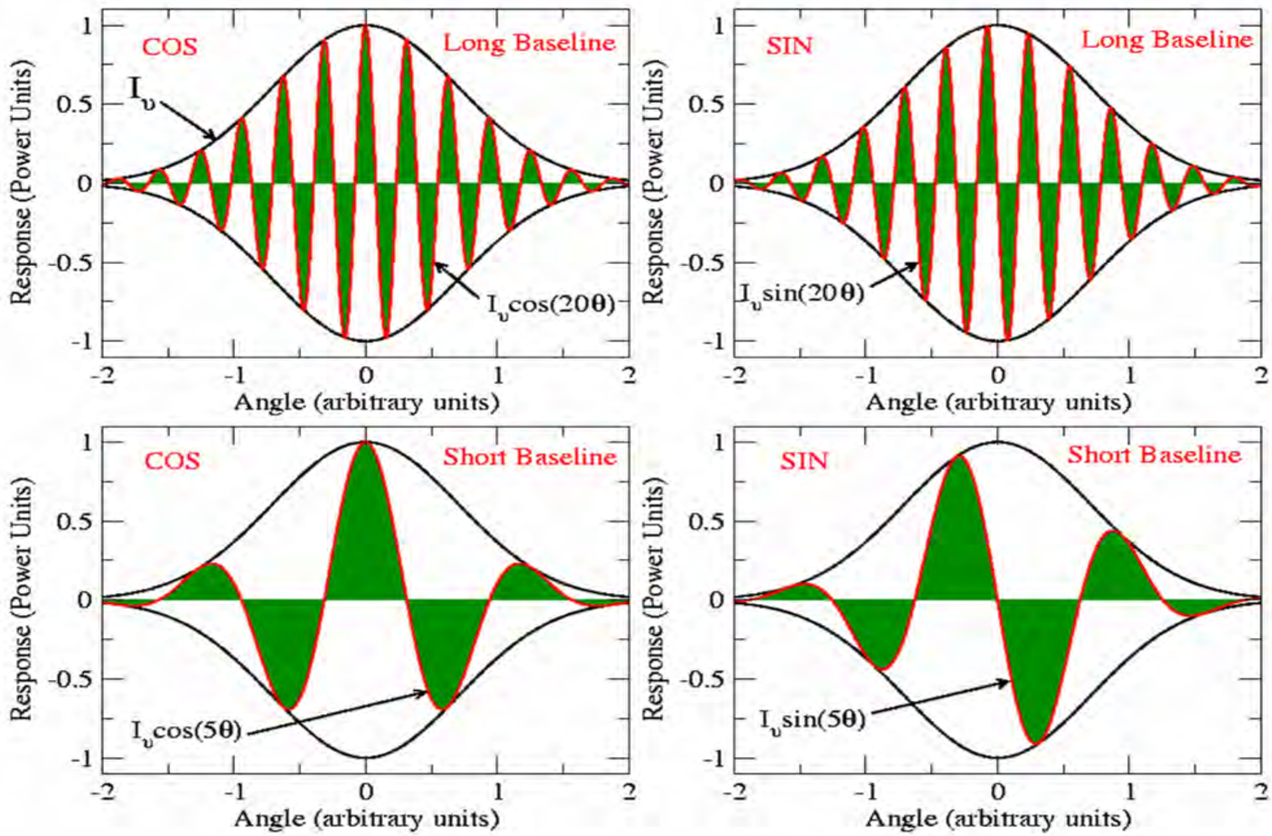
Picturing the visibility

- The source brightness is Gaussian, shown in black.
- The interferometer 'fringes' are in red.
- The visibility is the integral of the product – the net dark green area.

R_C R_S

Long baseline (large b)

Short baseline (small b)



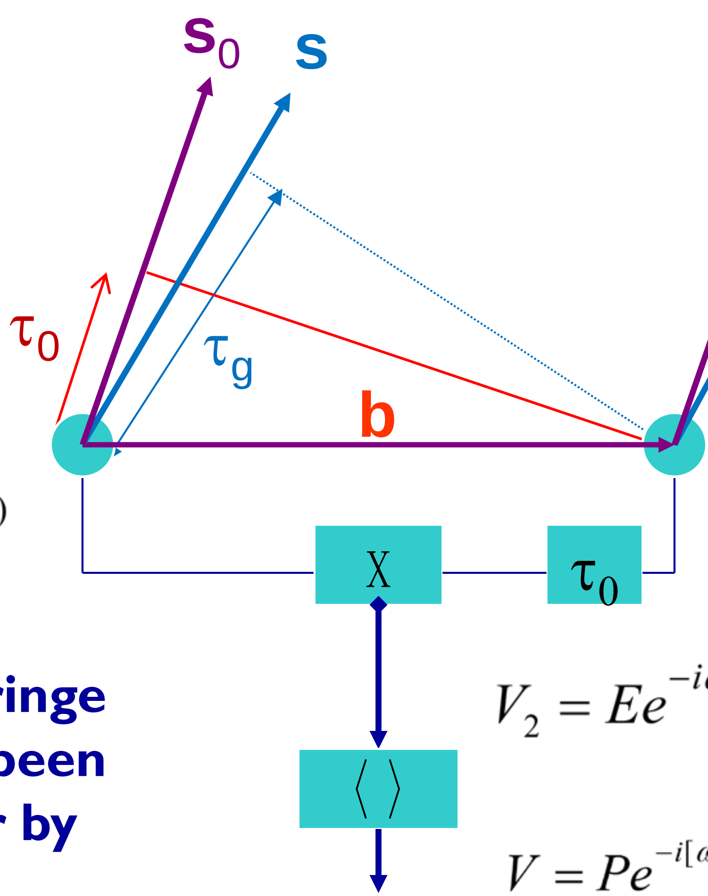
Observations from Earth

- Real interferometers are built on the surface of the earth – a rotating platform. From the observer's perspective, sources move across the sky.
- Since we know how to adjust the interferometer to move its fringe pattern to the direction of interest, it is a simple step to continuously move the pattern to follow a moving source.
- All that is necessary is to continuously slip the inserted time delay, with an accuracy $\delta\tau << 1/\Delta\nu$ to minimize bandwidth loss.
- A point source, at the reference position, will give uniform amplitude and zero phase throughout time (provided real-life things like the instrument, atmosphere, ionosphere, or geometry errors don't mess things up)

Adding time delay

$$\tau_g = \mathbf{b} \cdot \mathbf{s} / c$$

$$\tau_0 = \mathbf{b} \cdot \mathbf{s}_0 / c$$



s_0 = reference (delay) direction

s = general direction

$$V_1 = Ee^{-i\omega(t-\tau_g)}$$

$$V_2 = Ee^{-i\omega t}$$

$$V_2 = Ee^{-i\omega(t-\tau_0)}$$

$$V = Pe^{-i[\omega(\tau_0-\tau_g)]} = Pe^{i2\pi[\nu\mathbf{b}\cdot(\mathbf{s}-\mathbf{s}_0)/c]}$$

The entire fringe pattern has been shifted over by angle

$\sin\theta = c\tau_0/b$

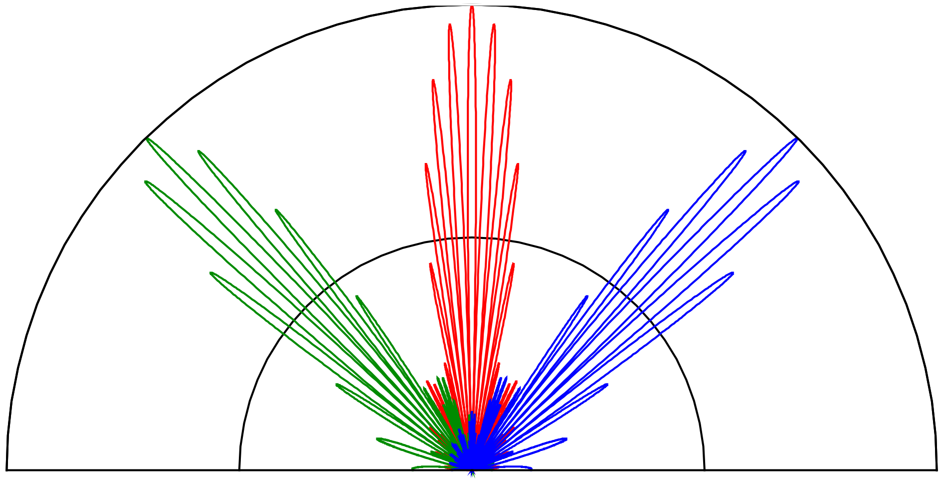
Illustrating delay tracking

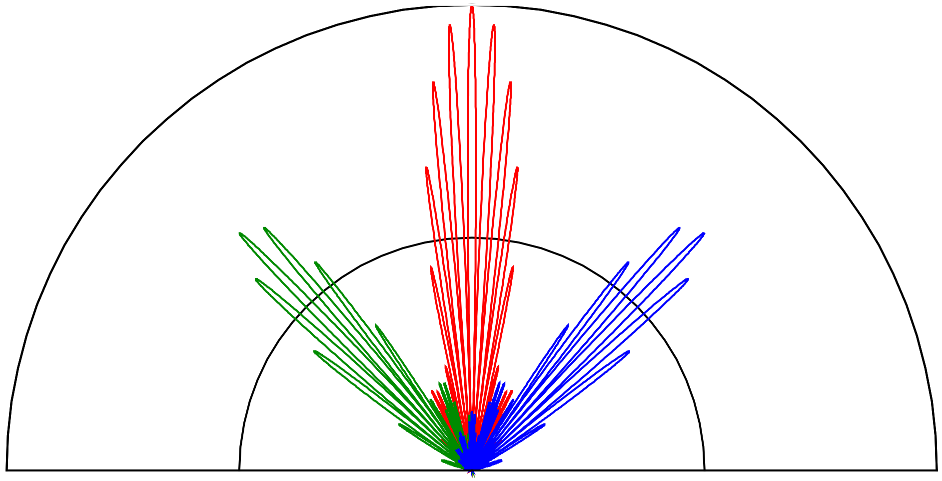
- Top Panel:

  Delay has been added and subtracted to move the delay pattern to the source location.

- Bottom Panel:

  A cosinusoidal sensor pattern is added, to illustrate losses from a fixed sensor.





The three centers in interferometry

1. **Beam tracking (pointing) center**: Where the antennas are pointing to (or, for phased arrays, the phased array center position).
2. **Delay tracking center:** The location for which the delays are being set for maximum wide-band coherence.
3. **Phase tracking center:** The location for which the phase is set in order to track the coherence pattern.

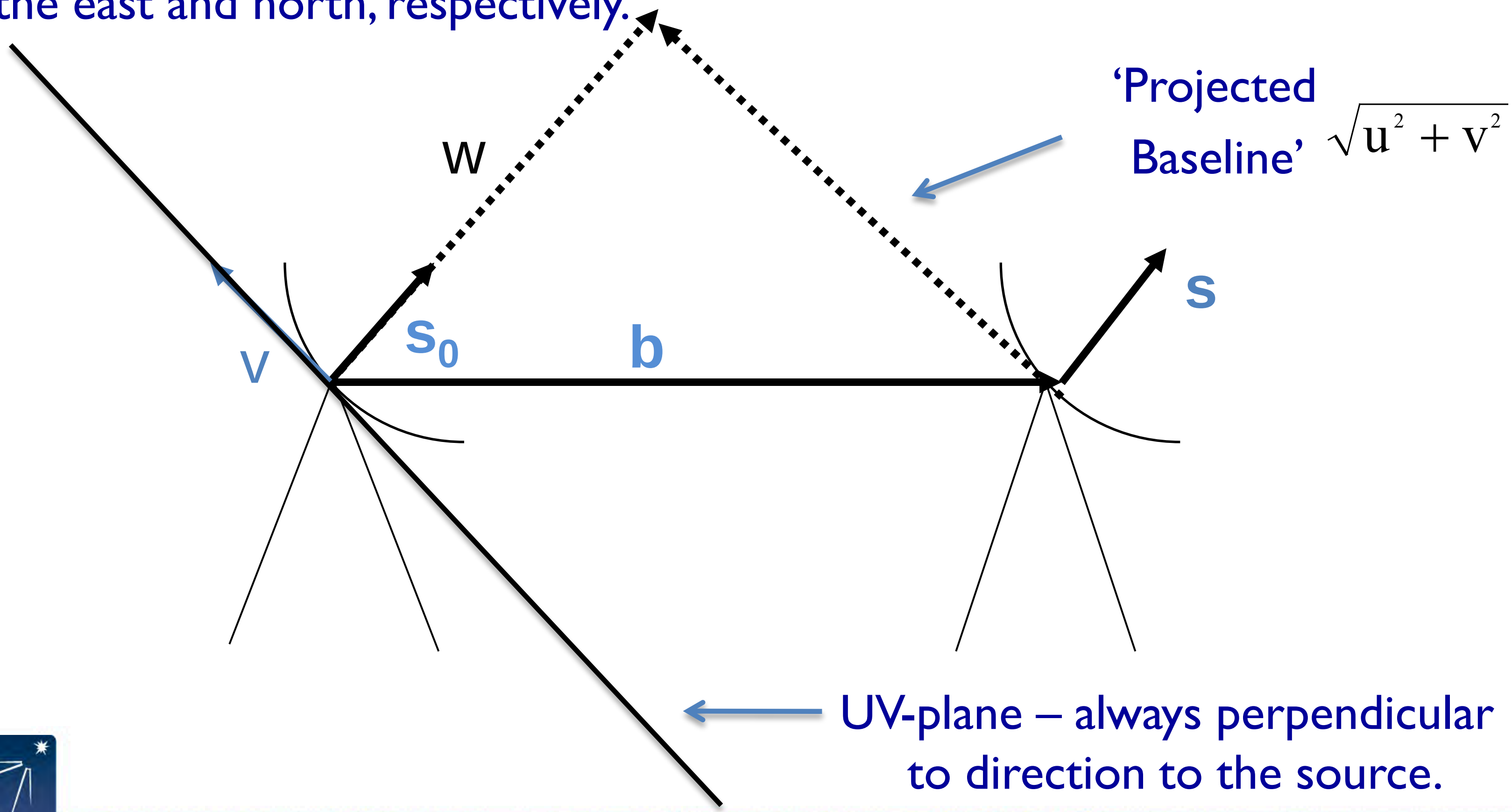
- Note: generally, we make all three the same.

Part 2: Towards imaging

- UV-coverage and dirty beam
- Weighting

General coordinate system

- This is the coordinate system in most general use for synthesis imaging.
- **w** points to, and follows the source, **u** towards the east, and **v** towards the north celestial pole. The direction cosines *l* and *m* then increase to the east and north, respectively.



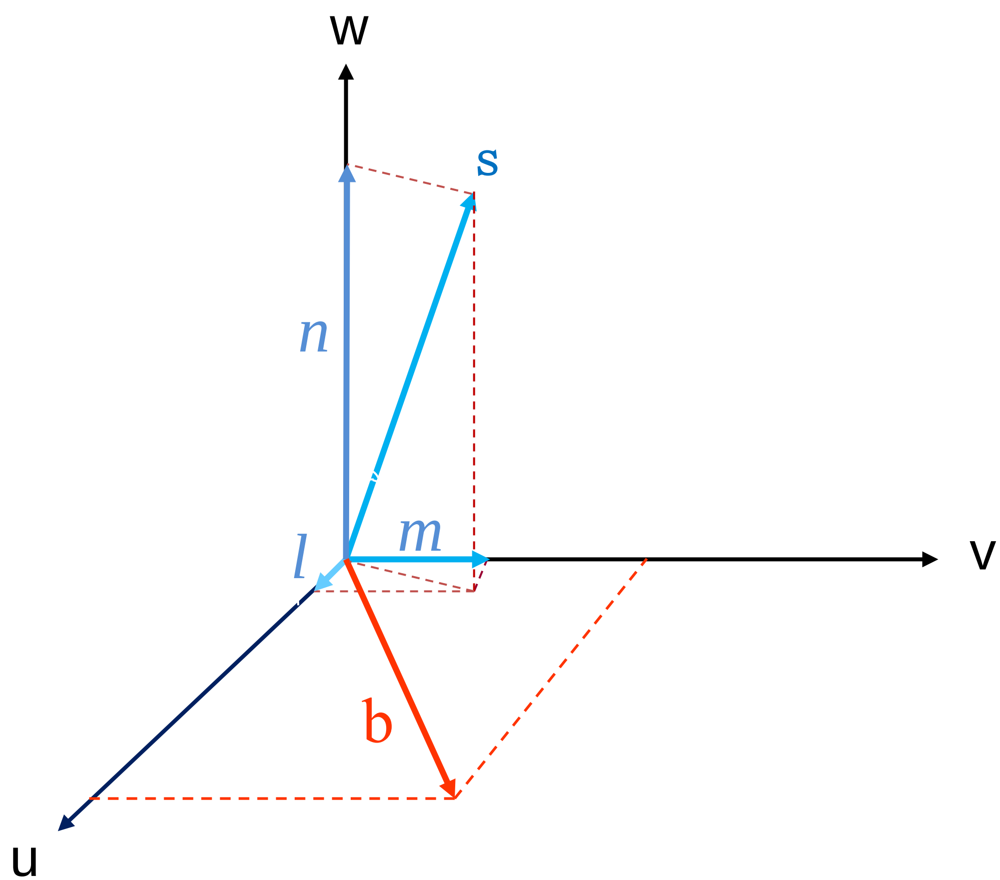
Direction cosines – describing the source

The unit direction vector **s** is defined by its projections (l,m,n) on the (u,v,w) axes. These components are called the **direction cosines**.

$$l = \cos(\alpha)$$

$$m = \cos(\beta)$$

$$n = \cos(\theta) = \sqrt{1 - l^2 - m^2}$$



The angles, α, β, and θ are between the direction vector and the three axes.

Coverage of the UV-plane

- Obtaining a good image of a source requires adequate sampling ('coverage') of the UV-plane.
- Adopt an earth-based coordinate grid to describe the antenna positions:
  - X points to H=0h, $\delta=0^\circ$ (intersection of meridian and celestial equator)
  - Y points to H = -6h, $\delta = 0^\circ$ (to east, on celestial equator)
  - Z points to $\delta = 90^\circ$ (to NCP).
- Then denote by (Bx, By, Bz) the coordinates, measured in wavelengths, of a baseline in this earth-based frame.

- (Bx, By) are the projected coordinates of the baseline (in wavelengths) on the equatorial plane of the earth.
- By is the East-West component
- Bz is the baseline component up the Earth's rotational axis.

Coverage of the UV-plane (2)

- Then, it can be shown that

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \sin H_0 & \cos H_0 & 0 \\ -\sin\delta_0 \cos H_0 & \sin\delta_0 \sin H_0 & \cos\delta_0 \\ \cos\delta_0 \cos H_0 & -\cos\delta_0 \sin H_0 & \sin\delta_0 \end{pmatrix} \begin{pmatrix} B_X \\ B_Y \\ B_Z \end{pmatrix}$$

- The u and v coordinates describe E-W and N-S components of the projected interferometer baseline.
- The w coordinate is the delay distance in wavelengths between the two antennas. The geometric delay, τ_g is given by

$$\tau_g = \frac{\lambda}{c} w = \frac{w}{\nu}$$

- Its derivative, called the fringe frequency ν_F is

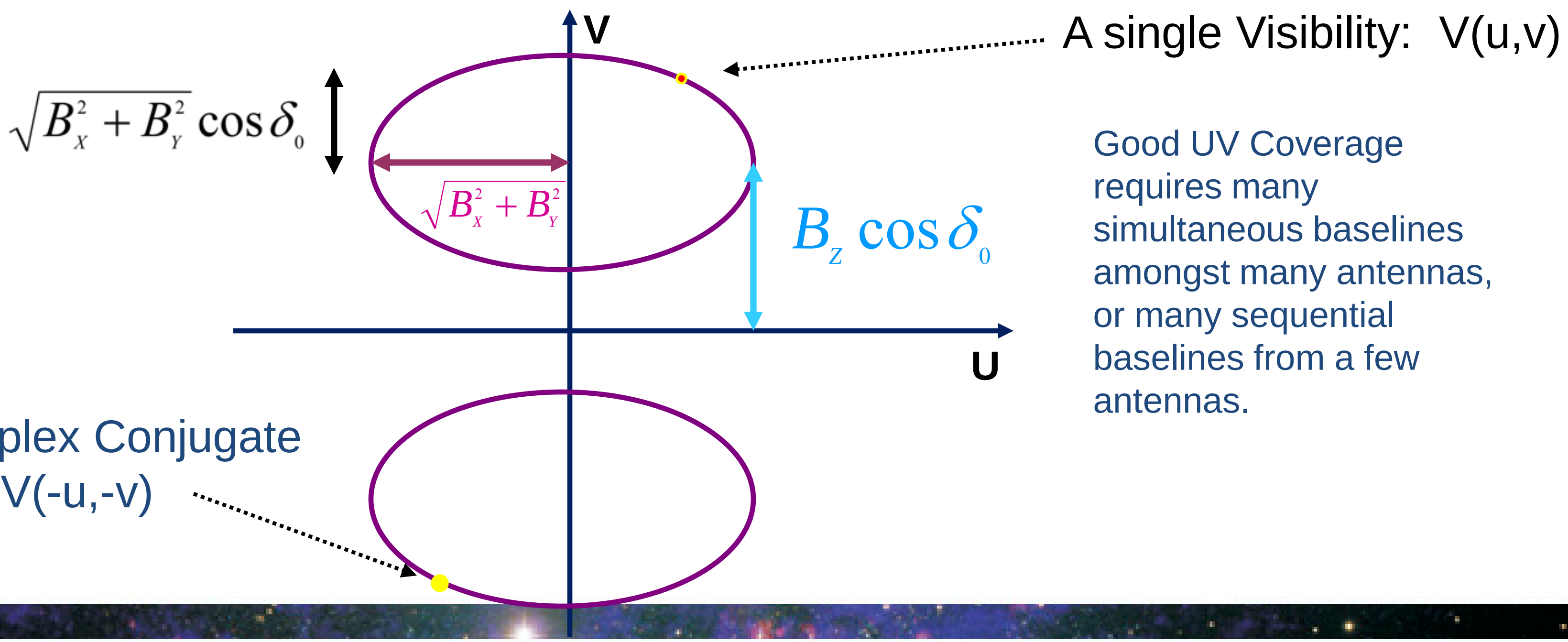
$$\nu_F = \frac{dw}{dt} = -\omega_E u \cos\delta_0$$

Baseline locus – the general case

- Each baseline, over 24 hours, traces out an ellipse in the (u,v) plane:

$$u^2 + \left(\frac{v - B_Z \cos\delta_0}{\sin\delta_0} \right)^2$$

- Because sky brightness is real, each observation provides us a second point, where: V(-u,-v) = V*(u,v)
- E-W baselines (Bx = Bz = 0) have no 'v' offset in the ellipses.



Good UV Coverage requires many simultaneous baselines amongst many antennas, or many sequential baselines from a few antennas.

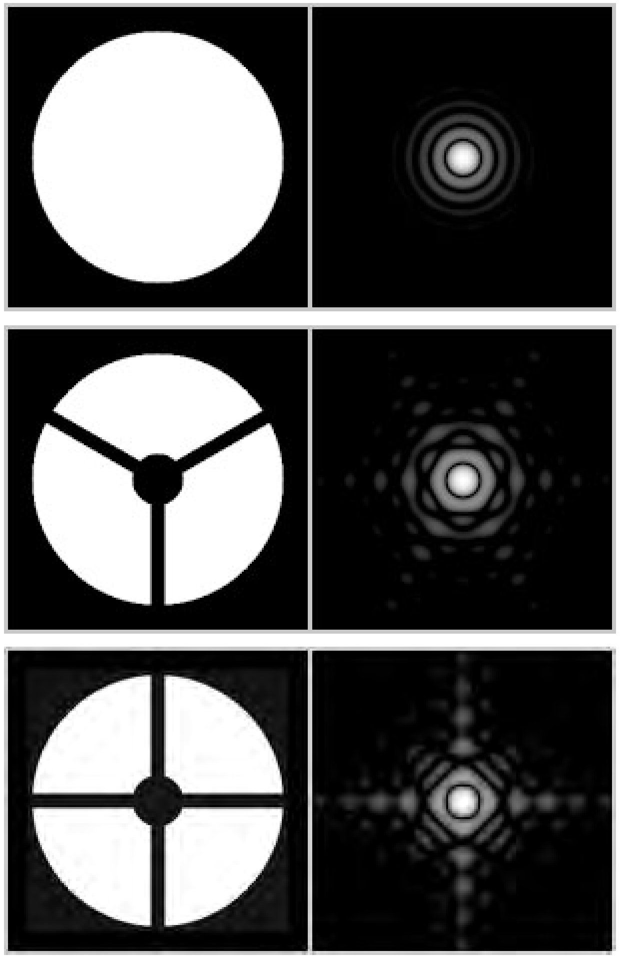
NRAO

Incomplete aperture filling

Example from optical telescopes

- A perfectly filled circular aperture
  The response to a point source (delta function) is an airy disk
- A partially blocked circular aperture
  The response to a point source (delta function) is a more complex pattern

In principle, the point spread function (PSF) can be calculated upfront, and removed from the image (deconvolution)

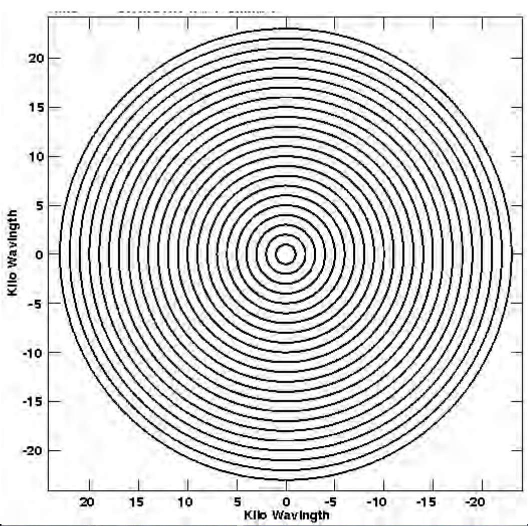


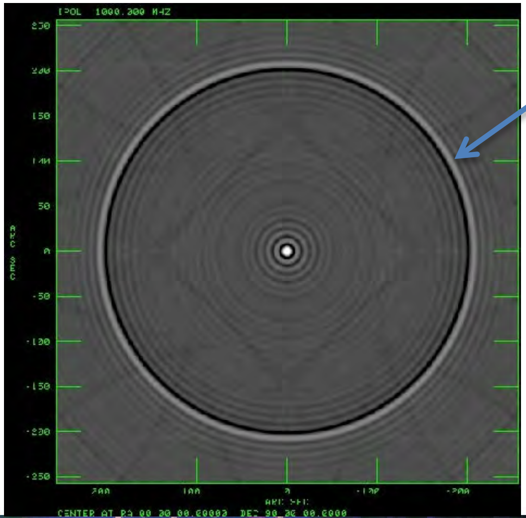
E-W array coverage and beams

- The simplest case is for E-W arrays, which give coplanar coverage.
- Then, Bx = Bz = 0
- Consider a 'minimum redundancy array', with eight antennas located at

  0, 1, 2, 11, 15, 18, 21 and 23 km along an E-W arm.

- Of the 28 simultaneous spacings, 23 are of a unique separation.
- The U-V coverage (over 12 hours) at δ = 90, and the synthesized beam are shown below, for a wavelength of 1m.

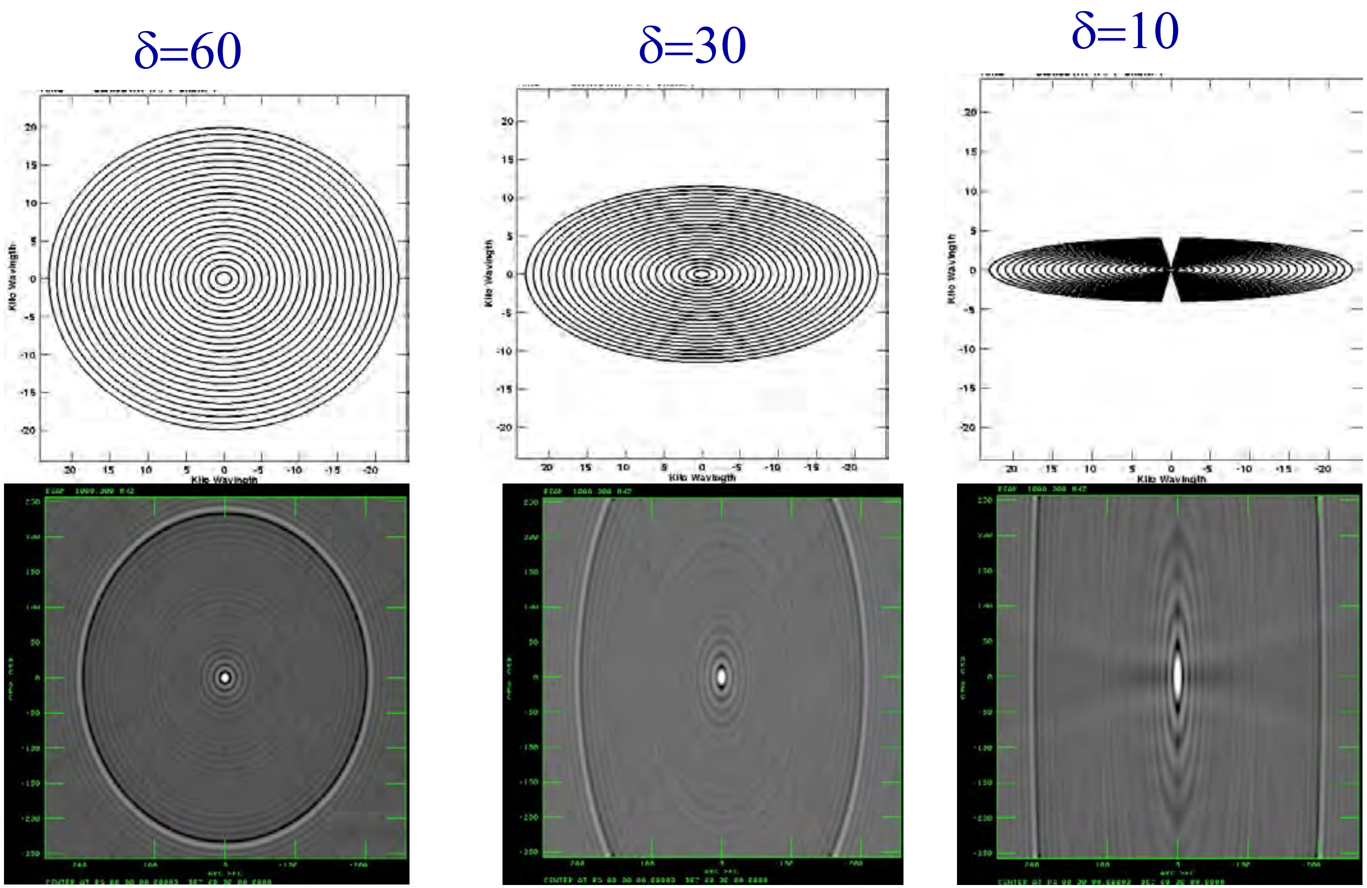




grating lobe

E-W arrays and low-DEC sources.

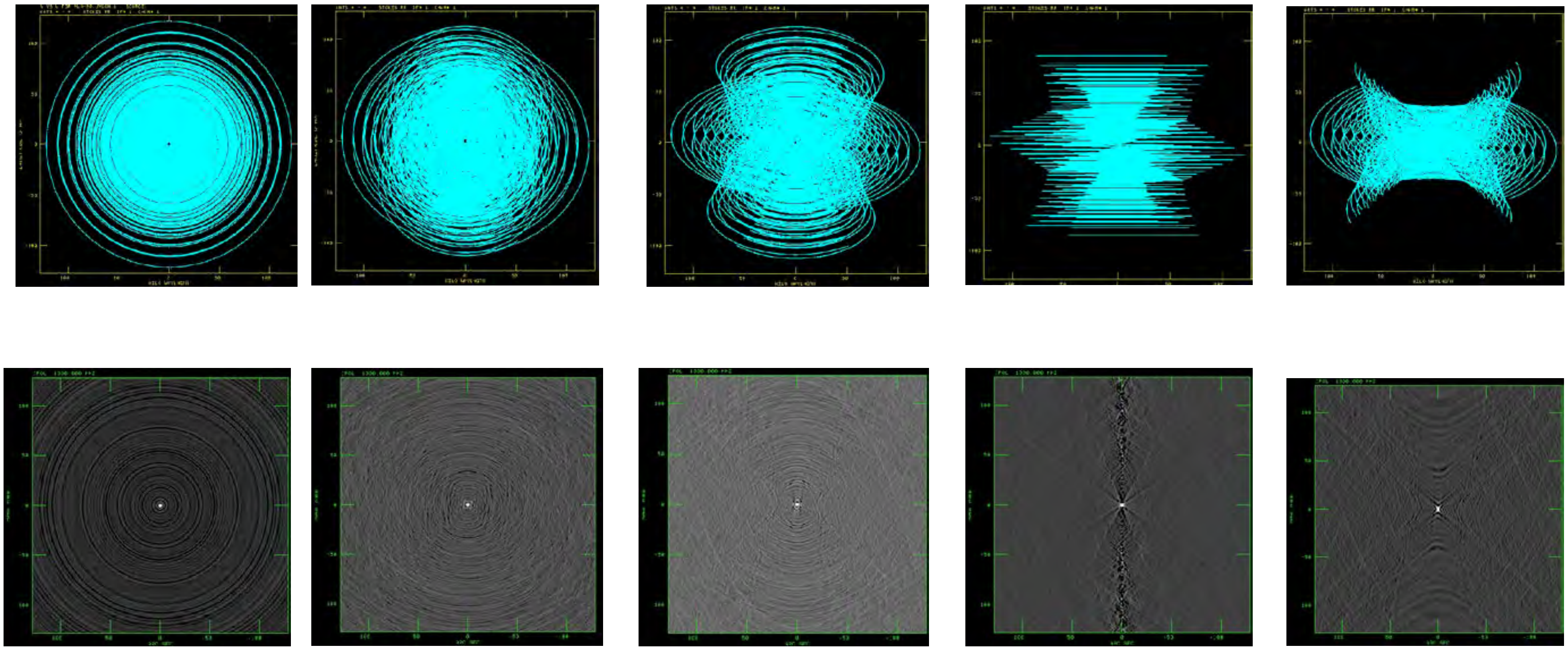
- But the trouble with E-W arrays is that they are not suited for low-declination observing.
- At δ=0, coverage degenerates to a line.



Getting good coverage near $\delta = 0$

- The only means of getting good 2-d angular resolution at all declinations is to build an array with N-S spacings.
- Many more antennas are needed to provide good coverage for such geometries.
- The VLA was designed to do this, using 9 antennas on each of three equi-angular arms.
- Built in the 1970s, commissioned in 1980, the VLA vastly improved radio synthesis imaging at all declinations.
- Each of the 351 spacings traces an elliptical locus on the (u,v) plane.
- Every baseline has some (N-S) component, so none of the ellipses is centered on the origin.

VLA UV-coverage and dirty beams



δ=90 δ=60 δ=30 δ=0 δ=-30

- Good coverage at all declinations, but troubles near δ=0 remain.

Non-coplanar interferometers

- What if the interferometer does not measure the coherence function on a plane, but rather does it through a volume? In this case, we adopt a different coordinate system. First we write out the full expression:

$$V_\nu(\mathrm{u},\mathrm{v},\mathrm{w}) = \iint \frac{I_\nu(l,m)}{\sqrt{1-l^2-m^2}} e^{-2i\pi(\mathrm{u}l+\mathrm{v}m+\mathrm{w}n)} dl dm$$

(Note that this is not a 3-D Fourier Transform).

- We orient the w-axis of the coordinate system to point to the region of interest. The u-axis point east, and the v-axis to the north celestial pole.
- We introduce phase tracking, so the fringes are 'stopped' for the direction l=m=0. This means we adjust the phases by $e^{2i\pi w}$
- With $n^2 = 1 - l^2 - m^2$ we get:

$$V_\nu(u,v,w) = \iint \frac{I_\nu(l,m)}{\sqrt{1-l^2-m^2}} e^{-2i\pi[ul+vm+w(\sqrt{1-l^2-m^2}-1)]} dl dm$$

From 3-d back to 2-d

- The expression is not a proper Fourier transform.
- We can get a 2-d FT if the third term in the phase factor is sufficient small.

- The third term in the phase can be neglected if it is much less than unity:

$$w\left[1-\sqrt{1-l^2-m^2}\,\right]=w(1-\cos\theta)\sim w\theta^2/2<<1$$

- This condition holds for small angles (small images):
  (angles in radians)

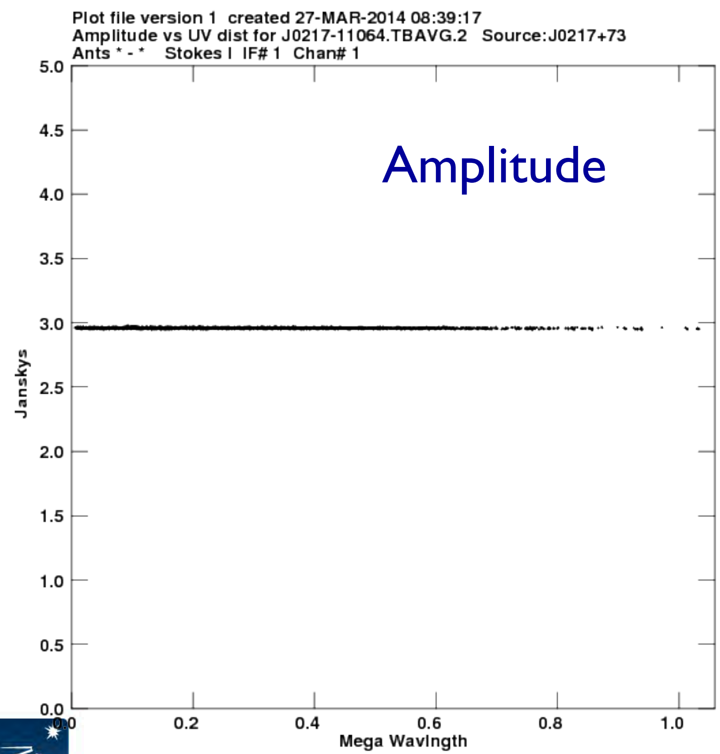
$$\theta_{\max}<\sqrt{\frac{1}{2\mathrm{w}}}\sim\sqrt{\frac{\lambda}{\mathrm{B}}}\sim\sqrt{\theta_{syn}}$$

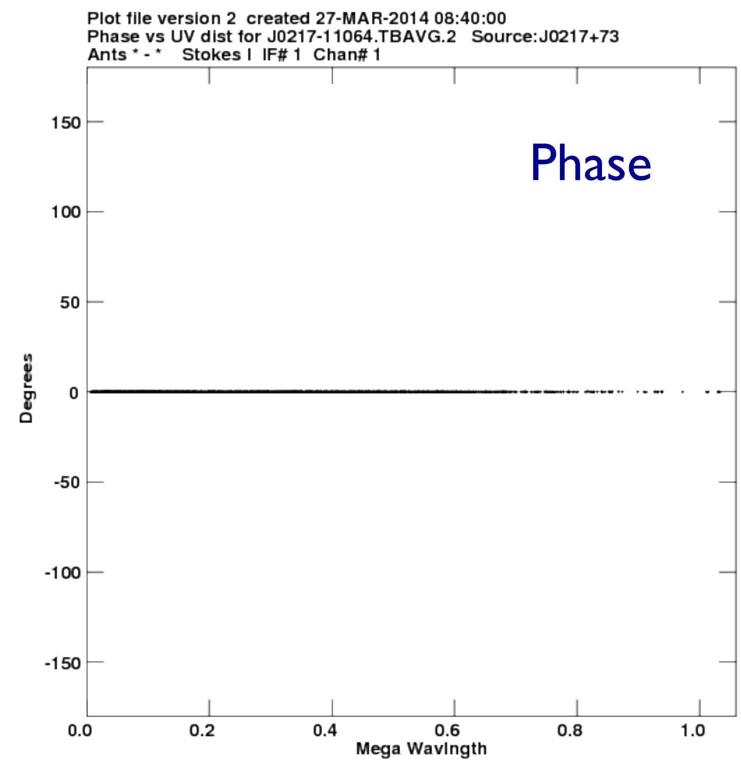
- If this condition is met, then the relation between the Intensity and the Visibility again becomes a 2-dimensional Fourier transform:

$$V_\nu'(u,v)=\iint I_\nu(l,m)e^{-2i\pi(ul+vm)}\,dldm$$

Examples of visibilities

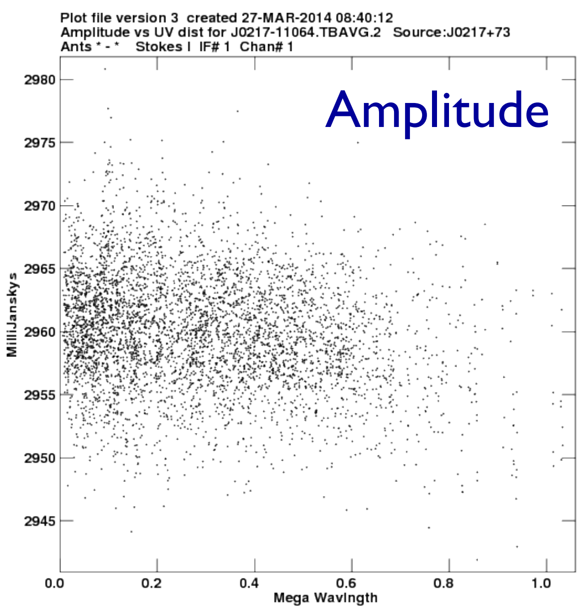
- Suppose we observe an unknown object.
- What is its shape, flux density and position?

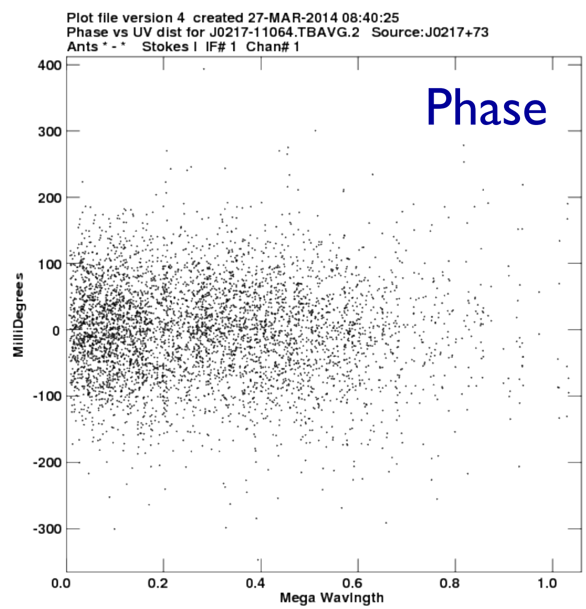


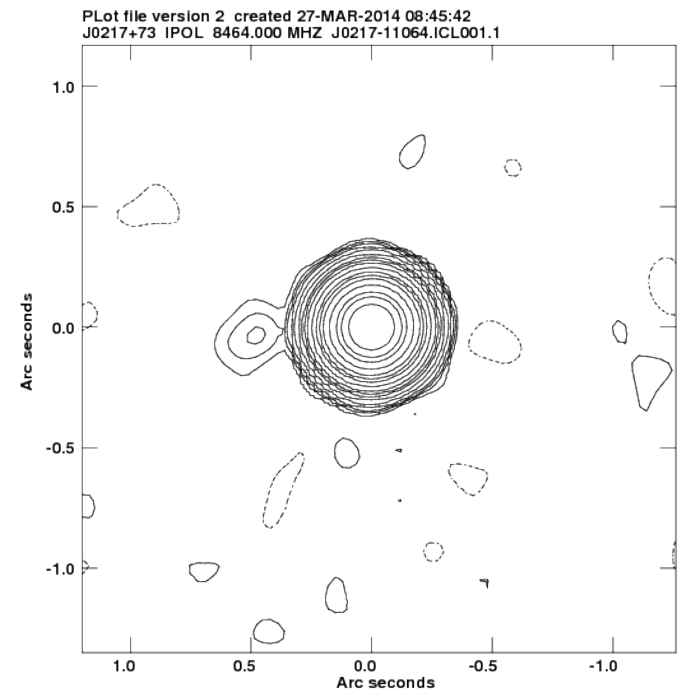


And the image

- The source is unresolved, but with a tiny background object.
- Dynamic range: 50,000:1
- Good for astrometry reference



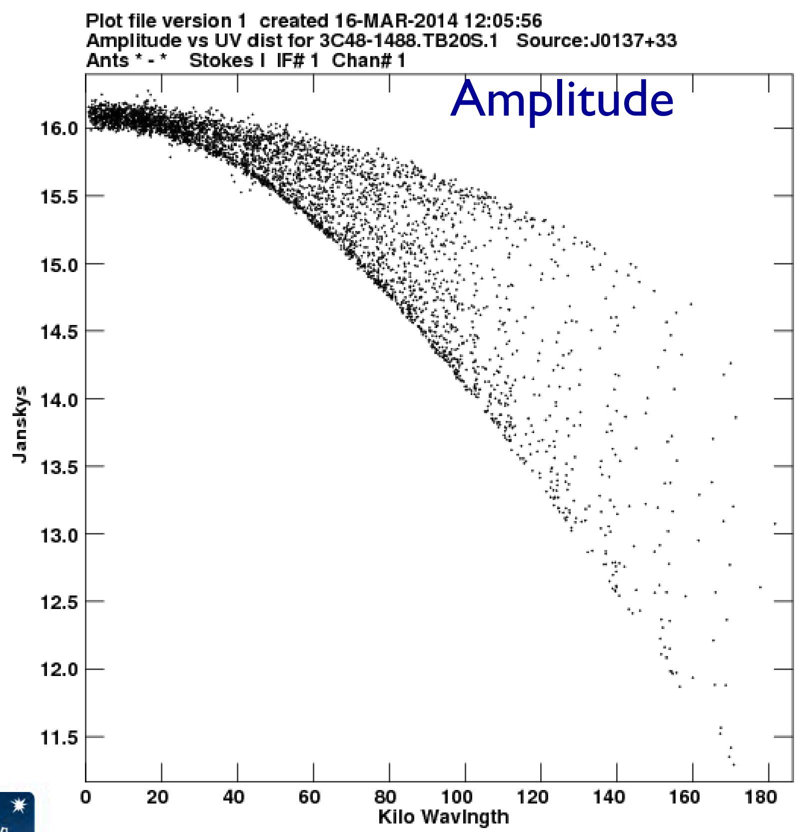


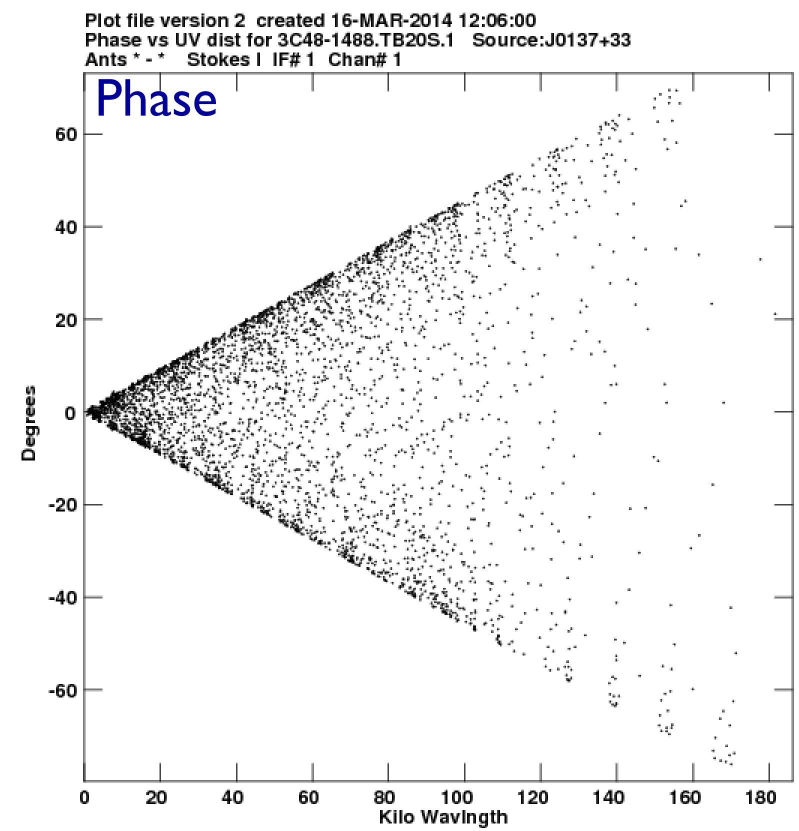


NRAO

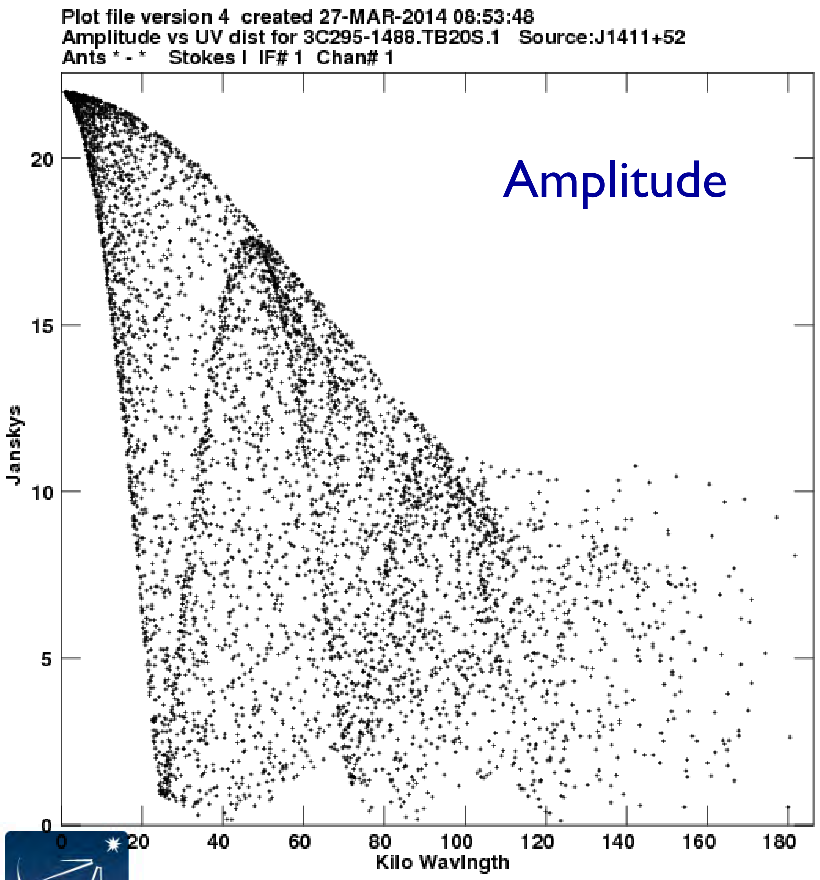
Examples of visibilities – 3C48, a slightly resolved object.

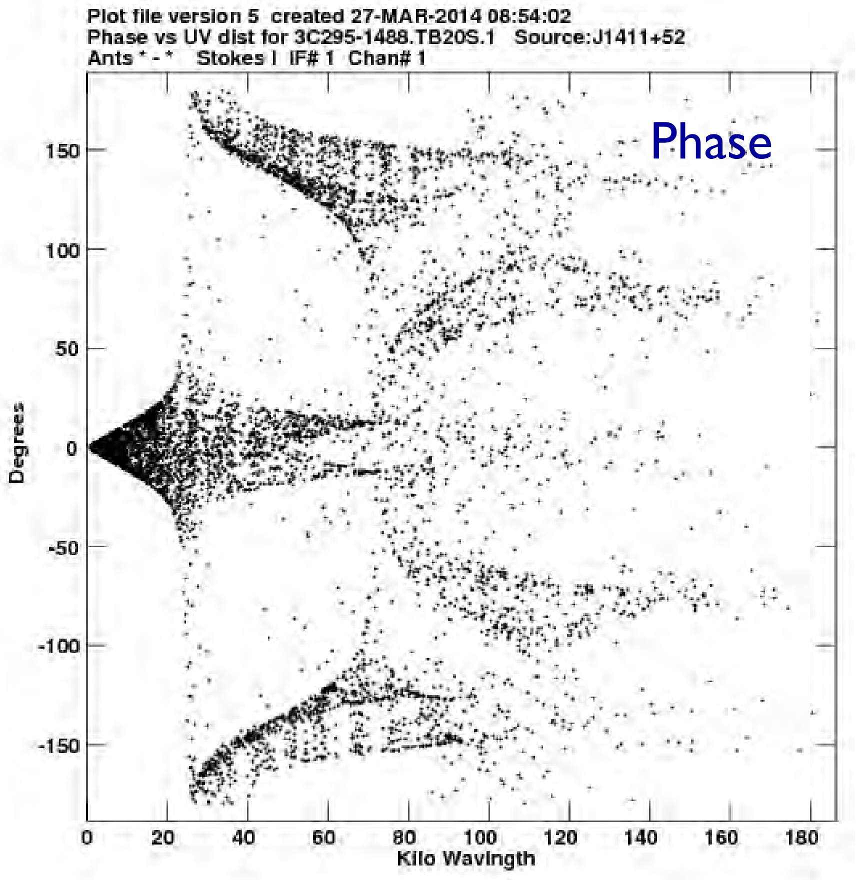
- Bright source – good for flux reference.





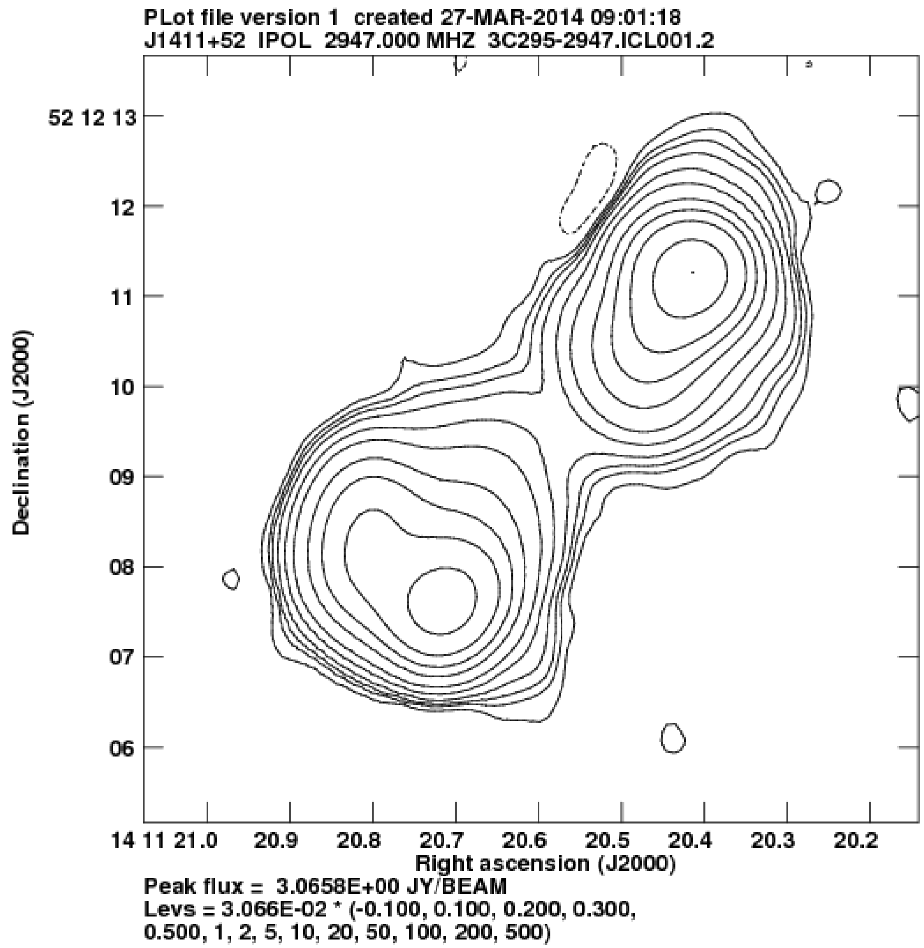
Examples of visibilities – 3C295, a well-resolved object





3C295 image

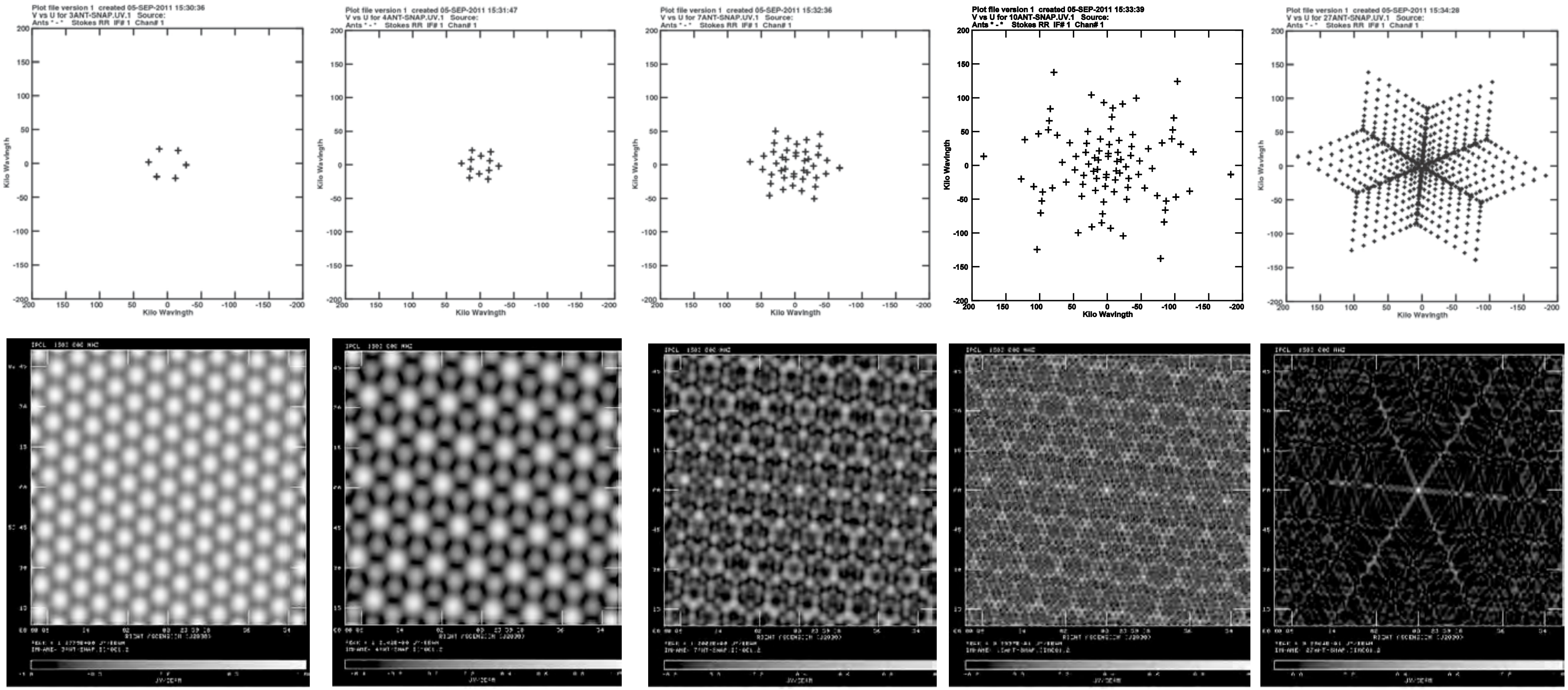
- A 5-arcsecond double.



Improve UV-coverage – add antennas

- A single observation of a point source.

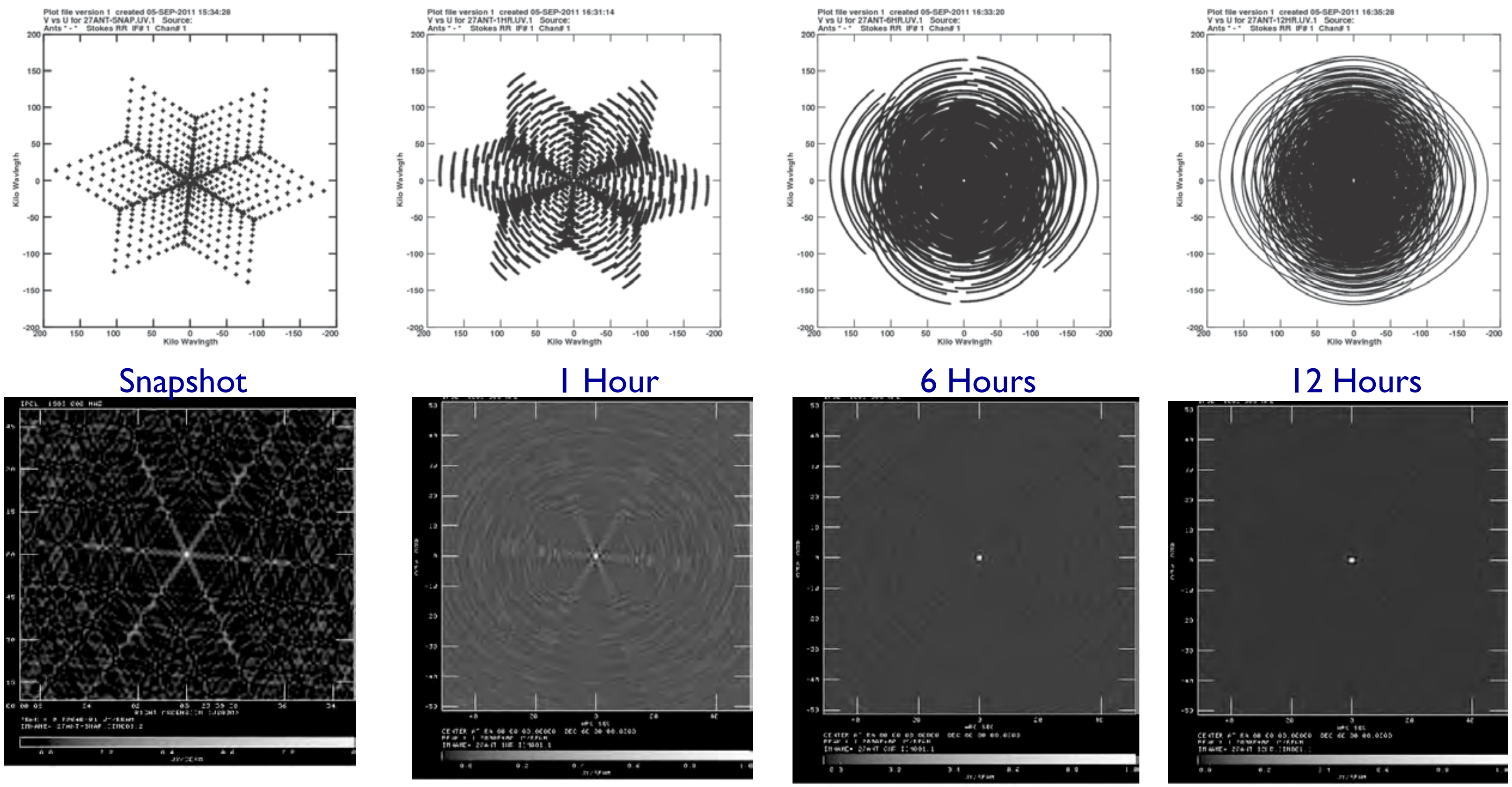
3 ant, 3baselines	4, 6	7, 21	10, 45	27, 351	



100%	100%	93%	82%	41%	= Pk. sidelobe
71%	41%	15%	11%	4%	= rms.
-280%	-104%	-32%	-23%	-10%	= Pk. neg.

Improve UV-coverage – observe over time

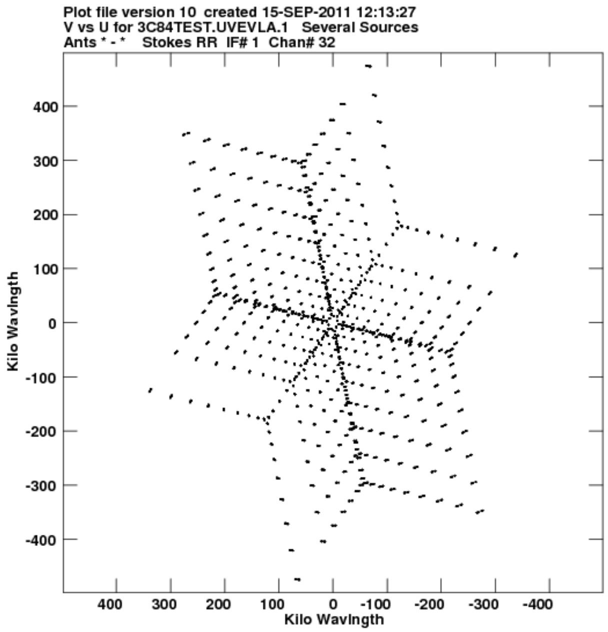
- UV coverage and Images for a point-source for 0,1, 6, and 12 hours' observing, with the 27-antenna VLA, at δ = 60.

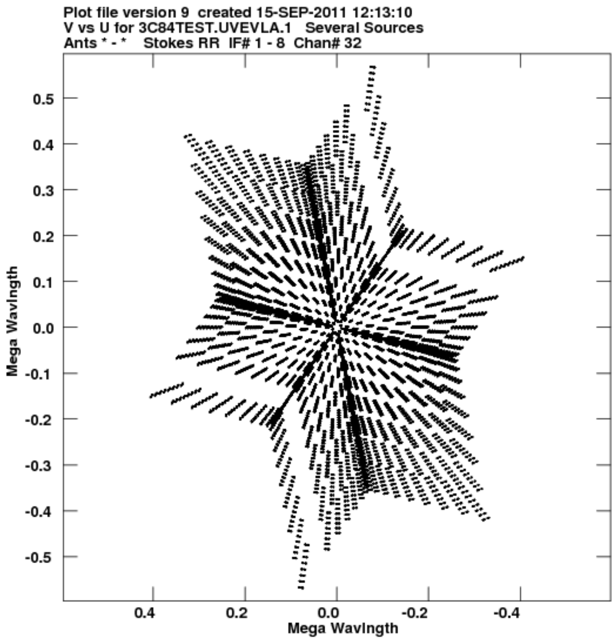


	Snapshot	1 Hour	6 Hours	12 Hours
Pk Sidelobe	41%	14%	2%	1%
rms	11%	4%	0.5%	0.3%
Max. Neg.	-10%	-10%	-5%	-4%

Improve UV-coverage – add frequencies

- 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.

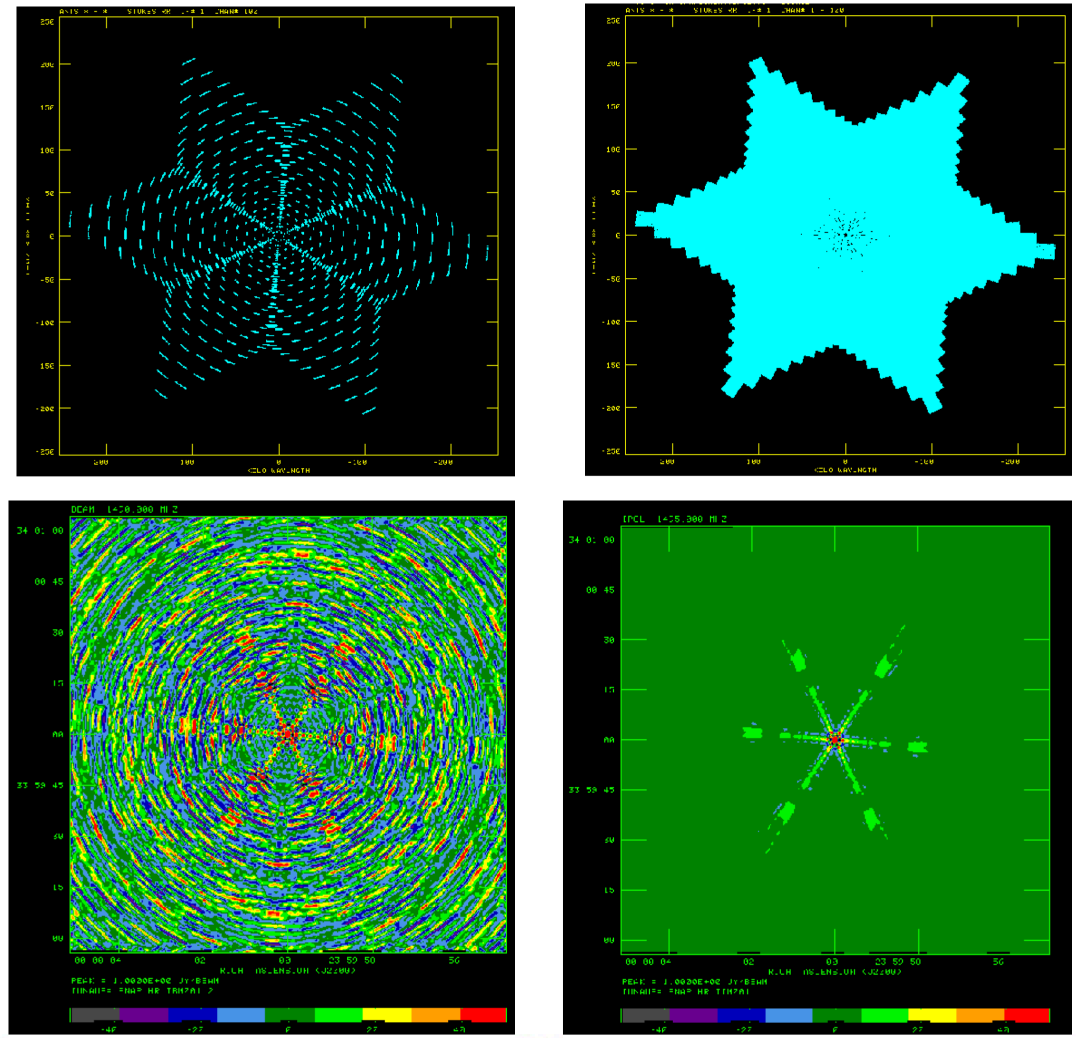




Imaging using wide bandwidth

- 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%



UV-coverage and imaging fidelity

- Although the VLA represented a huge advance over what came before, its UV coverage (and imaging fidelity) is far from optimal.
- The VLA's Y-shaped design was dictated by its 220 ton antennas, and the need to move them. Railway tracks were the only answer.
- The high density of samplings along the arms (the 6-armed star in snapshot coverage) results in 'rays' in the images due to small errors.

- A better design is to 'randomize' the location of antennas within the span of the array, to better distribute the errors.
- Future major arrays will utilize smaller, lighter elements which should not be positioned with any regularity.

Problems with UV-plane sampling

- You will note that the UV sampling by the interferometer is very non-uniform.
- There are large holes, and typically 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 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.

UV-plane weighting

- For imaging through FFT, the visibilities are gridded onto a regular grid. We can "weigh" the gridded values according to our scientific needs.
  - **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 versus 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 weighting'.
  - 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'.

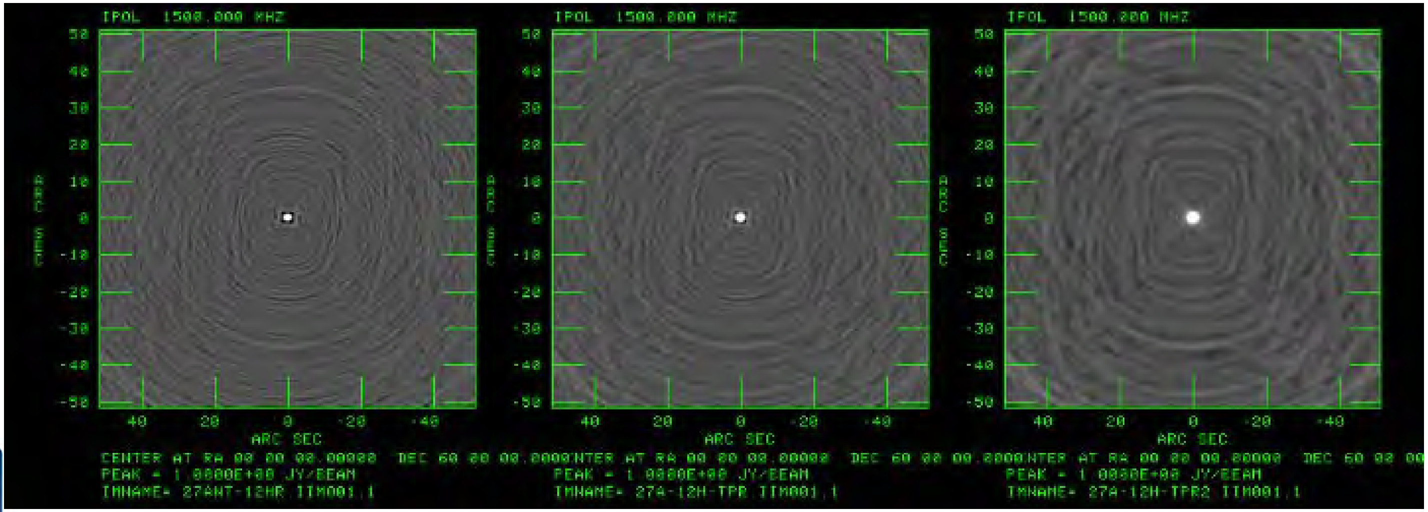
‘Natural’ versus ‘uniform’ weighting

- 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 UV-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 kλ	30% at 80 kλ
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%



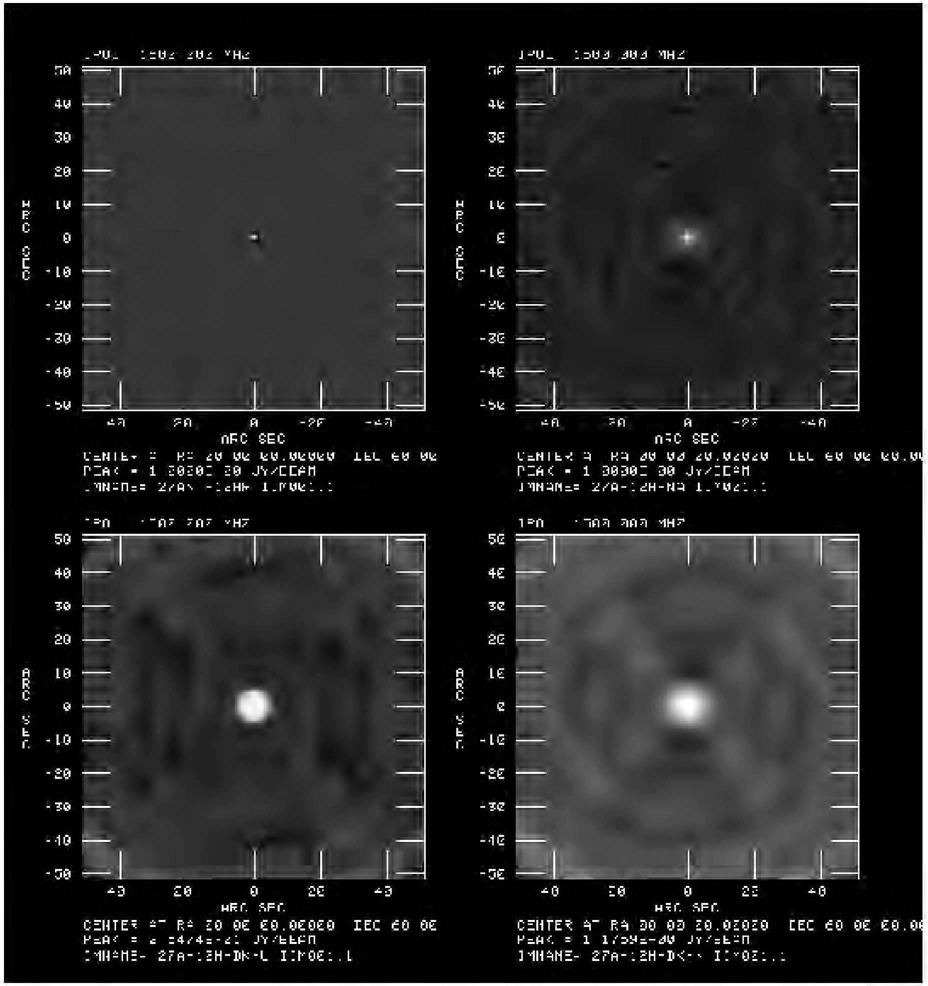
Examples of weighting

Four 'dirty' images:

- Upper row: point source.
- Lower pair: 20" disk.

- The effect of 'natural' weighting is to 'soften' and broaden the response.
- The peak brightness of the disk is much higher, due to the larger beam.
- The peak brightness of the unresolved source remains the same.

'Uniform' Wgt.   'Natural' Wgt.



NRAO

Part 3: Imaging

- UV-coverage and dirty image
- Deconvolution

The formal relations

- We learned, from parts 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 $V_\nu(u,v)$ by a Fourier transform:

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

- The formal inversion of this equation is:

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

- This is very simple (and very pretty), but cannot directly be used, since the functions concerned are analytic – known for all angles and distances.
- Thus, there is missing information, and this means we cannot guarantee a correct determination of $I_\nu(l,m)$.

The sampling function and the dirty map

- We define the 'dirty image' by:

$$\mathrm{I_D}(l,m) = \iint \mathrm{S(u,v)V_\nu(u,v)}e^{i2\pi(ul+vm)}\,\mathrm{d}l\,\mathrm{d}m$$

where S(u,v) is the 'sampling function', which describes the actual locations in the UV-plane where samples of the visibility are taken.

- We can write this as: $$S(\mathrm{u,v}) = \sum_{k=1}^{M} \delta(\mathrm{u-u_k, v-v_k})$$
- Then, using a property of the Dirac delta function, we find

$$\mathrm{I_D}(l,m) = \frac{1}{M}\sum_{k=1}^{M} \mathrm{V}(\mathrm{u}_k,\mathrm{v}_k)e^{i2\pi(\mathrm{u}_k l+\mathrm{v}_k m)}.$$

- This function is often called the 'principle solution'.

Deconvolution

- 'Deconvolution' is the process of removing artifacts caused by the irregular and/or incomplete sampling in the UV-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 FT 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 holes in the UV-plane for which we have no measurements.
- The true sky has characteristics which we can use to assist us in 'guessing' the values in the holes.
- Deconvolution is effectively the process of estimating the visibilities for the locations where we have no observations.

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 in case of total intensity mapping.

Major forms of deconvolution

- There are two major classes of deconvolution algorithms:

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

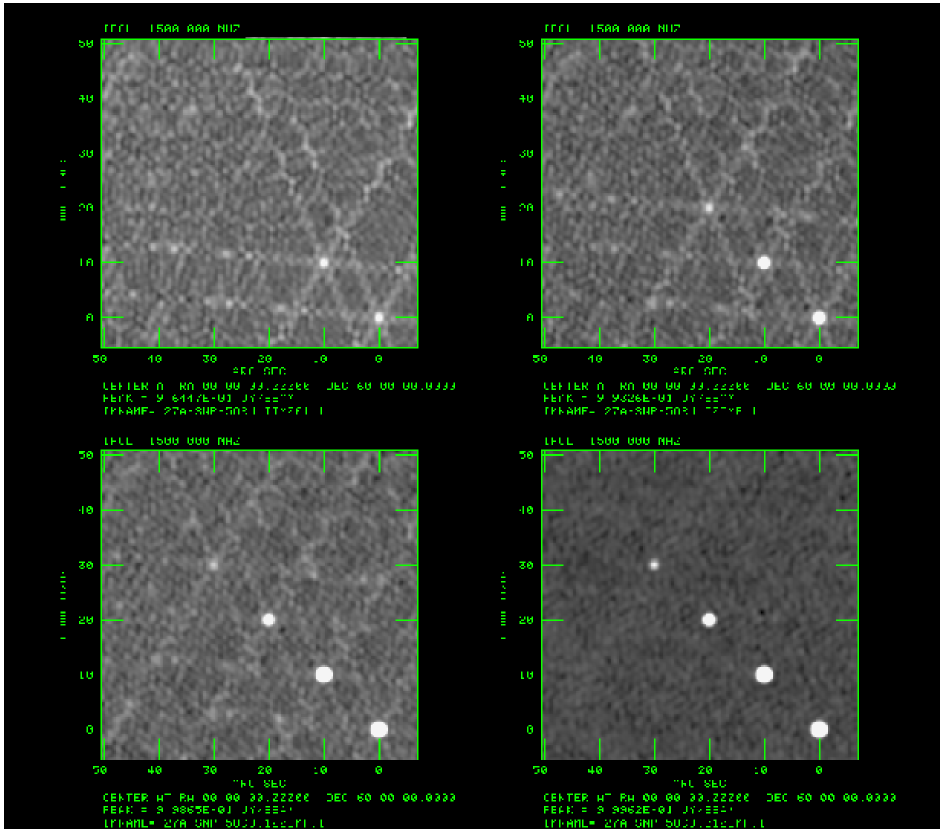
- The 'CLEAN' algorithm was developed by Hogbom in 1974. Various improvements (mostly for speed) have been developed since then.
- In essence, CLEAN assumes the emission to be a set of unresolved or partially resolved, well-separated objects.
- CLEAN remains the most commonly used procedure. Despite its simplicity, CLEAN works remarkably well.
- The procedure is:
  1. Locate the highest point of emission in the image, and subtract from the entire image the dirty beam, scaled by some 'loop gain' $\gamma<1$. Store the location and peak brightness removed.
  2. After several repetitions of step 1, subtract the FT of the point sources from the visibilities. Update the 'residual' image.
  3. Repeat steps 1 and 2 until some level of noise is reached.
  4. Rebuild the image, using the list of locations and peaks, adding a 'clean beam' back into the residual image in place of the psf.

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 = 0.2.
- Simulated snapshot with VLA, with suitable noise added.

- Top left: dirty Image
  - σ = 53 mJy/beam
- Top right: after 50 comp.
  - σ = 5.7 mJy/beam
- Bottom left: after 150 comp.
  - σ = 0.74 mJy/beam
- Bottom right: done (310 comp).
  - σ = 0.19 mJy/beam

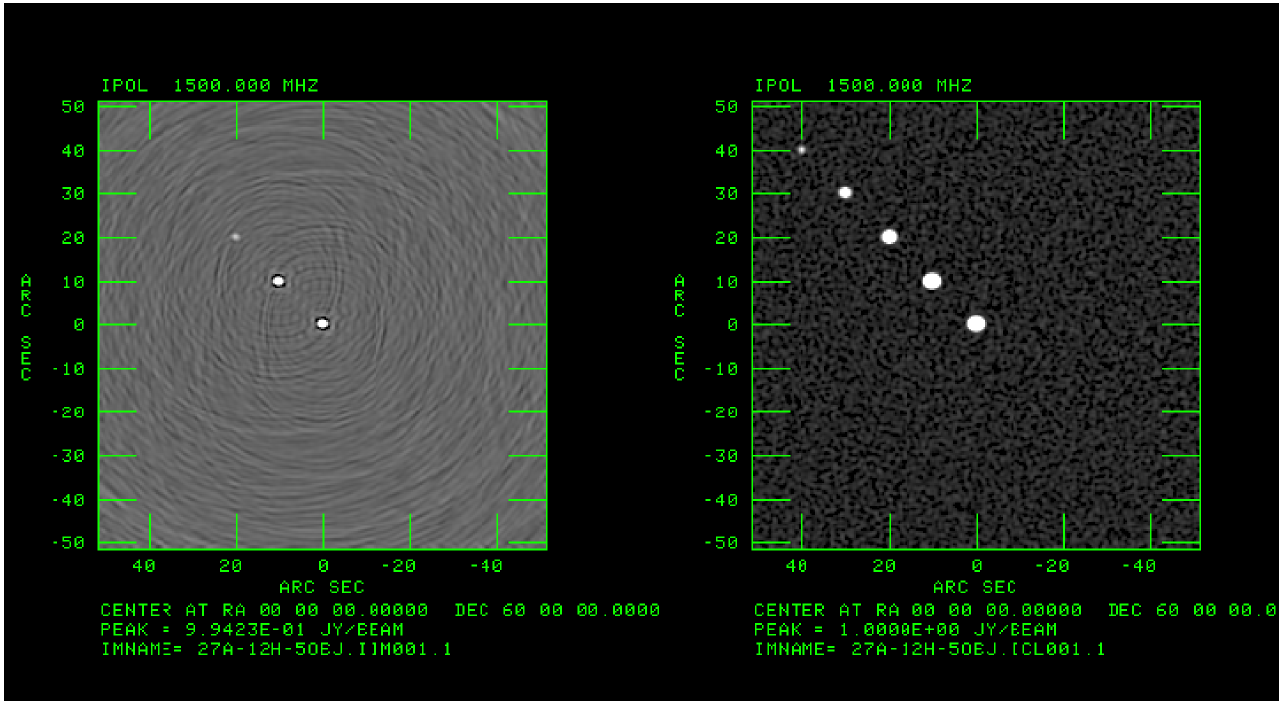
- Note: the weakest source is not visible – below the noise.



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/beam
- Clean image rms: 0.012 mJy/beam



CLEAN – some remarks

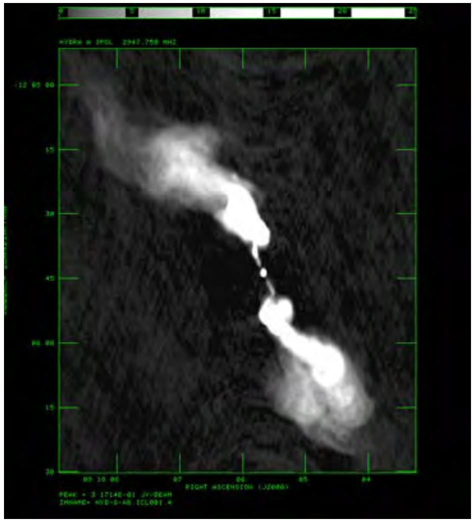
- There are a number of variants of 'CLEAN': the 'Clark' and 'Cotton-Schwab' algorithms are variants which permit much faster searching and subtracting. Multi-scale CLEAN helps on extended objects.
- CLEAN works naturally best with compact emission.
- CLEAN often does a poor job on low-brightness sources – the typical signature being 'rippling' in the restored image. The procedure can diverge in these situations, especially when the UV-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 interations, and the location of boxes, to provide the 'best' image, is a matter of experience.

Multi-scale CLEAN

- Two 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' (MS) 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.

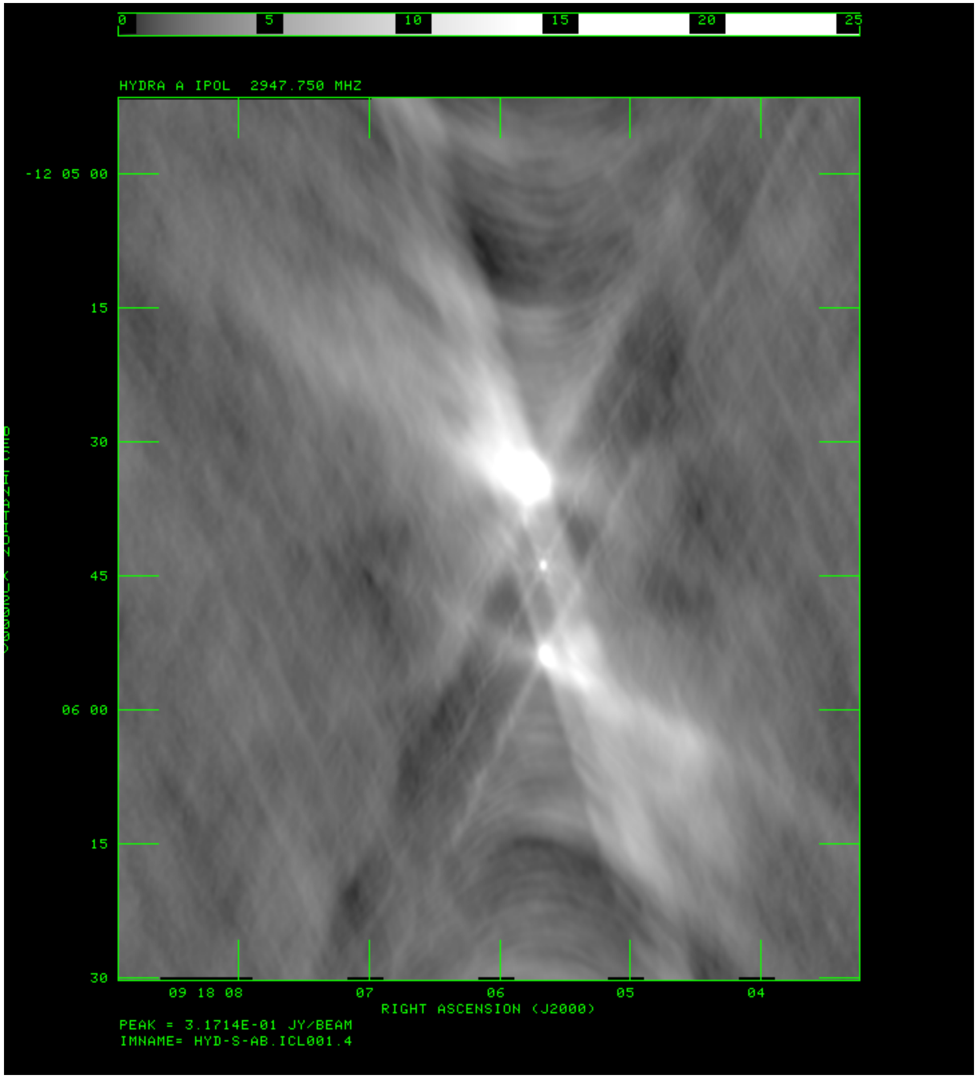
A comparison using real data

- 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 narrow-band (10 MHz BW) S-band (3 GHz) image, utilizing three of the four VLA 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.



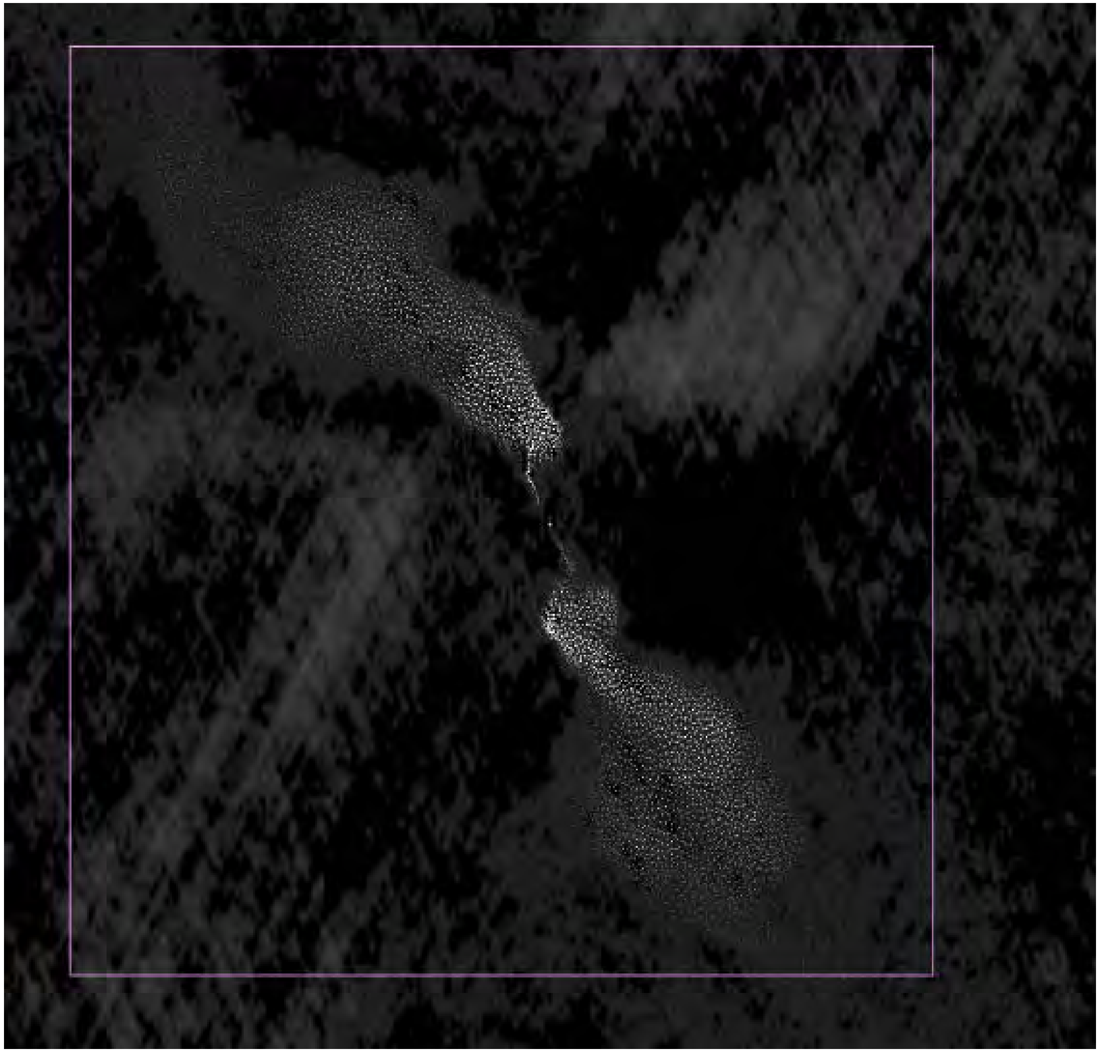
The ‘dirty’ image

- 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.



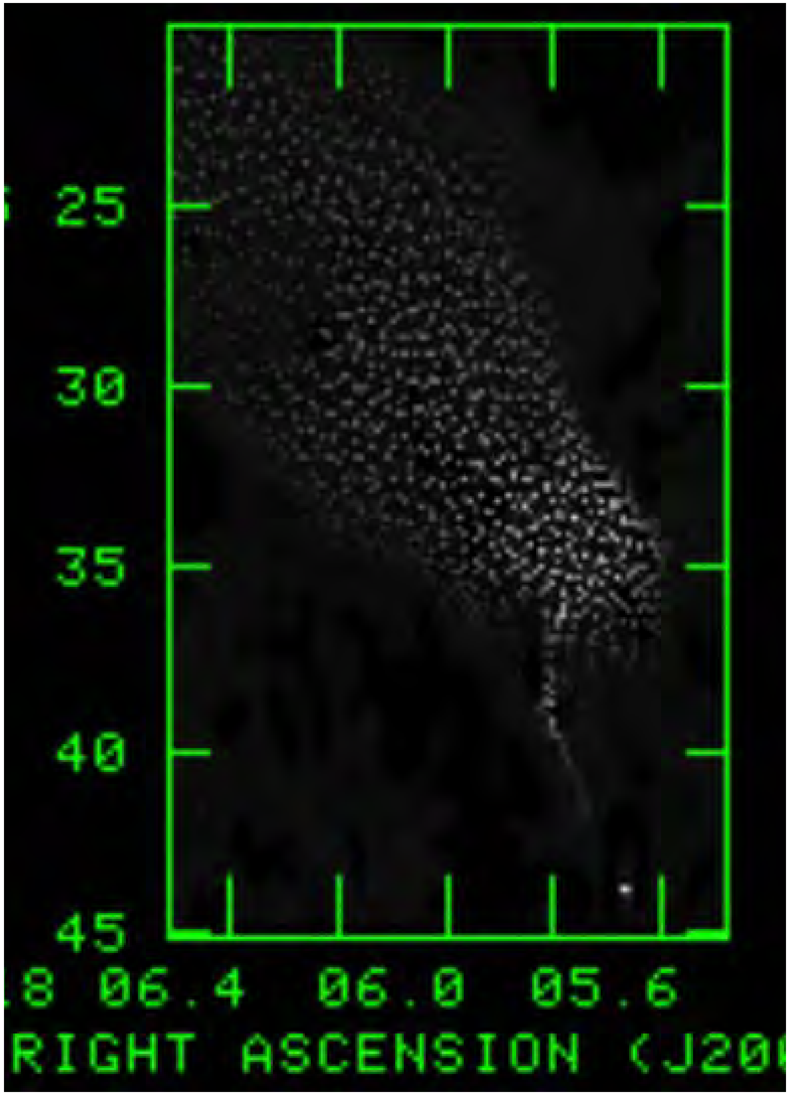
The raw clean solution

- Shown here are the 50,000 Clean components, with the residuals added.
- This mass of point components is actually a *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.



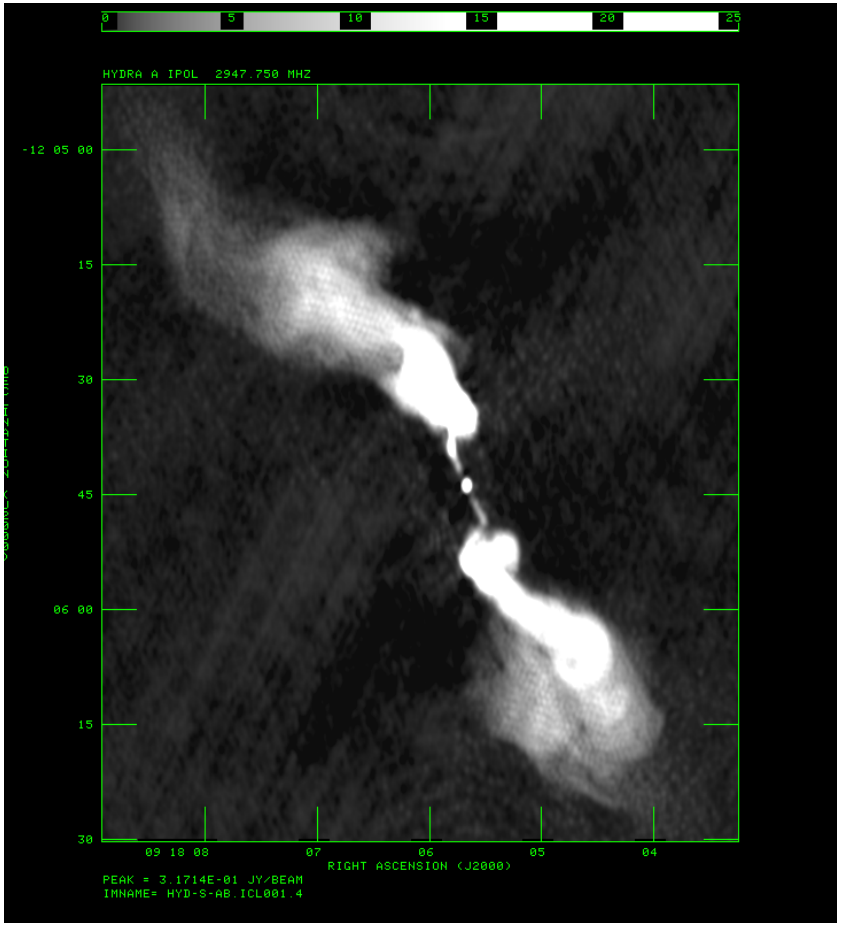
Zoom into the northern lobe

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

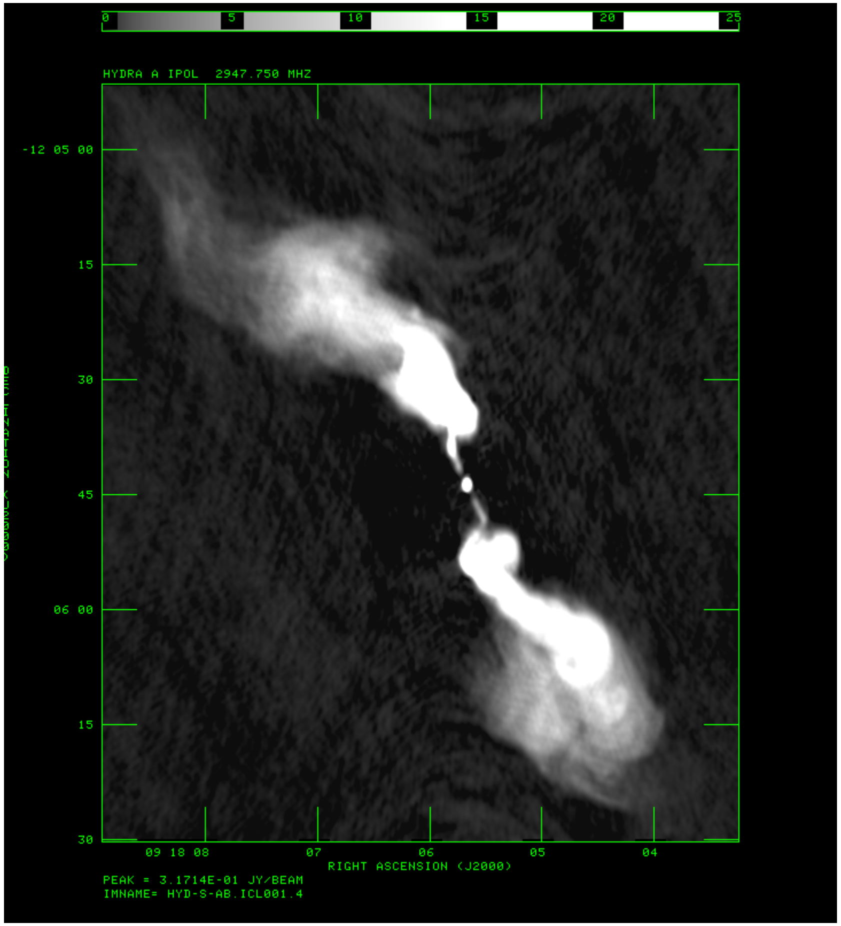


Standard CLEAN and Multi-Scale CLEAN

Standard CLEAN

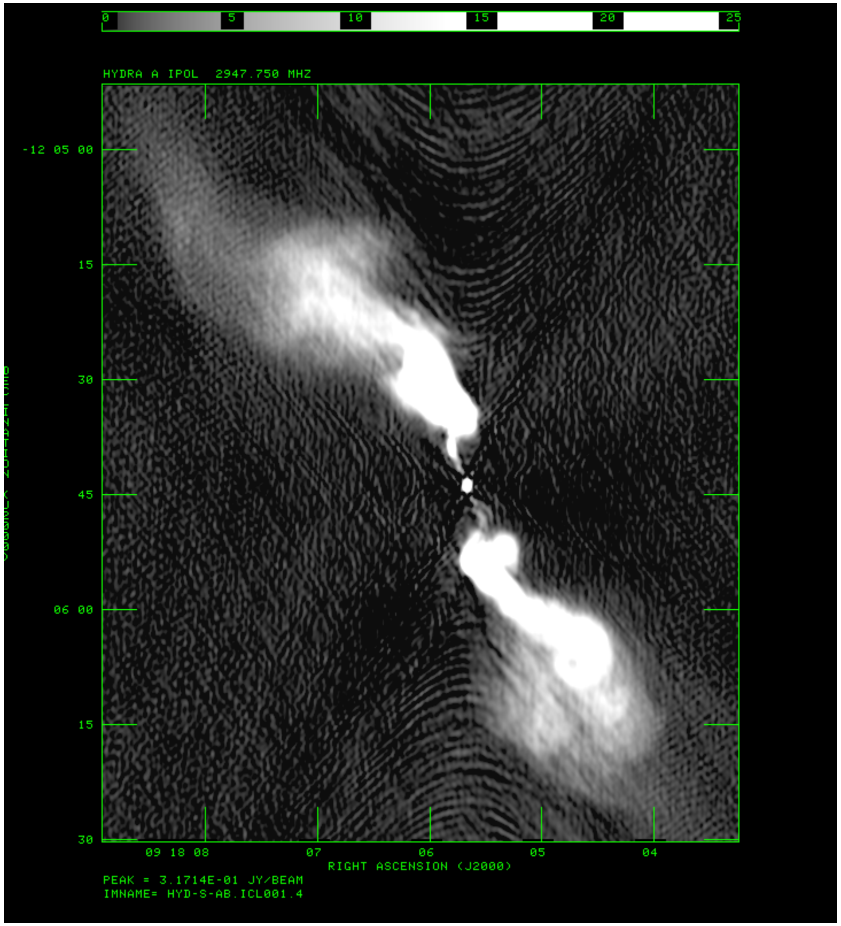


Multi-Scale CLEAN



Maximum Entropy Method (VTESS)

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

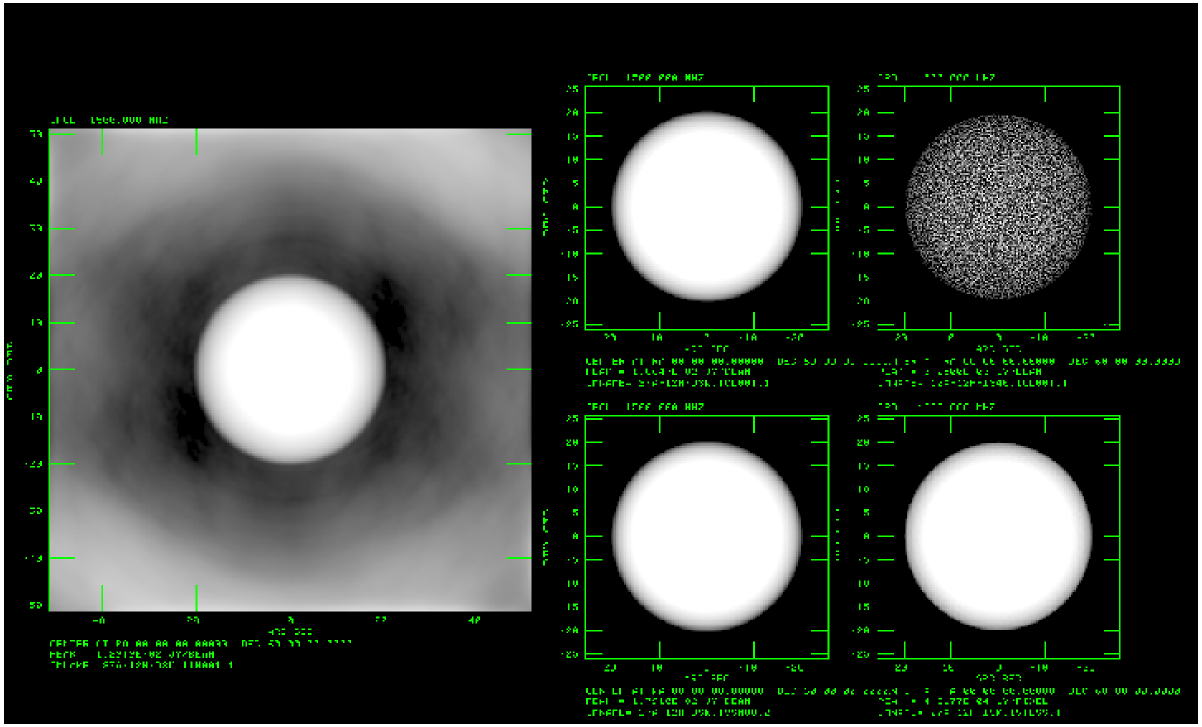


MEM versus CLEAN – big disk

- 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:

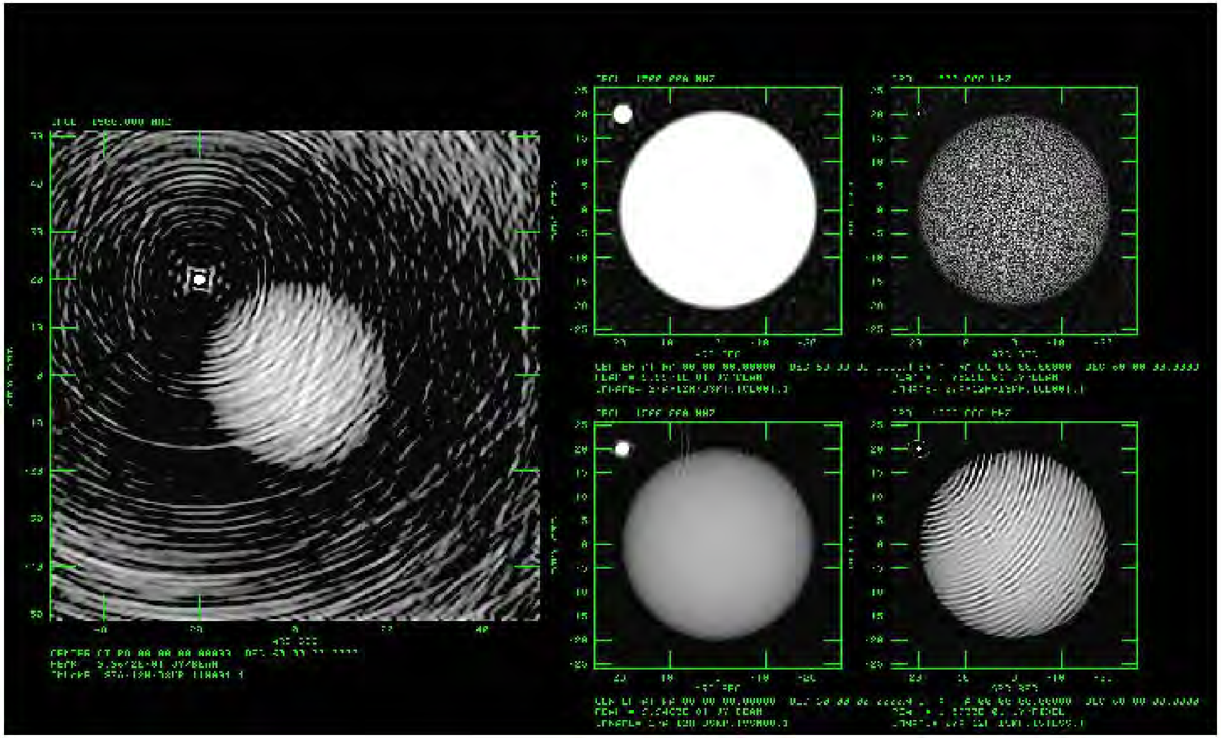
- Top Left: 100,000 component CLEAN, with restore.
- Top Right: Same CLEAN, but no clean beam restore.
- Bottom Left: 40-cycle VTESS, with gaussian smoothing.
- Bottom Right: Same VTESS, but no smoothing.
- Disk brightness: ~20 mJy/bm.



MEM 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, by adding a 1 Jy point source to the 10 Jy big disk.

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



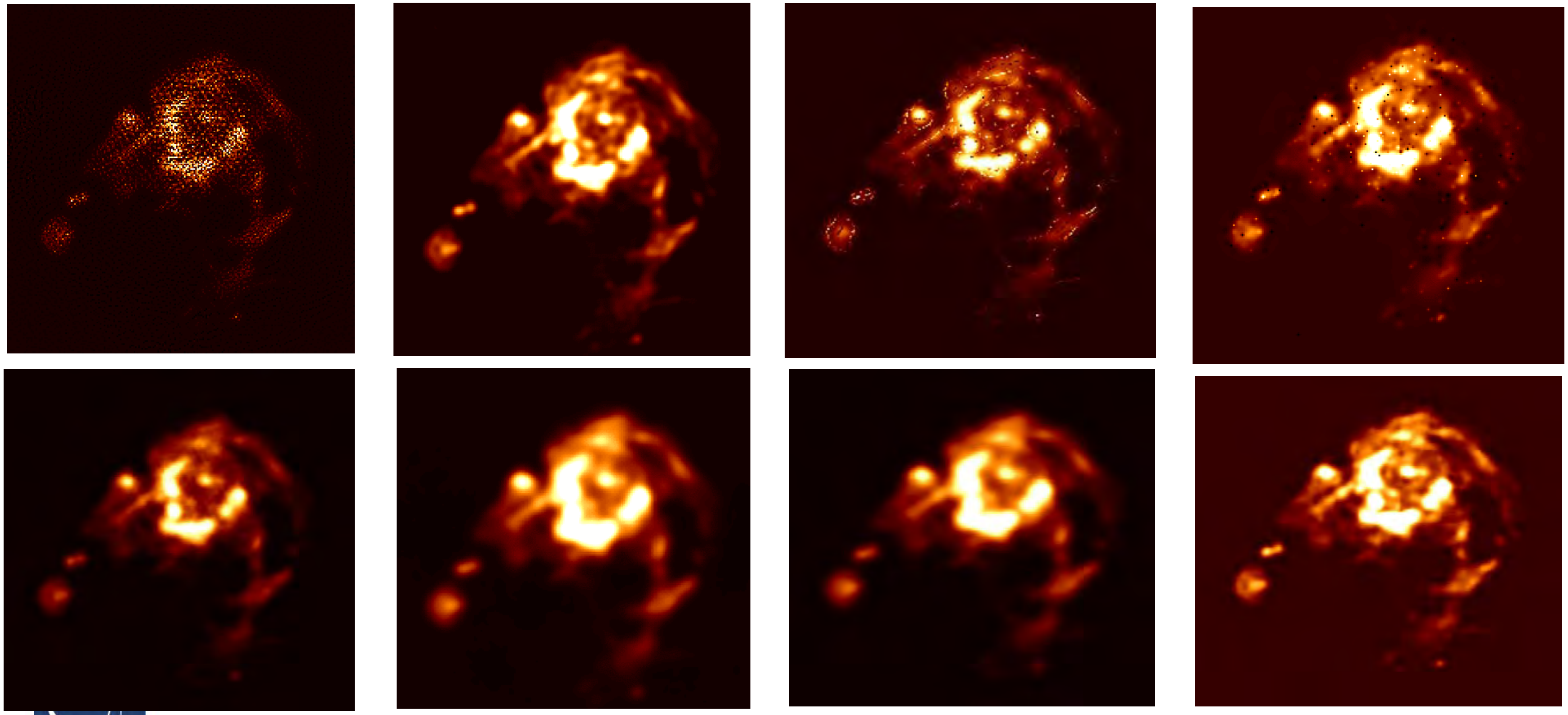
New image reconstruction efforts

- Deconvolution is the most critical, but uncertain, 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 the 'ASP' algorithm – similar to MS-CLEAN, but automatically adjusts the size of the best-fit gaussians during the subtraction process.

NRAO

Comparison of deconvolution algorithms

CLEAN	MEM	MS-CLEAN	ASP
Minimize L2 (assume sparsity in the image)	Minimize L2 subject to an entropy-based prior (e.g. smoothness)	Minimize L2 (assume a set of spatial scales)	Minimize L2 with TV-based subspace searches



I^m

I^{out}

Comparison of deconvolution algorithms

CLEAN

Minimize L2 (assume sparsity in the image)

MEM

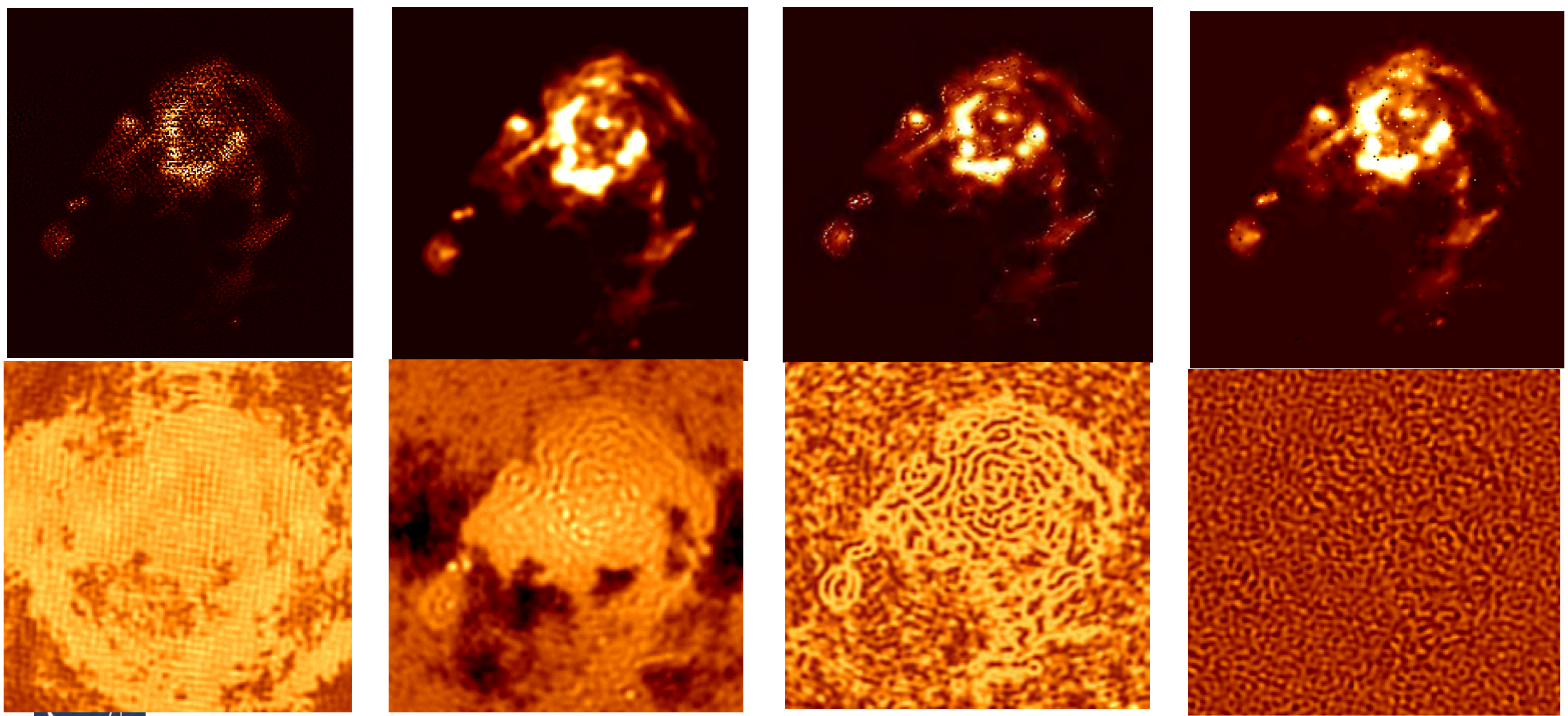
Minimize L2 subject to an entropy-based prior (e.g. smoothness)

MS-CLEAN

Minimize L2 (assume a set of spatial scales)

ASP

Minimize L2 with TV-based subspace searches



I^m

I^{res}

Next: Data Analysis with CASA

Possible lab discussion topics:

- Calibration / self-calibration
- Linear and circular feeds
- Excising bad data
- Wide-field imaging
- Wide-bandwidth imaging
- Bandwidth / time-averaging smearing