Modelling the Imaging Pipeline in the Science Data Processor

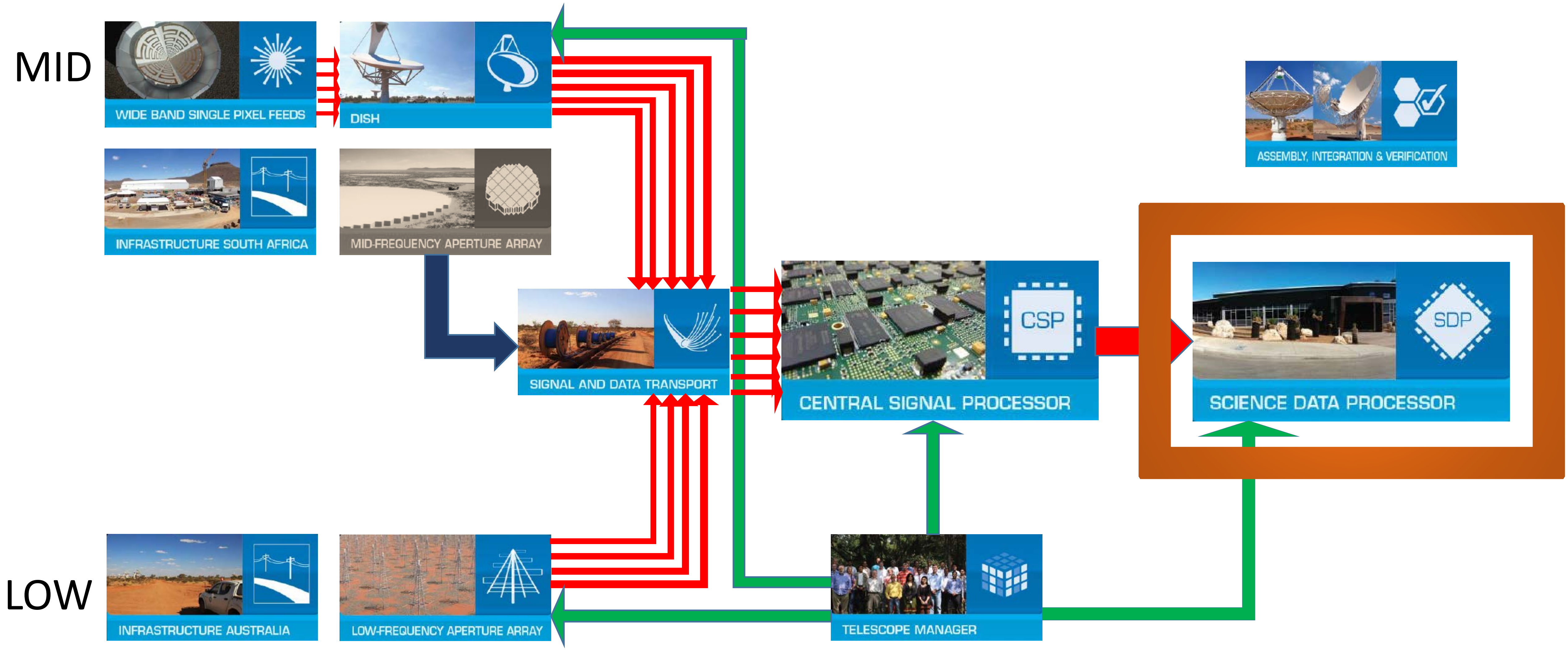
11 February 2016

Anthony Griffin

HIGH PERFORMANCE COMPUTING RESEARCH LAB
SCHOOL OF COMPUTER & MATHEMATCIAL SCIENCES

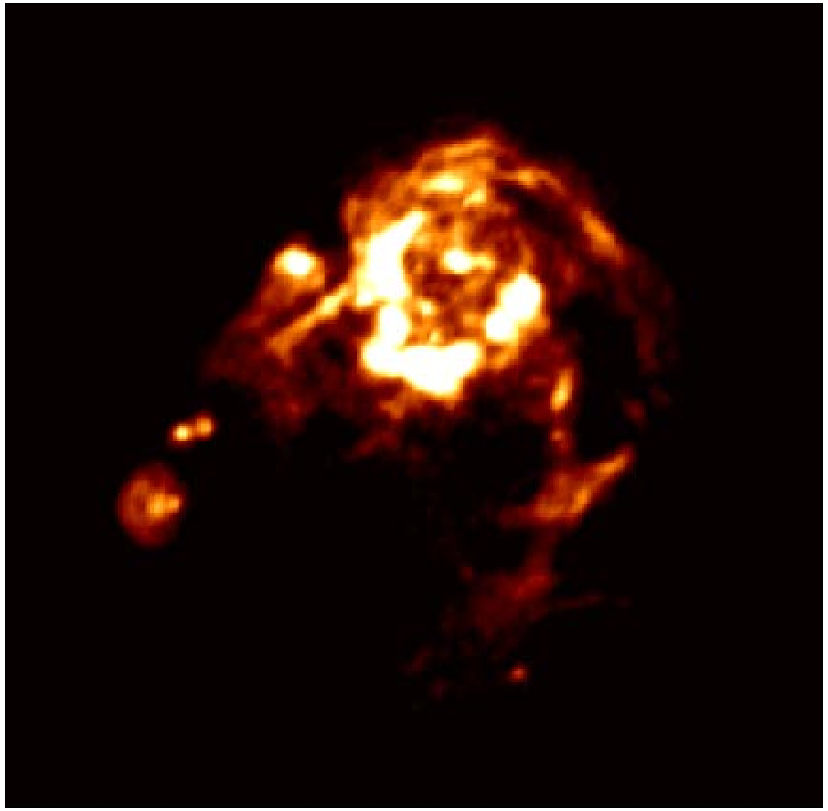
SKA SQUARE KILOMETRE ARRAY SDP SCIENCE DATA PROCESSOR

SKA Phase One Data Flow and Consortia Teams



Imaging Pipeline

- Major function of the Science Data Processor (SDP)
- Takes the output of the Channel Signal Processor (CSP)
  - Visibilities (uv-plane)
  - (Measurements in the Fourier domain)
- Processes the Visibilities
- Produces an image of a region of the sky



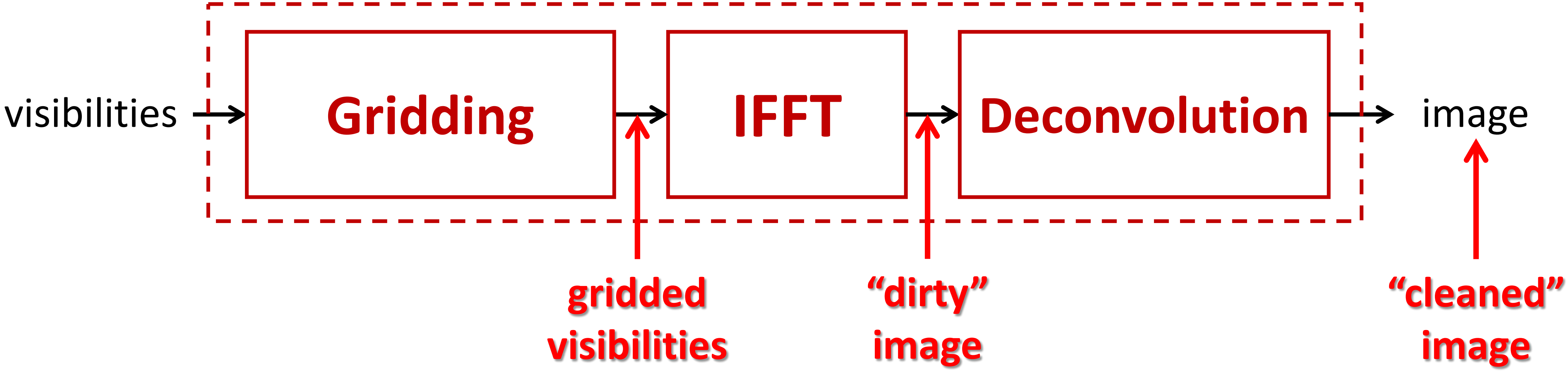
Why are we modelling the imaging pipeline?

- Give us a reference model for implementation variations
- Allow us to try out various imaging algorithms
- Determine the required mathematical precision:
  - Double- or single-precision floating point?
  - Fixed-point?
- Part of the prototyping plan for the SDP

How are we modelling the imaging pipeline?

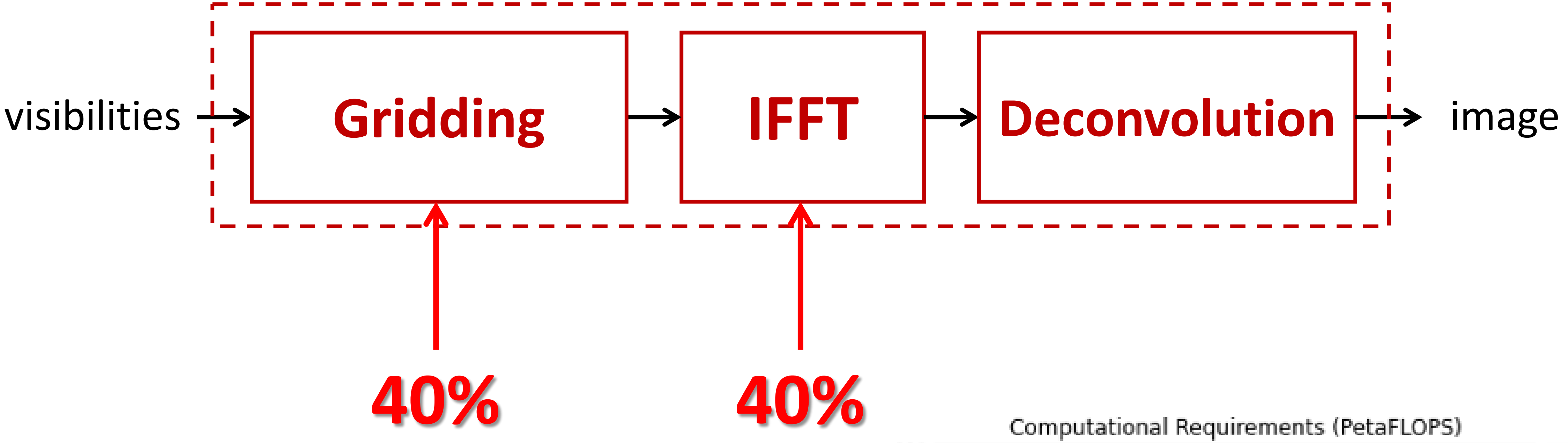
- Based on ASKAPsoft (our “golden model”)
  - Software that ASKAP runs (Australian SKA Pathfinder)
  - Written in C++, and runs in parallel processors
- Why?
  - Was an SKA pathfinder
  - State-of-the-art system
  - Source code
- Will be written in Matlab (translated into Matlab)
  - Very flexible
  - Easy to test precision, different implementations
  - Good software engineering principles (unit tests, etc)

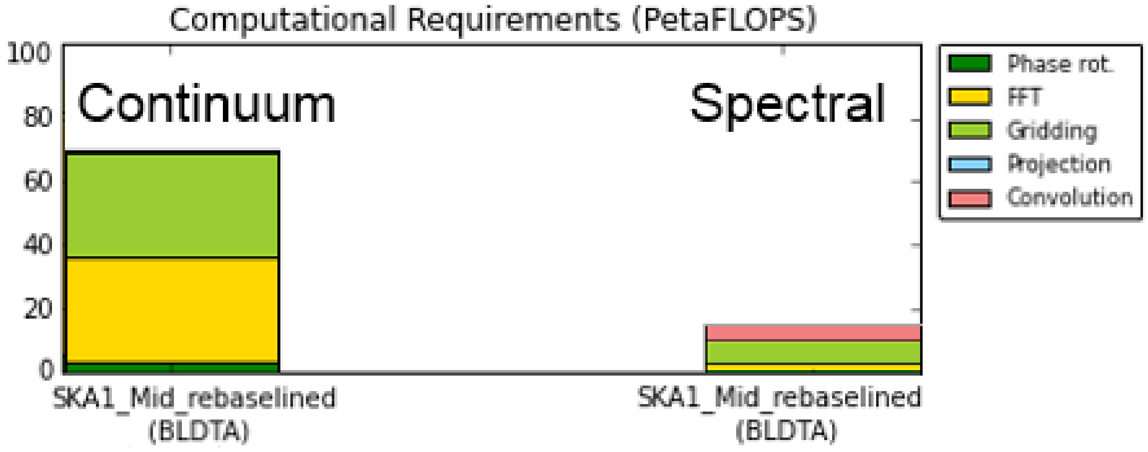
Imaging Pipeline



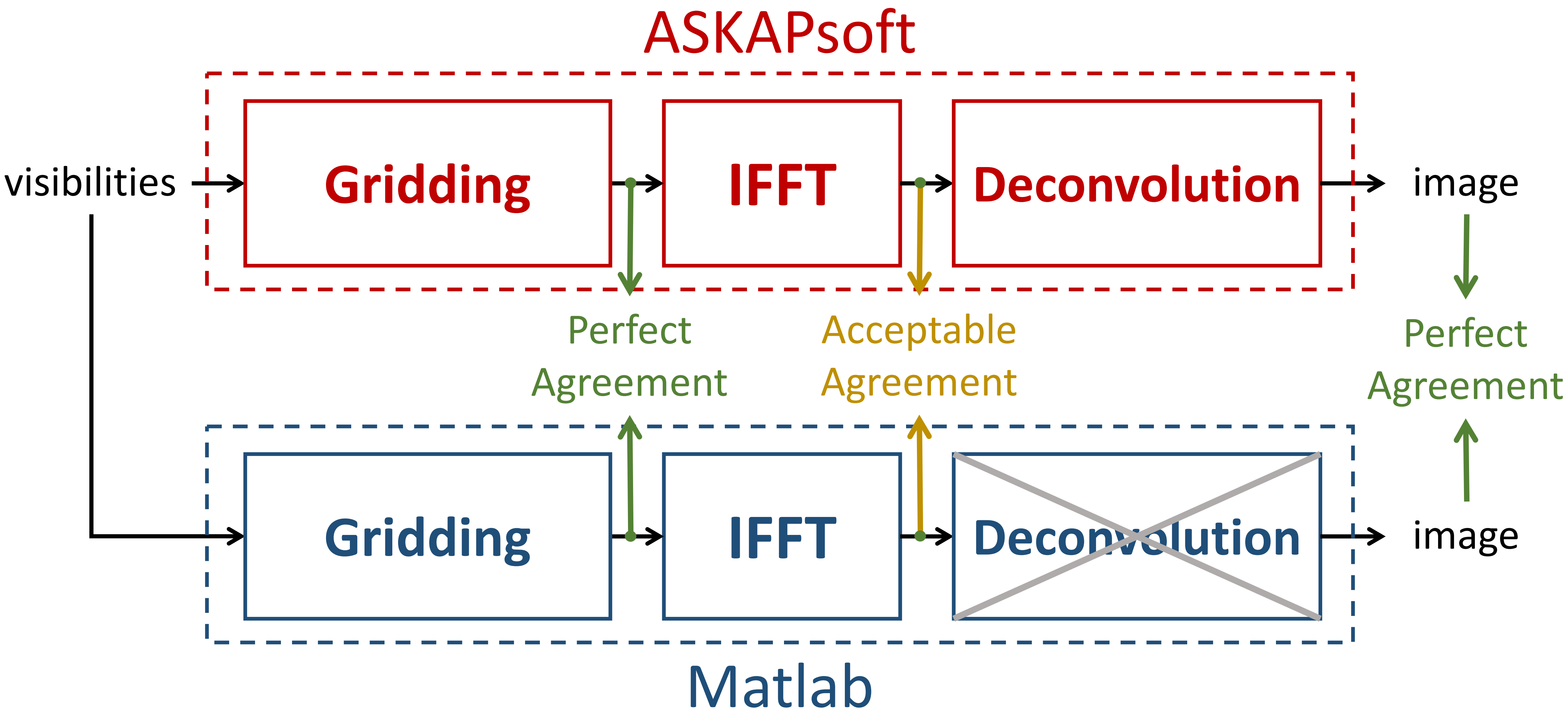
- Visibilities can occur anywhere in the uv-plane
  - continuously-distributed, but extremely under-sampled
- IFFT requires its input to lie on an evenly-spaced grid
- Gridding is the process of placing the visibilities on the IFFT’s input grid

Imaging Pipeline





Matlab Imaging Pipeline Modelling Goal



Comparison Metric

- Root mean square error

$$\text{RMSE}(\boldsymbol{X}, \widehat{\boldsymbol{X}}) = \sqrt{\frac{\sum_i (X_i - \widehat{X}_i)^2}{N}} = \frac{\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\|_2}{\sqrt{N}}$$

where

$\boldsymbol{X}$ is the “true” value

$\widehat{\boldsymbol{X}}$ is the estimated value

N is the size of $\boldsymbol{X}$ and $\widehat{\boldsymbol{X}}$

and

$\|\boldsymbol{X}\|_2 = \sqrt{\sum_i (X_i)^2}$

What is the problem with using the RMSE as a metric?

The RMSE will scale with the energy of the signals in question.

For double precision, acceptable agreement is an rRMSE of about 1×10^{-16}

For single precision, acceptable agreement is an rRMSE of about 1×10^{-7}

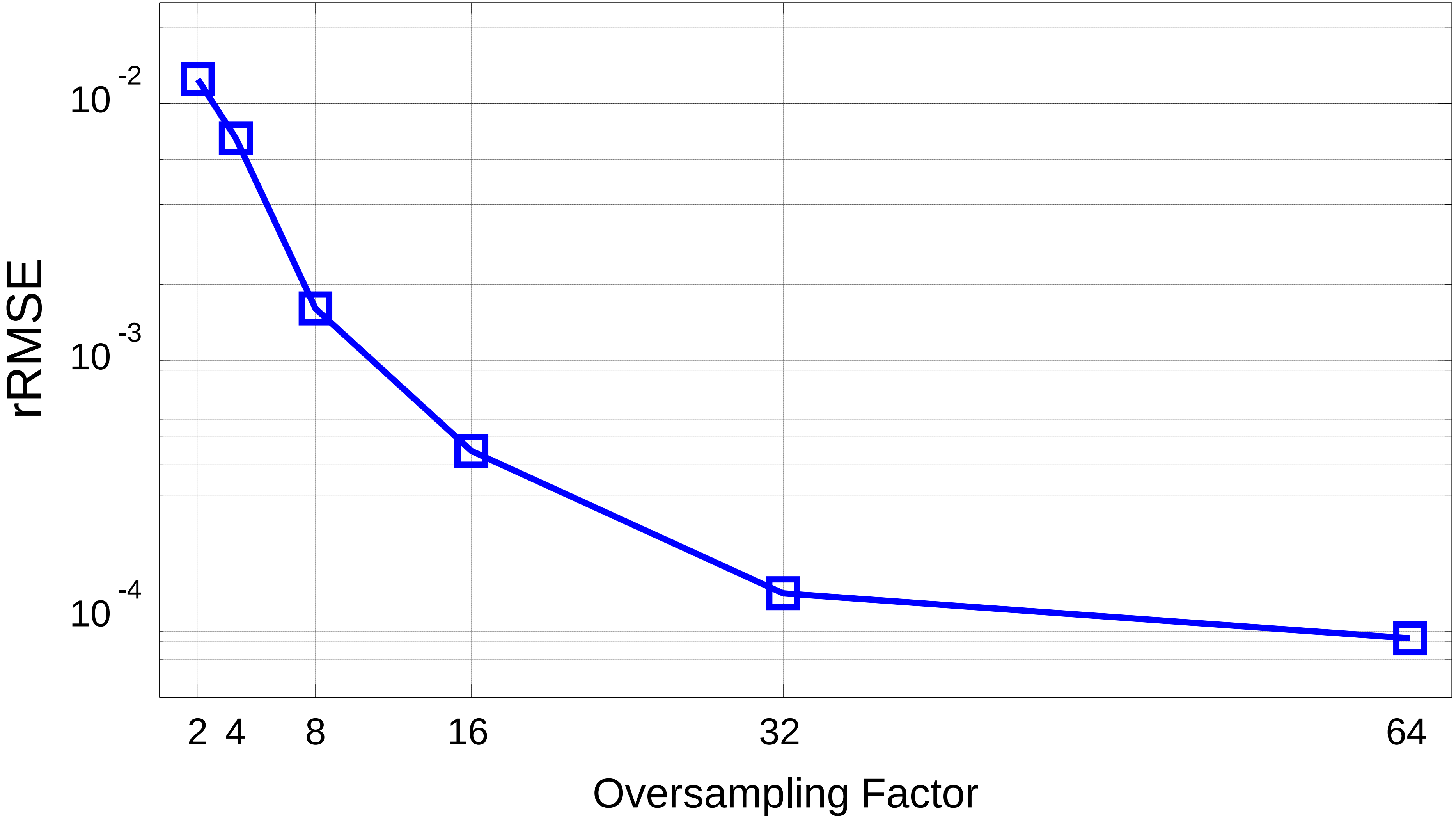
- Relative RMSE

$$\text{rRMSE}(\boldsymbol{X}, \widehat{\boldsymbol{X}}) = \sqrt{\frac{\sum_i (X_i - \widehat{X}_i)^2}{\sum_i (X_i)^2}} = \frac{\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\|_2}{\|\boldsymbol{X}\|_2}$$

“energy” of the error

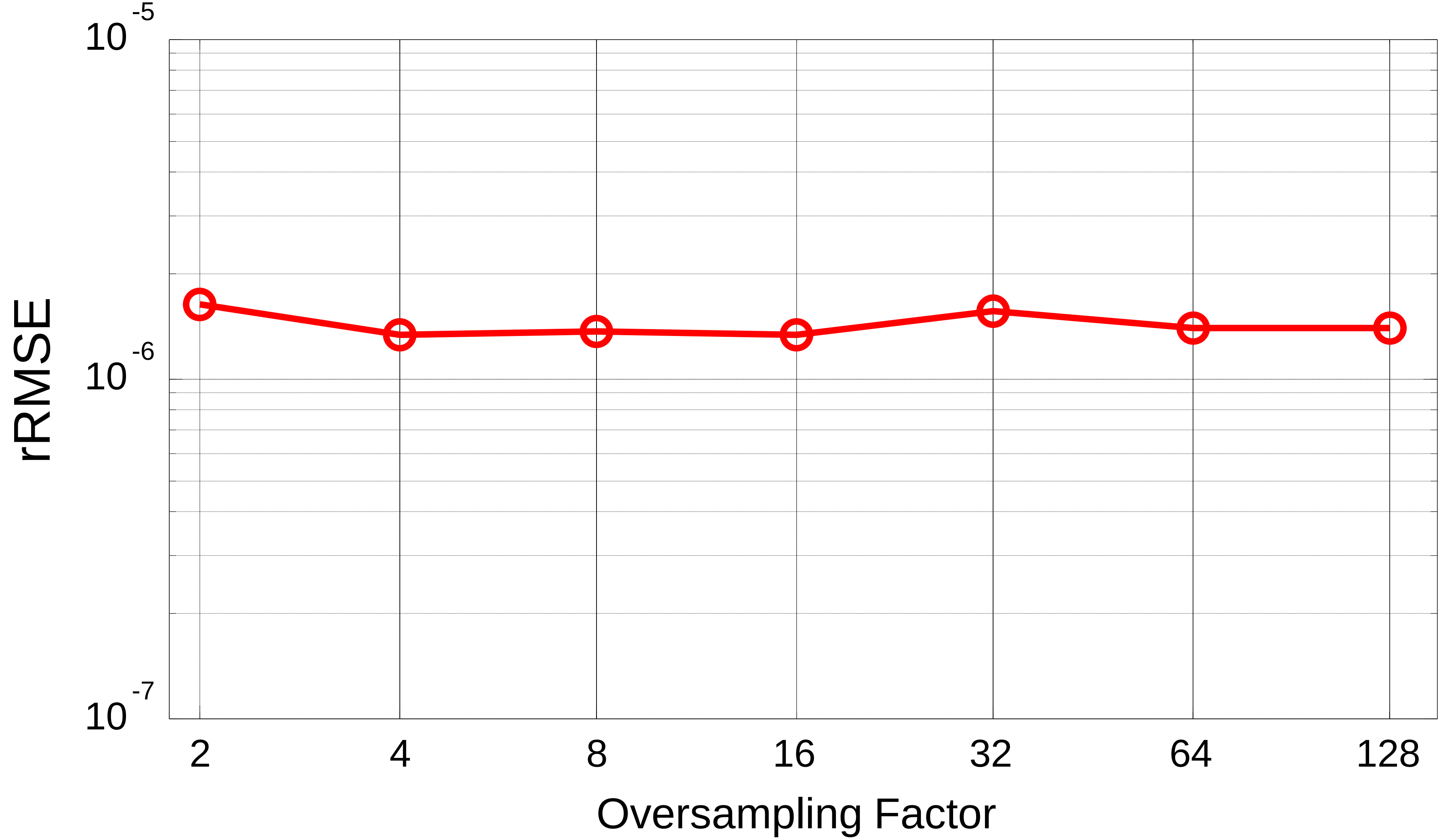
“energy” of the true signal

rRMSE of gridded visibilities vs oversampling factor



- higher oversampling is better
- reference gridded visibilities use 128 oversampling

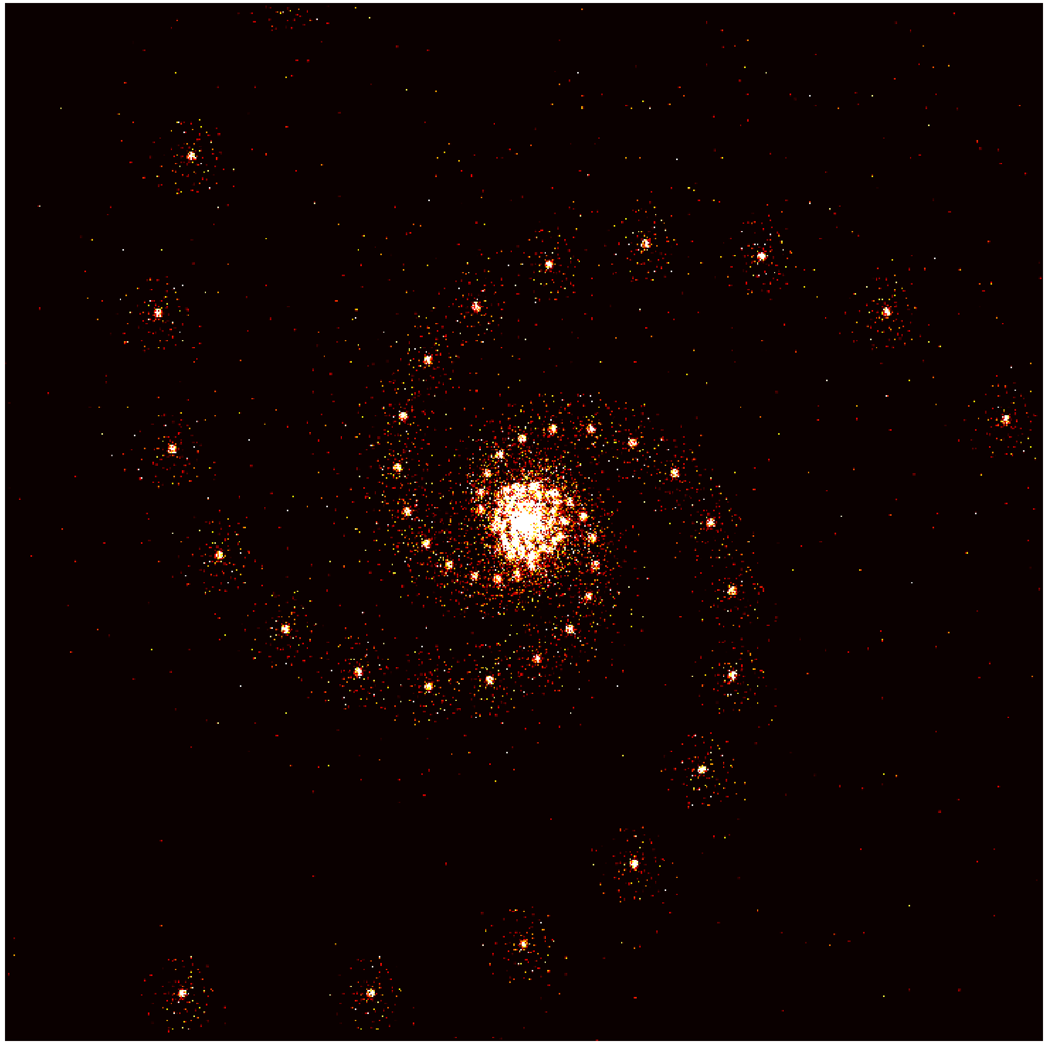
rRMSE between gridded visibilities using single- and double-precision



- Single-precision gridding is compared with double-precision gridding using the same oversampling
- e.g., single-precision with 4 times oversampling is compared with double-precision with 4 times oversampling

Example with SKA MID simulated data

gridded visibilities



IFFT

“dirty” image

Summary

- The Matlab Imaging Pipeline gridder has been implemented
  - Its output agrees perfectly with that ASKAPsoft.
  - It is now generating useful results.
- Next step is to implement the deconvolution stage.
- May need to look at speed.
  - Matlab is about 100x slower than ASKAPsoft running in a Linux virtual box!
- The Matlab source will be released within SKA.