Thursday Lecture Series
July – Aug. 2011, Socorro

Lecture 1: Plan, Intro., Deconvolution

July 14, 2011

S. Bhatnagar

Motivation

- Consolidate collective knowledge about processing EVLA data
  - What is possible (focus on post-processing with CASA)
  - What to practically expected
  - Highlight the inevitable significant differences from the VLA-era
- Highlight and disperse information about
  - New techniques required to realize instantaneous sensitivity
  - Issues due to large bandwidth and resulting high instantaneous sensitivity
    - Data volume, RFI
    - Wide-band, wide-field imaging
- Help keep the scientific staff up to date with why and how of post-processing in this era (EVLA and ALMA)
  - Deeper understanding than black-box processing (important, we think, at least for user support)

Plan

• Lecture 1: Basics of imaging and deconvolution		[Bhatnagar]
• Lecture 2: Wide-band imaging,	July 21st	[Rau]
• Lecture 4: Wide-field imaging,	July 28th	[Bhatnagar]
• Lecture 3: Data editing (RFI),	Aug. 4th	[Rau]
or		
Calibration		[Moellenbrock]
• Lecture 5: Mosaicking,	Aug. 11th	[Golap]
• Lecture 6: HPC,	Aug. 18th	[Golap]

- Attempt to keep the plan and content agile with audience feedback
  - More like (moderated) discussion sessions
  - Is this sufficient? Useful? Did we miss something important?
  - Does this help in spreading the information and understanding about EVLA post-processing issues? Among local scientific staff? Among external users via user support group?

Test Data Sets

- During the week, apply and learn more about what we discuss in the lectures
- 3C286 field
  - L-Band, ~1-2 GHz, 30 min. integration,   ?? GB
  - Why: Isolates wide-band issues independent of multi-scale, or PB-correction issues
  - How: What is possible in CASA and how.   What's lacking and coming...
  - Possibly demonstrate strength and weakness of auto-flagging algorithms
- Field with extended emission (Galactic SNR)
  - L-Band, ~1-2 GHz, multi-snapshot (smaller GBs) or several hour integration (larger GBs, appreciation for HPC needs :))
  - Why: Motivates need for wide-field, multi-scale imaging, PB-corrections
  - How: What's possible, what's lacking, what's coming...
- Mosaicking data: (Demo science?)
  - All of the above... and more

Introduction

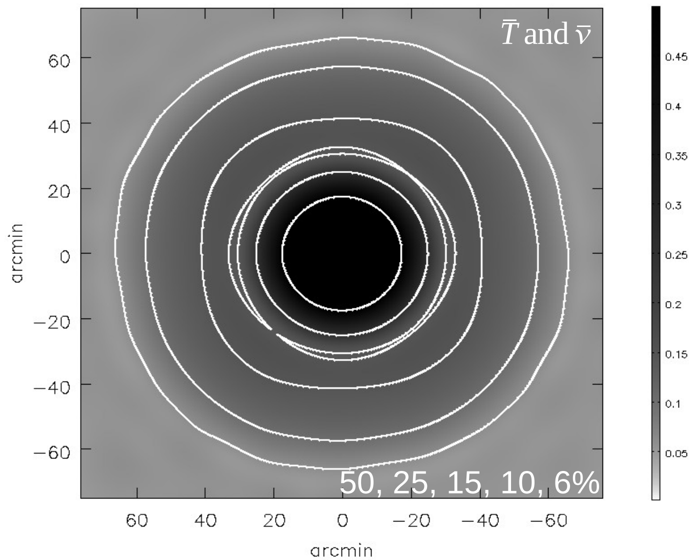
- What's different in the EVLA-era from a post processing point of view
  - Instantaneous 2:1 bandwidth ratio of the EVLA
  - WIDAR capabilities (high time and frequency resolution)
- Large bandwidth provides high instantaneous sensitivity
  - ...and a lot more data....and RFI
  - Wide-field imaging is required
    - Higher bandwidth-smearing, higher time-smearing
  - Time-dependent effects increase in magnitude farther out in the beam
    - PB rotation, Freq. & polarization dependence
  - Wide-band imaging is required
    - Frequency dependent effects (instrumental and sky) become significant
- Larger instantaneous high resolution frequency coverage
  - Data volume, image analysis, visualization
    - Image-data volume can remain large

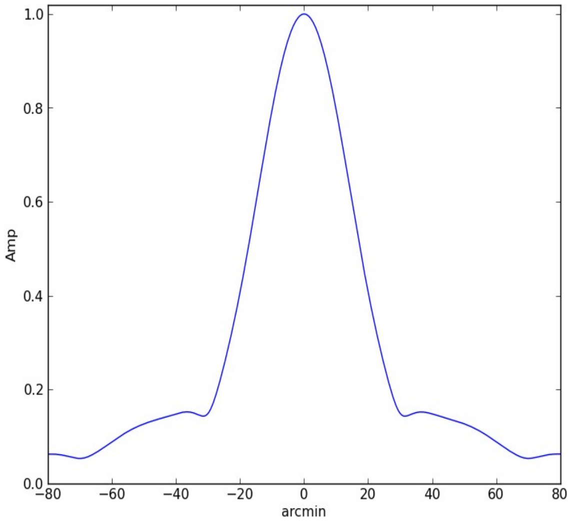
Wide-field Issues

- Large bandwidth provides high instantaneous sensitivity

$$\propto \frac{T_{sys}}{\sqrt{T_{\mathrm{Int}} \Delta\nu}} \qquad \Delta\nu = 1-2GHz$$

- For the same integration time, EVLA is sensitive to emission farther out





- The exact shape of the roll-off depends on the (1) SPWs used, (2) length of observation, and (3) to a lesser extent, on the data weights
- E.g. @ L-Band, PB gain ~1 deg. away can be up to 10%

Wide-field Issues

- Source strength as seen through the wide-band sensitivity pattern

$$S_{eff} = S(r) * PB(r)$$

- E.g. S = 1 Jy, r = 1^o, PB(r) ~ 0.1, S_{eff} ~ 100 mJy

- Error at the center of the image, due to a source of strength S at a distance r

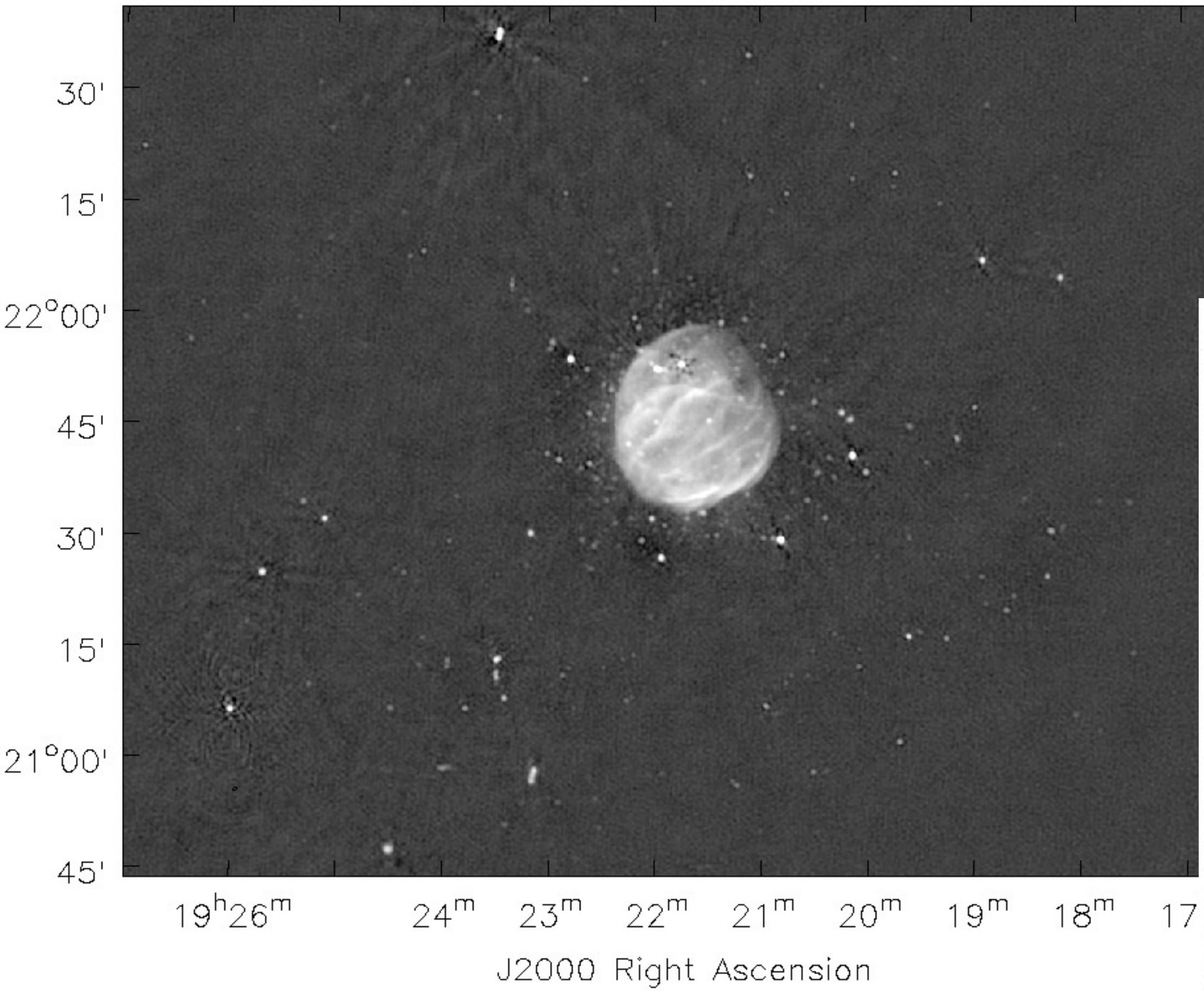
$$\Delta S = S(r) \times PB(r) \times PSF(r)$$

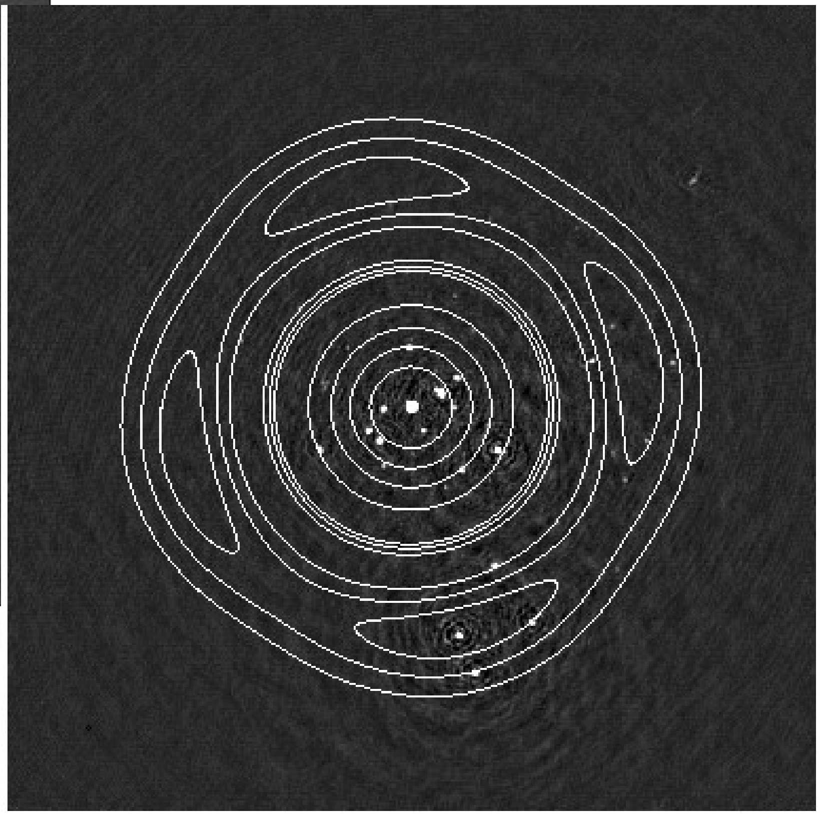
E.g. PSF(1^o) ~ 1.0—0.1%, $\Delta S = 1 mJy - 100\,\mu Jy$

- More precise estimates depends on frequency and time coverage

- Bottom line
  - Noise limited imaging of even compact sources, may need wider field imaging
    - Function of brightness distribution, required dynamic range, max. baseline, bandwidth

Wide-field Issues

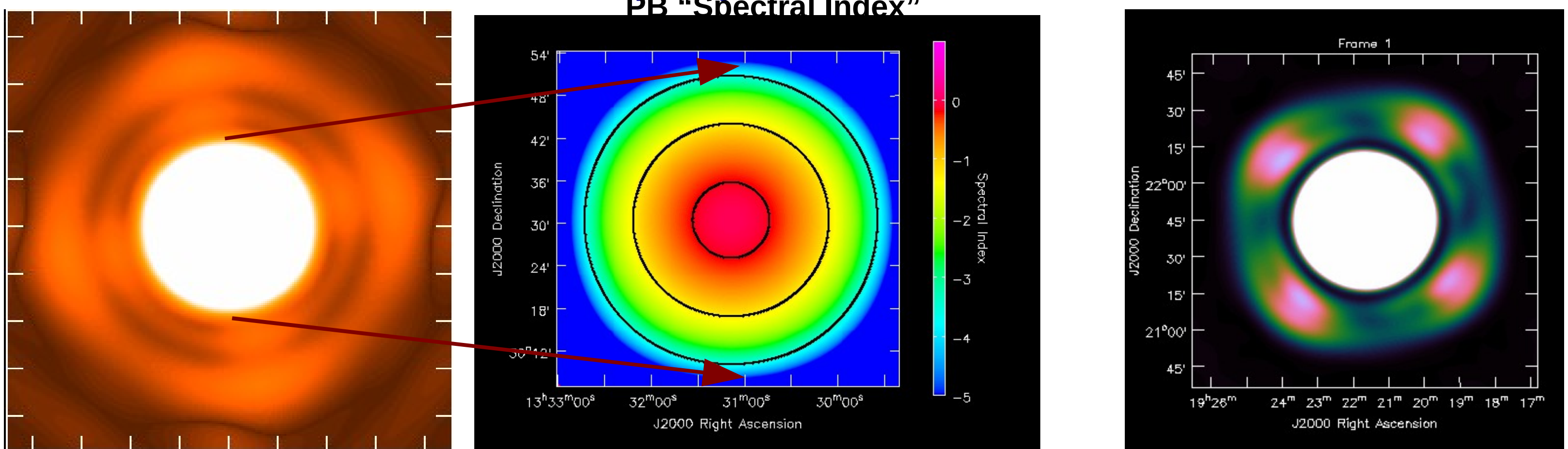




Wide-band Issues

- $I_{Continuum} = \int I_o PB(\nu)(\nu/\nu_o)^{\alpha(\nu)} d\nu\, dt = \int I_o (\nu/\nu_o)^{\alpha_{pb}(\nu,t)+\alpha(\nu)} d\nu\, dt$
- PB “spectral index” due to frequency dependence of the PB
  - Varies with time for rotationally asymmetric PBs

PB “Spectral Index”



- Sky spectral index assumed static in time, but varies with direction
  - Varies with frequency over EVLA bandwidths
- Bottom line
  - Frequency dependence of the sky needs to be accounted for even for the “inner” part of the beam (deconvolution minor cycle)
  - PB frequency and time dependence needs to be accounted for wider field imaging (deconvolution major cycle)

Data volume

- Data Volume is proportional to $N_{baselines}\ N_{channels}\ N_{pol}\ (T_{total}/T_{Int})$
- Loss in amplitude due to bandwidth smearing for continuum imaging

$$Amp.loss \quad \propto \quad \frac{\Delta\nu}{\nu}\frac{FoV}{Resolution}$$

$$\Delta\nu \quad = \quad \frac{\nu_o \eta D}{N_{PBSidelobes} B_{max}}$$

- N=2, B_{max}=D-array, BW=1GHz @ L-Band, 5% tolerance $\Delta\nu = 1\,MHz$
- No. of channels needed = 1000
- Integration time = 1-5 sec due to RFI and time-smearing limits
- Bottom line
  - Large data volume is an inevitable consequence of improved sensitivity
  - More computing cycles using large data is also an inevitable consequence of increased instantaneous sensitivity

Deconvolution

- Process of removing the effects of the emission in one part of the image on another part of the image
  - PSF sidelobes couples distant, otherwise independent pixels
  - Mathematically, even for an image with only multiple point sources, the Hessian is not diagonal (or diagonally dominant)
- Only average quantities are available in the image domain
  - Time and frequency averaging to realize higher sensitivity
  - Averaging across uv-plane
- Purely image-plane based deconvolution applicable only for the static case (along time, frequency and polarization axis)
  - **Hogbom Clean**: Static case, limited by quantization errors
  - **Clark Clean**: Static case + partially handle quantization errors
  - **Cotton-Schwab (CS) Clean**: Static case + handle quantization errors
  - **Multi-Term MFS (minor) + CS-Clean (major)**: Time-static, Freq-dynamic case
  - **Projection (major cycle) + MT-MFS (minor)**: Time- and Freq-dynamic case

Deconvolution

- Data prediction (predict data for the given image)

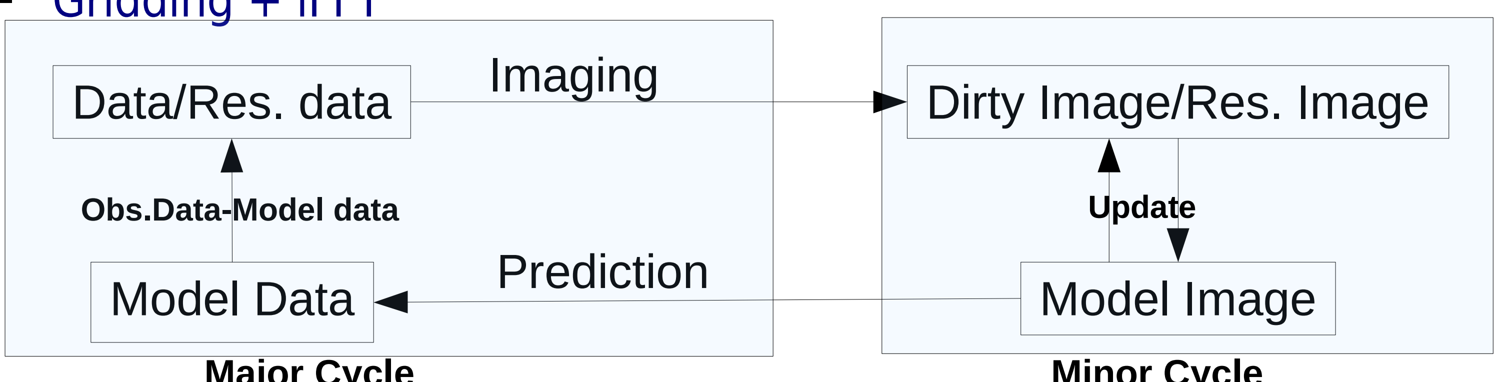
$$\boldsymbol{V = A\, I^{o} + N} \qquad\qquad V_{ij} = deGrid_{ij}\, FT(I)$$

  - A: Linear transform between Image and Visibility domain
    - FFT+ de-Gridding
  - V, I and N are the visibility-, image- and noise-vectors
- Imaging (make an image from the given data)

$$\boldsymbol{A^{T} V = A^{T} A\, I^{o} + A^{T} N} \qquad I^{Dirty} = PSF * I + PSF * Noise$$

$$\boldsymbol{I^{d} = B I^{o} + B N}$$

  - $A^{T}A$: Convolution with the PSF
    - Gridding + iFFT

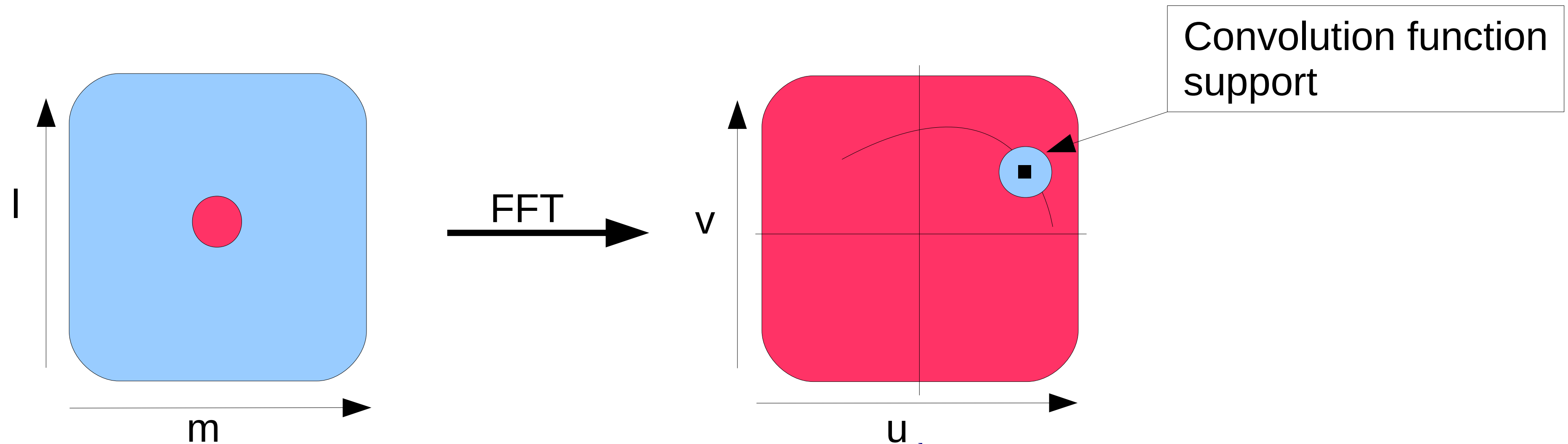


Gridding/de-Gridding (Re-sampling)

- Measurement Equation for a perfectly calibrated interferometer

$$\boldsymbol{V = A I^o + N} \qquad\qquad V_{ij} = deGrid_{ij}\, FT(I)$$

  - A: Linear transform between Image and Visibility domain
    - FFT+ de-Gridding



$$V(u_{ij}) = \sum_{du=-s}^{du=+s} V(u_{ij}+du)\, G(u_{ij}+du)$$

    - **A is the “FTMachine” in casapy-lingo**
    - **ftmachine=”ft” ---> Convolution function = Prolate Spheroidal**

Deconvolution as ChiSq Minimization

- $\boldsymbol{V^M = A\, I^M + N}$ $\qquad V_{ij} = deGrid_{ij}\, FT(I)$
- Linear equation, parametrized by I^{model}
  - However, A is singular!
- Non-linear solver, to solve for the Model Image
  - Compute residuals: $\mathbf{V^{Obs} - AI^M}$ (data domain)

    $\mathbf{I^d - BI^M}$ (image domain)
  - Make Residual Image $\mathbf{I^{res}}$
  - Find update direction: Steepest Descent Algorithm

$$I^c = max\left(-2[I^{Res}]\frac{\partial \chi^2}{\partial\, Param}\right)$$

  - Update model: $I_i^M = T(I_{I-1}^M)$ $\qquad$ *for our discussions this is* $= I_{i-1}^M + \alpha * I_i^c$
  - α Is the loop-gain/step-size. Ideally should be f(Hessian). But typically set to fixed small value (<1.0).

Major Cycle
(always expensive)

Minor Cycle
(can be expensive)

Deconvolution as ChiSq Minimization

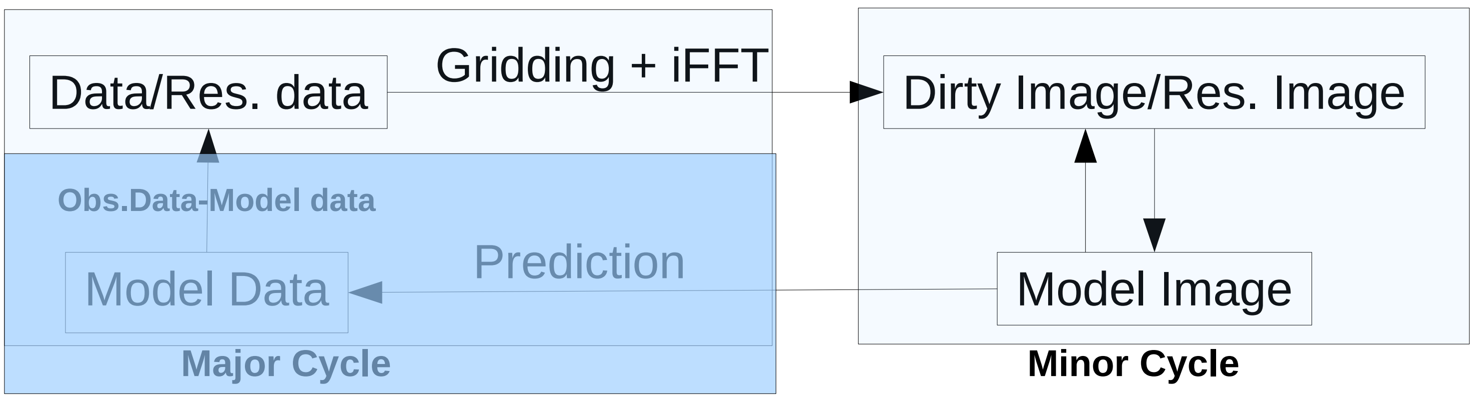
- Gridding/de-Gridding is required to use FFT for Fourier Transform

- **Minor cycle**: Data to Image via Gridding + FFT is inherently inaccurate
  - Aliasing, gridding interpolation errors, quantization (pixelation)
- **Major cycle**: Most often used algorithms control error propagation by periodically computing the residuals at full accuracy in the data domain
  - Computing cost: 2 FFTs + Gridding + de-Gridding
  - I/O cost: full data access per Major cycle

- Most accurate and, for most cases also the most expensive transform is DFT (Direct Fourier Transform)
  - Sometimes used to accurately remove the brightest sources

Hogbom Clean

- Purely image plane deconvolution
- Does not iterate between data and image domain



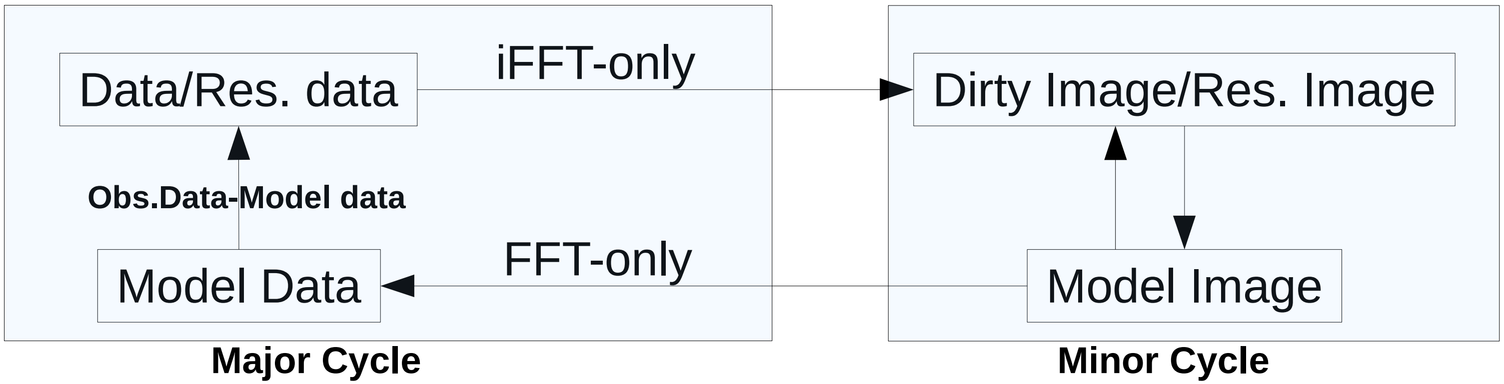
- Uses full PSF to compute residuals in the image domain
- Fastest but least accurate (not useful with modern telescopes)

$$I_i^M = I_{i-1}^M + \alpha\, max\left(I_i^{Res}\right); \qquad I_i^{Res} = I^d - \alpha\, I_i^M$$

- Typically constitutes the Minor Cycle of modern algorithms

Clark Clean

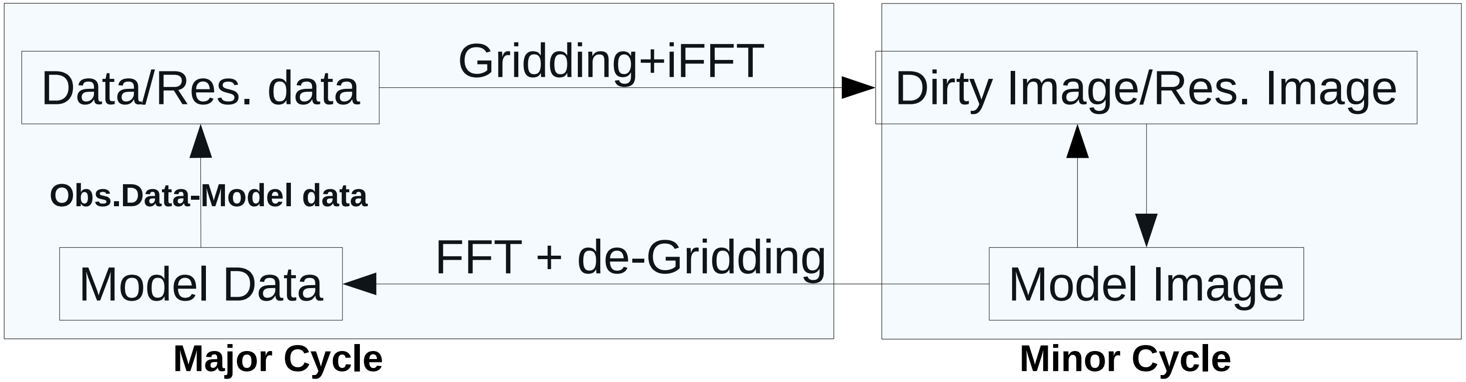
- Compute residual on a grid
  - No Gridding and de-Gridding during major-cycle iterations



- Hogbom Clean-style minor cycle: Uses a truncated PSF
  - Stopping criteria determined by the highest PSF sidelobe and “cyclefactor”
  - Beware of slower convergence with “bad PSFs” (e,g. ALMA 7-antenna)
- More accurate than Hogbom Clean, but not good enough for modern telescopes
- Sometimes used to reduce the number of expensive major cycles

CS Clean

- Compute residual using original data
  - Needs Gridding and de-Gridding during major-cycle iterations



- Most commonly used algorithm
- Every major cycle access the entire data base
  - Significant increase in I/O and computing load
- Assumes, co-planar, time- and freq-independent Measurement Equation
- Cannot account for wide-field wide-band and time variability issues

Major-Minor Cycles

- Minor cycle is usually all in the image domain
  - Always with gridded data (Images or gridded visibilities)
  - The Model Image update step – determines the “Clean Components”
  - The “location cycle”
- Major Cycle computes the residuals in the visibility domain
  - Clark Clean: Uses gridded visibilities
  - CS-Clean: Uses the visibility database
    - Computing cost: gridding and de-gridding + 2FFT
    - I/O cost: full data access per major cycle
- User control on the number of major and minor cycles
  - Minor cycle ends when threshold or max. iterations (“niter”) is achieved
  - A major cycle is triggered if the max. residual > cf*(max. PSF sidelobe)
    - The “cyclefactor” is a user parameter to trigger more or less major cycles
    - Direct control on number of major cycles via interactive mode

MS Clean

- A Minor Cycle algorithm
- Models the sky as a collection of “blobs” (think of tapered Gaussians)
  - More expensive than scale-less algorithms (Hogbom, Clark), but more accurate for extended emission
- Major cycles used, typically in the CS-Clean way
  - DFT prediction possible
- Most commonly used when deconvolution of extended emission is important
  - Also for high dynamic range imaging in the presence of strong compact sources
- Can be combined with major cycle algorithms to account for PB effects, non co-planar arrays (A-, W-, AW-Projection)

- Memory footprint an issue for very large images

Multi-term Clean

- Minor cycle algorithm to account for frequency dependence
- Models the sky as a collection of components whose amplitude follow a polynomial (in frequency for MFS)
  - Computing load scales linearly with number of terms
  - Memory footprint scales linearly with number of terms
- Most accurate where wide-band time-invariant ME is appropriate
- MS-MFS: Uses MS-Clean for each term of the polynomial expansion
  - Computing load and memory footprint scales as $N^2_{terms} * N^2_{scales}$
- Can use any of the major cycle algorithms to account for time variability, non co-planar issues (A-, W-, AW-Projection)
- Minor cycle can be as expensive as major cycles
- Memory footprint a bottleneck for large images

Projection Algorithms for Major Cycle

- Incorporates direction dependent effects as part of the gridding function **(ftmachine = "?project")**
  - $V_{ij} = A_{ij} I^o + N_{ij}$
  - Construct D, such that $D_{ij}^T A_{ij} \approx 1$
  - Imaging $D_{ij}^T V = D_{ij}^T A_{ij} I^o + D_{ij}^T A_{ij} N_{ij}$
  - Prediction: $V_{ij}^M = D_{ij} I^M$
- W-Projection: $D = FT[e^{2\mathrm{pi}\,\iota\sqrt{w-1}}]$ ftmacine="wproject"
- A-Projection: $D_{ij} = FT[PB_{ij}(t, \nu, pol)]$
- AW-Projection: $D_{ij} = A_{ij} * W$
  - High frequency imaging require A-only Projection
  - Low frequency imaging requires AW-Projection
    - Need to sample the frequency and W-axis separately
- Larger convolution kernel support: Increases computing load

Summary: Controlling the run-time

- User parameters that impact total run-time
  - No. of minor cycle iterations ("niter")
  - Loop gain:
    - A function of how bad is the PSF. Too small increases run-time and stability
    - Can (should?) be increased (0.5 – 0.7) for more expensive MS-, MS-MFS without sacrificing stability
  - Cycle factor: Controls when a major cycle is triggered
  - No. of scales in MS-Clean: Computing and memory foot print scales linearly with no. of scales
  - No. of terms in MT-MFS: Computing and memory footprint scales as square of no. of Taylor terms
    - Depends on how complex is the frequency dependence of the data
  - No. of W-Planes: Weak impact on run-time for W-only Projection
    - Strong dependence on run-time for AW-Projection
  - No. PA-steps: Weak impact on run-time for A- or AW-Projection for EVLA/ALMA