Grant number: 101093934

WP5: Data processing toolkit for Advanced Radio Astronomy

M. Verkouter (JIVE) & R. Beswick (e-MERLIN, UKSRC / JBCA, University of Manchester)

on behalf of WP5 team

Grant number: 101093934

WP5 Overview & objectives

- Main Objectives: Development of modular, open-source and flexible toolkit components for associated workflows. Components to enable rapid, reproducible and scalable analysis tools. Strong emphasis on translation of knowledge & developments between different but adjacent RIs (*pooling, sharing and developing*)

- Partners: 16 partners (UNIMAN, ASTRON, JIV-ERIC, CSIC, MPG, VIRAC, ULEI, SDU, INAF, SKAO, EPFL, Radboud, UPretoria, Heidelberg, RATT, ICRAR, OAN)

New associate partners are joining WP5

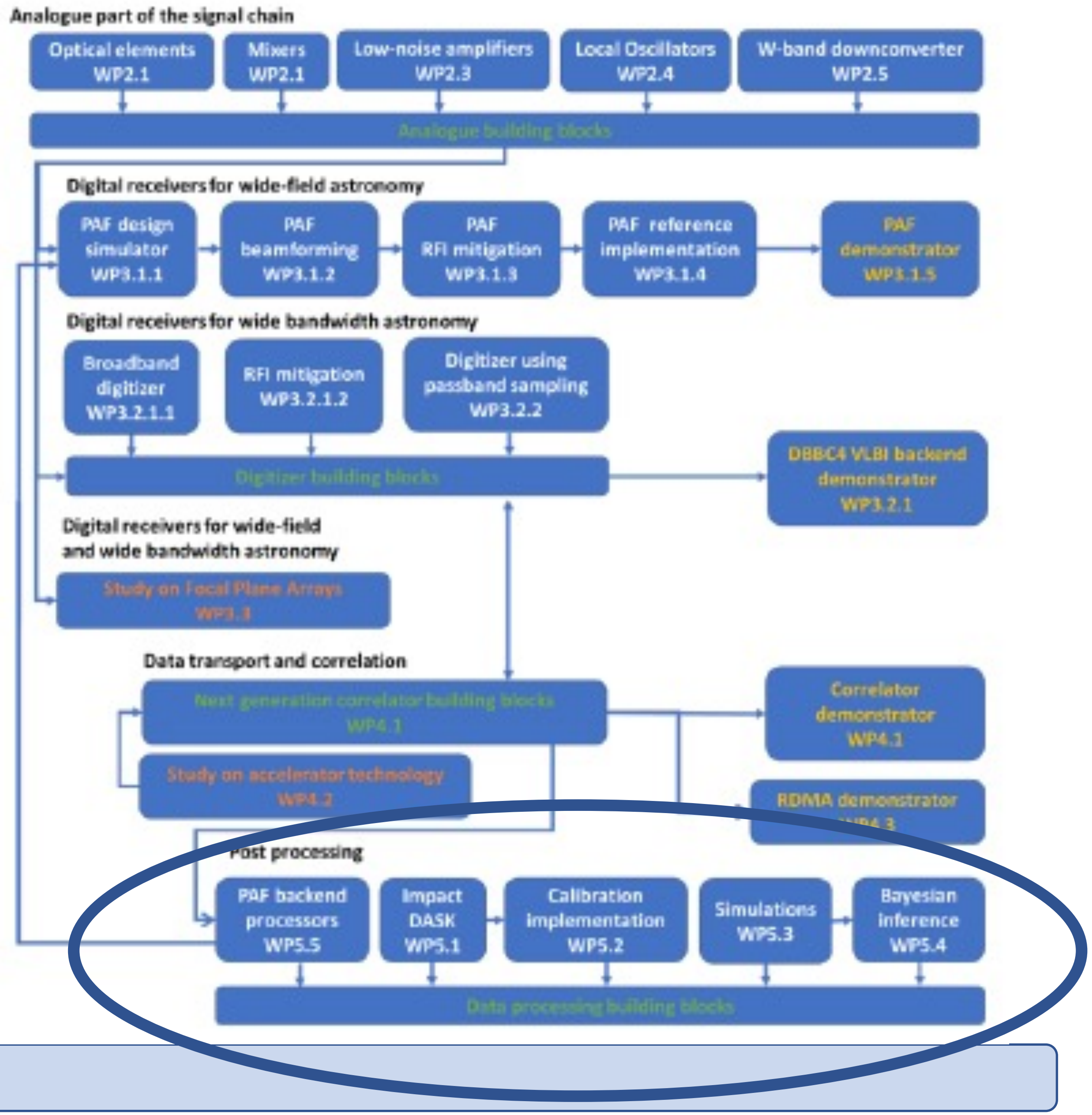
– Increased outputs, visibility and expand experience of the team

WP5: Advanced Data processing

WP5 provides key elements in of the e-2-e facility signal chain

- key part of the full stack
- Essential to maximise outputs from facilities
- supports and integrates with WP3 & WP4 (in particular)

Essential toolkit for users of radio astronomy facilities



WP 5 : Advanced Data processing

Task 5.1	Task 5.2	Task 5.3	Task 5.4	Task 5.5
DASK workflows	Scalable fringe fitting	Optimizing calibration	Bayesian inference	PAF processing toolkit

- Developing the foundation blocks for future processing tools for multiple radio astronomy infrastructures - inc. EVN/VLBI, EHT, LOFAR, e-MERLIN, Effelsberg, SKA

WP5 structure & tasks:

5 key task areas:

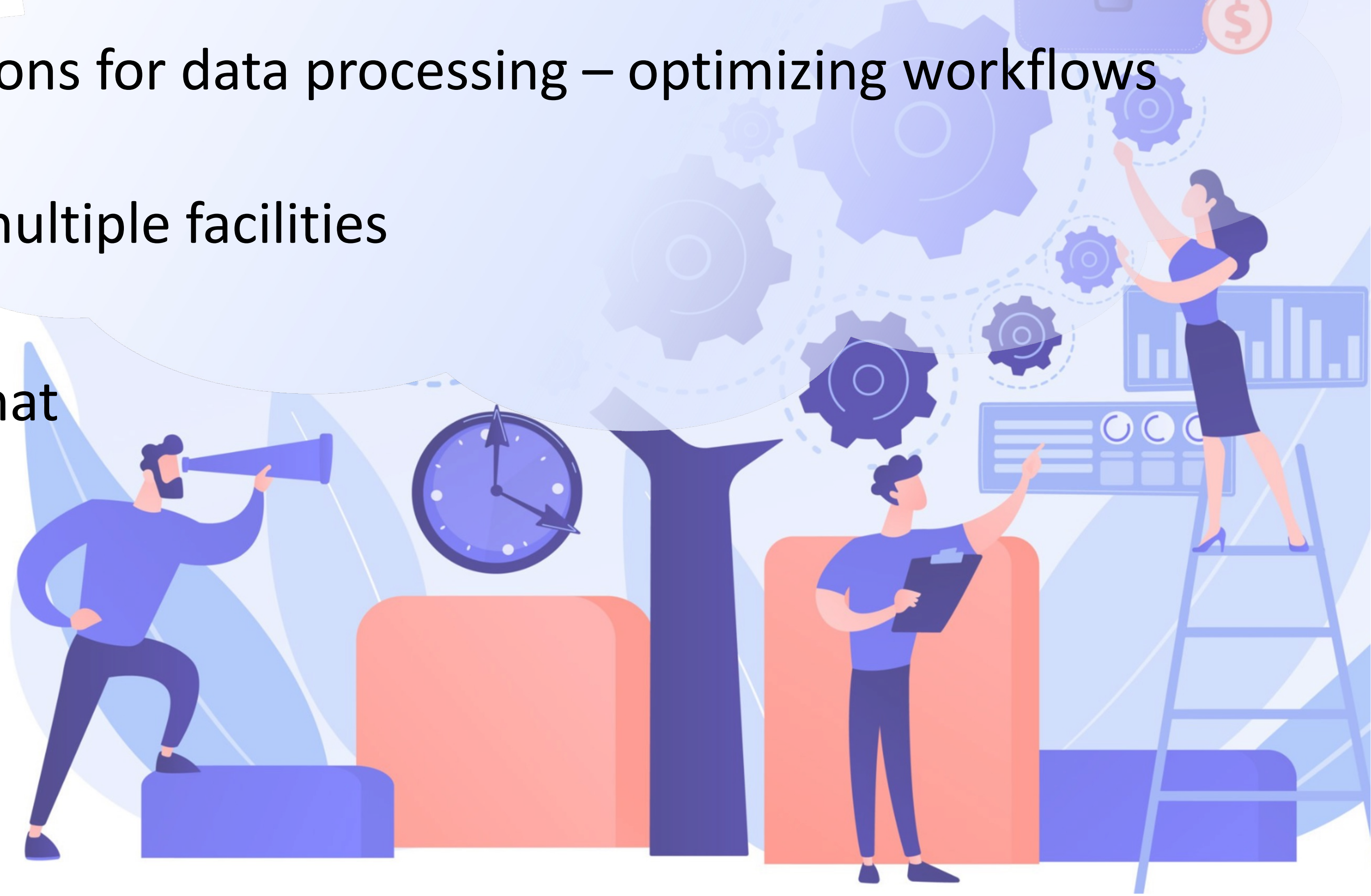
- Task 5.1 : *The impact of DASK on automated processing workflows for Radio Astronomy data (**ASTRON**,VIRAC, UNIMAN, EPFL, SKAOb, RATT)*
- Task 5.2 : *Develop a generic and scalable fringe fit calibration implementation in the Dask framework (**JIVERIC**)*
- *Task 5.3 : Simulations for optimising calibration and parameter extraction (**JIV-ERIC**, UNIMAN, Radboud, CSIC, UP, ICRAR, OAN)*
- *Task 5.4 : Bayesian inference for sparse visibility data (**ULEI, SDU**, INAF)*
- *Task 5.5 : Modular PAF Backend Processors toolkit (**MPG**)*

- Each task coordinated by independent teams but brought together under single umbrella to share knowledge & expertise.

- Modular, open-source and flexible components to process interferometry data

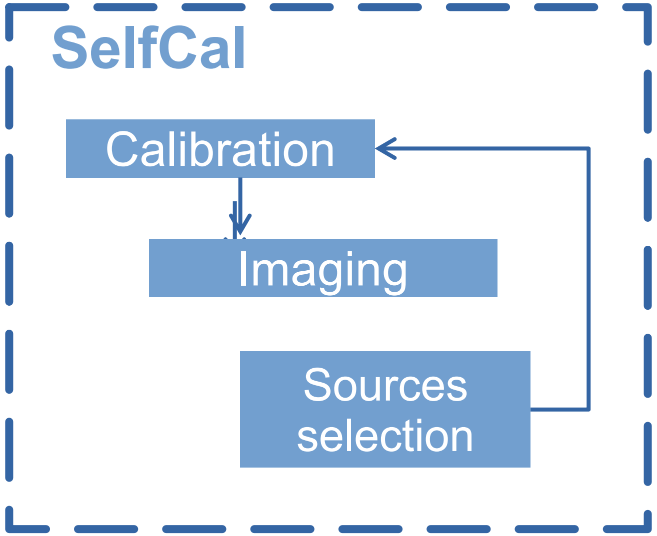
WP 5.1 & 5.2 DASK for automated workflows

- FAIR and scalable solutions for data processing – optimizing workflows and pipelines
- Uniform approach for multiple facilities
- DASK'y' fringe fitter
- Discussion on data format



WP5.1: Optimizing pipelines execution

From the general picture

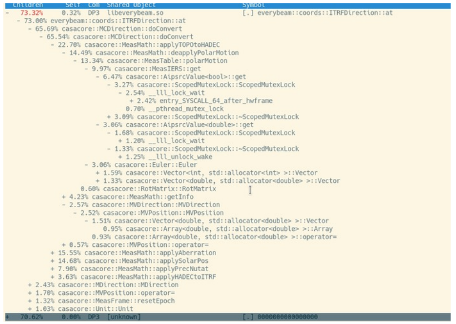


We are using reframe[1] as to ensure reproducibility of the results.
Together with perfmon[2] which provides timeseries of:

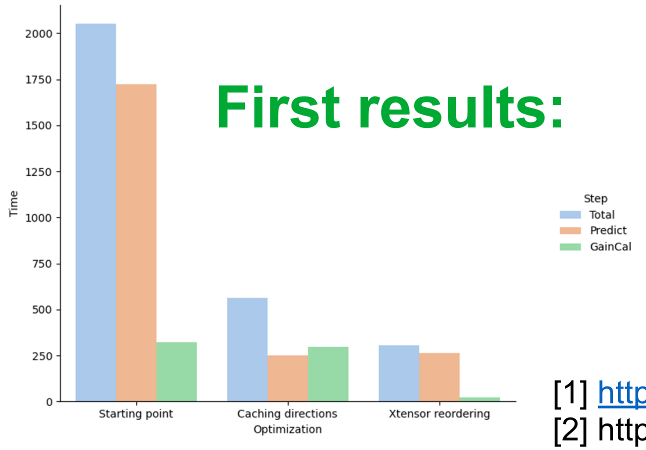
- CPU usage
- network traffic
- memory usage
- power usage

… single step performance…

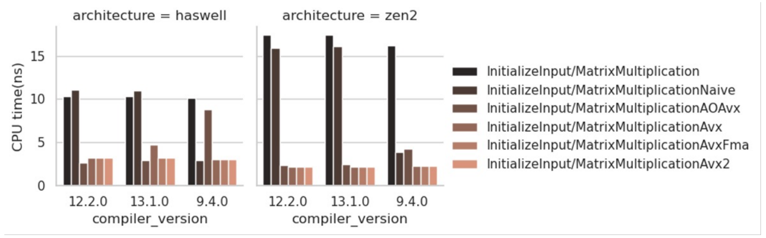
Using perf we analyzed the details of the single step to increase the performances



First results:



… to inspect a single function



Using the google benchmark framework we explored different implementations on different compilers/architectures

[1] https://git.astron.nl/mancini/ska-sdp-benchmark-tests - extends SKA repo
[2] https://gitlab.com/ska-telescope/sdp/ska-sdp-perfmon

***ASTRON**,VIRAC, UNIMAN, EPFL, SKAOb, RATT*

WP5.2 Fringe fit in DASK framework

- Mission: write code
- Since project start: rudimentary FFT based `fringefit` in Dask
- have a rudimentary least-squares `fringefit` too
  - Both for one spectral window
  - Both suitable for parallel graphs
- We make and execute Dask graphs
- There is much still to do!
  - Combining spectral windows
  - Calibration tables
- Lot of collaboration with NRAO (USA)'s ngCASA/AstroVIPER/xradio/RADPS
  - Very close to "core development team"
  - Involved in data format definition (MSv4) and e.g. calibration table format
    - Exploring cloud-native data formats

WP5.3 Optimizing calibration

- VLBI data processing is perceived as DIFFICULT
- Relies on human experience: it is also biased

--> Automate calibration choices where possible

- Synthetic data generation
- Dynamic imaging of sparse datasets (EHT)

Use cases: EVN, SKA-VLBI

FEATURES	SYMBA	VLBI Simulator	OSKAR	pyuvsim	EHTsim
Telescope or array settings					
antenna position	✔	✔	✔	✔	✔
frequency/multi-frequency	228 GHz	1.4 GHz	1 GHz	1.4 GHz	230 GHz
bandwidth	2 GHz	2 GHz	2 GHz	2 GHz	2 GHz
channels	64	64	64	64	64
observing time	24 hr	12 hr	12 hr	12 hr	24 hr
visibility structure	ms / uvfits	ms	ms	uvdata	uvfits
Input source structure					
image	✔	✔	✔	✘	✔
variability	✔	✘	✘	✘	✔
input model format	ascii / fits	txt	ascii	yaml	txt
Direction independent effects					
thermal noise	✔	✔	✔	✔	✔
calibration effects	✔	✔	✔	✔	✔
signal corruption	✔	✔	✔	✔	✔
Direction dependent effects					
beam	✔	✔	✔	✔	✔
troposphere	✔	✘	✘	✘	✘
ionosphere	✔	✘	✘	✘	✘
pointing errors	✔	✔	✔	✔	✔
geometrical effects	✔	✔	✔	✔	✔
Data output format					
ms	✔	✔	✔	✔	✘
hdf5	✘	✘	✘	✔	✘
uvfits/fits-id	✔	✘	✔	✔	✔
ascii	✔	✔	✘	✘	✘
xarray	✘	✘	✘	✘	✘
Polarization structure					
full stokes	Yes	Yes	No	Yes	Yes
faraday rotation	Yes	No	No	No	No
Hardware/software requirements					
python	✔	✔	✔	✔	✔
docker	✔	✘	✔	✘	✔
singularity	✔	✔	✔	✘	✘
Additional requirements to consider					
wide field	✘	✔	✘	✔	✘
subarraying	✘	✘	✔	✘	✘
multiple beams	✘	✔	✔	✔	✔

Milestone 5.3

***JIV-ERIC**, UNIMAN, Radboud, CSIC, UP, ICRAR, OAN*

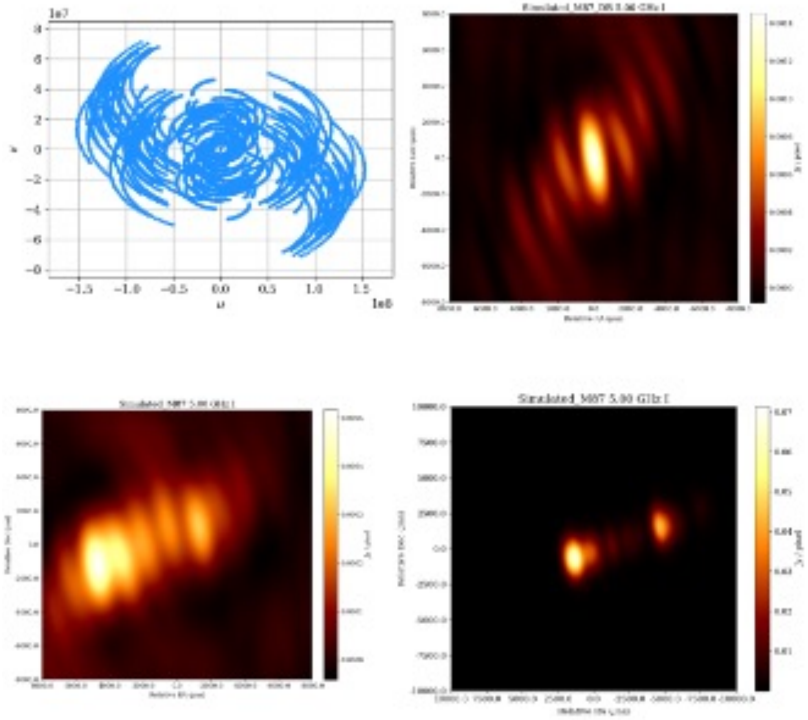
Synthetic data generation

- Software selection
- Benchmark datasets
- Astronomical datasets with realistic errors
- Vary calibration settings
- Compare results

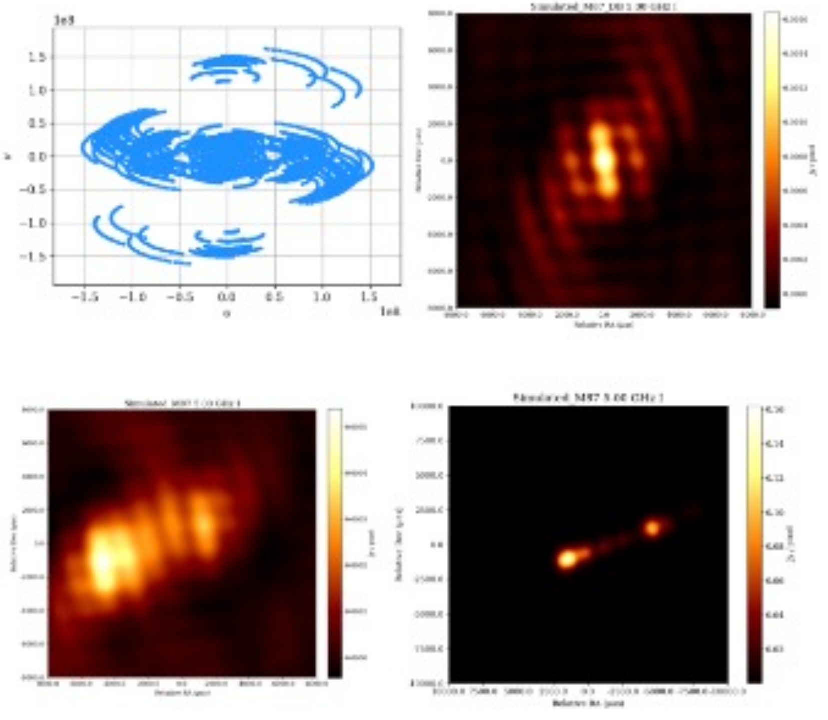
Grant number: 101093934

Synthetic data generation

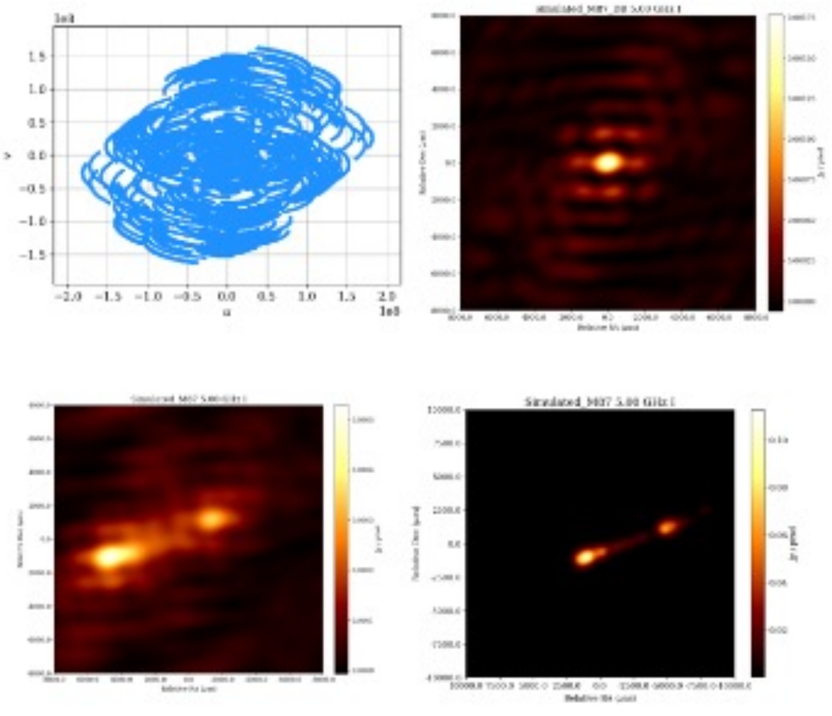
CASE 1: EVN (only NE antennas)



CASE 2: EVN + SKA core (4 km)



CASE 3: EVN + SKA + AVN



WP5.4 Bayesian Inference imaging – workplan

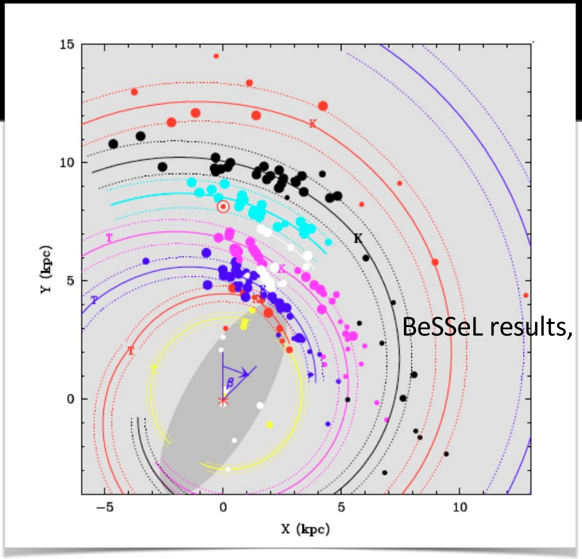
- Objectives
  - Development of modular, open-source and flexible toolkit components for associated workflows
  - deliver error distributions (posteriors) on fitted parameters
    -> provide quantitative assessments of model predictions
  - Assess performance of these methods using simulated datasets
  - Strong emphasis on translation of knowledge & developments between different but adjacent RIs
- Partners/Resources
  - Leiden, SDU->Heidelberg(MPI), INAF
- Milestones
  - Expand user base by porting existing methods to more user-friendly implementations
  - Apply methods to a wider VLBI context (EVN, ...)
- Planning
  - Deliverable M48
  - But could depend on finding right people

***ULEI, SDU**, INAF, Heidelberg*

- Early planning and code design phase
  - Settle on set of “wanted” code features for VLBI modeling code [mature state, coding in fall]
    - VLBI data products: Visibility amplitudes, Various closure quantities, Complex visibilities
    - Calibration components: Gain solvers, D-terms, etc
    - Powerful/versatile samplers for a Bayesian method returning value and uncertainty
  - Clearly understand and define “interfaces” to other RADIOBLOCKS [this meeting]
    - Data formats
    - Workflow
    - Parallelization strategy (DASK?)
    - GPUs and other computer architectures
    - Code development [git, (github)]
- Hires:
  - Pascal Keller (PhD student with Huib)
  - SDU postdoc [delayed]
- Regular meetings have begun
  - Push towards first Bayesian VLBI astrometry

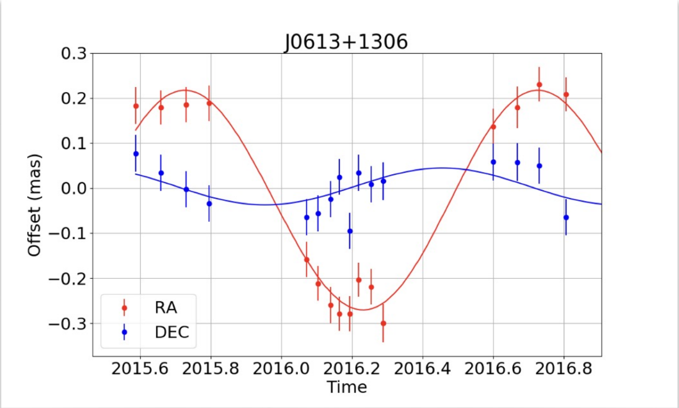
Case study: (Maser) astrometry

- Large campaigns, like BeSSeL rely on phase referencing + imaging
    - Estimate accuracy of phase referencing from residuals
- Could be done more robustly
- Explored in pilot project

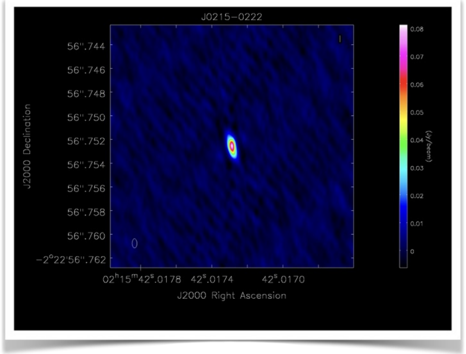


BeSSeL results, Reid et al., 2019

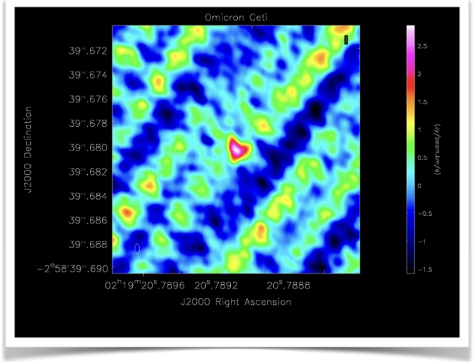
Phase ref monitoring each source for multiple epochs, e.g. S269 Quiroga-Nunez et al., 2019

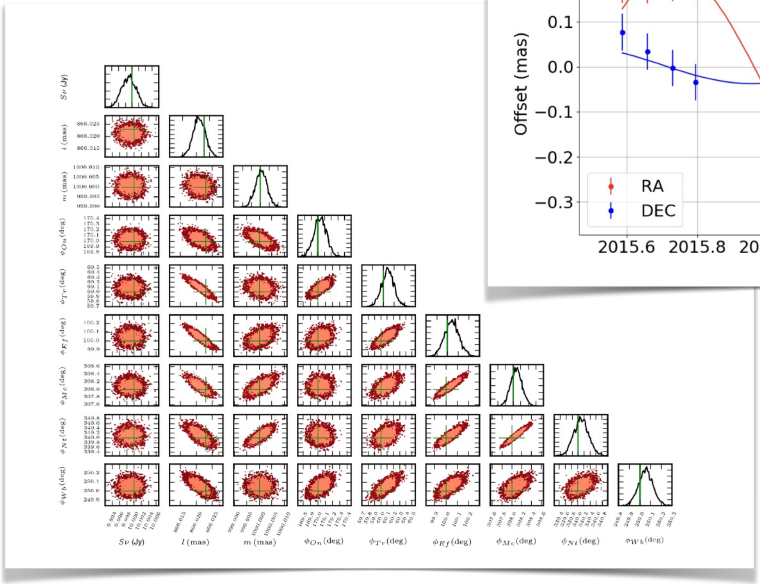


Reference quasar



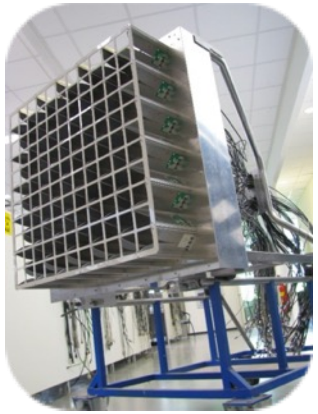
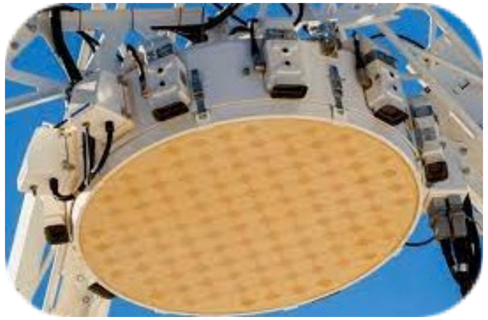
SiO maser star

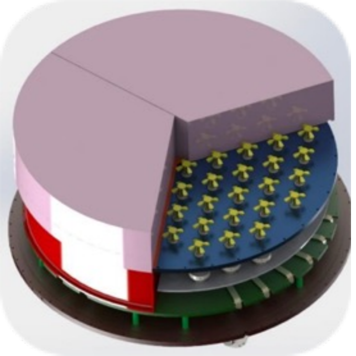




MCMC inference of position and phase calibration
Artificial data, van Langevelde priv com.

WP5.5: Modular PAF Backend Processor Toolkit



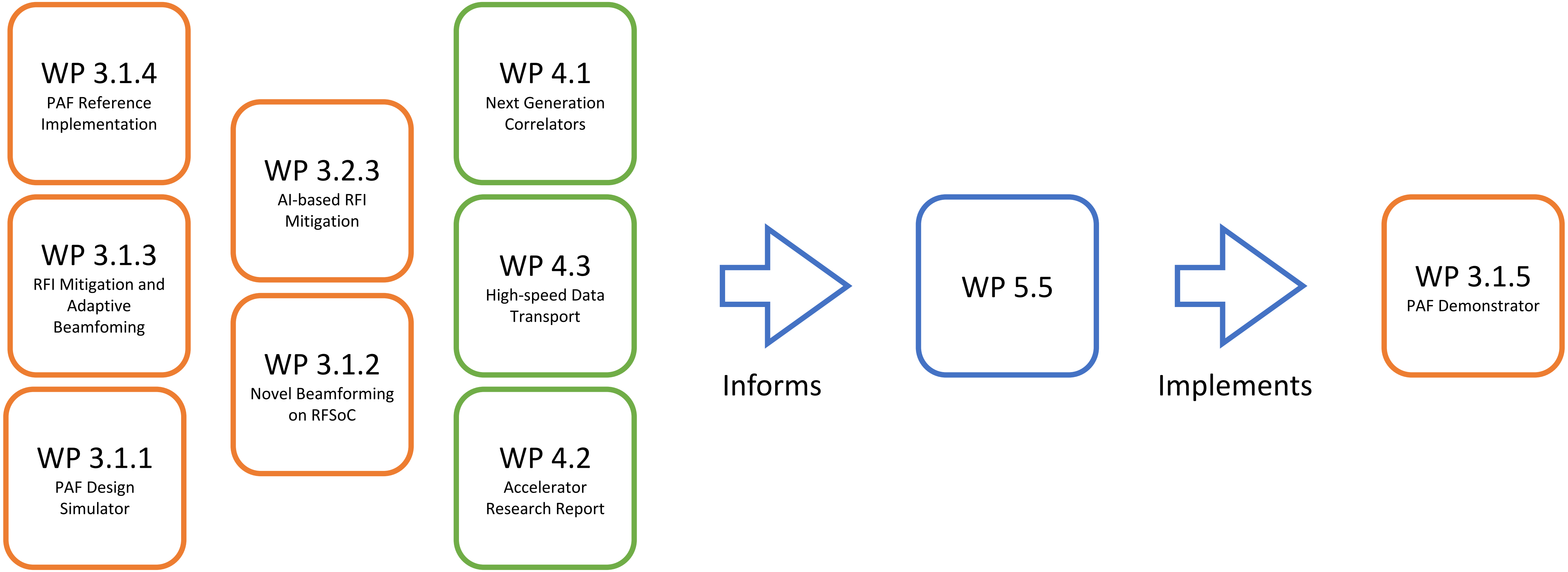
Currently deployed PAFs, such as those used by RIs as the SKA pathfinder and precursor facilities APERTIF and ASKAP have custom backend solutions to support beamforming and correlation. Whilst these address the needs of the specific facilities, they do not provide a common, efficient, nor open-source solution that can be easily reused by other facilities.

[Task 5.5 will] development of a general PAF backend solution based on commercial-off-the-shelf (COTS) hardware. The design goal is to produce a modular and scalable real time PAF backend processing solution, encompassing modules for PAF beamforming and the production of typical radio astronomical data products (e.g. full-Stokes spectra, pulsar timing and search data).

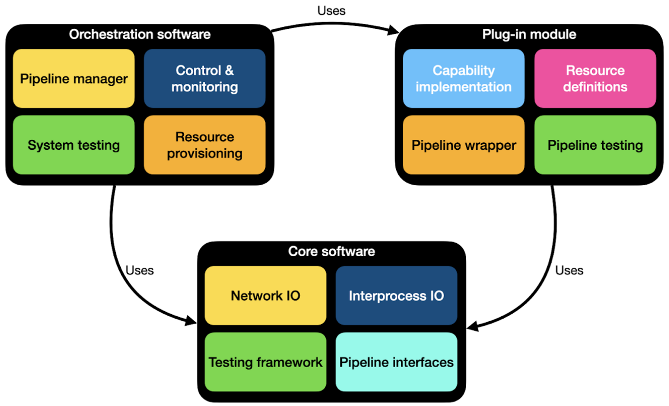
D5.7: Software/firmware repositories containing developed code and documentation (Task 5.5, month 48)

MPIfR

WP5.5 Synergies and links across RadioBlocks

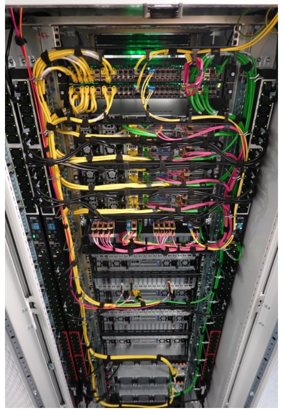


WP5.5 Progress



Framework design for PAF processing toolkit, based on core / plug-in / orchestration architecture.

Initial module development and testing for beam former, correlator and VLBI processor.



Design, purchase, deployment and installation of the EDGAR computing cluster @ Effelsberg for PAF testing and development.

32 GPU servers + 8 VM hosts
80 x AMD EPYC 9354 CPUs
64 x Nvidia L40 GPUs
Nvidia Spectrum-4 800 GbE switch
Nvidia Quantum-2 400 Gb/s Infiniband switch

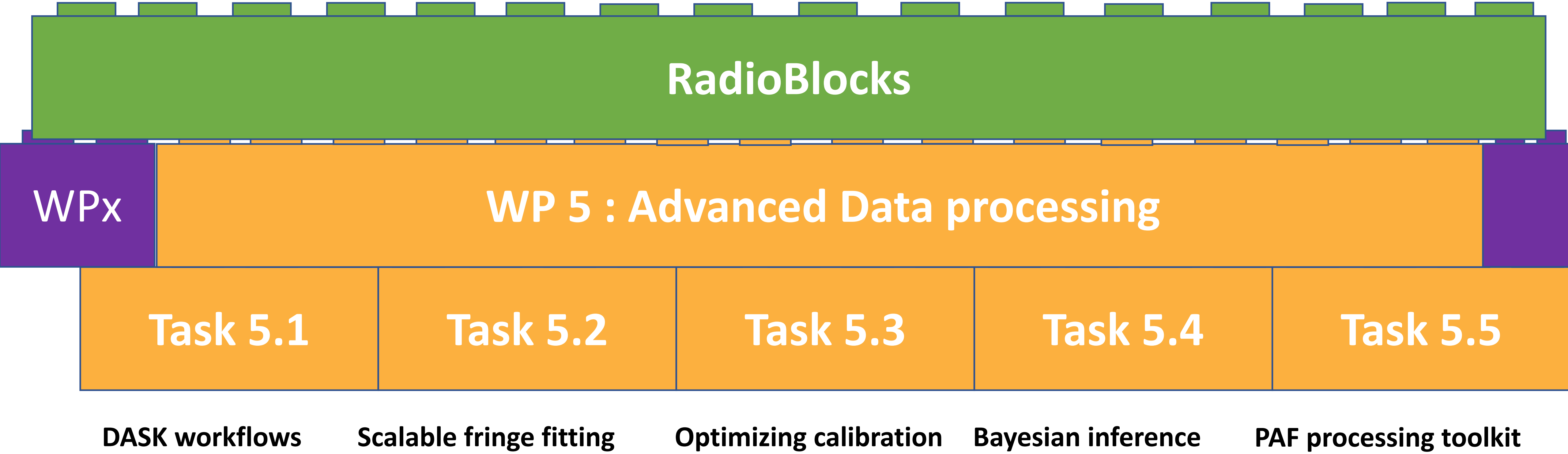
Demonstration of GPU-based VLBI module using a single pixel feeds on Effelsberg and the Thai National Radio Telescope.

First demonstration of VLBI with a telescope in Thailand.

Deliverables & Milestones

- *D5.1: Library of DASK-accelerated interface to analyse radio astronomy data formats (EPFL, Task 5.1, month 12) –* ***Complete***
- *Milestone 5.1: Report on granularities suitable for efficient fringe fitting (TP5.2) –* ***completed***
- ***Milestone 5.2 :*** *Interface & conversion software (CASA MS format and Python GPU DataFrame library cuDF) (TP5.1) –* ***Completed***
- **Milestone 5.3:** *Report on generic VLBI simulation Strategies (TP5.3) –* ***Completed***
- **Milestone 5.4**: *Report on modular architecture (TP5.1) – Month 28*
- **Milestone 5.4:** *Assessment of Bayesian package links (TP1,2,3) – Month 36*
- ***Milestone 5.4 :*** *Advocating package for early-adopters (ALL TPs) – Month 42*
- ***D5.2:*** *Implementation of a fringe fit algorithm in the DASK framework (JIV-ERIC, Task 5.2, month 48) -*
- ***D5.3:*** *Prototype processing workflow functionality using software modules (e.g. AOflagger, Image-Domain-Gridder) under DASK framework (ASTRON, Task 5.1, month 42)*
- ***D5.4:*** *Port of Bayesian package to make use of DASK-like HPC methods (SDU, Task 5.4, month 48)*
- ***D5.5:*** *Demonstrated usage of end-to-end simulation tool for single RI (JIV-ERIC, Task 5.3, month 48)*
- ***D5.6:*** *Full-Stokes dynamical imaging algorithm for time-variable VLBI sources (CSIC, Task 5.3, month 48)*
- ***D5.7:*** *Software/firmware repositories containing developed code and documentation (MPG, Task 5.5, month 48)*
- ***D5.8:*** *DASK accelerated data reading & GPU processing for radio astronomy tools (EPFL, Task 5.1, month 48)*

Number of deliveries are documented s/w, code for open access. Largely at end of project.



- Developing the foundation blocks for future processing tools for multiple radio astronomy infrastructures - inc. EVN/VLBI, EHT, LOFAR, e-MERLIN, Effelsberg, SKA

Grant number: 101093934

All-Hands Splinter sessions & Wp5 activities

Further plenary talks related to WP5:

- The Africanus Software Ecosystem: An Overview – Tuesday Afternoon
- Dynamic Imaging – Wednesday Morning
- GPUs and Radio Astronomy – Wednesday Morning
- ICRAR contribution to RadioBlock – Wednesday Morning

Multiple Splinter session- all welcome

- **Wednesday afternoon & Thursday all-day**
  - Including – Wp4+5 crossover, DASK hackathon, DASK presentation (Simon Perkins), WP5 – near/mid-term planning, Representative Data sets and formats (WP4 linkage) and Simulations (WP5)

Grant number: 101093934

End?

WP5 structure & tasks:

- **Task 5.1 :** ***The impact of DASK on automated processing workflows for Radio Astronomy data*** *(**ASTRON**,VIRAC, UNIMAN, EPFL, SKAOb) - XXPM*
    - DASK impact investigations – toolkits
    - Leverage and augment LOFAR related developments.
    - Availability and deployment for multiple facilities
    - Scalability for large data-sets and compatibility with HPC
    - Open-source available code and repositories.

- **Task 5.2 :** ***Develop a generic and scalable fringe fit calibration implementation in the Dask*** *framework (**JIVERIC**)*
    - *Extension to Fringe fitting*
    - *DASK implementation (linking to 5.1) – portability and flexibility*
    - *scalability from single CPU to cluster environment*

Each task coordinated by independent teams but brought together under single umbrella to share knowledge & expertise.

WP5 structure & tasks:

- ***Task 5.3 : Simulations for optimising calibration and parameter extraction*** *(**JIV-ERIC**, UNIMAN, Radboud, CSIC, UP)*
  - *Tools to automate and improve VLBI calibration parameter estimation (ML)*
  - *Via end-2-end simulations characterise errors and propagation through calibration – feed into Bayesian inference T5.4)*
  - *Expand existing tools (e.g. mm-VLBI/EHTC) to include generic applications for cm-VLBI etc.*
  - *Scalability (links to T5.1, T52.) and implementation of ML techniques*
  - *Link to SKA-VLBI Task force for joint development of simulation toolkits.*
- ***Task 5.4 : Bayesian inference for sparse visibility data*** *(**ULEI, SDU**, INAF)*
  - *Bayesian inference methods for VLBI*
  - *Build on and extend EHTC analysis*
  - *Develop as scalable application (link to T5.1-5.3)*
- ***Task 5.5 : Modular PAF Backend Processors toolkit*** *(**MPG**)*
  - *Modular & Scalable analysis toolkit for PAFs.*
  - *Directly link to WP3*

Each task coordinated by independent teams but brought together under single umbrella to share knowledge & expertise.

Resources & Partner spread:

Partners	UNI MAN	ASTRON	JIVE	CSIC	MPG	ULEI	VIRAC	SDU	INAF	SKAO	EPFL	Radb oud	UP
Task 5.1	x	**X**					x			x	x		
Task 5.2			**X**										
Task 5.3	x		**X**	x								x	x
Task 5.4						**X**		**X**	x				
Task 5.5					**X**								
Total effort (PM)	10	42	42	22	43	16	19.2	24	3	0	24	8	36

*Non-cost associated partners:
UNIMAN – funded via UKRI Horizon Guarantee scheme
EPFL – via Swiss, UP - South Africa, ICRAR, OAN, RATT
Plus SKAO