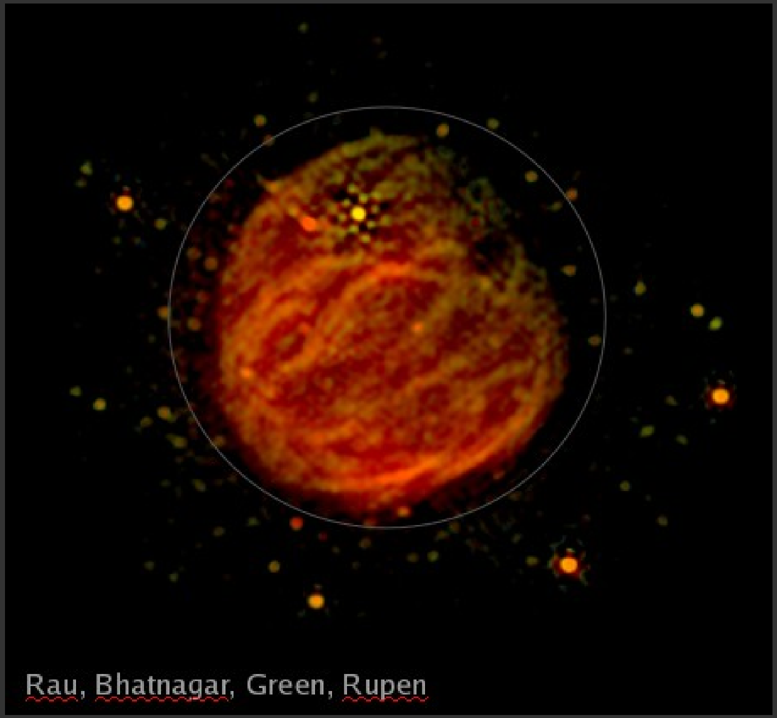
# A systems approach to calibration and imaging

**Feb. 12th, 2021**


Rau, Bhatnagar, Green, Rupen

S. Bhatnagar

Algorithms R&D Group

# Plan for the talk

- Introduction

- Theoretical framework for the calibration of, and imaging with interferometric radio telescopes
    - Constraints vs Parameters
    - Parameter separation: sky vs instrumental vs hybrid

- Principles to navigate the space of algorithms–software–scientific requirements
    - Algorithm architecture
    - Best vs Good-enough
    - Affordable vs desired (or Astronomy/AstroPh vs Computing focus)

- Process of estimating size-of-computing

# Introduction

- NRAO
  - A NSF funded national observatory
  - To build and operate large radio astronomy facilities
    - VLA (~300MHz + 1-50GHz), ALMA (~100-900GHz), VLBA (1000-Km baseline class), GBT (single dish)
  - Central Development Lab (CDL), CV
  - Scientific software : CASA (AIPS)
    - Open source, widely used in RA community world-wide
    - Runs on laptops/desktops, clusters, external clusters ("cloud" – AWS)
    - In progress: Use of dis-joined computers, fine(r)-grain parallelism Massively parallel/special h/w

- Sanjay Bhatnagar
  - Scientist at NRAO, Socorro, NM, USA
  - Head of the NRAO Algorithms R&D Group (ARDG)
    - Full-time scientific staff, Senior Software Engineers, Postdoc, Students. In-kind contributions to/from CASA
    - Collaborations: ngVLA, CASA, MeerKAT/IDIA, NRAO-SCG, CHTC, UW-Madison, Sandia Labs/Los Alamos, NVIDIA

# The Measurement Equation

- ME for full-pol treatment including DI and DD effects

$$\overrightarrow{V}_{ij}^{Obs} = G_{ij} w_{ij} \int PB_{ij}(\vec{s}) \quad I(\vec{s}) \quad e^{\iota(\vec{s}.\vec{b}_{ij})} d\vec{s} \quad + \quad n_{ij}$$

  - $G_{ij}$ is the DI, multiplicative gain.
  - $PB_{ij}$ is the DD antenna-based (complex) gain.
  - $n_{ij}$ is the additive measurement noise.

  - $I(\vec{s})$, $PB_{ij}$, $G_{ij}$ are the unknowns
  - $\overrightarrow{V}_{ij}^{obs}$ and the noise model are the only measured constraints.

- Imaging is the process of solving for a best-fit model for $I(\vec{s})$

- Calibration is the process of solving for a best-fit model for $PB_{ij}$ and $G_{ij}$

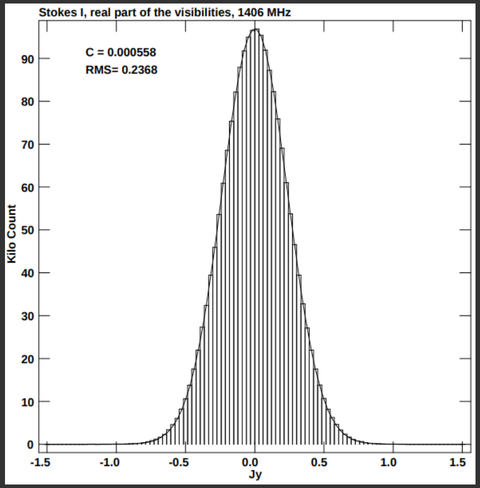
# The Measurement Equation

- Re-writing the ME in the data domain

$$\overrightarrow{V_{ij}^{Obs}} = w_{ij} G_{ij} . \boldsymbol{F} I \quad + \quad n_{ij} \qquad \text{DI only}$$

$$\overrightarrow{V_{ij}^{Obs}} = w_{ij} \left[ G_{ij} A_{ij} \right] * \boldsymbol{F} I \quad + \quad n_{ij} \qquad \text{DI + DD}$$

- Noise is normally distributed, provided
  - the entire signal chain is operating in the linear regime
    - » A strong h/w design constrain
  - post processing maintains the Hermitian property of the integral
    - » A strong algorithm/software design constrain (or at least should be!)



Momjian, EVLA Memo 204, 2017

# Parameterization

- As with *any* modeling, optimal results are when the models are expressed in their most compact form
  - Minimum possible parameters
  - Smallest extent in the domain
- $I(\vec{s})$ in the image domain
  - Amplitude, shape and location
  - Frequency, polarization (and time) dependence
  - Basis functions that minimize the number of terms
- $A_{ij}$ and $G_{ij}$ in the visibility domain
  - Separable as antenna-based parameters

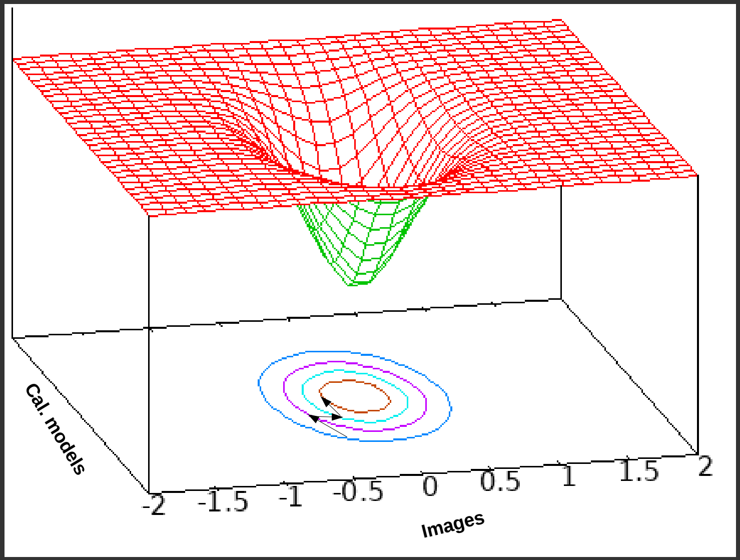
$$\boldsymbol{A}_{ij} = \boldsymbol{A}_i \otimes \boldsymbol{A}_j^T \qquad \otimes \equiv \text{Outer convolution}$$

  - Optimal SNR per parameter

- Any part of the signal not separable as antenna-based quantities is either an un-modelled sky signal or a coupling in the hardware
  - Usually a powerful discriminator of the sky signal

# Post processing

- Top-level view: Optimization in a 2D space of plausible images and calibration models
  - Imaging + SelfCal assumes an orthogonal space.
    - Solve for I, keeping calibration model fixed
    - Solve for calibration model, keeping I fixed



- Joint optimization possible
  - Probably necessary for high-DR science

# Numerical optimization

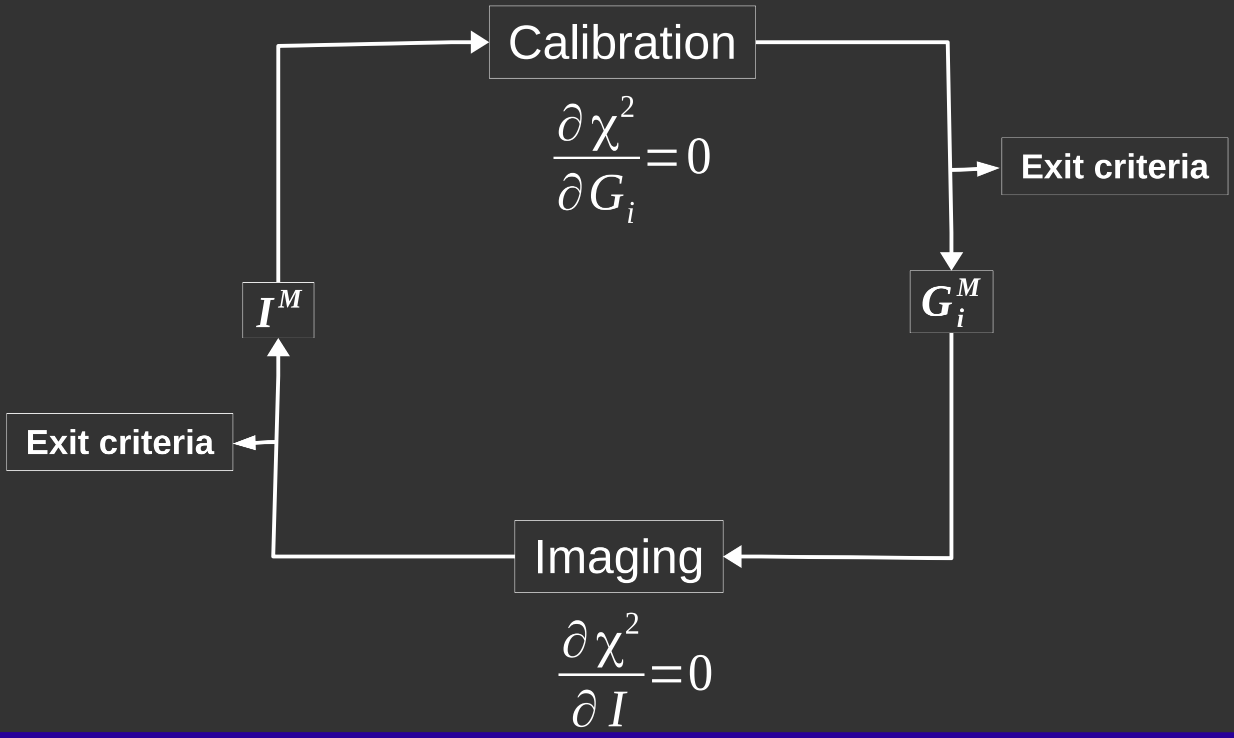
- Imaging itself is fundamentally iterative
  - Hessian is not positive definite
  - High-D optimization problem
    - Algorithms that reduce the dimensionality give better performance

- In general, a wide-band full-polarization treatment necessary
  - VLA, ALMA, ngVLA: Hybrid Mueller treatment is sufficient.
  - Full Mueller may be necessary for receivers with strong, time-variable leakages

# Calibration-Imaging Iterations

- The Chisq function is the optimal estimator for normal distribution
  - A variety of numerical optimization to choose from (many discussed in the RCI series). Differences in the results are insignificant.
  - Complexity (run-time, algorithmic, software) perhaps a better discriminator.

$$\chi^2 = \sum_{i,j} \left|\overrightarrow{V^{Obs}_{ij}} - w_{ij}\left[G_{ij} A_{ij}\right] * \boldsymbol{F}\, I\right|^2$$



Calibration

$$\frac{\partial \chi^2}{\partial G_i} = 0$$

Exit criteria

$$\boldsymbol{G}_i^M$$

Imaging

$$\frac{\partial \chi^2}{\partial I} = 0$$

Exit criteria

$$\boldsymbol{I}^M$$

# DI Calibration model

- Solving the set of equations

$$\frac{\partial \chi^2}{\partial G_i}=0 \quad \text{where } G_i = J_i \otimes J_j^T \quad , \quad J_i = \begin{bmatrix} J_p & J_{p\to q} \\ J_{q\to p} & J_q \end{bmatrix}_i$$

$$J_i^n = J_i^{n-1} + \alpha\left[\frac{\sum_{j,j\neq i} X_{ij} w_{ij} J_j^{n-1}}{\sum_{j,j\neq i} w_{ij} |J_j^{n-1}|^2} - J_i^{n-1}\right] \qquad X_{ij} = V_{ij}^{M^{-1}} V_{ij}$$

**Hamaker et al. notation**

  - Real-time solvers possible (and in use!) with *simple, low-order* predictors.
  - Implementations robust to data corruption.  RFI detection

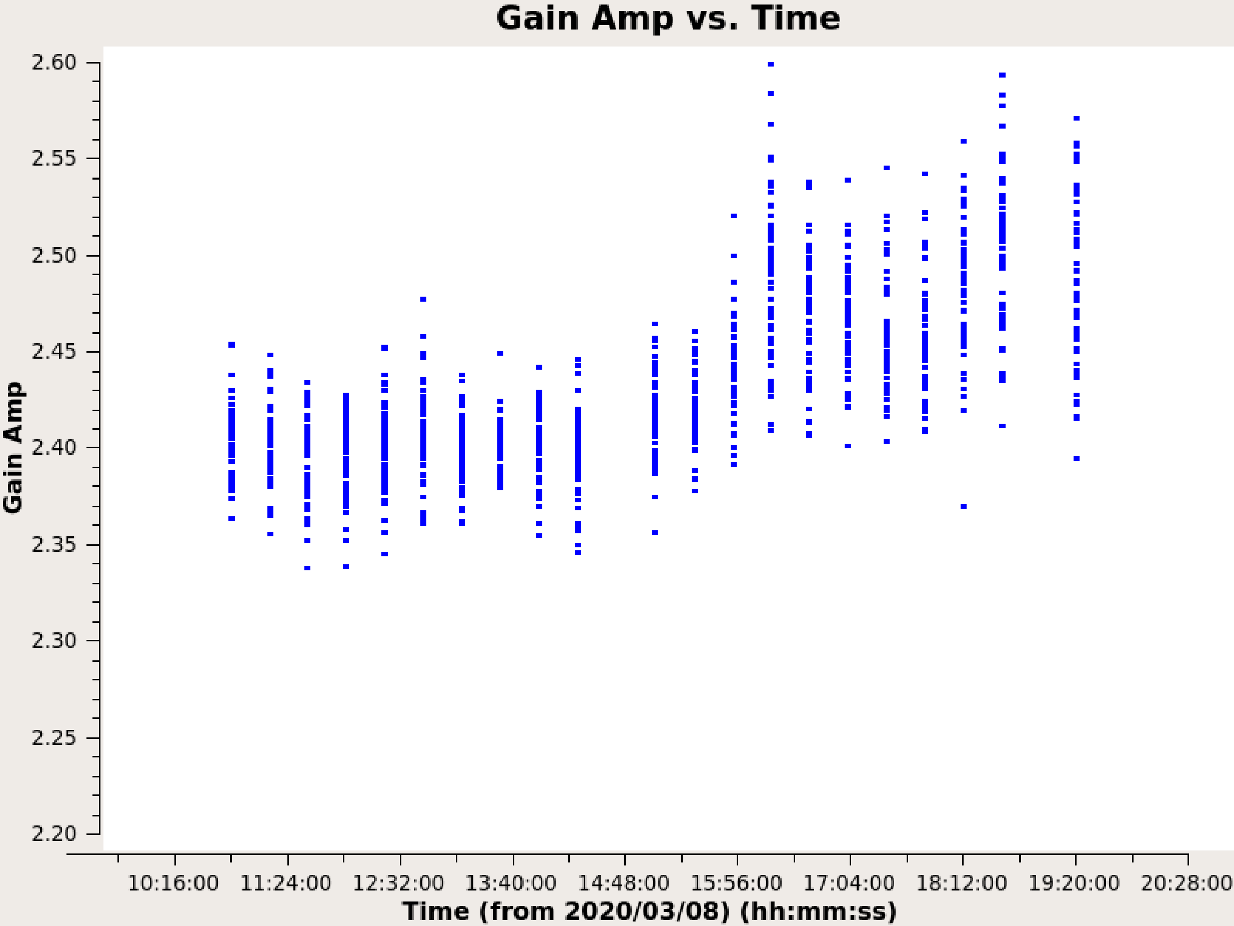
- For insight, above can be derived from weighted average of closure quantities.
      - A note by Sri Kulkarni?  Or Cornwell?

**Thompson A. R. and Daddario L. R. 1982 RaSc 17 357**

http://www.aoc.nrao.edu/~sbhatnag/GMRT_Offline/antsol/antsol.html

# DI Calibration model

- Slowly varying functions
  - Real-time solvers possible with *simple, low-order* predictors.



**Thompson A. R. and Daddario L. R. 1982 RaSc 17 357**

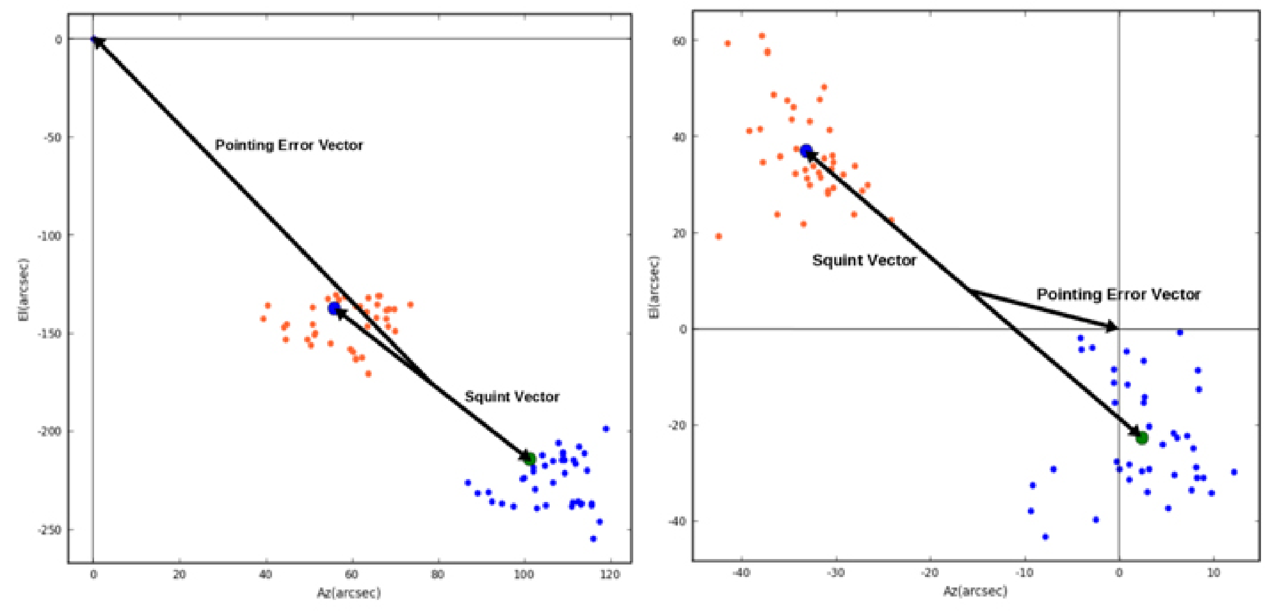
http://www.aoc.nrao.edu/~sbhatnag/GMRT_Offline/antsol/antsol.html

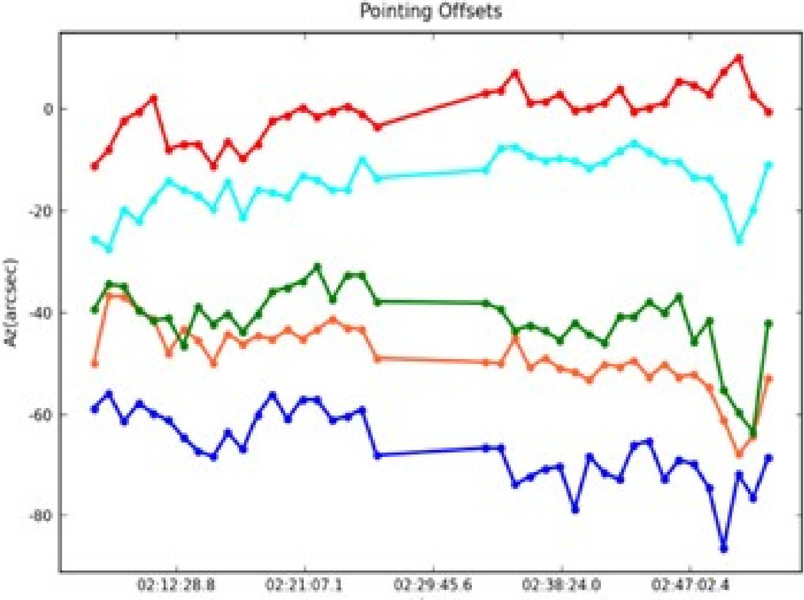
# DD Calibration model – I: Pointing SelfCal

- $\chi^2 = \sum_{i,j} |\overrightarrow{V_{ij}^{Obs}} - w_{ij}\left[A_{ij}\right] * \boldsymbol{F}\, I|^2$ where $A_{ij} = A_i \otimes A_j^T$ $(\otimes \equiv \text{outer convolution})$

Parameterize $A_i$ for antenna aperture (shape, pointing offsets,...). Then solve

$$\frac{\partial \chi^2}{\partial A_i}\frac{\partial A_i}{\partial a_i} = 0$$

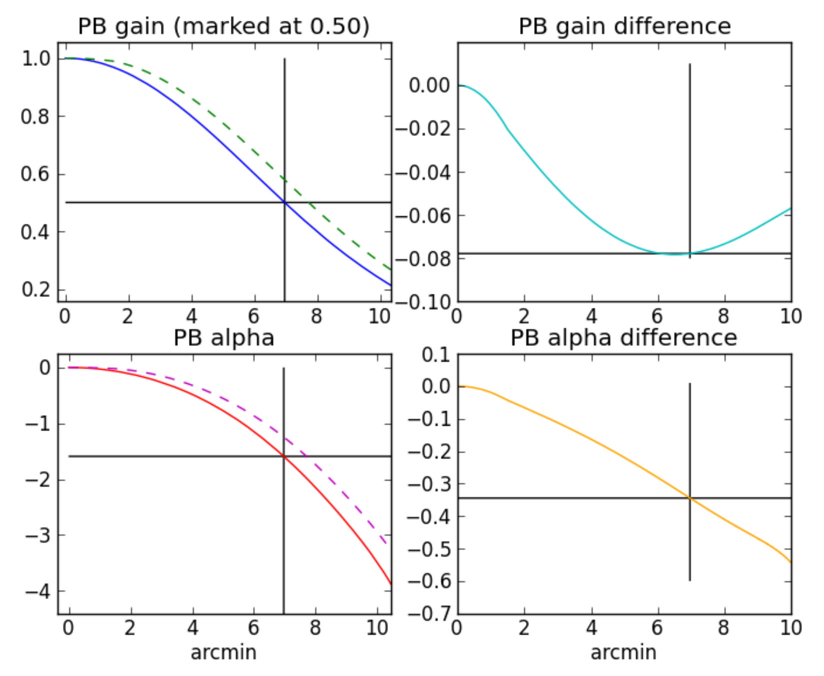


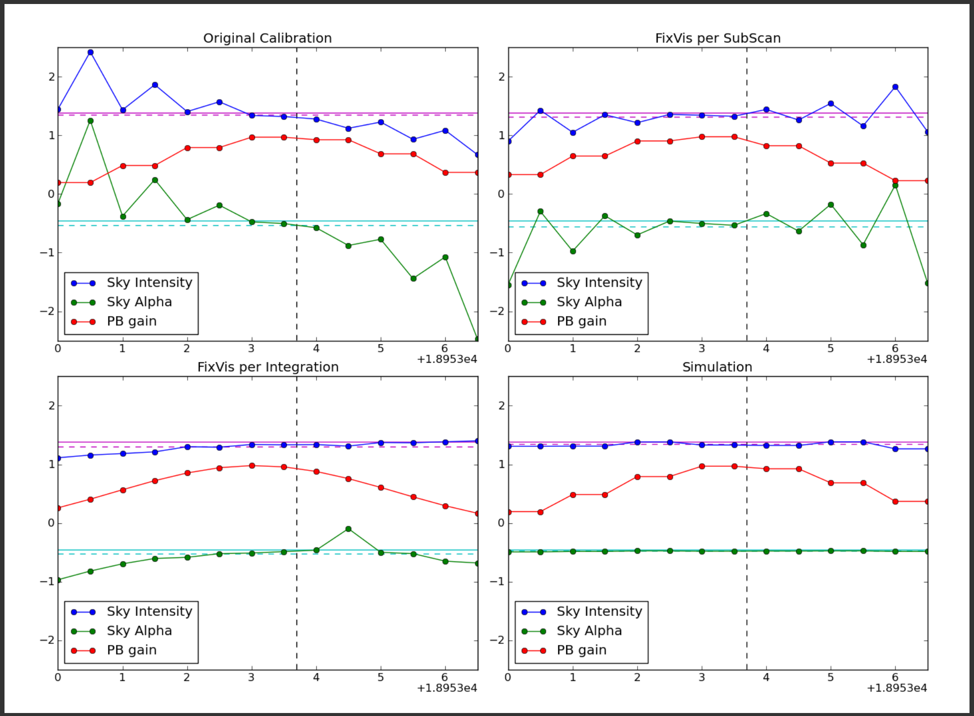


S. Bhatnagar and T. J. Cornwell 2017 AJ 154 197

# DD Calibration model - II

- Pointing matters! Pointing correction in wide-band mosaic imaging with the VLA



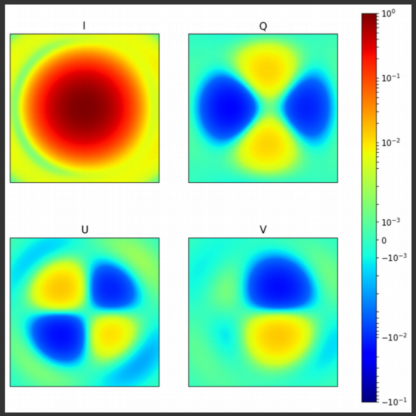


**U. Rau & S. Bhatnagar, ARDG Memo, 2018**

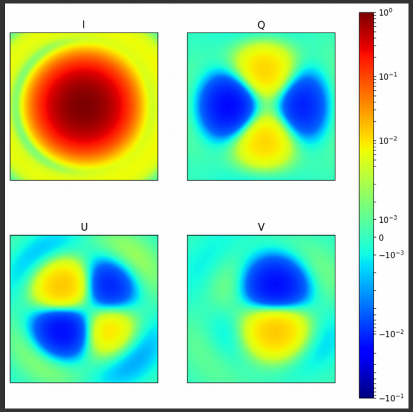
https://safe.nrao.edu/wiki/pub/Software/Algorithms/WebHome/ARDG_VLASS_Imaging_Report_v2.pdf

# DD Calibration model - III

- Solvers for shape, DD leakage, time-, elevation-variability etc. are possible (and coming).



**Measurement**



**Parametric Model**

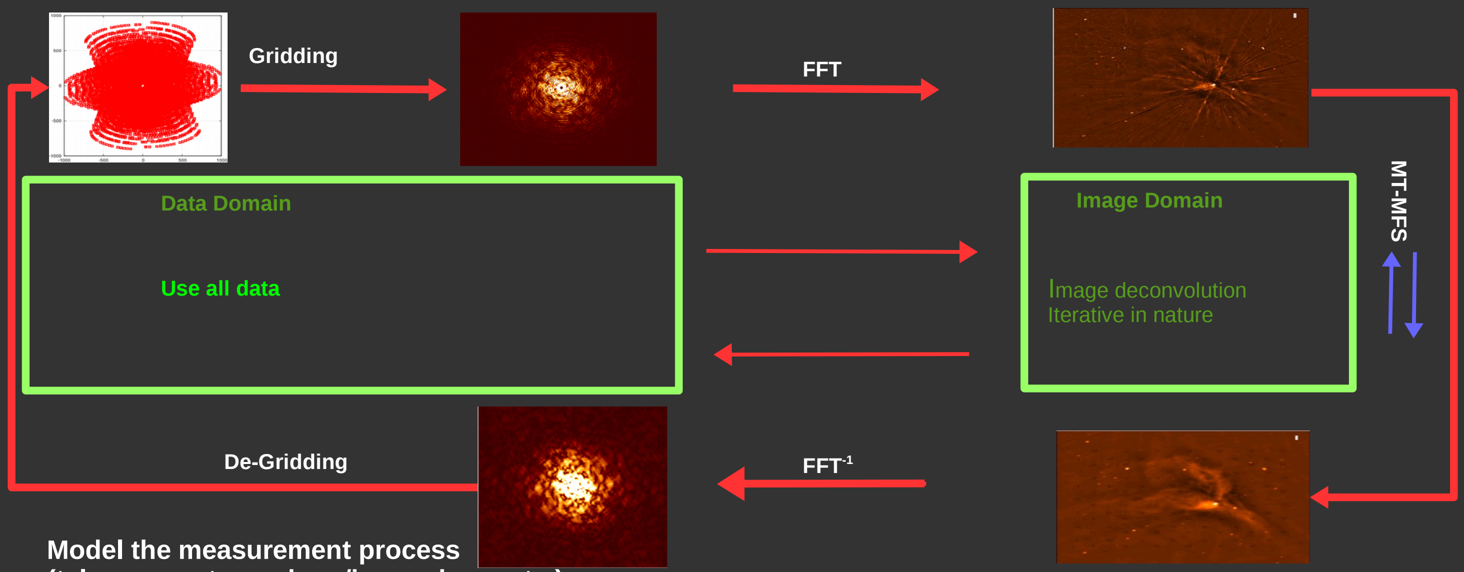
- Solvers for ionospheric, atmospheric effects?
  - High computational complexity, but getting cheaper to compute
- Naturally works with A-Projection for DD calibration during imaging
  - Slowly varying functions (Antennas. Yey!)
  - Low computational complexity, and getting cheaper

**Srikrishna et al, in prep**

# Imaging

- The process of inverting the measurement equation

$$\boldsymbol{F}^{T}\, T\, \overrightarrow{V_{ij}^{Obs}} = \boldsymbol{F}^{T}\, T\;\; W_{ij}\, A_{ij} * F\, I \quad + \quad F^{T}\, T\, N$$



Gridding

FFT

MT-MFS

Data Domain

Use all data

Image Domain

Image deconvolution
Iterative in nature

De-Gridding

$FFT^{-1}$

**Model the measurement process
(telescope, atmosphere/ionosphere, etc.)**

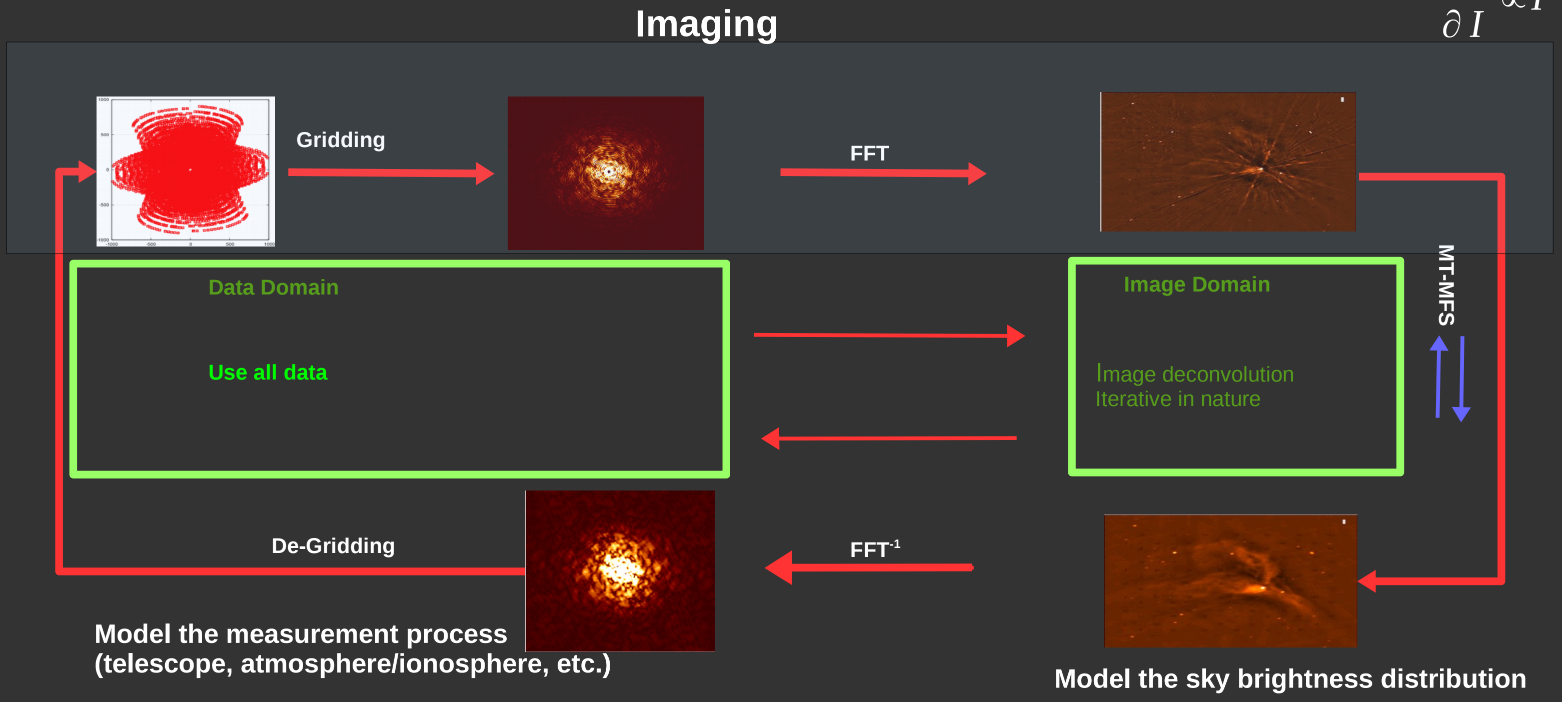
**Model the sky brightness distribution**

# Imaging

- The process of inverting the measurement equation

$$F^T T \overrightarrow{V^{Obs}_{ij}} = F^T T \; W_{ij} A_{ij} * F I \quad + \quad F^T T N$$

$$\frac{\partial \chi^2}{\partial I} \propto I^R$$



**Imaging**

Gridding

FFT

Data Domain

**Use all data**

Image Domain

Image deconvolution
Iterative in nature

**MT-MFS**

De-Gridding

$FFT^{-1}$

**Model the measurement process (telescope, atmosphere/ionosphere, etc.)**

**Model the sky brightness distribution**

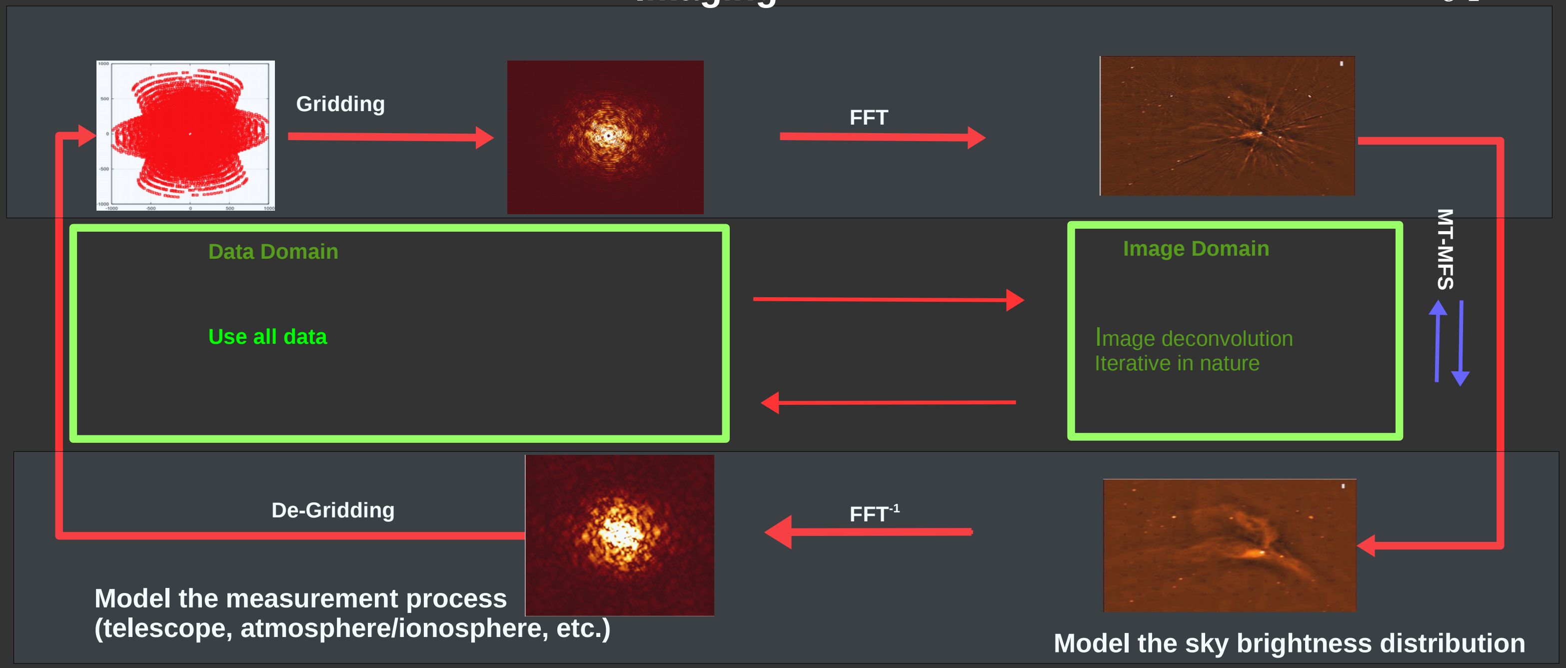
# Imaging

- The process of inverting the measurement equation

$$\boldsymbol{F}^T T\, \overrightarrow{V^{Obs}_{ij}} = \boldsymbol{F}^T T\ W_{ij} A_{ij} * F I \quad + \quad F^T T N$$

**Imaging**

$$\frac{\partial \chi^2}{\partial I} \propto I^R$$



**Data Domain**

**Use all data**

**Image Domain**

Image deconvolution
Iterative in nature

**Model the measurement process (telescope, atmosphere/ionosphere, etc.)**

**Model the sky brightness distribution**

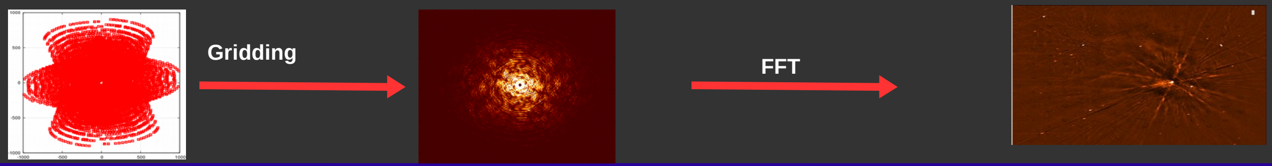
**Prediction/Simulation/Forward modeling**

# Imaging algorithms

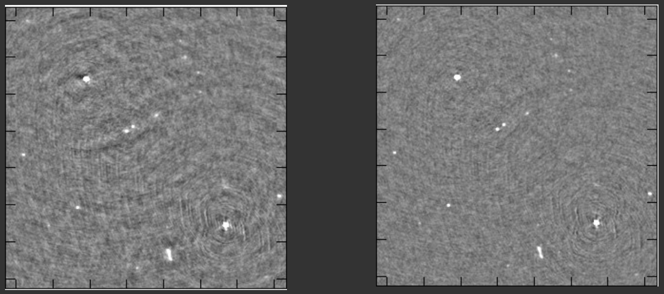
- Projection algorithms: Modified transform to include DD corrections

$$\boldsymbol{F}^T T \, \overrightarrow{V_{ij}^{Obs}} = \boldsymbol{F}^T T \;\, W_{ij} A_{ij} * F I \quad + \quad F^T T N$$

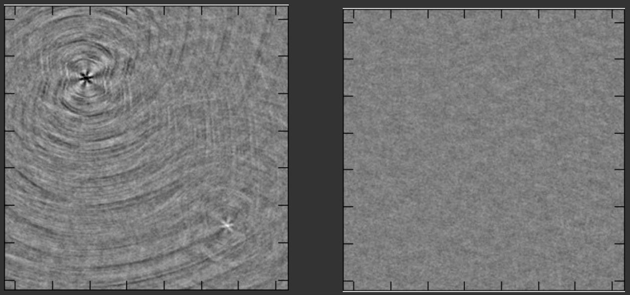
  - Make images free of DD effects.  SNR optimal
  - Design T as a (pseudo) inverse for the DD operators
    - » W-Projection: **Cornwell et al, IEEE Selected topics in Sig Proc.2008**
    - » (WB) A-Projection: **Bhatnagar et al., A&A 487, 419-429 (2008)**
      **ApJ (2013,Vol.770, No. 2, 91)**
    - » W-Snapshot: **Cornwell, et al.,  2012, SPIE Conference Series Vol. 8500, Image Reconstruction from Incomplete Data VII**
      **Ye et al., https://arxiv.org/pdf/2101.11172.pdf**
  - Decouples imaging and image reconstruction algorithms
  - Naturally and simply enables otherwise complex scientific capabilities:
    - » E.g., Wide-field wide-band full-polarization mosaic imaging with heterogeneous arrays and pointing correction + SD-Int. algorithms



# Full-Mueller A-Projection



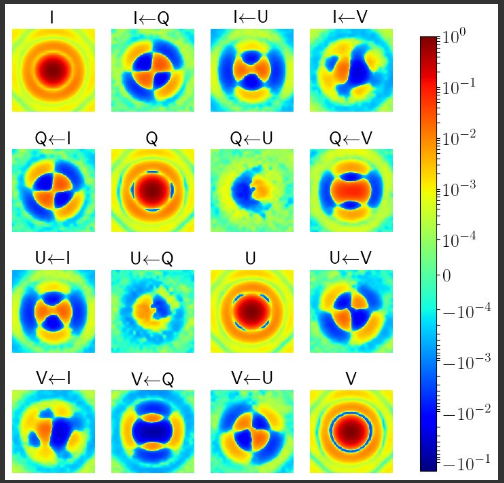
**Stokes-I**



**Stokes-V**

- $$\begin{bmatrix} I_I^{Obs} \\ I_Q^{Obs} \\ I_U^{Obs} \\ I_V^{Obs} \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{bmatrix} \cdot \begin{bmatrix} I_I^o \\ I_Q^o \\ I_U^o \\ I_V^o \end{bmatrix}$$

- Diagonal: "pure" poln. Products
- Off-diagonal: Include poln. leakage



**Jagannathan et al, in prep**

# Imaging algorithms: Faceting+W-Proj.

- Image-plane faceting has an equivalent transform in the visibility domain

$$\boldsymbol{F}[I(\boldsymbol{C}\vec{s})] = |det\,\boldsymbol{C}|^{-1} \quad V(\boldsymbol{C}^{-1}\vec{b})$$

Sault et al., 1996, A&A?

- Combination of UV-domain faceting and W-Projection may be a good fit for massively parallel h/w
  - Modified visibility and co-ordinates gridded to a single grid
  - Residual W-term effect corrected with W-Projection with a reduced max. W.

- Available in CASA (since the AIPS++ days in fact!)

# Image modeling (aka “deconvolution”)

- $\chi^2 = \sum_{i,j} |\overrightarrow{V_{ij}^{Obs}} - A_{ij} * w_{ij} F I|^2 = \sum_{ij} |\overrightarrow{V_{ij}^{Obs}} - [B I]_{ij}|^2$

$$\frac{\partial \chi^2}{\partial I} = 0 \qquad \text{where} \qquad I(\vec{s}) = \sum P_k(\{a_0, a_{1,\ldots}\})$$

- Normal equations: $I^D = [B^T B] I^0 + [B^T B] N$

  $[B^T B]$ is the convolution with the PSF

- Some observations:
  - Noise in the image domain has a non-zero correlation length
  - $[B^T B]$ Is formally not inveritable.
    - RA jargon: “missing spacing”, “solutions non-unique”,…

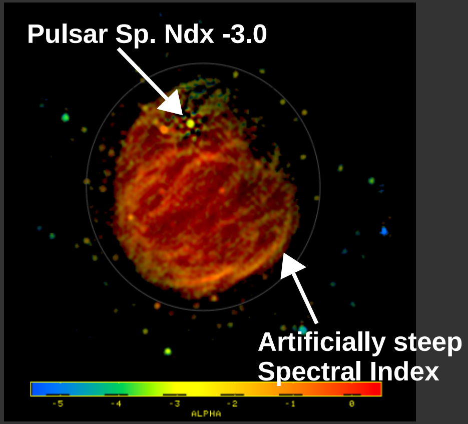
  - Modern telescopes make the Eigen value spectrum more tractable
    - Are faster approximations possible now?
  - O($10^{12-15}$) constraints are now typical. Non-uniqueness of solution argument over-played?
    - Solution differences insignificant
    - How scientifically useful are expensive algorithms for “error estimates”?

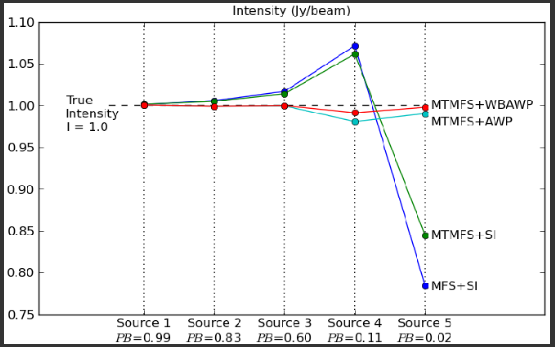
# Image modeling (aka “deconvolution”)

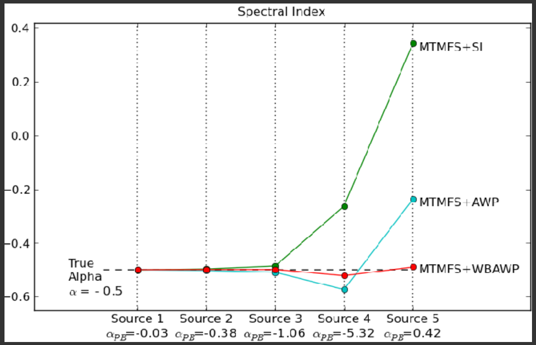
- Solvers in use in RA
  - Most, if not all, are formally already CS algorithms!
  - Fundamentally iterative in nature
  - Most can be understood as differing in
    - basis functions use for modeling the emission, prior information, constraints/regularization
    - A complete basis set is ruled out: Imposing constraints like “positivity” isn’t necessary (or even correct)
- Hogbom
  - $I(\vec{s}) = \sum_k F_k \; \delta(x_k)$
  - Ignores correlated noise
  - “Boxes” as regularizers
- MS-Clean, MT-MFS
  - $I(\vec{s}) = \text{A dictionary of scales}$
  - Regularization: best-fit largest scale
- Bayesian methods
  - With and without scale-sensitive basis
  - Formal approach to deriving regularization/constraints

# Image modeling (aka "deconvolution")

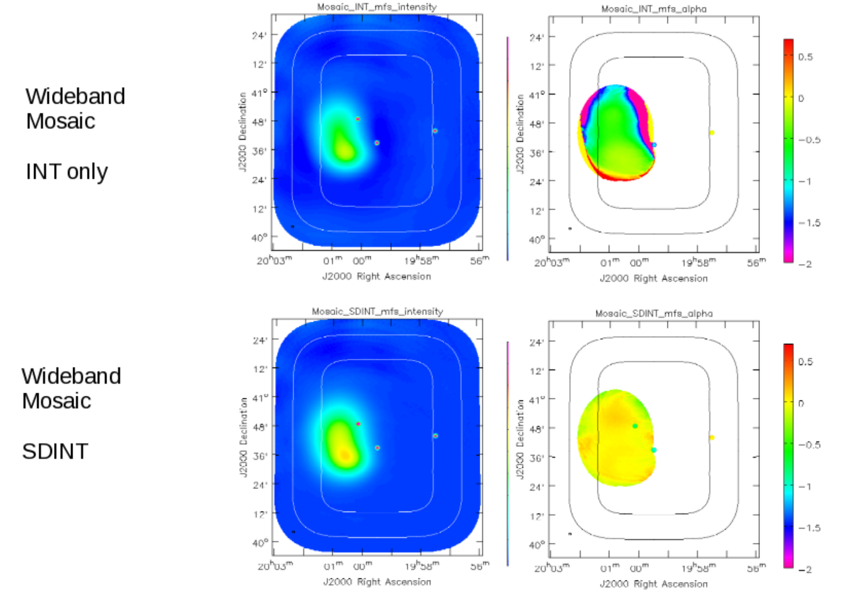
- Wide-field wide-band imaging with scale-sensitive reconstruction







- SD+Int Mosaic Imaging

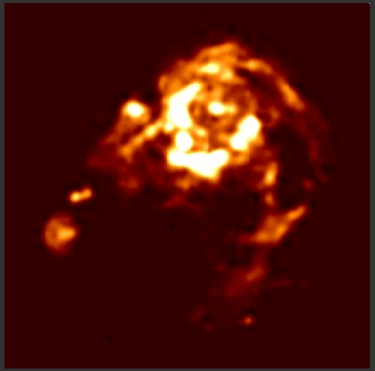


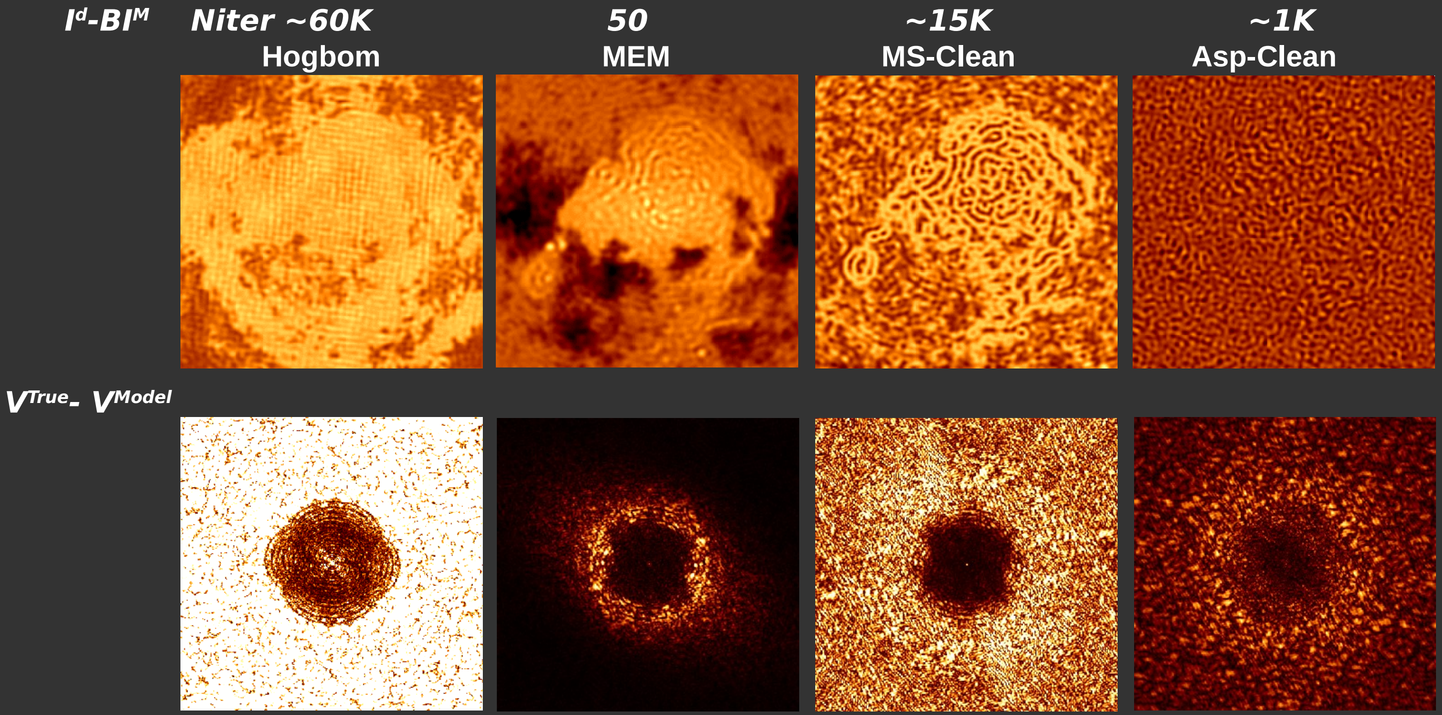
Wideband Mosaic : gridder='mosaic' with specmode='cube' and deconvolver='mtmfs' with specmode='mfs' with conjbeams=T implemented as part of the SD+INT merging process (in python). [[ Used PySynthesisImager prototyping interface ]]

Rau, et al., AJ, Vol.158, No. 1, 2019

# Image modeling (aka “deconvolution”)

- Asp-Clean
  - Automatically discover the scale-sensitive basis set
  - Formally regularized with constraints
  - Embedded heuristics
- New implementation with improvements under validation/verification for release in CASA





Hsieh et al., in prep

# Computing costs

- Need an integrated systems approach for designing post-processing system
  - Scientific requirements
  - Algorithm architecture
  - Computing architecture: software + hardware

- Develop a size-of-computing (SofC) estimate based on the *required* algorithms/procedures for the key science drivers

- Develop a computing cost model based on available software implementations
  - Establish a measurable, and a *practical* procedure to estimate SofC

- Pipeline processing is an important driver for *realistic* SofC, algorithm, software and computing architectures
  - Heuristics can add significantly to SofC

# Procedure for SofC estimates for ngVLA

- Establish the algorithms *required* for the key science drivers

- Develop a theoretical model for the computational complexity of the required algorithms
  - Develop scaling laws
  - Identify the highest-nail(s) in an end-to-end cost of computing

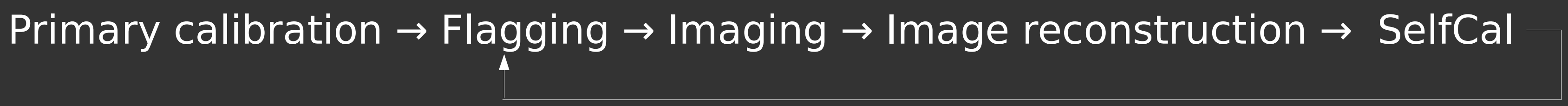
- Measure the *single-core* code efficiency of *an* implementation. Compare against industry standard to normalize the s/w and h/w used.

- Estimate and verify the SofC based on code efficiency.

- Develop a computing architecture
  - Large scale parallelization, fault-tolerant software on hybrid hardware
  - Implications of pipeline processing for SRDP
  - Operational costs

# Procedure for SofC estimates for ngVLA

- Baseline end-to-end processing steps



Primary calibration → Flagging → Imaging → Image reconstruction → SelfCal

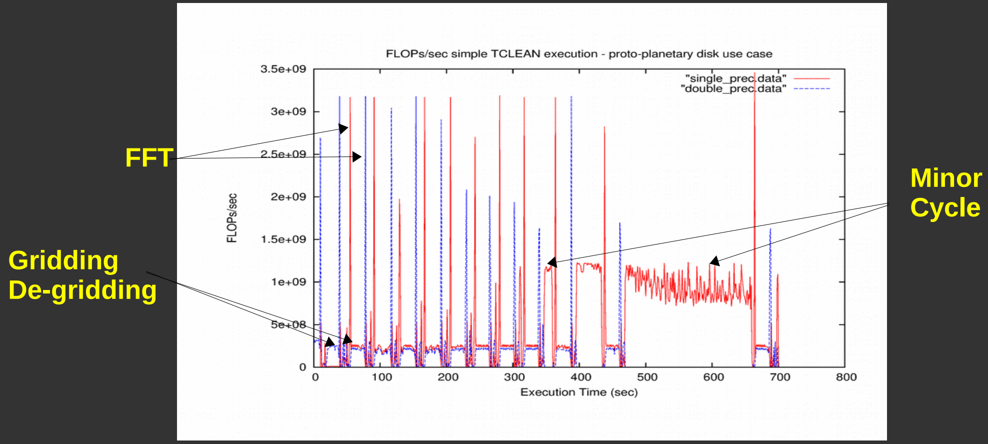
  - Automatic processing can add orders of mag to SofC

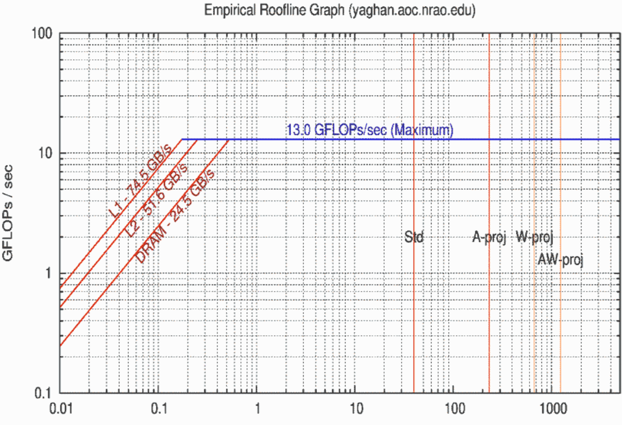
- Calibration (primary and SelfCal) and imaging are iterative optimization problems. Identify the computing hot-spots (in the Fermi Problem style)

- As it typical in general, derivative computation is the highest nail
  - Calibration: O($N_{vis}$) algorithm
  - Imaging: O($N_{vis} \times CF^2$) algorithm
  - Image reconstruction: Scales roughly as image size
  - Flagging: Cost assumed to be smaller than imaging but larger than calibration

# Size of computer

- Imaging, and in-turn gridding/degridding identified as the cost driver





- Computing on CPU-core is compute bound.
  - Conclusion: large scale parallelization will solve the problem. But...
- Order of mag more CPU cores required than even “industry standard” for 24x7 operations!
- FLOP per data point is high O(1000). Memory accesses per data point O(10000).
  - Conclusion: Massively parallel h/w will help
  - GPU, TPU: few x 1000 cores, 900 GB/s bandwidth, O(TeraFLOPS)

- An efficient implementation gets to memory bandwidth-bound computing. Reduces the coarse parallelization width by orders of mag
- **Within reach!**

**Bhatnagar, Pokorny, Hiriart, ngVLA**