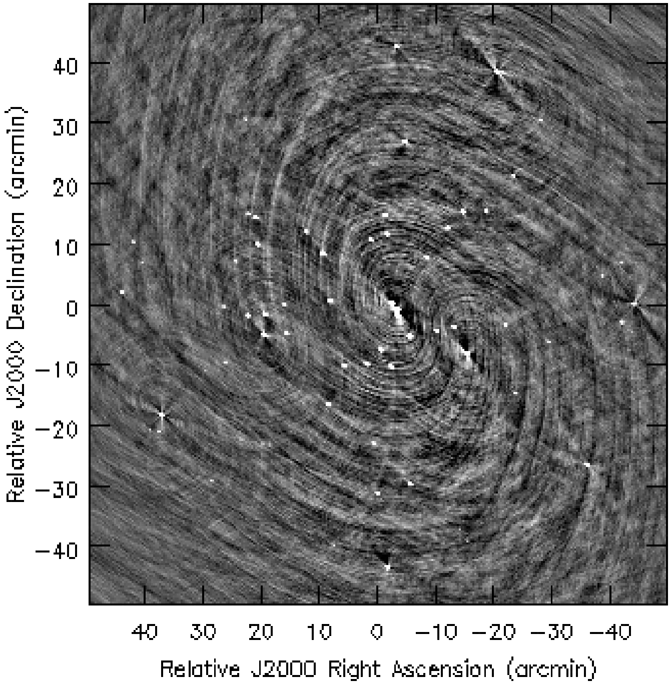
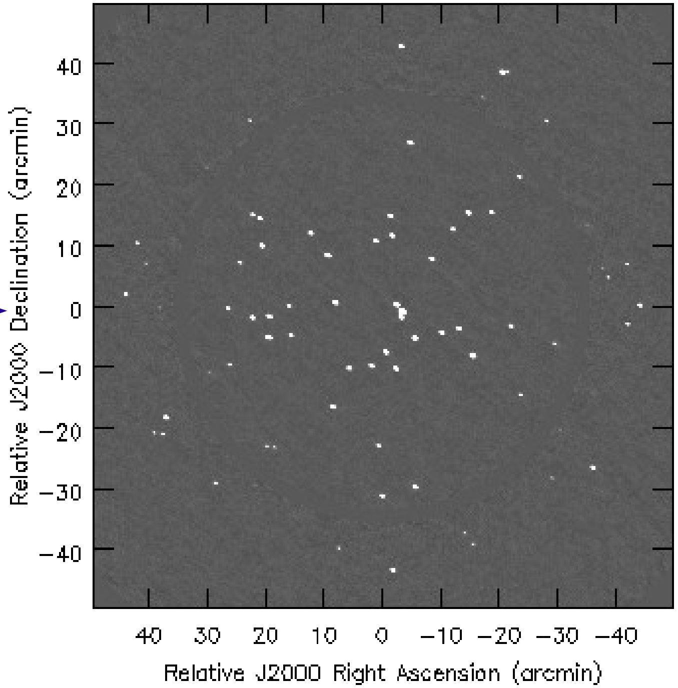
Interferometric Imaging and Direction Dependent Effects



RMS ~15μ Jy/beam

Primary Beam Corrections



RMS ~1μ Jy/beam

S. Bhatnagar
NRAO

High sensitivity imaging

- Sensitivity $\propto \frac{N_{ant} A_{ant} \sqrt{N_t \tau N_{chan} \Delta \nu}}{T_{sys}}$
- Higher sensitivity is achieved using larger collecting area ($\propto$ $\mathbf{N_{ant}}$), wider band-widths ($\mathbf{N_{channels}}$) and longer integrations in time ($\mathbf{N_t}$)
  - Data volume $\propto$ $\mathbf{N^2_{ant}\ N_{channels}\ N_t}$
- Implications for high dynamic range imaging
  - Wider field imaging required → finer sampling in time and frequency
    - $\mathbf{N_{channels}}$ = 1-10GHz/MHz and $\mathbf{N_t}$ = 10hr/(1-10sec)
    - Wider range of angles on the sky (==> Direction Dependence)
  - Smaller scale variations over larger parameters space to be accounted for
  - Algorithm efficiency remains a critical parameter
    - 10-100x increase in the number of samples to achieve the required sensitivities

Synthesis Imaging

- Measurement Equation

$$V_{ij}^{Obs}(\nu) = M_{ij}(\nu,t) W_{ij} \int M_{ij}^{S}(s,\nu,t) I(s,\nu) e^{2\pi\iota(b_{ij}.s)} ds$$

$b_{ij} = b_i - b_j$: The Baseline Vector; *i* and *j* represent the two antennas

$M_{ij}(\nu,t) = J_i(\nu,t) \otimes J_j^*(\nu,t)$:Direction <u>independent</u> gains

$M_{ij}^{S}(s,\nu,t) = J_i(s,\nu,t) \otimes J_j^*(s,\nu,t)$:Direction <u>dependent</u> gains

- Unknowns in are the M_{ij}, M^s_{ij} and I
- ***Combined RHS determines the 'time constant"over which averaging will help***

Some observations

- Conventionally, calibration and imaging are treated as independent processes
  - Solving for calibration terms: Solve for ***M*,** using observations of a field with known structure
  - Calibration: Make $V^{Calibrated} = M^{-1} V^{obs}$
  - Imaging: Keep the calibration terms fixed, solve for ***I***

- When Direction Dependent (DD) effects are significant, imaging and calibration can no more be treated as orthogonal
  - Correction for Direction Dependent (DD) effects cannot be separated from imaging

$$V(b_{ij}) = \int M_{ij}^{S}(s) I(s) e^{2\pi\iota(b_{ij}.s)} ds$$

  - Algorithms must fundamentally separate ***I(s)*** from M^s

NRAO

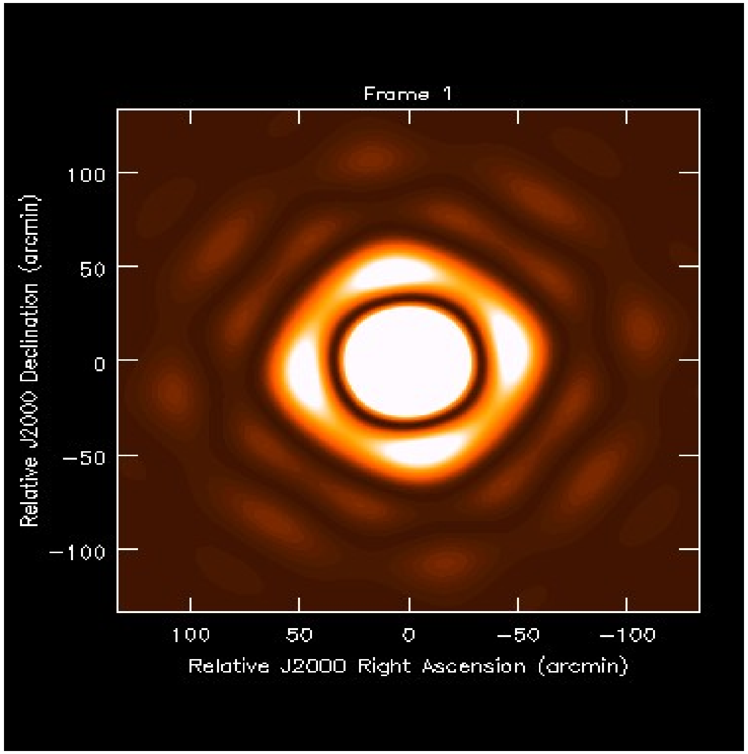
Parametrized Measurement Equation

- Need more sophisticated parametrization of the ME
  - Better parametrization of the J_i, J_i^S *and the* Sky *(I^M)*
  - Solver for the (unknown) parameters
  - Forward and reverse transform that account for the DD terms
  - Efficient run-time implementation

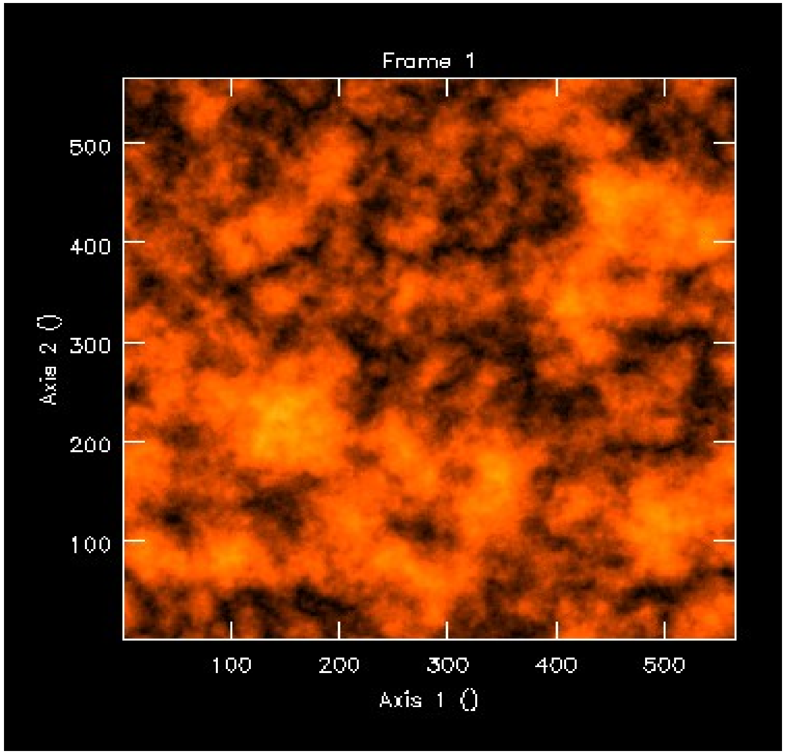
- Useful parametrization:
  - Which models the effects well and with minimum DoF
  - For which efficient solvers can be implemented
  - Which optimally utilizes the available SNR
    - Noise on the solved parameters: $\sigma(p) = \left[\frac{2 k_b T_{sys}}{\eta_a A \sqrt{N_{ant} \nu_{corr} \tau_{corr}} \sqrt{N_{SolSamp}}} \right] \frac{1}{S}$

$$\text{where } S = \int \frac{\partial E_i(\mathbf{s}, p)}{\partial \mathbf{s}} E_j^*(\mathbf{s}, p) I^M(\mathbf{s}) e^{2\pi\iota \mathbf{s}.\mathbf{b}_{ij}} d\mathbf{s}$$

Example of DD effects

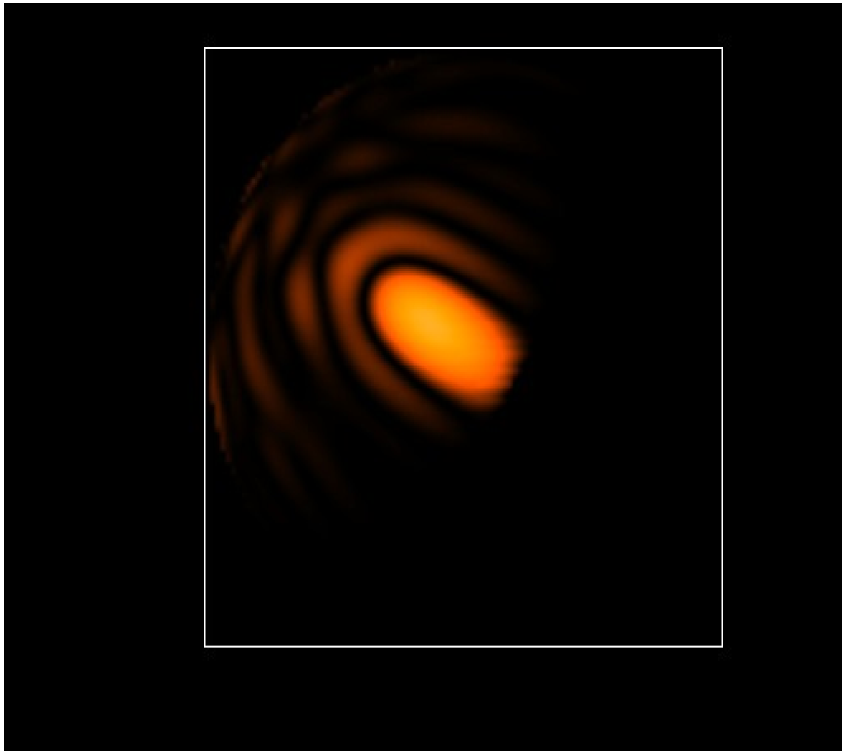


Time and DD Primary Beam: EVLA

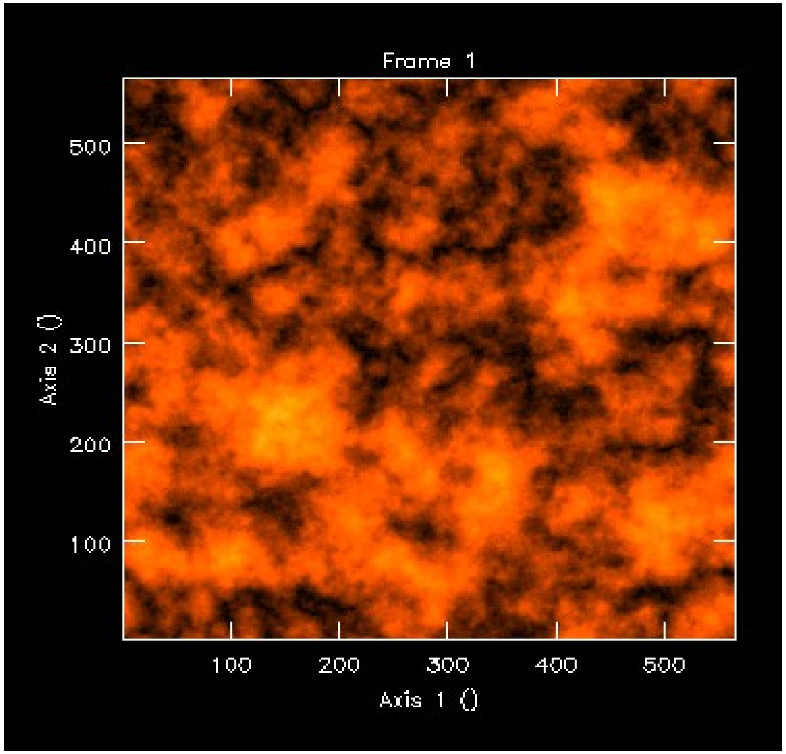


Ionospheric Phase Screen

Examples of DD effects



Time and DD Primary Beam: LWA



Ionospheric Phase Screen

Deconvolution and Calibration: Theory

- Calibration and image deconvolution operations can be described as function optimization

$$V^{Obs} = M\, A\, M^S\, I^{True} + N$$

- N is independent gaussian random variable – in the data domain
- Image deconvolution (CLEAN, MEM,...) $\chi^2 = \left| M^{-1} V^o - A I^M \right|^2$ where $I^M = \sum_k P_k$; P_k is the Pixel Model

$$\frac{\partial \chi^2}{\partial I^M} \equiv \text{Dirty Image} \quad = FT\left[V^{Corr} \right]$$

$$I^M_{Iter} = I^M_{Iter-1} + \alpha \Delta \chi^2$$

- Calibration ("antsol", "self-cal") $\chi^2 = \left| V^o - M\, A\, I^M \right|^2$

$$\frac{\partial \chi^2}{\partial M} \equiv \text{Update direction}$$

$$M_{Iter} = M_{Iter-1} + \beta \Delta \chi^2$$

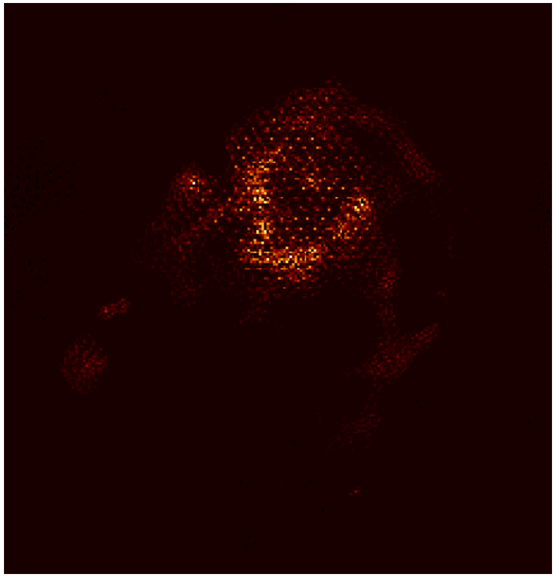
- Requires: (1) Residual = Obs. Data – Current Model $\left(V^o - Model\left(M, M^S, I \right) \right)$

  (2) Derivative computation (approx.)

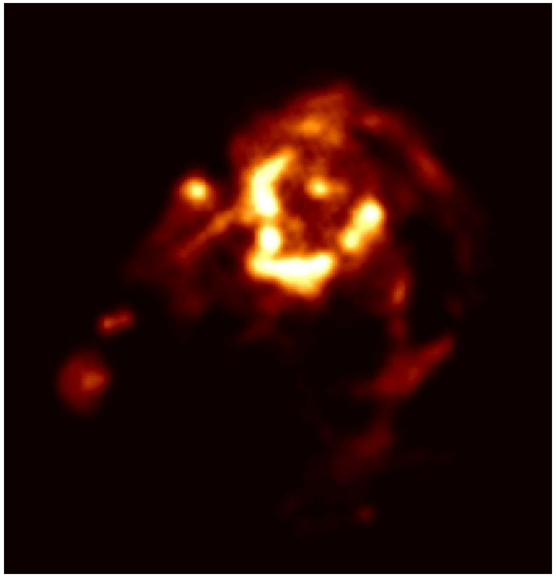
Advances in Calibration and Imaging Techniques in Radio Astronomy, Rau et al., Proc. IEEE, Vol. 97, No. 8, Aug.2009, 1472

Deconvolution: Parametrization of the emission

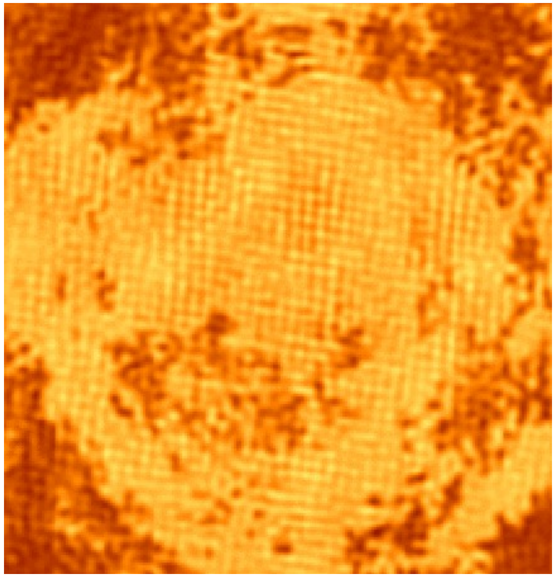
- Scale-less deconvolution algorithms:
  - $I^M = \sum_k A_k \delta(x - x_k)$:Treat each pixel as an independent DoF
  - CLEAN (and its variants), MEM (and its variants)



Component Model



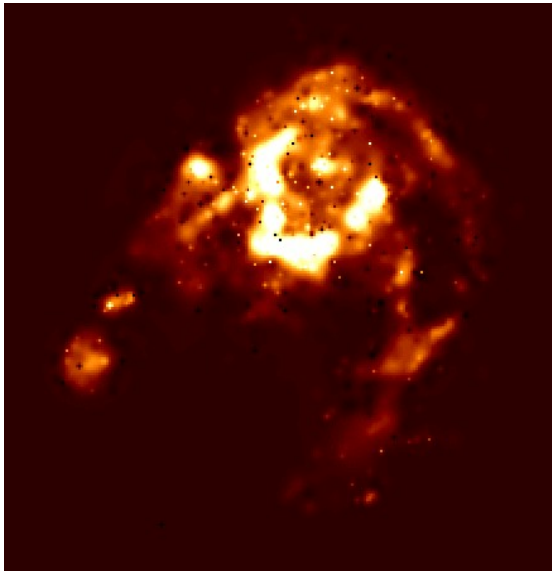
Restored Model



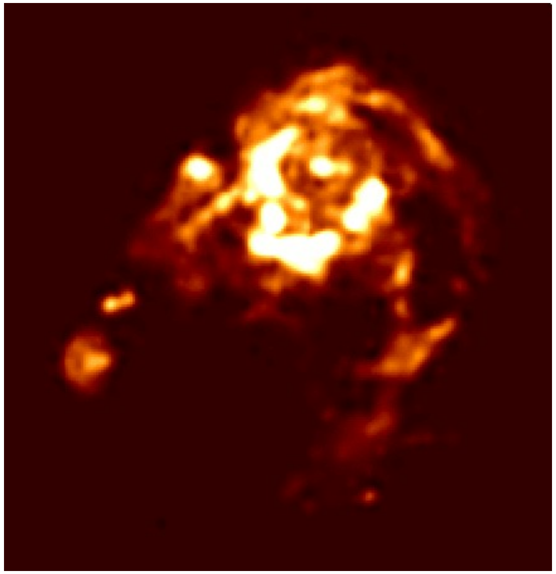
Residuals

Deconvolution: Parametrization of the emission

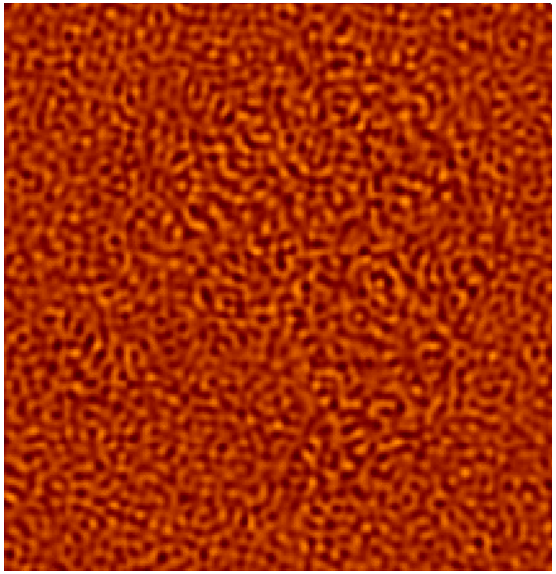
- Scale-sensitive deconvolution algorithms:
  - $I^M = \sum_k A_k f(Scale, Position)$: Decompose the image in a scale-sensitive basis
  - Asp-Clean (A&A, 747, 2004 (astro-ph/0407225), MS-Clean (IEEE JSPSP, Vol. 2, No. 5, 2008)



Component Model

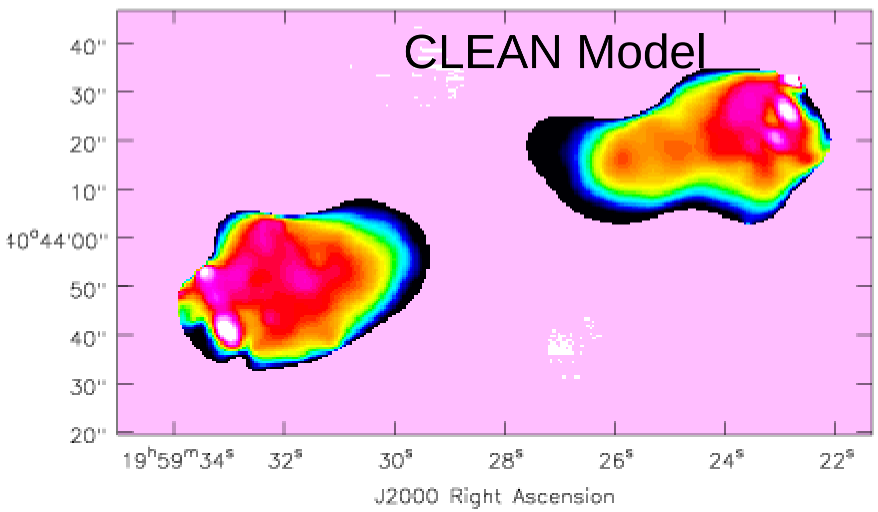


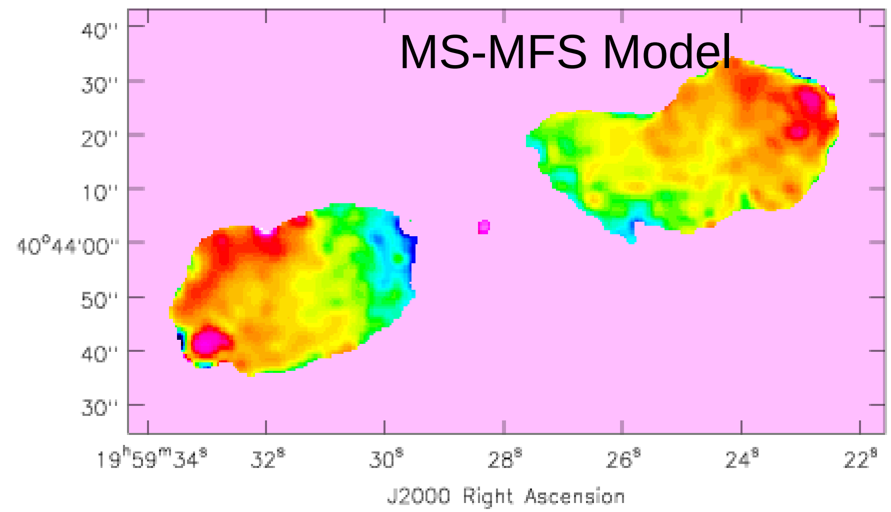
Restored Model



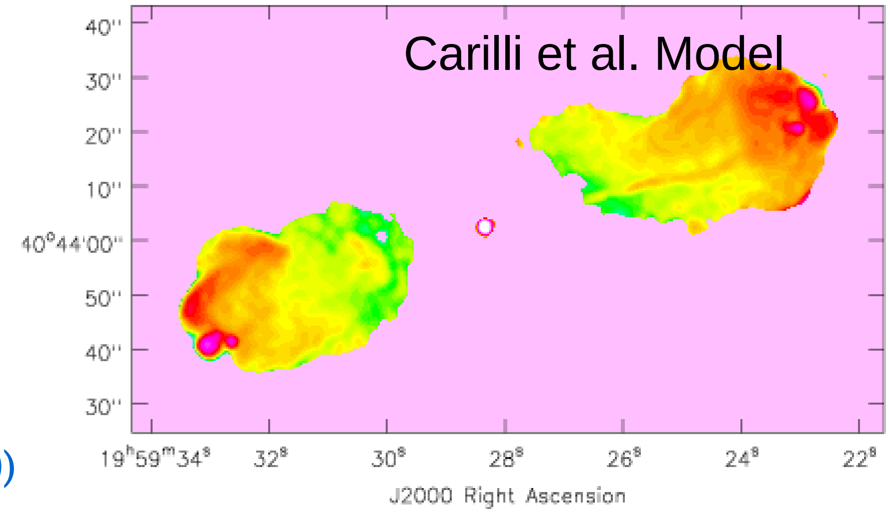
Residuals

Source structure: Frequency and spatial





- MS-MFS: Multi-scale modeling of the source structure + Taylor expansion along frequency axis
- Snap shot uv-coverage per Spectral Window (10x50 MHz)
- Carilli et al.: Full synthesis in multiple VLA configuations.



(PhD Thesis, Rao Venkata, NRAO/NMT, 2010)

Direction Dependent Effects

- Difficult to "correct" raw data for DD terms.

$$V(b_{ij}) = \int M_{ij}^{S}(s) I(s) e^{2\pi\iota(b_{ij}.s)} ds$$

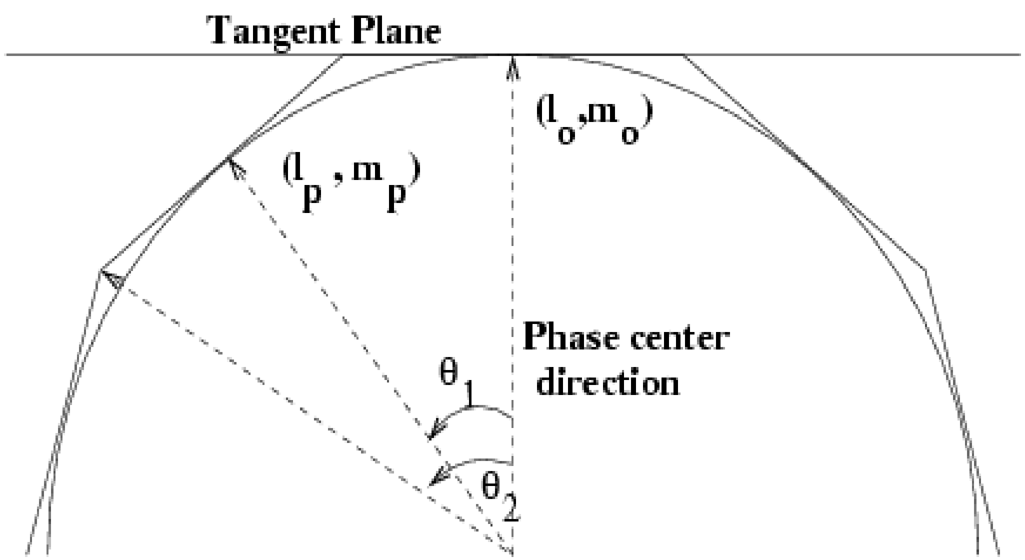
- Removal of DD effects then fundamentally couples calibration and imaging processes
  - Include DD effects as part of imaging/deconvolution process.
- Pieces required
  - Efficient algorithms to correct for DD effects during imaging
  - Develop well constrained parametrized models that described DD effects
  - Efficient algorithms to solve for the parametrized the DD terms

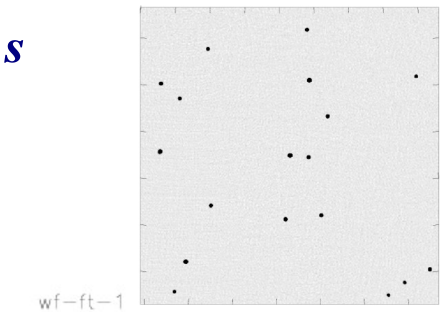
- 100s GB data sets, higher computing needs of new algorithms ==> integrated approach for largely automated/pipeline processing
- All effects of the same order **must** be accounted for simultaneously to validate the algorithms

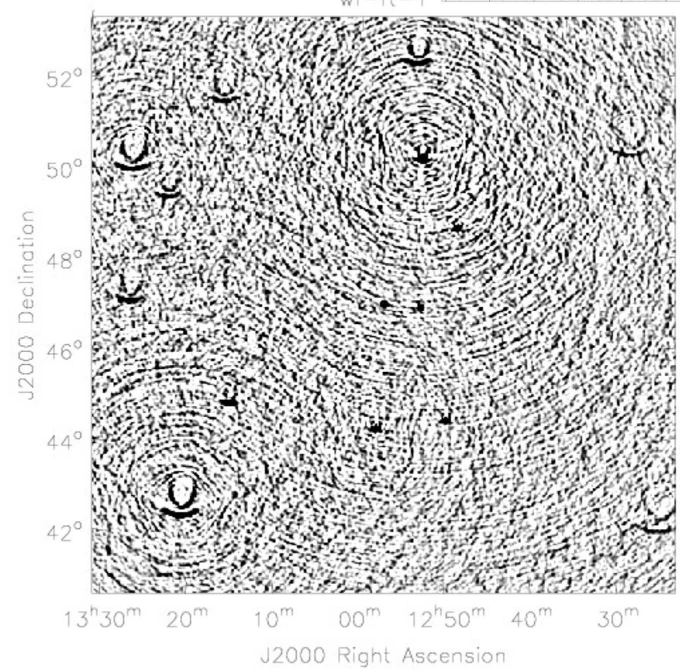
DD Effects: Non co-planar baselines

- 2D Fourier Transform approximation is not valid: The "w-term" error

$$V^{Obs}(b_{ij}) = \int P^{S}_{ij}(s, \nu, t)\, I(s, \nu)\, e^{2\pi\iota[u_{ij} l + v_{ij} m + w_{ij}(\sqrt{1-l^2-m^2}-1)]}\, ds$$

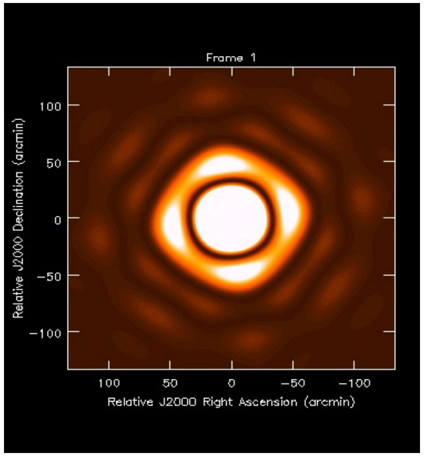


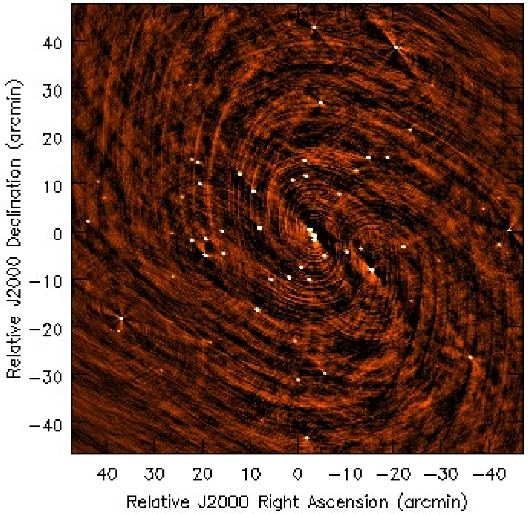




DD Effects: Time varying Primary Beams

- Variations with time due to
  - Rotation with Parallactic Angle for El-Az mount antennas (GMRT, EVLA, ALMA)
  - Frequency and polarization dependence (most telescopes)
  - Projection effects (Aperture Array elements)
  - Pointing errors (all telescopes), structural deformation
  - Heterogeneous antenna arrays



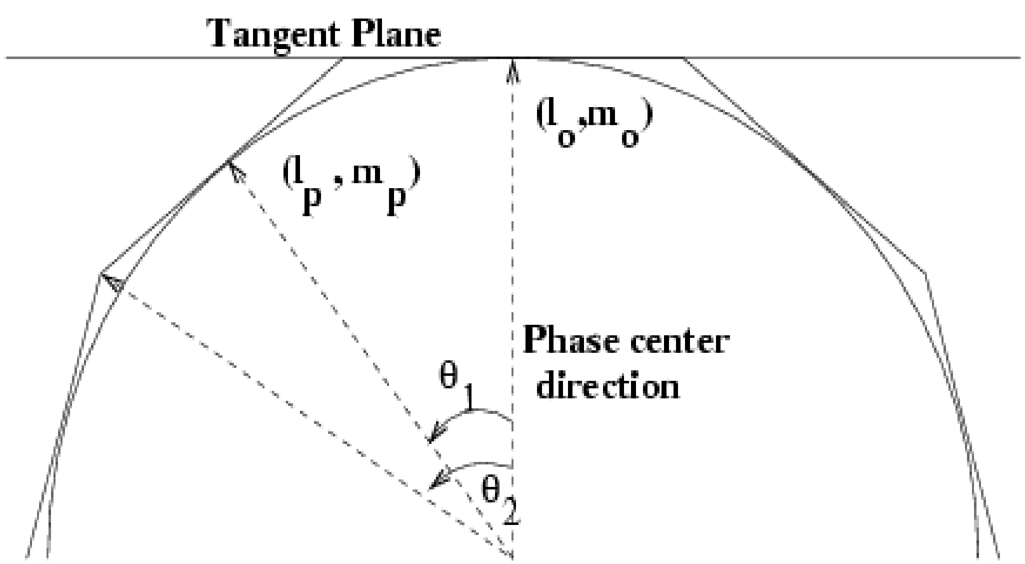


Faceted imaging: Piece-wise constant approx.

- A fundamentally measurement effect, modeled as an image domain effect

$$V^{Obs}(b_{ij}) = \int P^{S}_{ij}(s,\nu,t)\, I(s,\nu)\, e^{2\pi\iota[u_{ij}l + v_{ij}m + w_{ij}(\sqrt{1-l^2-m^2}-1)]}\, ds$$

- Combination of FFT (on the flat facets) + DFT



- Trouble with extended emission
- Multiple intermediate images
- High algorithmic complexity
- Difficult to interface with other algorithms to account for other effects
- Computationally sub-optimal

Projection methods

- The Measurement Equation in linear algebra notation

$$V^{Obs} = F^T M^S I = E V$$

  - *E* represents direction dependent effects in the visibility (data) domain

- If an operator ***K*** that models ***E*** and has appropriate properties can be constructed, K^T *and* ***K*** can be used in forward and inverse transforms to produce distortion free images (a.k.a. "operater pre-conditioning").
  - Forward transform: $FK^T V^{Obs} = I^d$ *(Imaging)*
  - Inverse transform: $KF^T I^M = V^M$ *(Model Data computation)*

- *For computational efficiency, **E** should have a finite support size*
- *Residual and derivative computation done at approximately FFT efficiency*

The W-Term: Optics interpretation

- $$V^{Obs}(b_{ij}) = \int P^S_{ij}(s,\nu,t)\, I(s,\nu)\, e^{2\pi\iota[u_{ij}l+v_{ij}m+w_{ij}(\sqrt{1-l^2-m^2}-1)]}\, ds$$
  - Visibility and the image are not related by a 2D Fourier Transform
    - 2D FT only if (1) small field-of-view, or (2) small baseline length

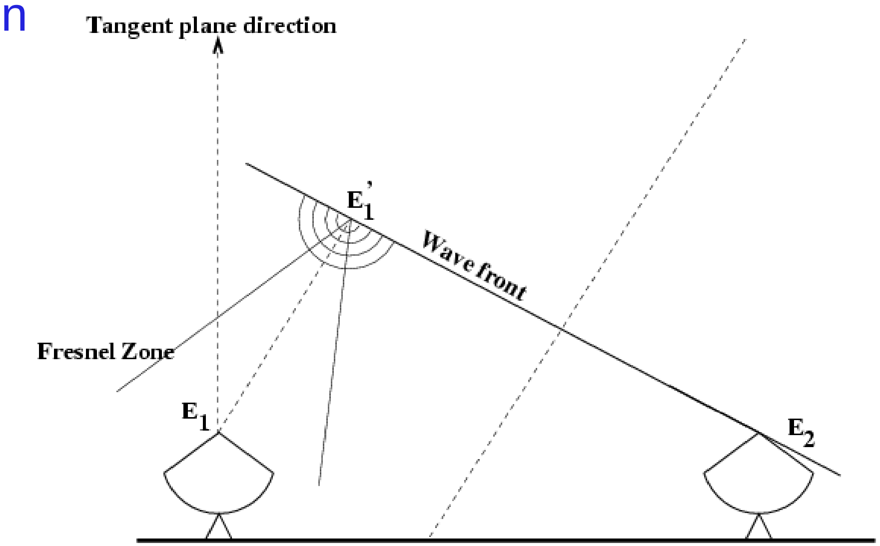
- What we measure is <$E_1 E^*_2$>
- What want to measure: <$E'_1 E^*_2$>

- E_1 is E'_1 propagated using the Fresnel Diffraction

$$V^o(u,v,w) = G\, V(u,v)$$

$$V^o_{12} = \int e^{2\pi\iota w_{12}\sqrt{1-l^2-m^2}-1}\, I(l,m)\, e^{2\pi\iota[u_{12}l+v_{12}m]}\, dl\, dm$$

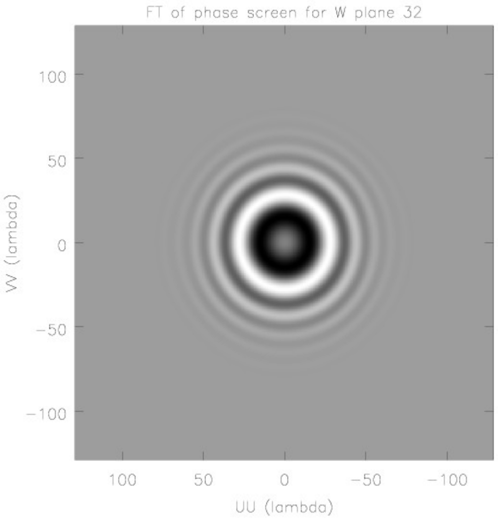
$$G(u,v,w) \;=\; Fresnel\ Propaga$$



W-Projection algorithm

- $G = FT\left[e^{2\pi\iota\sqrt{1-l^2-m^2}-1}\right]$

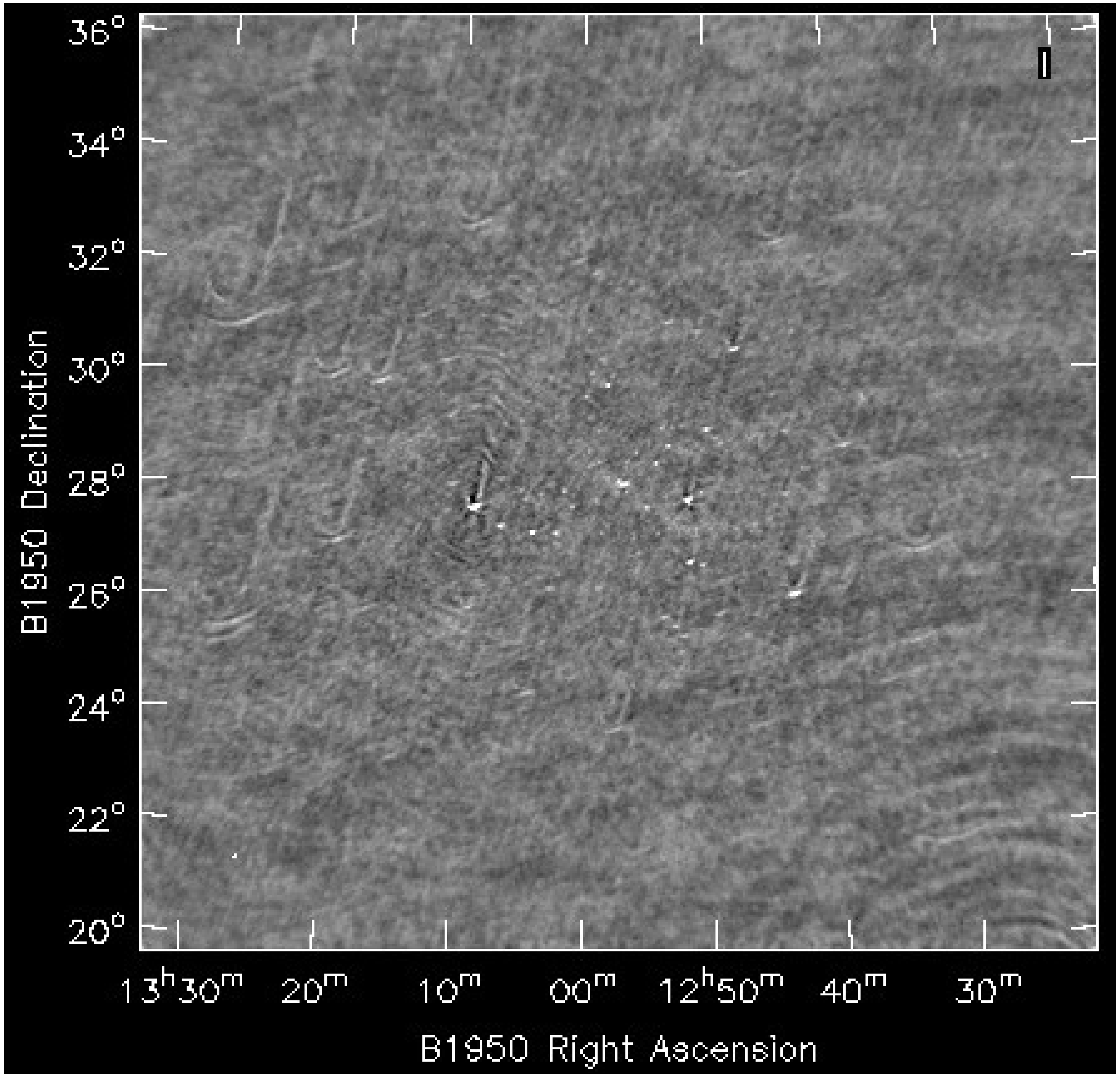
  $V^o(u,v,w) = V^M(u,v) * G(u,v,w)$



- About 10x faster than faceted imaging algorithm
- Straight forward to combine with wide-band imaging and algorithms to correct for other DD effects

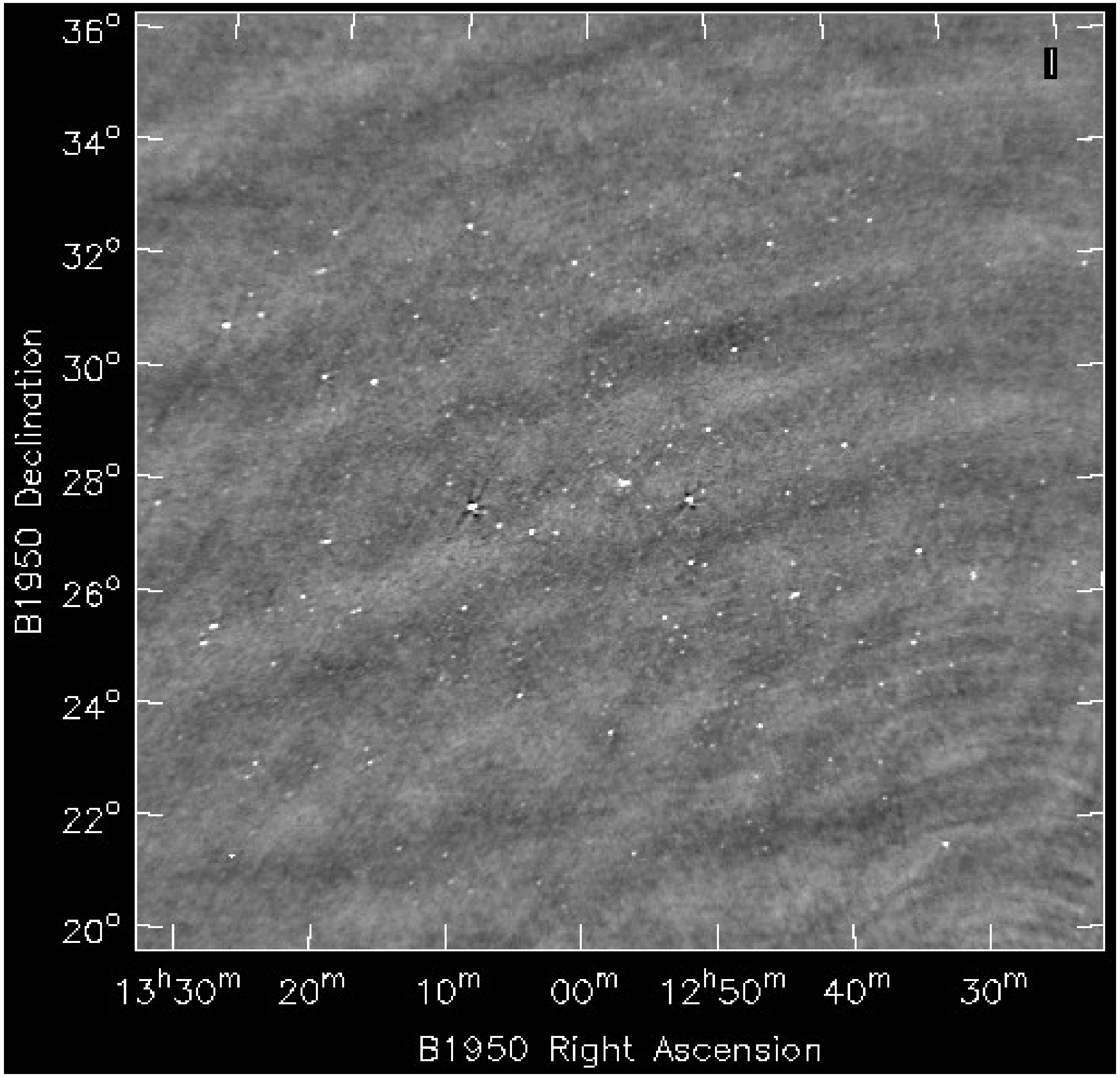
The non-coplanar baseline effect: The W-Projection algorithm, IEEE JSTSP, 2008

Without W-term correction



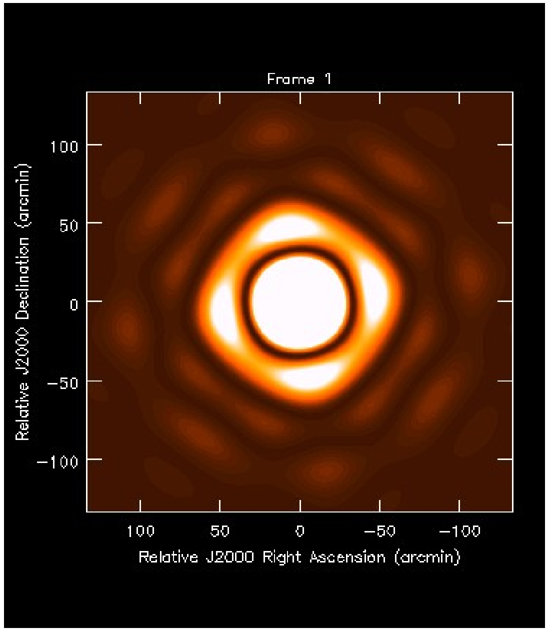
The Coma Cluster
At 74 MHz

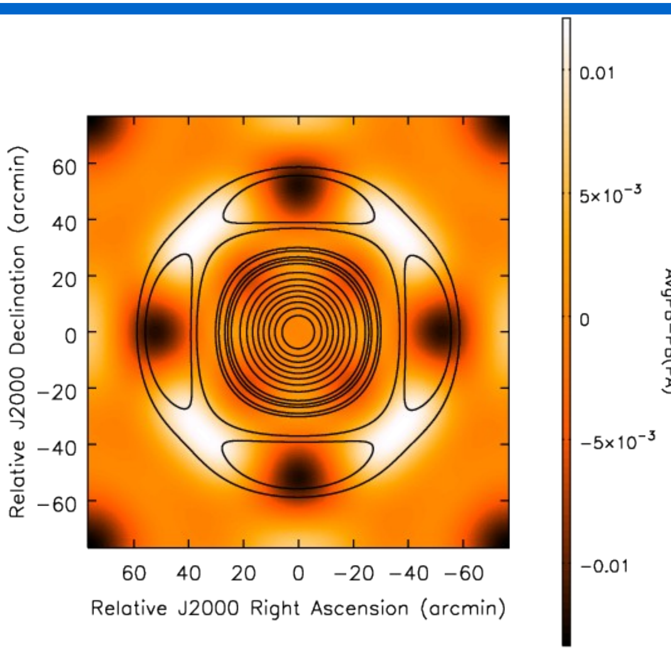
With W-Projection algorithm



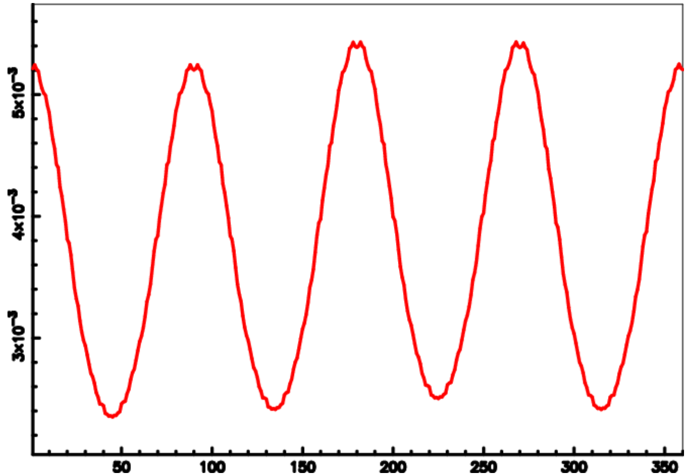
The Coma Cluster
At 74 MHz

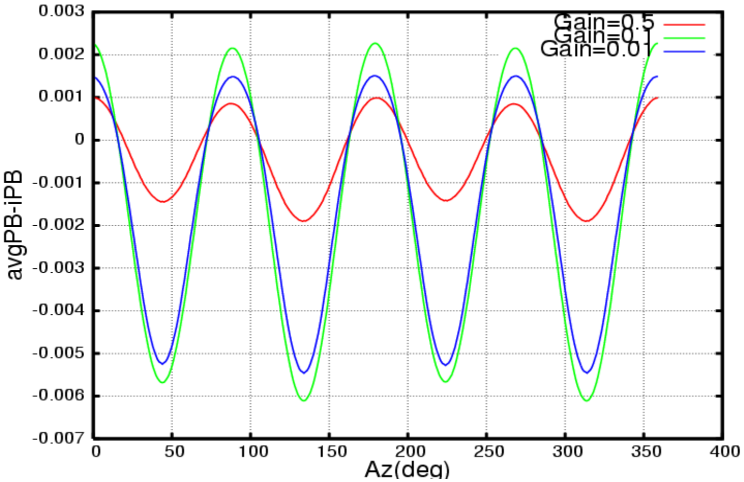
Time varying DD gains due to PB





ΔPB

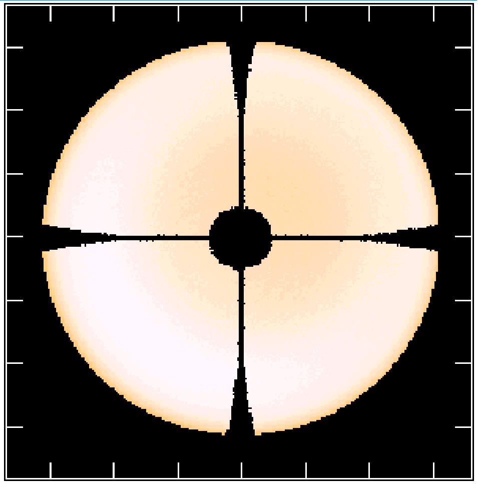




$$\Delta I = PSF * (I \Delta PB)$$

A-Projection algorithm

- $V^o(u,v,w) = V^M(u,v) * G(u,v;Time,Poln.)$



Model for EVLA aperture illumination (real part)

- Modified forward and reverse transforms:
  - No assumption about sky properties
  - Spatial, time, frequency and polarization dependence naturally accounted for
  - Done at approximately FFT speed

- Combining with W-Projection or image plane part of the various deconvolution algorithms is straight forward (algorithm complexity is lower)
- Efficient solvers to solve for more precise parametrized models (Pointing SelfCal and its extensions)

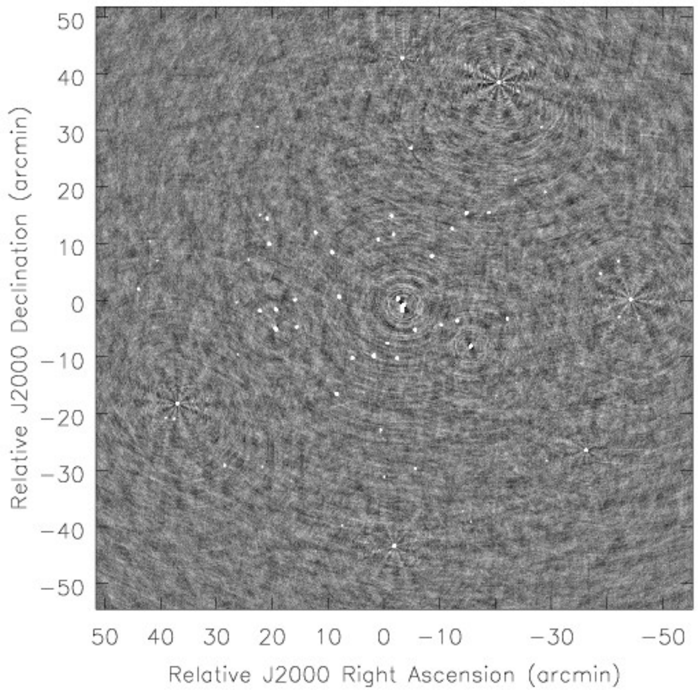
A-Projection algorithm, A&A 2008

Full beam imaging

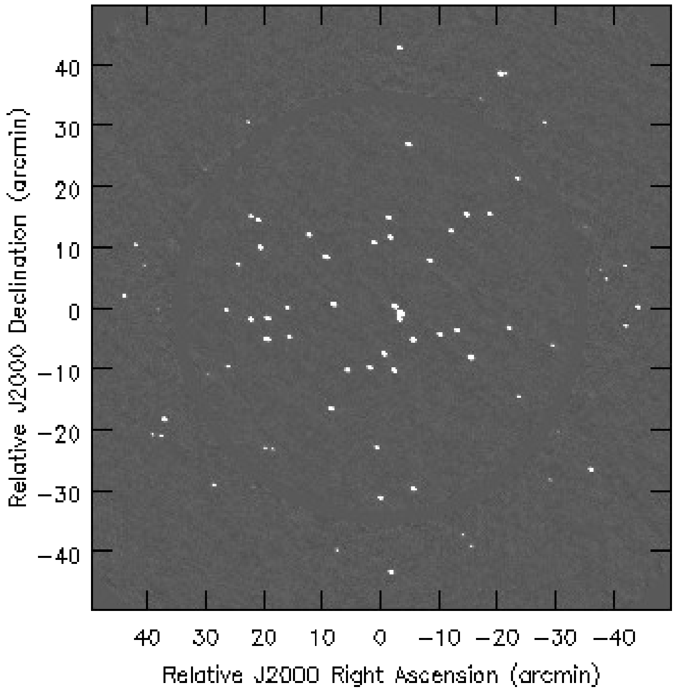
- Limits due to the rotation of asymmetric PB
  - Max. temporal gain variations @ ~10% point
  - DR limit: few X 10^4:1
- Limits due to antenna pointing errors
  - In-beam error signal max. @ 50% point
  - DR limit: few X 10^4:1
  - Limits for mosaicking would be worse
    - Significant flux at half-power point
    - Significant flux in the side-lobes for most pointing
- Approach taken
  - Algorithm R&D (SNR per DoF, error propagation, computing requirements,....)
  - Proof-of-concept tests with realistic simulation
  - Apply to real data

A-Projection algorithm: PB corrections

Before Correction



After Correction

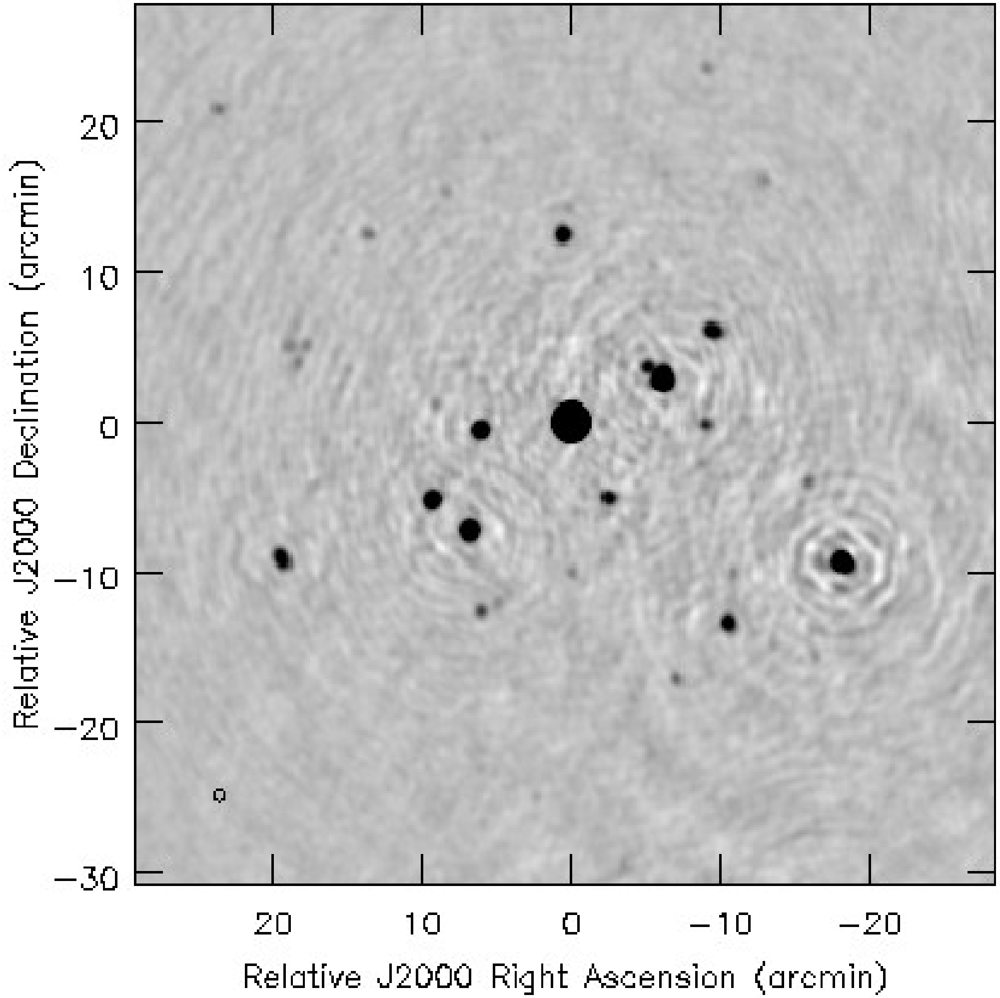


$$\textit{Minimize}: V_{ij}^{O} - E_{ij} * [F I^{M}] \quad w.r.t. \quad I^{M}$$

Goal: Full-field, full-polarization imaging at full-sensitivity

A-Projection: Bhatnagar et al., A&A,487, 2008

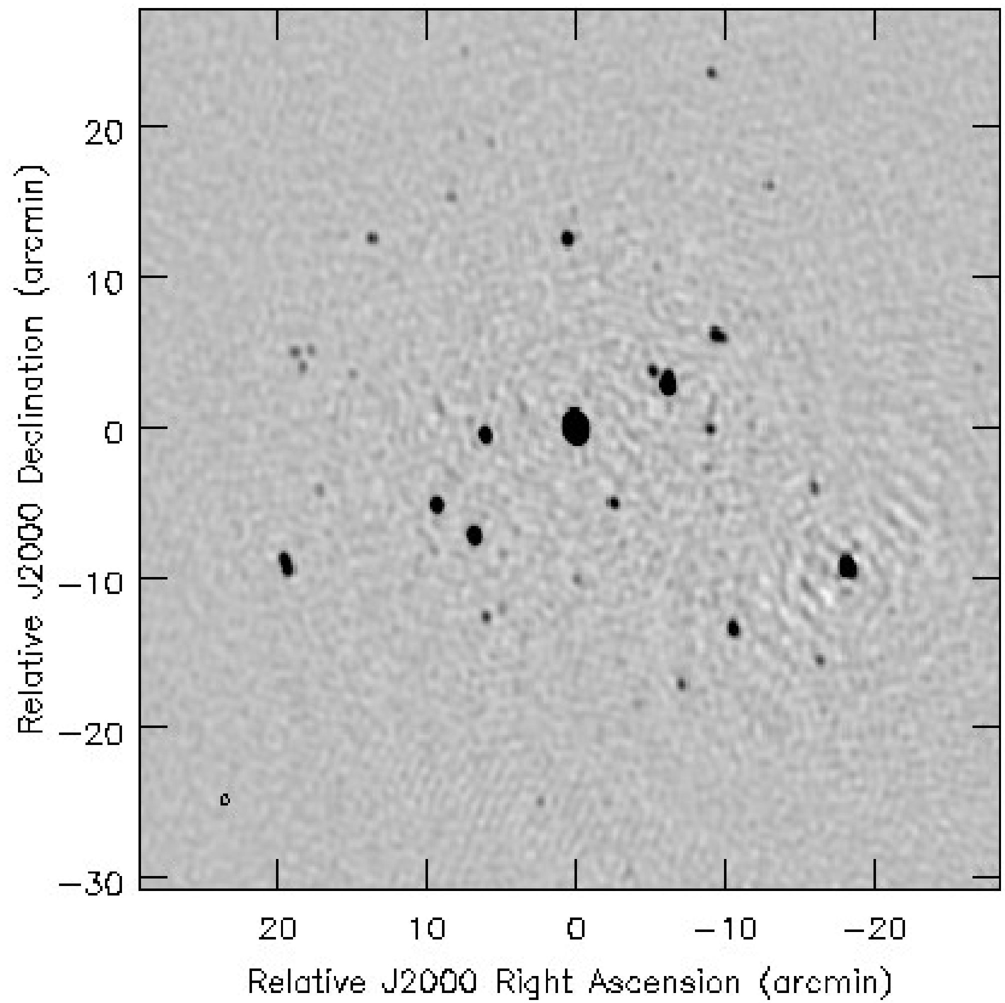
EVLA L-Band Stokes-I: Before correction



- 3C147 field at L-Band with the EVLA
- Only 12 antennas used
- Bandwidth: 128 MHz
- ~7 hr. integration

- Dynamic range: ~700,000:1

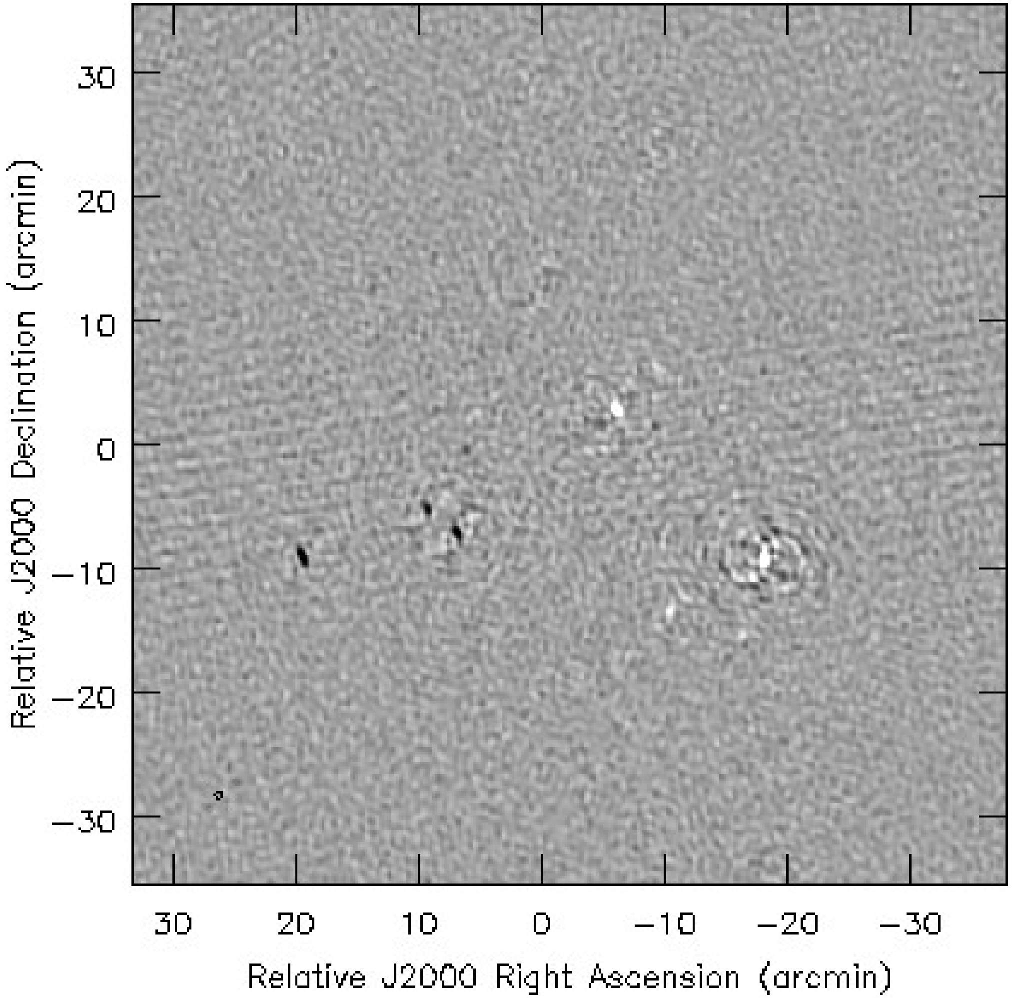
EVLA L-Band Stokes-I: After correction



- 3C147 field at L-Band with the EVLA
- Only 12 antennas used
- Bandwidth: 128 MHz
- ~7 hr. integration

- Dynamic range: ~700,000:1

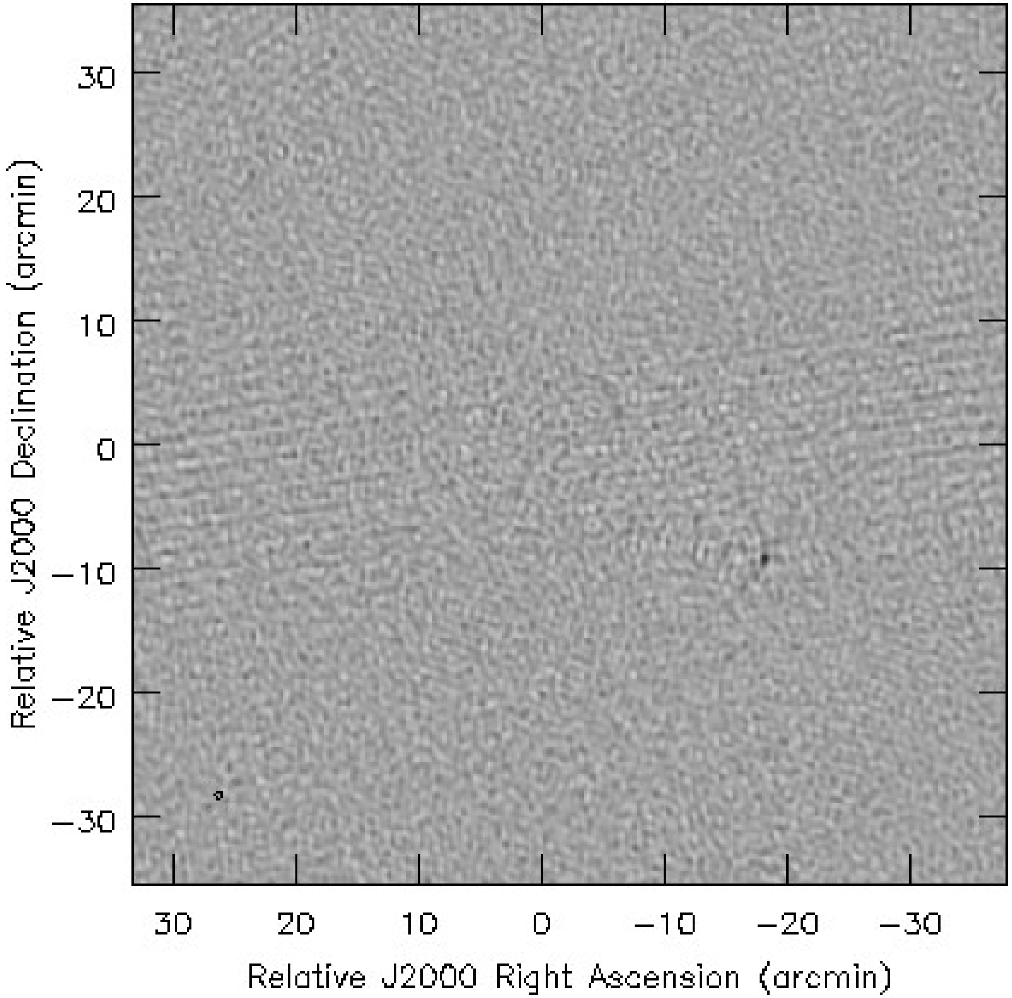
EVLA L-Band Stokes-V: Before correction



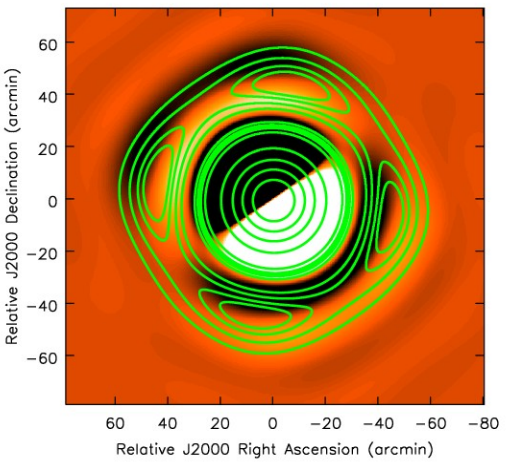
Is it $\boldsymbol{M}(\boldsymbol{s}, \boldsymbol{Poln})$ *?*

Or is it $\boldsymbol{I}(\boldsymbol{s}, \boldsymbol{Poln})$ *?*

EVLA L-Band Stokes-I: After correction

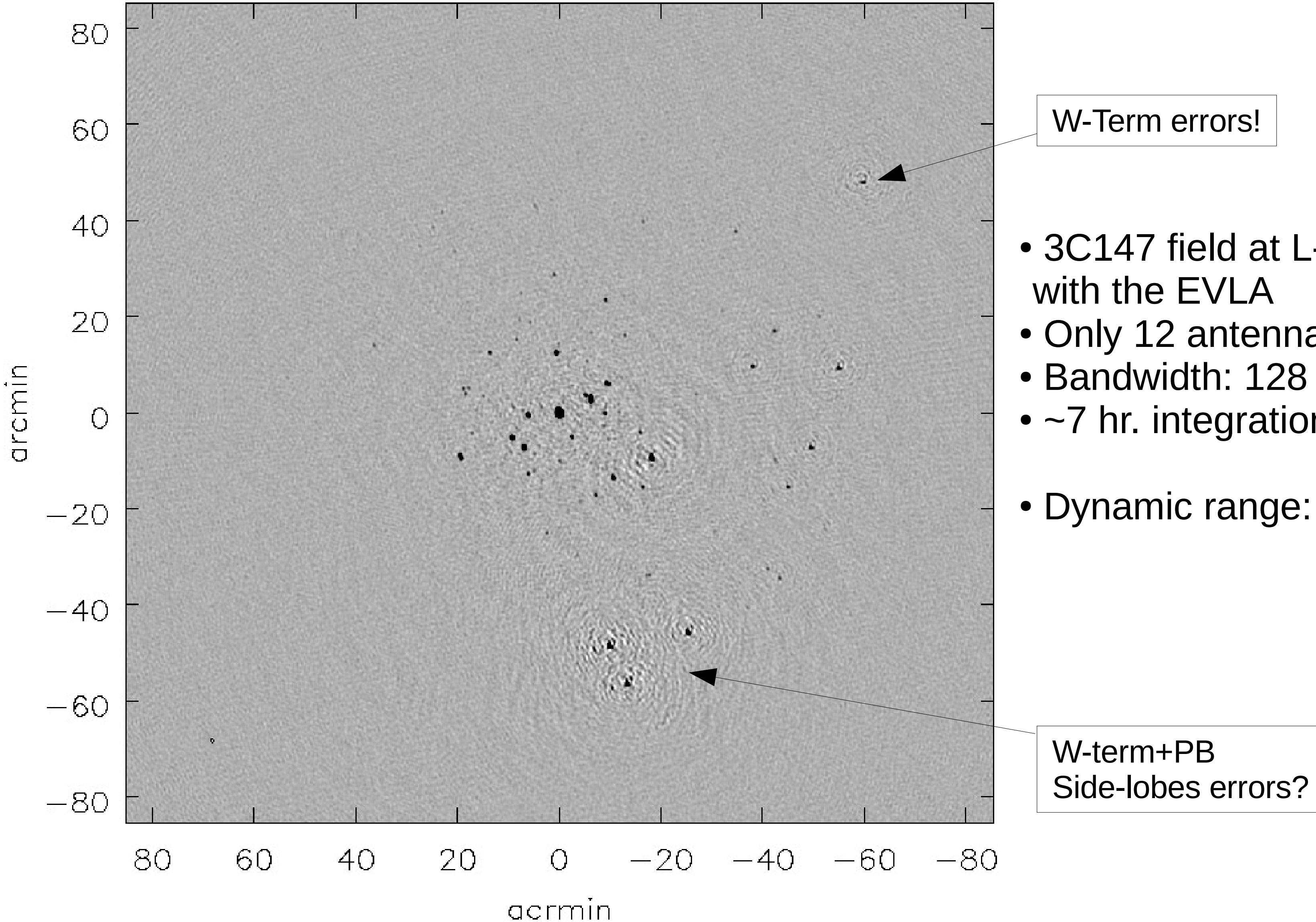


Use physical model for the Stokes-V pattern:



Contours: Stokes-I power pattern
Colour: Stokes-V power pattern

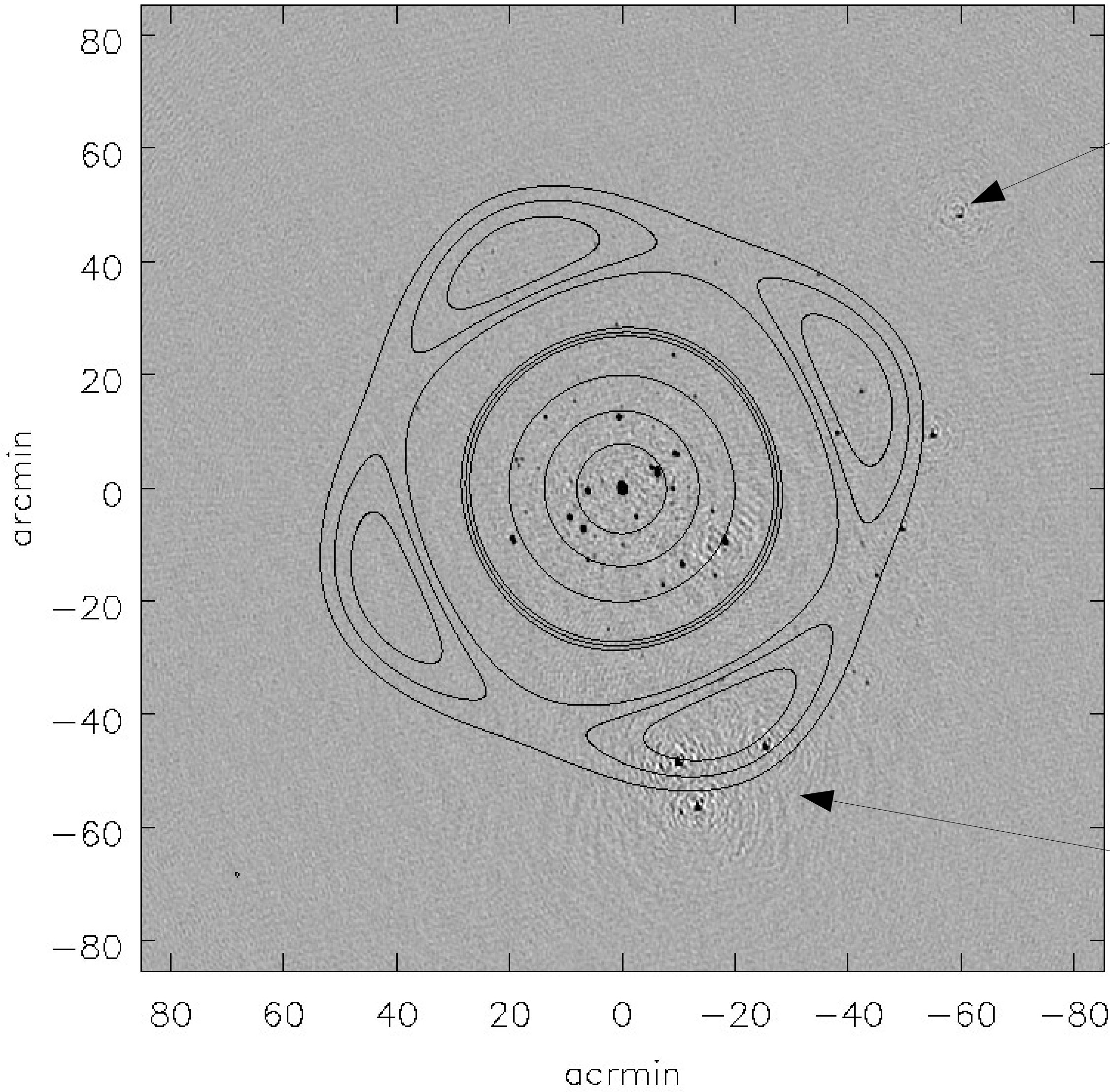
3C147: Full field



- 3C147 field at L-Band with the EVLA
- Only 12 antennas used
- Bandwidth: 128 MHz
- ~7 hr. integration

- Dynamic range: ~700,000:1

3C147: Full field



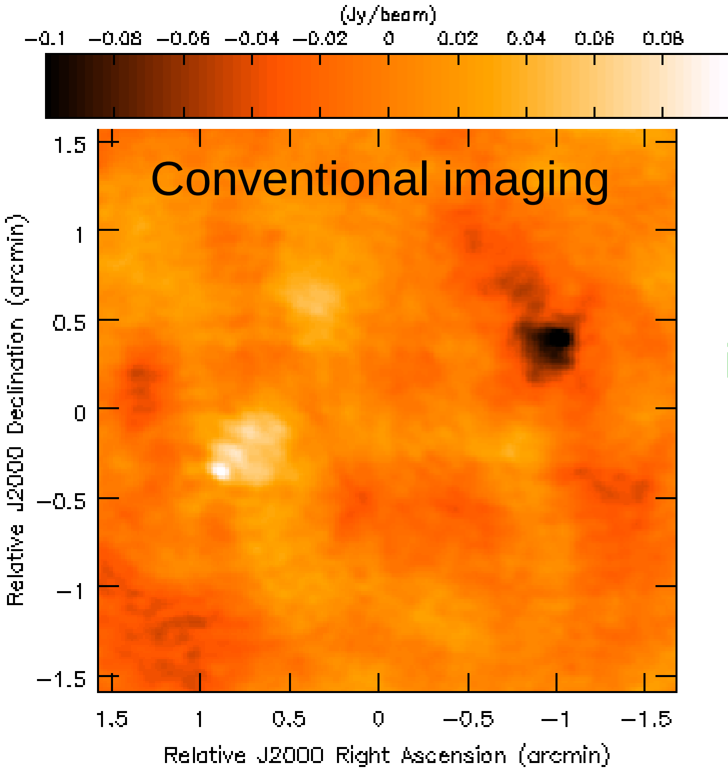
W-Term errors!

- 3C147 field at L-Band with the EVLA
- Only 12 antennas used
- Bandwidth: 128 MHz
- ~7 hr. integration

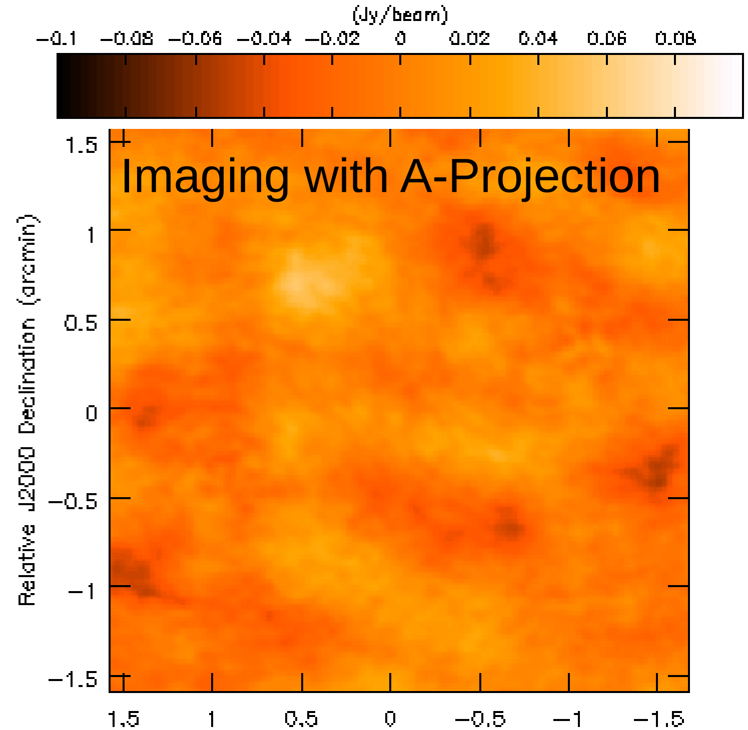
- Dynamic range: ~700,000:1

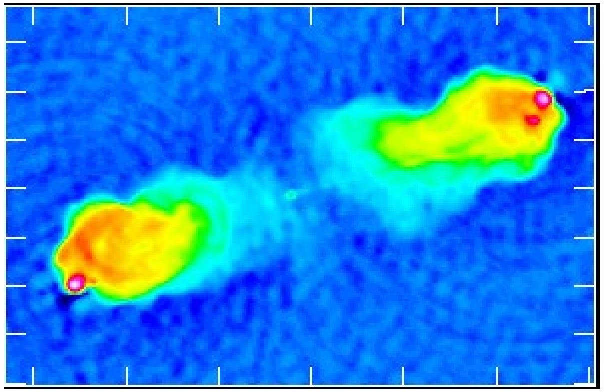
Errors due PB side-lobes?

Example: Imaging extended emission



Run time for both is comparable.



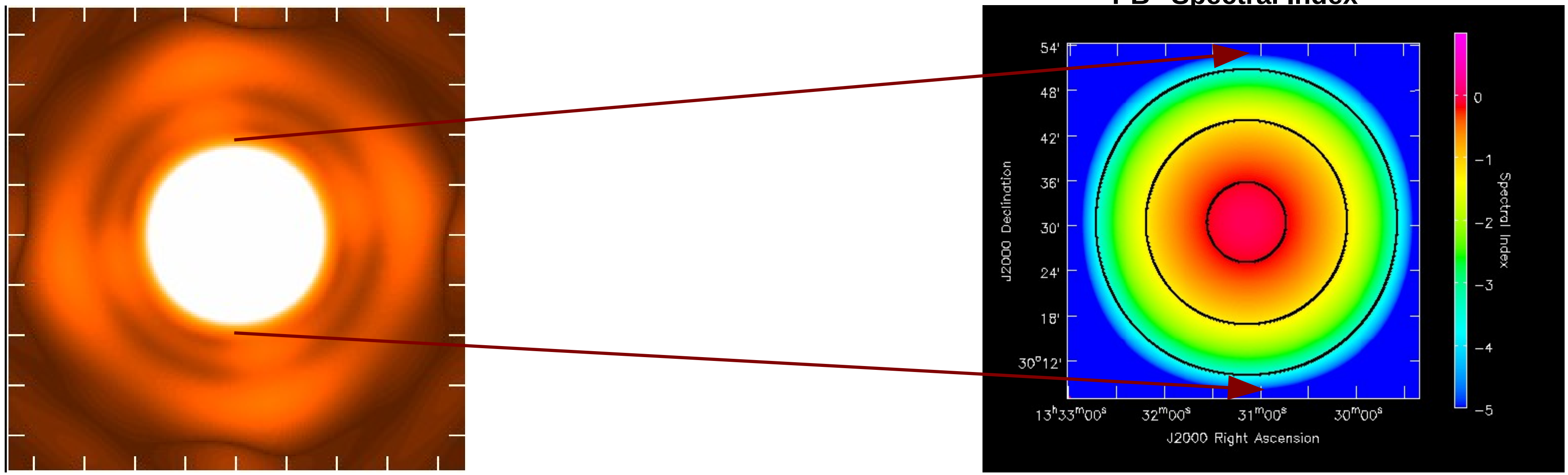


- Stokes-V imaging of extended emission
- Algorithms designed for point sources will not work

Wide-band PB effects

- For wide-band observations, frequency dependence of the PB is a first order effect

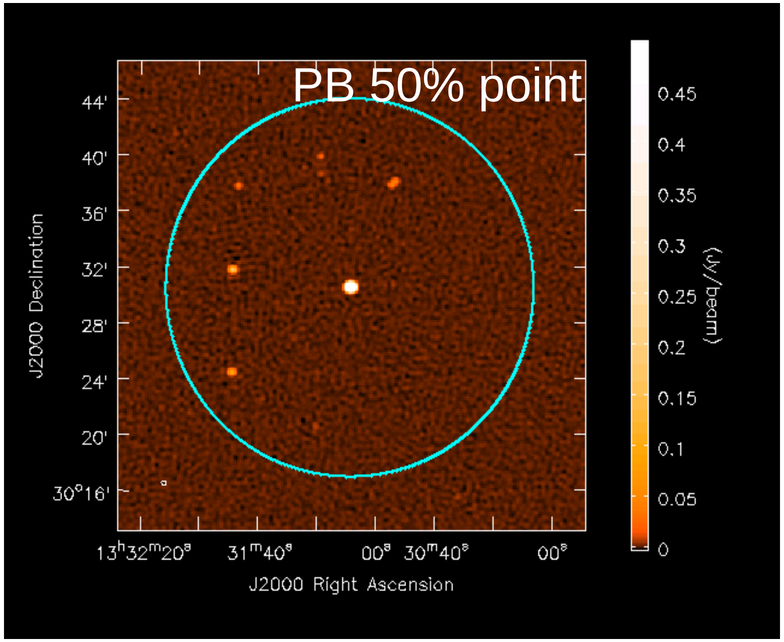
PB “Spectral Index”



$$V(b_{ij}) = \int M^S_{ij} I(s) e^{2\pi\iota(b_{ij}\cdot s)} ds$$

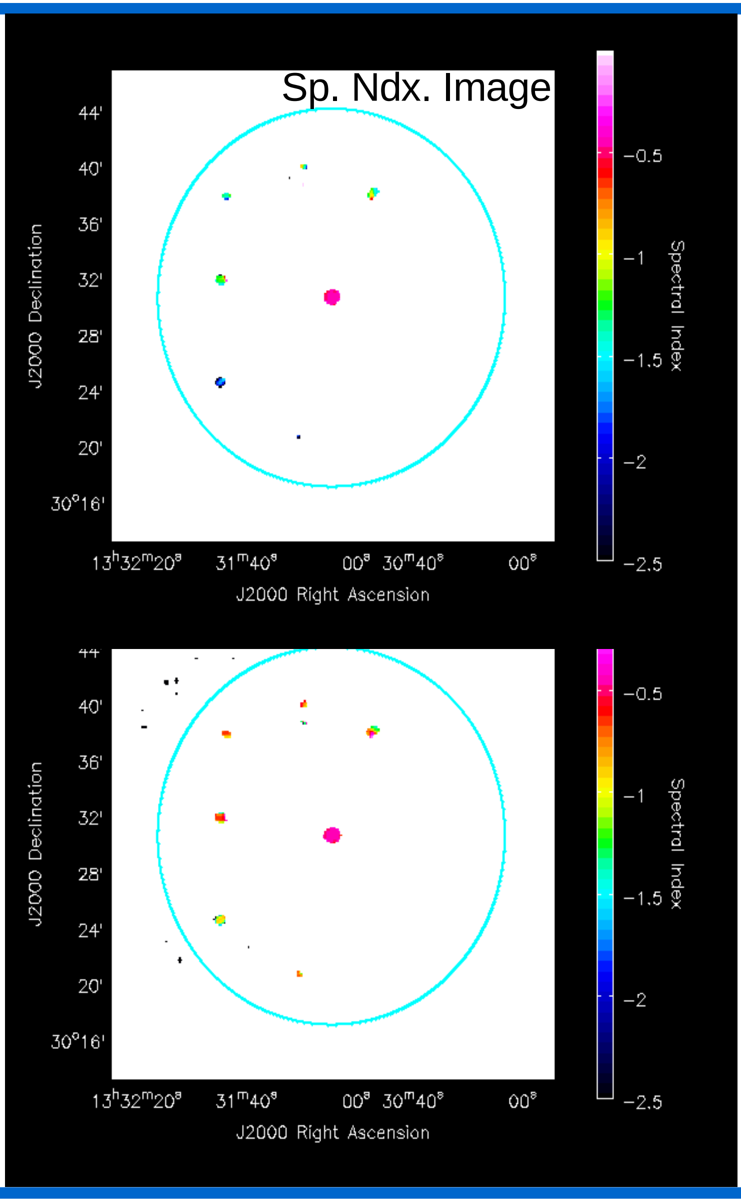
- Is it $M(s,\nu)$ or is it $I(s,\nu)$?
- Fundamental separation: Include PB as part of the measurement process (include its effect as part of forward and reverse transforms)

Wide band imaging of 3C286 field



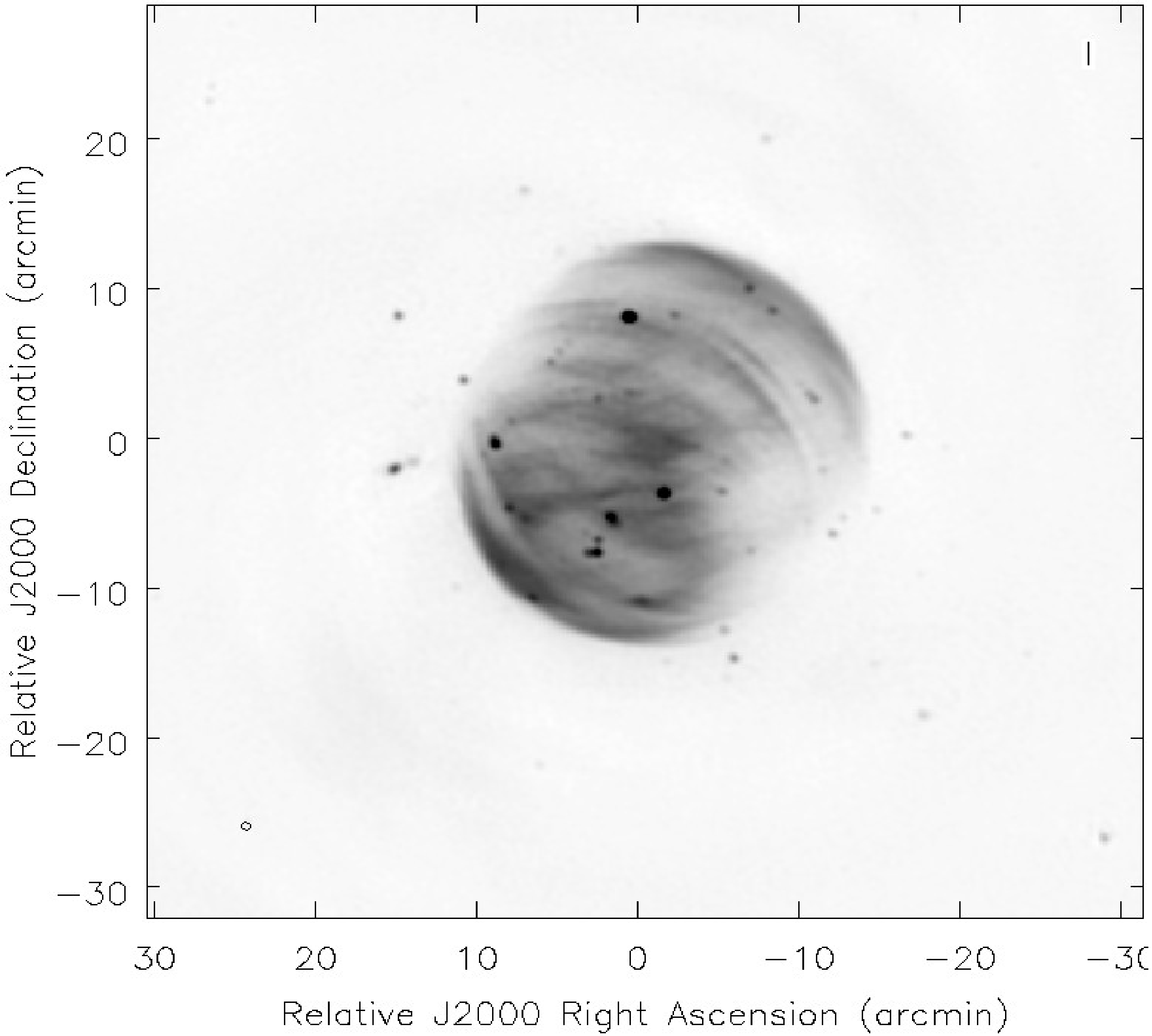
Without PB correction

With PB correction



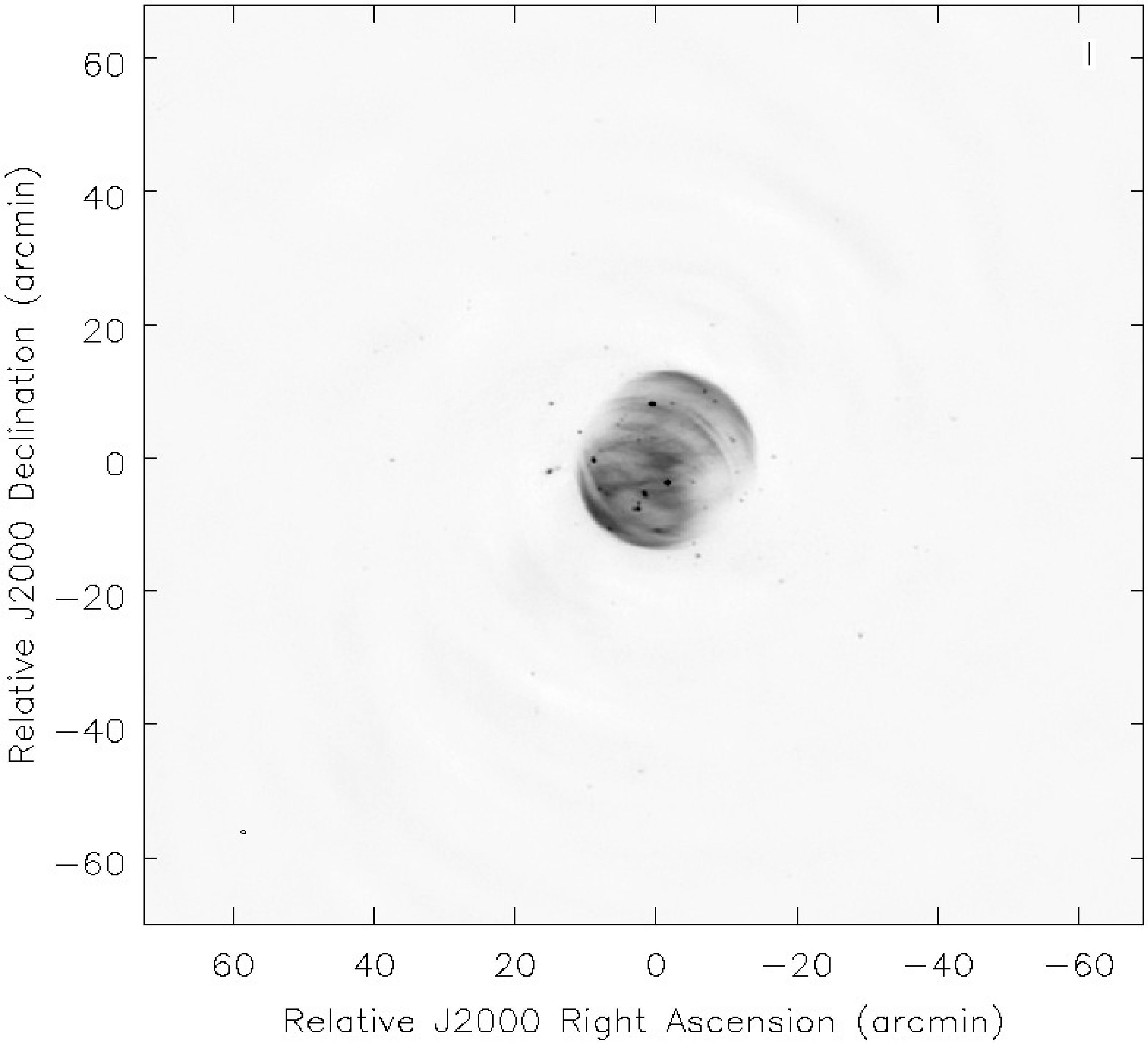
(Rao Venkata, PhD Thesis, 2010)

Wide band imaging with the EVLA

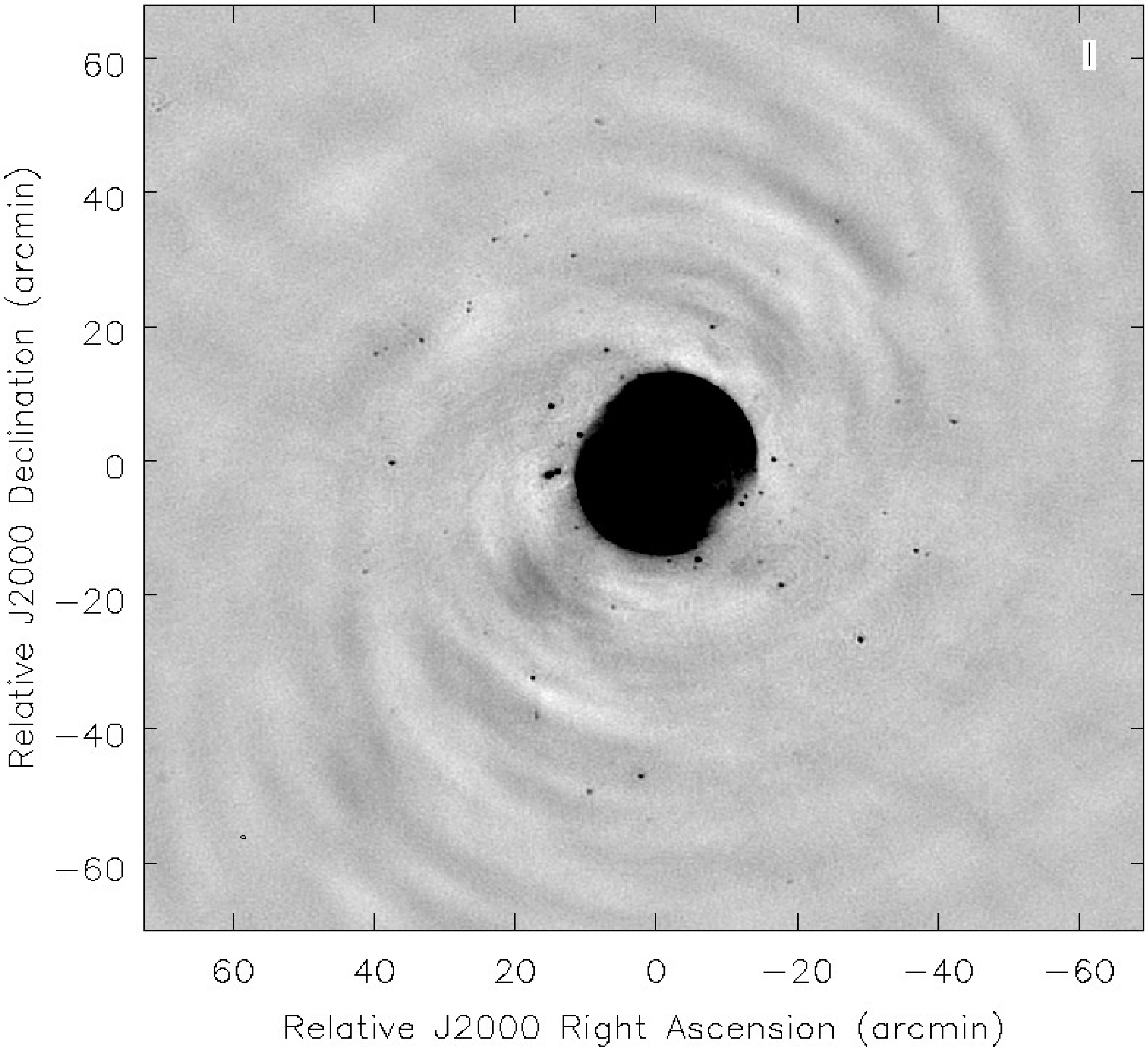


- 1.2-1.8GHz
- ~40 microJy/Beam
- RSRO Projects
  (AB1345, Bhatnagar et al.
  AT374, Taylor et al.)
- Scientific goals
  Spectral Index imaging
  RM Synthesis

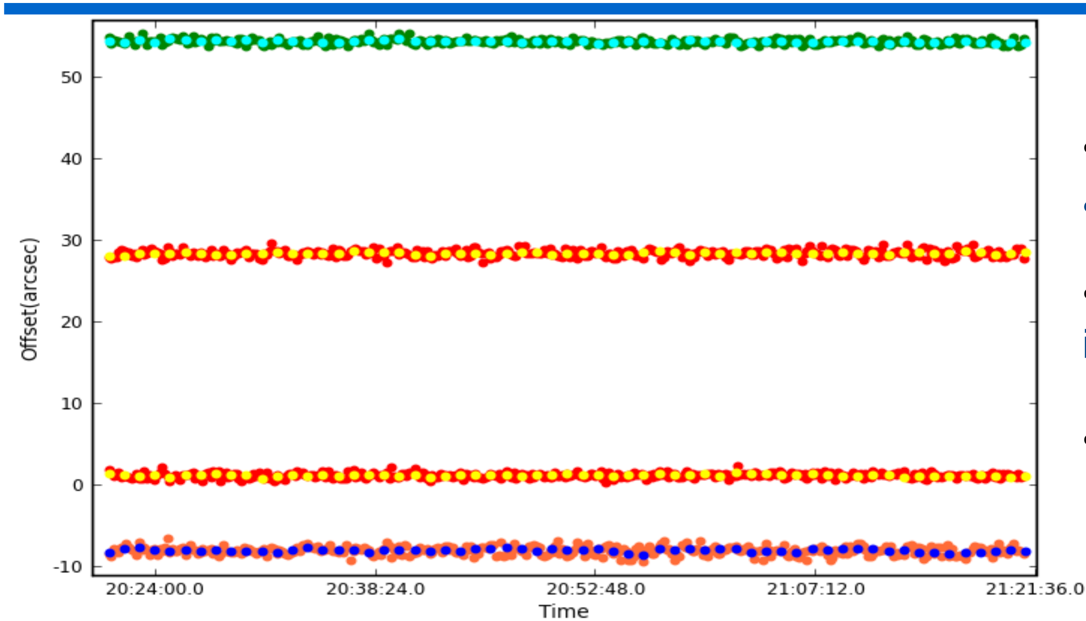
Wide band imaging with the EVLA



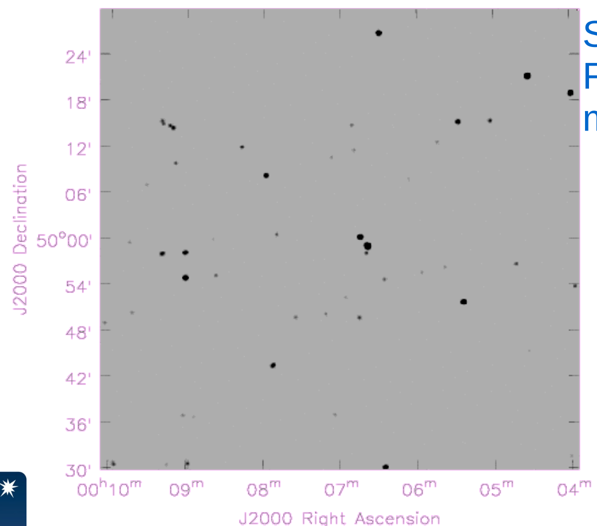
Wide band imaging with the EVLA



Pointing SelfCal: DD SelfCal algorithm



- Typical antenna pointing offsets for VLA as a function of time
- Over-plotted data: Solutions at longer integration time
- Noise per baseline as expected from EVLA

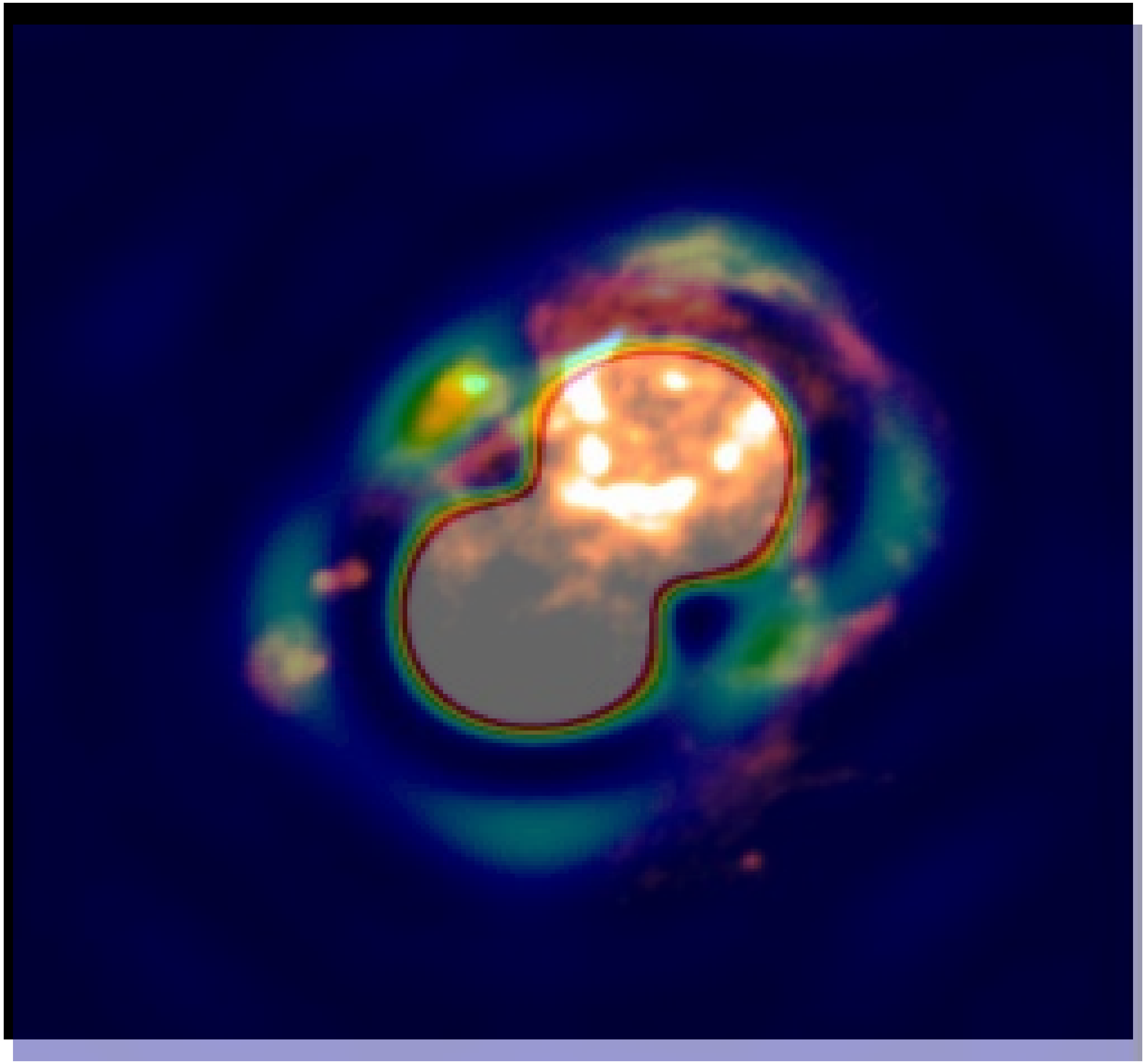


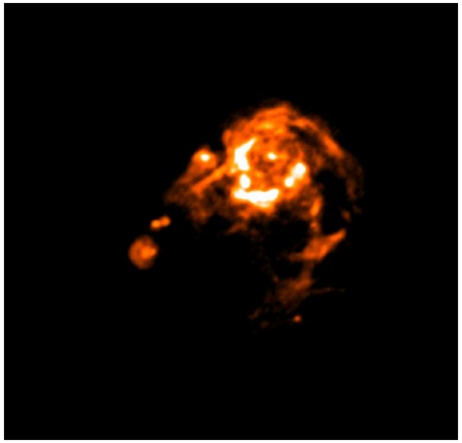
Sources from NVSS. Flux range ~2-200 mJy/beam

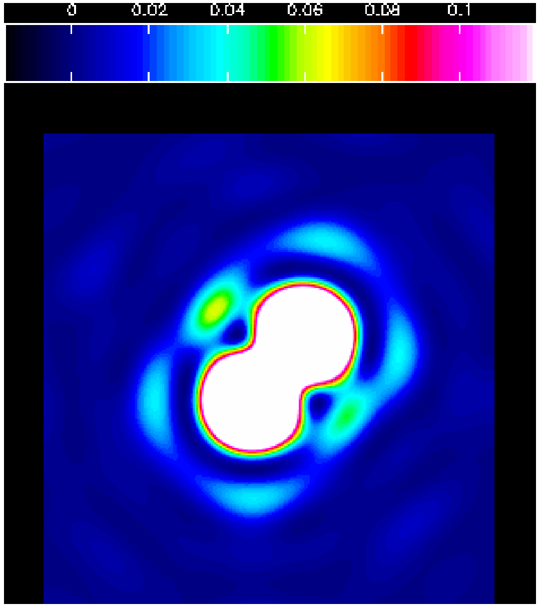
$$\boldsymbol{Minimize : V_{ij}^{O} - E_{ij} * V_{ij}^{M} \quad w.r.t. \quad E_i}$$

(Bhatnagar et al., EVLA Memo 84, 2004)

PB effects in mosaicking: Wide(r) field

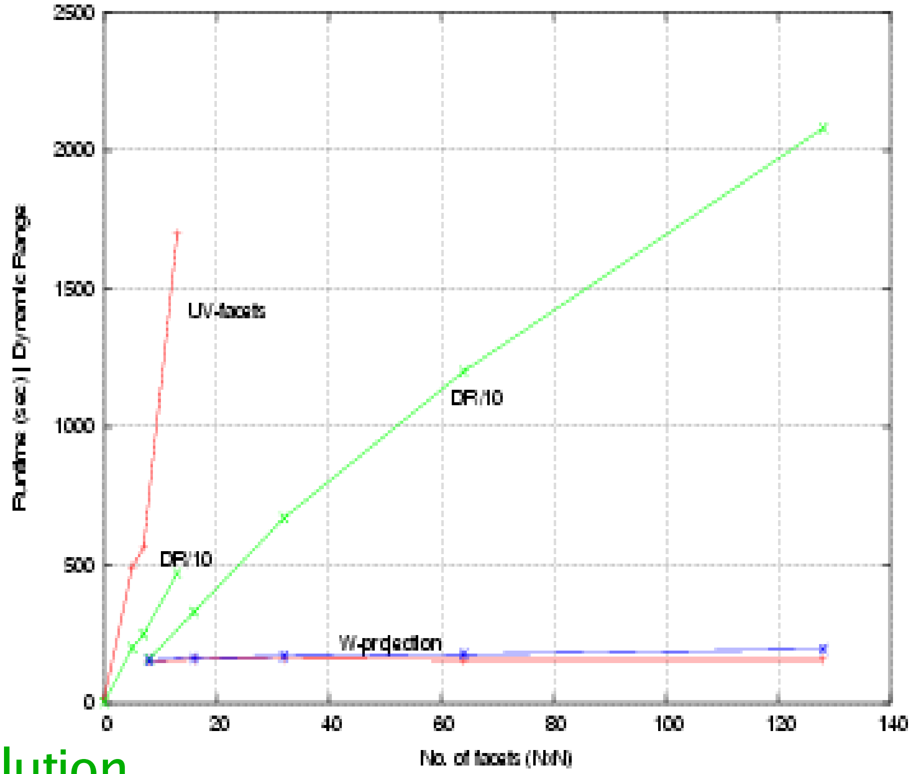






Computing load

- Scaling laws for imaging
  - Non co-planar baseline correction
    - W-Projection: $(N^2_{wproj} + N^2_{GCF})N_{VIS}$
    - Faceting: $N^2_{facets} N^2_{GCF} N_{vis}$



  - Combine with Scale-sensitive deconvolution

- N_{vis}: $10^{10\text{-}12}$, N^2_{GCF}: 7-50 , N_{comp}: $10^{4\text{-}5}$

I/O load

- Near future data volume (1-2 years)
  - Statement about an year ago: "100-1000 GB/12hr by mid-2010"
    - Recent data with the EVLA: 120 GB
- Next 5 years
  - 100X increase (in volume and effective I/O)

- Non-streaming data processing
  - Expect 20-50 passes through the data (flagging + calibration + imaging)
    - Effective data i/o: few TB
  - Exploit data parallelism
    - Distribute normal equations (SPMD paradigm looks promising)
  - Deploy *computationally efficient* algorithms ('P'of SPMD) on a cluster

Computing challenges

- Calibration of direction dependent terms
  - As expensive as imaging
- Significant increase in computing for wide-field wide-band imaging
  - E.g. convolution kernels are larger (up to 50x50 for single facet EVLA A-array, L-band imaging)
  - E.g. Multiple terms for modeling sky and aperture for wide-band widths
- Terabyte Initiative: 4K x 4K x 512 x 1Pol tests using 200 GB data set
  - Timing
    - Simple flagging : 1h
    - Calibration (G-Jones) : 2h15m
    - Calibration (B-Jones) : 2h35m
    - Correction : 2h
    - Imaging : 20h
  - Compute : I/O ratio : 2:3

Parallelization: Initial results

- Spectral line imaging: (8GB RAM per node)
  - Strong scaling with number of nodes & cube size
  - Dominated by data I/O and handling of image cubes in the memory
  - 1024 x 1024 x 1024 imaging
    - Run-time with 1-Node : 50hr
    - Run-time with 16-nodes : 1.5 hr
- Continuum imaging: (No PB-correction or MFS)
  - Requires inter-node I/O (Distribution of normal equations)
  - Dominated by data I/O
  - 1024 x 1024 imaging:
    - 1-node run-time : 9hr
    - 16-node run-time : 70min (can be reduced up to 50%)

Parallelization: Initial results

- Image deconvolution is the most expensive step
- Matching data access and in-memory grid access patterns is critical
- Optimal data access pattern for imaging and calibration are in conflict
  - Freq-Time ordered data optimal for imaging
  - Time-ordered data optimal for calibration
- DD calibration comparable to imaging in computing
- SS deconvolution + MSF might make FLOPS per I/O higher: A good thing!

- Production Cluster
  - 32 nodes, 2x4 cores, 12 GB RAM, InfiniBand inter-connect
  - Data served via a Luster FS
    - Measured I/O throughput: 800-900 MB/s

Conclusions:

- For EVLA and ALMA, the dominant (similar order) DD effects:
  - Time dependence of the PB: rotation with PA, pointing errors
  - Frequency dependence of the PB
  - Frequency dependence of the sky
  - Modeling of extended emission deconvolution (certainly for mosaic imaging)
- Solutions: AW-Projection (PB effects) + MS-/Asp-Clean + MFS (frequency dependence of the sky)
- Additionally –needs to deal with 100GB –1TB datasets
  - Need efficient algorithms which can be deployed on parallel computers
- Work in progress on extensions and generalization of algorithms for SKA-class imaging and calibration problems.
- All of this work is/will be available in the CASA libraries

Conclusions:

- Three fundamentally different approaches being pursued
  - Corrections in the data domain (FFT based transforms)
    - AW-Projection, Pointing SelfCal, Mosaicking,...
    - Challenges: Controlling the propagation of errors
  - Corrections in the image plane (DFT based transforms)
    - Faceting
    - Challenges: Curse of dimensionality, run time efficiency with realistic data volumes
  - Linear Algebra methods (Least-Square All Sky Imaging)
- Hybrid:
  - Image- (or UV-) plane faceting + Projection algorithms
  - Facets on a few strong sources only

References:

- W-Projection: IEEE Journal of Selected Topics in Signal Processing, Vol. 2, No. 5, 2008
- A-Projection: A&A, 487, 419, 2008 (arXiv:0808.0834)
- Pointing SelfCal: EVLA Memo. 84, 2004
- Scale sensitive deconvolution of astronomical images: A&A, 426, 747, 2004 (astro-ph/0407225)
- MS-Clean: IEEE Journal of Selected Topics in Signal Processing, Vol.2, No.5,2008
- Advances in Calibration and Imaging in Radio Interferometry: Proc. IEEE, Vol. 97, No. 8, 2008
- Calibration and Imaging challenges at low frequencies: ASP Conf. Series, Vol. 407, 2009

- Parametrized Deconvolution for Wide-band Radio Synthesis Imaging; PhD Thesis, Rao Venkata; NMT, 2010

- http://www.aoc.nrao.edu/~sbhatnag , http://www.aoc.nrao.edu/~rurvashi
- Home pages of SKA Calibration and Imaging Workshops (CALIM), 2005, 2006, 2008, 2009