Co-design for SKA
Project status update

Atelier technique ECLAT

Shan Mignot – Observatoire de la Côte d'Azur

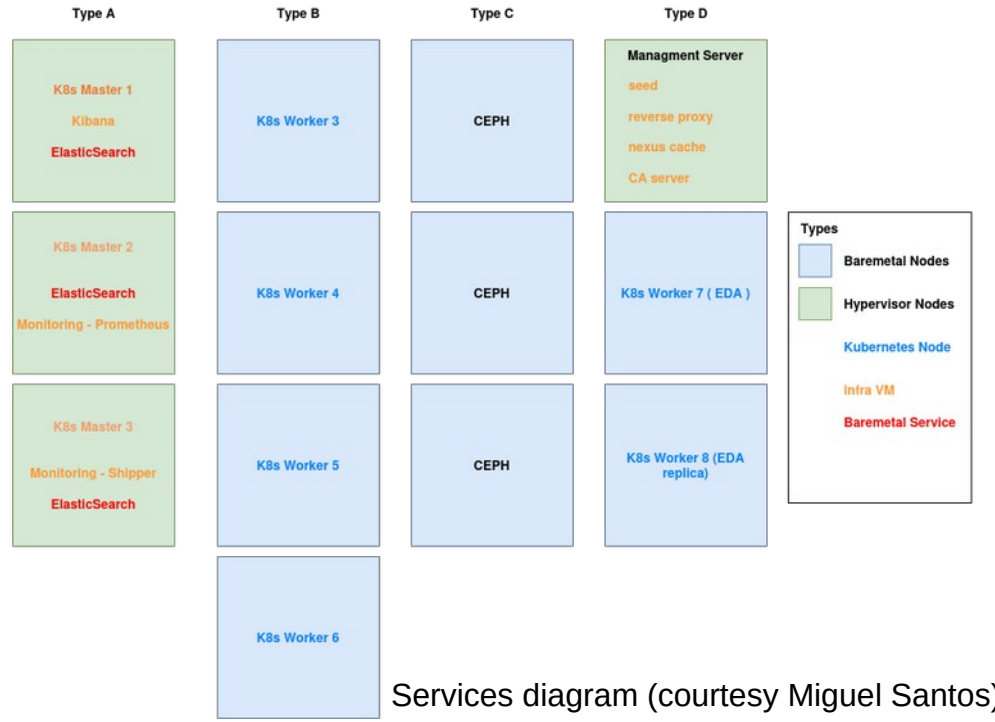
2024-11-28

Construction: AA0.5 (I)

- Early testing of antennas: use of software available in the community
- Antennas
  - Mid: 1st antenna on site, most parts of dishes in South Africa
  - Low: all four stations normally handed over to AIV
- Processing hardware
  - Mid in Karoo array processing building (on site) & Low in Pawsey (in Perth)
  - identical for Mid and Low: 13 HPE ProLiant DL385 Gen10 Plus v2 Servers to form 4 server types (A, B, C, D)

Milestone Event (earliest)		SKA-Mid	SKA-Low
Construction Approval		2021 Jul	2021 Jul
AA0.5 AIV start	4(3) dishes 4 stations	2025 Jun	2024 Jul
AA0.5 end	4(3) dishes 4 stations	2025 Dec	2024 Dec
AA1 end	8 dishes 18 stations	2026 Jul	2025 Nov
AA2 end	64 dishes 64 stations	2027 Jun	2026 Oct
AA* end	144 dishes 307 stations	2028 Apr	2028 Jan
Operations Readiness Review		2028 Jul	2028 Apr
End of Staged Delivery programme		Formal end of construction (including contingency): 2029 Mar	
AA4	197 dishes 512 stations	TBD	TBD

Schedule (courtesy Robert Lang, SKAO)



Services diagram (courtesy Miguel Santos)

Construction: AA0.5 (II)

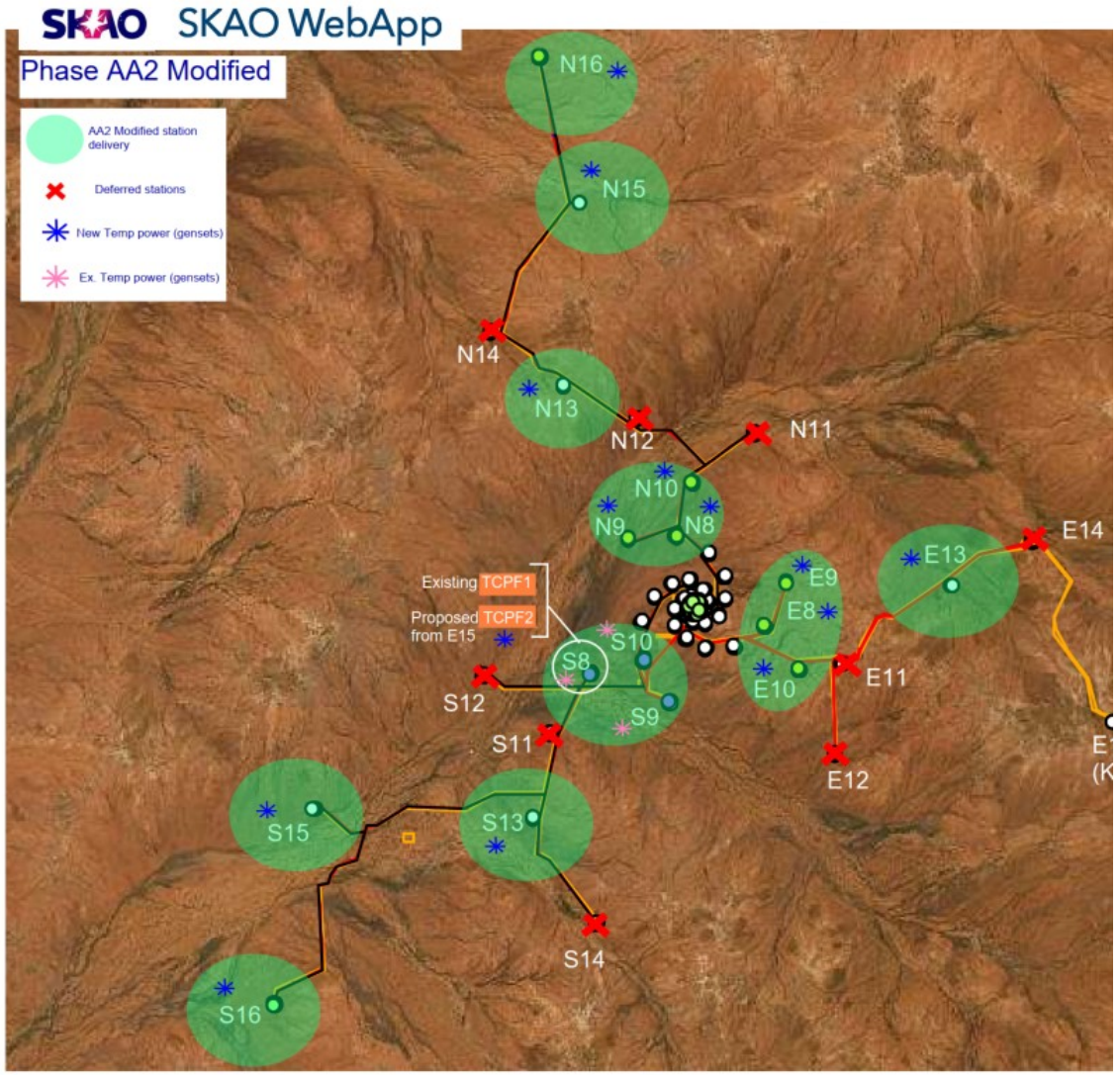
- Servers A, B and D intended for workloads
- Servers C intended for storage
- Kubernetes cluster
  - masters: 3 virtual A nodes
  - workers
    - 2 virtual A nodes
    - 6 bare metal B and D nodes

	Server Type A	Server Type B	Server Type C	Server Type D
CPU	64 Cores / 128 T (2 cpus)	64 Cores / 128 T (2 cpus)	64 Cores / 128 T (2 cpus)	64 Cores / 128 T (2 cpus)
RAM	256 GB (16 x 16 GB)	512 GB (32 x 16 GB)	512 GB (32 x 16 GB)	128 GB (16 x 8 GB)
SSD	1.92 TB (2 x 0.96 TB) 7.68 TB (2 x 3.84 TB)	1.92 TB (2 x 0.96 TB) 3.84 TB	1.92 TB (2 x 0.96 TB)	1.92 TB (2 x 0.96 TB)
DISK	-	-	200 TB (10 x 20 TB)	1 TB
Network	10Gb 2-port SFP+ BCM57412 OCP3 Adapter (2 ports) ConnectX-5 100GE 1P NIC (1 port)	10Gb 2-port SFP+ BCM57412 OCP3 Adapter (2 ports) ConnectX-5 100GE 1P NIC (1 port)	10Gb 2-port SFP+ BCM57412 OCP3 Adapter (2 ports) Broadcom P225p NetXtreme-E Dual-port 10Gb/25Gb Ethernet PCIe Adapter - NIC (2 ports)	10Gb 2-port SFP+ BCM57412 OCP3 Adapter (2 ports) ConnectX-5 100GE 1P NIC (1 port)
	x3	x4	x3	x3
CPU	192 Cores / 384 T	256 Cores / 512 T	192 Cores / 384 T	192 Cores / 384 T
RAM	768 GB	2048 GB	1536 GB	384 GB
SSD	5.76 TB 23.04 TB	7.68 TB 15.36 TB	5.76 TB	5.76 TB
DISK	-	-	600 TB (200 TB with ceph mirroring)	3 TB

Server types fo AA0.5 (courtesy Miguel Santos)

Progress on AA2

- Modification of Low AA2 layout
  - Central Processing Facility and Power and Signal distribution will be delivered late
  - move some AA2 stations from core to spiral arms to use available resources (Remote Processing Facilities)
  - impact on resolution (long baselines) & dynamics (fewer core stations) expected to lead to increased computing
- SDP processing hardware
  - staged delivery to manage risk
  - first delivery expected for November 2025
  - Pawsey has a power limit at 300 kW for AA2



Revised station layout for AA1 (courtesy SKAO)

SEAC: pipeline development review

- Review triggered by Council (~1 year ago)
- Review documentation delivered this summer, meeting in October, final report expected early 2025
- Preliminary report shared by panel is quite critical concerning
  - status of development
  - feasibility risks for AA2 and AA*
  - incompatibility of intended use of the telescopes with the hardware budget
  - insufficient use of existing community solutions

Next steps

- Reorganisation
  - Miles Deegan to focus on SDP and join the DP ART
  - co-design teams to merge: SCOOP (in-kind) and RACOON (contracted)
  - co-design team to work as a regular SAFe team to allow better inclusion in overall effort
- Tension between producing feature-complete software (TBD for MVP) and benchmarking / optimisation effort
  - SCOOP to endorse this effort (TBC)
  - request for SCOOP to work on optimisation
    - ownership issue remains (eg. following the findings and proposal made by Clément Devatine during his internship at Eviden)
    - continue working with teams SCHAAP & PANDO who have a longer history of working with core developers of DP3 and WSClean

Low Goal 5

- Deliver an MVP for end-to-end processing for continuum imaging by the end of PI25 (March 2025)
  - intended in a realistic SDP context (monitoring and control) but first planned as manual (Slurm script)
- Test platform from AWS and benchmarking intended to inform November 2025 hardware order
  - head node: single t3a.xlarge instance
  - different queues to use different hardware (bare metal) and scale
  - 2.4 TB Lustre high performance storage at 480 MBs/s
  - 2 S3 bucket for input/output data

AWS queues (courtesy Bojan Nikolic)

Queue	Instance Type	vCPUs	RAM	Network Performance	Queue Max Intances	Hourly Cost (On Demand)	Hourly Cost (Spot)
c7i-metal-24xl-ondemand / spot	c7i-metal-24xl	96	192 GB	37.5 Gbit	6	$5.09	$1.43
c7i-metal-48xl-ondemand / spot	c7i-metal-48xl	192	384 GB	50 Gbit	3	$10.18	$2.82
r7i-metal-24xl-ondemand / spot	r7i-metal-24xl	96	768 GB	37.5 Gbit	6	$7.46	$2.01
r7i-metal-48xl-ondemand / spot	r7i-metal-48xl	192	1536 GB	50 Gbit	3	$14.92	$4.02

Development strategy

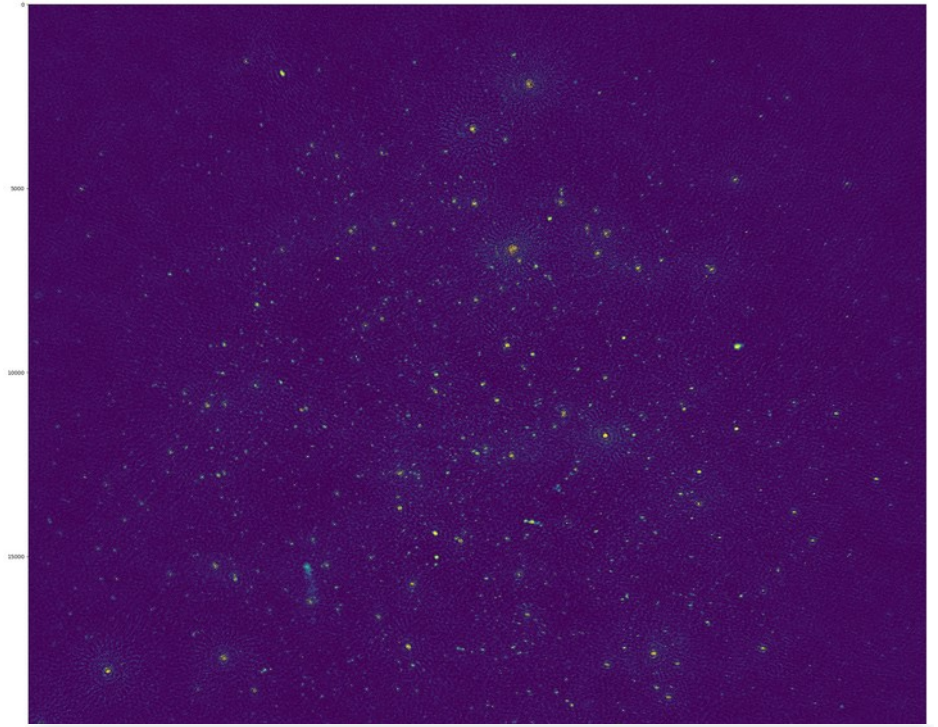
- DP3 and WSClean as building blocks for AA2
  - meet schedule constraint
  - build on existing software
- DP ART uncertain on whether is can scale further
  - AA* is 2 years later with numbers of visibilities increased by ~5 (Mid) and ~23 (Low)
  - AA4 and subsequent extensions are still larger
- Develop intrinsically distributable pipelines
  - MSv4 & xradio (based on xarray) to replace MSv2
  - Imaging swiFTly scheme: distributed Fourier transforms and w-towers (Wortmann et al, 2024)

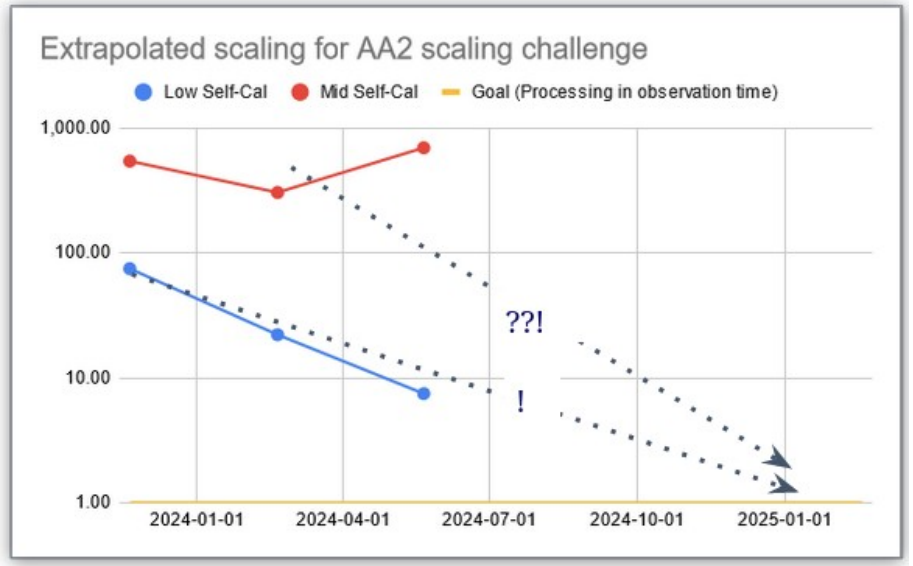
Status of pipelines for AA2

- Test with simulated data based on AA1 configuration (as will be available by November 2025)
  - flux calibration MS, polarisation MS (x4), bandpass, delay, complex gain MS, target observation MS
- Pipelines
  - instrumental calibration: delay, bandpass, complex gain, polarisation leakage, polarisation angle, [illegible] density
  - pre-processing: flagging, a priori calibration, averaging (frequency, time)
  - self-calibration and continuum imaging pipeline
- Risks
  - integration
  - performance

ICAL pipeline

- ICAL: self-calibration pipeline still under development
  - derived from Rapthor
  - runs 9 cycles (7 originally in Rapthor)
  - DP3 for calibrate and predict
  - WSClean for imaging
  - parallelisation on time and frequency by running separate instances of DP3 and WSClean via Dask workers
- Tested on calibration field (LOFAR observation)
- Recent benchmarking and optimisation
  - still ~7x behind the objective of processing data as fast as it is acquired





Performance improvement (courtesy Peter Wortmann)

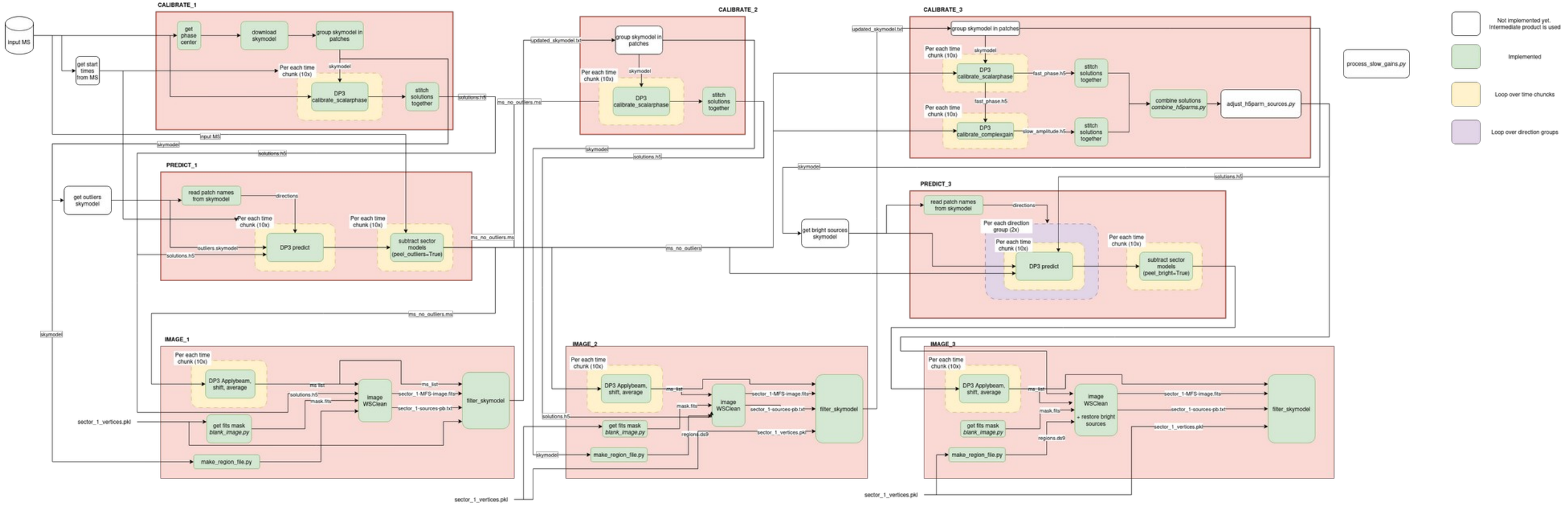
Rapthor pipeline (I)

- Cycle 1
  - Calibrate_1: based on the phase center of the MS, the skymodel is downloaded and grouped in patches Those are used for a phase-only calibration.
  - Predict_1: the sources which lie outside the field we want to image are predicted and subtracted in the visibility space. This produces the Ms: ms_no_outliers, which is further used throughout the pipeline.
  - Image_1: The dataset is imaged and the solutions in calibrate_1 are applied while imaging.
- Cycle 2
  - Calibrate_2: using the new skymodel produced by image_1, run a phase_only calibration.
  - Image_2: Same as image_1, whith updated solutions.

Rapthor pipeline (II)

- Cycles 3, 4, 5, 6:
  - Calibrate_*: using the new skymodel produced by image_2, run two DDECal: scalarphase and complexgain. The solutions are then combined.
  - Predict_*: extract the bright sources from the skymodel and predict them. Those are subtracted from the MS in the visibility space. This creates the MS: ms_no_outliers_no_bright_sources, wh is further used for imaging.
  - Image_*: the dataset is imaged. Afterwards, the bright sources are put back in the image and in the output skymodel.
- Cycles 1-6 run on 20% of data and 7th on 100%

Rapthor pipeline (III)



Schematic view of the first 3 cycles of Rapthor (courtesy Chiara Salvoni)

ECLAT contributions to the SDP challenge

- SKAO not really looking at alternative community software
  - study DDF-pipeline and establish a fair basis for comparison for risk management
- Longer term proposal for SKAO's scalable pipelines
  - baseline partitioning for parallelisation (Sunrise)
  - NumPEx tools (Exa-DoST, Exa-AToW)
- Sustainable computing (also applicable to SRC-FR (with additional constraint of distributed storage and computing)
  - power consumption and environmental impact internships @ Avalon/Lagrange
  - optimise constrained execution
    - compile-time mapping and scheduling: PREESM and sim-sdp @ Dark Era
    - multi-node mapping, scheduling and resource management @ Exa-AToW

SRCNet and SRC-FR

- update from Chiara Ferrari

We recognise and acknowledge the Indigenous peoples and cultures that have traditionally lived on the lands on which our facilities are located.

SKAO

www.skao.int