RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cea cnrs Inria

FRANCE 2030
PROGRAMME DE RECHERCHE
NUMÉRIQUE POUR L'EXASCALE

ExaDoST: WP4 – SKA Illustrator

Damien Gratadour (OP & CNRS)
Antsa Rasamoela, Valentin Hazard, Iheb Becher (LAB)
Shan Mignot, Sunrise Wang (OCA)

Outline

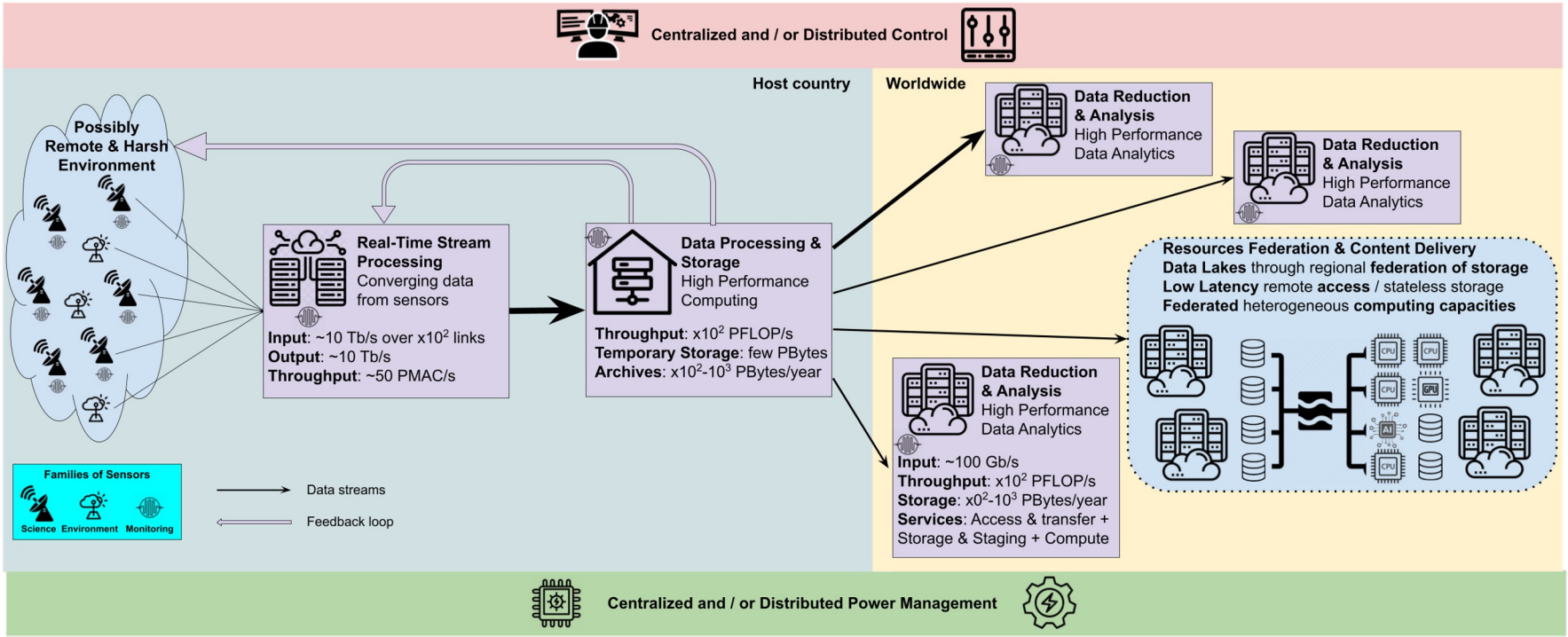
1. Introduction

1. Benchmarking activities

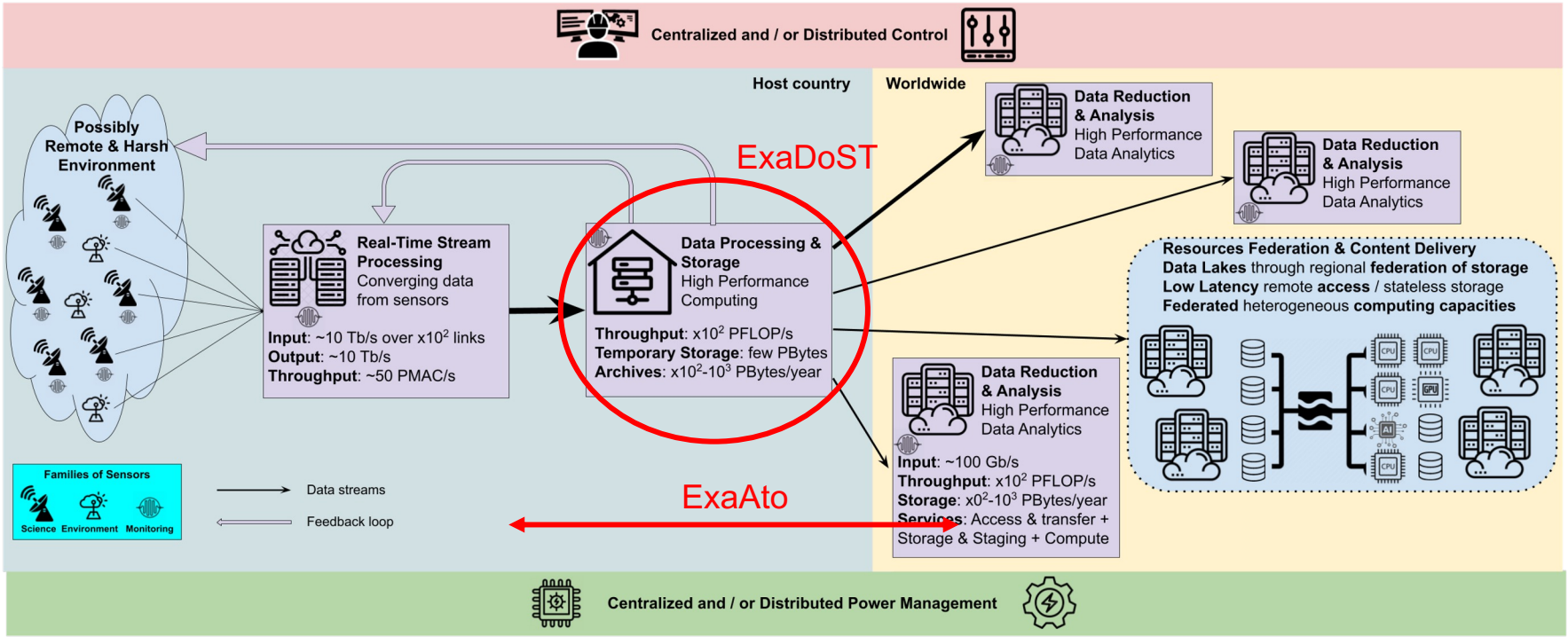
1. New algorithms development

1. Introduction

SKA: global experiment across the continuum



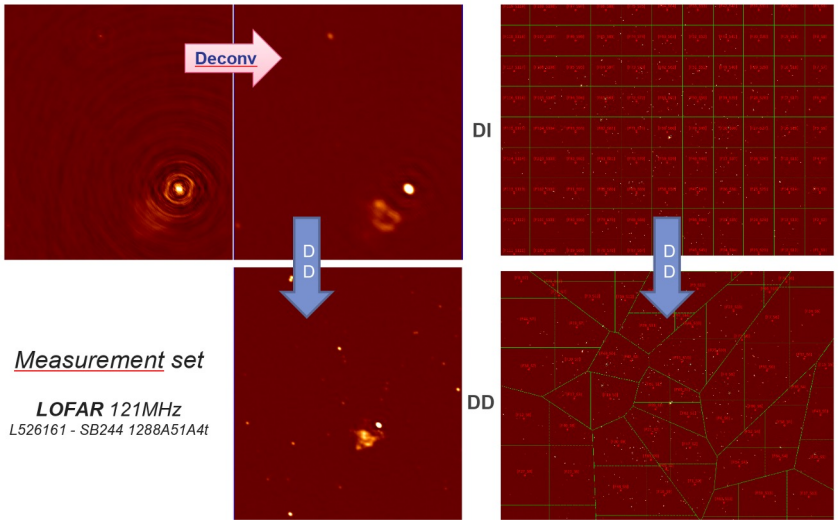
SKA studies in NumPEx

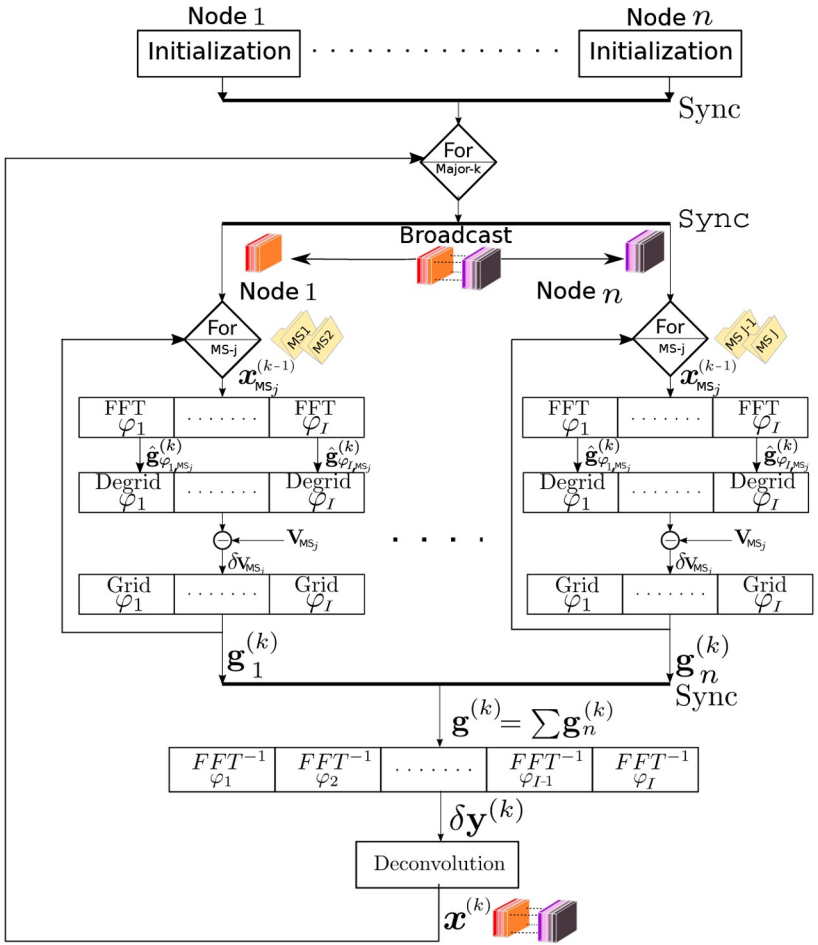


2. Benchmarking activities

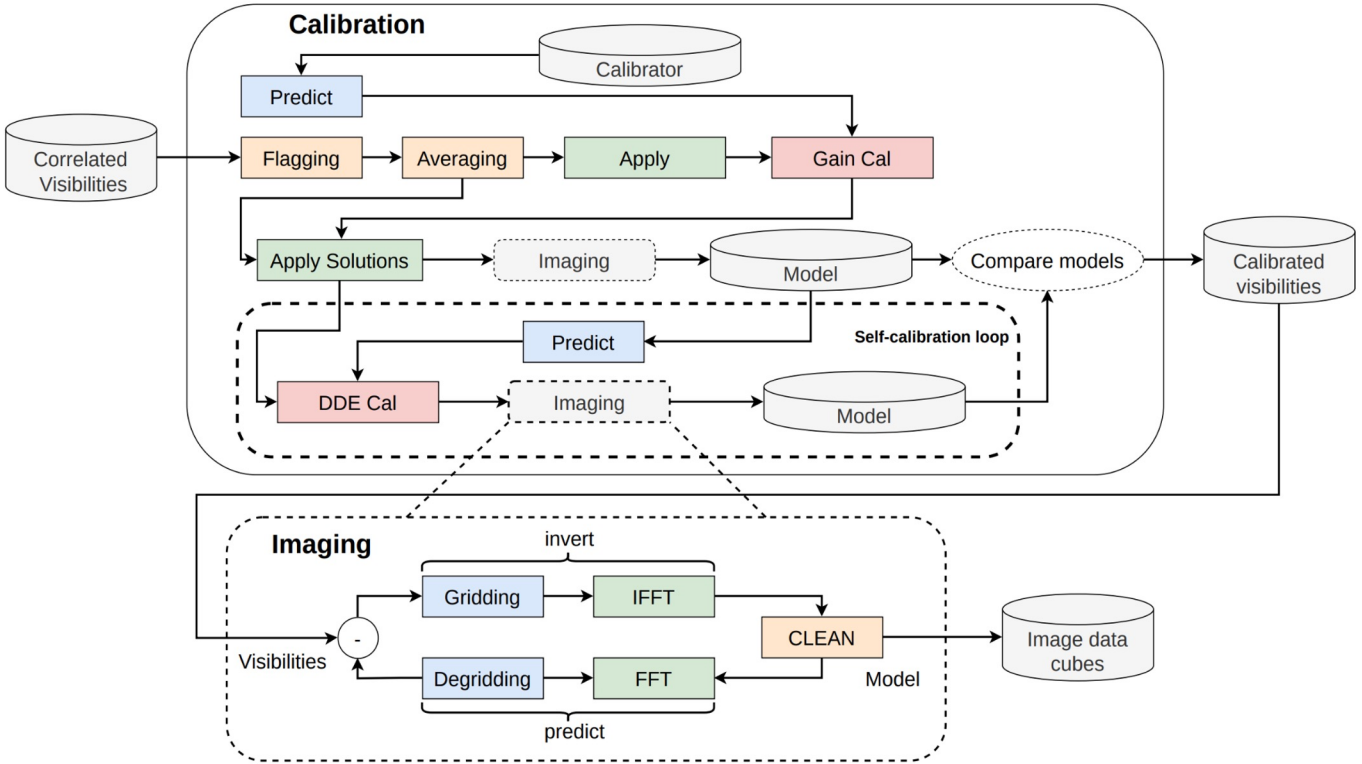
DDF pipeline (KillMS + DDfacet)

SotA software used to process large surveys data
Python based + multi-node parallelization using MPI



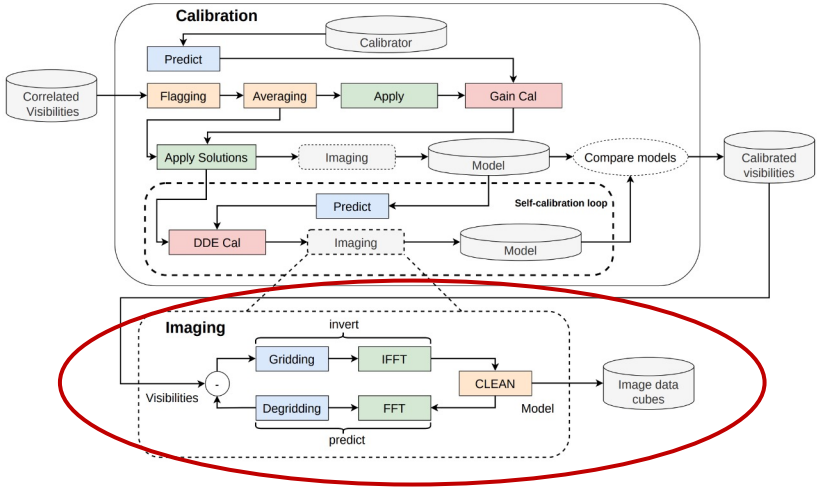


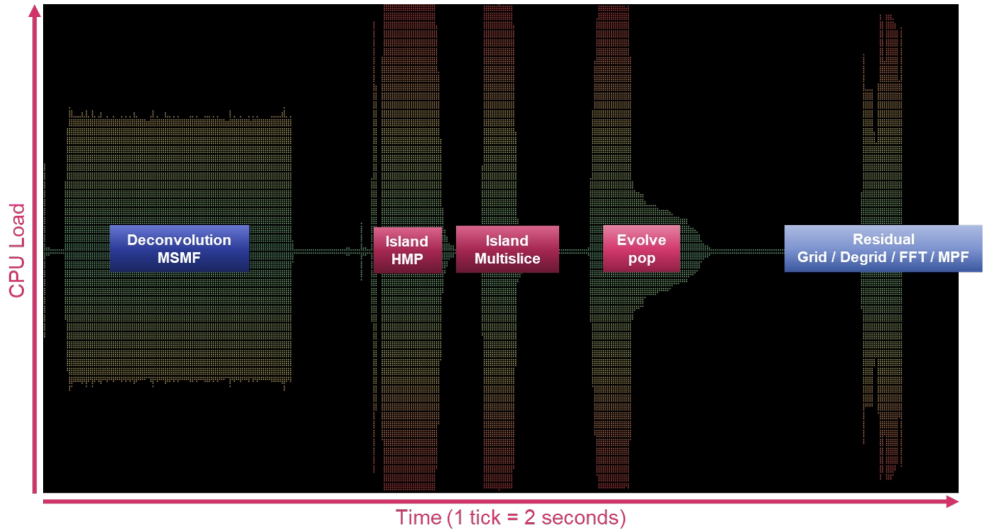
Typical pipeline: complex workflow with iterative sequences



Initial benchmarking @ LAB (CPU load, single node)

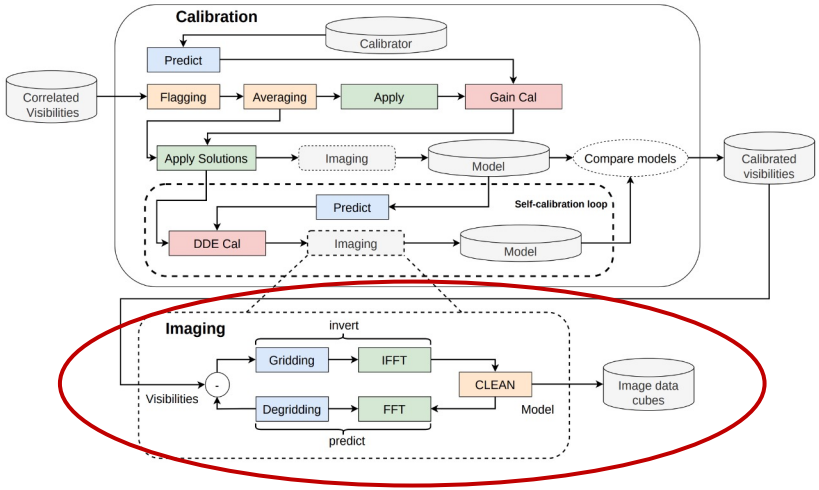
Results with Dool (~1.5h execution)

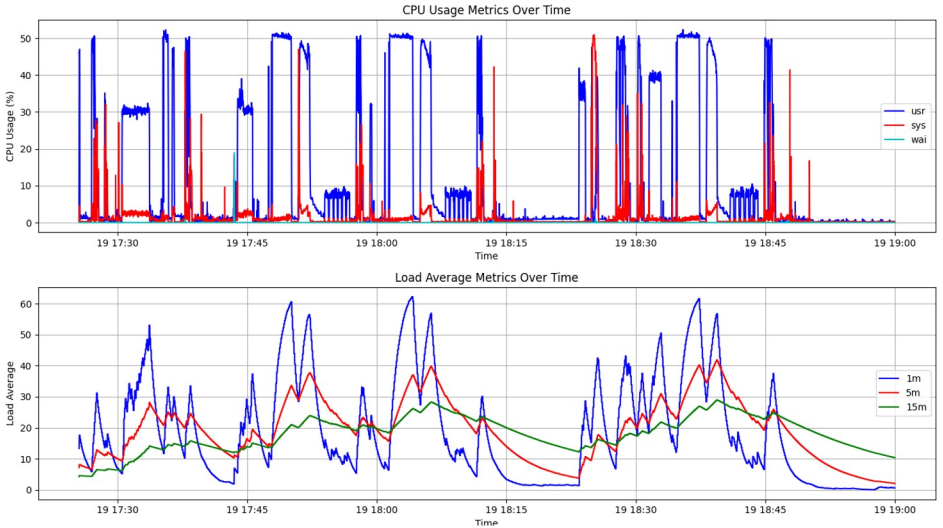




Initial benchmarking @ LAB (CPU load, single node)

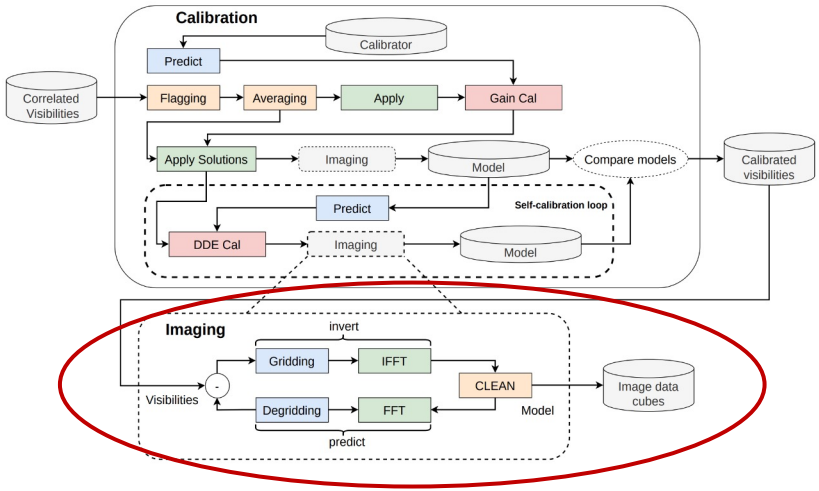
Results with Dool (~1.5h execution)





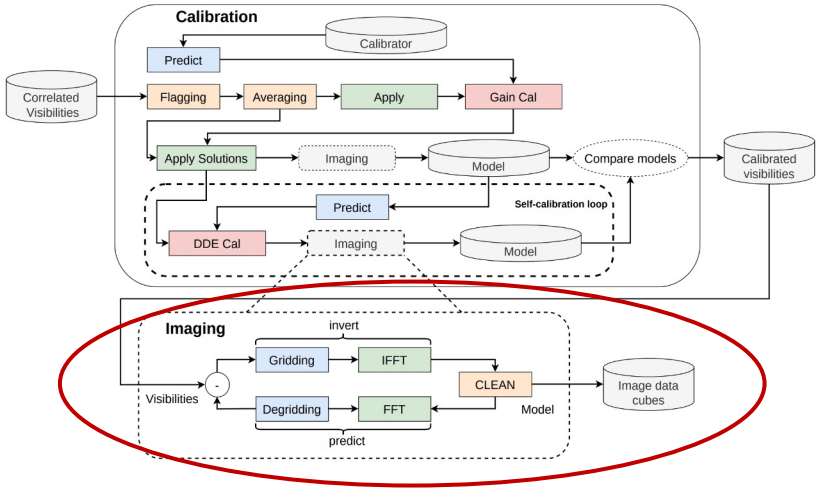
Initial benchmarking @ LAB (memory, single node)

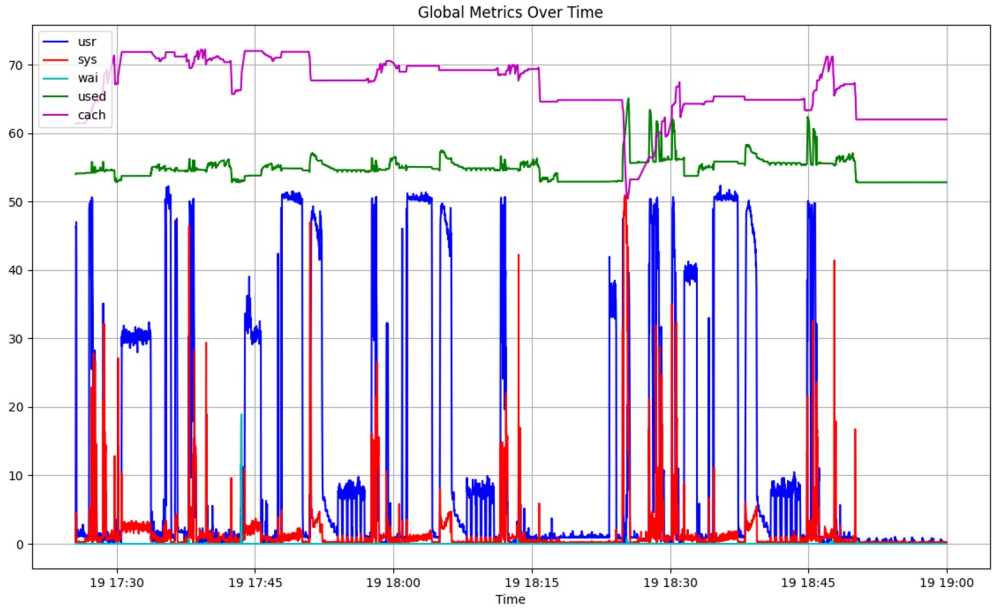
Results with Dool (~1.5h execution)



Initial benchmarking @ LAB (all combined, single node)

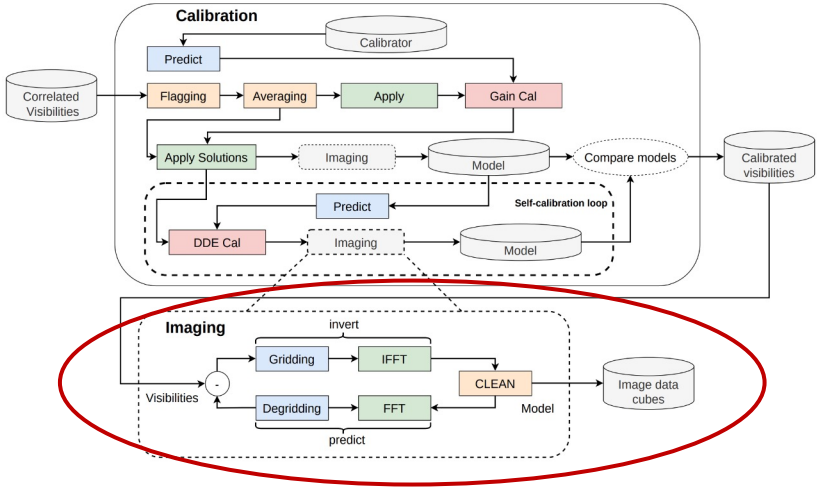
Results with Dool (~1.5h execution)

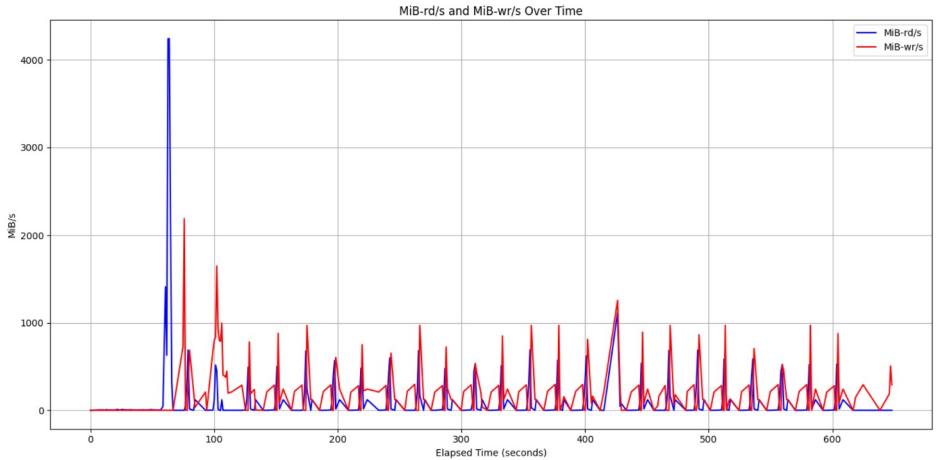




Initial benchmarking @ LAB (I/O, single node)

Preliminary results with Darshan (warning: shorter time period !)



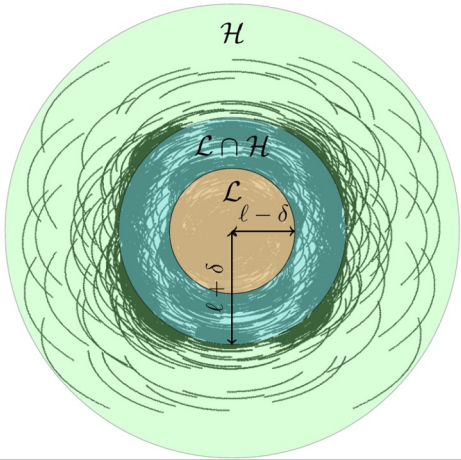


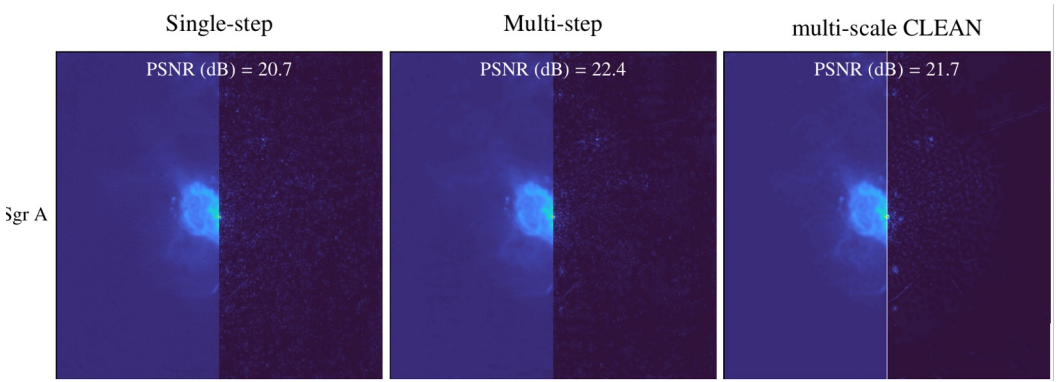
3. New algorithms developments

Multi-step recontruction

Proposal from OCA for new algorithm with interesting properties:

- Consider subsets of visibilities (patitionned by baseline length
- Reconstruction in 2 steps
- Advantages over “classical” methods:
  - Alleviate memory concerns (grid / degrid)
  - Flexible data distribution in a cluster
  - Progressive reconstruction (link with WP2 in situ)





4. Conclusions

Slowly ramping up !

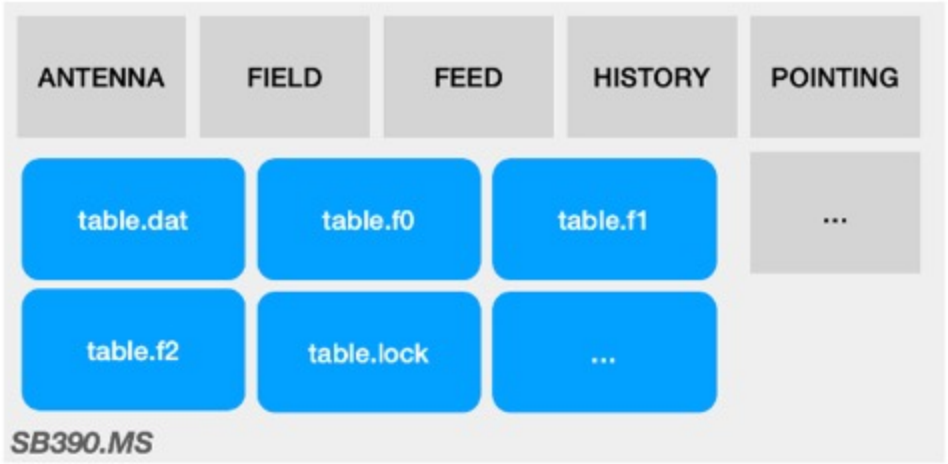
Funding just made available ! (and first contract to start soon @ OCA)

- Regular meetings with the KerData and TADaaM teams and within the “astro team” (OP, LAB, OCA)
- ½ week workshop on SKA pipelines in Jan. 2024 with MeerKAT collaborators (South Africa) in Paris and follow-up remote meetings
- 2 internships started in 2024 and first benchmarking results
- Ongoing collaboration with SCOOP team @ SKA (lead Shan Mignot) + ExaAto (contact: Mathis Certenais)
- Trying to coordinate a more global effort (also including ECLAT joint lab)

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

cea cnrs Inria

- Standard of the community (NRAO) to store interferometric visibility measurements
- 1 MS file = 1 directory containing binary tables and subdirectories = 1 black-box



Contains

Observation setup

Time vector

Frequency band

Antenna positions

Measurements

History

...

- 1 MS must be accessed **indirectly** through dedicated librairies and tools
- 1 MS **integrity** is key to avoid corruption (tar.gz for shipping, no direct intervention inside)
- 1 MS = **1 frequency band** of the instrument *A given scientific observation = a set of MS files*
- 1 MS generally ranges from **few MB** up to **several tens of GB**

File type	Min size	"Max" size
1 MS file	Few 10s-100s MB	~100s GB
e.g. 1 observation (244 MS) (T=1h – df=1 ch/SB – dt=8s)	15 GB	
e.g. 1 observation (244 MS) (T=1h – df=64 ch/SB – dt=1s)	~10 TB	
e.g. 1 observation (244 MS) (T=8h – df=1 ch/SB – dt=8s)	~120 GB	
e.g. 1 observation (244 MS) (T=8h – df=64 ch/SB – dt=1s)	~80-100 TB	