Imaging and Deconvolution II

Advanced Topics

Daniel Mitchell
2018 ICRAR/CASS Radio School

CSIRO ASTRONOMY AND SPACE SCIENCE
www.csiro.au

Imaging and Deconvolution II

- How and why do the basic assumptions break down?
- Techniques used to extend imaging and deconvolution
- Major cycles and minor cycles

Calibration/Imaging Basic Assumptions

Calibrated visibilities sample the same 2D FFT of the same sky

$$V_{jk}(t,\nu) = g_j(t,\nu) g_k^*(t,\nu) S_{jk}(t,\nu) \iint I(l,m)\, e^{-i2\pi(ul+vm)} dl dm$$

Gains are antenna-based and independent of direction

Sky is fixed over the course of an observation

2D Fourier transform between sky and gridded visibilities

How and why do the basic assumptions break down?

How do the assumptions break down?

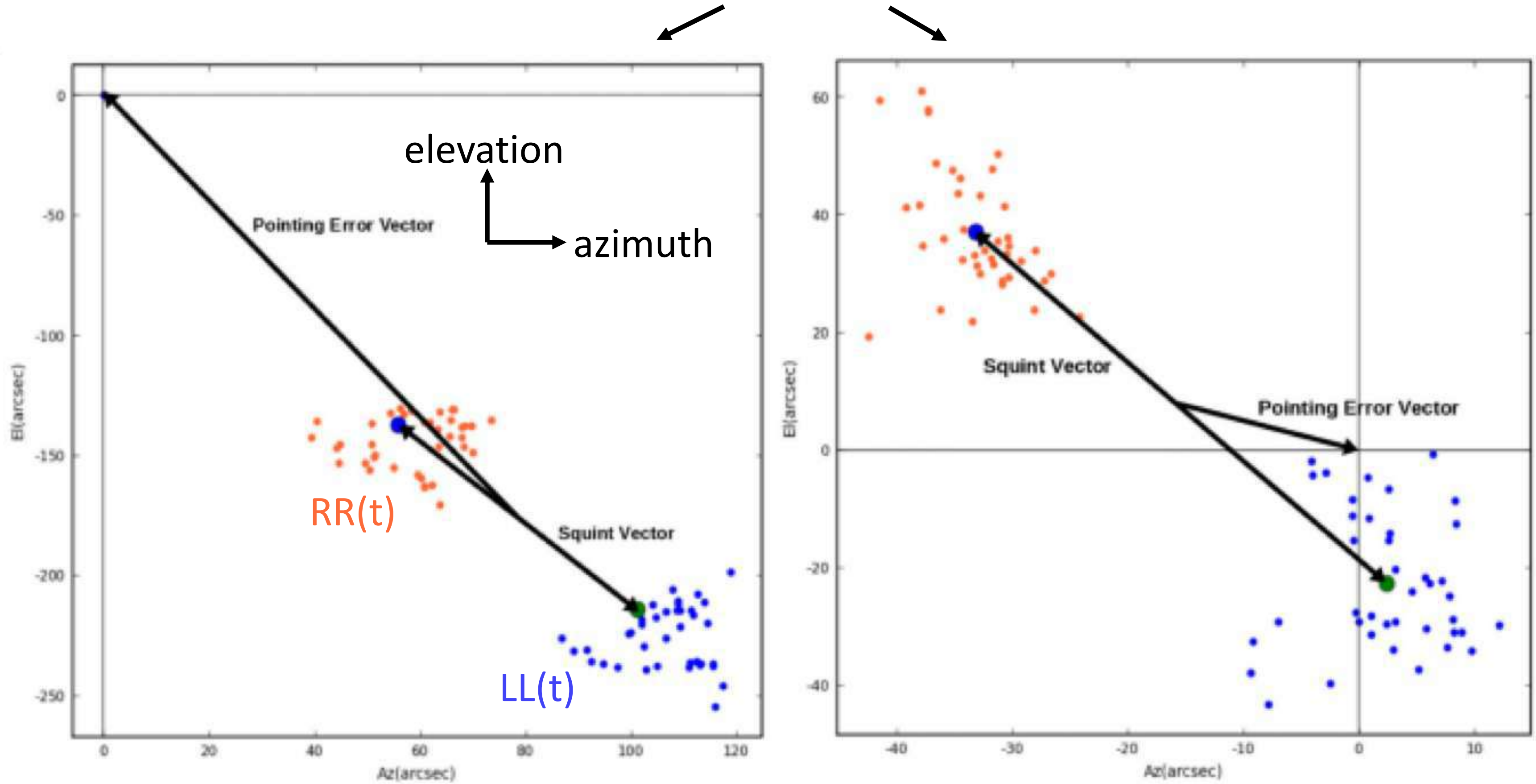
$$V_{jk}(t,\nu) = G_{jk}(t,\nu)S_{jk}(t,\nu) \iint A_{jk}^{inst}(l,m,t,\nu,p)\, I(l,m)\, e^{-i2\pi(ul+vm)} dldm$$

When visibilities have a different response to the sky

- i.e. different primary beams
  - With time (e.g. rotating antenna beams)
  - With frequency (e.g. beam width ∝ λ/D)
  - With antenna (e.g. pointing errors, beam-former variability)
  - With baseline (e.g. decorrelation)
  - With beam (for multi-beam or phase-array feed systems)
  - With polarisation (e.g. different XX and YY beams when forming Stokes I)

JVLA Primary Beam Variability

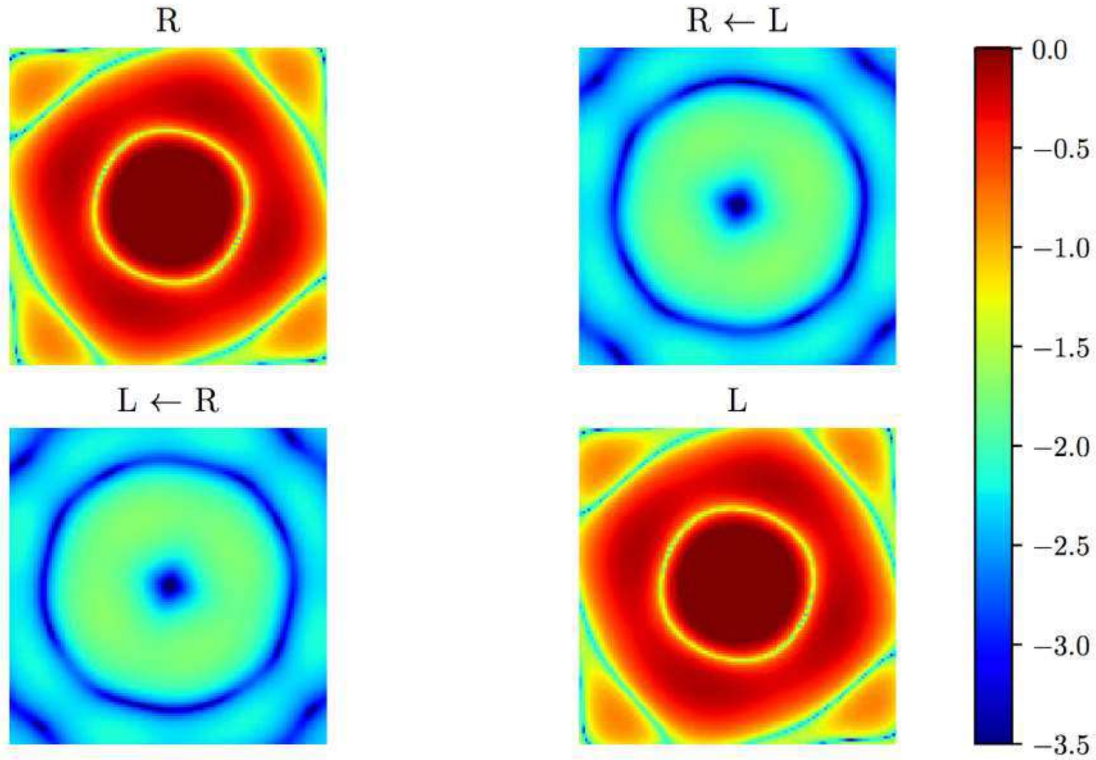
Pointing offsets for two JVLA antennas



Bhatnagar & Cornwell 2017 (arXiv:1808.04516)

JVLA Primary Beam Variability

Azimuthal asymmetries rotate on the sky when tracking a field with alt-az dishes



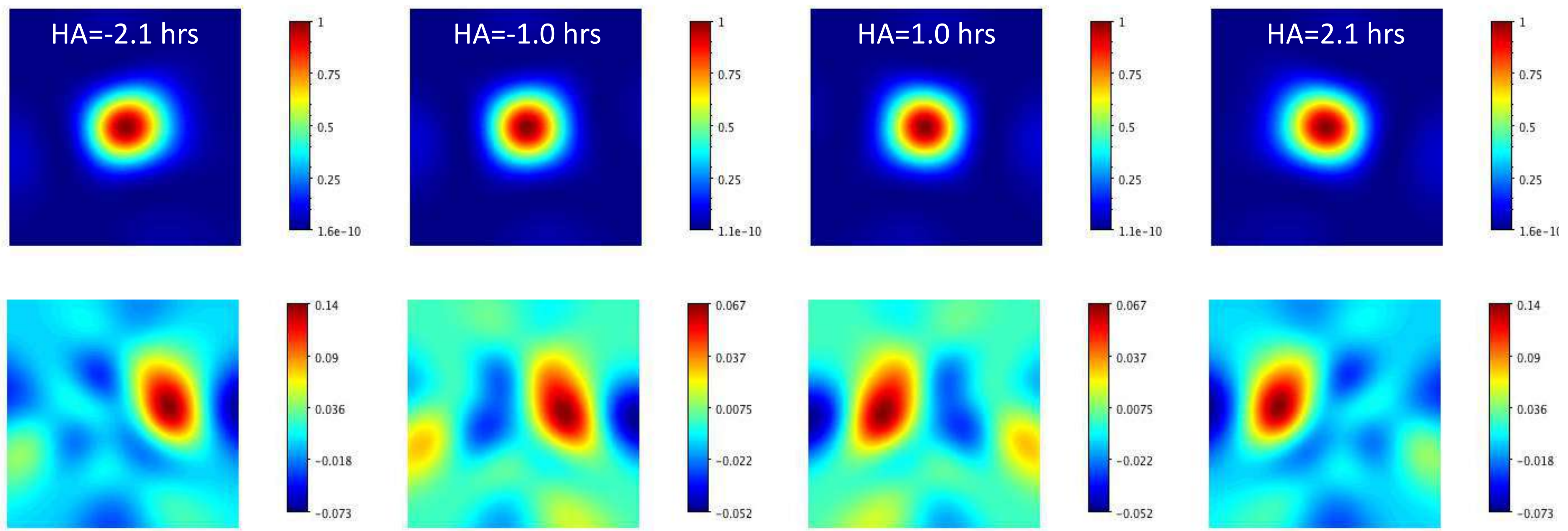
Jagannathan et al. 2017 (arXiv:1706.01501)

CSIRO

MWA Primary Beam Variability

Phased array beams from a tile of dipoles that are fixed to the ground can change shape as they track

Time-dependent tile response

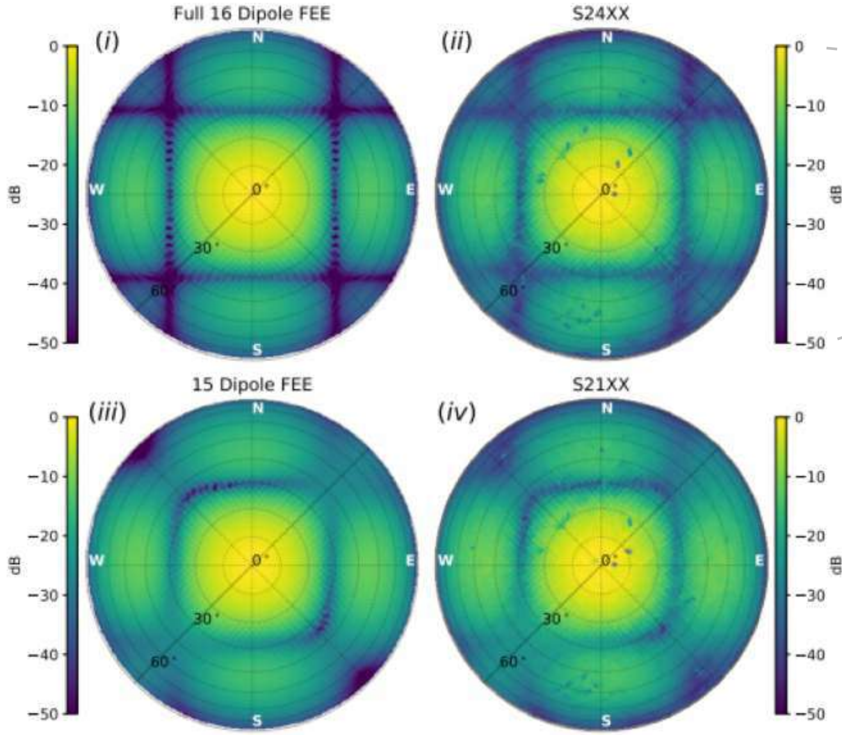


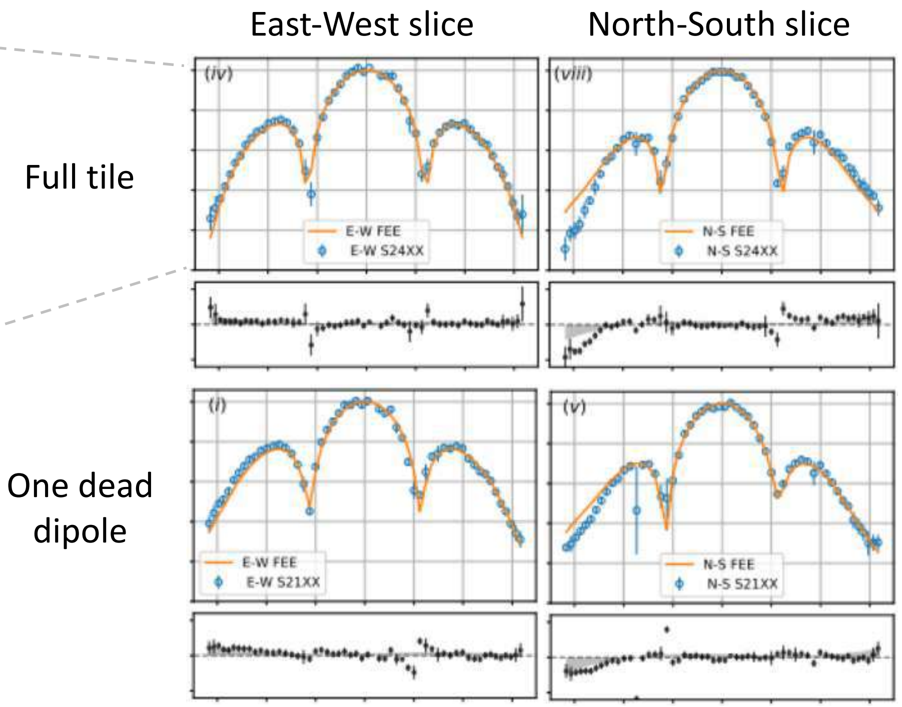
Time-dependent error

Simulated MWA primary beams

MWA Primary Beam Variability

And can change from tile to tile



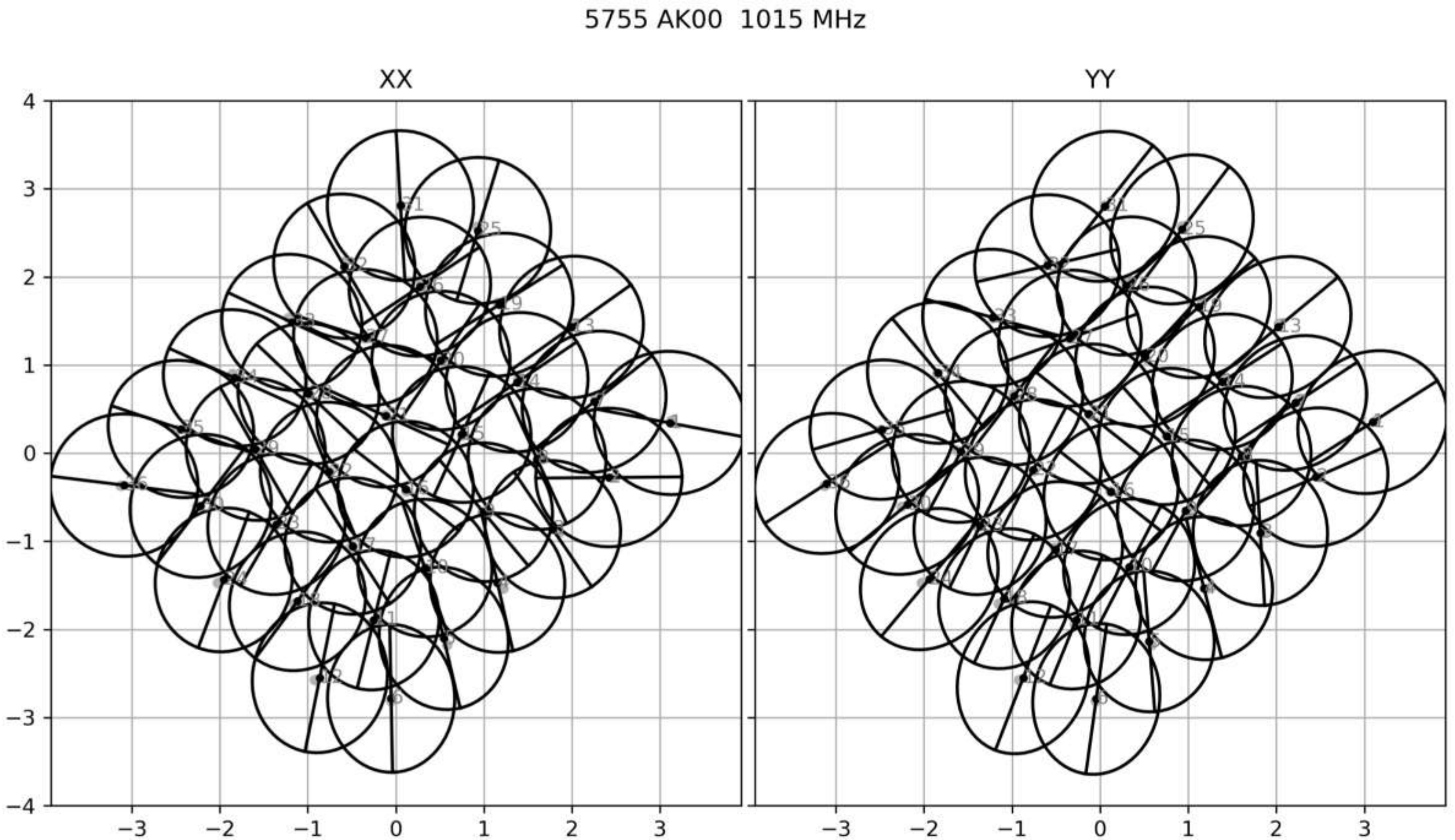


Full Embedded Element primary beam models. Sokolowski et al. 2017 (arXiv:1710.07478)

Measured primary beams

Beam measurements at ~ 138 MHz using ORBCOMM satellites
Line et al. 2018 (arXiv:1808.04516)

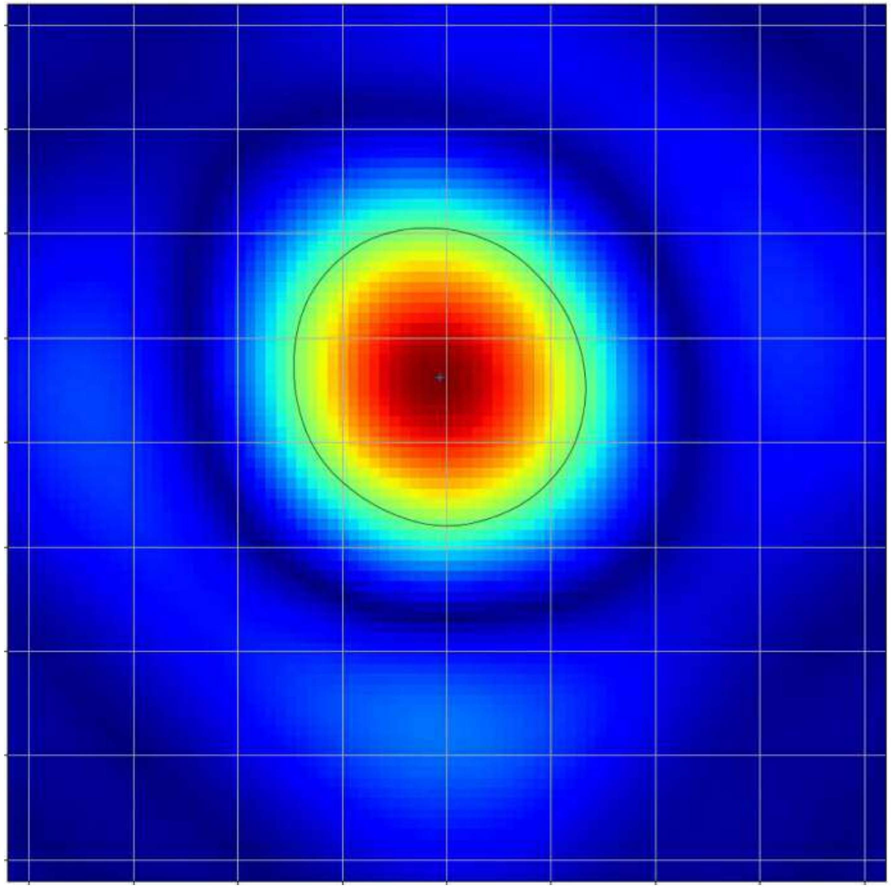
ASKAP Primary Beam Variability



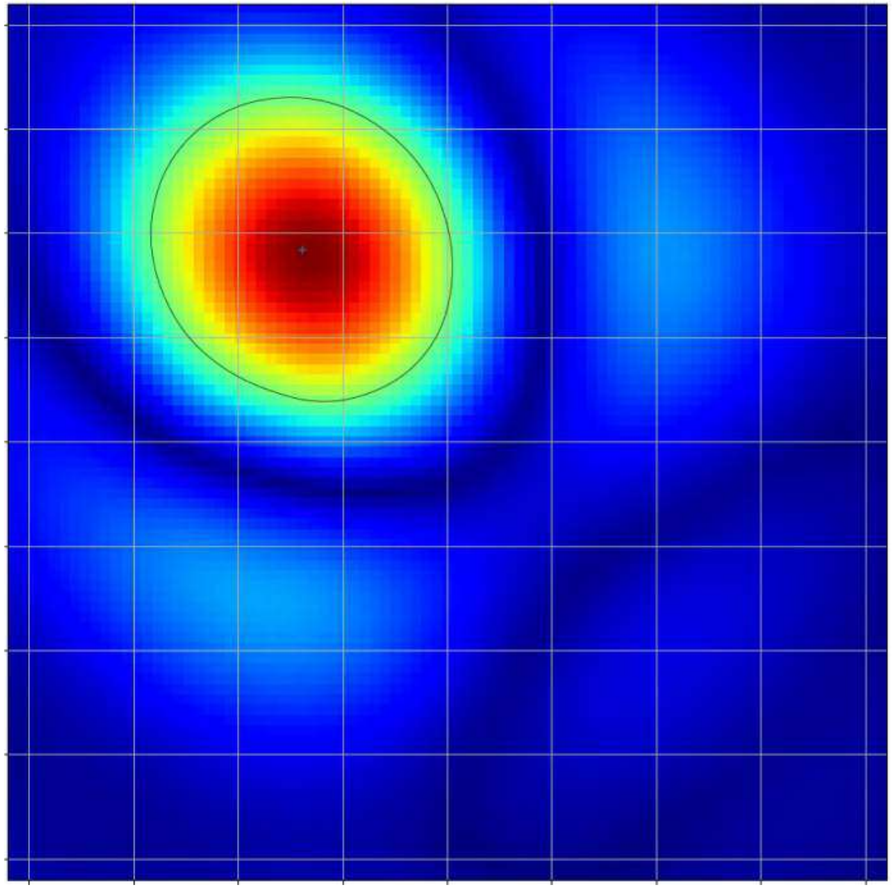
Thanks to Dave Mcconnell

ASKAP Primary Beam Variability

AK05 beam 00 792MHz, XX



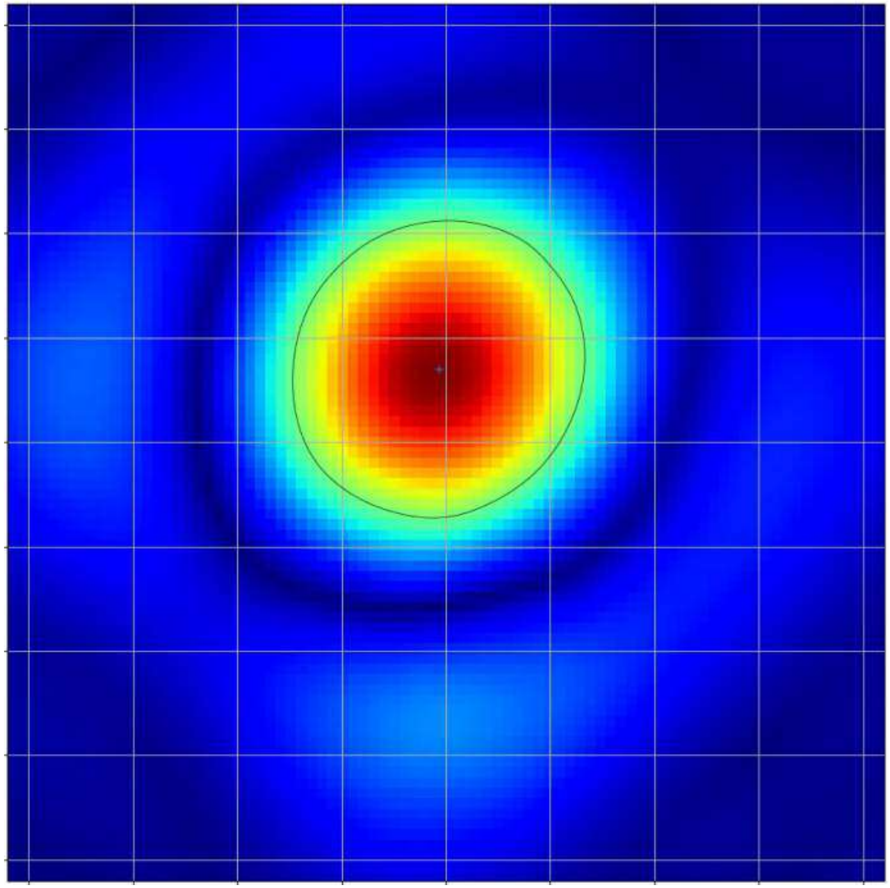
AK05 beam 18 792MHz, XX



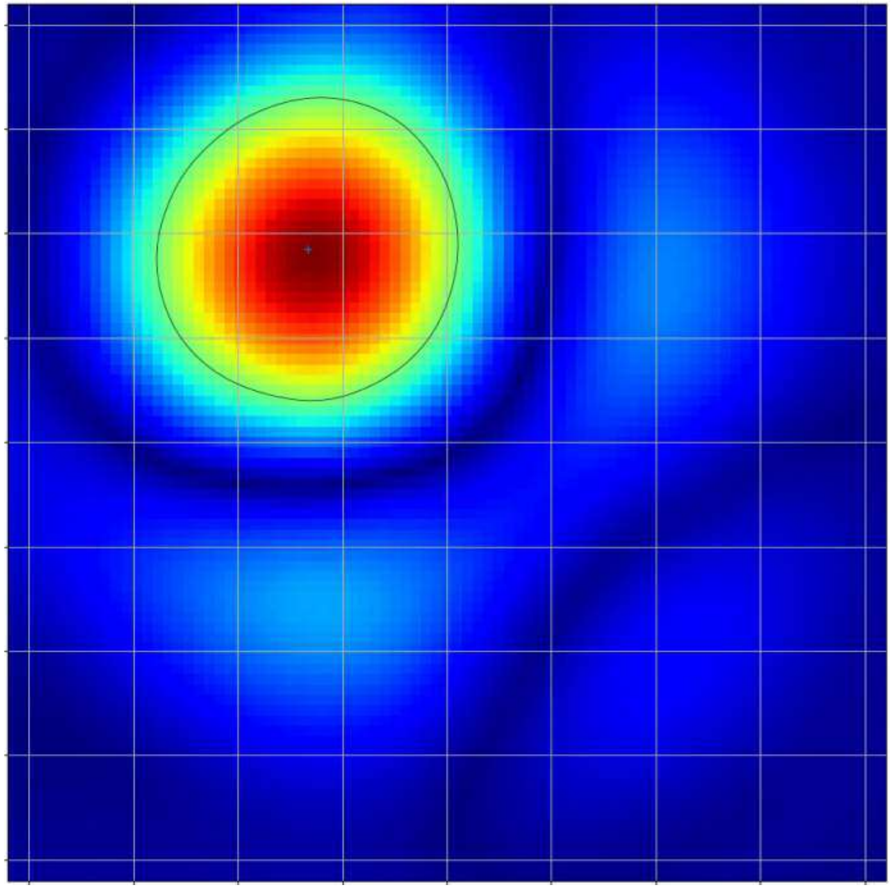
Thanks to Aidan Hotan

ASKAP Primary Beam Variability

AK05 beam 00 792MHz, YY



AK05 beam 18 792MHz, YY



Thanks to Aidan Hotan

How do the assumptions break down?

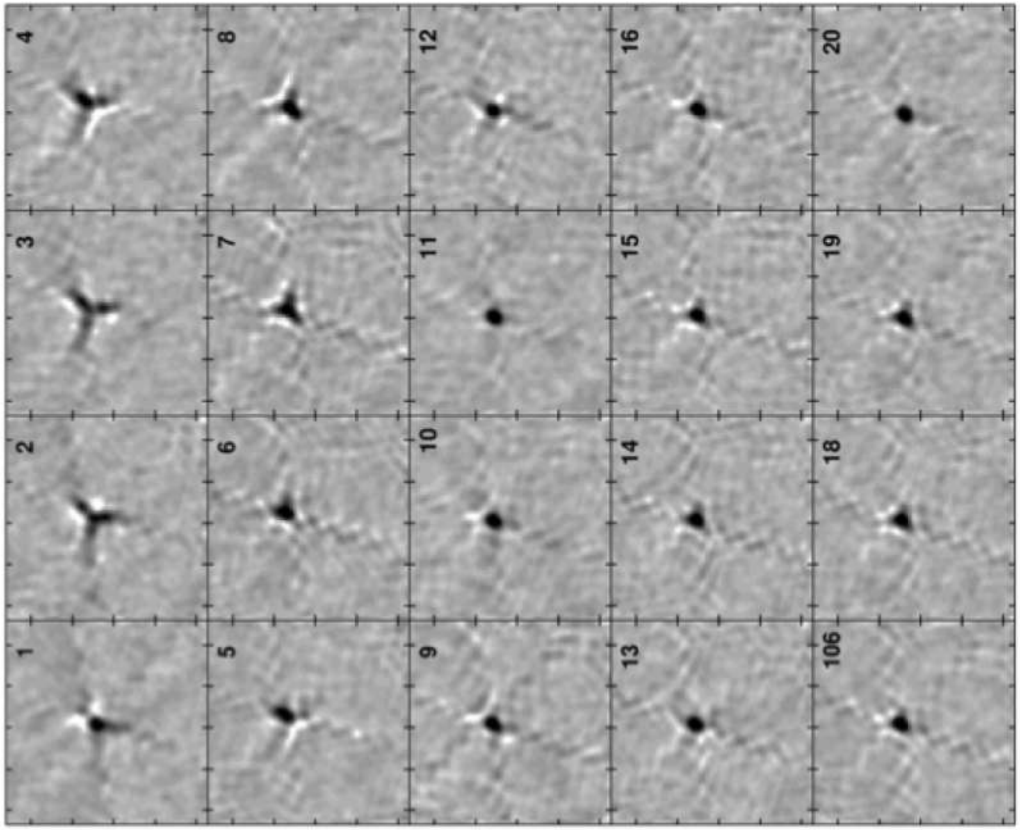
$$V_{jk}(t,\nu) = G_{jk}(t,\nu)S_{jk}(t,\nu)\iint A_{jk}^{atm}(l,m,t,\nu,p)\, I(l,m)\, e^{-i2\pi(ul+vm)}dldm$$

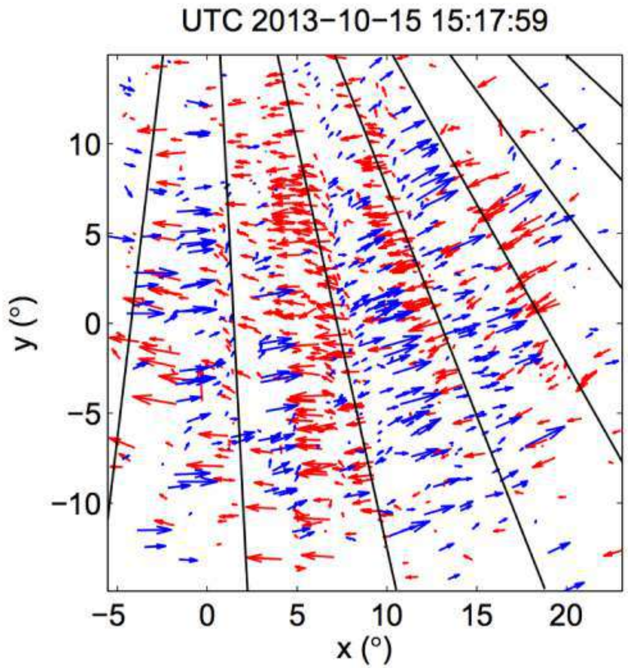
When visibilities see a different atmosphere

- Ionospheric refraction
- Ionospheric Faraday rotation
- Troposphere (at GHz frequencies)

Ionospheric Refraction Variability

Cotton (2004) ASP Conf. Series 345, 74 MHz, 1-min VLA snapshots



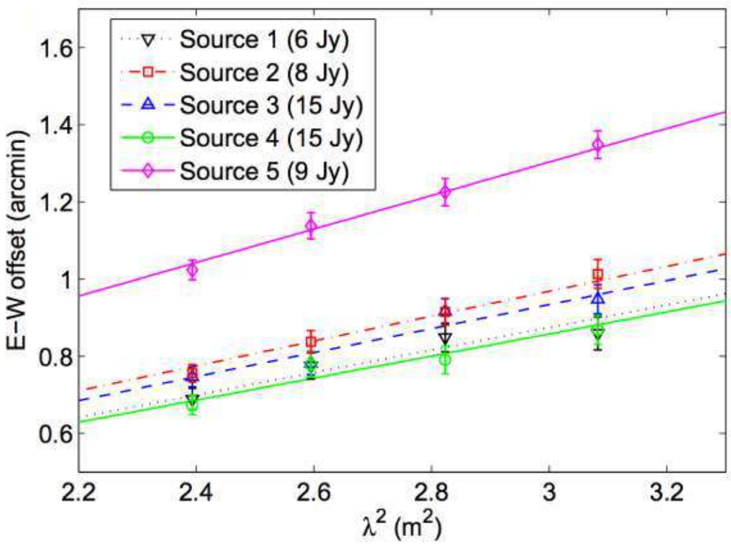


Loi et al. (2015) arXiv:1504.06470

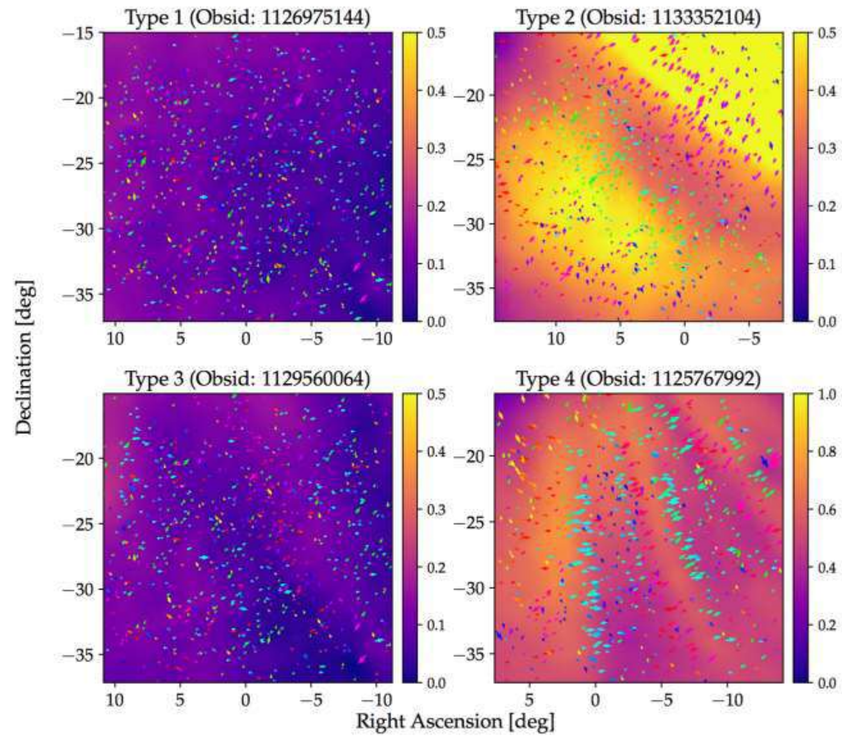
Ionospheric Refraction Variability

- The refraction has a known dependence on frequency.
- For a small array like MWA, the refraction often results in time-dependent position shifts.
- See Natasha's lecture on Wednesday.

Loi et al. 2015 (arXiv:1504.06470)

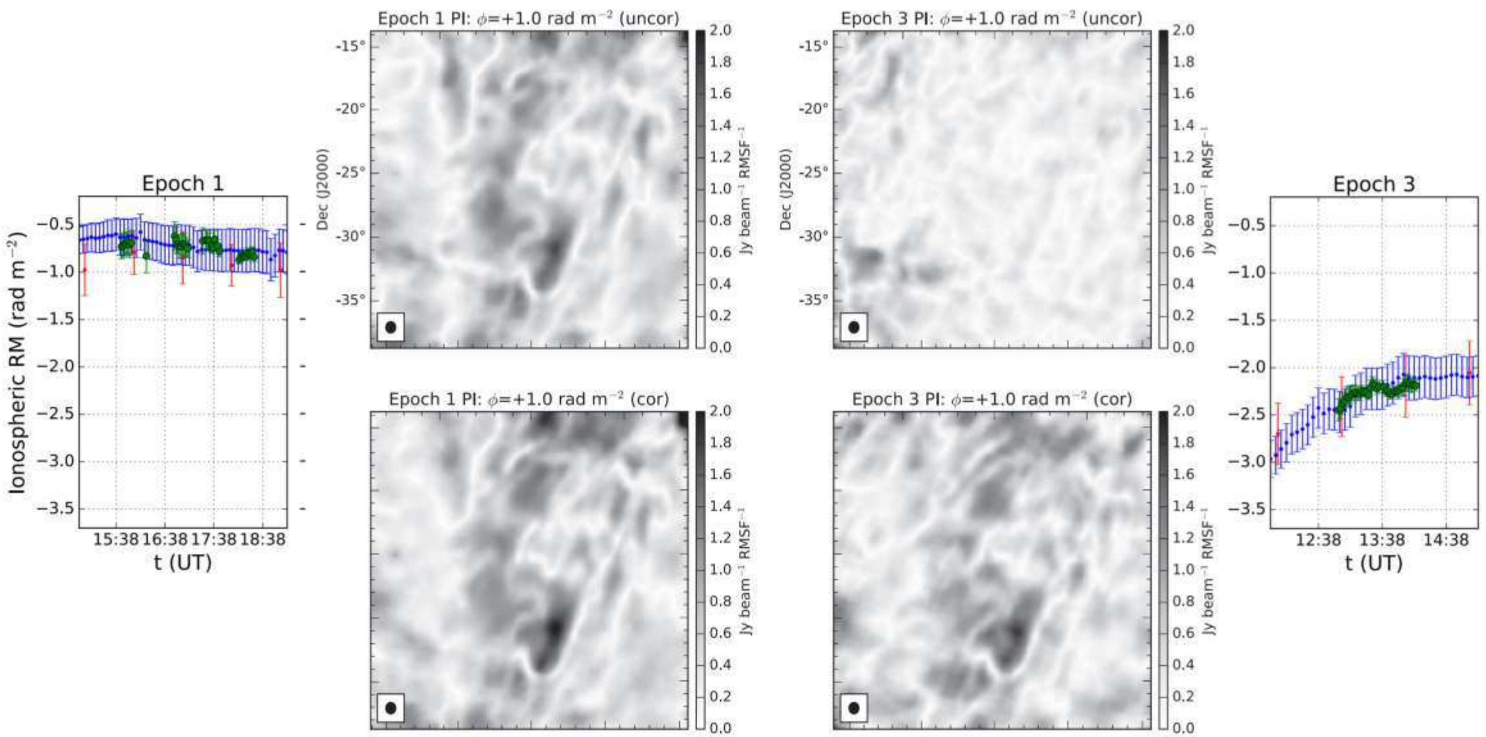


Jordan et al. 2017 (arXiv:1707.04978)



Ionospheric Faraday Rotation Variability

- The angle of linear polarisation generally rotates as a function of frequency.
- Transform to “Rotation Measure” space to average in frequency
- Ionospheric Faraday rotation changes Rotation Measure spectra.
- See Emil’s lecture on Friday.



Lenc et al. 2017 (arXiv:1607.05779)

How do the assumptions break down?

$$V_{jk}(t,\nu) = G_{jk}(t,\nu)S_{jk}(t,\nu)\iint I(l,m,t,\nu)\,e^{-i2\pi(ul+vm)}dldm$$

When visibilities see a different sky

- Radio transients
- Spectra of radio sources
- Faraday rotation of linearly polarised objects
- Near by objects: the sun, planets, etc.

How do the assumptions break down?

$$V_{jk}(t,\nu) = G_{jk}(t,\nu)S_{jk}(t,\nu)\iint I(l,m)\, e^{-i2\pi(ul+vm+w(n-1))}\,dldm$$

When the visibility-sky relationship is not 2D

- w-terms.
- Particularly bad for large fields of view.

Why push the 2D limits?

- To improve sensitivity and/or uv coverage.
  - By averaging in time, frequency, polarisation, etc.
- To carry out large surveys.
  - Large FoV to increase survey speed.
  - Large FoV to increase instantaneous footprint.
- To achieve high dynamic range at low frequencies.
  - The low-frequency sky is crowded → need to image the full FoV.
  - Many smaller antennas for ionospheric calibration → large FoV.
- To get to most out of your instrument.

Techniques used to extend imaging and deconvolution

Convolutional Gridding

- To use a FFT, need to interpolate visibilities onto the uv grid.
- Use a convolution kernel with a desired image-domain response
  - recall: $\mathfrak{F}(F \times G) = \mathfrak{F}(F) \ast \mathfrak{F}(G)$
  - e.g. $\mathfrak{F}($ prolate spheroidal window fn $)$ = prolate spheroidal gridding kernel
  - can be different for each visibility

Convolutional Gridding

- To use a FFT, need to interpolate visibilities onto the uv grid.
- Use a convolution kernel with a desired image-domain response
  - recall: $\mathfrak{F}(F \times G) = \mathfrak{F}(F) * \mathfrak{F}(G)$
  - e.g. $\mathfrak{F}$(prolate spheroidal window fn) = prolate spheroidal gridding kernel
  - can be different for each visibility

visibility

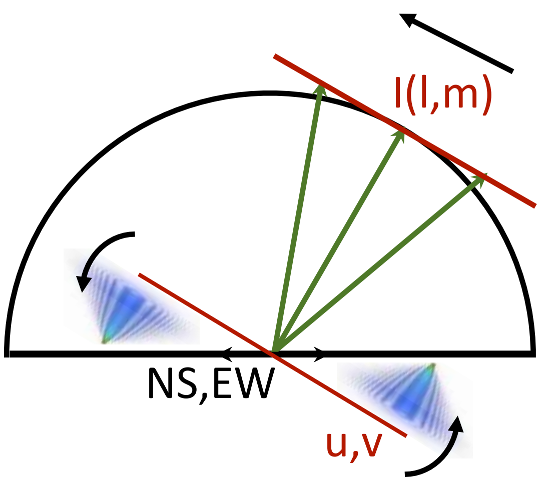
The sky in the field of view

spheroidal window function to limit aliasing from outside the field of view

spheroidal gridding kernel

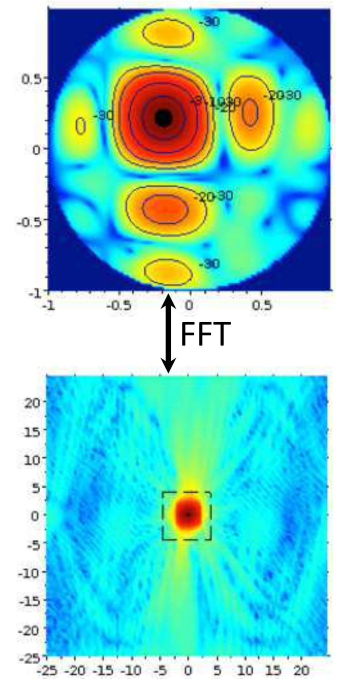
Convolutional Gridding

- To use a FFT, need to interpolate visibilities onto the uv grid.
- Use a convolution kernel with a desired image-domain response
  - recall: $\mathfrak{F}(F \times G) = \mathfrak{F}(F) * \mathfrak{F}(G)$
  - e.g. $\mathfrak{F}($ prolate spheroidal window fn $)$ = prolate spheroidal gridding kernel
  - can be different for each visibility
- W-Projection
  - Convolutional gridding with $\mathfrak{F}\left(e^{i2\pi w_{jk}\left(\sqrt{1-l^2-m^2}-1\right)} \right)$
  - Fresnel propagation to a common plane
  - Kernels become very large for wide field of view ($\propto \theta^4$)
    - Expensive to generate → cache ($\propto \theta^6$)
    - Algorithms can become limited by memory or memory-bandwidth.



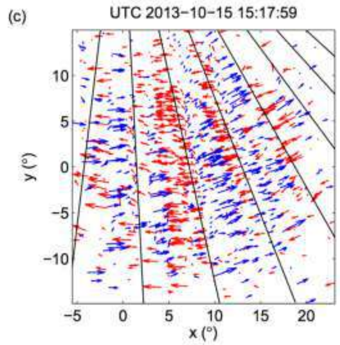
Convolutional Gridding

- To use a FFT, need to interpolate visibilities onto the uv grid.
- Use a convolution kernel with a desired image-domain response
  - recall: $\mathfrak{F}(F \times G) = \mathfrak{F}(F) * \mathfrak{F}(G)$
  - e.g. $\mathfrak{F}($ prolate spheroidal window fn $)$ = prolate spheroidal gridding kernel
  - can be different for each visibility
- A-Projection
  - Convolutional gridding with $\mathfrak{F}(A_{jk}^{inst*}(l, m, t, \nu, p))$
  - Cancel direction-dependent phases, square amplitudes.
  - Optimal weighting for mosaicking, snapshot stacking, etc.
  - Can grid multiple beams or mosaic pointings on a single uv grid
    - recall: $\mathfrak{F}(A(l+dl,m+dm)) = \mathfrak{F}(A(l,m)) \exp(-i2\pi(u.dl+v.dm))$
    - exacerbates W-Projection issues (increased field of view)



Convolutional Gridding

- To use a FFT, need to interpolate visibilities onto the uv grid.
- Use a convolution kernel with a desired image-domain response
  - recall: $\mathfrak{F}(F \times G) = \mathfrak{F}(F) * \mathfrak{F}(G)$
  - e.g. $\mathfrak{F}$(prolate spheroidal window fn) = prolate spheroidal gridding kernel
  - can be different for each visibility
- I-Projection
  - Convolutional gridding with $\mathfrak{F}(A_{jk}^{iono*}(l, m, t, \nu, p))$
  - Like A-Projection, but correcting for ionospheric phase shifts.
  - Very fast update rate (≈ 10 sec)
  - Not widely used, but one of only a few options for large arrays.



Loi et al. (2015)
arXiv:1504.06470

Convolutional Gridding

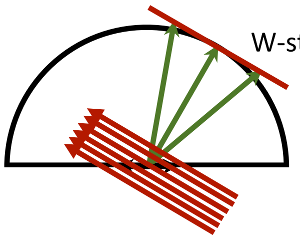
- To use a FFT, need to interpolate visibilities onto the uv grid.
- Use a convolution kernel with a desired image-domain response
  - recall: $\mathfrak{F}(F \times G) = \mathfrak{F}(F) * \mathfrak{F}(G)$
  - e.g. $\mathfrak{F}($ prolate spheroidal window fn $)$ = prolate spheroidal gridding kernel
  - can be different for each visibility
- $\mathfrak{F}(A \times B \times C \times \cdots) = \mathfrak{F}(A) * \mathfrak{F}(B) * \mathfrak{F}(C) * \mathfrak{F}(\cdots)$
  - e.g. convolutional gridding with $\mathfrak{F}(A_{jk}^{inst*} \times A_{jk}^{iono*} \times e^{i2\pi w_{jk}\left(\sqrt{1-l^2-m^2}-1\right)})$

Convolutional Gridding

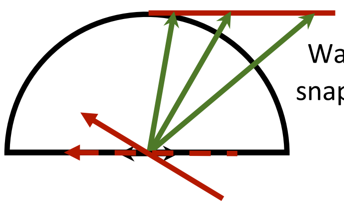
- To use a FFT, need to interpolate visibilities onto the uv grid.
- Use a convolution kernel with a desired image-domain response
  - recall: $\mathfrak{F}(\ F \times G\ ) = \mathfrak{F}(\ F\ ) * \mathfrak{F}(\ G\ )$
  - e.g. $\mathfrak{F}($ prolate spheroidal window fn $)$ = prolate spheroidal gridding kernel
  - can be different for each visibility
- Image-domain gridding: transform small regions of the uv plane back to the image plane and apply the convolutions as multiplications
- e.g. $\mathfrak{F}(\ A_{jk}^{*} \times e^{i2\pi w_{jk}\left(\sqrt{1-l^2-m^2}-1\right)} \times e^{i2\pi(ul+vm)} \times V_{jk}))$

Visibility Segmentation

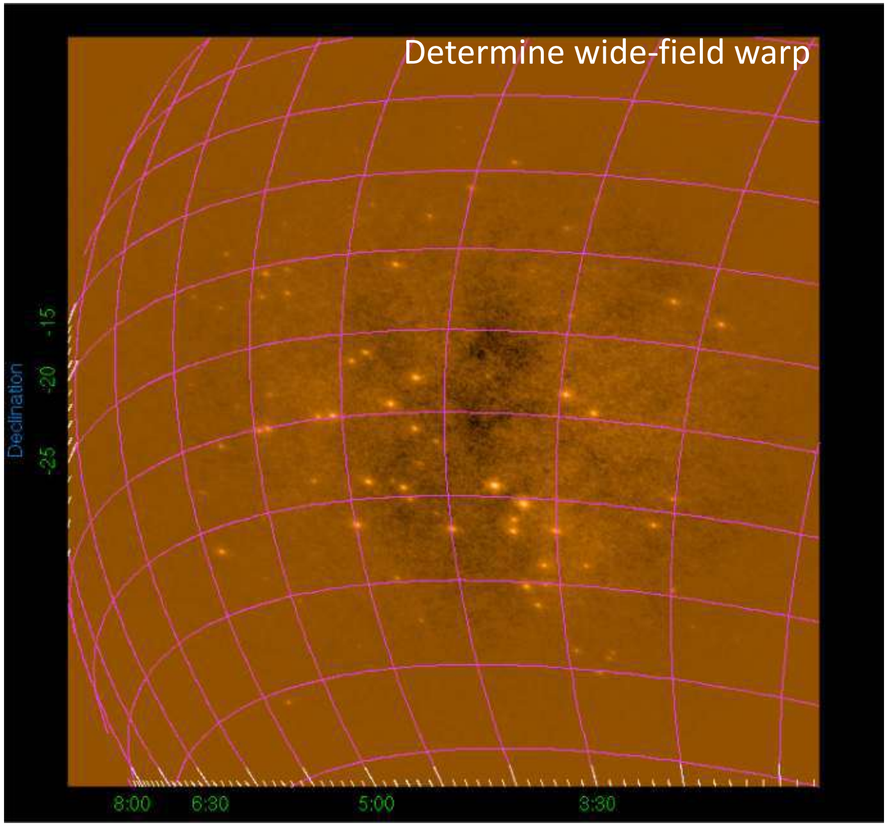
- Split the visibilities into multiple 2D transforms and stack images
  - Reduce w-term effects
- W-Stacking
  - Grid visibilities to their nearest "w-plane"
  - FFT each separately
  - Multiply images by $e^{-i2\pi w\left(\sqrt{1-l^2-m^2}-1\right)}$ and stack

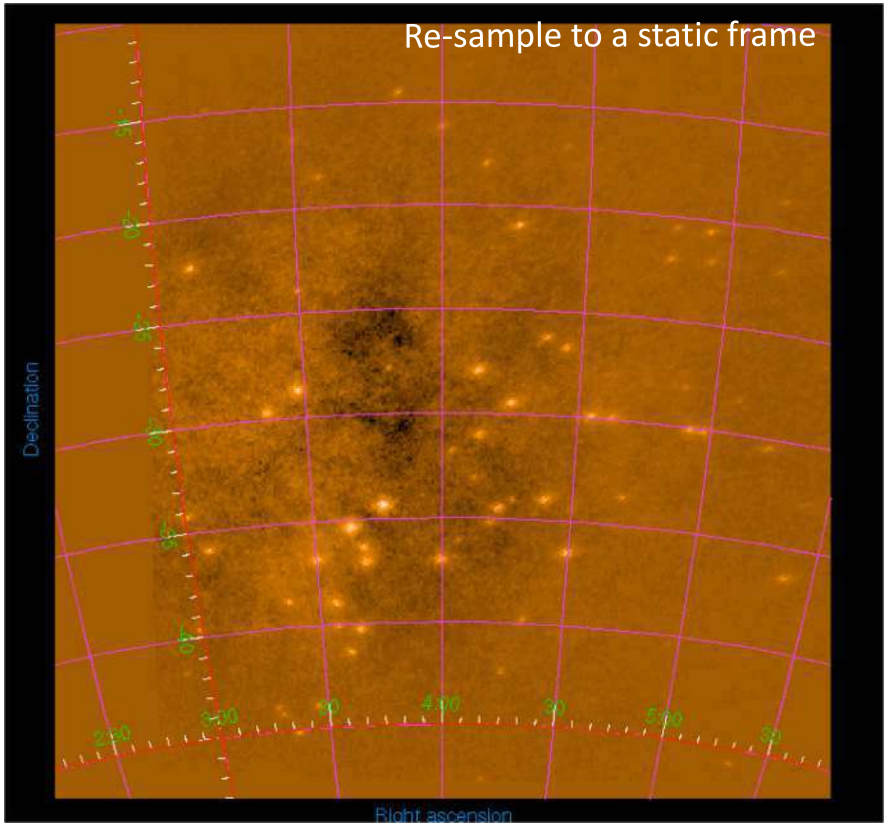


- Warped Snapshots
  - Grid visibilities to a best-fit 2D plane for short snapshots
  - FFT each separately
  - Regrid images to a common frame and stack
  - Many A^{inst} and A^{iono} terms are approx. constant for a given snapshot
    - Can be applied to the snapshot images, rather than during gridding.



Visibility Segmentation — Snapshots





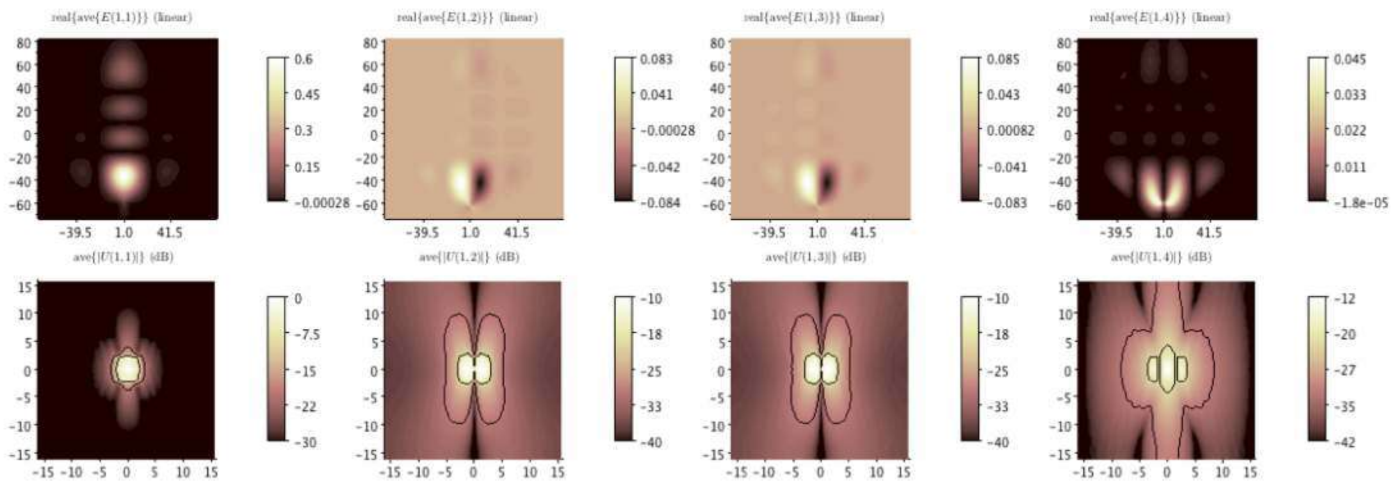
Simulated data: field centre: -3.5 to +3.5 hrs

Visibility Segmentation — Snapshots

NN←xx NN←xy NN←yx NN←yy

All-sky polarised tile response

Fourier response



Visibility Segmentation — Snapshots

	NN←xx	NN←xy	NN←yx	NN←yy
All-sky polarised tile response				
Fourier response				
Deal with curved sky in the image domain				
Remaining Instrument Fourier response				

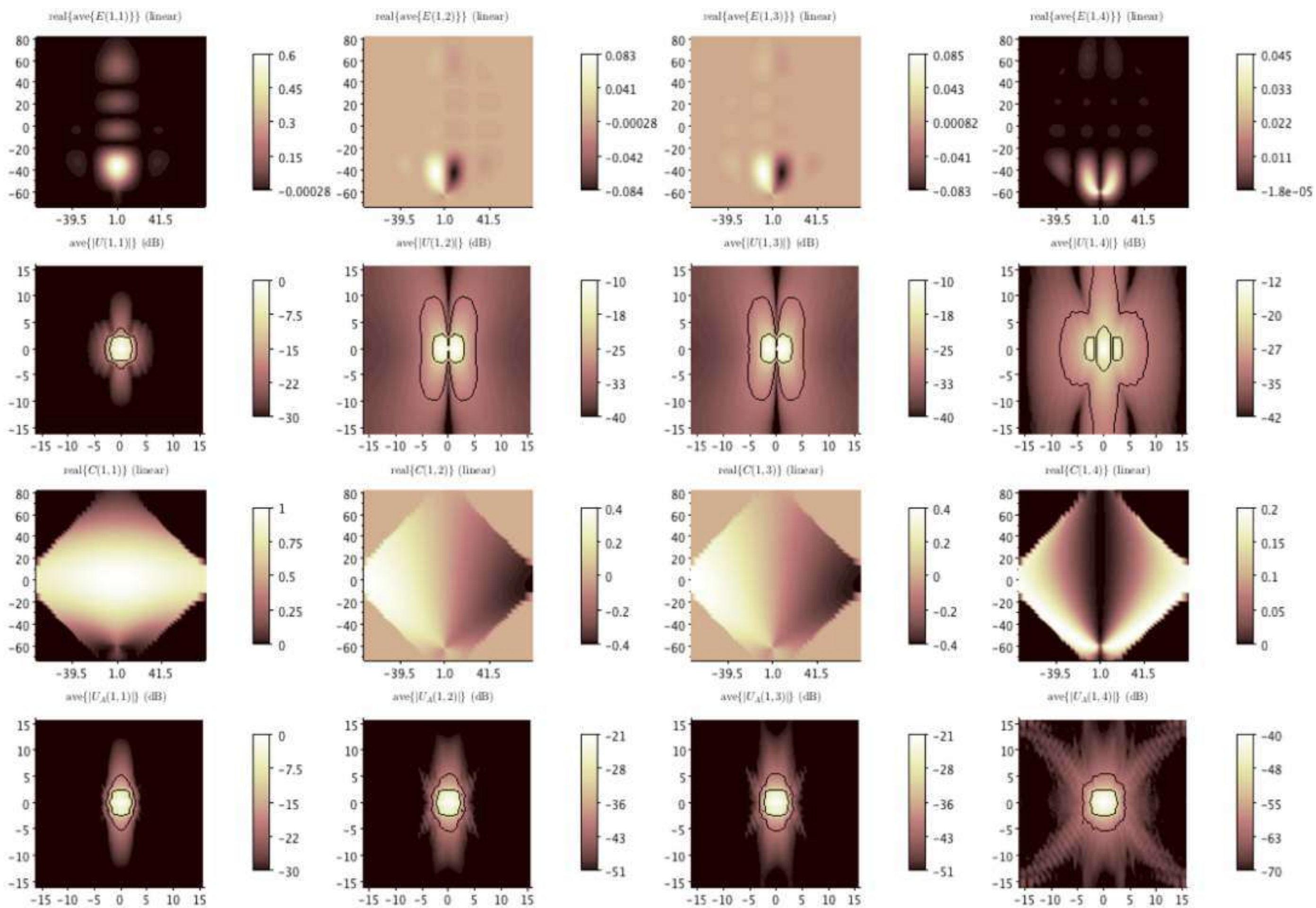
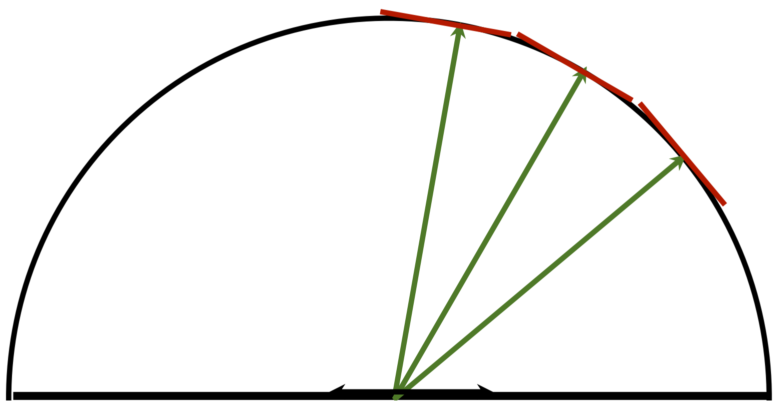


Image Facets

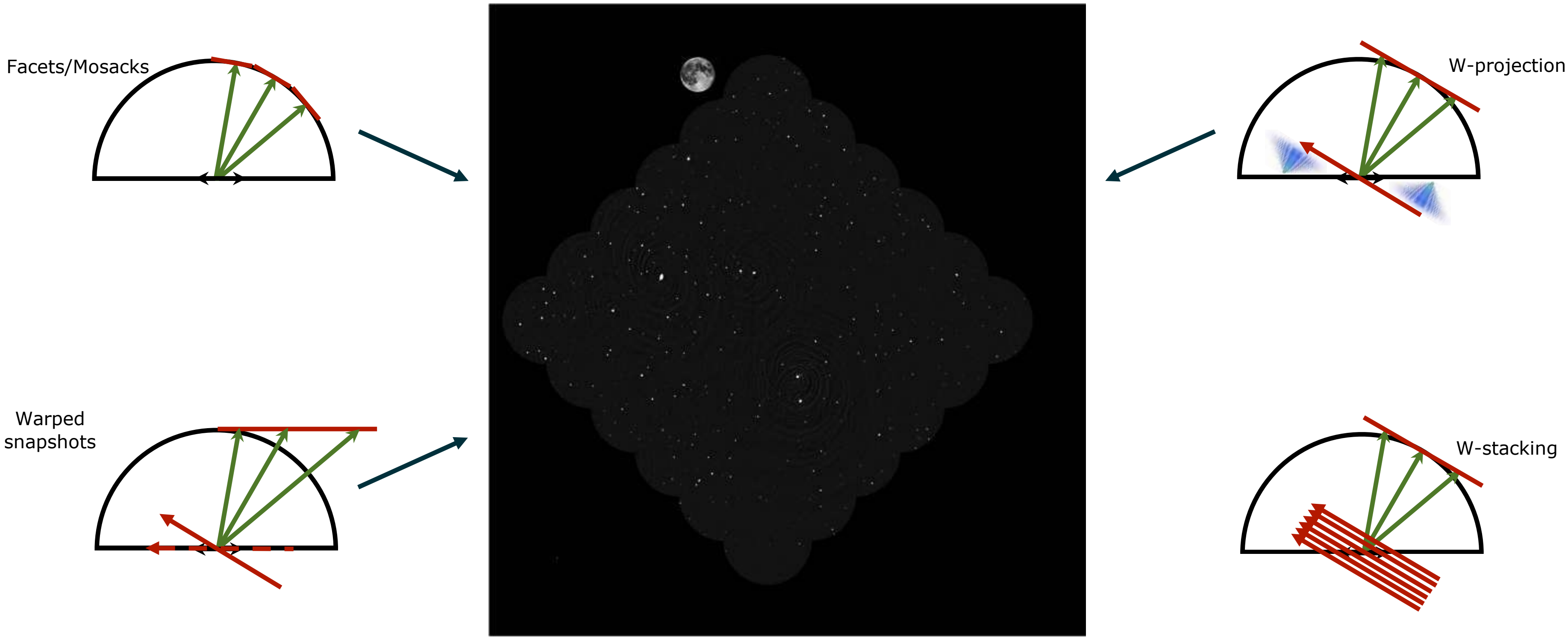
- Split the sky into smaller segments.



- Wide-field calibration and imaging factors can be applied to each separately.

Combined Approaches

Most packages support multiple approaches that can be / have been blended together to meet computing and/or scientific needs.



ASKAPsoft example

Wide-Band Imaging

- Increase image sensitivity: reduces as $1/\sqrt{B}$
- Increases uv coverage
- Better angular resolution at higher frequencies

- But cannot simply average visibilities: decorrelation and smearing!

- And the sky and the instrument change with frequency
  - Although many cosmic sources vary smoothly

Multi-Frequency

Decompose model image into spectral Taylor terms:

$$I_\nu^{mdl} = \sum_{term\, t}^{N_t} \left(\frac{\nu \; - \; \nu_0}{\nu_0}\right)^t I_t^{cc}$$

If $I_\nu = I_{\nu_0} \left(\frac{\nu}{\nu_0}\right)^{I_\alpha + I_\beta \log\left(\frac{\nu}{\nu_0}\right)}$, then approximately:

- $I_0^{cc} = I_{\nu_0}$
- $I_1^{cc} = I_\alpha \; I_{\nu_0}$
- $I_2^{cc} = \left(\frac{I_\alpha \left(I_\alpha - 1\right)}{2} - I_\beta \right) I_{\nu_0}$

Multi-Scale

Decompose model into multiple spatial scales:

$$I^{mdl} = \sum_{scale\ s}^{N_s} I_s^{shp} * I_s^{cc}$$

Spatial Basis (CASA, ASKAPsoft):

- tapered, truncated parabolas with widths proportional to s.
- prolate spheroidal wavefunction used for truncation

MS-MFS (or M&M)

Decompose model image into terms and scales:

$$I_\nu^{mdl} = \sum_{t}^{N_t} \sum_{s}^{N_s} \left(\frac{\nu - \nu_0}{\nu_0} \right)^t \left[I_s^{shp} * I_{s,t}^{cc} \right]$$

- Estimate the sky via joint deconvolution of the brightness at a reference frequency and Taylor terms at N_s spatial scales.
- Need to avoid CLEANing too deep!
  - In general neither scales nor terms are orthogonal
  - Can be a highly coupled system, often starting with a very poor model.

Rau & Cornwell 2011 (arXiv:1106.2745)

ASKAPsoft M&M: BasisFunctionMFS

- Based on the CASA algorithm.
- Jointly deconvolve MFS, separately deconvolve MS
- Normal equations for 2nd order expansion:

coupling matrix : $C_{t_1,t_2} = I^{psf}_{t_1,t_2}(0,0)$

- $$\begin{bmatrix} C_{0,0} & C_{0,1} & C_{0,2} \\ C_{1,0} & C_{1,1} & C_{1,2} \\ C_{2,0} & C_{2,1} & C_{2,2} \end{bmatrix} \begin{bmatrix} I_0^{sky} \\ I_1^{sky} \\ I_2^{sky} \end{bmatrix} = \begin{bmatrix} I_0^{dirty} \\ I_1^{dirty} \\ I_2^{dirty} \end{bmatrix}$$

$$I^{psf}_{t_1,t_2} = I_s^{shp} * \left\{ \sum_\nu \left(\frac{d\nu}{\nu_0}\right)^{t_1+t_2} I_\nu^{psf} \right\} * I_s^{shp}$$

$$I_t^{dirty} = I_s^{shp} * \left\{ \sum_\nu \left(\frac{d\nu}{\nu_0}\right)^t I_\nu^{dirty} \right\}$$

- Choose peak for this scale (max$\left(I_0^{dirty}\right)$ or other) and normalise
- Select dominant scale and for all t: CLEAN I_t^{dirty} and update I_t^{cc}

Major cycles and minor cycles

Image-Based Deconvolution Limitations

- Pixelisation and gridding effects
  - Need infinite CCs to CLEAN an unresolved source that lies between pixels
  - Out-of-field sidelobes from in-field sources alias back into the image
  - Convolutional gridding is not a pure interpolation — also smoothing
- Limited accuracy of PSF sidelobes
  - Due to wide-field effects and the imaging approach
  - Due to computational limitations (limited oversampling, w-planes, etc.)
- Coupling of MSMFS scales and terms

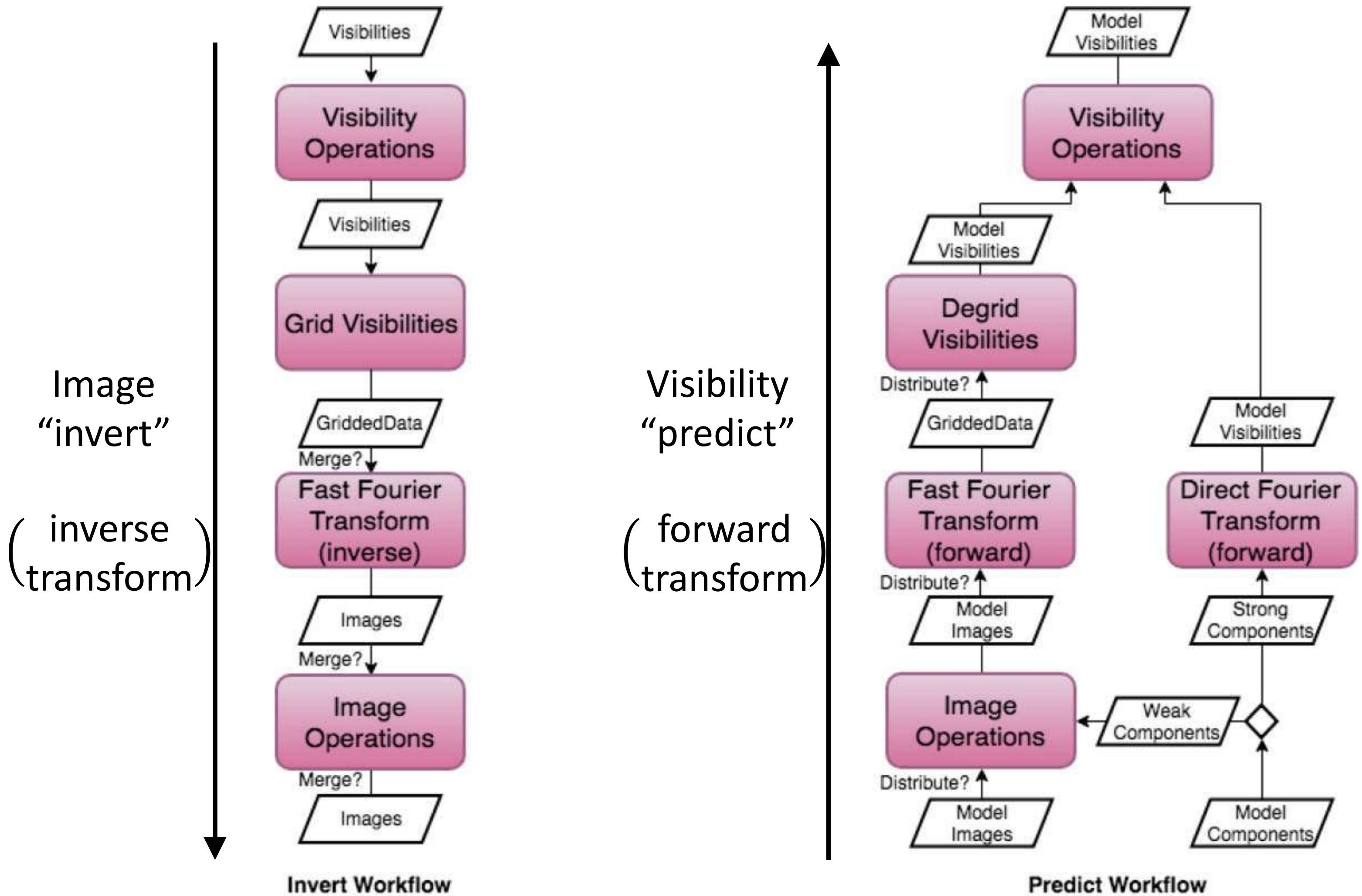
Image-Based Deconvolution Limitations

Common solution:

Image residual visibilities and iteratively improve sky model

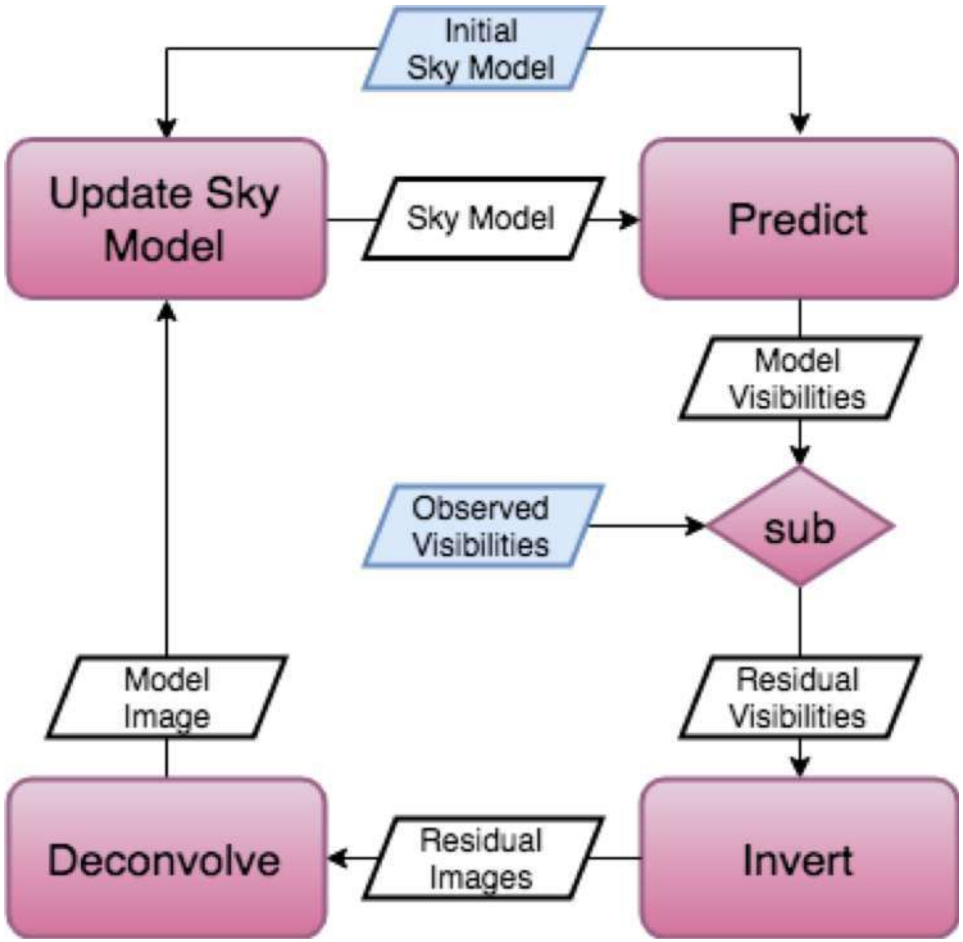
- e.g. Cotton-Schwab CLEAN:
  1. Do a shallow CLEAN
  2. FFT CLEAN component image to the uv domain
  3. "De-grid" to form model visibilities and subtract from observed visibilities
  4. Form residual image and return to 1 (unless converged)
- Visibility subtraction is accurate and avoids many of the limits
- Use a source finder on restored images to build model images?
  - Accurate centroiding.
  - Include residual flux in sky model.

Imaging and Predicting



Major Cycle

- Just pass residual CLEAN components?
- Accumulate CLEAN components?
- Run source finder on restored images?



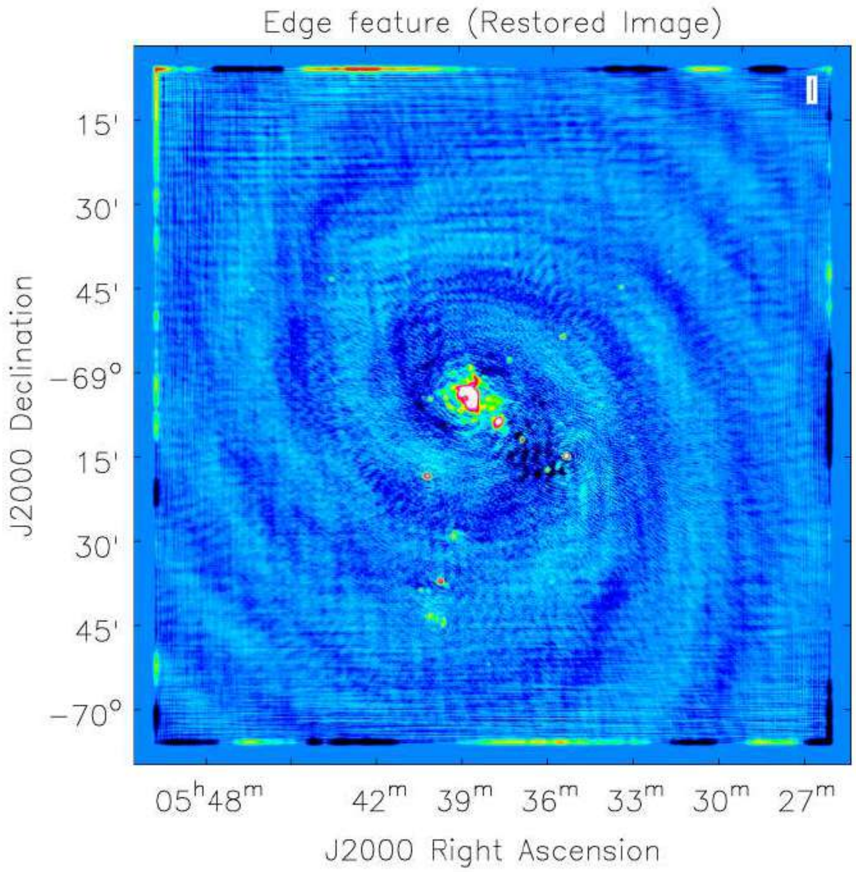
High-accuracy generation of model visibilities

Subtract model visibilities (from observed or previous residual set)

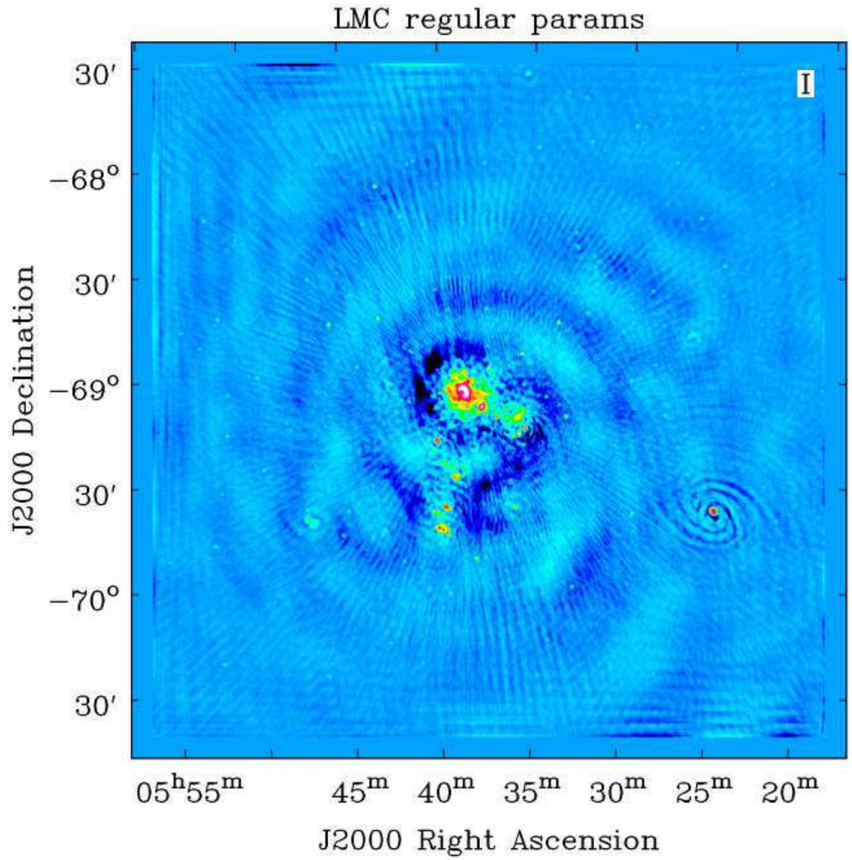
Image residual visibilities

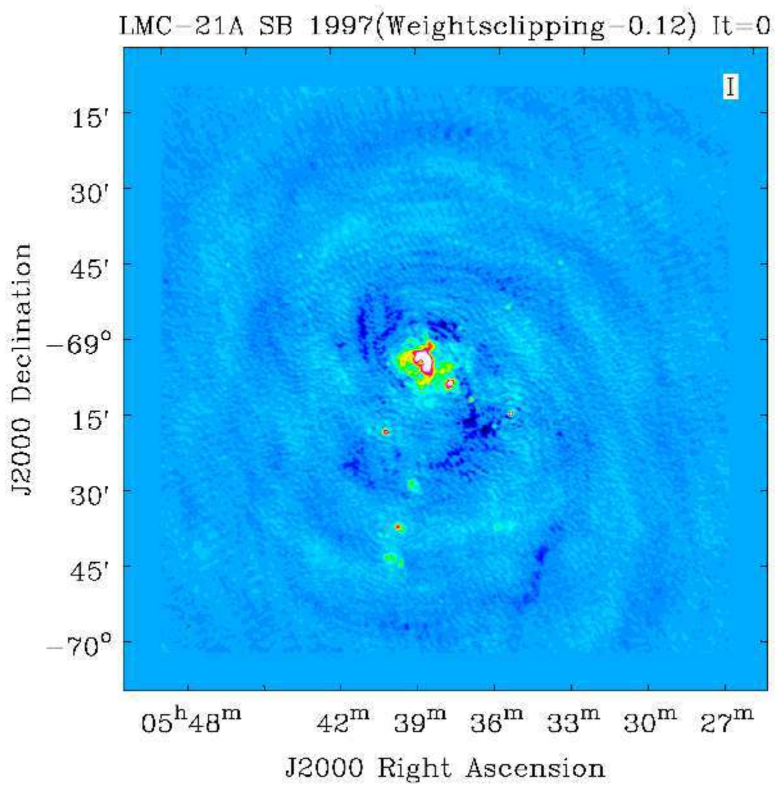
Minor Cycle: CLEAN down to the sidelobe level of brightest pixel in the residual image

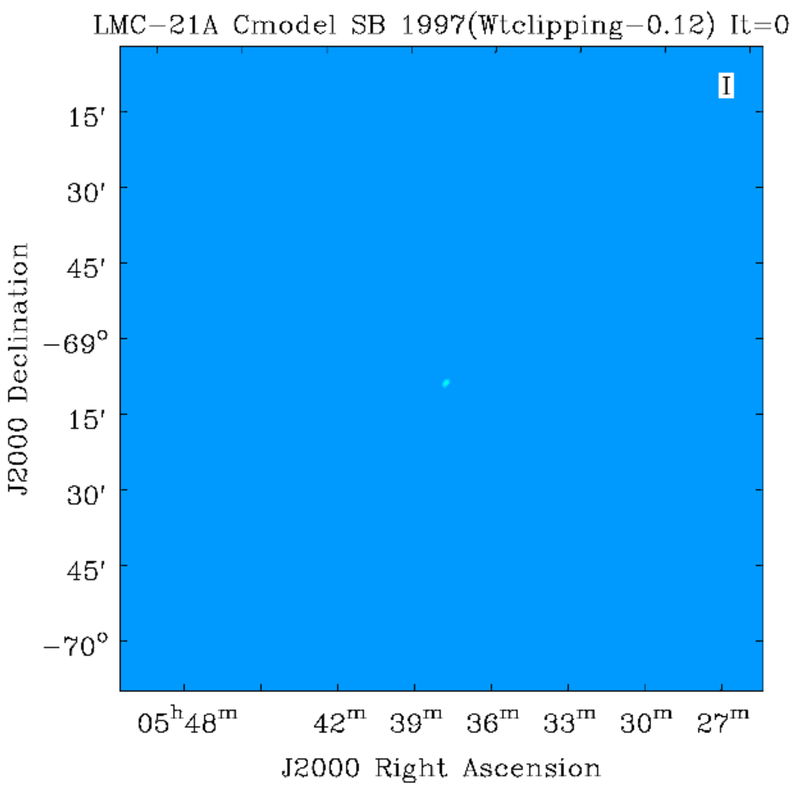
Example — Edge Features from diffuse emission



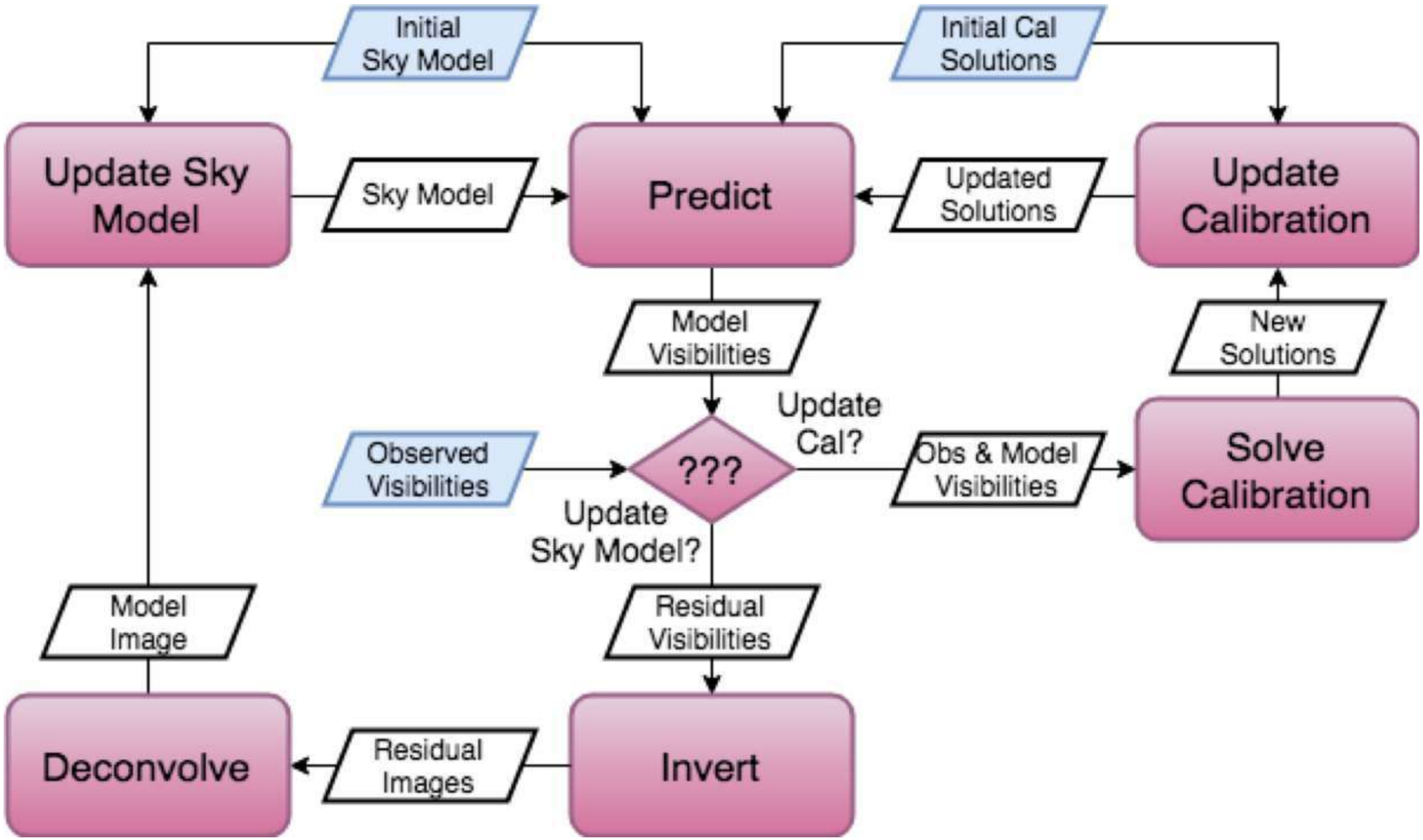
Example — Due to snapshotting & pre-conditioning



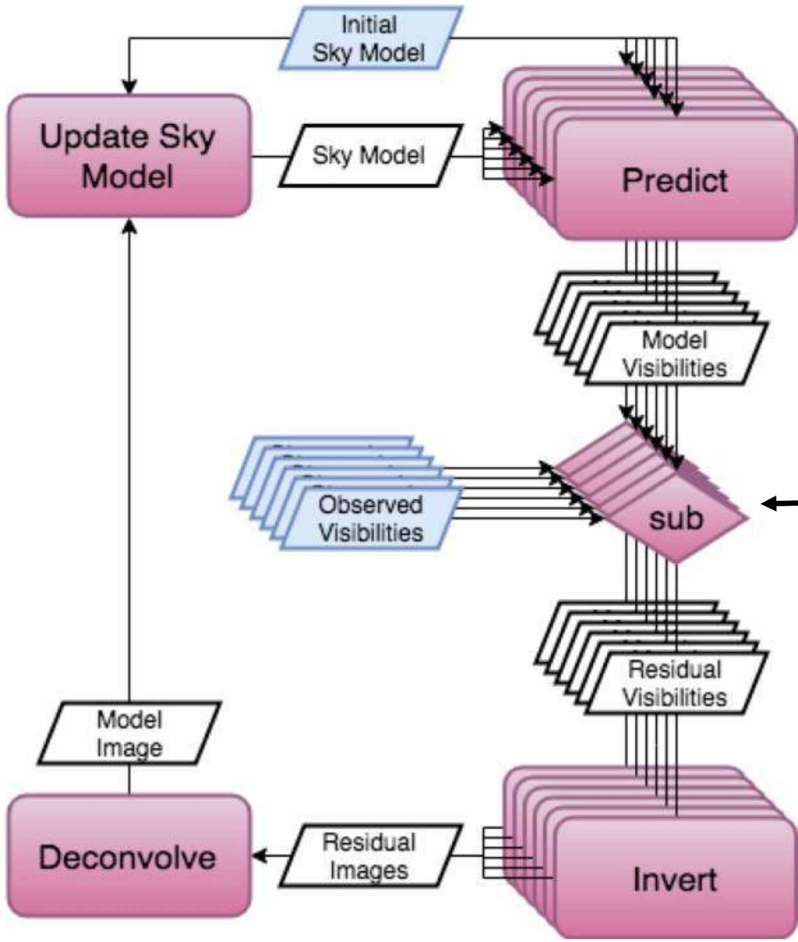

LMC−21A SB 1997(Weightsclipping−0.12) It=0
I
15′
30′
45′
−69°
15′
30′
45′
−70°
J2000 Declination
05h48m
42m
39m
36m
33m
30m
27m
J2000 Right Ascension


LMC−21A Cmodel SB 1997(Wtclipping−0.12) It=0
I
15′
30′
45′
−69°
15′
30′
45′
−70°
J2000 Declination
05h48m
42m
39m
36m
33m
30m
27m
J2000 Right Ascension

Self Calibration



High Performance Computing



Straightforward data parallelism over frequency in major cycle

Split into many sub-images (i.e. facets)?

Potential bottleneck at the minor cycle

Heavy use of multi-threading?

Summary

- Wide fields of view, wide bandwidths, low frequencies and high dynamic range complicate synthesis imaging and deconvolution.
- There are a wide variety of approaches to push the limits, most coming with computational overheads.
- More information to come on various topics:
  - ASKAPsoft and RTS lectures this afternoon
  - Wed: Ionosphere and calibration at low frequencies (Natasha)
  - Wed: Wide-field imaging and mosaicking (Tim)
  - Fri: High dynamic range imaging (Ian)
  - Fri: Polarimetry (Emil)
  - Fri: Spectral line processing (Karen)
  - Fri: Transient imaging and detection (Christene)
  - Fri: Source extraction and characterisation (Paul and Tobias)

A few other approaches that were skipped

- The Maximum Entropy Method (MEM) — diffuse structure
- Compressed Sensing — find “CLEAN components” in other sparse domains (e.g. the wavelet transform of an image)
- Peeling — generate local calibration solutions for strong sources
- Forward modelling — accurately calculate direction-dep. PSF
- RM synthesis — Image de-rotated “rotation measure” images