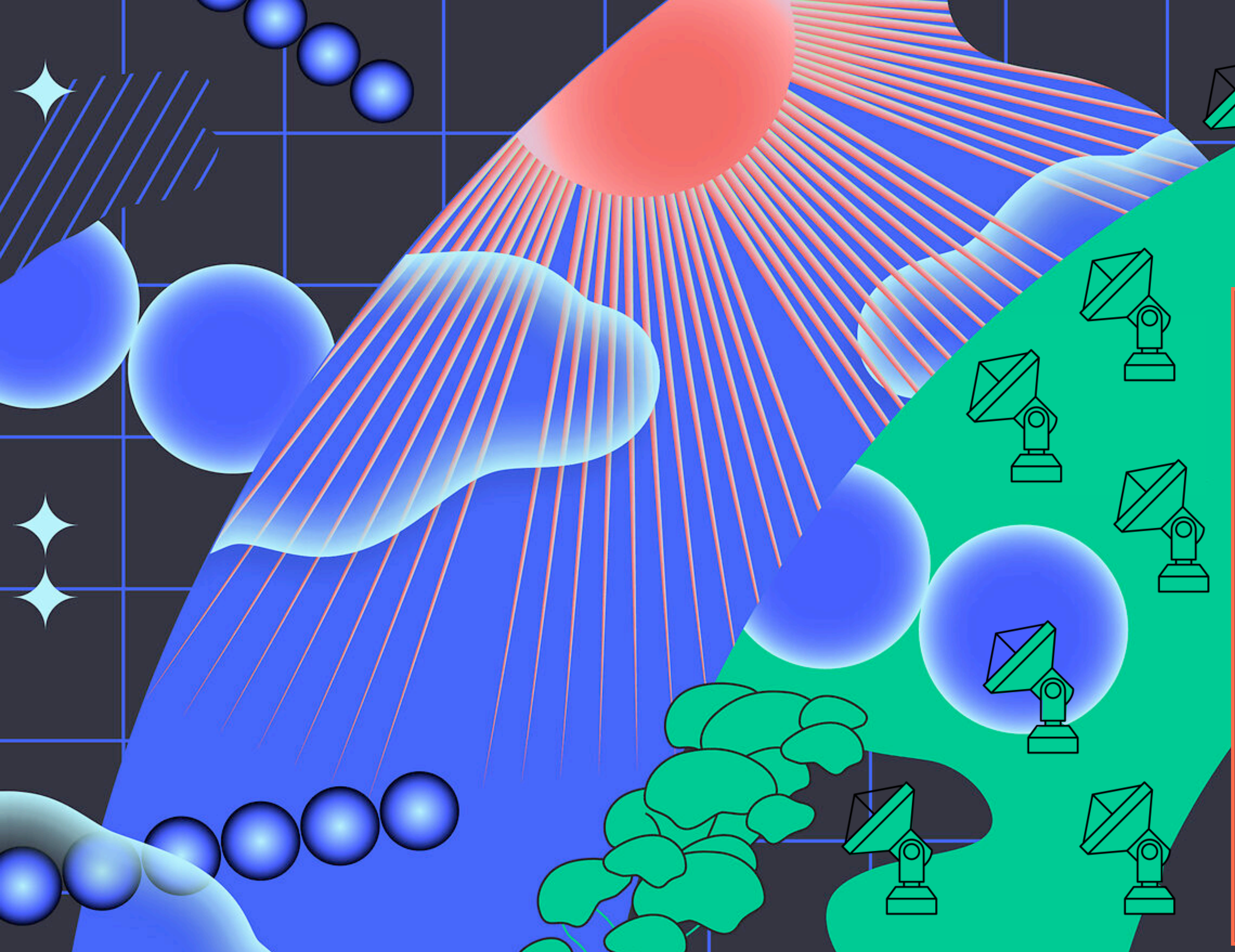


Illustration by Sandbox Studio, Chicago with **Ana Kova**

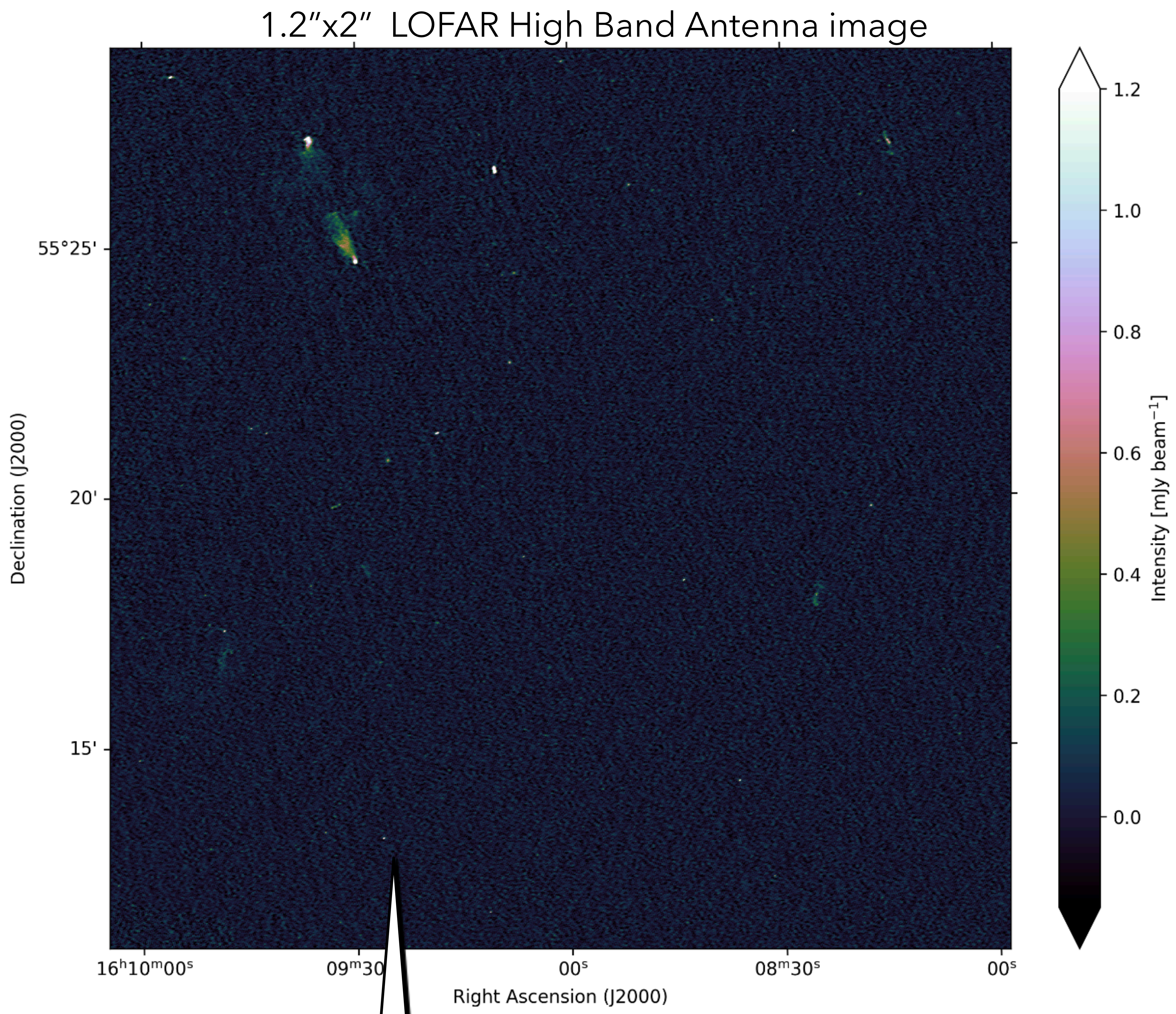
HPC & RADIO ASTRONOMY

Emma Tolley EPFL

SKACH Science Meeting

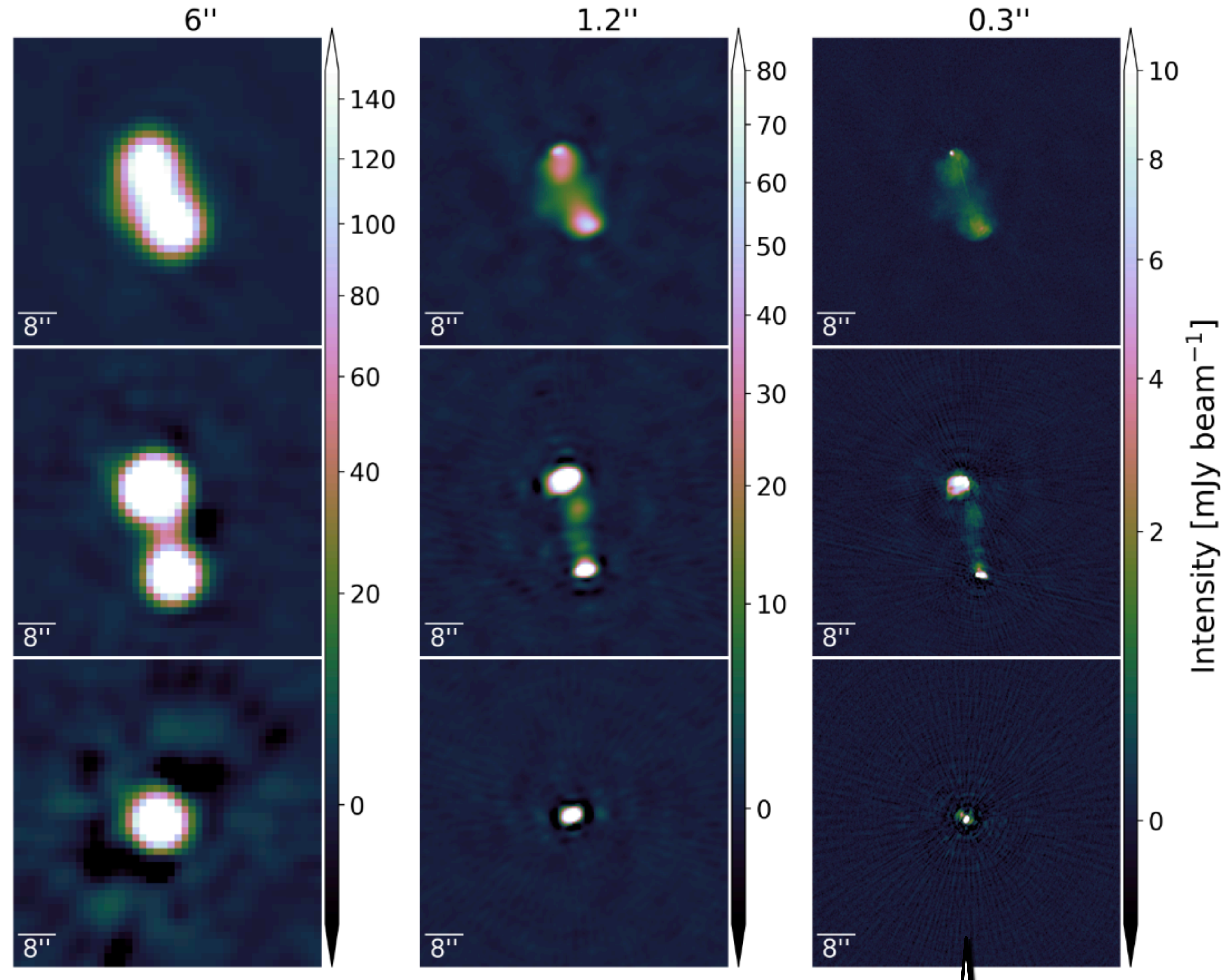
10 March 2024

THE COMPUTATIONAL CHALLENGE



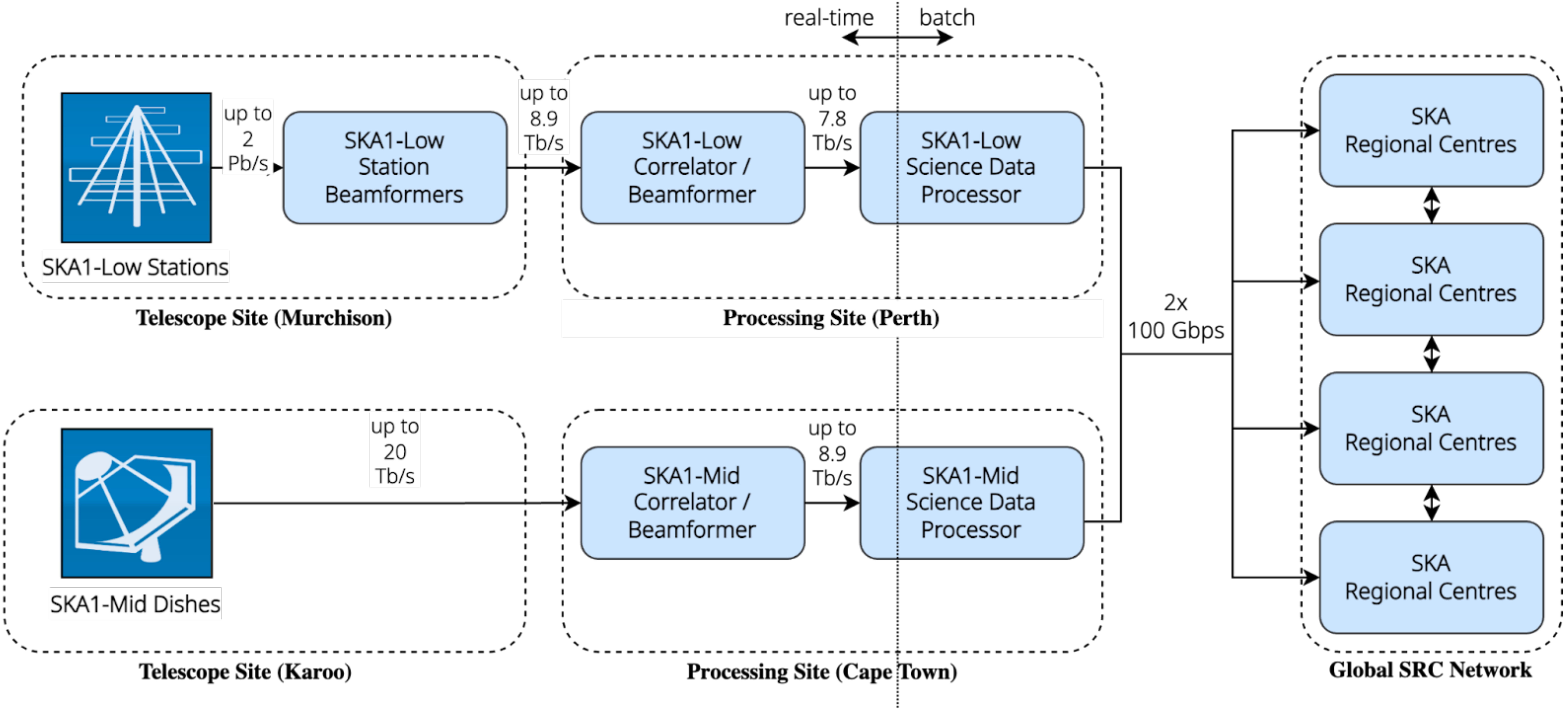
52,000 core hours (13,000 hours for imaging)
~6 days to image ~8 hour observation

[Ye+ 2023, arXiv:2309.16560]

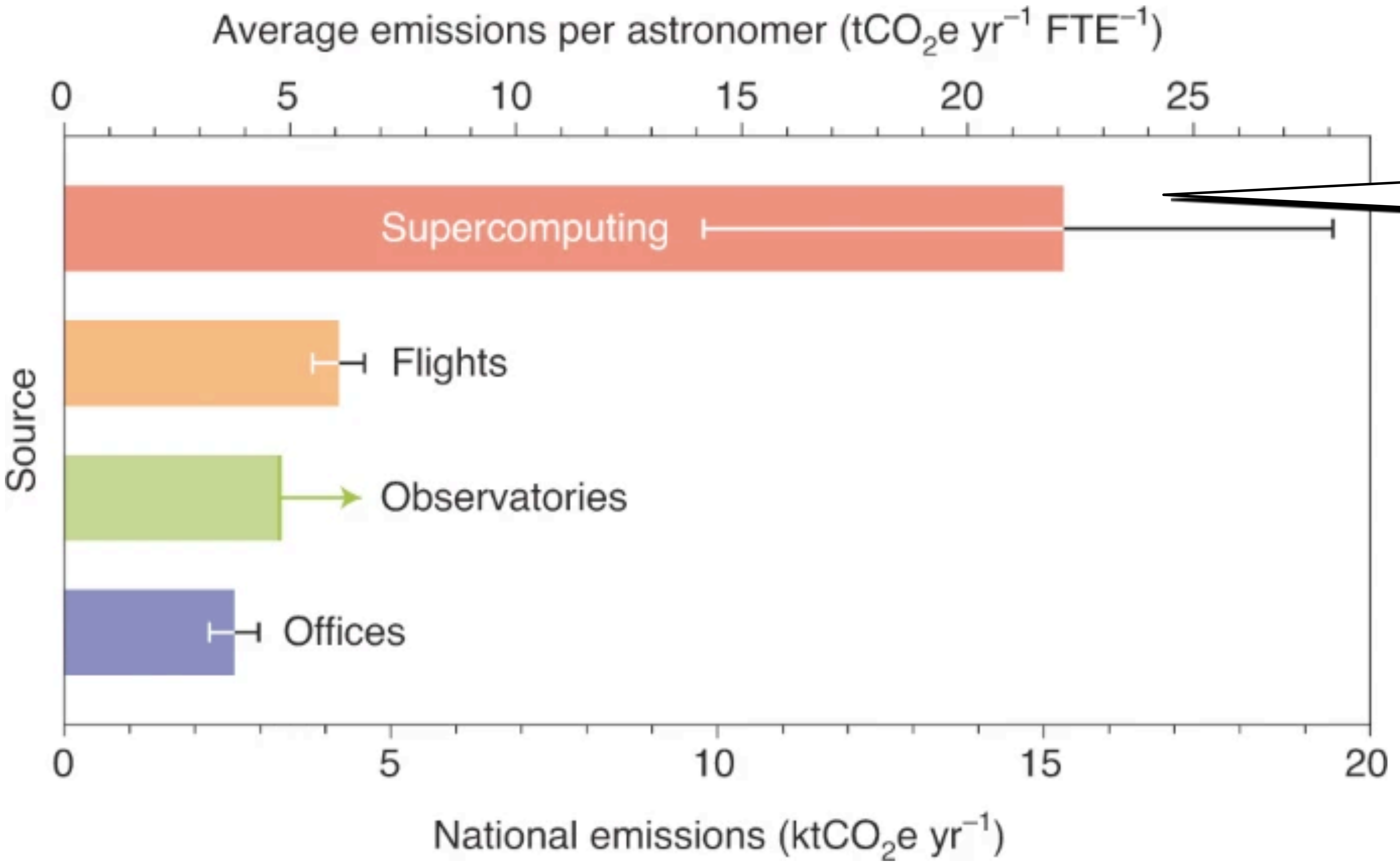


250,000 core hours for
0.3″ field

THE COMPUTATIONAL CHALLENGE



THE COMPUTATIONAL CHALLENGE



~40% Computational/theoretical astrophysics (simulations)
~30% radio astronomy data reduction
~30% optical & infrared

A. Stevens et al., "The imperative to reduce carbon emissions in astronomy" Nature Astronomy 2020

For reference, per-capita carbon emissions*….
in Australia: 17 tCO_2/yr
in Switzerland: 4.7 tCO_2/yr

*https://www.worldometers.info/co2-emissions/co2-emissions-per-capita/

...& SOLUTIONS

- **High Performance Computing**
  - GPU computing & parallelism (the focus of this talk)
- **Quantum computing**
  - Applications of the Quantum Fourier Transform for imaging (Brunet, Tolley,et al,2024)
- **Artificial intelligence**
  - For image reconstruction (Tajja, Aghabiglou, Tolley, et al,2025)
  - For data discovery (Misha, Tolley, et al, 2025)
  - Physics-informed neural networks for cosmological simulations (Korber, Tolley, Bianco, Kneib, 2023)

LIGHTNING OVERVIEW

Emma Tolley: Assistant Professor in the laboratory of Astrophysics and leader of the radio astronomy HPC group

Piyush Panchal

HPC development of MWA pulsar search pipeline

Swiss **SKA Regional center** platform development (See Rohini's talk!)

Hamza Chouh

Adrien Devresse

SKAO contract for **co-design** of the science data processor

HPC development of radio imaging software

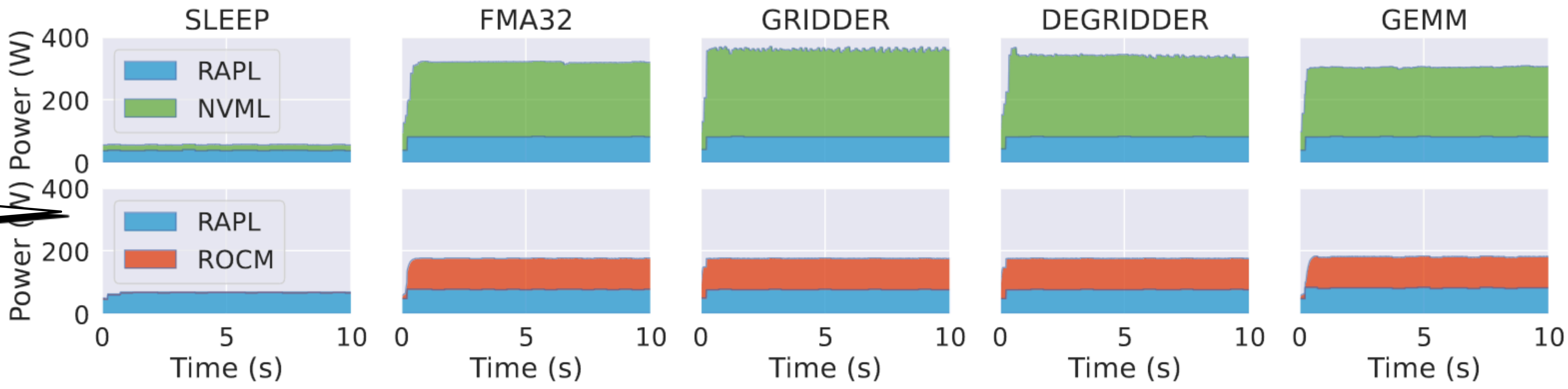
Workflow parallelism & data formats

Pablo Llopis

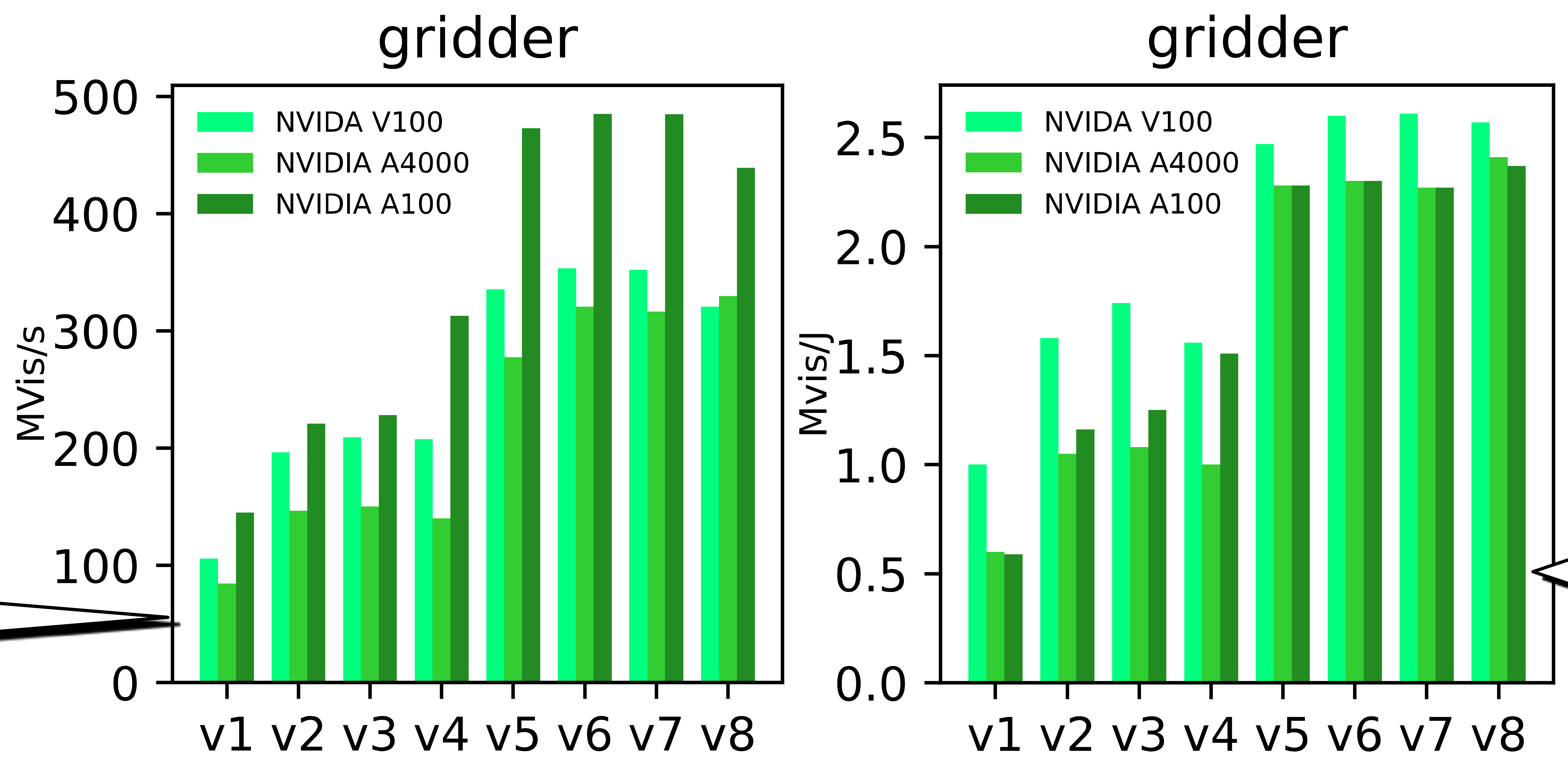
Carolina Lindqvist

Etienne Orliac

BENCHMARKING



PMT library for **Energy benchmarking** (arXiv:2210.03724)



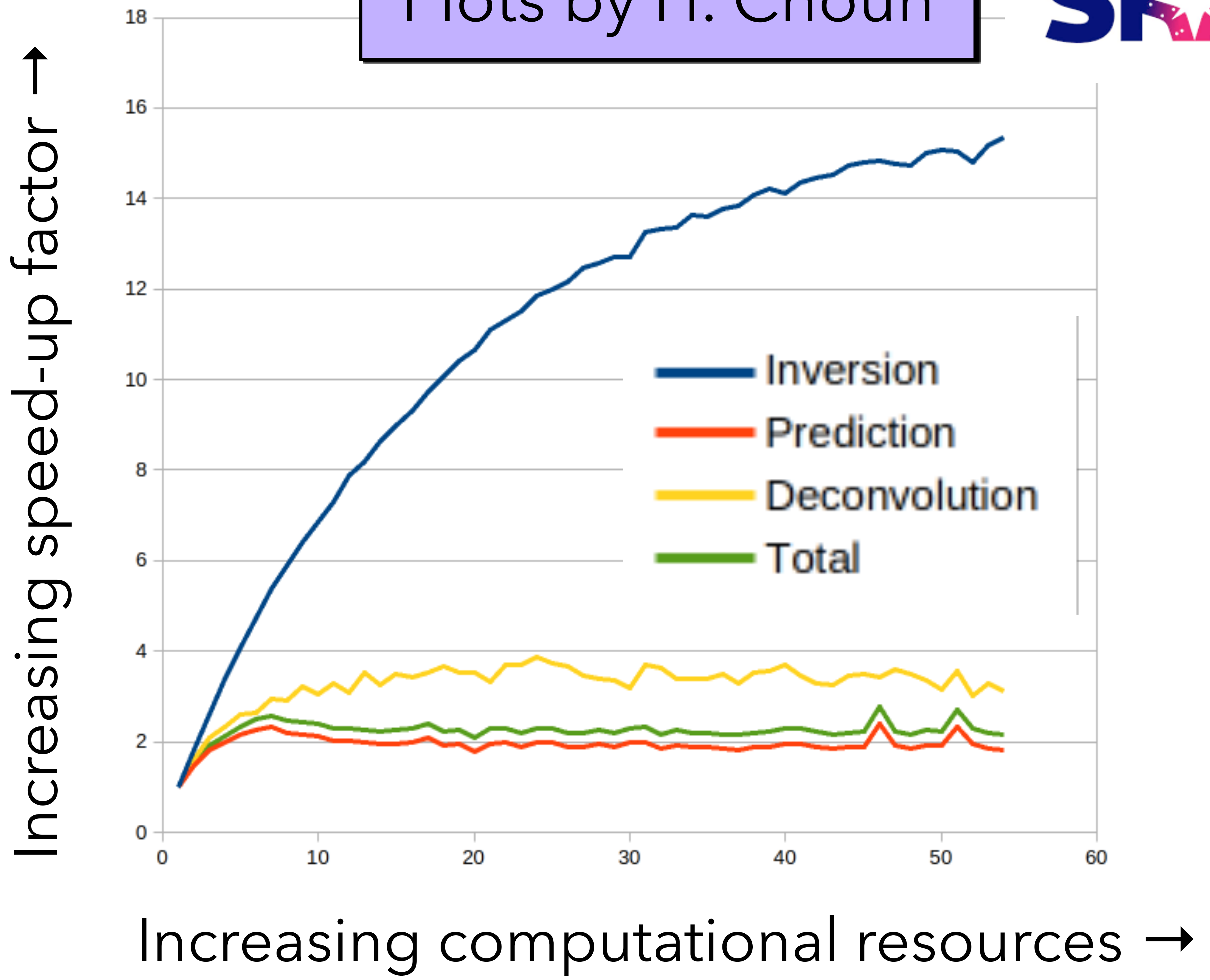
A100s can process more data/s…

… but V100s are more efficient

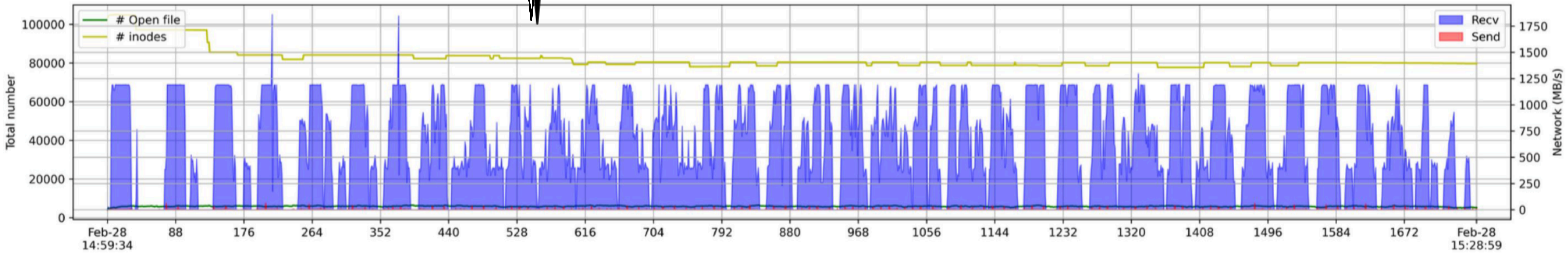
BENCHMARKING

Plots by H. Chouh

Not all parts of the workflow can scale to larger computing systems



Often I/O bound



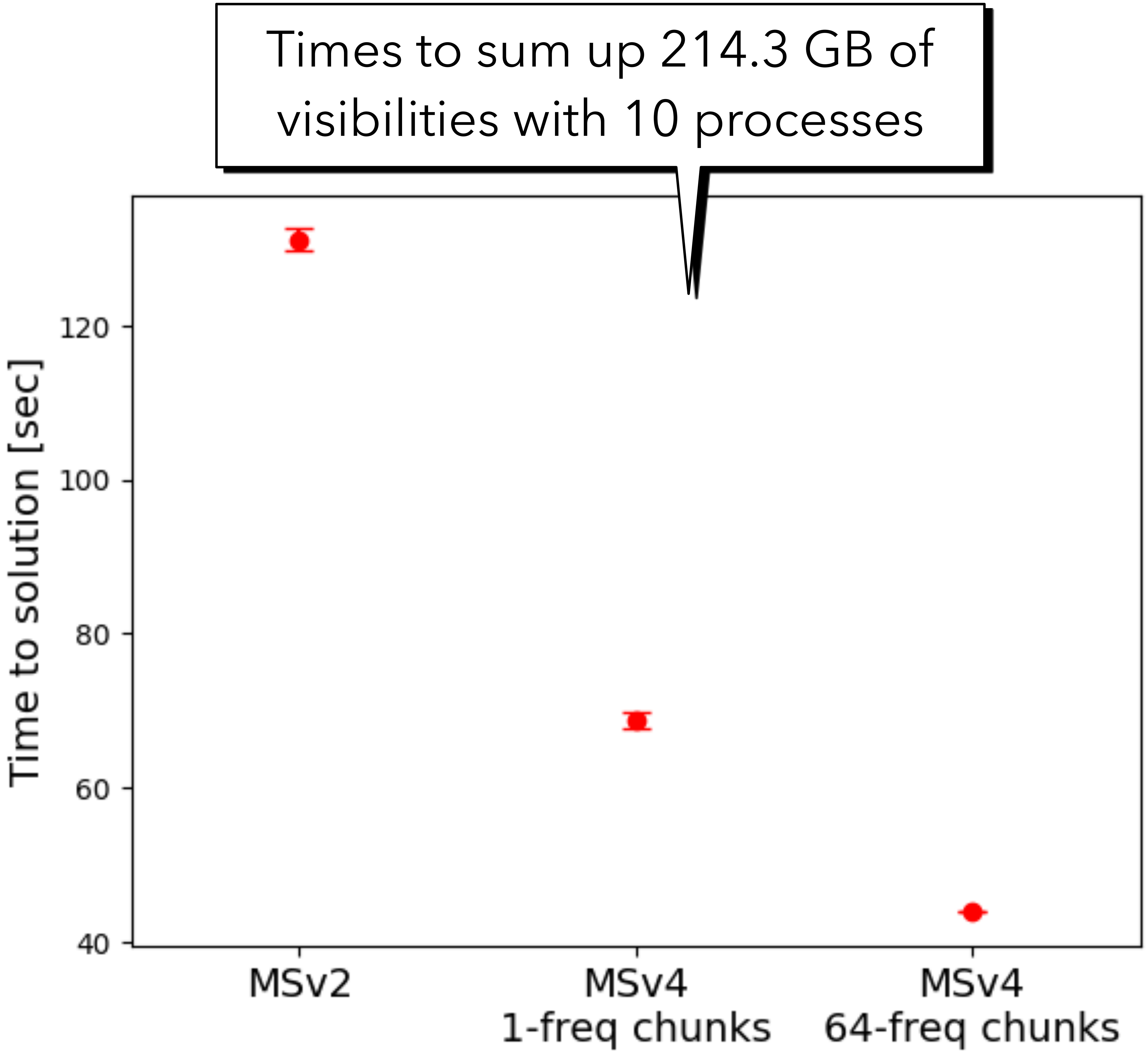
Batch preprocessing pipeline I/O usage (single node, 8 tasks)

A NEW DATA FORMAT?

Plot by E. Orliac

RADIO BLOCKS radioblocks.eu

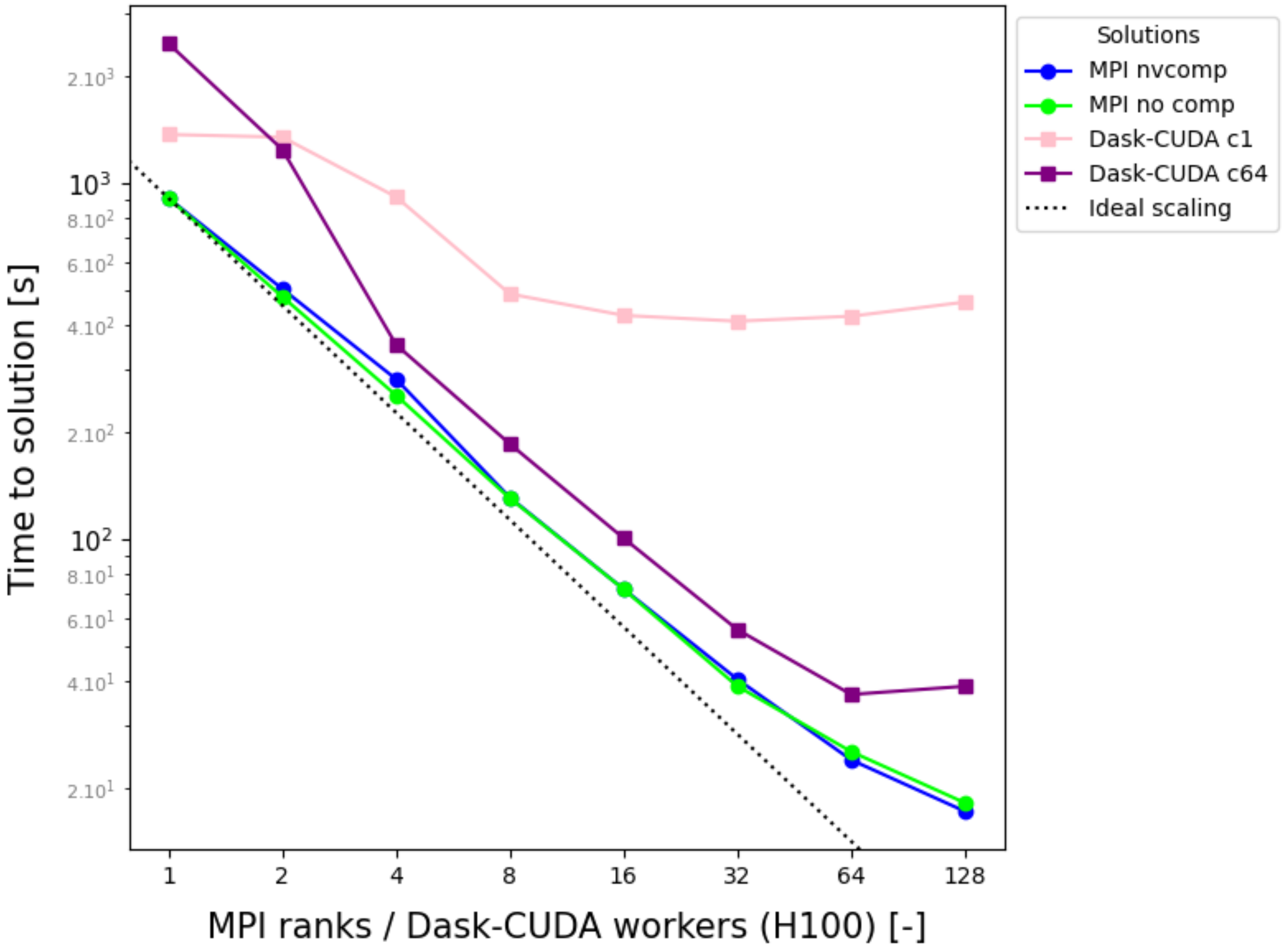
- **MSv2**: the familiar CASA tables data format
- **MSv4**: Zarr storage based on labeled n-dimensional arrays (xarray) datasets, under development as part of the ppen-source Python project **XRADIO** (Xarray Radio Astronomy Data IO)



A NEW WORKFLOW STRATEGY?

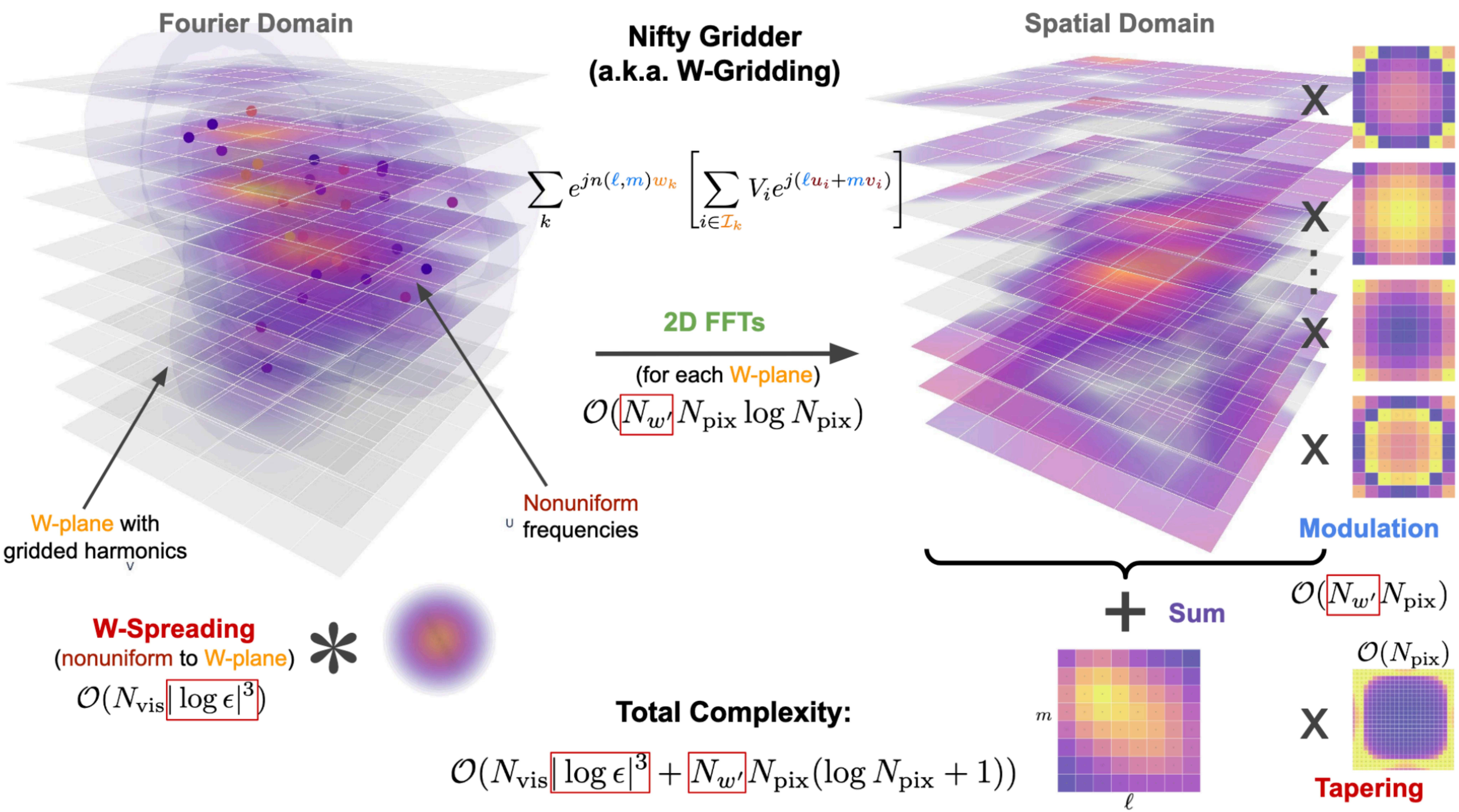
RADIO BLOCKS radioblocks.eu

Plot by E. Orliac



IMAGING

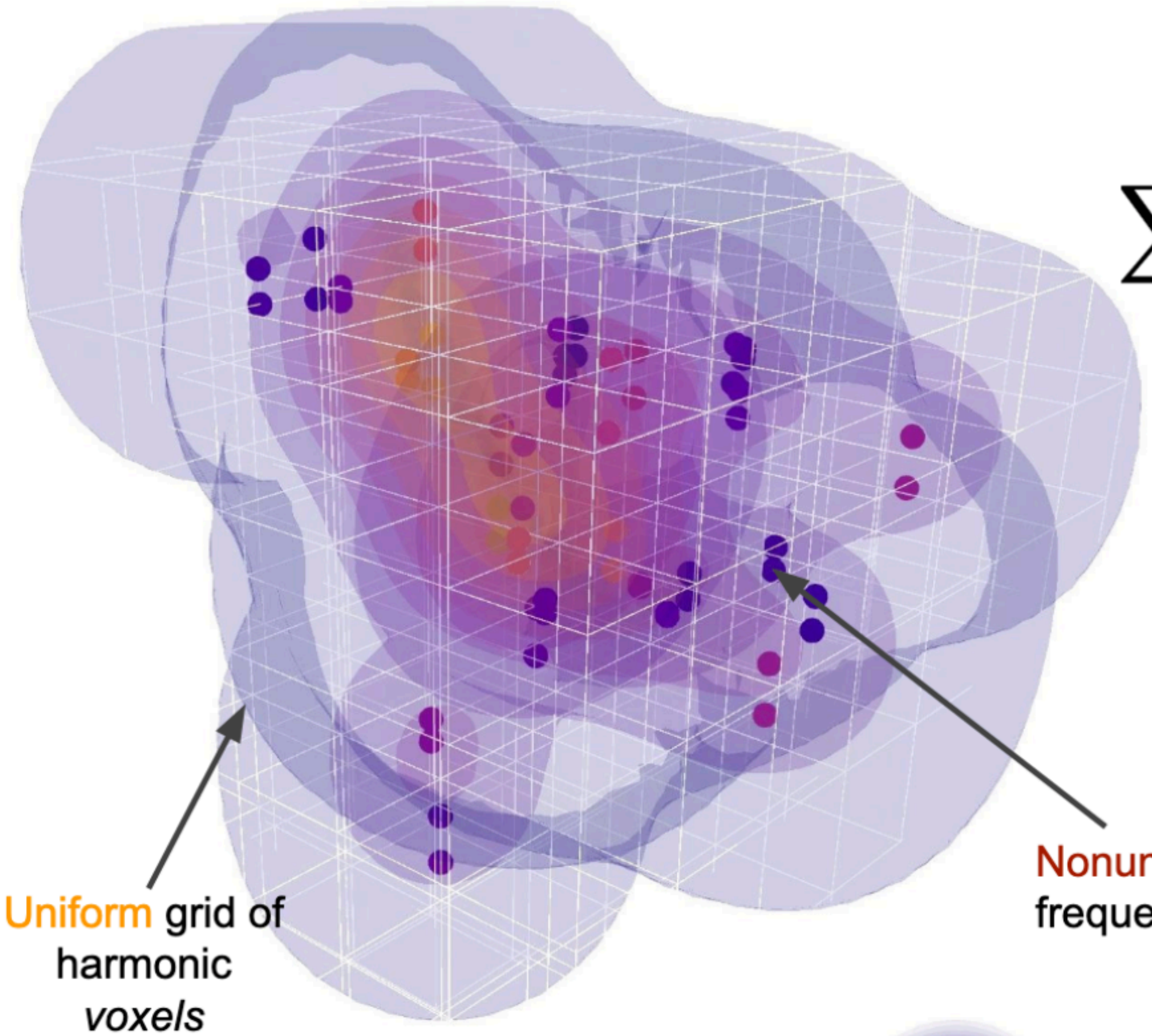
Kashani+, 2023, arXiv:2306.06007



IMAGING

Kashani+, 2023, arXiv:2306.06007

Fourier Domain



3D NUFFT (Type 3)

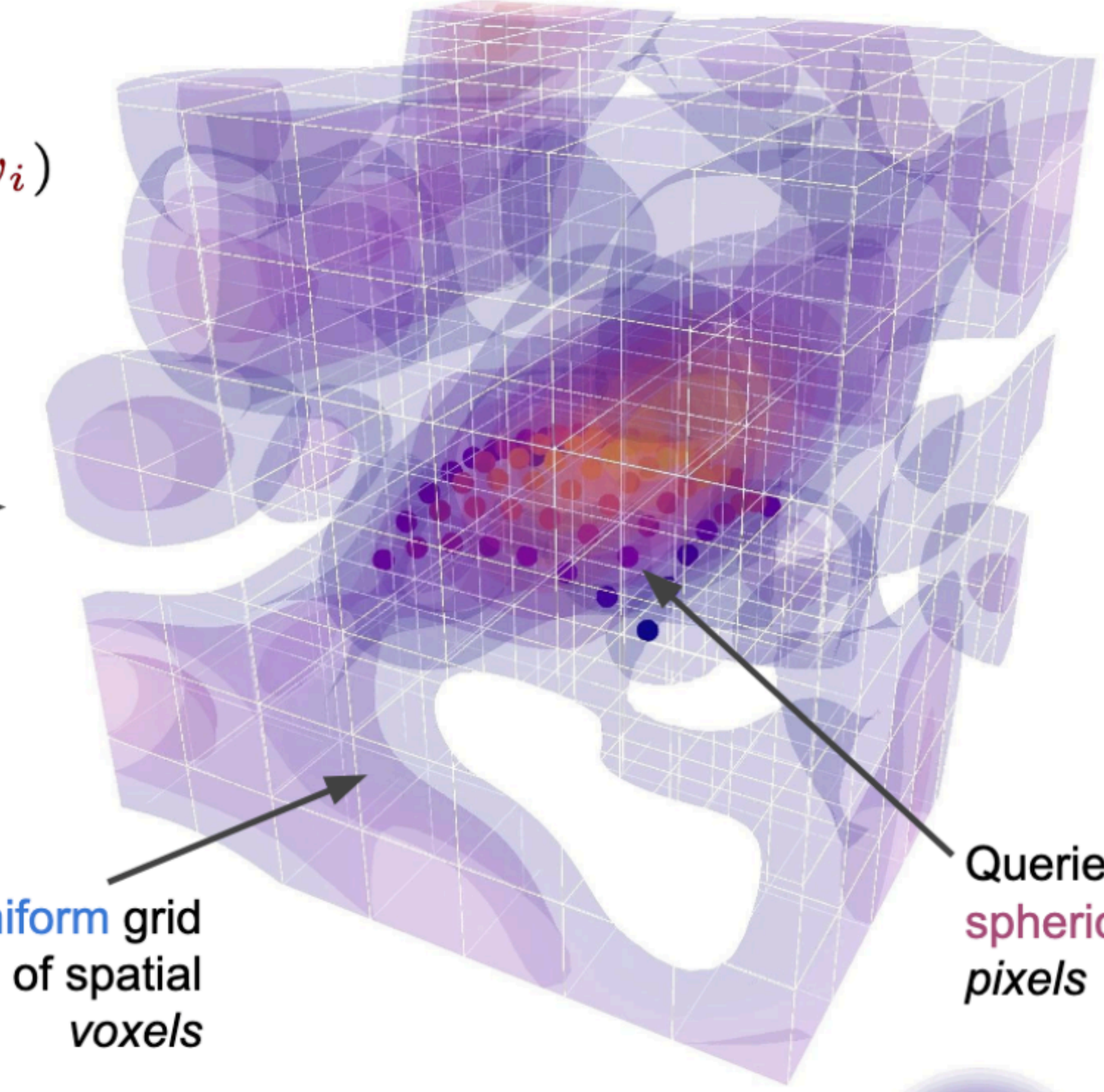
$$\sum_i V_i e^{j(\ell u_i + m v_i + n w_i)}$$

3D FFT

(uniform to uniform)

$$\mathcal{O}(N_{\text{mesh}} \log N_{\text{mesh}})$$

Spatial Domain



Spreading (+Tapering)
(nonuniform to uniform)

$$\mathcal{O}(N_{\text{vis}} |\log \epsilon|^3 + N_{\text{mesh}})$$

$*$ 

Interpolation (+Tapering)
(uniform to nonuniform)

$$\mathcal{O}(N_{\text{pix}} |\log \epsilon|^3 + N_{\text{pix}})$$

$*$ 

Total Complexity:

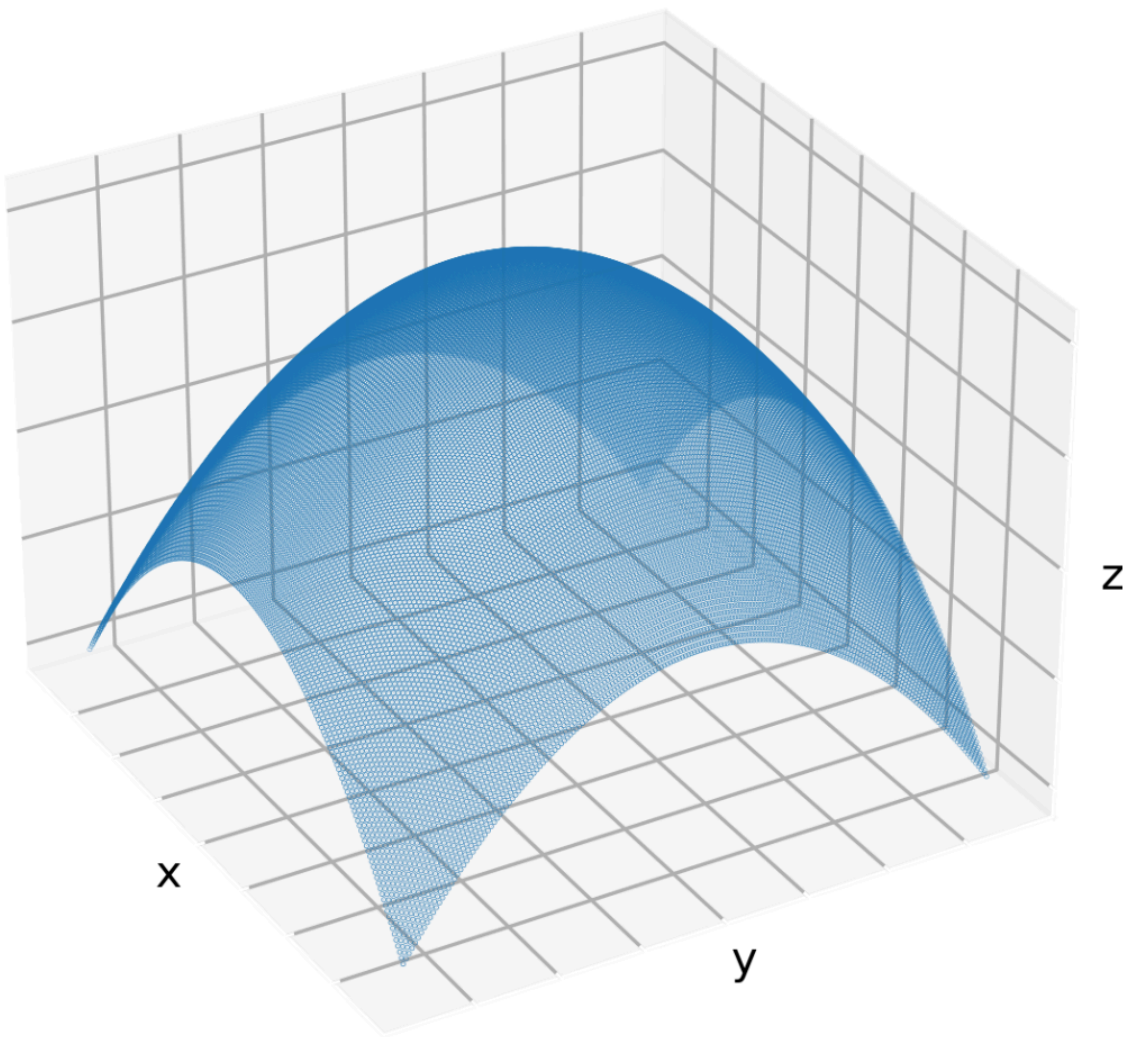
$$\mathcal{O}((N_{\text{vis}} + N_{\text{pix}}) |\log \epsilon|^3 + N_{\text{mesh}}(\log N_{\text{mesh}} + 1) + N_{\text{pix}})$$

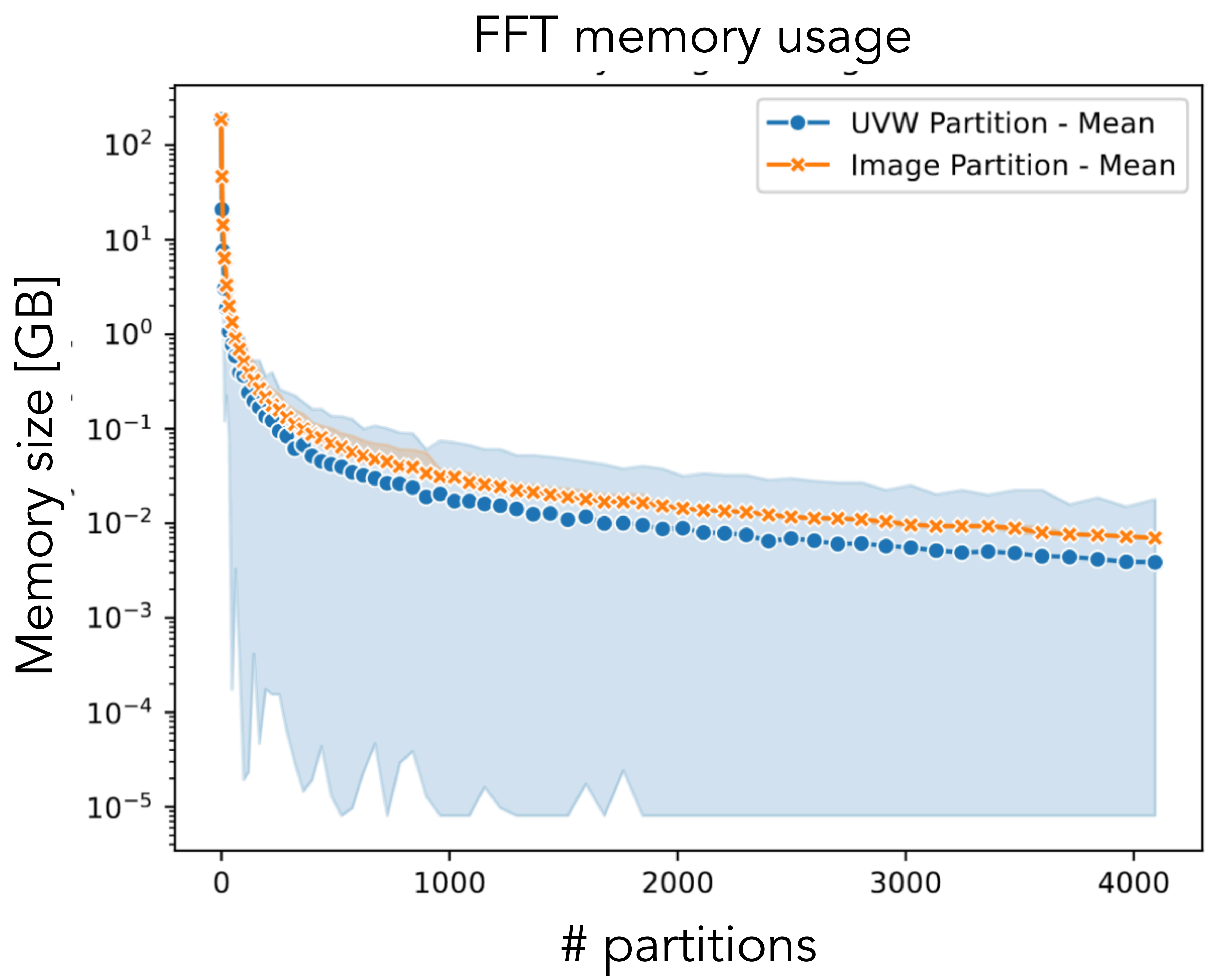
HPC IMPLEMENTATION

Plots by S. Frasch (CSCS)

Platform for Advanced Scientific Computing

Directly reconstruct images on the celestial sphere using 3D type-3 non-uniform FFT (NUFFT)





Can use **domain partitioning** to take advantage of sparse input & output domains, reduce memory consumption, improve performance

Bluebild Imaging++ (BIPP): CPU & GPU implementation with CUDA & HIP, funded by PASC 2021-2025, collaboration between EPFL & CSCS. Now published: [Tolley et al, j.ascom.2024.100920]

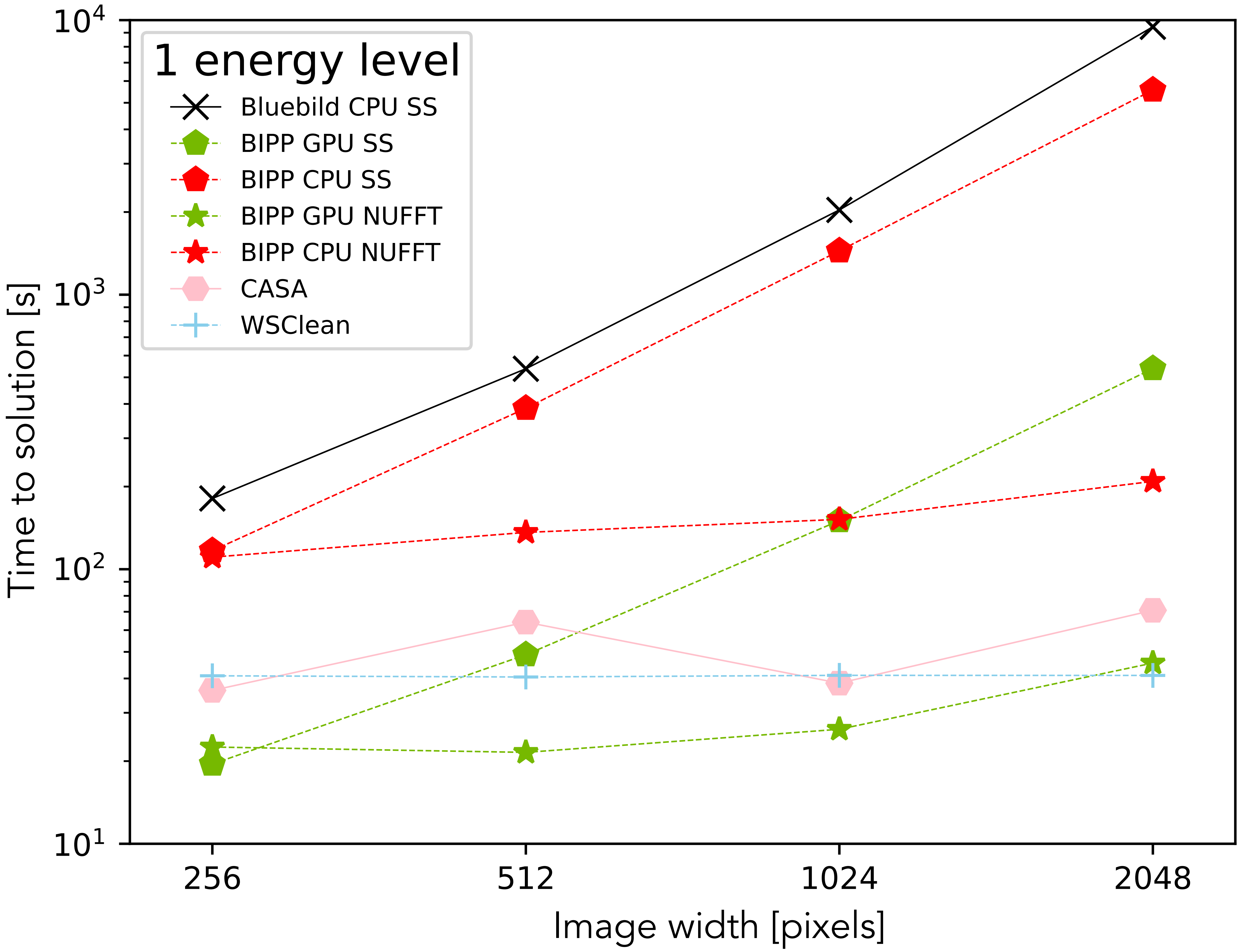
Platform for Advanced Scientific Computing

pasc-ch.org

HPC IMPLEMENTATION

Plot by E. Orliac

Platform for Advanced Scientific Computing



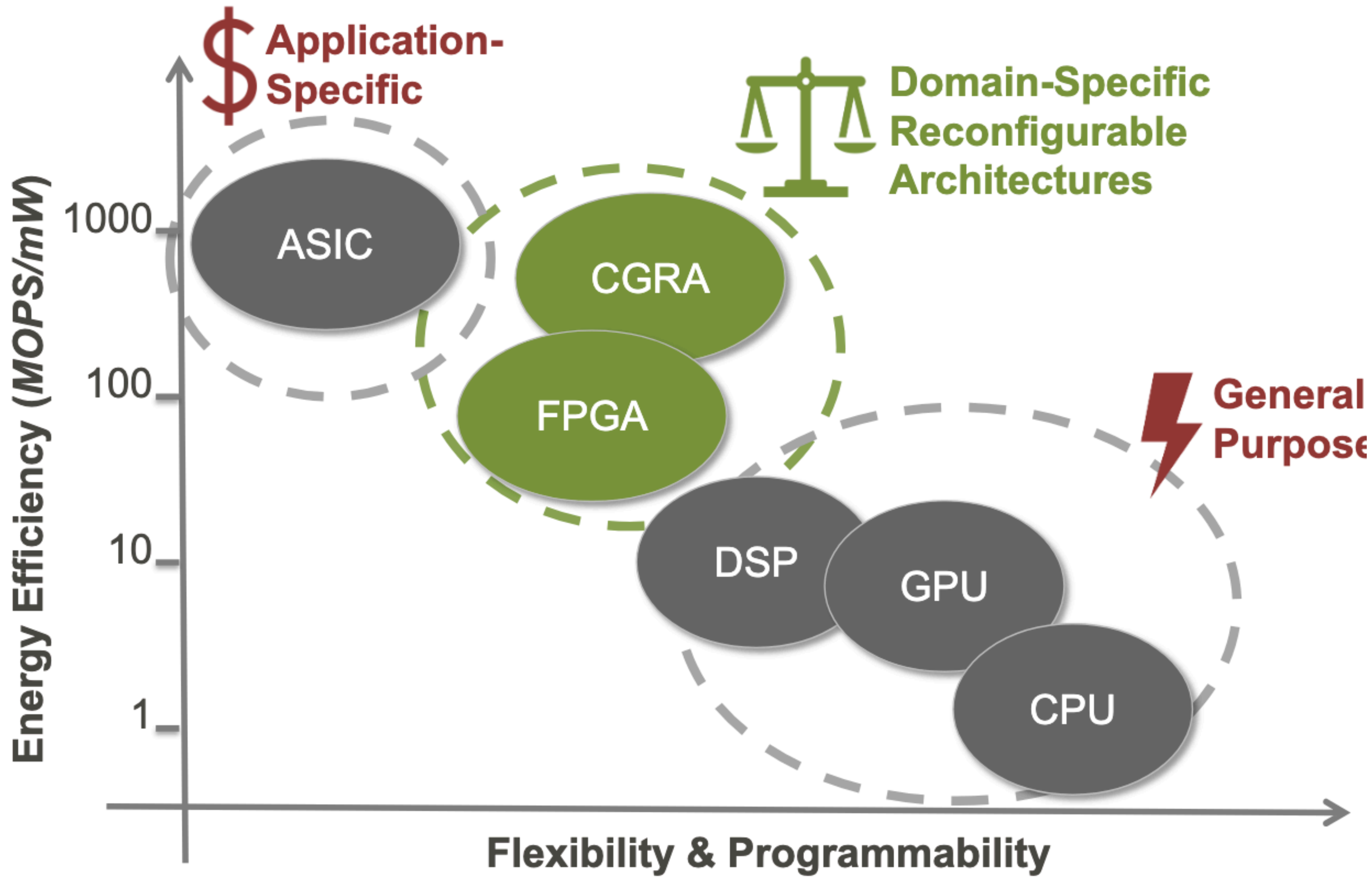
Process	Time [s]	% Total
Total	45.57	100.0
Parameter estimation	4.42	9.7
Reading visibilities	1.91	4.2
Processing	2.50	5.9
Eigen decomposition	0.60	1.3
Imaging	41.03	90.0
Reading visibilities	1.84	4.0
Processing	39.19	86.0
Eigen decomposition	0.50	1.1
Other	0.12	0.3

Continuing to improve the 3D NUFFT with a dedicated HPC library **neonufft**

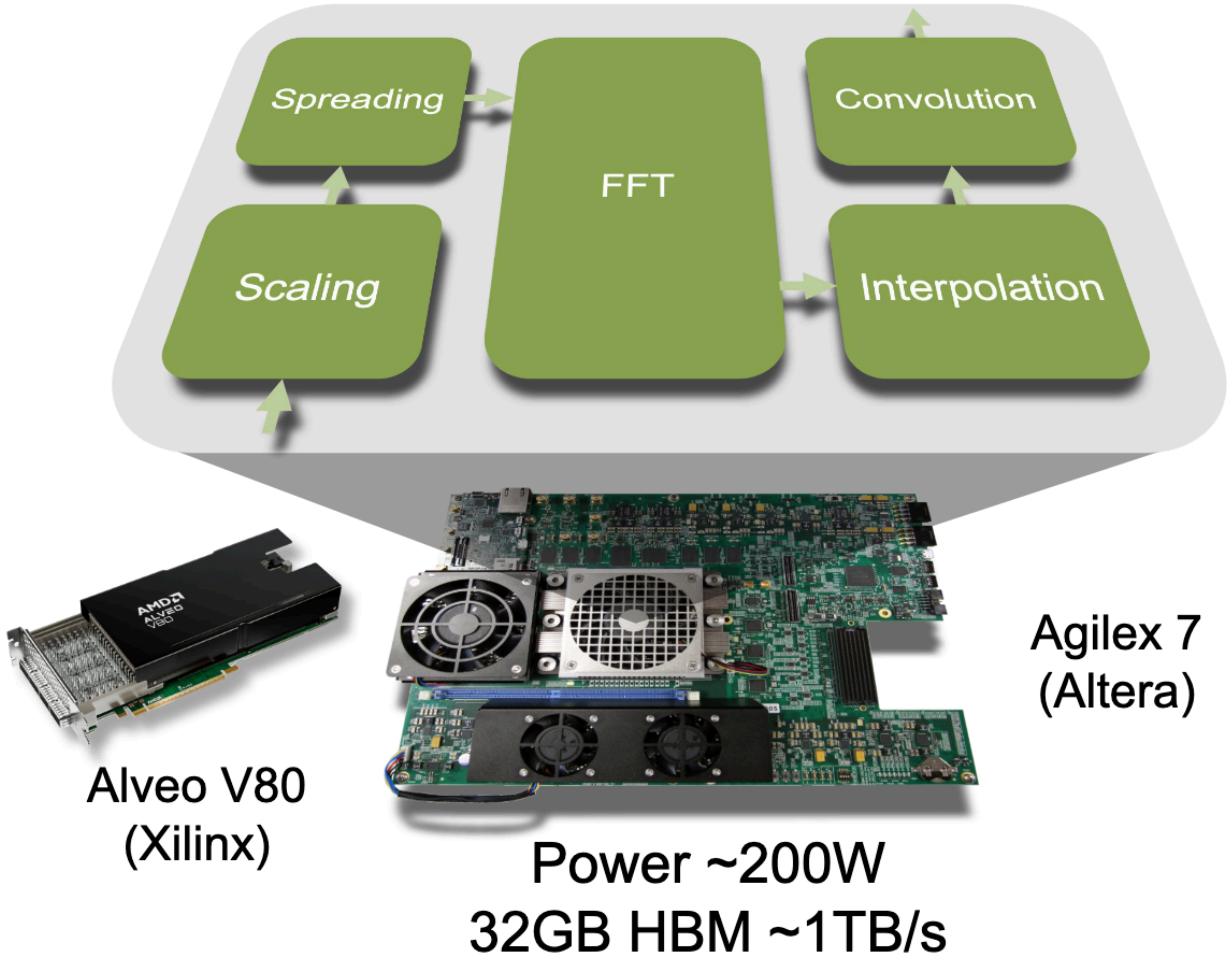
Bluebild Imaging++ (BIPP): CPU & GPU implementation with CUDA & HIP, funded by PASC 2021-2025, collaboration between EPFL & CSCS. Now published: Tolley et al, j.ascom.2024.100920

SPECIALIZED HARDWARE

Specialization for Energy



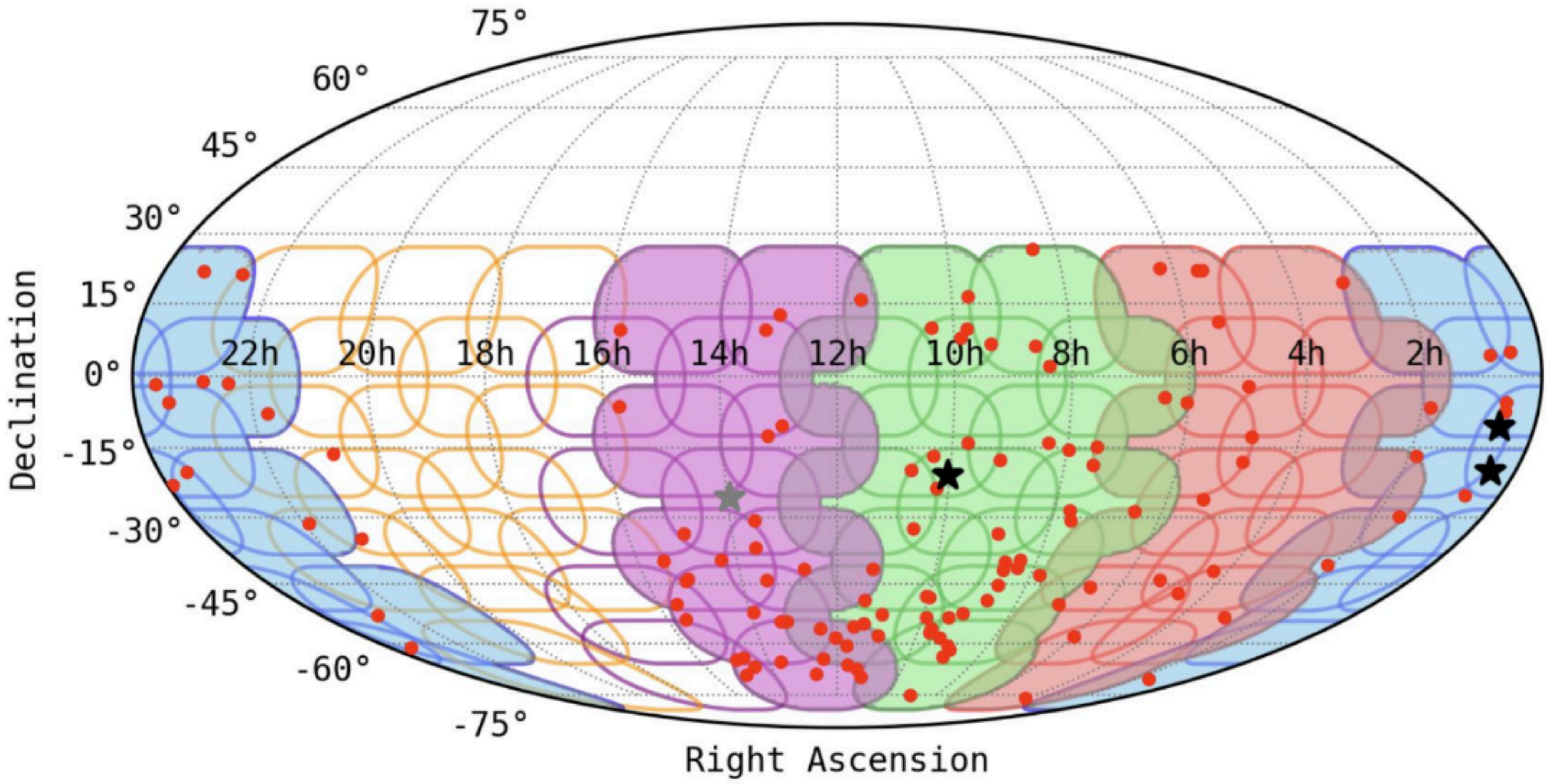
Dataflow Design Paradigm



Rubén Rodríguez Álvarez, Denisa Constantinescu, Miguel Peón-Quirós (EPFL)

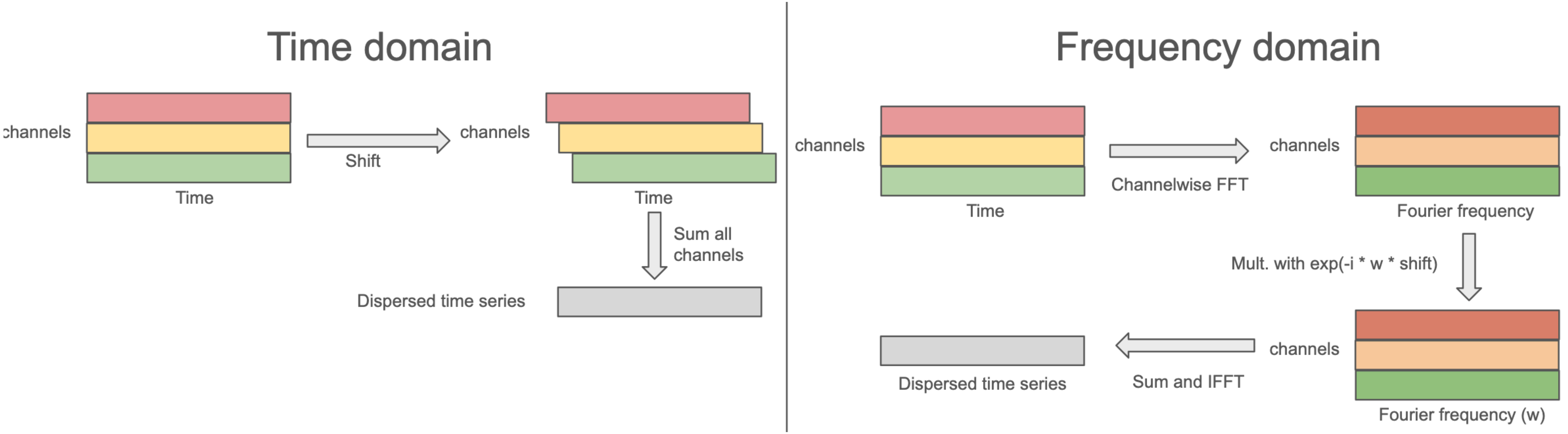
PULSAR SEARCH

- 4 PB data collected in SMART Pulsar survey (MWA)
- Shallow processing (~10% data) using Presto: ~300K CPU hours
- Problem: Full processing using Presto in ~23 years
- Motivation: Pulsars help studying extreme physics
- Goal: GPU acceleration to enable full/faster processing



PULSAR SEARCH

Results by P. Panchal



	Dedisp (FDD, GPU)	PrestoZL (TDD, GPU)	Presto (TDD, CPU)
1000 DMs	28.95 s	> 139.77 s (800 DM)	8 hrs+
500 DMs	23.29 s	98 s	4.75 hrs

Kuma (H100): 15000 DMs, 116.223 s!

WHAT'S NEXT?


Exposures over 30 minutes from latitude 51° N
Image © Alan Dyer/AmazingSky.com

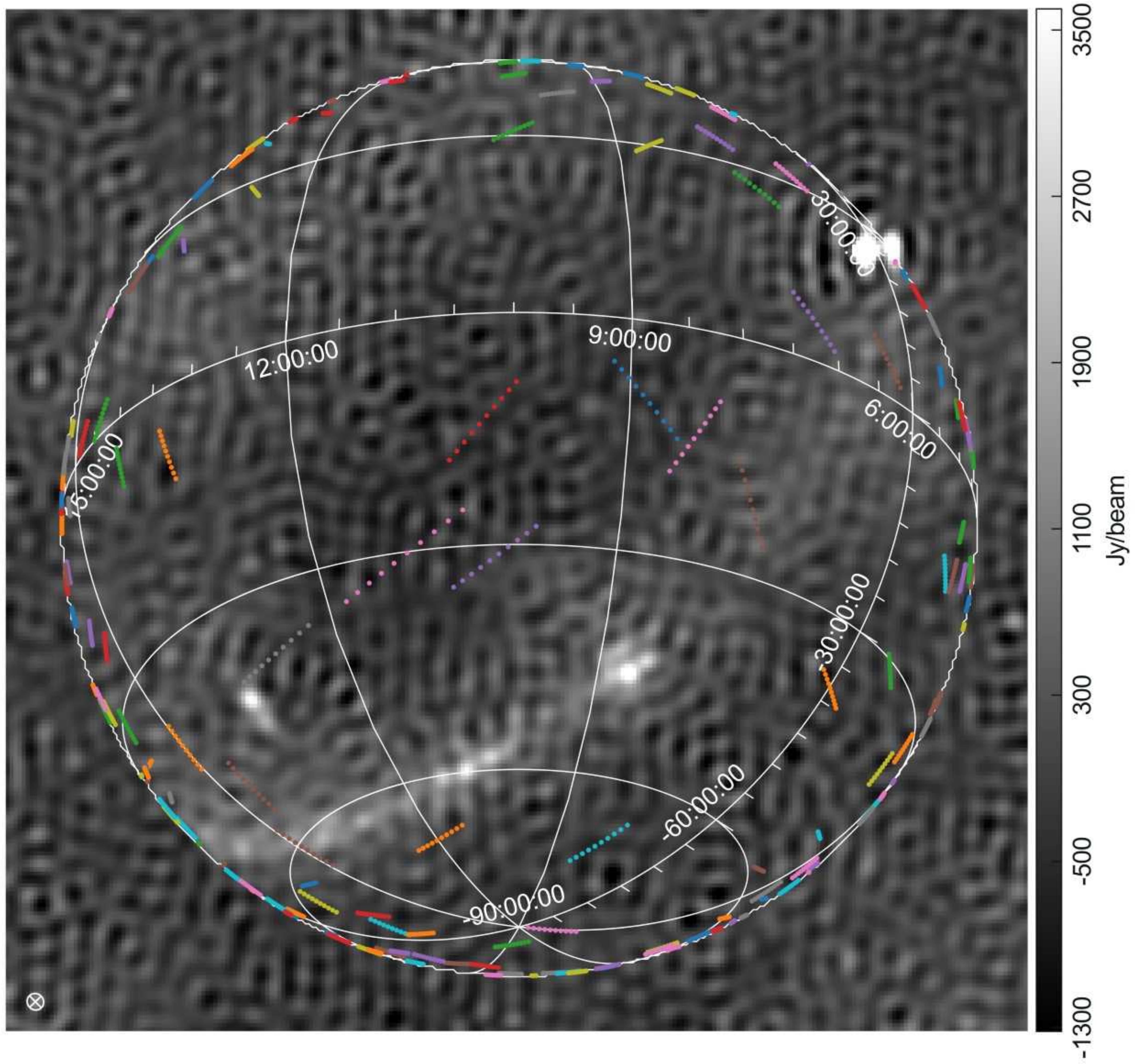
Image © Alan Dyer/AmazingSky.com

RFI FROM SATELLITES

Expected starling satellites visible in a 20s interval for the SKA Engineering Development Array 2

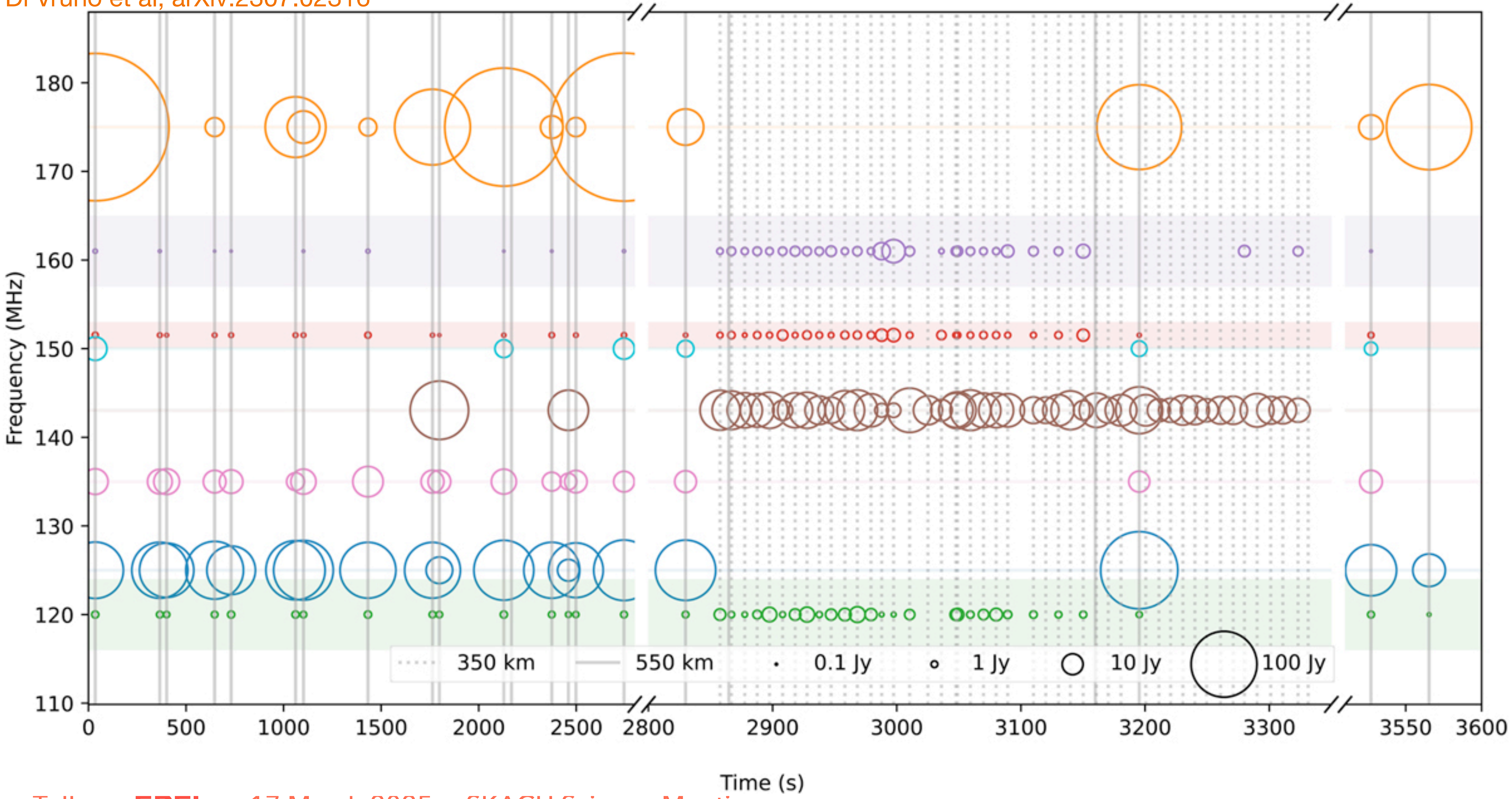
- **Intentional radio emission** (communications, remote sensing, radionavigation, etc.)
  - Regulated by the ITU-R (2020) Radio Regulations (RR)
  - Starlink assigned transmission frequencies: 10.7–12.7 GHz
- **Unintended electromagnetic radiation** (UEMR) generated by electrical circuits on the satellite
  - Not explicitly regulated at the ITU-R level
- Reflections of radio emission from terrestrial or other external sources

Figure credit: H. Thums (EPFL)



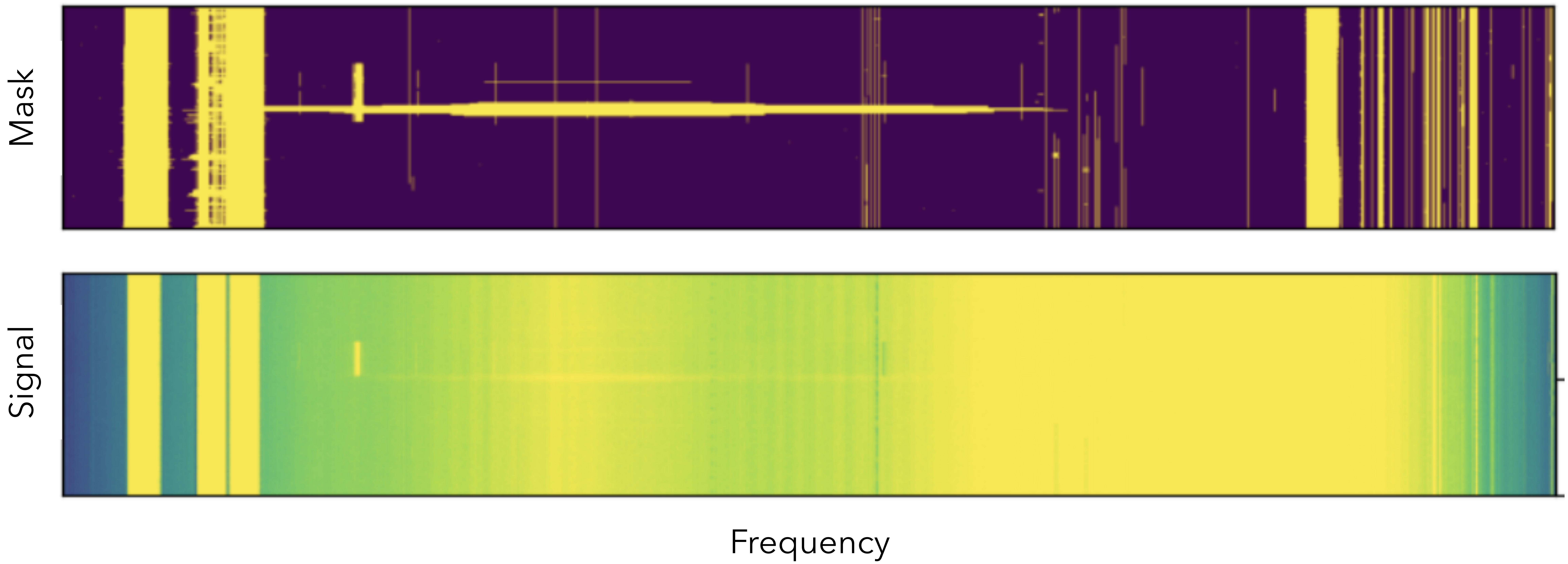
UEMR FROM STARLINK SATELLITES

F. Di Vruno et al, arXiv:2307.02316



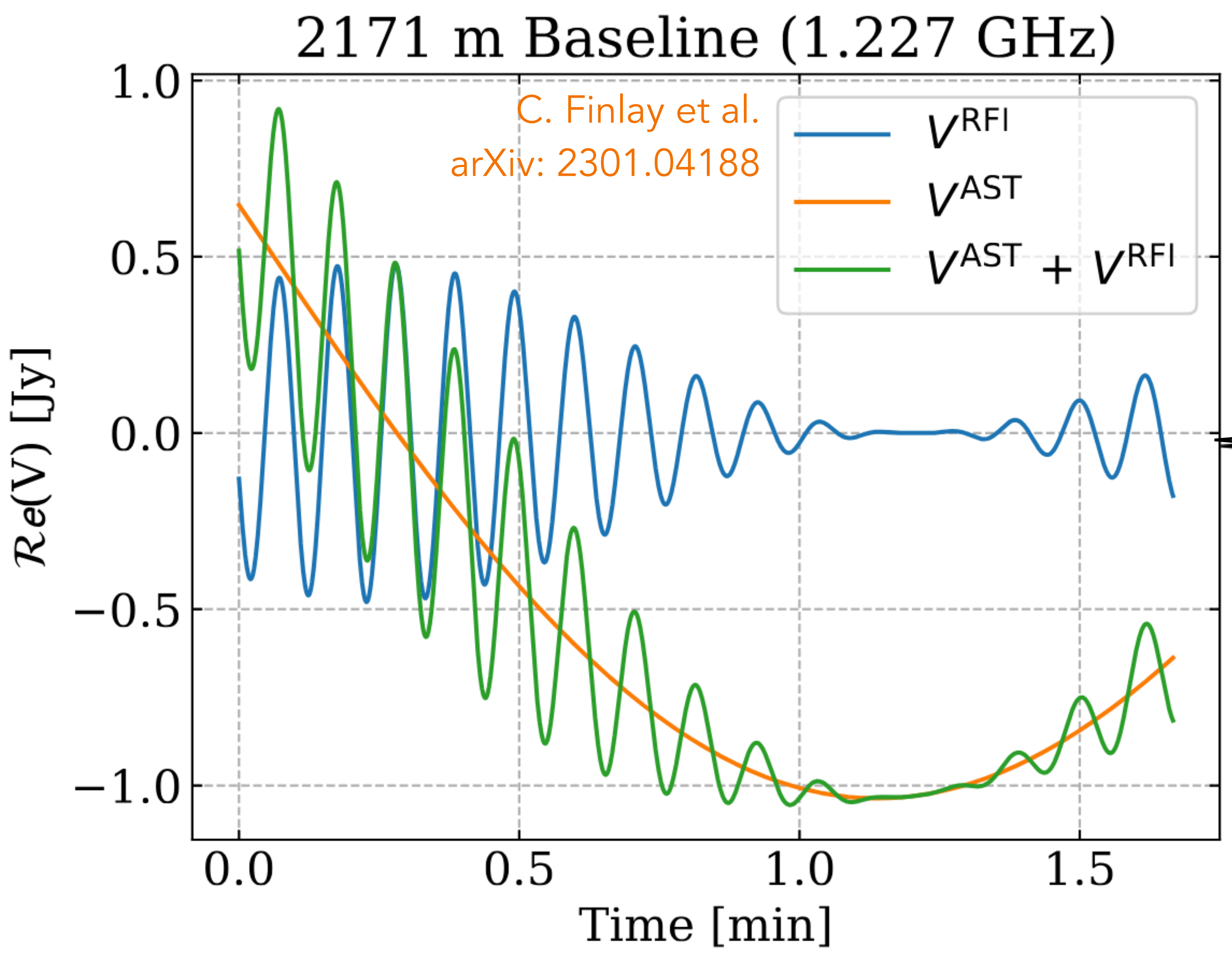
UEMR FROM STARLINK SATELLITES

Evaluating the impact on science use cases… see talk at upcoming 2025 SKAO General Science Meeting in Görlitz, Germany



Plots by N. Cerardi (EPFL)

SUBTRACTING SATELLITE RFI



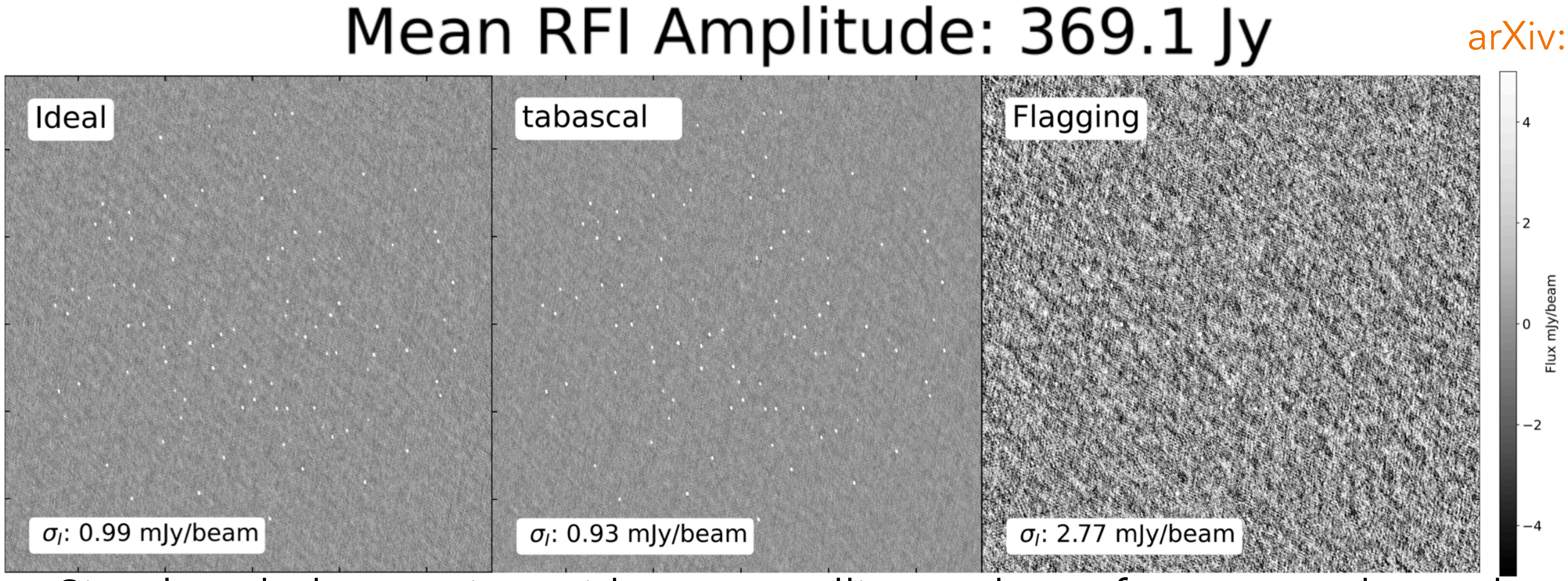
Can we **fit for** and **subtract** the RFI components and recover astrophysical signal instead of discarding data?

$$\tilde{V}_{pq}(t) = G_p(t)\left(V_{pq}^{\text{AST}}(t) + V_{pq}^{\text{RFI}}(t)\right)G_q^H(t)$$

`tabascal` code developed by C. Finlay et al.: arXiv: 2301.04188

IMAGING WITH ESTIMATED VISIBILITIES

C. Finlay et al.
arXiv: 2301.04188



Simulated observation with one satellite and one frequency channel

- Can we use Bayesian forward modeling to fit and subtract satellite signal?
    - Need to demonstrate that this **scales** to the problem size of the SKA & works on **real data**
    - Beginning to work on HPC implementation of `tabascal`, funded by the Swiss Platform for Advanced Scientific Computing (PASC) & EPFL (2025-2028)
    - Applying to real observations, extending model to multi-frequency, etc

IN SUMMARY

- Team is working a wide range of topics in HPC & computing for radio astronomy:
    - **SRCNet** platform development
    - **Benchmarking** of SKAO pipelines
    - Exploring **new data formats and new workflow management** strategies
    - Development and acceleration of new algorithms for **imaging** and **pulsar searching**
- So far getting excellent performance from GPUs over CPUs, cannot discount the power of accelerators! Should be seriously considered for SDP procurement