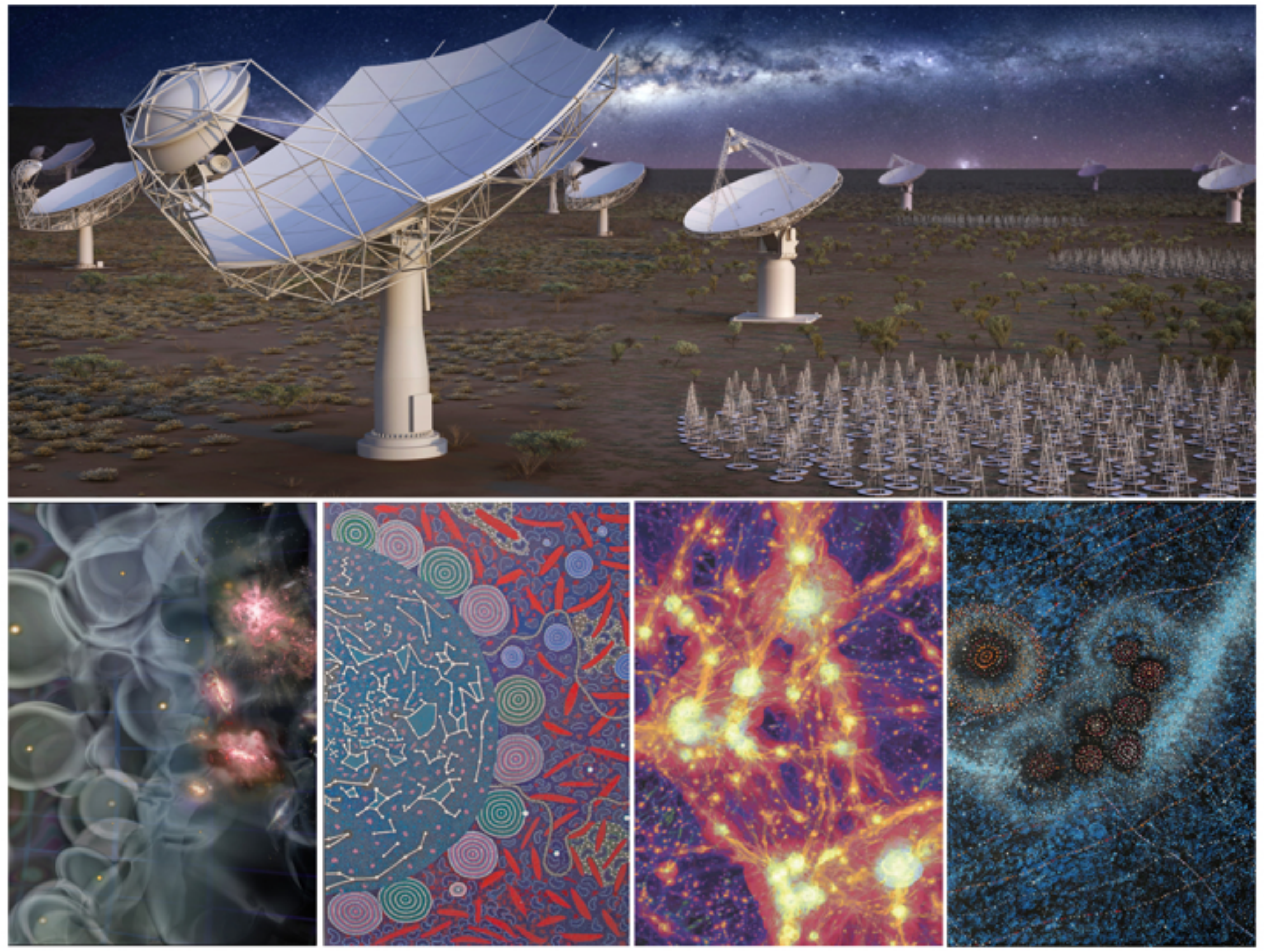
NumPEx

Exascale computing

The French NumPEx project and its many roles in the SKA project

Jean-Pierre Vilotte

Scientific Deputy (DS) for HPC/HPDA/AI & open science,
Institut des Sciences de l'Univers (INSU), CNRS (France)

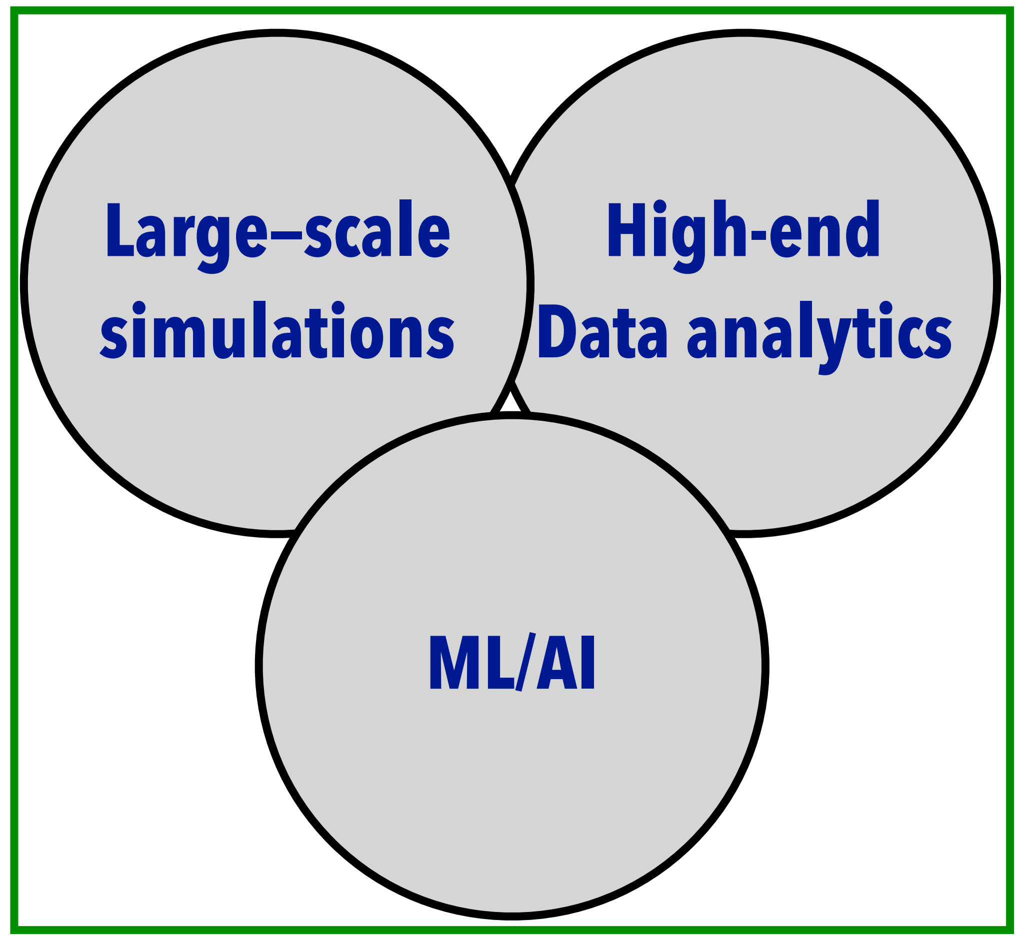


HPC challenges for new extreme scale applications
Aristote, Paris, March 6-7, 2023

NumPEx: Capable Exascale challenges

NumPEx: 5 years national project (CNRS, CEA, INRIA, Universities) - 40,8 M€

Coordinators: J.Y. Berthoud (INRIA), J. Bobin (CEA), M. Daydé (CNRS)

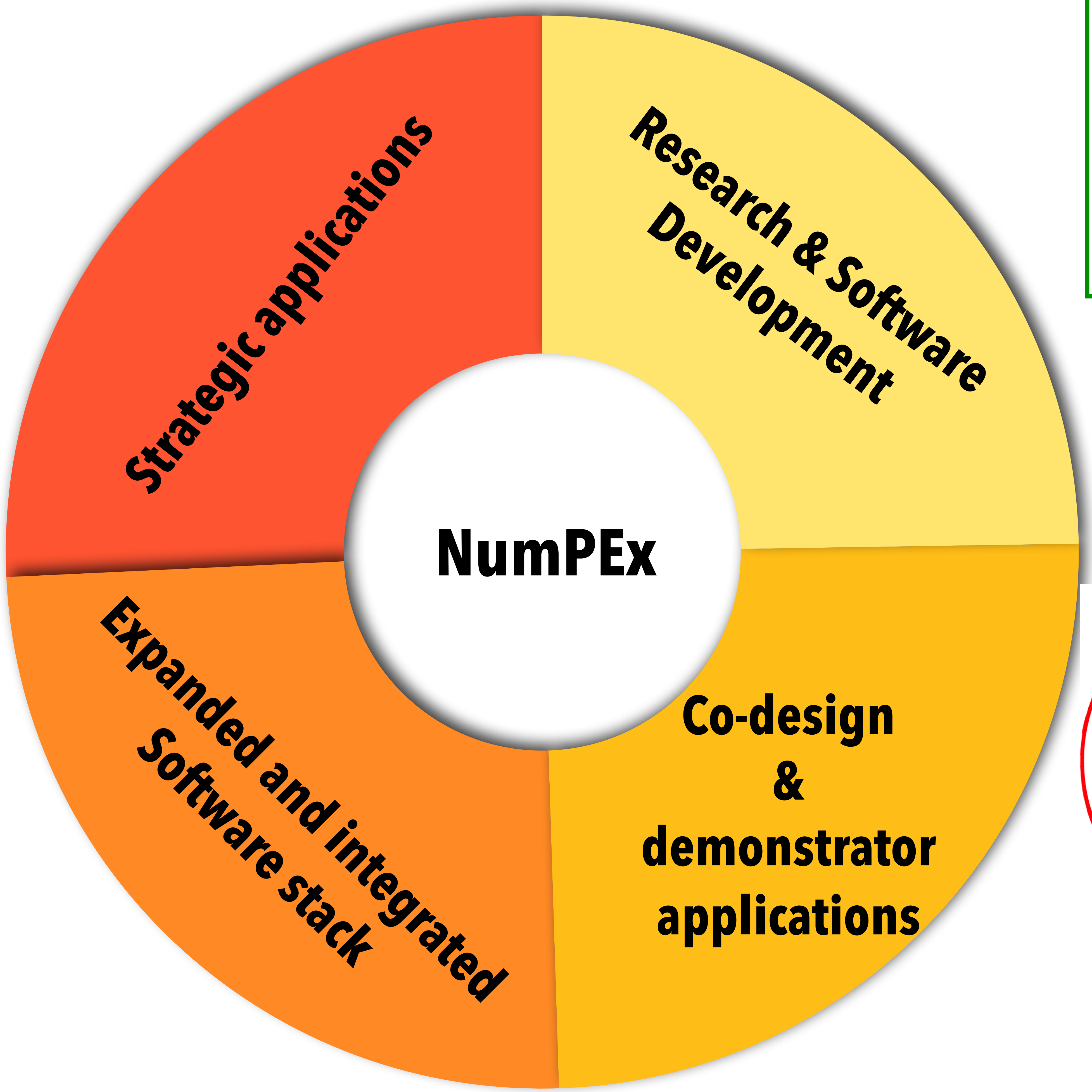


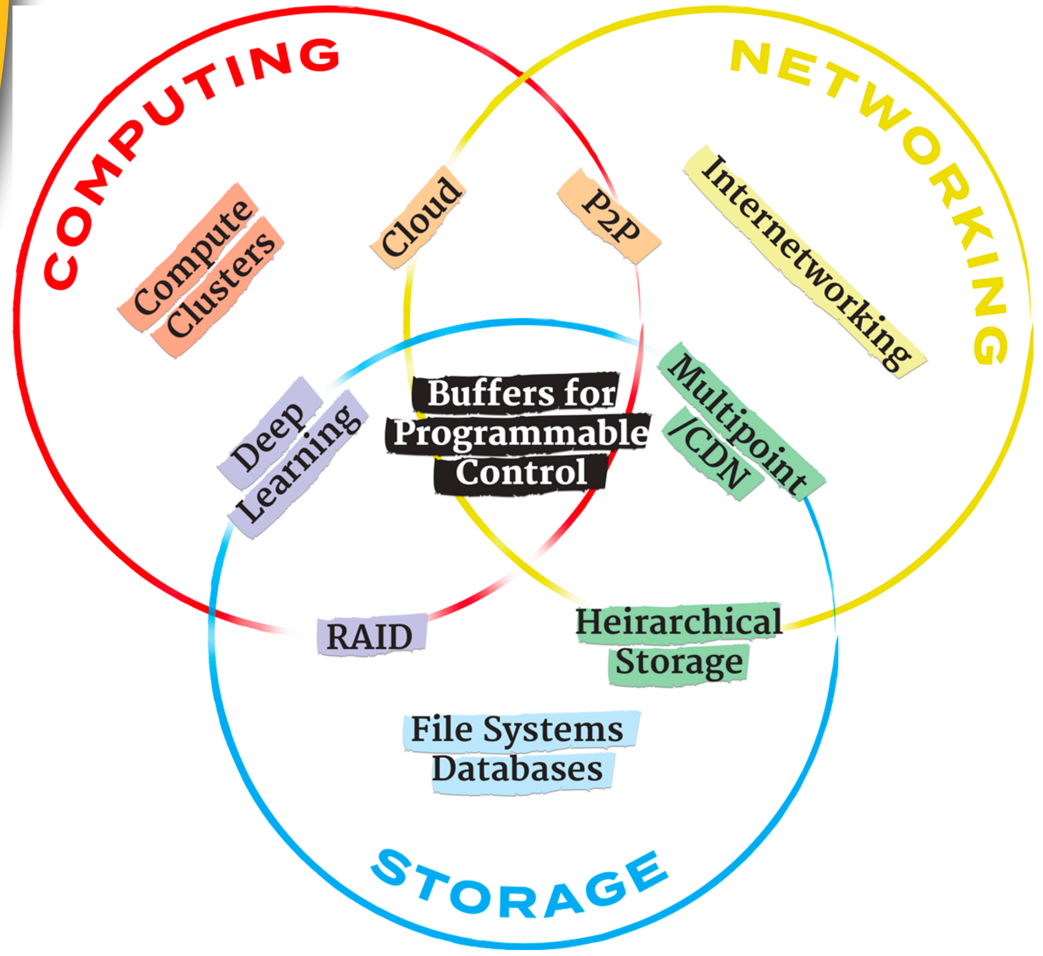
Aggregate the French HPC/HPDA/AI community, foster new collaborations and synergies

Co-develop, integrate, validate and deliver an expanded exascale software stack to accelerate strategic Exascale applications productivity and sustainability

Contribute and accelerate the emergence of a European sovereign exascale software stack and performant strategic exascale applications

Build a multidisciplinary national workforce and develop training to improve CSE application development and software integration methodologies



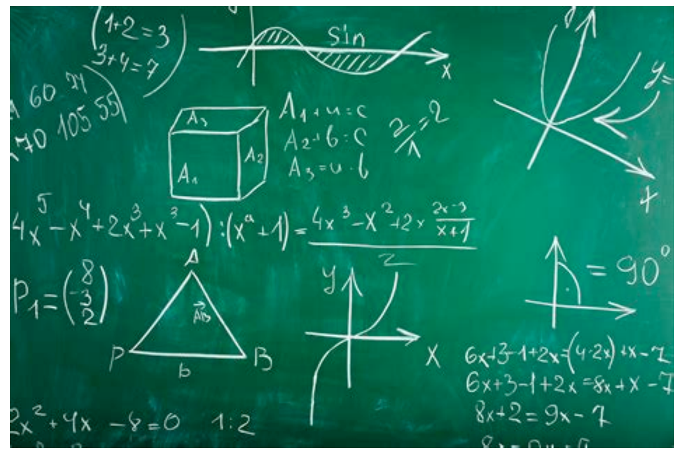


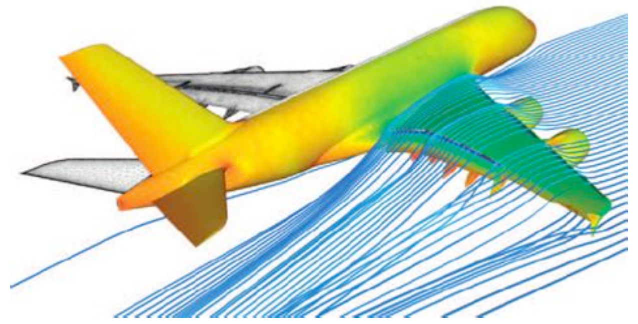
NumPEx
Exascale computing

NumPEx: Capable Exascale challenges

A co-design and co-development approach

Research





Hardware technologies

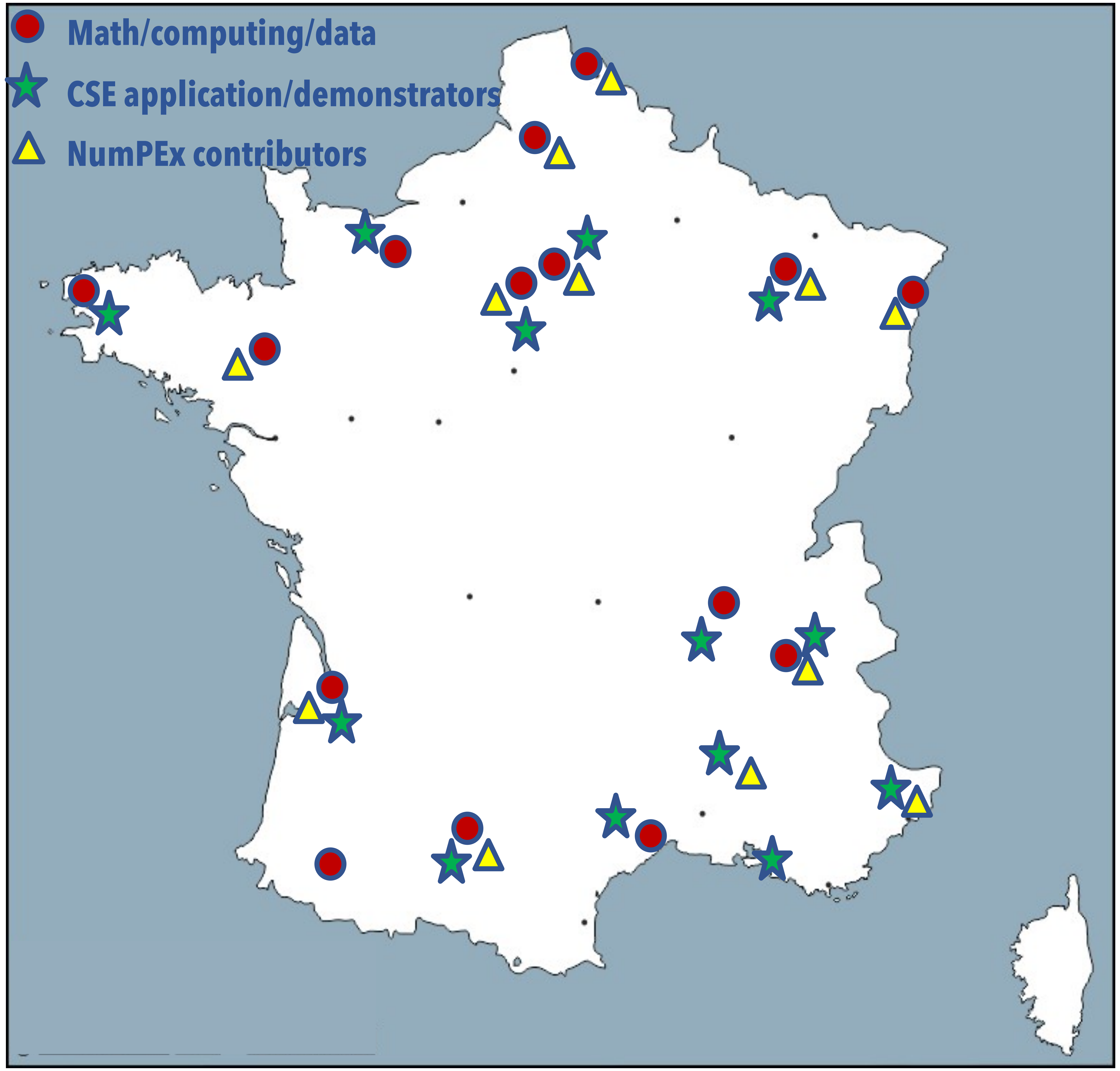
NumPEx

Applications

Software technologies

NumPEx by numbers

NumPEx
Exascale computing



5 Years
41 M€*

2023-2027
* 500 man.year non permanent staff + 150 man.year permanent staff

Core research institutions

Core national Research Institutions: CNRS, CEA, INRIA, Universities, Engineer schools, Industry

3 Focus area

Software stack development (PC 1-3)
Wide-area workflows and architecture (PC 4)
Integration and CSE application co-development (PC 5)

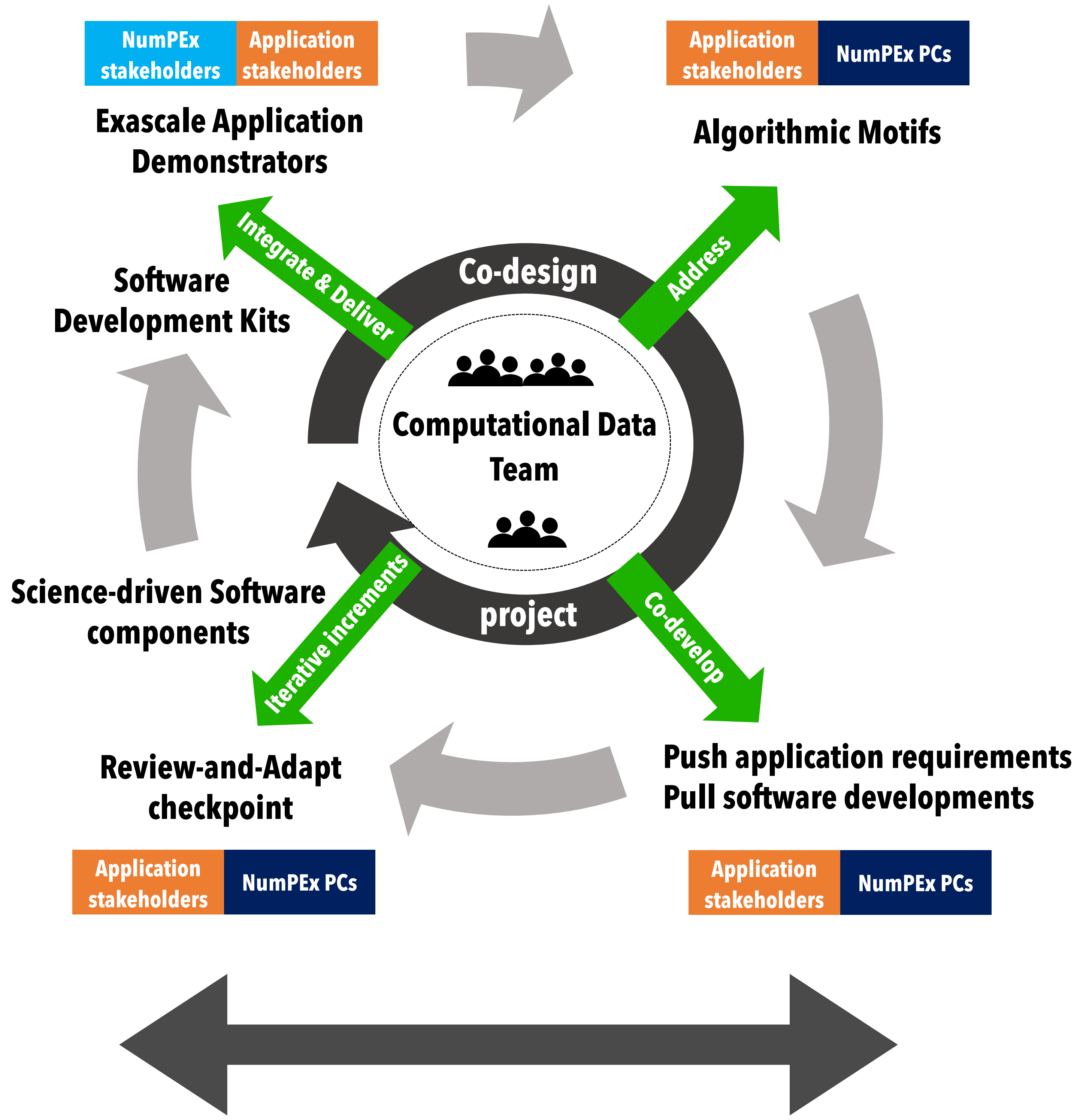
80 teams
500 researchers

Research & software Development
Multi-disciplinary computational and data team enabling co-development

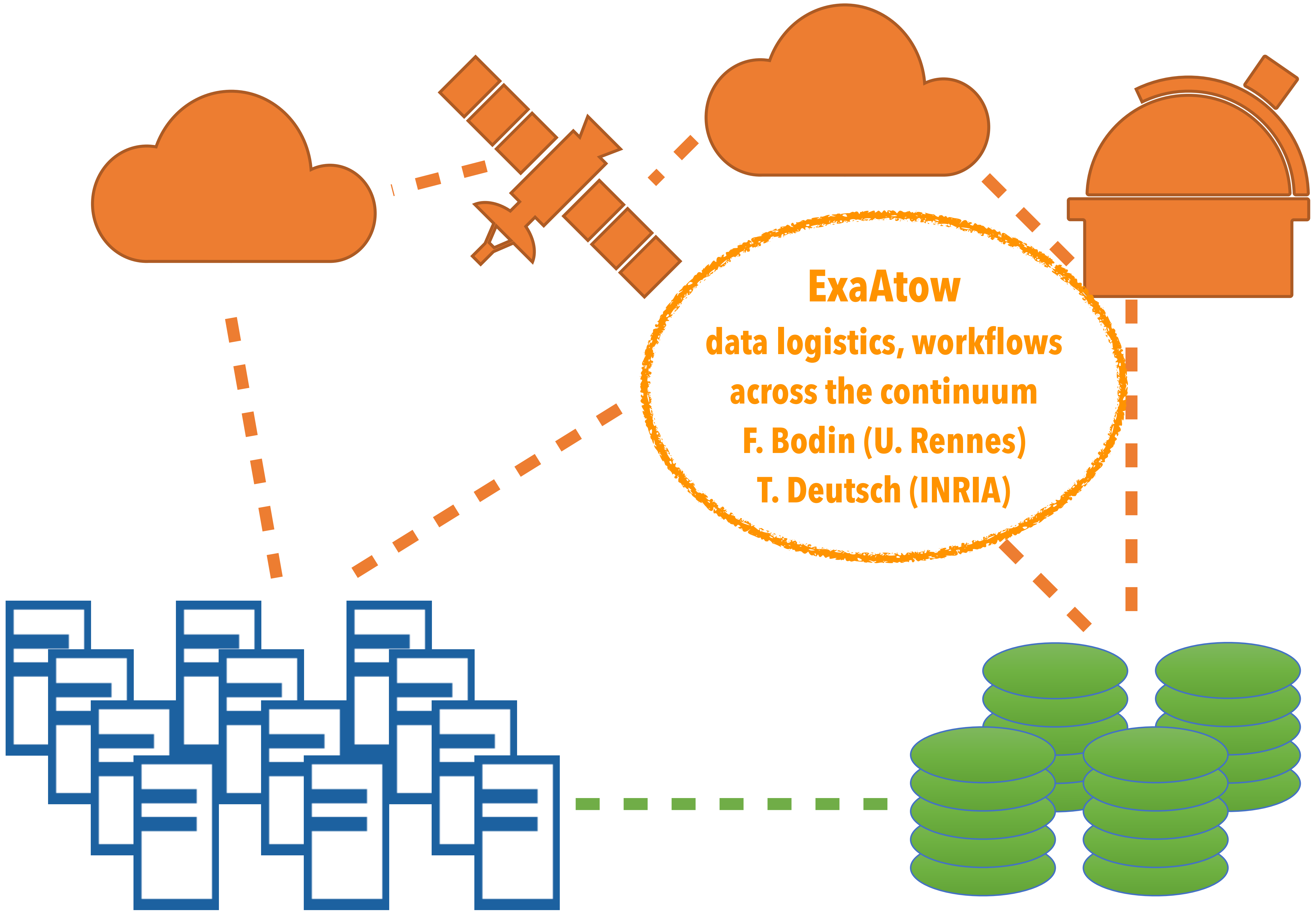
NumPEx
Exascale computing

NumPEx work plan

Applications



ExaDIP
software co-design & co-development, integration and delivery
J.-P. Vilotte (CNRS)
V. Brenner (CEA)



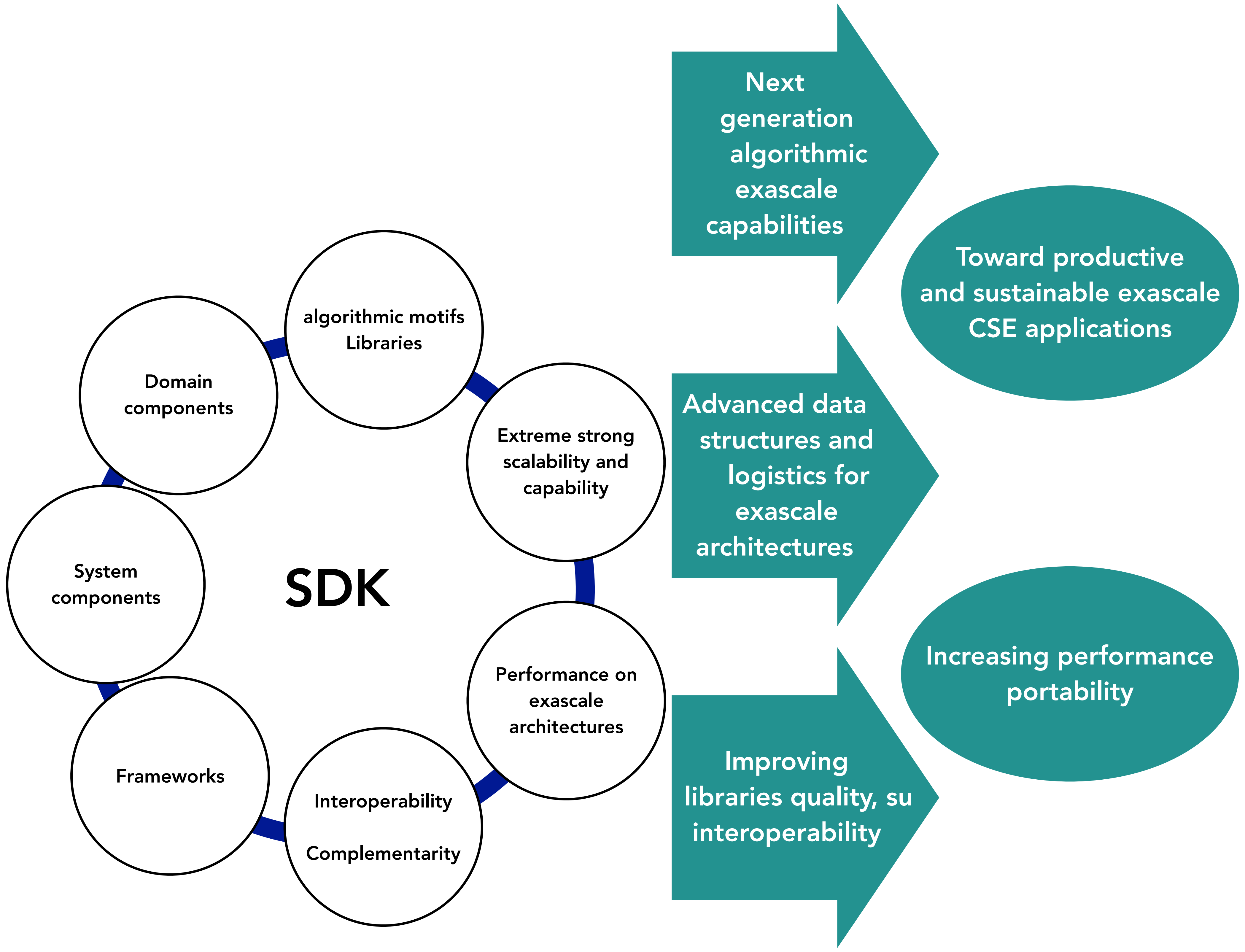
ExaMA
algorithms, math libraries
C. Prud'homme (UNISTRA)
H. Barucq (INRIA)

ExaSoft
parallel programming and execution environments
R. Namyst (INRIA)
A; Buttari (CNRS)

ExaDost
in-situ data reduction and analytics, storage, IO
G. Antoniu (INRIA)
J. Bigot (CEA)

NumPEx
Exascale computing

NumPEx: ExaDIP



Timeline: SDK release 1 · SDK release 2 · .... · SDK release n

Crosscutting algorithmic motifs

- Accelerate the exascale development and productivity of exascale ADs by developing science-driven mathematical models and software components;
- Investigate crucial performance trade-offs between software components.
- Instantiated in many ways depending of the ADs contexts, each of them having unique requirements.

Community Software Policies

- Improve software quality, usability, access and sustainability;
- Provide foundation for deeper levels of interoperability
- Establish a certification process to label software (maturity, portability, compliance)

Software Development Kits

- Logical collections of value-added interoperable software components as needed by ADs,
- Integrated and delivered using meta-builder and container systems enabling a combined deployment on exascale systems and combination as needed by CSE applications

Software packaging and deployment technologies

- Promote common Meta-builder systems (e.g. SPACK, GUIX, NIX) and container technologies (e.g. Singularity)
- Extend/harden new capabilities enabling deployment on exascale systems and regression testing

Software Integration hub

- Enable access to externally managed software integration and testing platforms
- Synergetic collaborations with national computing facilities, vendors and other initiatives

NumPEx
Exascale computing

SKA Demonstrator

SKA FRANCE
SQUARE KILOMETRE ARRAY

Extracting new insights on fundamental questions from massive Sky observations flux over a wide range of radio frequencies

Observations characterised by:
- high sensitivity
- high resolution: time, frequency, position/size
- multi-frequency coverage

Different observation modes:
- imaging, non-imaging

Advanced high-end data analytics
- radio data reduction and calibration
- science data analysis

SKA1-LOW (Australia)
- 130,000 long-periodic antennas
- Maximum baseline: ~65 km
- Frequency range: 50 MHz - 350 MHz
- ~2 Pb/s

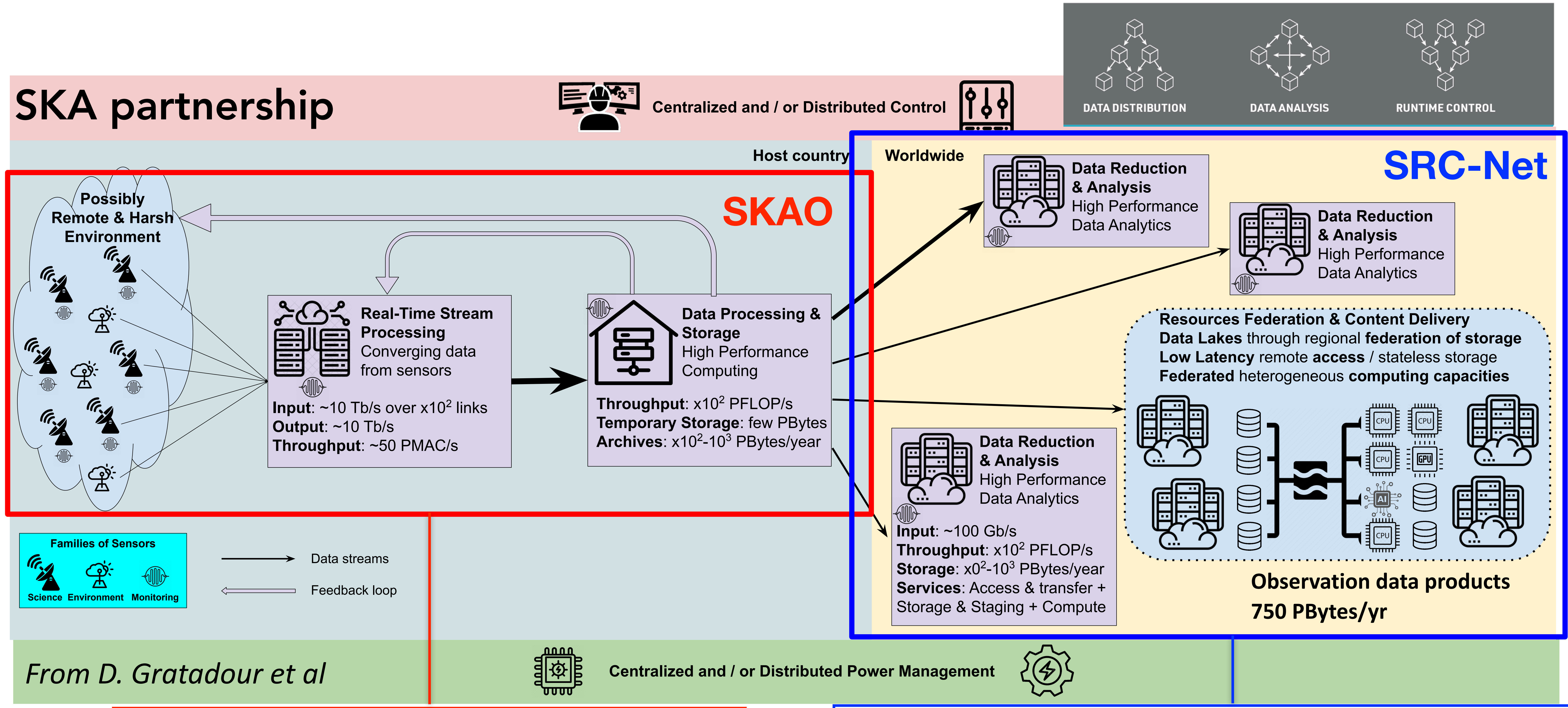
SKA1-MID (South Africa)
- 197 dishes
- Maximum baseline: 150 km
- Frequency range: 350 MHz - 15 GHz
- ~8.8 Tb/s

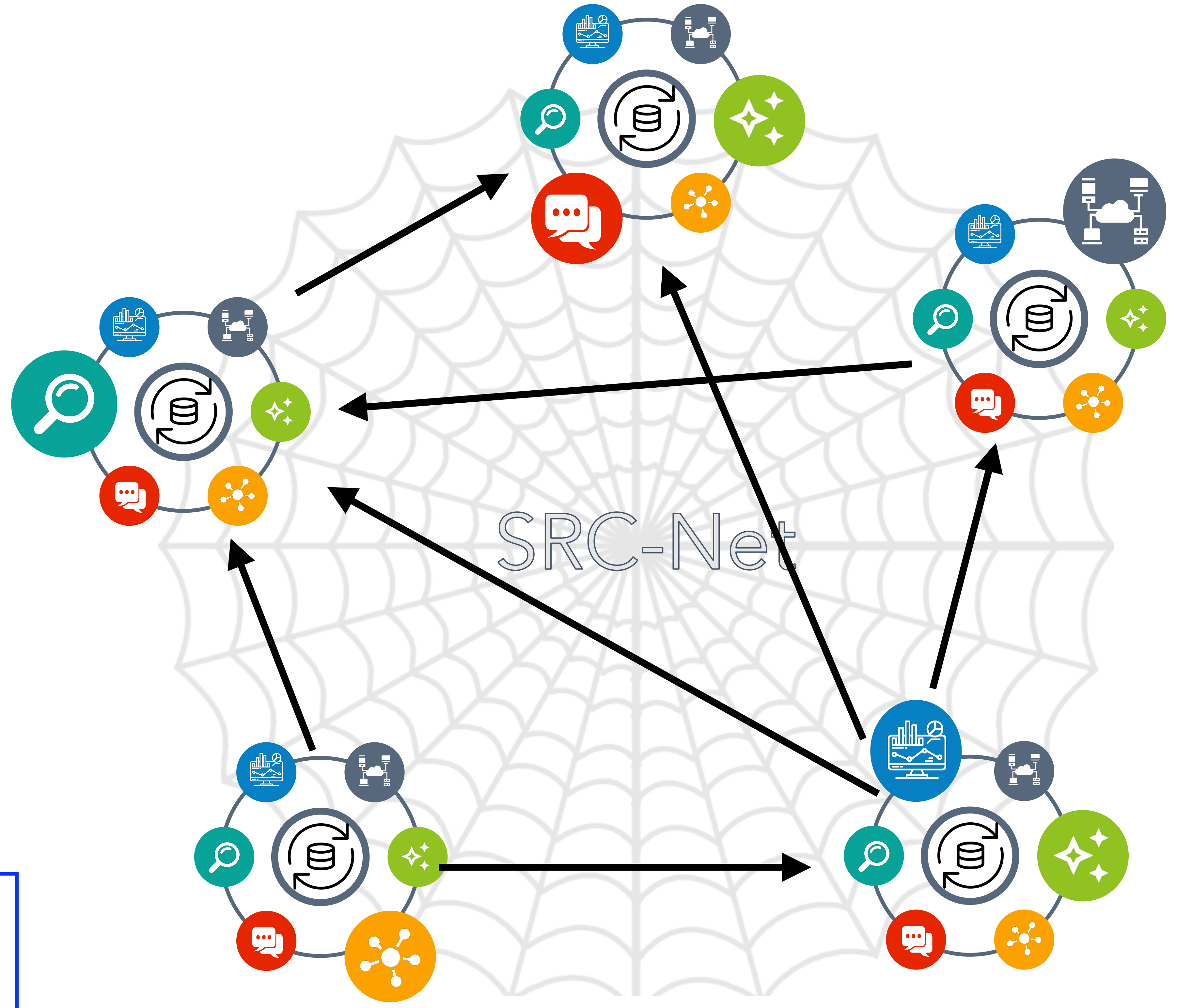
50 MHz — 350 MHz — 15 GHz

NumPEx
Exascale computing

SKA use case

SKA FRANCE
SQUARE KILOMETRE ARRAY





Observation data product delivery

- High-rate streaming data reduction and processing,
- Edge to centralised HPC (SPC) computing/AI and content delivery networks

Distributed Data archive and Data analysis

- Federated Exascale and hybrid Cloud platforms: storage, networking, compute, wide-area workflows, data logistics
- Logistical networking, cybersecurity and power management

The SRC-Net: a critical component of SKA

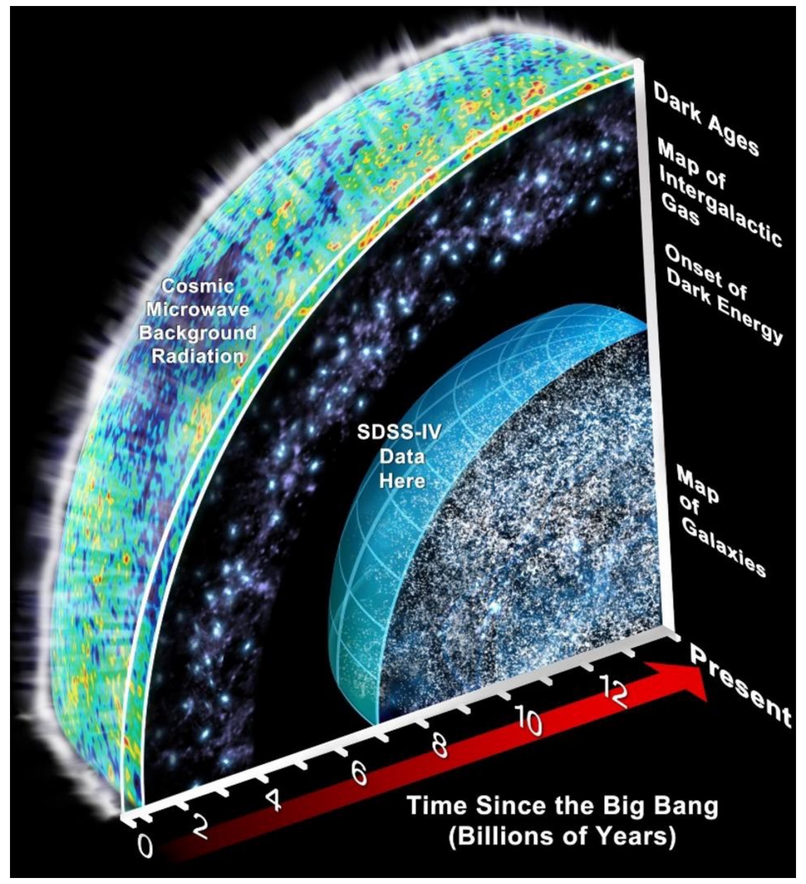
Federation of distributed ressources including exascale systems (storage, networking, computing) to fully process, archive, curate and scientifically use SKA observation data products

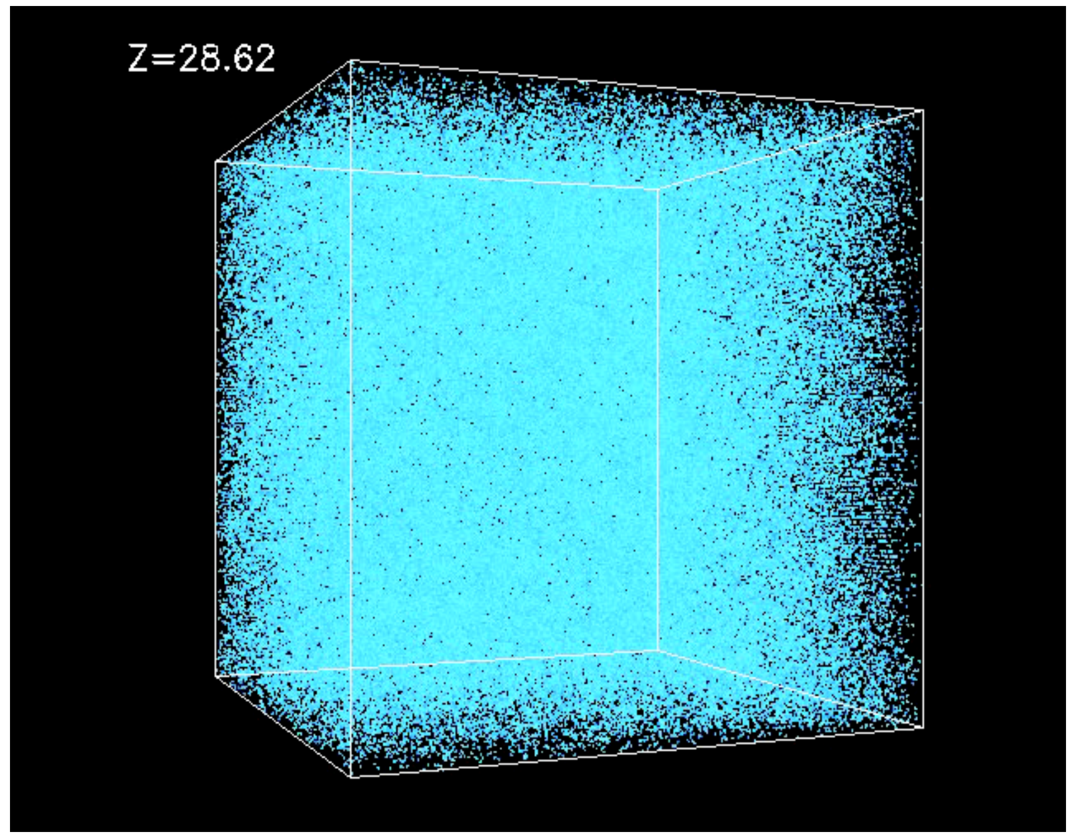
Science-driven capabilities: bridge organisational, and technological boundaries; foster major collaborative and interdisciplinary efforts across algorithmic research, software development and integration, data logistics, continuum of infrastructures and SKA science communities communities

No access to the SDPs nor to the raw SKAO data

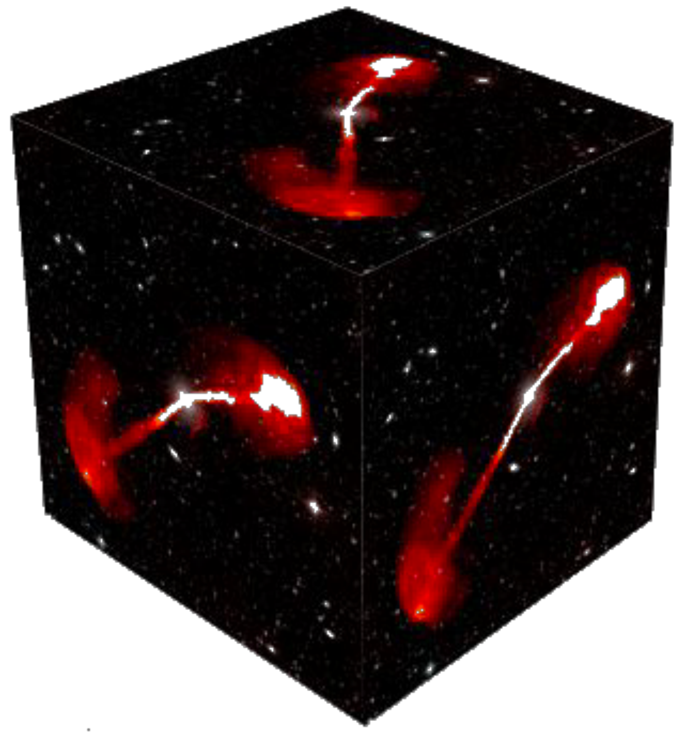
SKA Exascale science challenges

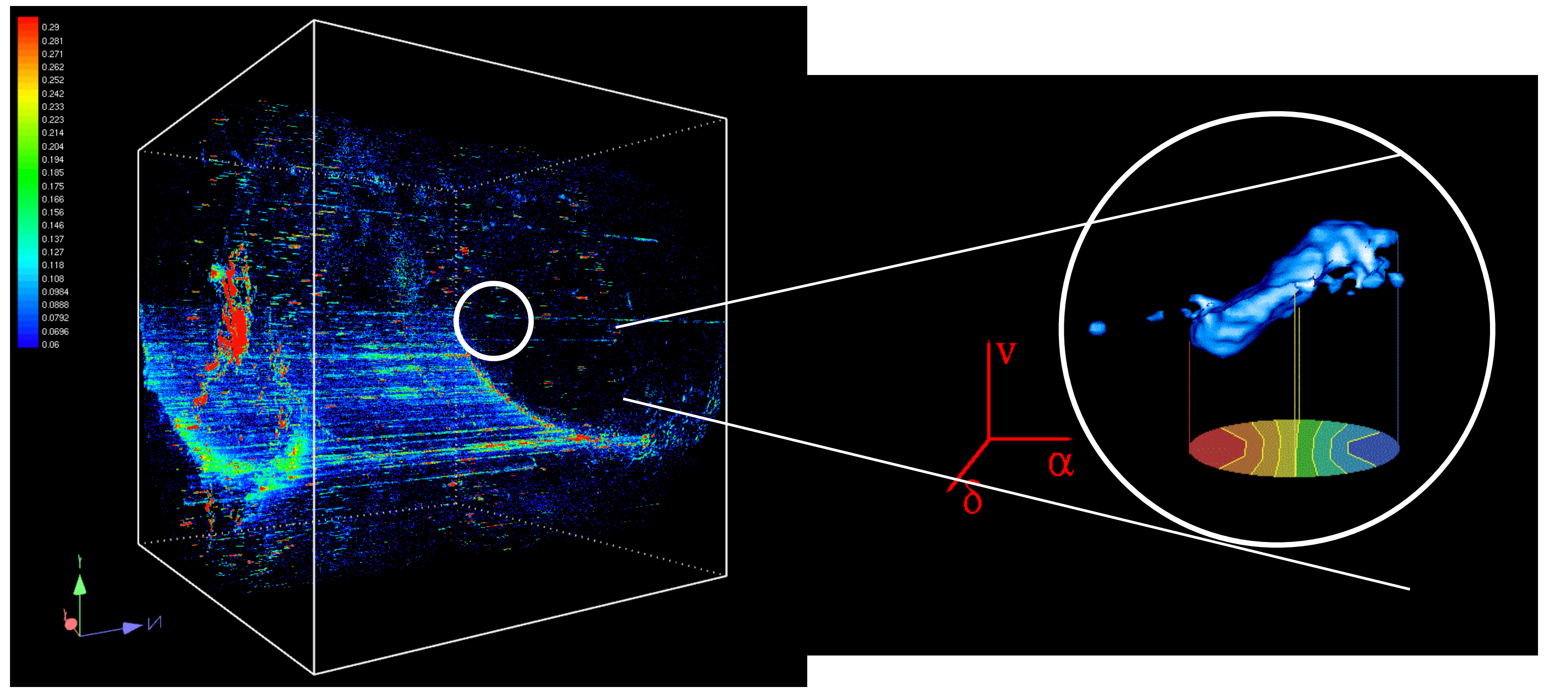
SKA FRANCE
SQUARE KILOMETRE ARRAY



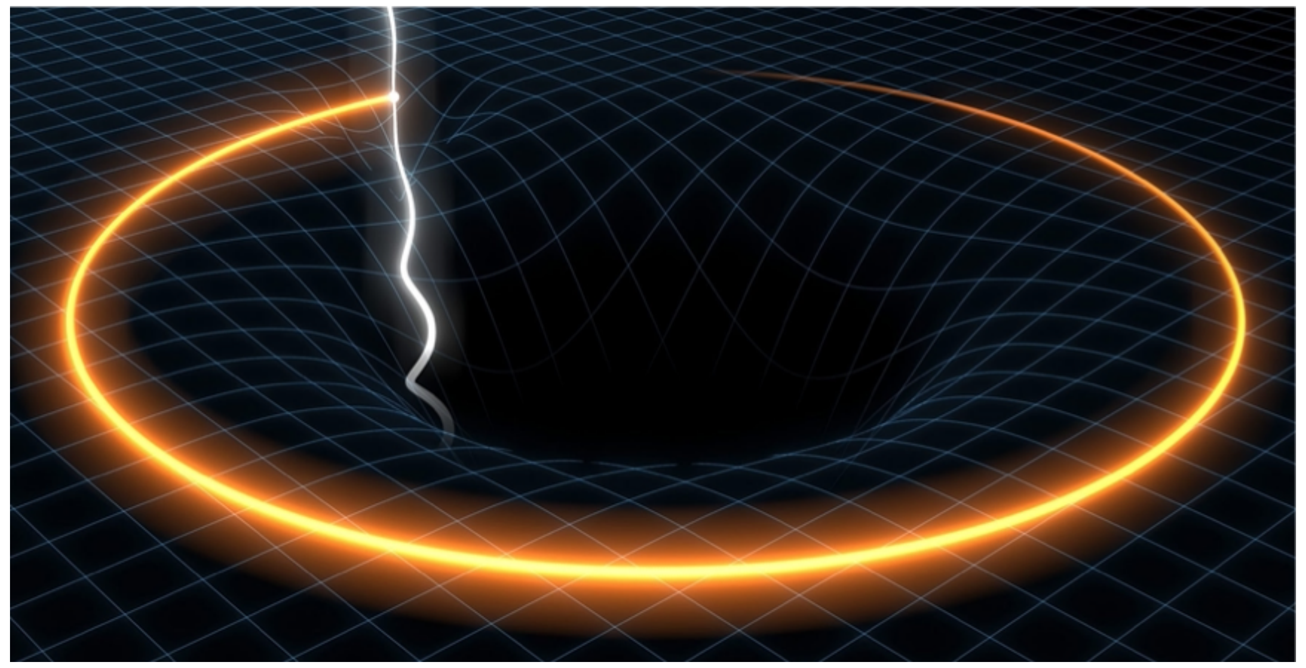


Detection, classification, simulation, inference in HI surveys

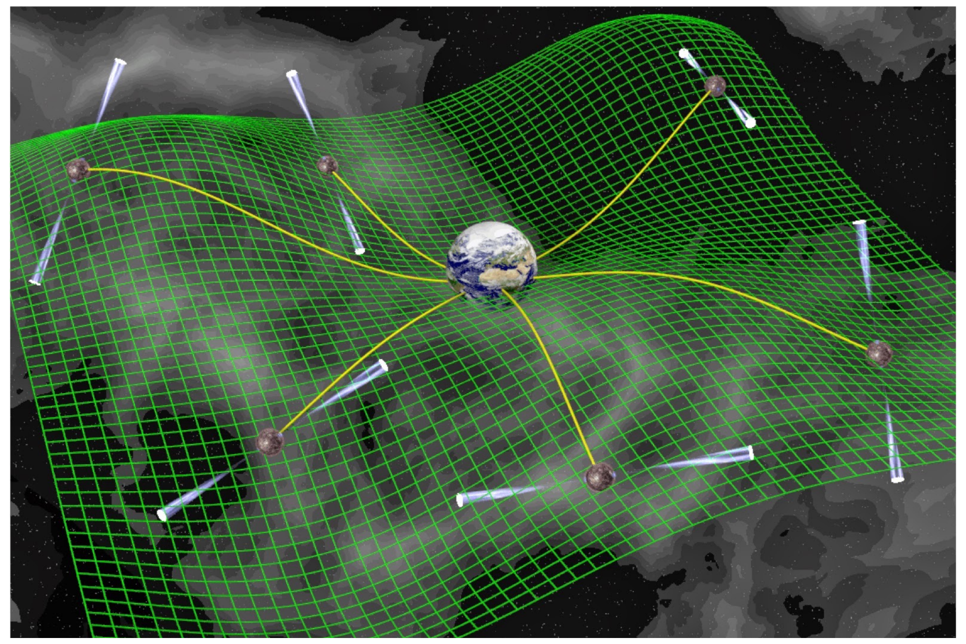




Pulsar/black-hole: general relativity test



Milliseconds Pulsars: pulsar timing array gravitational wave detection



SKA data era

- Sky surveys: ~10^{12} - 10^{18} bytes images
- Archived science-ready data: ~700 PB/yr
- SKA data cube: ~0.9 PB
- Catalog: ~10^8-10^9 objects (stars, g
- alaxies, etc.)

Cosmic dawn
(First stars & Galaxies)

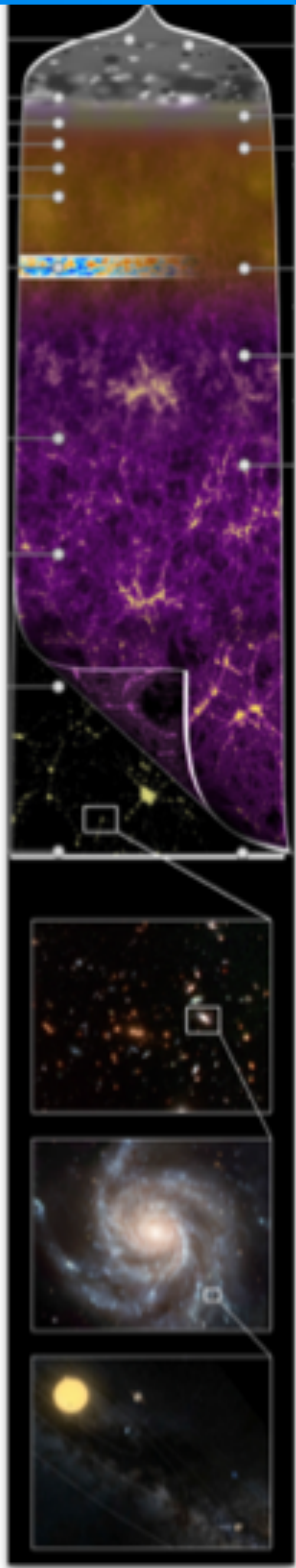
Cosmology
(Dark matter, Large-scale structures)

Galaxy evolution
(gas content & new stars)

Cosmic magnetism
(origin & evolution)

Fundamental physics
(gravitational waves & compact objects)

Cradle of life
(Planets, Molecules, SETI)



SKA observation mode: Non-Imaging

Faster

Pulsar/transient search & Pulsar Timing Array

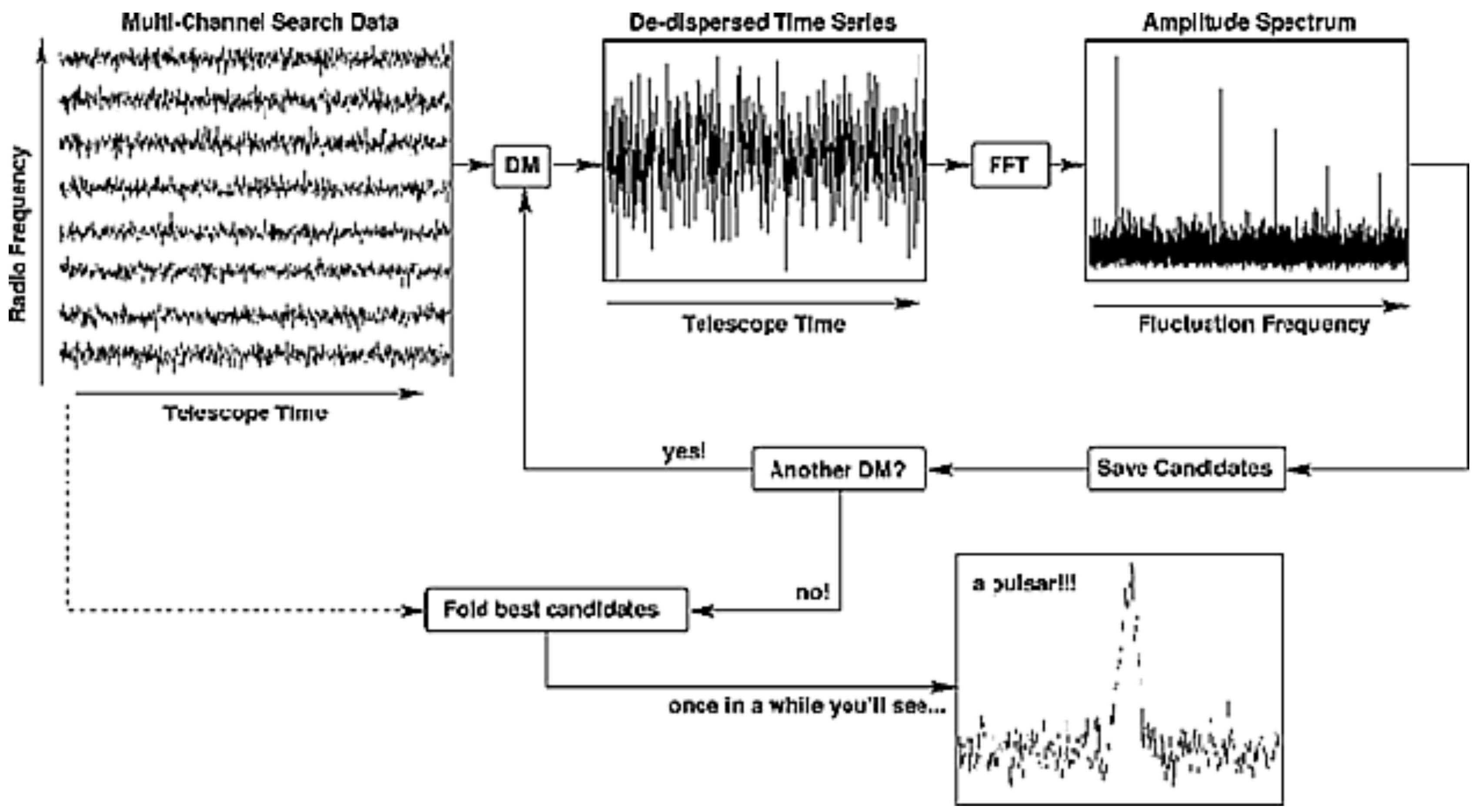
Tied-array beams: forms hundreds of beams within the dish/station beam

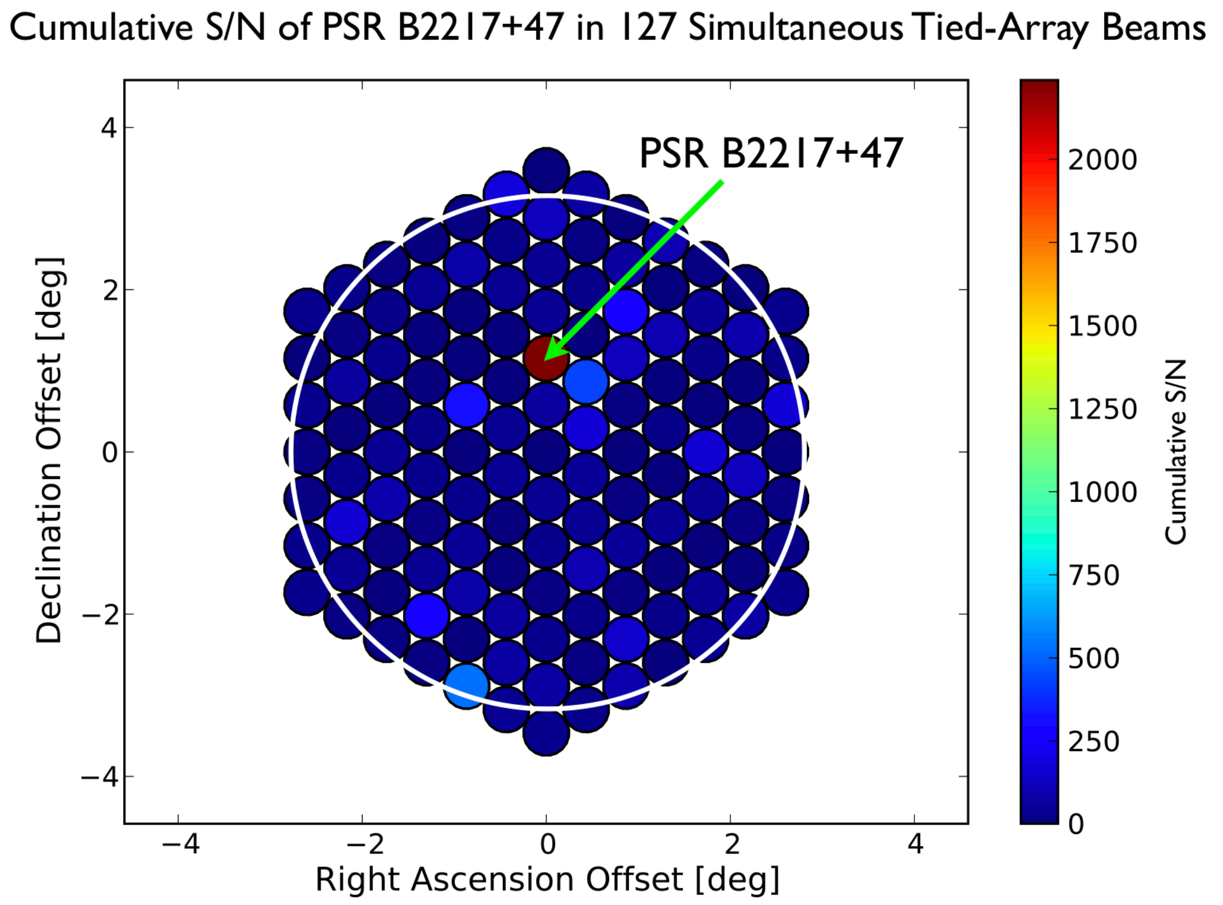
Time resolution ~60 - 100 μs, Data rate 800 GB/s

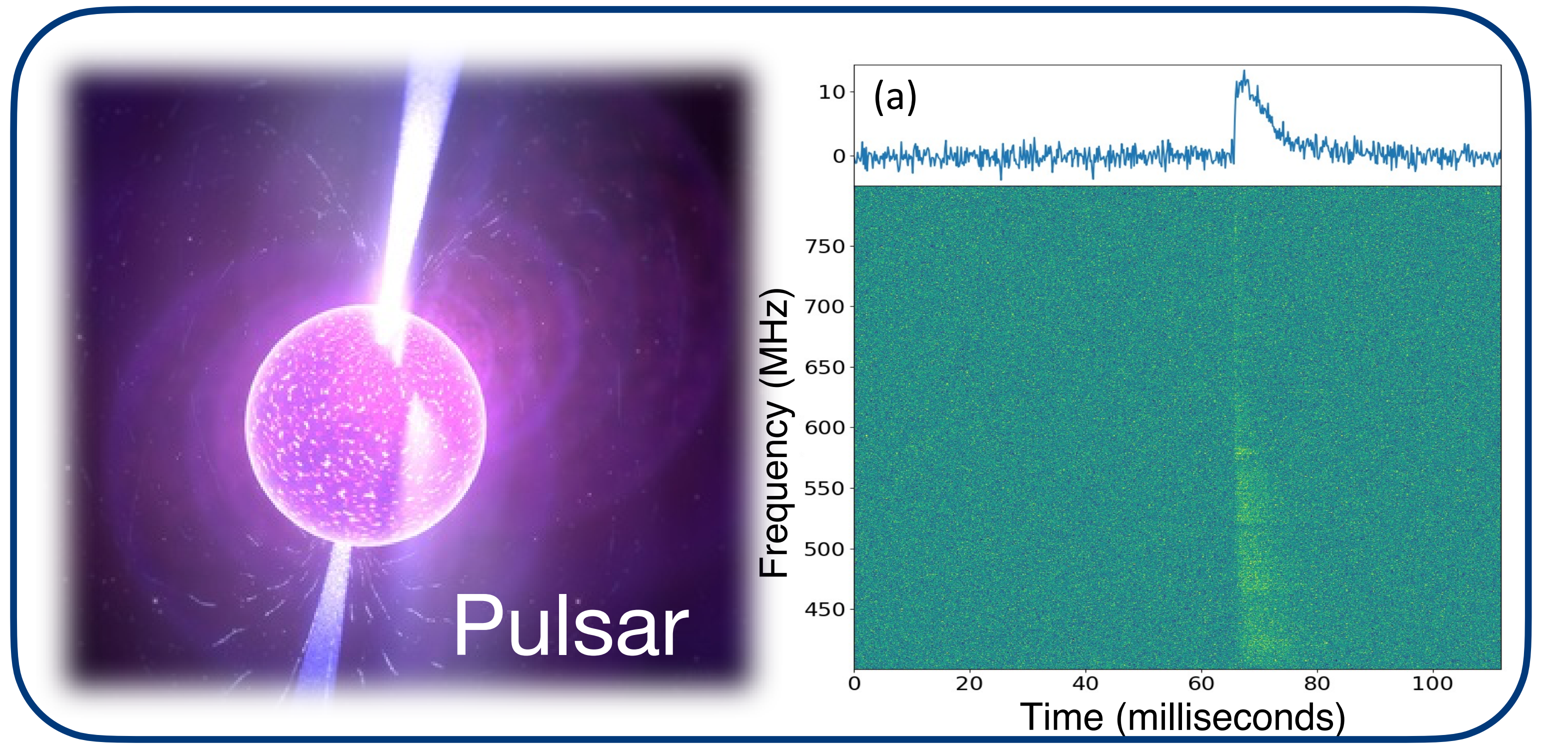
High-rate (~8 Tb/s) streaming data processing

Array multi-wavelength detection,

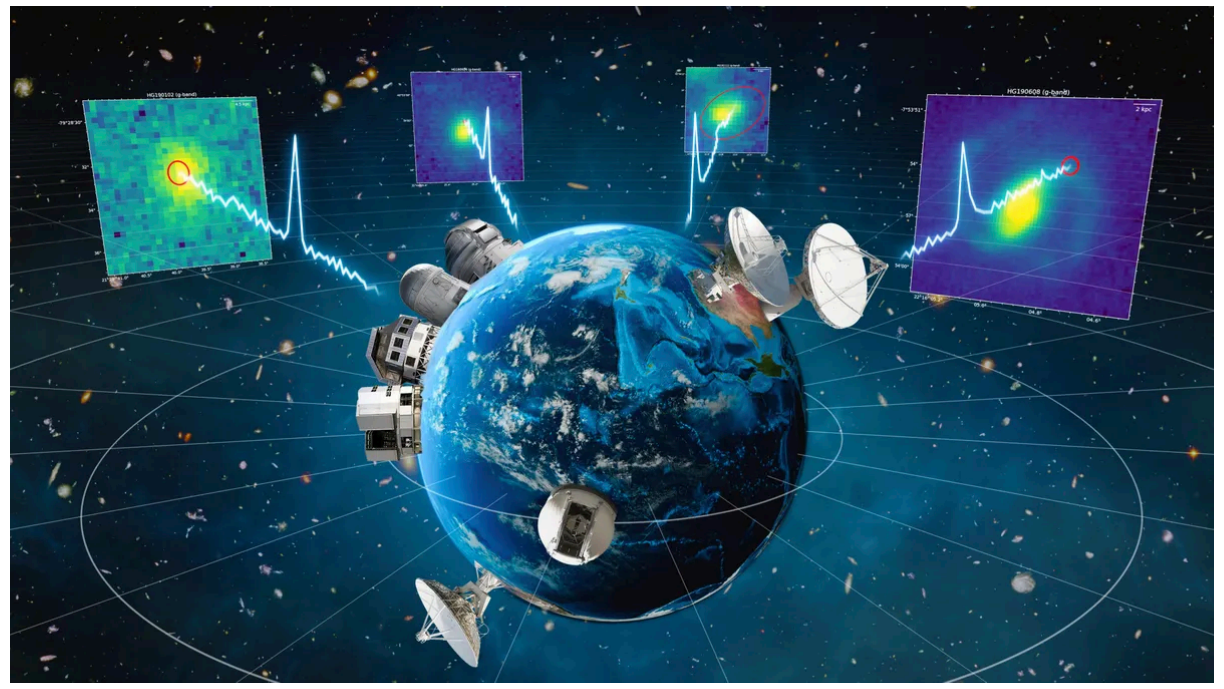
Distributed ML, correlations, transient noise, simulation-based inference

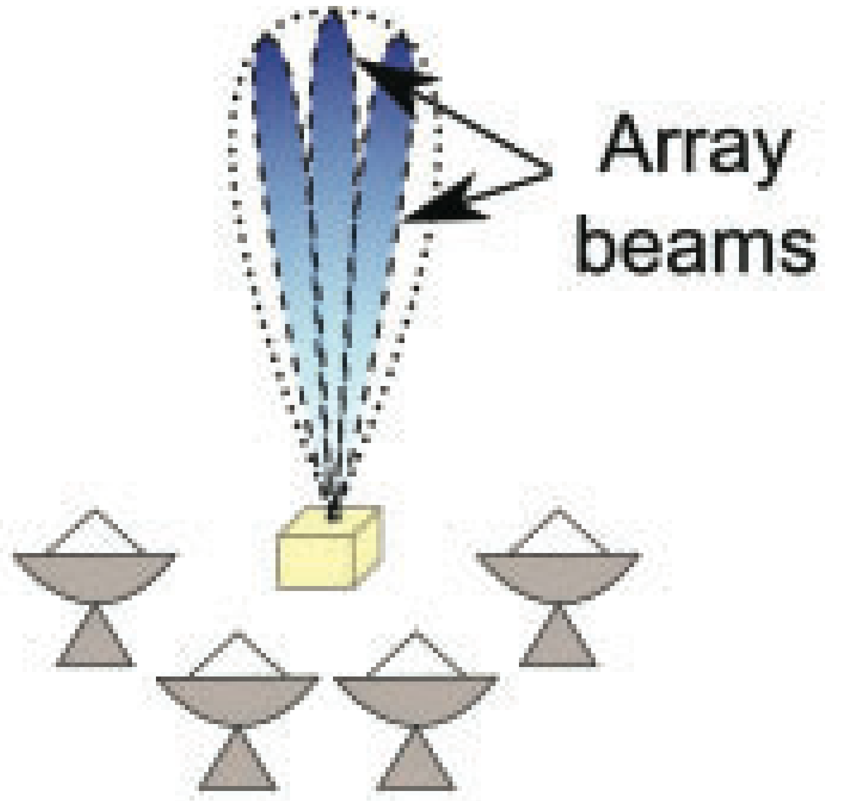


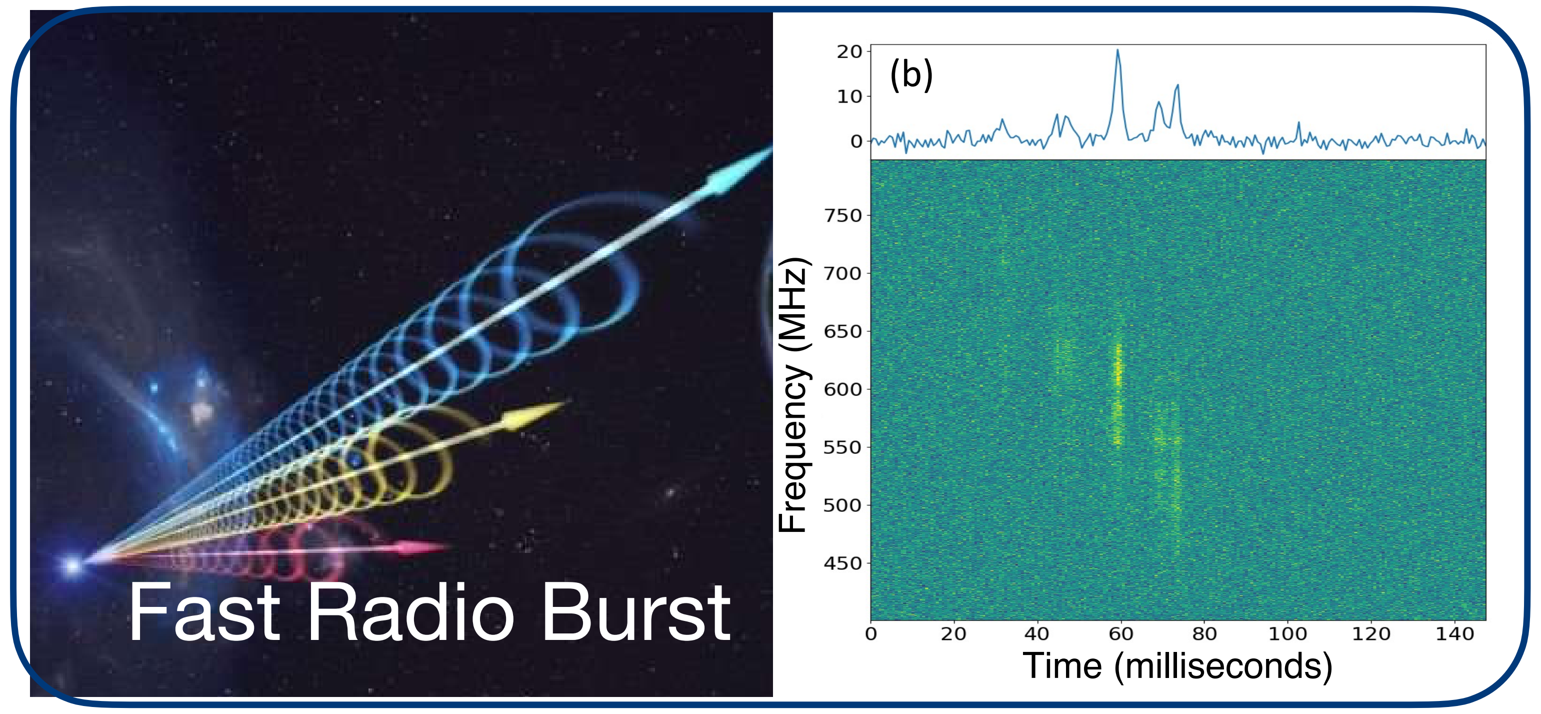




From Cherry Ng







Algorithmic motif:

ExaMa: Machine Learning to reduce number of false positive, search a large range of signal morphologies, scalable graph neural networks, simulation-based inference, high-fidelity surrogates

ExaSoft: GPU-parallelisation for beam forming, programmable accelerators (e.g., FPGA) high-rate streaming data processing, mixed low and high arithmetics, performance portability

ExaDow: Exascale ML-based transient detection, high-rate streaming data in-situ processing and reduction

ExaAtow: Distributed data archiving and fast data access, multi-telescope coordination, federated data processing across the SRCNet

SKA observing modes: Imaging

Fainter/Further

Most of the other Key Science Topics

(e.g., interstellar medium, weak lensing, large scale structure, CMB, 21cm signal from Cosmic Dawn)

4D images (RA, Dec, Frequency, Polarisation)

Image size: ~1PB; Archival: ~120 PB

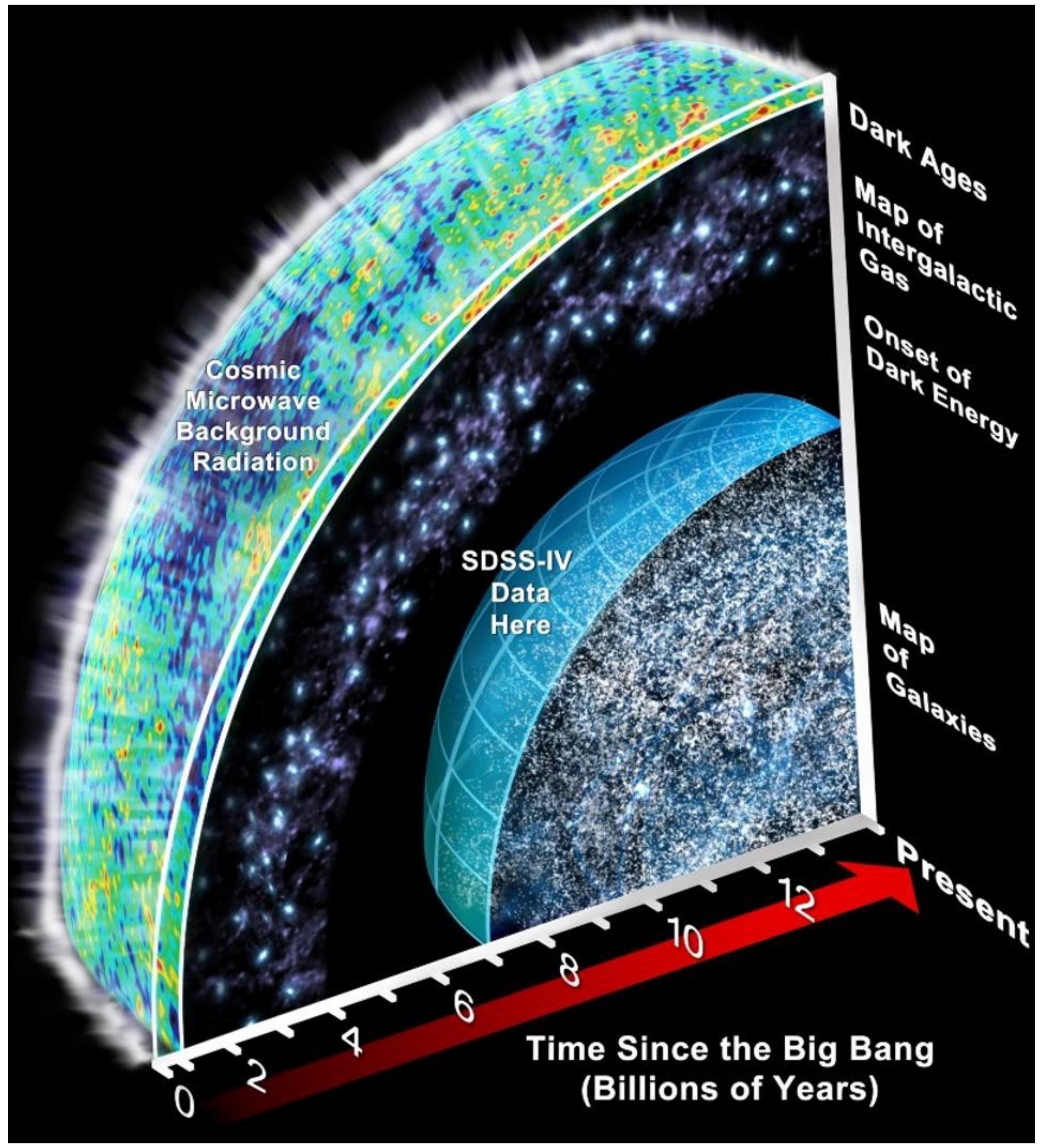
Large number of objects in each deep spectral line (~100 TBs - 1PB)

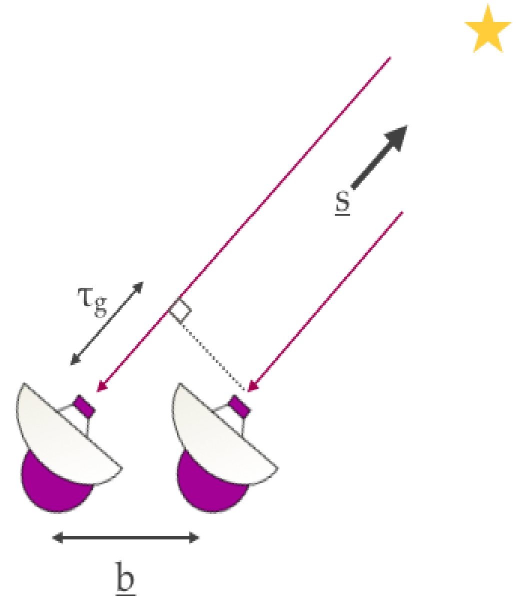
Clustering, ML source identification & classification

Simulation-based Bayesian inference large ensemble of simulations,

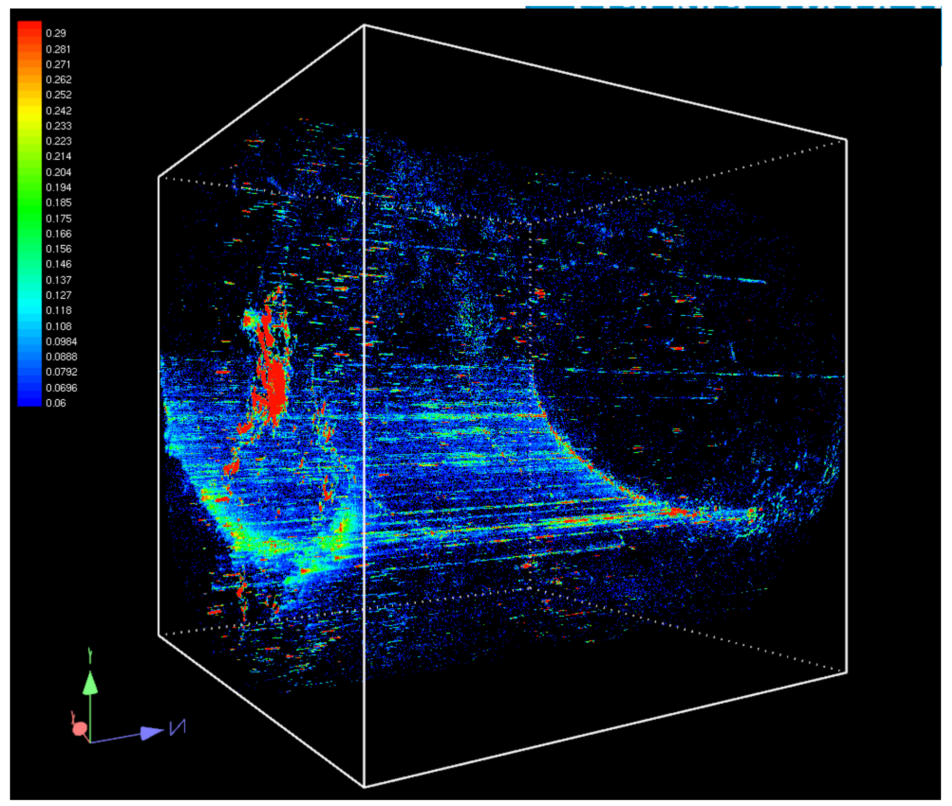
Physics-based ML, high-precision surrogate models

Non-Gaussian statistical estimators, generative models, components separation





$$T(x,y) = \int\int V(u,v) e^{-2\pi i(ux+vy)} du dv$$



Algorithmic motif:

ExaMa: Physics-based ML/DML, graph neural networks, statistical estimators, component separation & high-order model reduction, large range of source signal morphologies, simulation-based bayesian inference, high-fidelity surrogate models, multi-physics & multi-scale MHD/Particule simulations

ExaSoft: GPU-parallelisation, heterogeneous hardware, performance portability

ExaDow: Exascale ML-based source detection classification, distributed ML/DML, and data layout, in-situ data processing and reduction

ExaAtow: Distributed data archiving and fast data access, federated data processing and ML

NumPEx
Exascale computing

SKA observing modes: calibrated interferometric visibilities

Faster/Fuller

Epoch of Reionization

Data archive of: > 200 PB; Per observation: > 250 GB

Large number of objects in each deep spectral line (~100 TBs - 1PB)

Primary beam, ionospheric distortion, polarisation leakage

Complexity/processing time (number of baselines, directions, objects in the sky model, channels)

Post-processing ranges from needing to assess source catalogues to working with Fourier domain data

Data set can be split across radio frequency and time slots, but up to 30 PBs each (data object monster)

ML/DML, clusterisation, simulation-based inference, constrained optimisation (spectral/time/space)

Cost function:

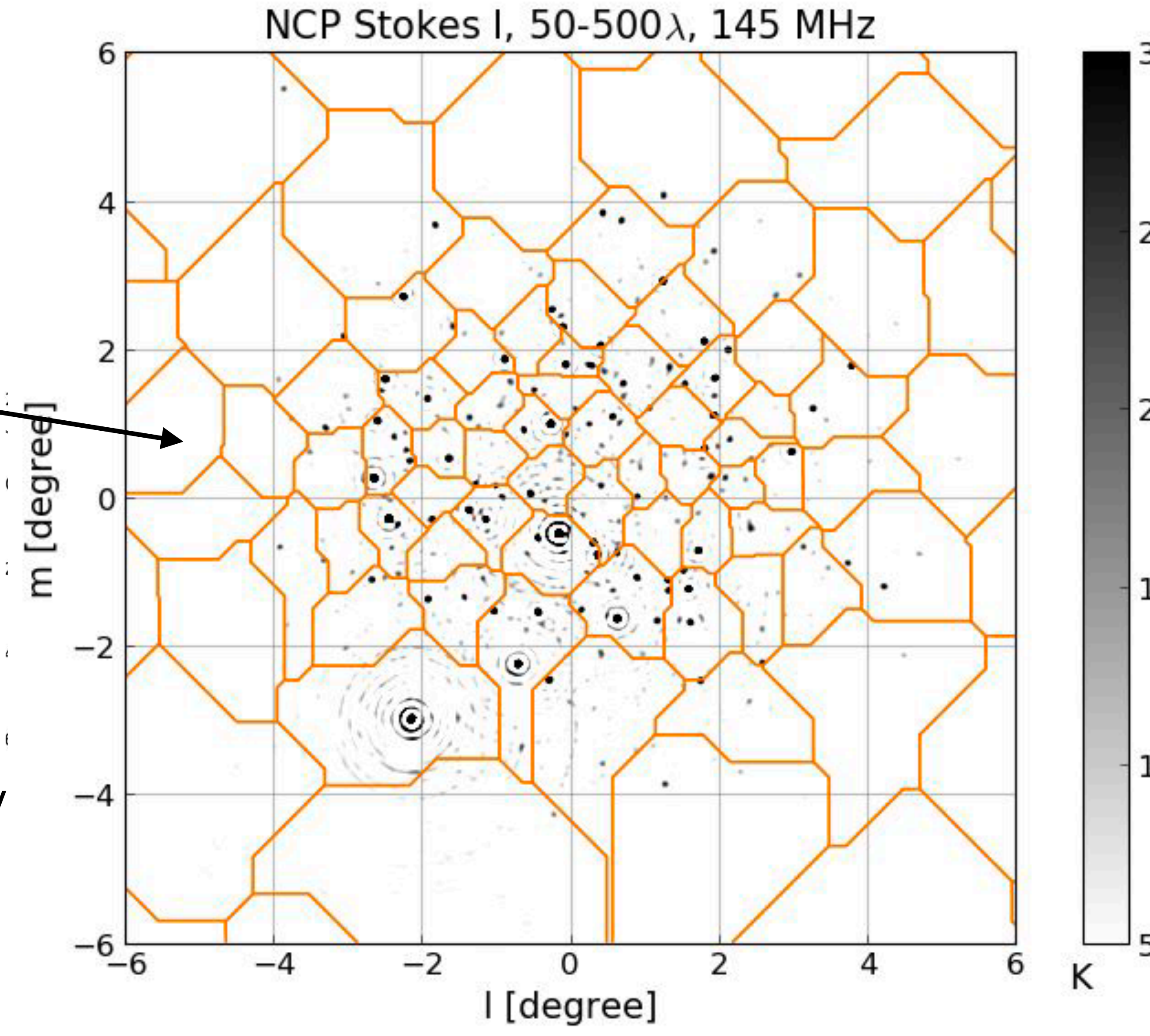
$$g(\boldsymbol{\theta}) = \sum_{v,t,i,j} \left\| \mathbf{V}_{ijvt} - \sum_{k=1}^{K} \mathbf{J}_{ik}(\boldsymbol{\theta})\, \mathbf{C}_{ijkvt}\, \mathbf{J}^{\mathrm{H}}_{jk}(\boldsymbol{\theta}) \right\|^2$$

Nbs of clusters

Measured visibility

Gains (Jones matrix)

Sky Model visibility



F. Mertens, C. Tasse, J. Girard

Algorithmic motif:

ExaMa: Distributed ML/DML, statistical estimators, simulation-based inference, constrained optimisation, high-fidelity surrogates

ExaSoft: GPU-parallelisation, heterogeneous hardware and performance portability

ExaDow: distributed ML/DML, clusterisation, in-situ data processing and reduction

ExaAtow: Distributed data archives and fast data access, distributed ML

NumPEx
Exascale computing

SKA: Exascale systems in the edge to Hybrid Cloud continuum

- **Federation of heterogeneous Resources (FoR):** what *governance and policies (non uniform)? to enable what services? to integrate in what sense distributed centralised Exascale, Hybrid Cloud, edge computing systems with communication? (ExaAtow)*
- **Distributed SKA data archives, management and access:** different *observation and science advanced data products layout, data-dependent access and services, visualisation and software libraries, tools from data lake to domain-oriented data mesh (ExaDost)*
- **Wide-area complex workflows** (HPC, HDA, AI, Visualisation): *diverse patterns depending on the science use case of when, where and how data are accessed, transformed, processed/analysed and intermediate results managed.*
- ***Control and flexibility of data and compute placement in time***: ability to run complex with changing computation "width", blur the lines between traditional silos (storage, networking, processing), more complete internet-based software stack including necessary networking, storage, processing as services in a meta-data driven approach.
- ***Resource flexibility:*** transparent dynamic stateful management of heterogeneities and resource complexities (aggregation/ disaggregation)and with uniform API, per-services security, FoR to best fit application needs and productivity
- ***Programming support:*** an abstraction layer isolating developers from underlying hardware, heterogeneity and complexity, graph of tasks, network on chips, system-level protocol standardise the interface on how to stream data
- ***Resource discovery:*** serving as light-weight global scheduler when mapping workflow onto the graph of available resources and topology
- ***Workflow portability and composability:*** stateful logistics in terms of operations (in-sity, in-locus, in-transit) on buffers with varying size and duration, the allocation and control of which is determined dynamically at run time.
- ***End-to-end simulation-based inference:*** AI/ML-driven multi-physics and multi-scale coupling, exploiting high and low arithmetics, AI/ML-driven Bayesian inference (in-situ data processing/reduction, high-order model reduction, large model space exploration), high-fidelity surrogate models
- **Cybersecurity and sustainability at the system level:** *hyper-vision of the set of resources and topology used in different workflow deployments.*