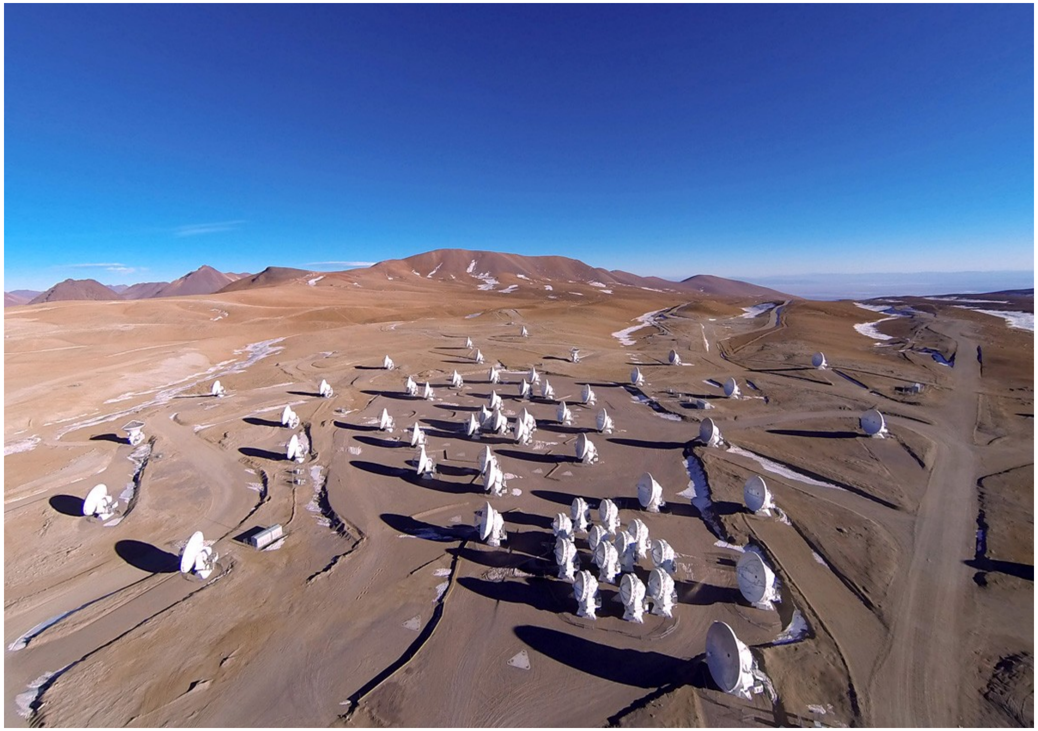
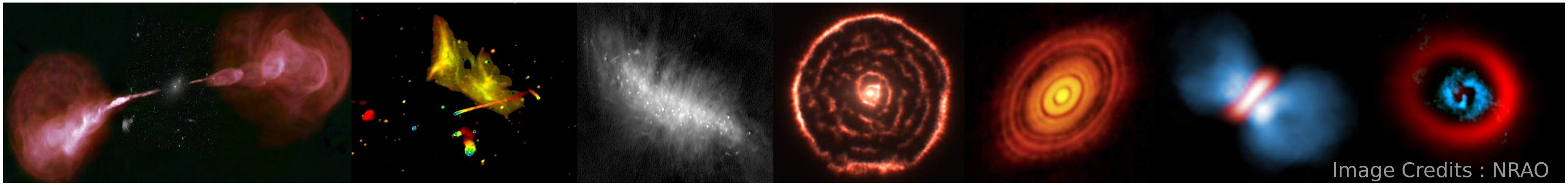
Introduction to Radio Interferometry – Algorithms and Computing





CENTER FOR HIGH THROUGHPUT COMPUTING

NRAO

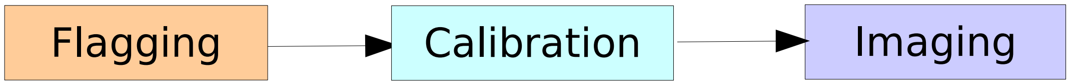
Urvashi Rau
National Radio Astronomy Observatory

CHTC / NRAO Meeting
26 August 2019

Outline

- Image formation with a radio interferometer

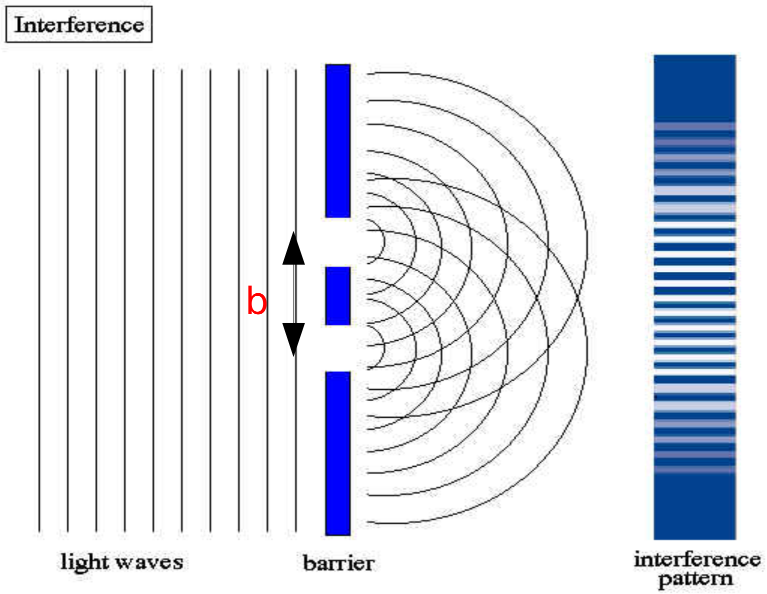
- Data processing steps



  - What / Why
  - Ability to partition data and algorithms
    - Data access patterns
    - I/O, compute and memory hot-spots
    - Complexity / flexibility required of software and frameworks

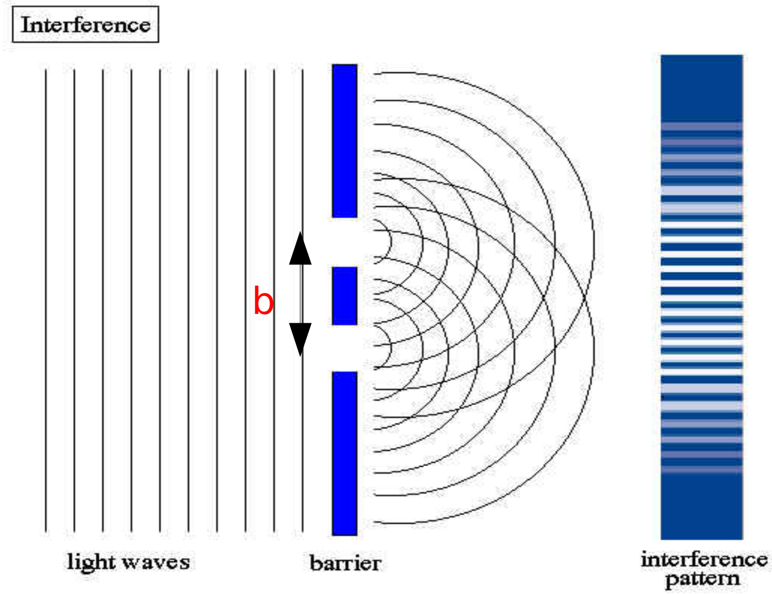
An interferometer is an indirect imaging device

Young's double slit experiment

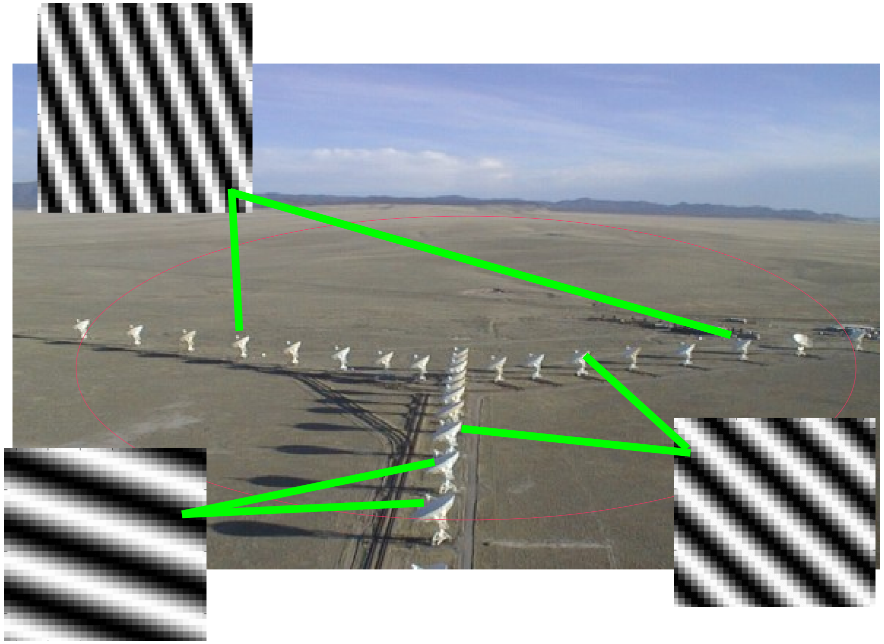


An interferometer is an indirect imaging device

Young's double slit experiment

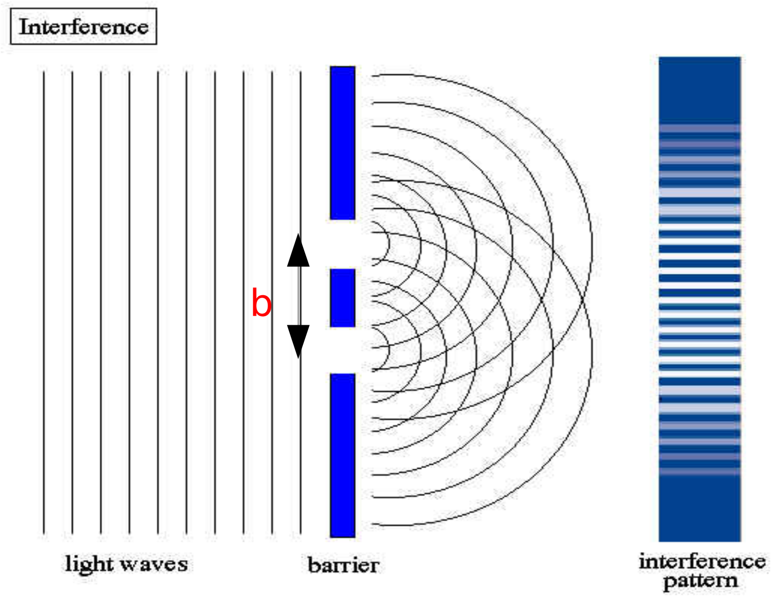


Each antenna-pair => one 2D fringe

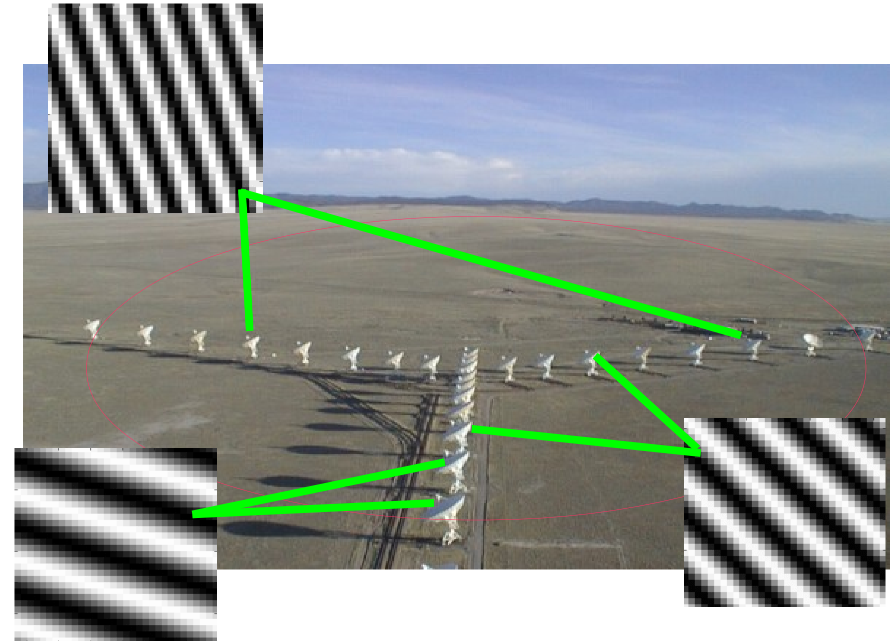


An interferometer is an indirect imaging device

Young's double slit experiment



Each antenna-pair => one 2D fringe



2D Fourier transform :

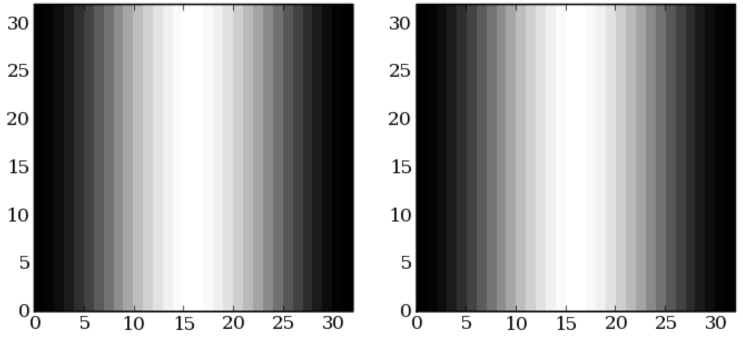
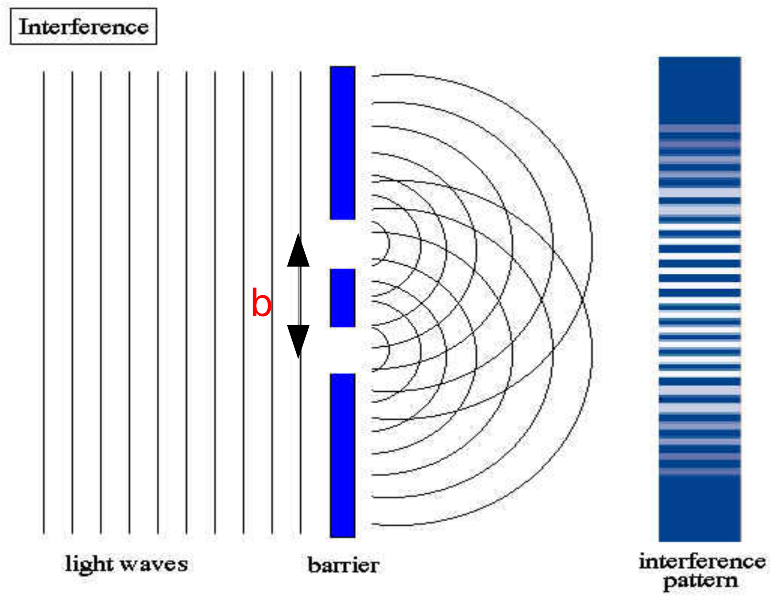


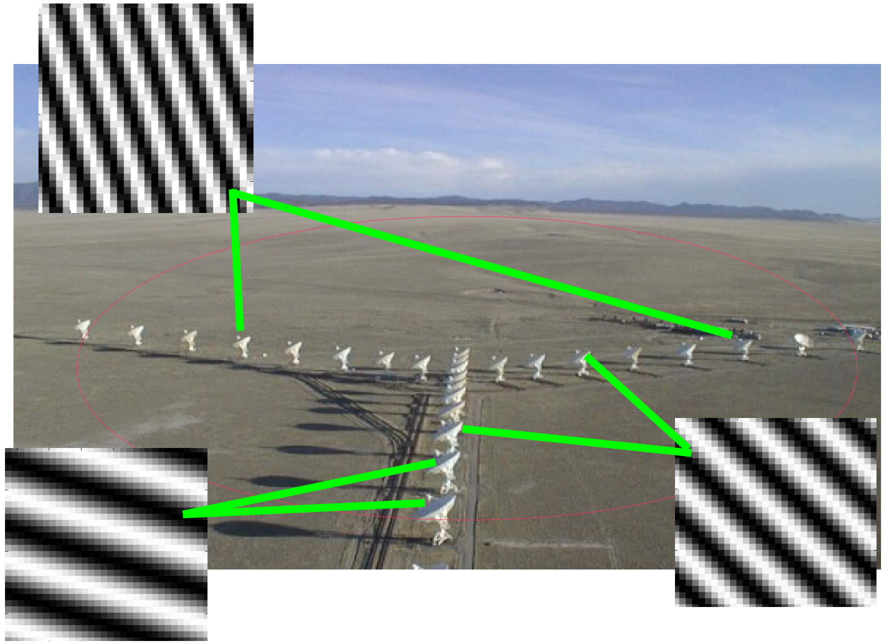
Image = sum of cosine 'fringes'.

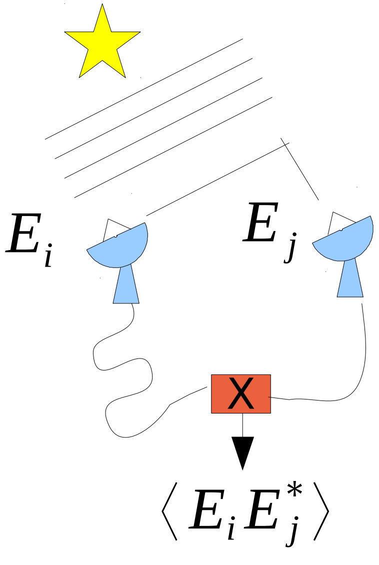
An interferometer is an indirect imaging device

Young's double slit experiment



Each antenna-pair => one 2D fringe





Measuring fringe parameters

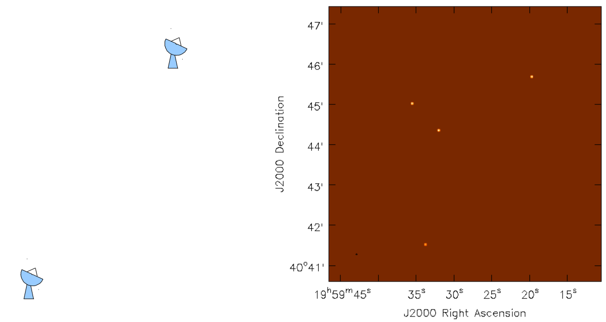
Amplitude, Phase : $\langle E_i E_j^* \rangle$ is a complex number

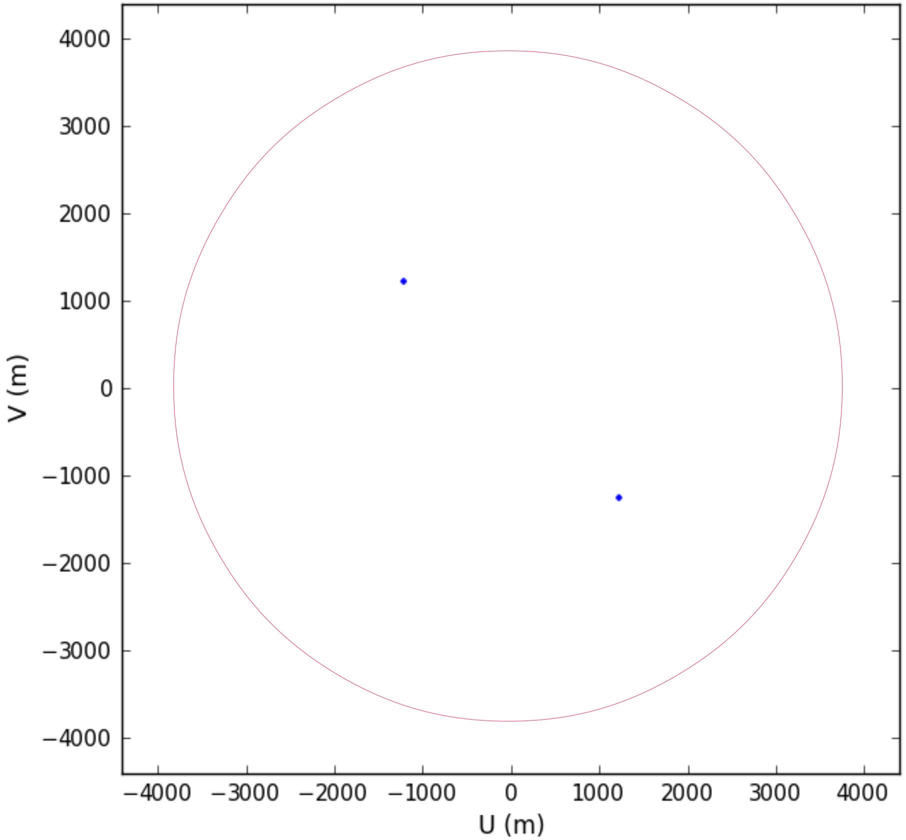
Orientation, Wavelength : Vector between each pair of antennas

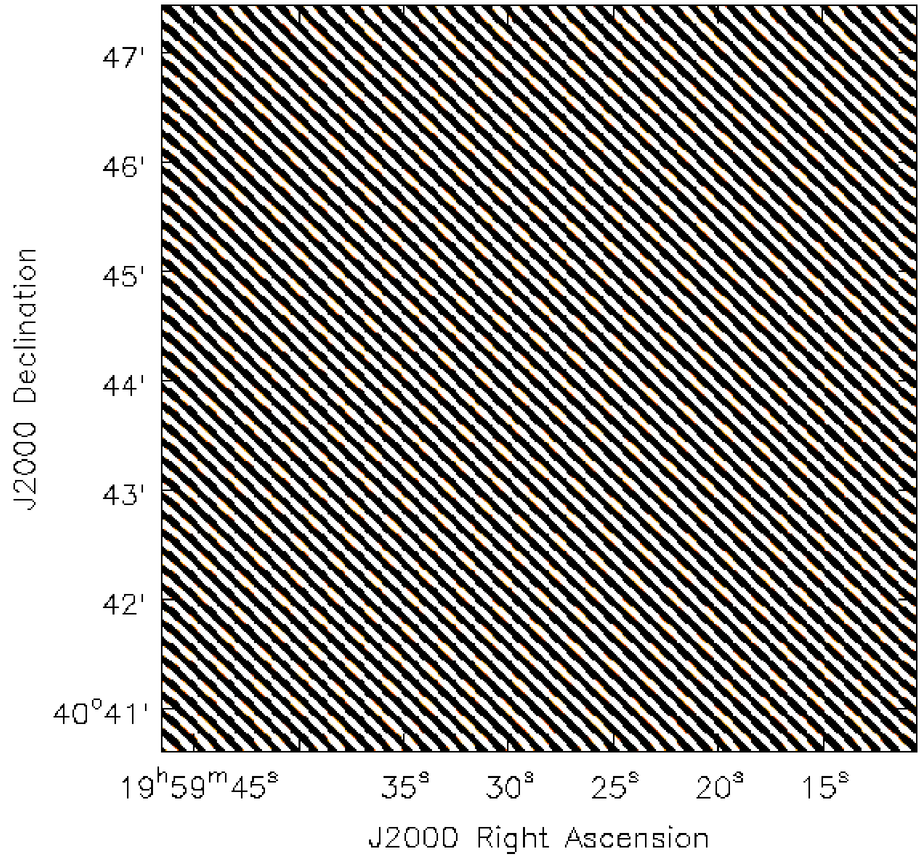
Goal : Measure as many distinct fringes as possible

Aperture Synthesis Imaging

Image with 2 antennas => 1 fringe





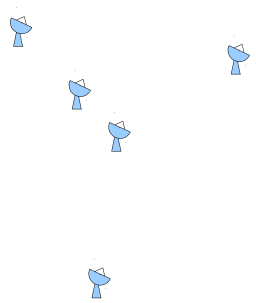


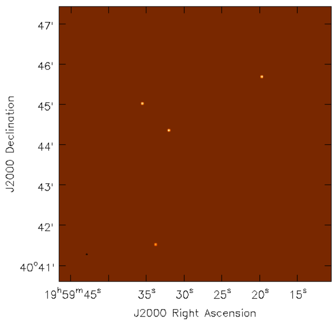
Synthesized Aperture

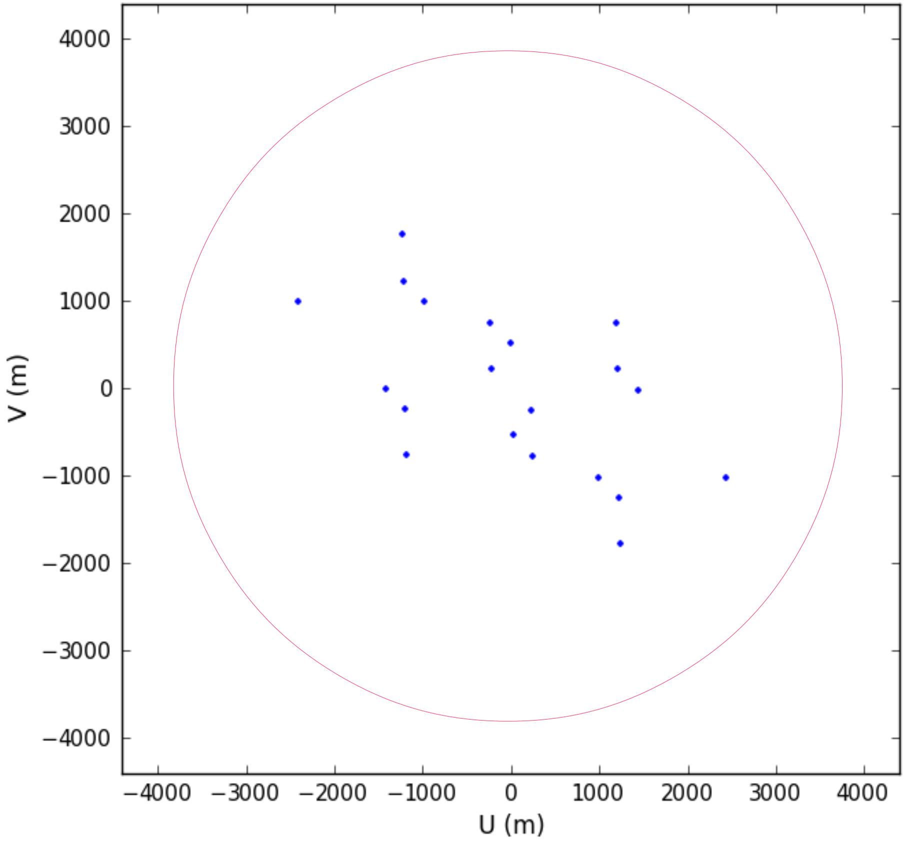
Observed Image

Aperture Synthesis Imaging

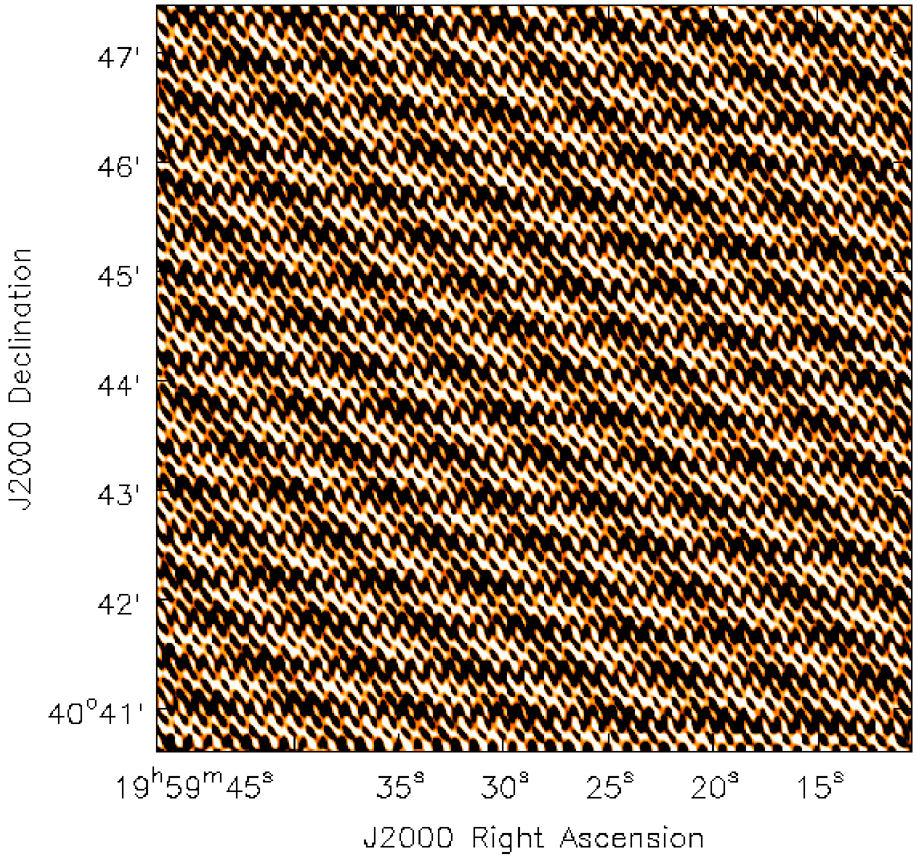
Image with 5 antennas => 10 fringes







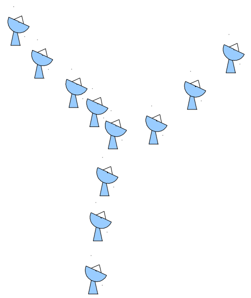
Synthesized Aperture

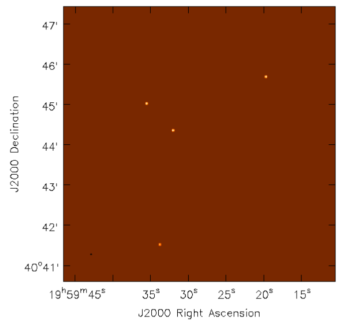


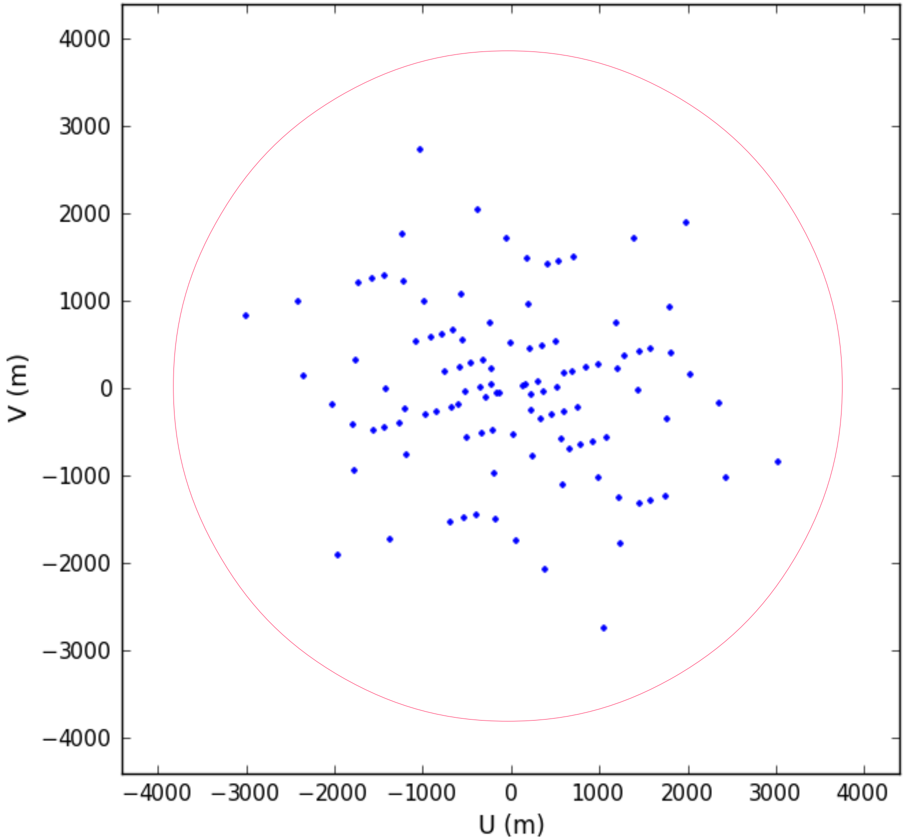
Observed Image

Aperture Synthesis Imaging

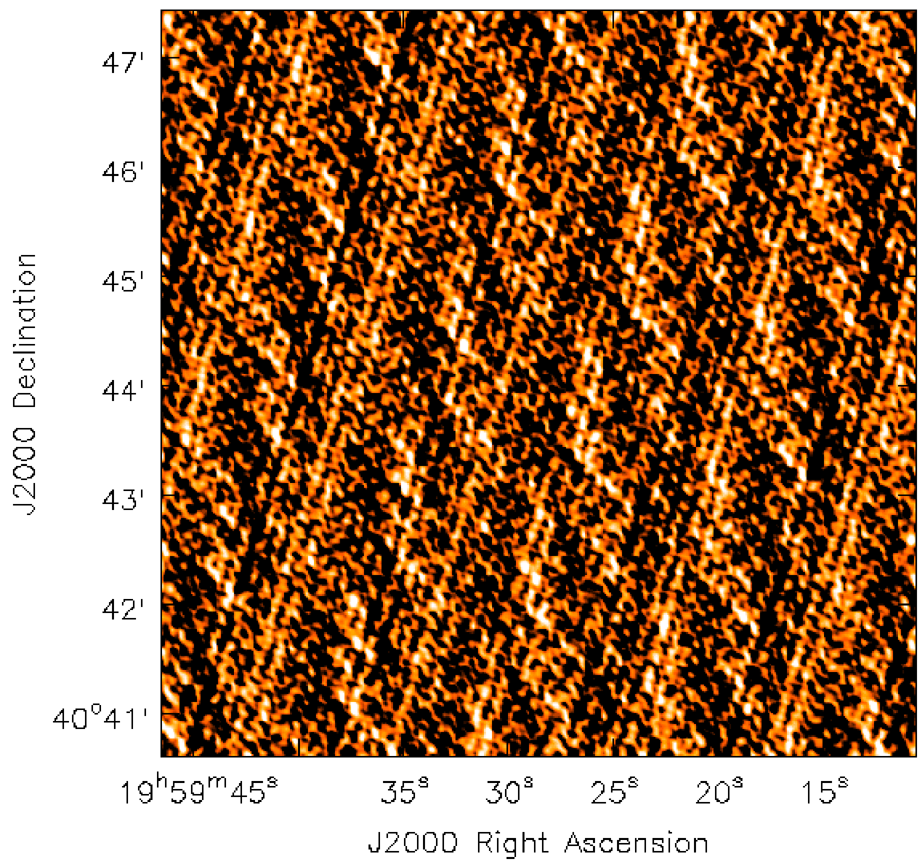
Image with 11 antennas => 55 fringes







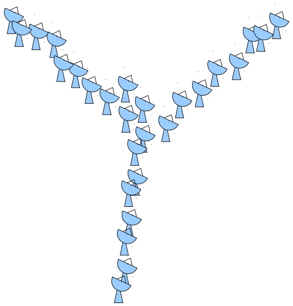
Synthesized Aperture

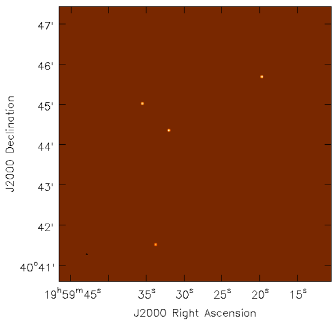


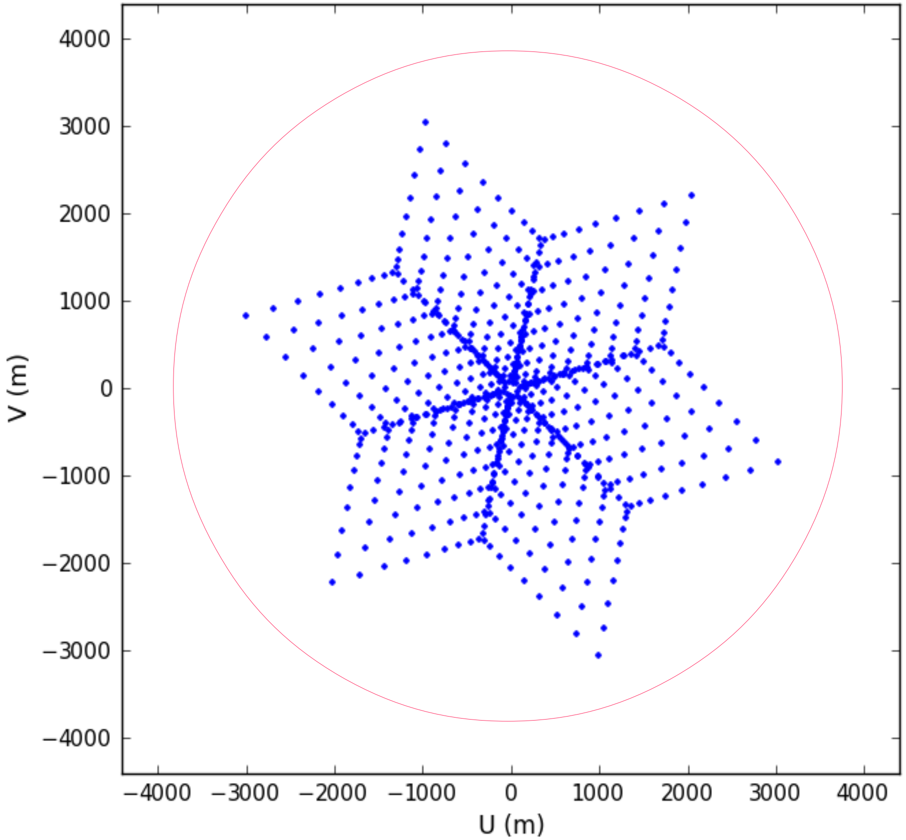
Observed Image

Aperture Synthesis Imaging

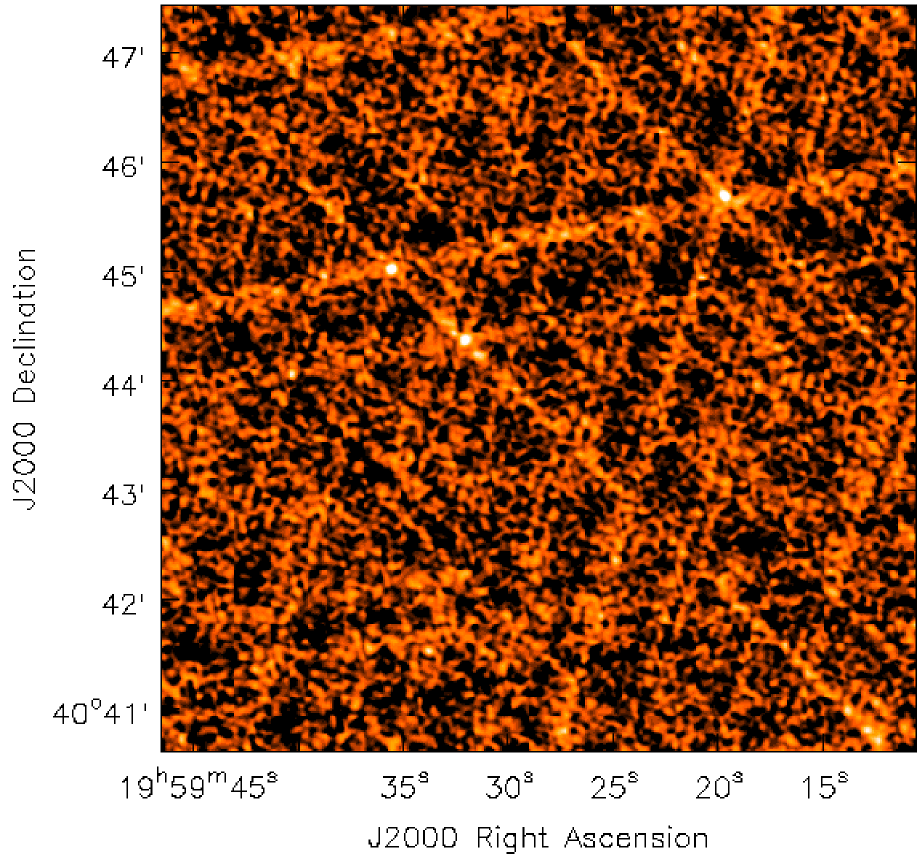
Image with 27 antennas => 351 fringes







Synthesized Aperture

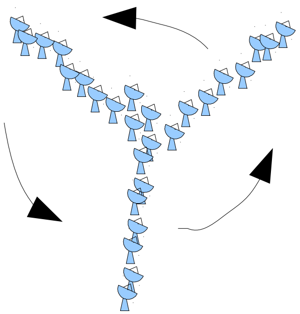


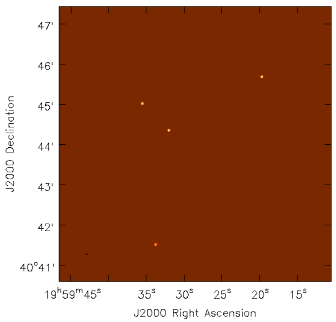
Observed Image

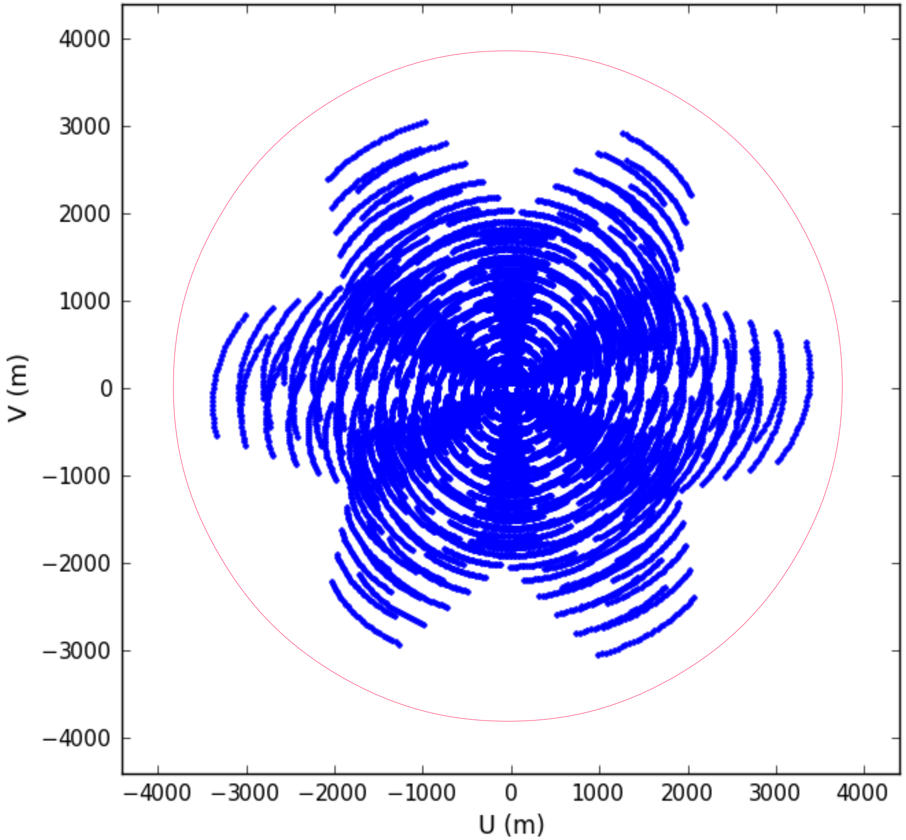
Aperture Synthesis Imaging

Image with 27 antennas over 2 hours

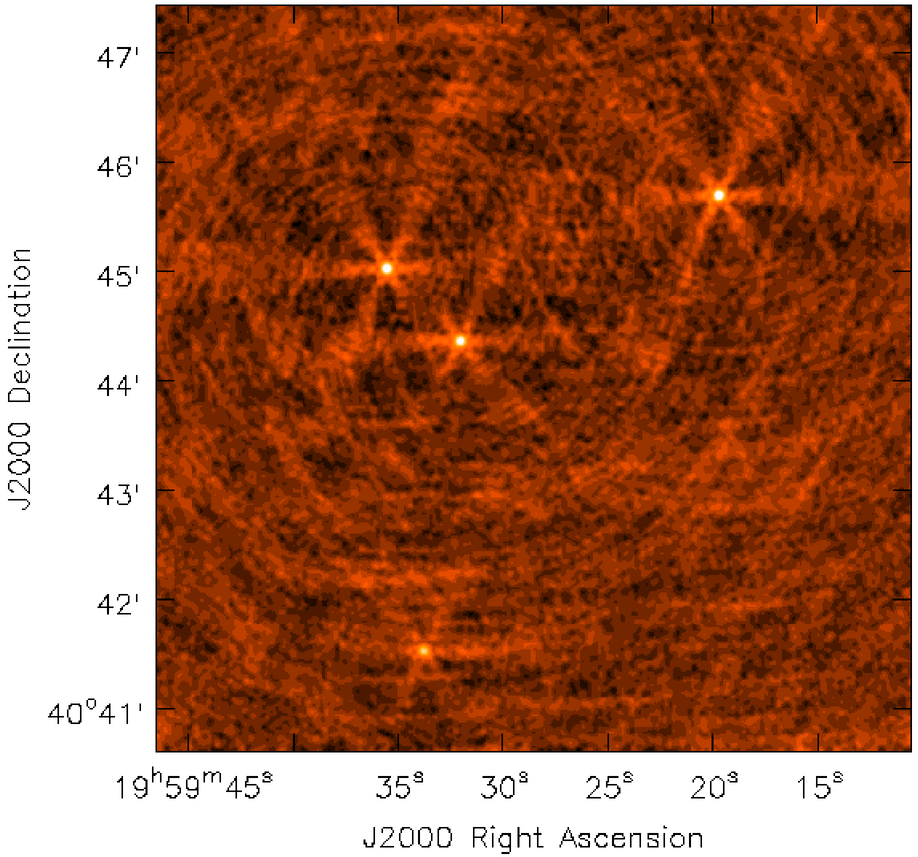
" Earth Rotation Synthesis "







Synthesized Aperture

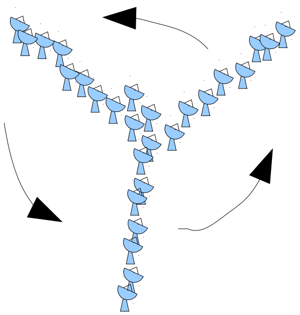


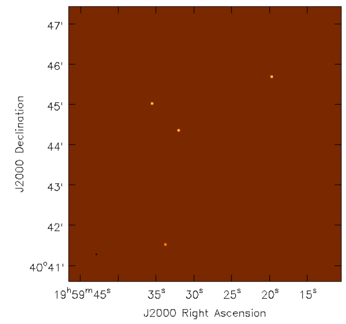
Observed Image

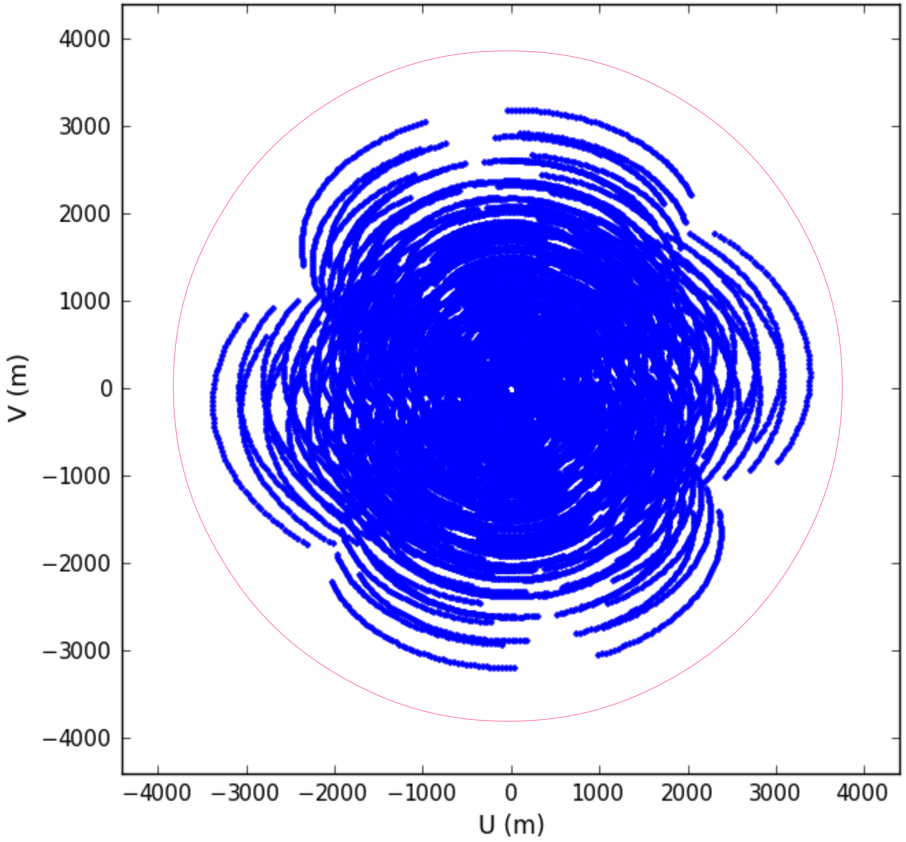
Aperture Synthesis Imaging

Image with 27 antennas over 4 hours

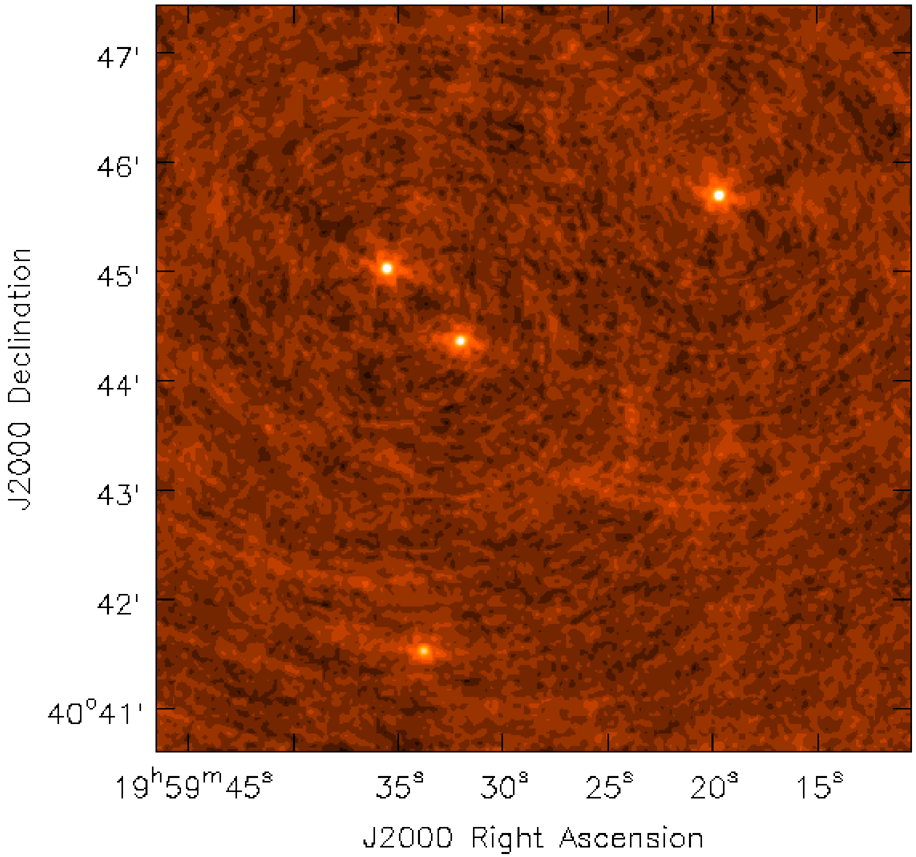
“ Earth Rotation Synthesis “







Synthesized Aperture

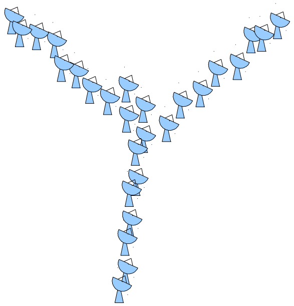


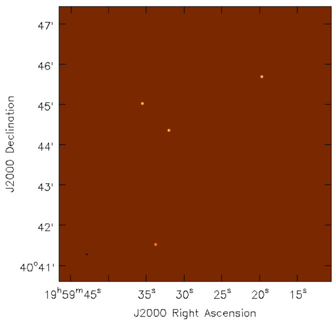
Observed Image

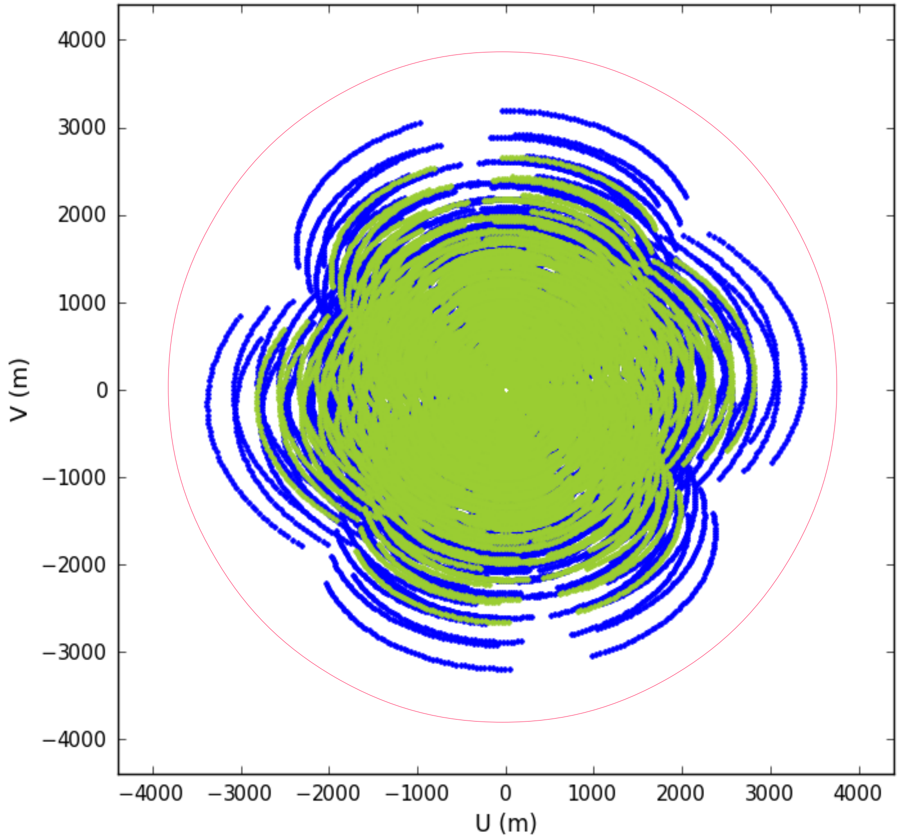
Aperture Synthesis Imaging

Image with 27 antennas over 4 hours
at 2 observing frequencies

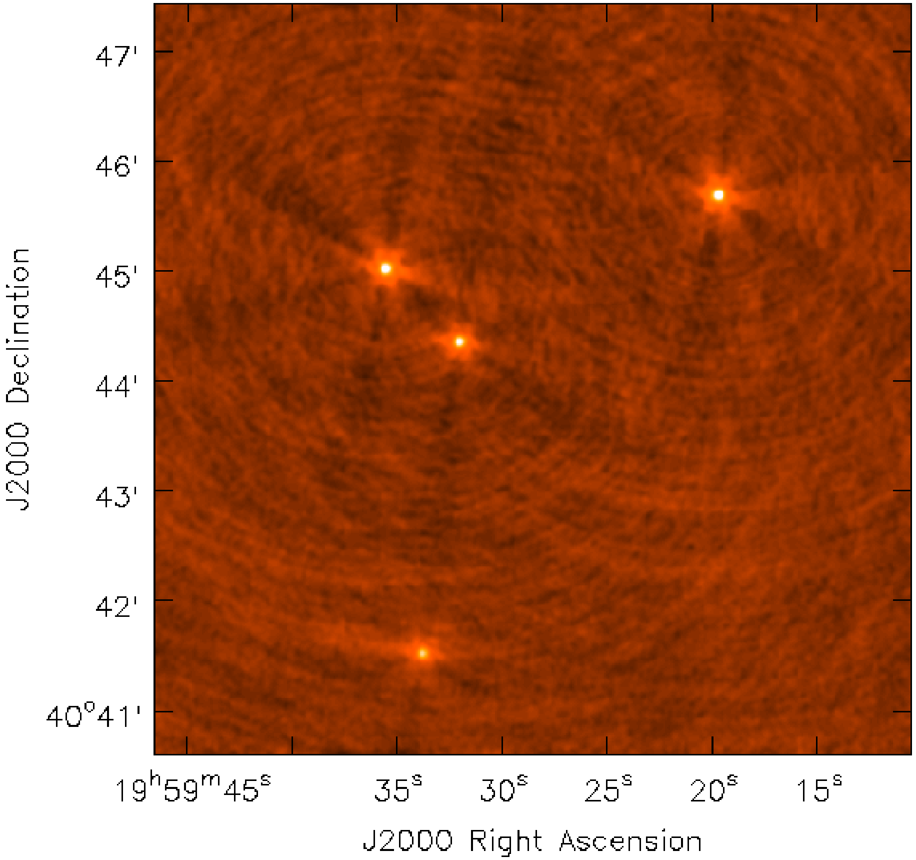
“ Multi-frequency Synthesis “







Synthesized Aperture

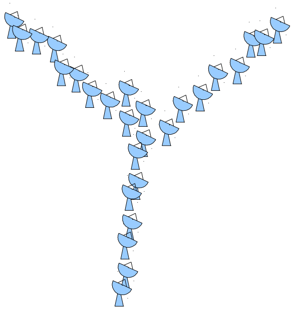


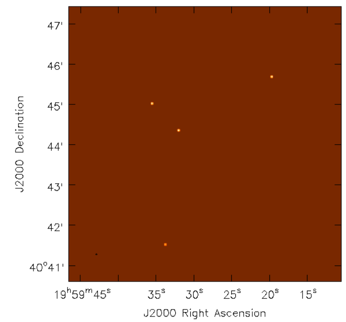
Observed Image

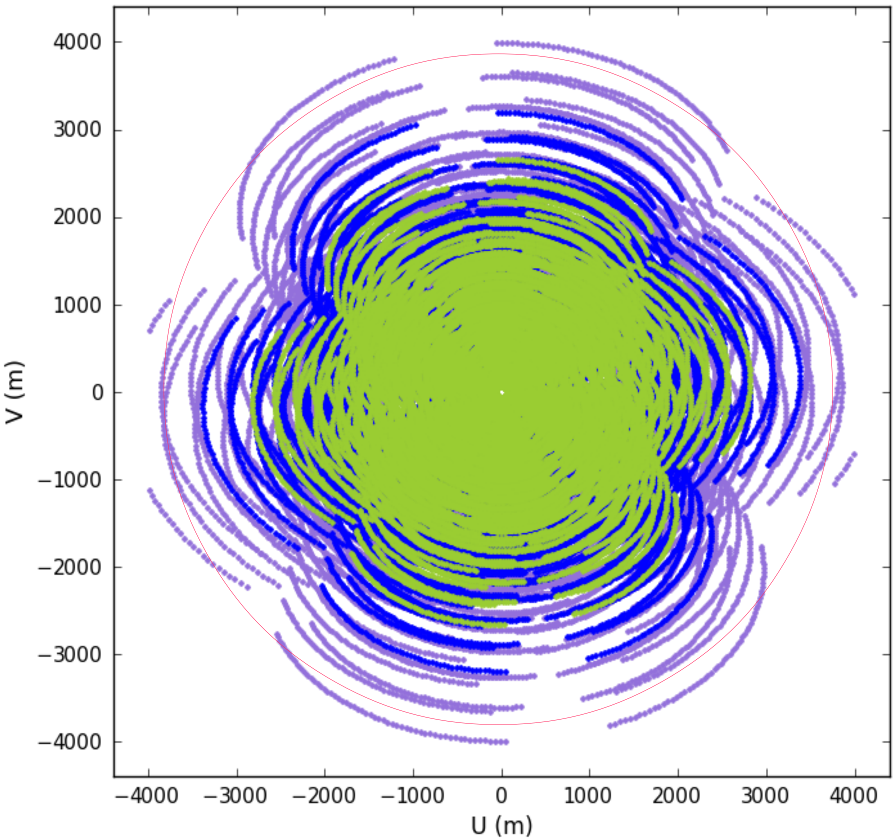
Aperture Synthesis Imaging

Image with 27 antennas over 4 hours
at 3 observing frequencies

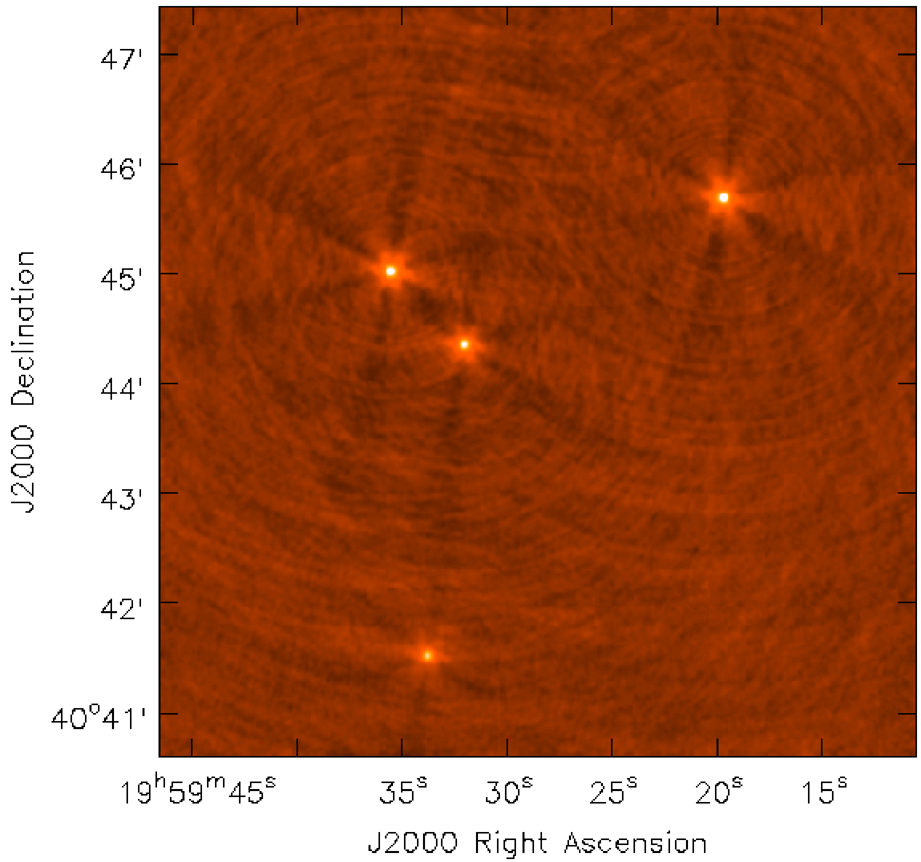
“ Multi-frequency Synthesis “





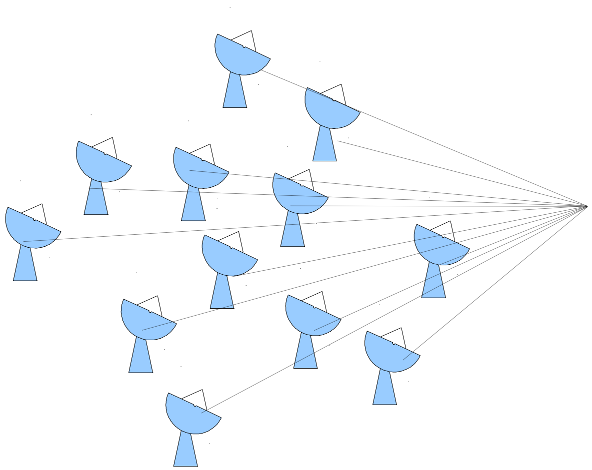


Synthesized Aperture



Observed Image

Data Acquisition and Analysis



Correlation (Real time system. FPGA/ASIC + backend cluster)

Time Series → Correlation → Spectral Channels → Integrate

Example Data rate : N(N-1)/2 * 1000 complex values per second

Data Archive (2.4 PB RAID storage)

Each observation is stored as a relational database

Example : VLA archive is 1.8 PB in size (+ 1 TB per day)

Post Processing – (1.6 PB Lustre FS, workstations, 90 node cluster, AWS)

Flagging → Calibration → Imaging

Identify and mask corrupted data (RFI, Instrument errors, etc)

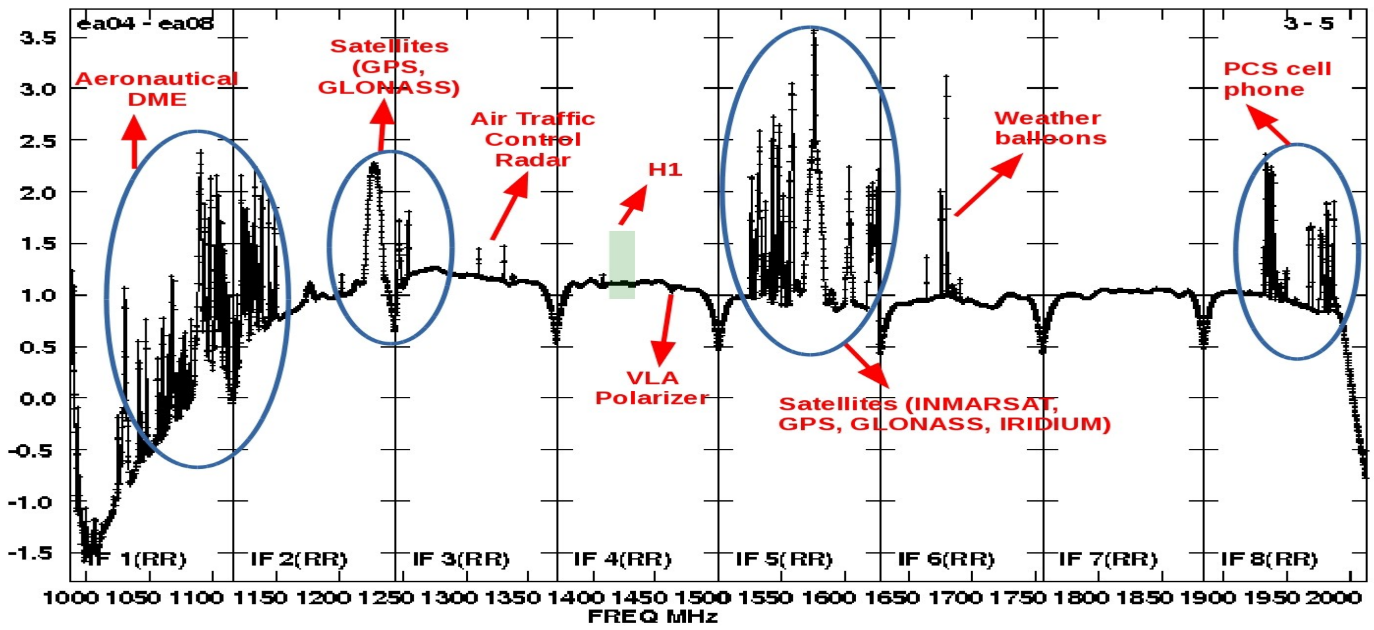
Derive and apply corrections to undo the effects of complex valued antenna gains

Reconstruct images by iterative model fitting while correcting for other instrumental effects

Thanks to J.Robnett for numbers

Radio Frequency Interference
- Cellular phones, aircraft radar, satellite comms, military radar, car radars, etc...

Instrumental flags
- Antenna tracking delays, glitches in signal processing, antenna dropouts, shadowing...



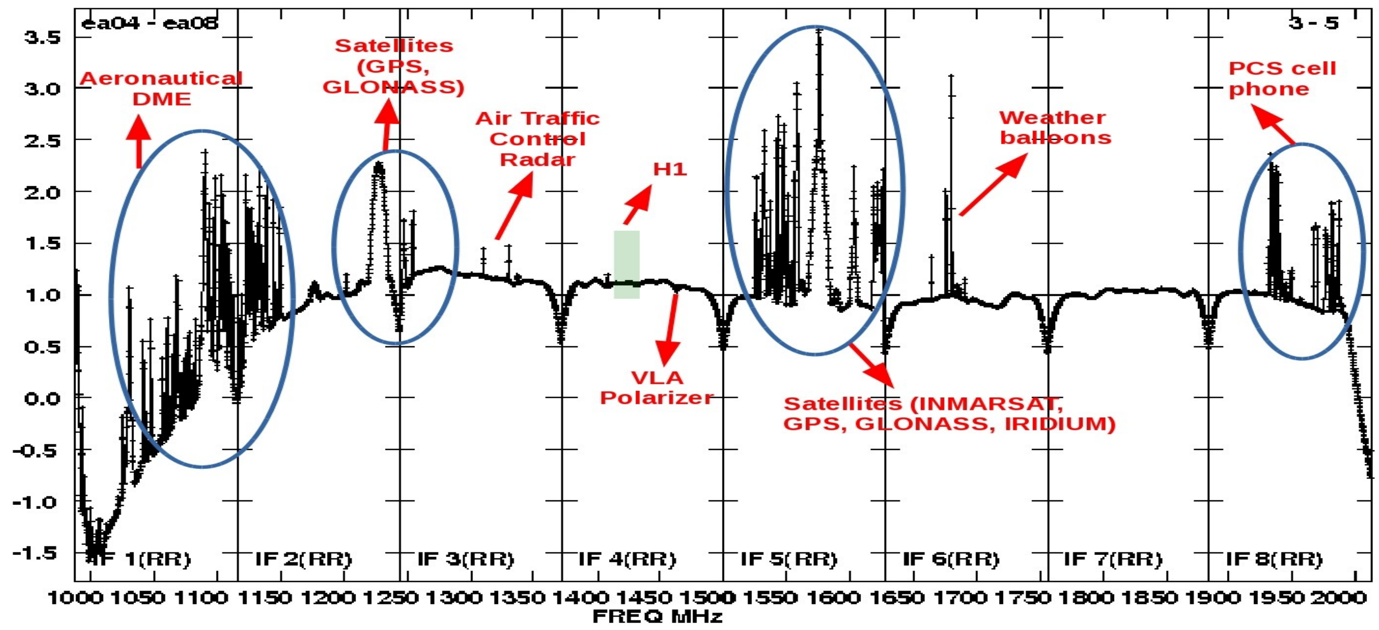
Flagging → Calibration → Imaging

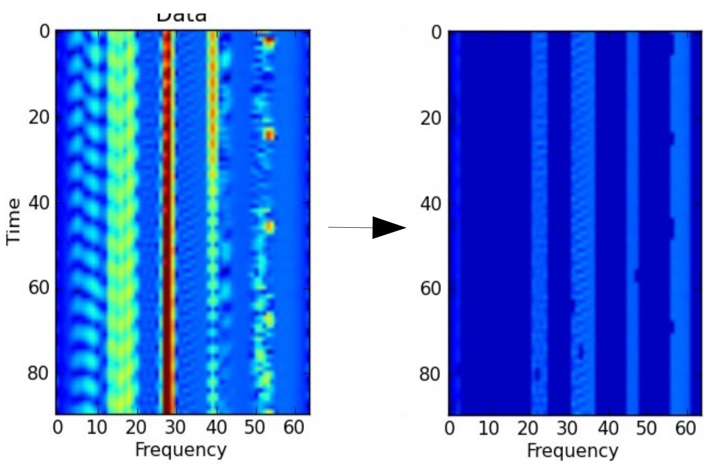
Radio Frequency Interference

- Cellular phones, aircraft radar, satellite comms, military radar, car radars, etc...

Instrumental flags

- Antenna tracking delays, glitches in signal processing, antenna dropouts, shadowing...





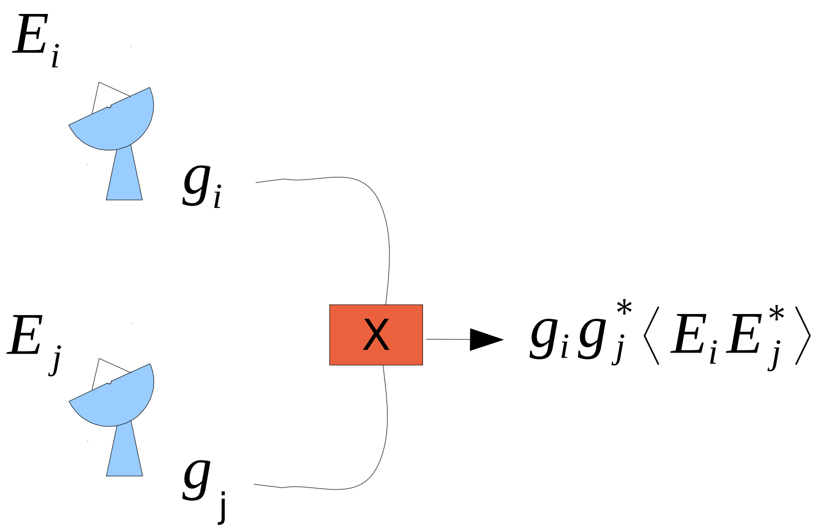
Flagging = masking unusable data

- Manual Flags : Select and mask known bad data
- Auto-Flags : Model based and statistical outlier detectors

- Parallelize by partitioning the data along axes where the algorithms are also separable.

  - Example : Consider 2D time-frequency slices, separately for each antenna-pair

The front-end electronics on each antenna introduces a multiplicative complex gain on the incoming signal. This must be removed.



(1) Observe a known source

$\langle E_i E_j^* \rangle$ is known

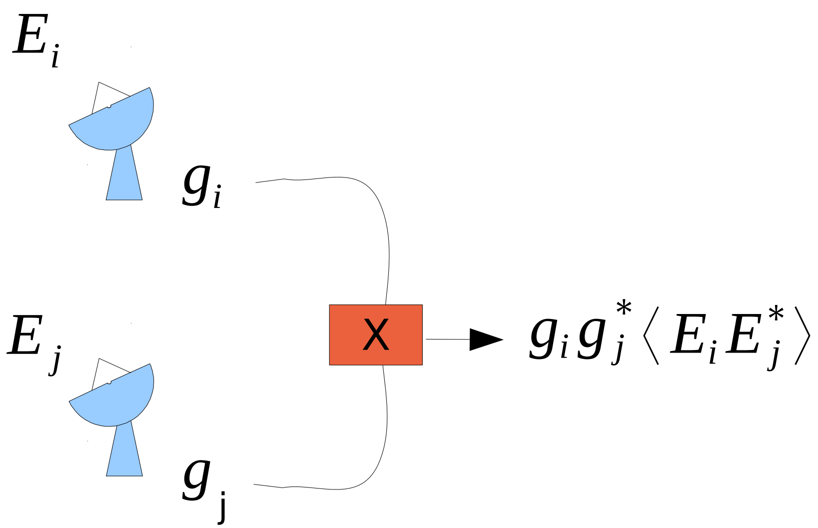
(2) Use data from all correlation pairs ij
Solve for complex gains g_i

(3) Apply corrections to target data :

$$\frac{g_i g_j^* \langle E_i E_j^* \rangle}{g_i g_j^*}$$

Flagging → **Calibration** → Imaging

The front-end electronics on each antenna introduces a multiplicative complex gain on the incoming signal. This must be removed.



(1) Observe a known source

$\langle E_i E_j^* \rangle$ is known

(2) Use data from all correlation pairs ij
Solve for complex gains g_i

(3) Apply corrections to target data :

$$\frac{g_i g_j^* \langle E_i E_j^* \rangle}{g_i g_j^*}$$

Calibration is easily parallelized but requires flexibility on the type of data views.

– Gain solution requires all baselines per time/freq to be visible together.
– Gain application can be done independently on each data point.

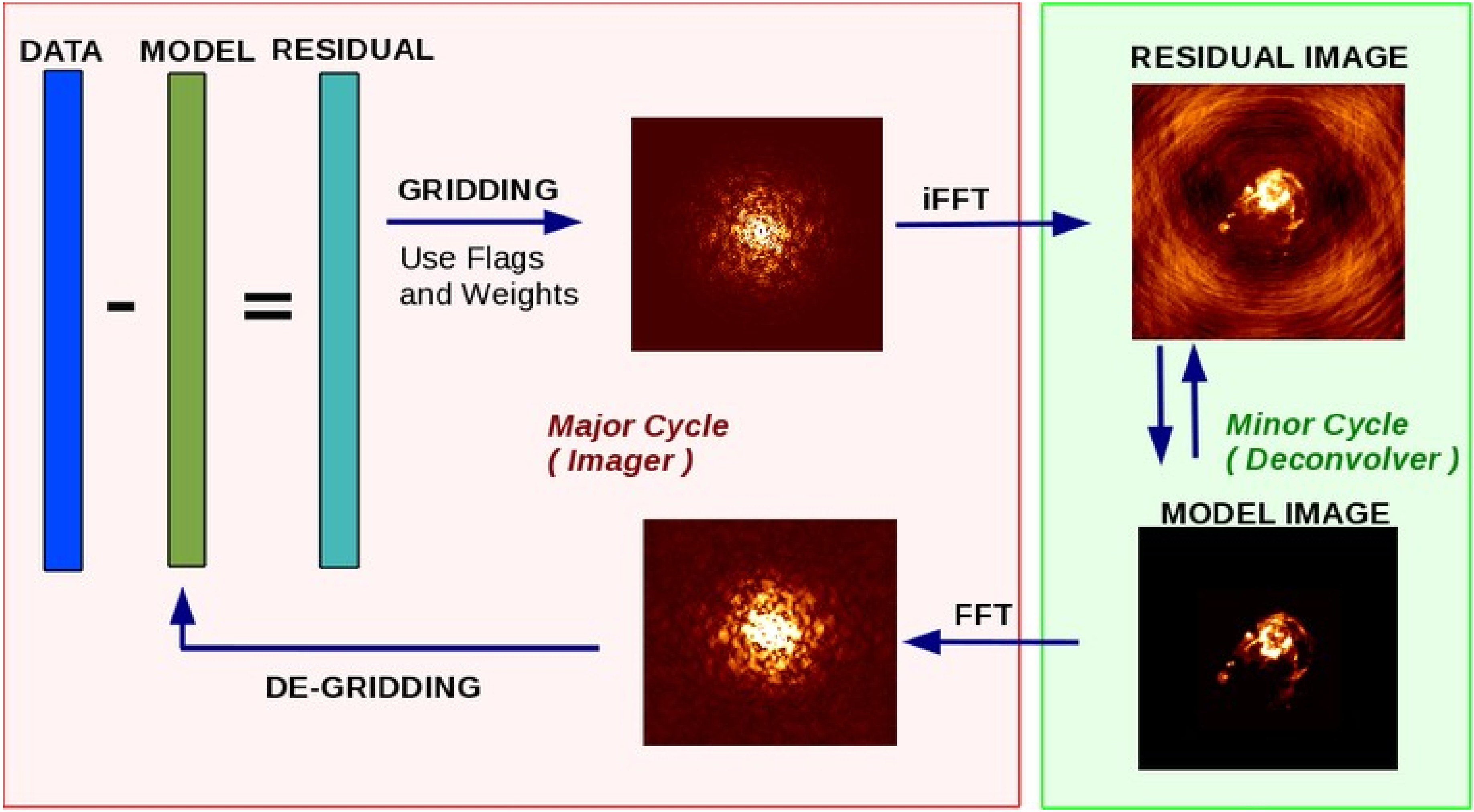
– A multi-stage process, requiring different pre-averaged views of the data

Flagging → Calibration → Imaging

Image reconstruction is an iterative model-fitting / optimization problem

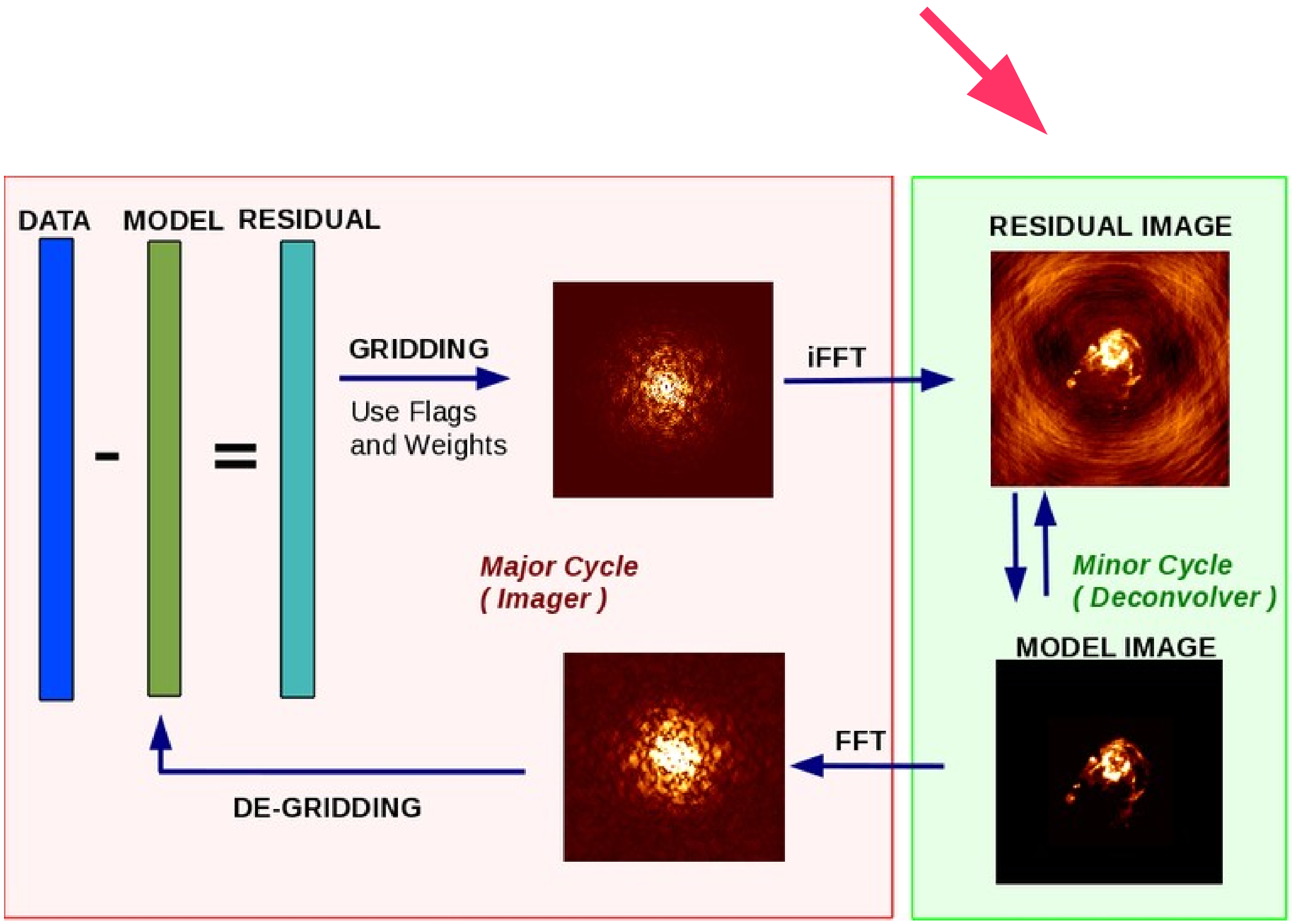
Measurement Eqn : $[A] I^m = V^{obs}$

Iterative solution : $I^m_{i+1} = I^m_i + g[A^T W A]^+ \left(A^T W \left(V^{obs} - A I^m_i\right)\right)$



Flagging → Calibration → Imaging

Minor Cycle : Deriving a sky model (image-domain)



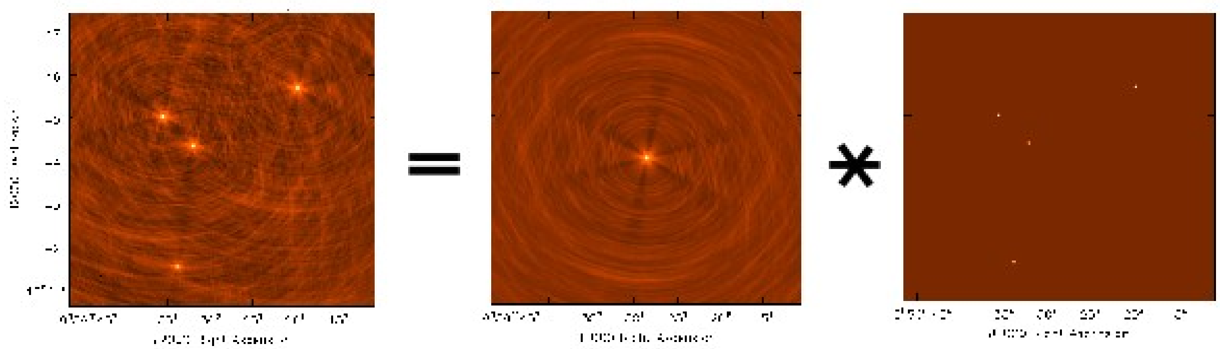
Flagging → Calibration → **Imaging**

Algorithms : Parameterized models + Iterative model fitting (L2, L1, TV..)

CLEAN : A greedy algorithm that solves a convolution equation

Needs to access all pixels in a few images per iteration.

Convolution Equation ==> Deconvolution

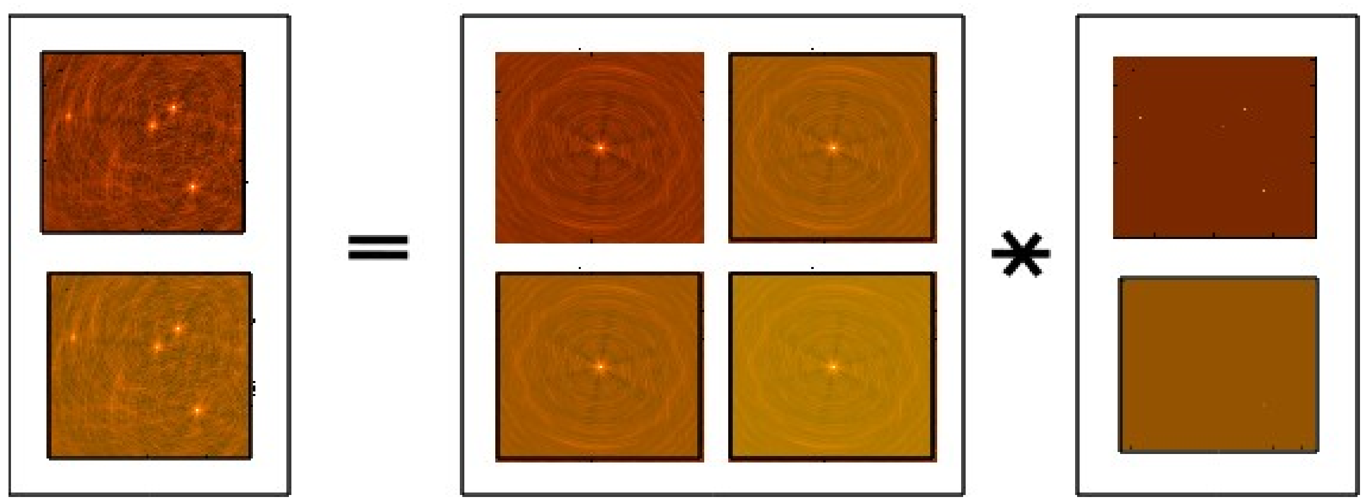


Multi-Term CLEAN : Solves a block-convolution equation.

Needs all pixels of O(Nblocks ^2) images to be accessed for each iteration.

With multi-scale and wide-band, Nblocks ~ 10.

Multi-Term Convolution Equation ==> Joint Deconvolution

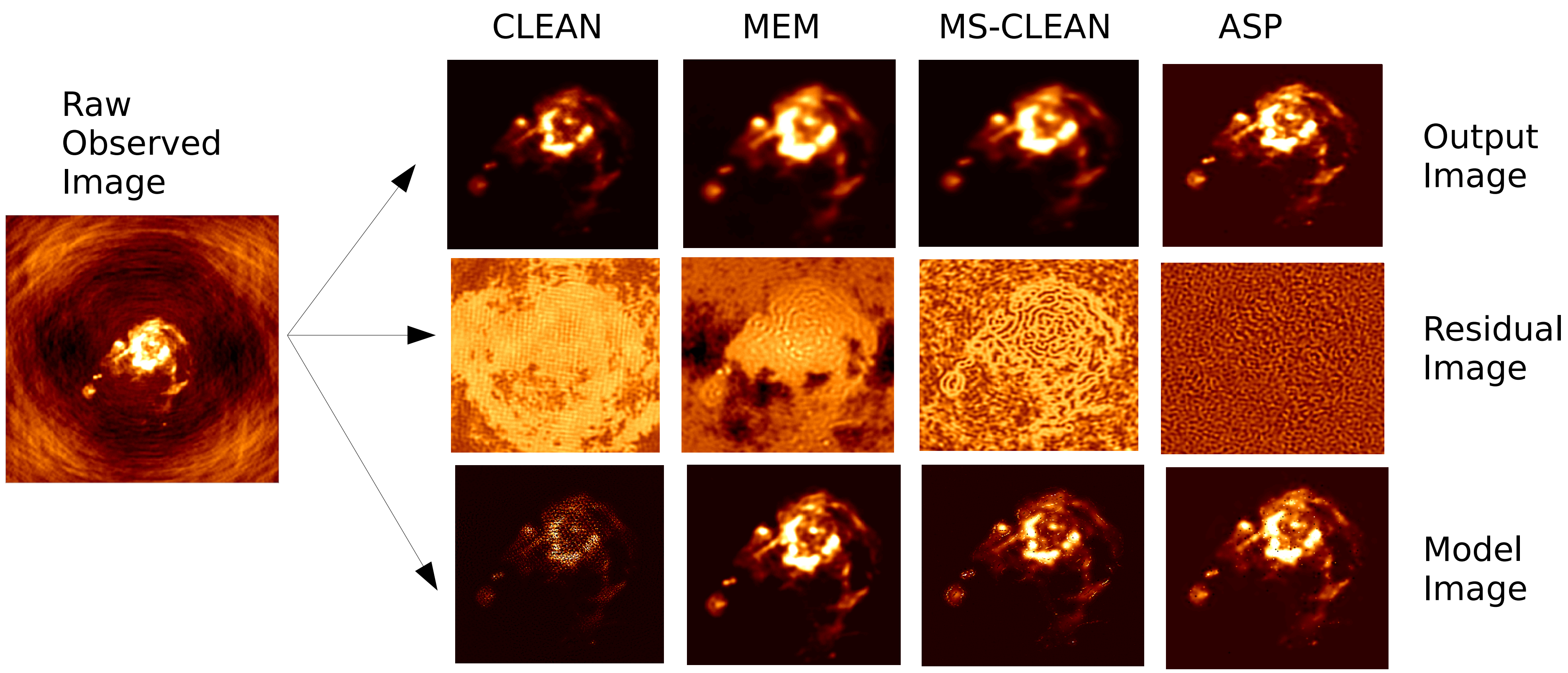


Flagging → Calibration → Imaging

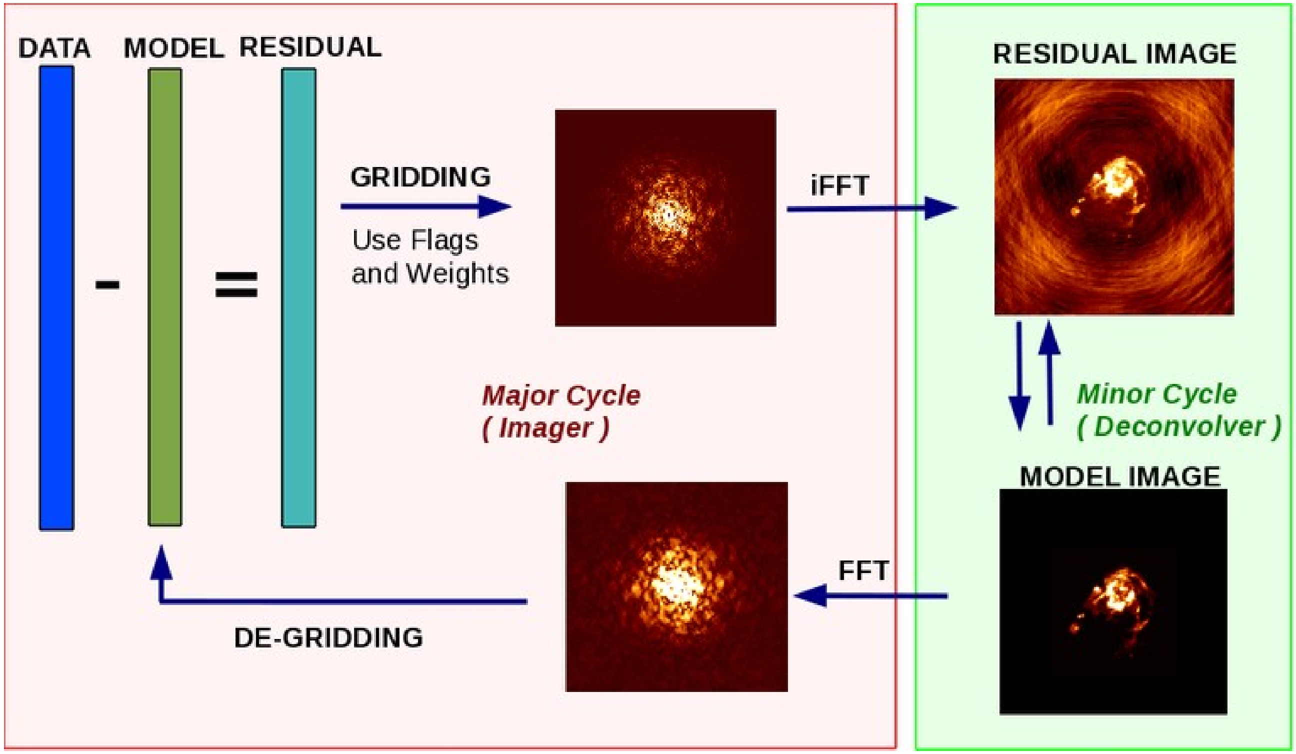
No unique solution (theoretically)

=> Algorithm choice depends on sky structure, data quality, target science

=> Different algorithms and parameters (e.g. convergence criteria, N_iterations) could result in orders of magnitude differences in computing load.



Major Cycle : Transforms between data and image/model spaces

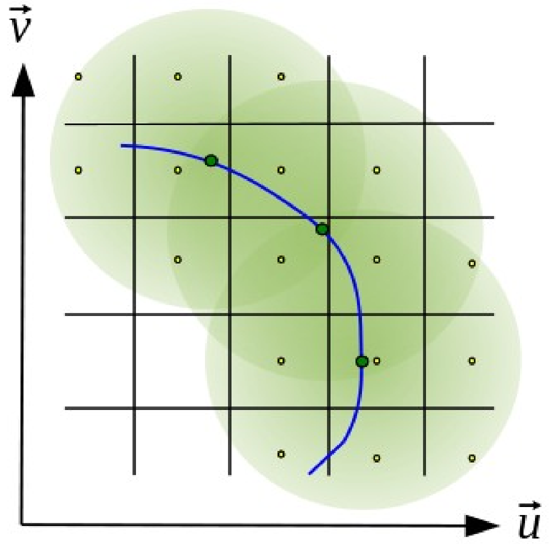


Flagging → Calibration → **Imaging**

Transforming the observed data into an image : Gridding + iFFT

Transforming the sky model into data-space : FFT + De-Gridding

Gridding = Convolutional Resampling of visibilities to a regular grid



Convolution in UV-domain (per vis)

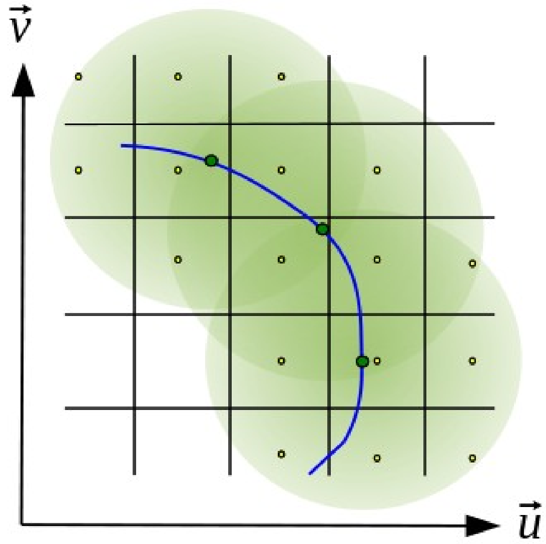
Visibility Prediction (de-gridding) :
Calculate a Forward Model

Flagging → Calibration → **Imaging**

Transforming the observed data into an image : Gridding + iFFT

Transforming the sky model into data-space : FFT + De-Gridding

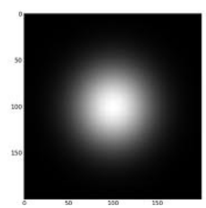
Gridding = Convolutional Resampling of visibilities to a regular grid



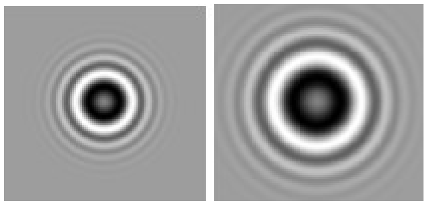
Convolution in UV-domain (per vis)

Visibility Prediction (de-gridding) :
Calculate a Forward Model

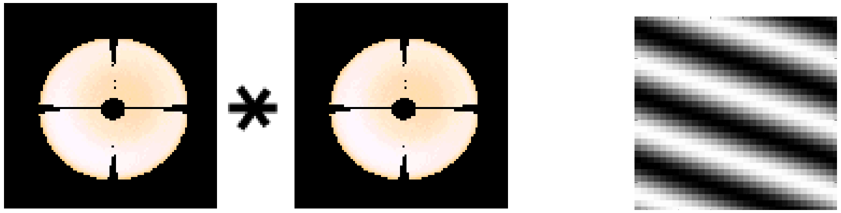
Standard Imaging :
Prolate Spheroidal



W-Projection :
FT of a Fresnel kernel



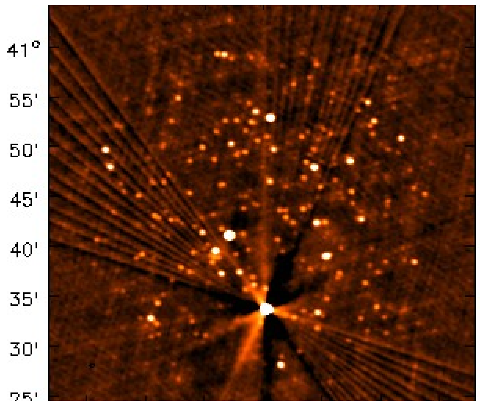
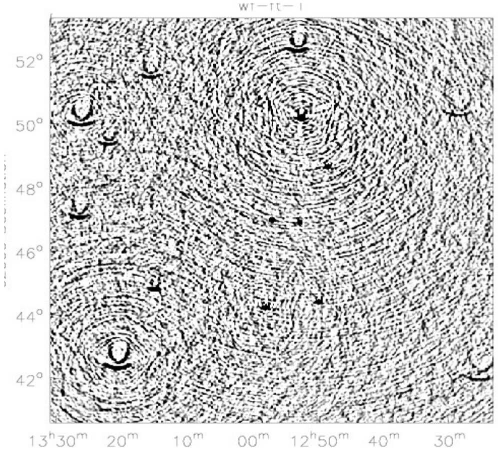
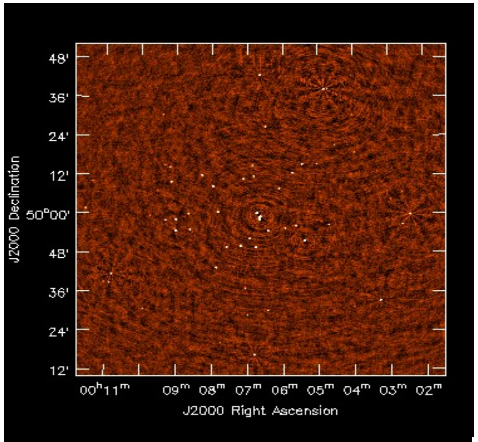
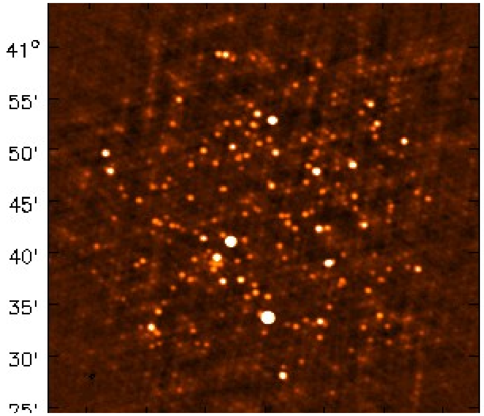
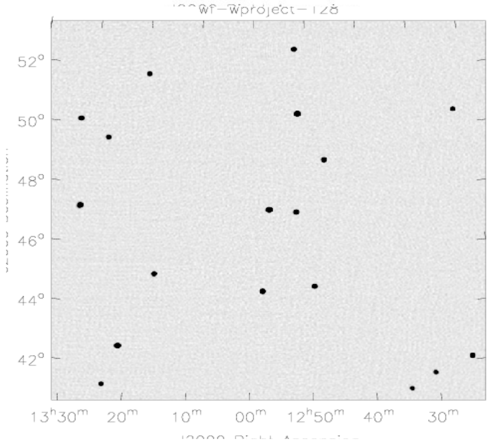
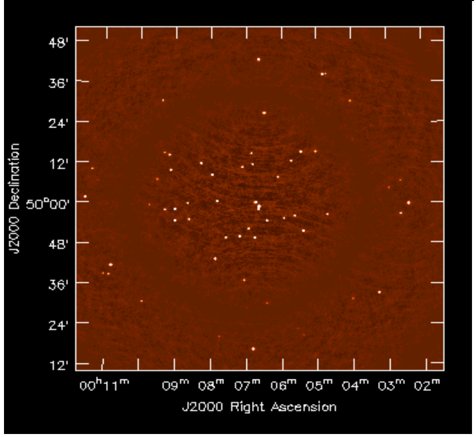
A-Projection : Baseline aperture illumination functions
+ phase gradients for pointing offsets
+ ionospheric refraction models



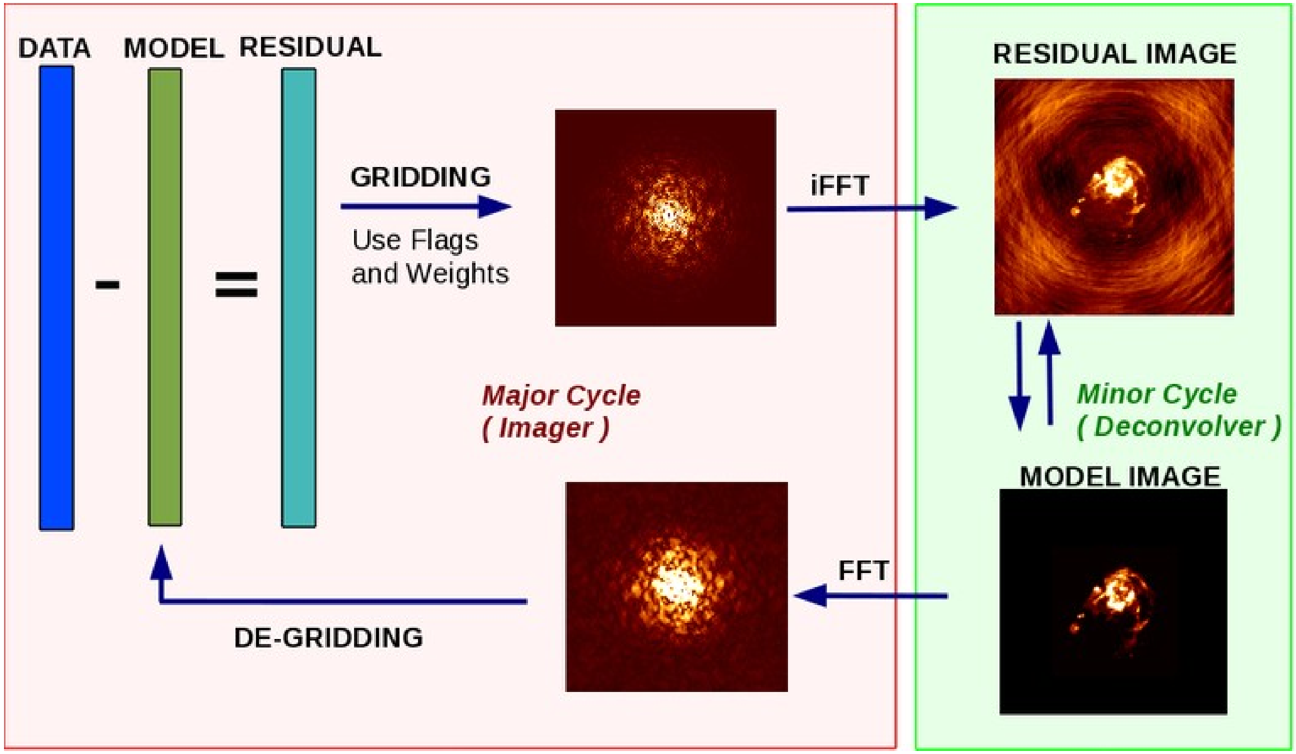
Gridding kernel sizes : 5x5 pix to 200x200 pix

Flagging → Calibration → **Imaging**

Examples of imaging improvements with the more expensive gridders

Kernel size :	Antenna Beam rotation with parallactic angle	Sky curvature, non-coplanarity (W-term)	Antenna-based, time-variable pointing offsets
Prolate spheroid 5x5 pix			
Wide-field effects 10x10 pix 200x200 pix			

Computing, Data Volume, Image Sizes, Memory Use,..



Flagging → Calibration → **Imaging**

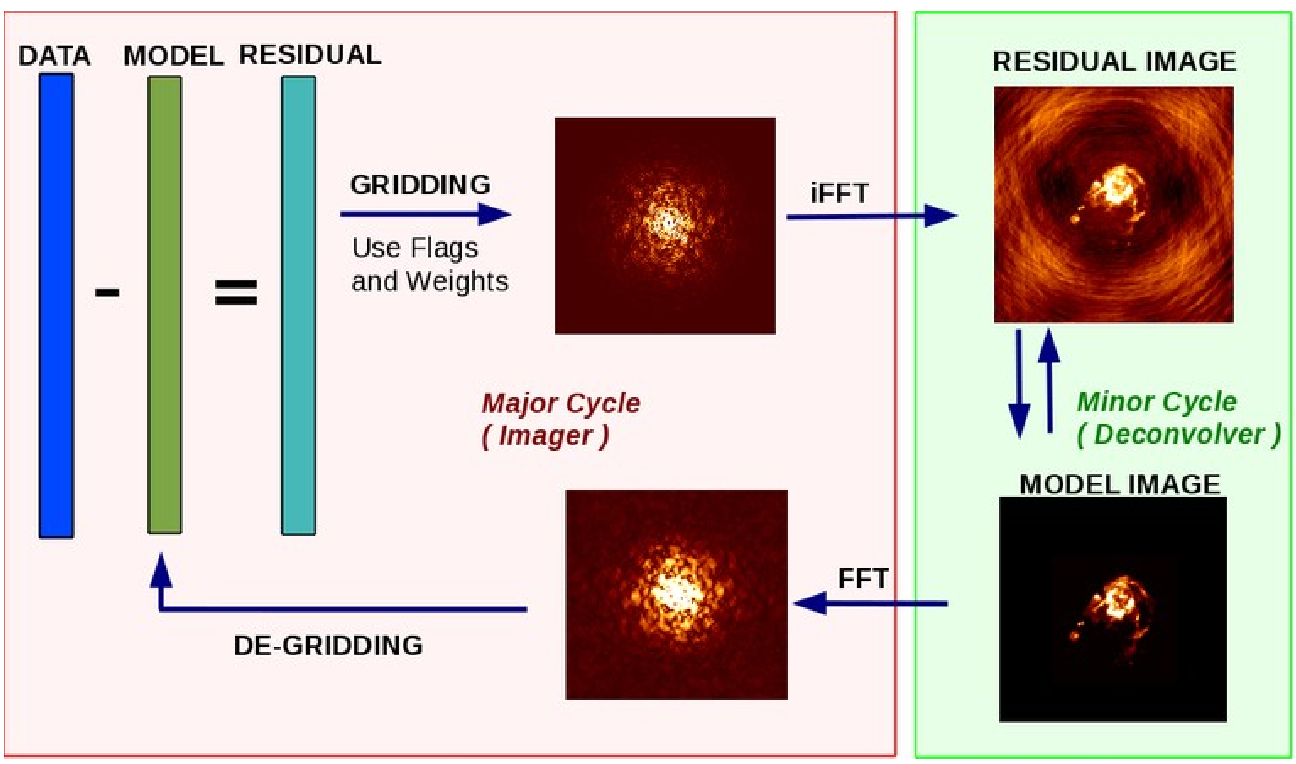
Data volume

N_data =
N_ant^2 x
N_chan x
N_pol x
N_time

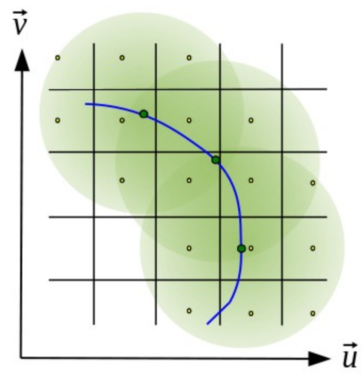
Complex numbers

Lustre I/O

Example :
8hr data
300 GB



Flagging → Calibration → **Imaging**



Gridding : Convolutional resampling

O(N_data) x (nxn) complex multiply/add (n=5 - 100)
=> Compute load : O(N_data) * 10^{2-5} flops

Data parallelization, Multi-threading, GPUs, etc...

Example : Major cycle : 1hr → 10 days (Diff Algorithms)

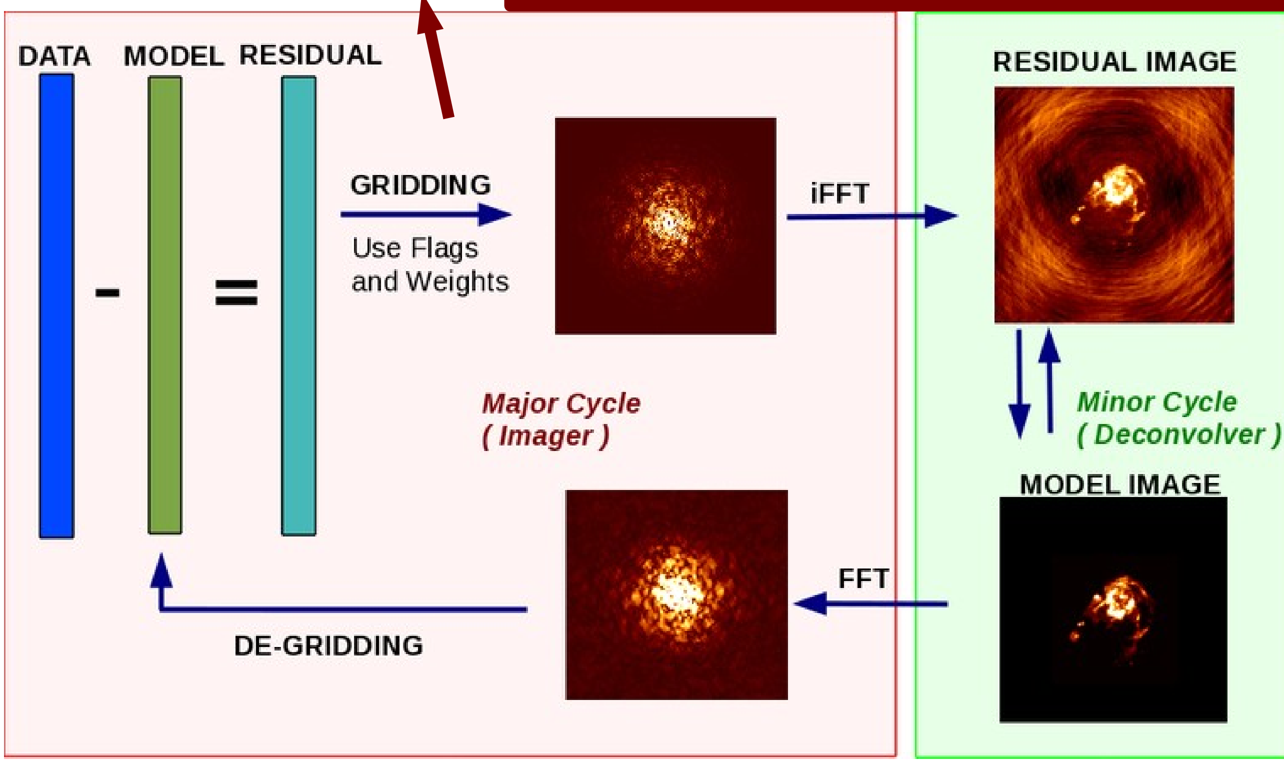
Data volume

N_data =
N_ant^2 x
N_chan x
N_pol x
N_time

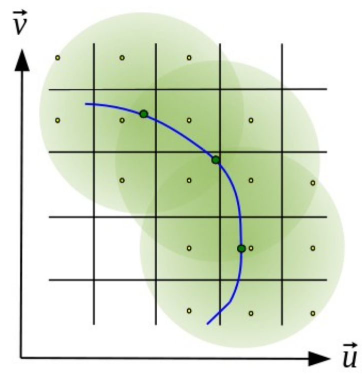
Complex numbers

Lustre I/O

Example :
8hr data
300 GB



Flagging → Calibration → **Imaging**



Gridding : Convolutional resampling

O(N_data) x (nxn) complex multiply/add (n=5 - 100)
=> Compute load : O(N_data) * 10^{2-5} flops

Data parallelization, Multi-threading, GPUs, etc...

Example : Major cycle : 1hr → 10 days (Diff Algorithms)

Data volume

N_data =
N_ant^2 x
N_chan x
N_pol x
N_time

Complex numbers

Lustre I/O

Example :
8hr data
300 GB

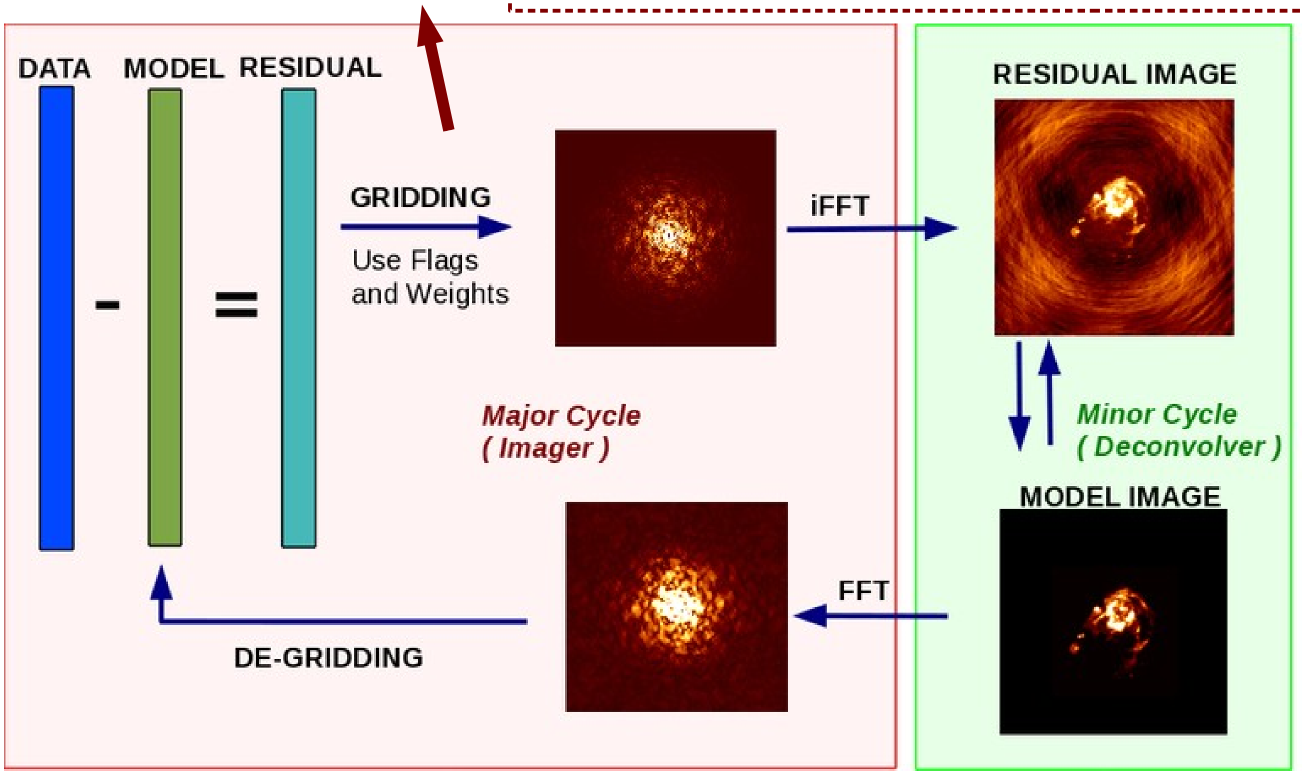


Image sizes

N_pix = Nx x Ny x
N_chan x N_pol

Real / Complex

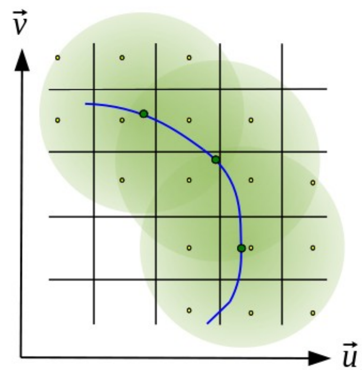
FFTs : O(NlogN)
Pixel math: O(N^2)
Mem : ~8 copies

Multi-threading
Chan parallelization

Nx : 1k → 40k
N_chan : 200 - 16K

Example :
2K x 2K x 1000
~15 GB per image

Flagging → Calibration → **Imaging**



Gridding : Convolutional resampling

O(N_data) x (nxn) complex multiply/add (n=5 - 100)
=> Compute load : O(N_data) * 10^{2-5} flops

Data parallelization, Multi-threading, GPUs, etc...

Example : Major cycle : 1hr → 10 days (Diff Algorithms)

Data volume

N_data =
N_ant^2 x
N_chan x
N_pol x
N_time

Complex numbers

Lustre I/O

Example :
8hr data
300 GB

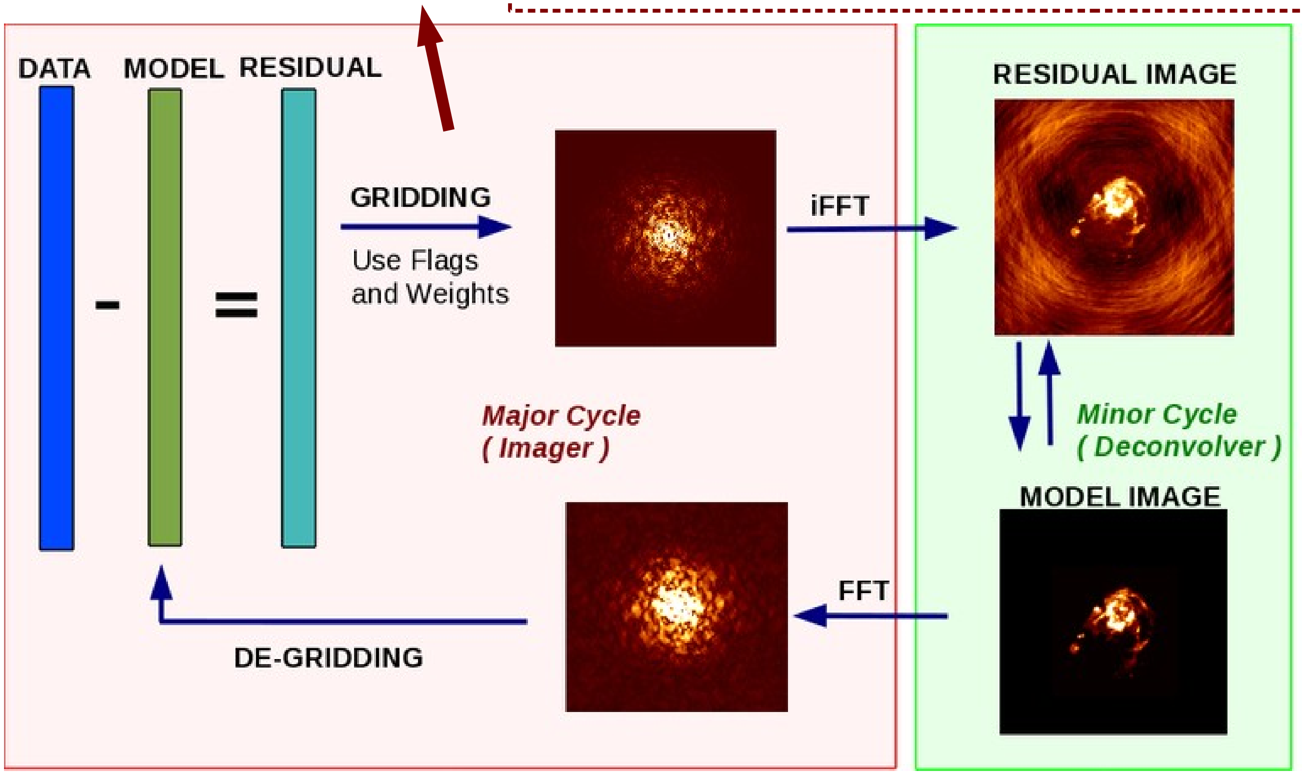


Image sizes

N_pix = Nx x Ny x
N_chan x N_pol

Real / Complex

FFTs : O(NlogN)
Pixel math: O(N^2)
Mem : ~8 copies

Multi-threading
Chan parallelization

Nx : 1k → 40k
N_chan : 200 - 16K

Example :
1K x 1K x 256
~1 GB per image

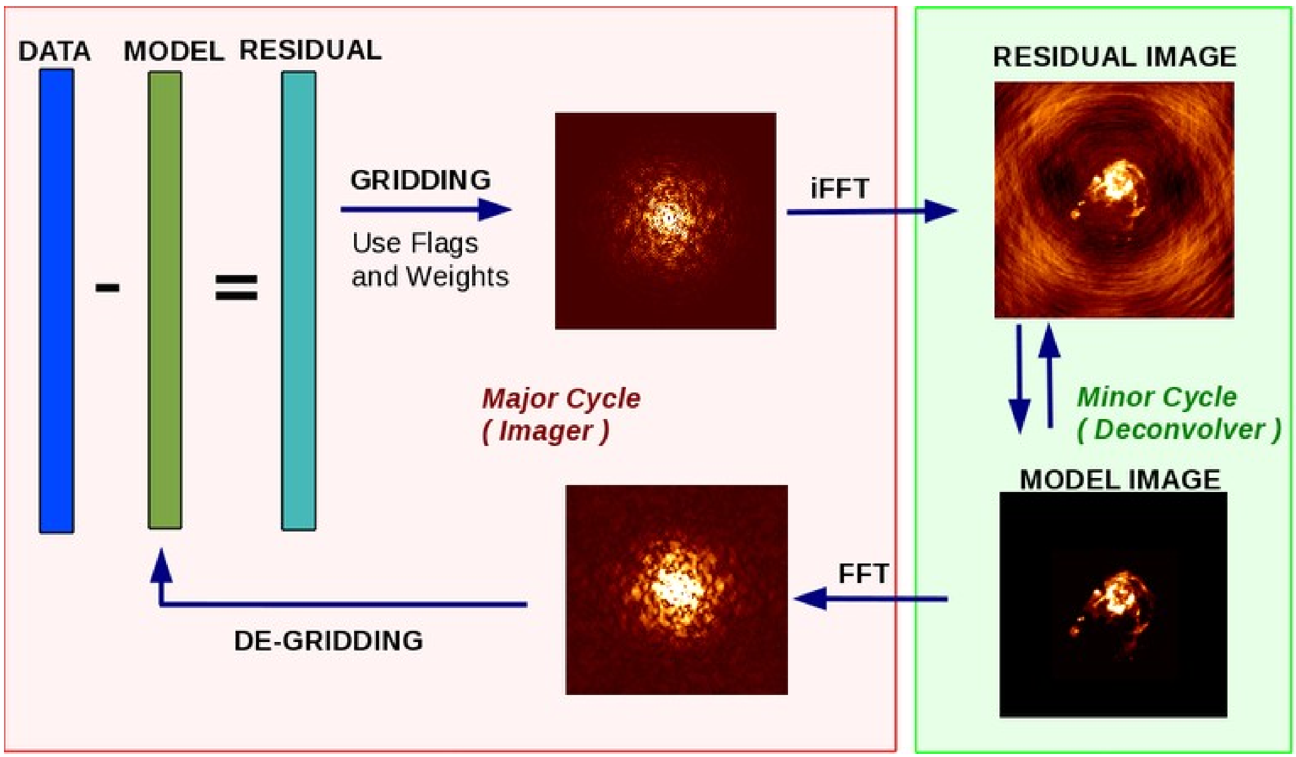
Number of iterations : 5 – 10 major cycle loops
10^2 to 10^4 minor cycle steps

Runtime varies by 1-2 orders of magnitude. Depends on data.

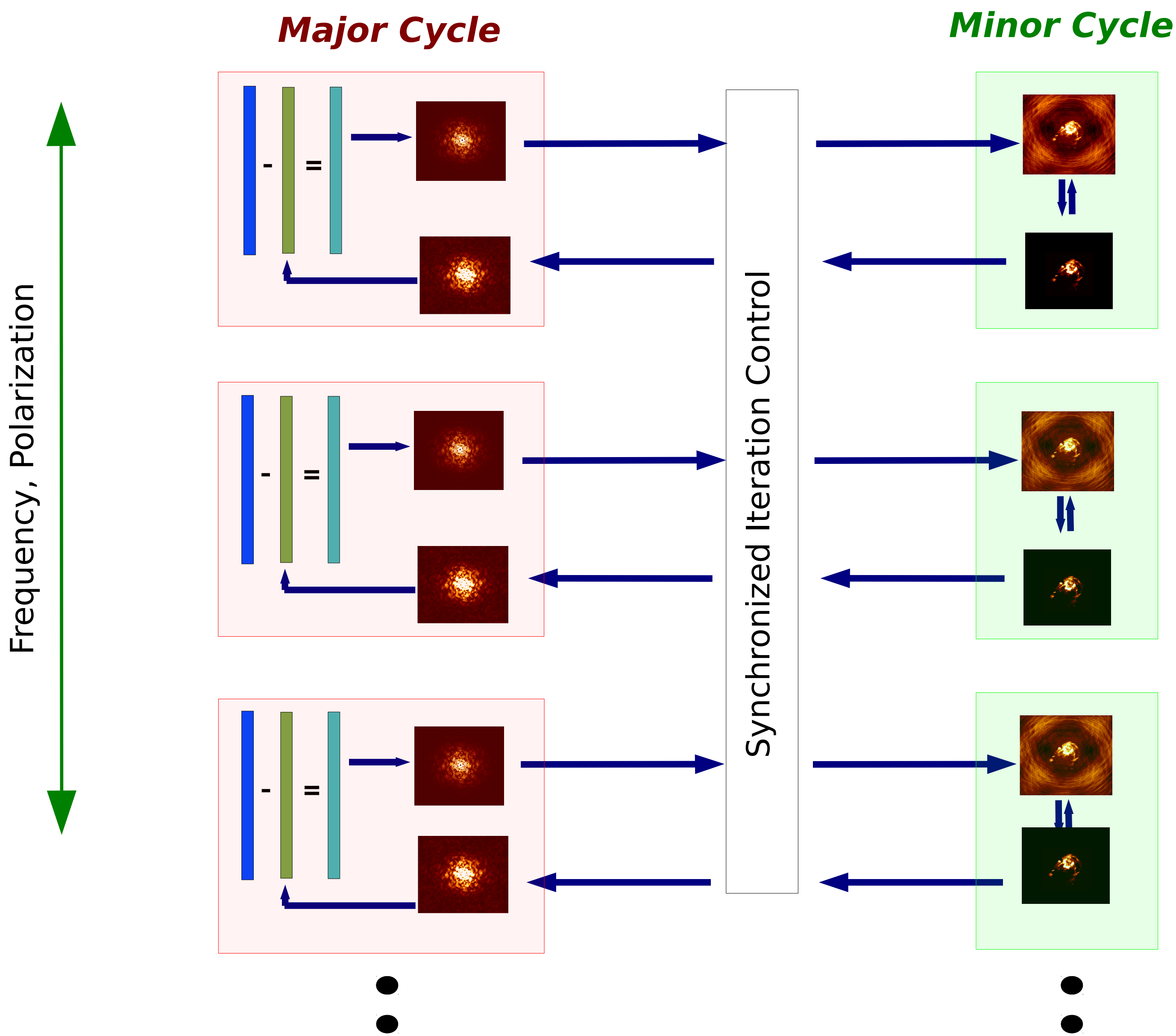
Our current approach to parallelization

- Cube Imaging

- Continuum Imaging



Cube Imaging : Partition along Frequency for Data, Images and Algorithms



Images are 3D or 4D

(X, Y, channel, polarization)

Algorithms can be run independently on each 2D XY plane.

Synchronization is needed for iteration control

Image pixels :
1kx1k to 80kx80k

Spectral channels :
1k to 16K

Pol planes :
1 to 4

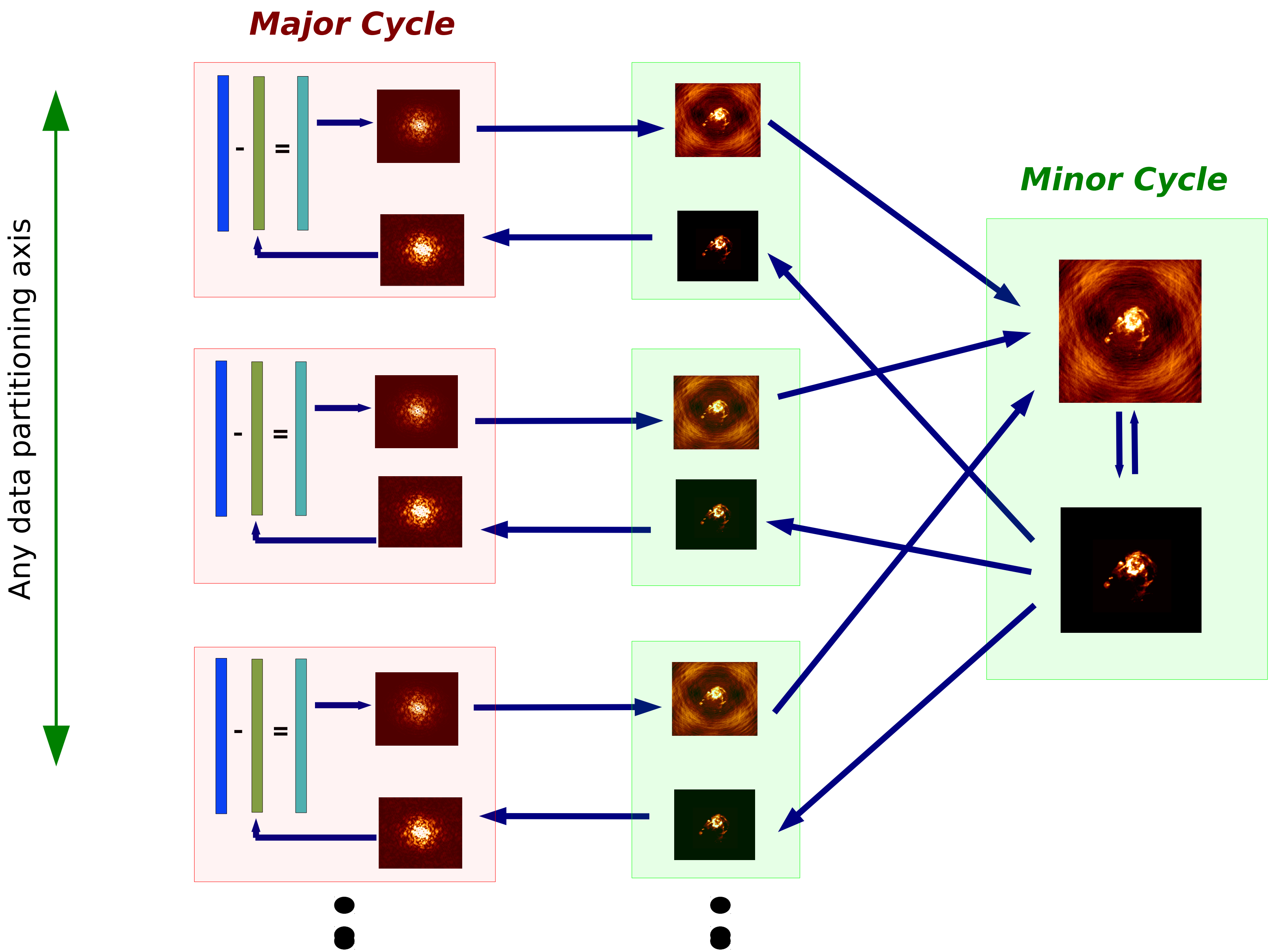
Number of images to access at each iteration :
4 ~ 20

Number of minor cycle iterations :
100 - 100000

Flagging → Calibration → **Imaging**

Continuum imaging :

Partition along any axis for major cycle. Combined minor cycle



Images are 2D (or 3D/4D)

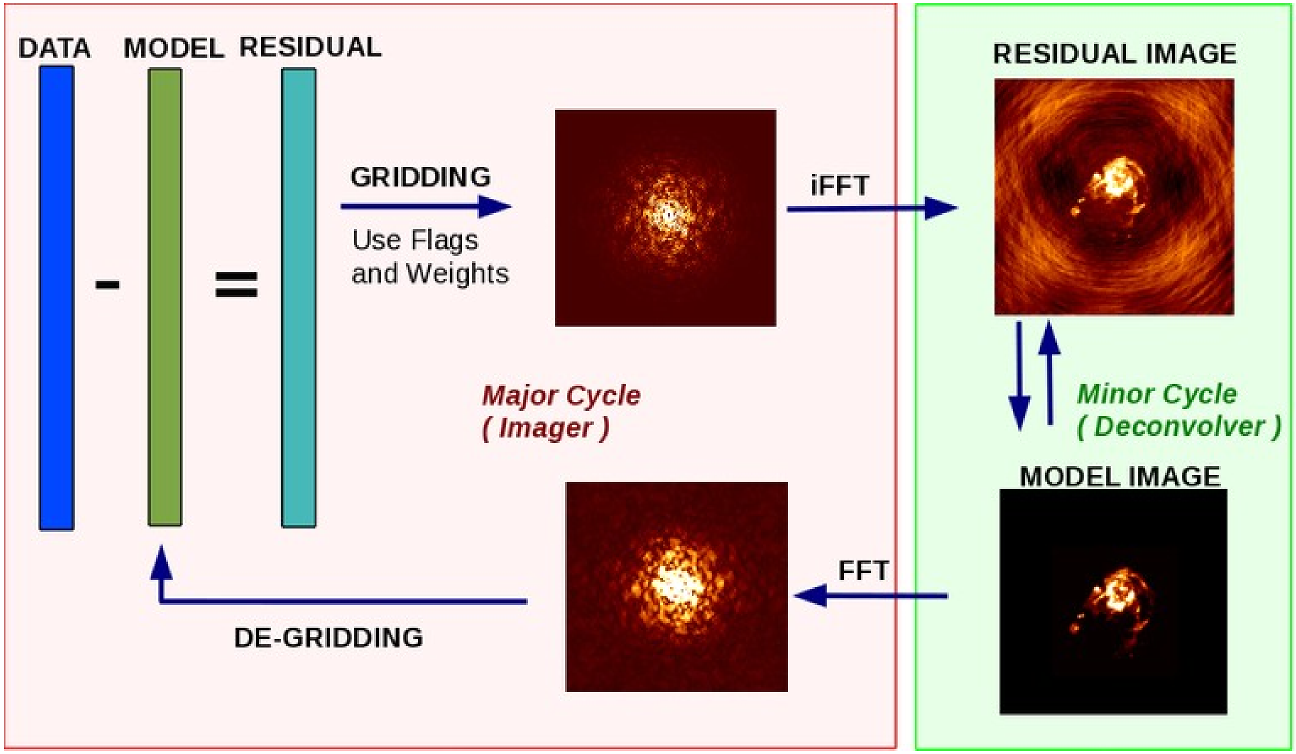
Image pixels : 1kx1k to 80kx80k

Number of images to access at each iteration : 4 ~ 50

Number of minor cycle iterations : 100 - 100000

Multiple Imaging Modes (cube, continuum + others)

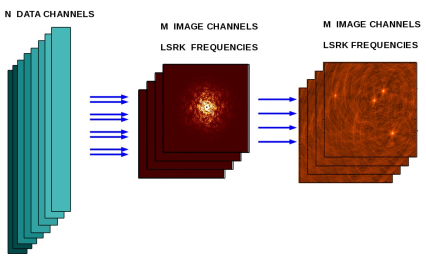
- Depends on target science



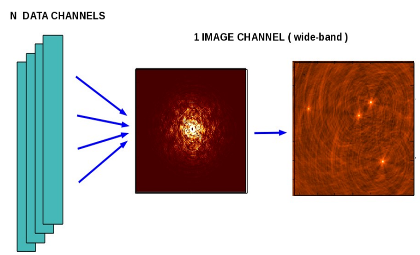
Flagging → Calibration → **Imaging**

Need flexible mapping of data shapes to image shapes (with parallelization)

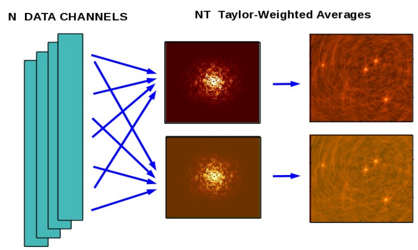
Spectral Cube



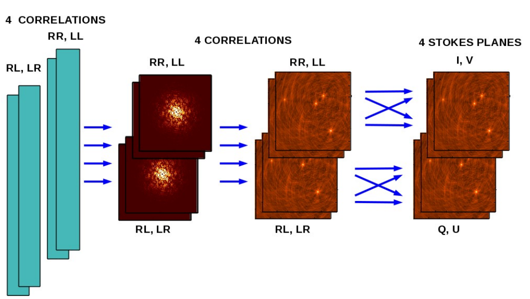
Continuum



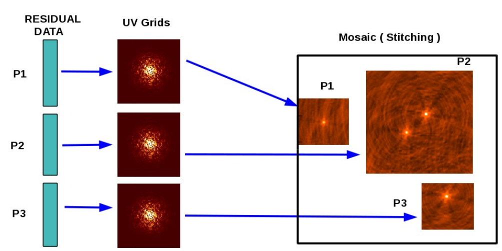
Wideband Continuum



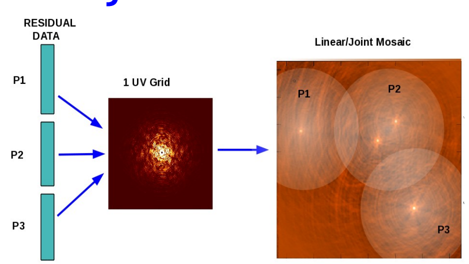
Stokes Planes



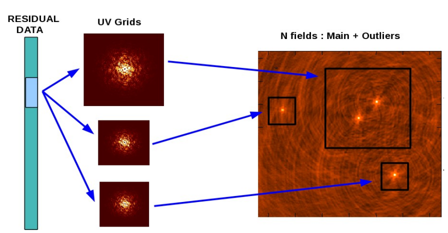
Stitched Mosaic



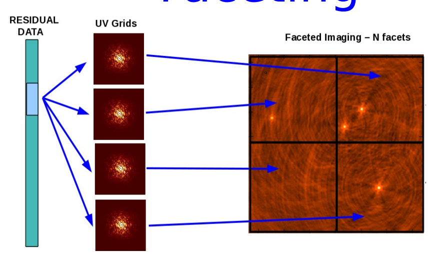
Joint Mosaic



Multi-Field



Faceting



Flagging → Calibration → **Imaging**

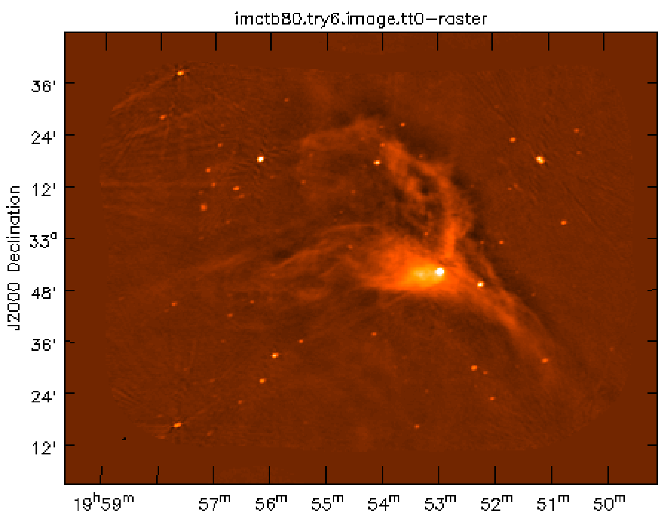
Data combination : Joint imaging using constraints from multiple observations

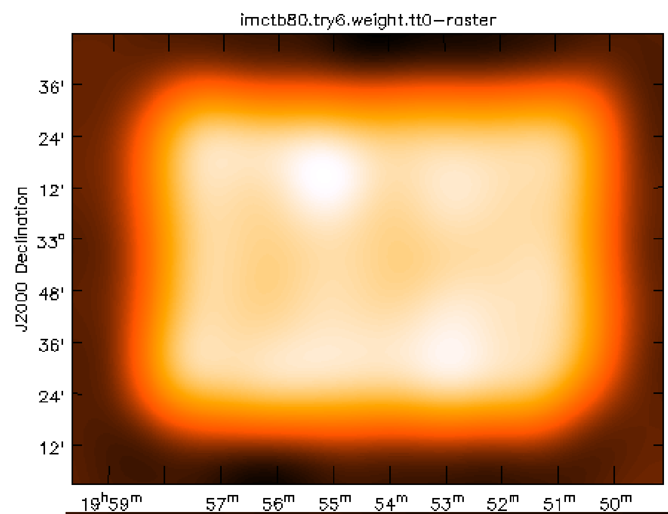
=> Multi-Spectral , Multi-pointing , Wide-field, Multi-instrument

Software : Must parallelize on logical axes and also support repeated combination

Interferometer joint mosaic intensity

(used phase gradients across aperture functions during gridding)

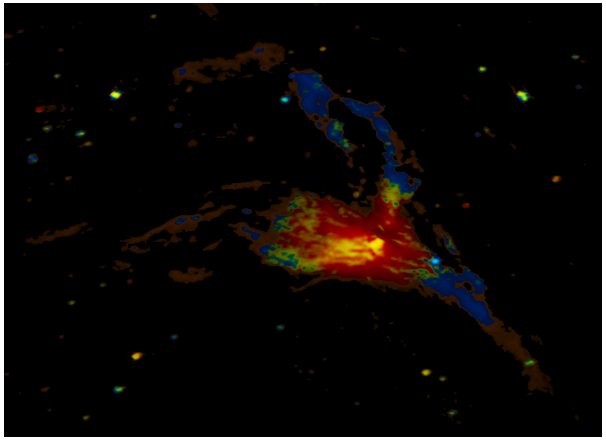


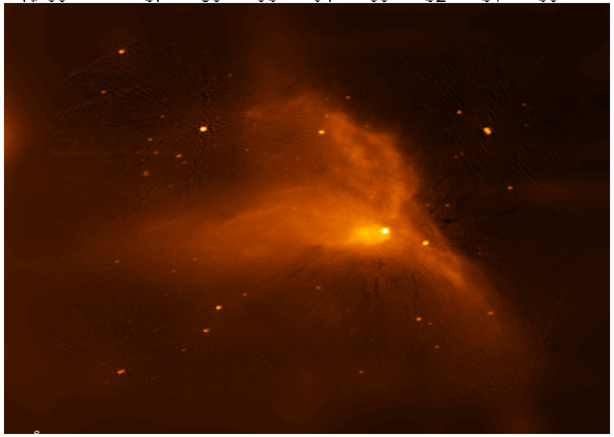


Joint mosaic primary beam from 106 VLA pointings

Wide-Field Spectral Index

(with corrections for PB spectrum)



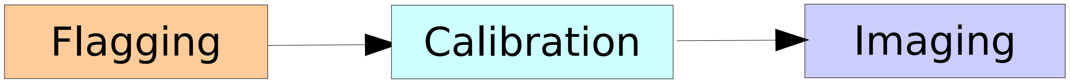


Interferometer + Single dish

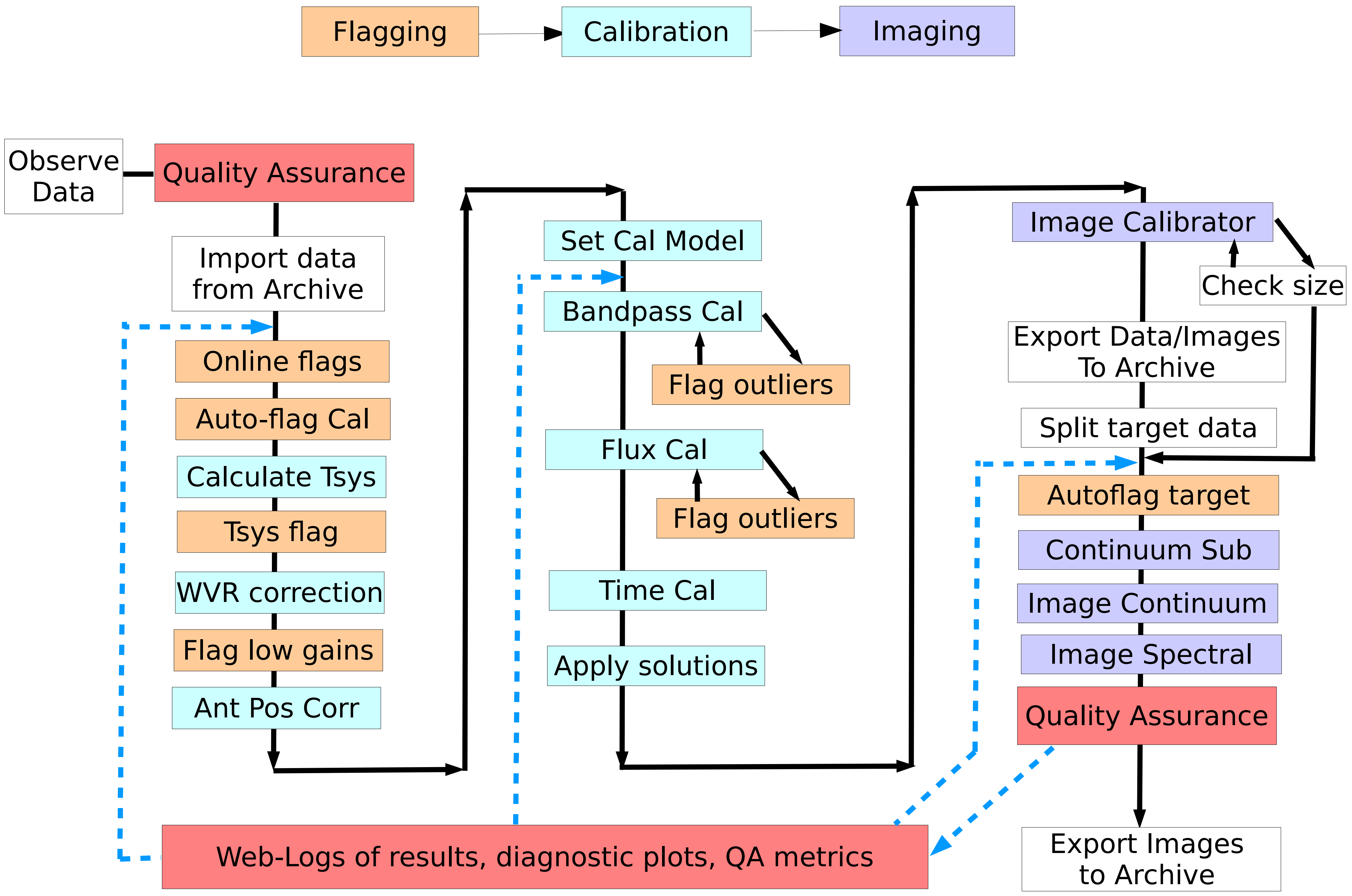
(intensity only)

An example of the current state-of-the-art ...

Science Ready Data Products – Automated Analysis Pipelines



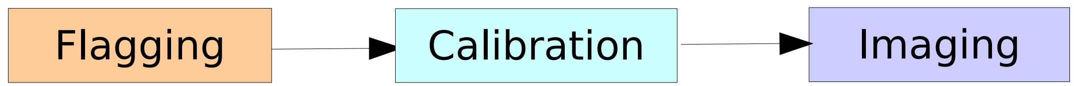
Science Ready Data Products – Automated Analysis Pipelines



Summary

An interferometer is an indirect imaging device.

Image formation :



- Some algorithms are easily parallelized (by data) and some are not
- Each step requires its own data-ordering and access pattern.
- Software must support a large variety of operational modes
  - Multiple axes of parallelization or data-views
  - Combinations of algorithms that require different views
- Total runtime is not always predictable (depends on each dataset)

Later Talks :

- Specific use cases for initial NRAO-CHTC collaboration
- Ongoing work to find better generic solutions
- Current vs Future telescopes to support