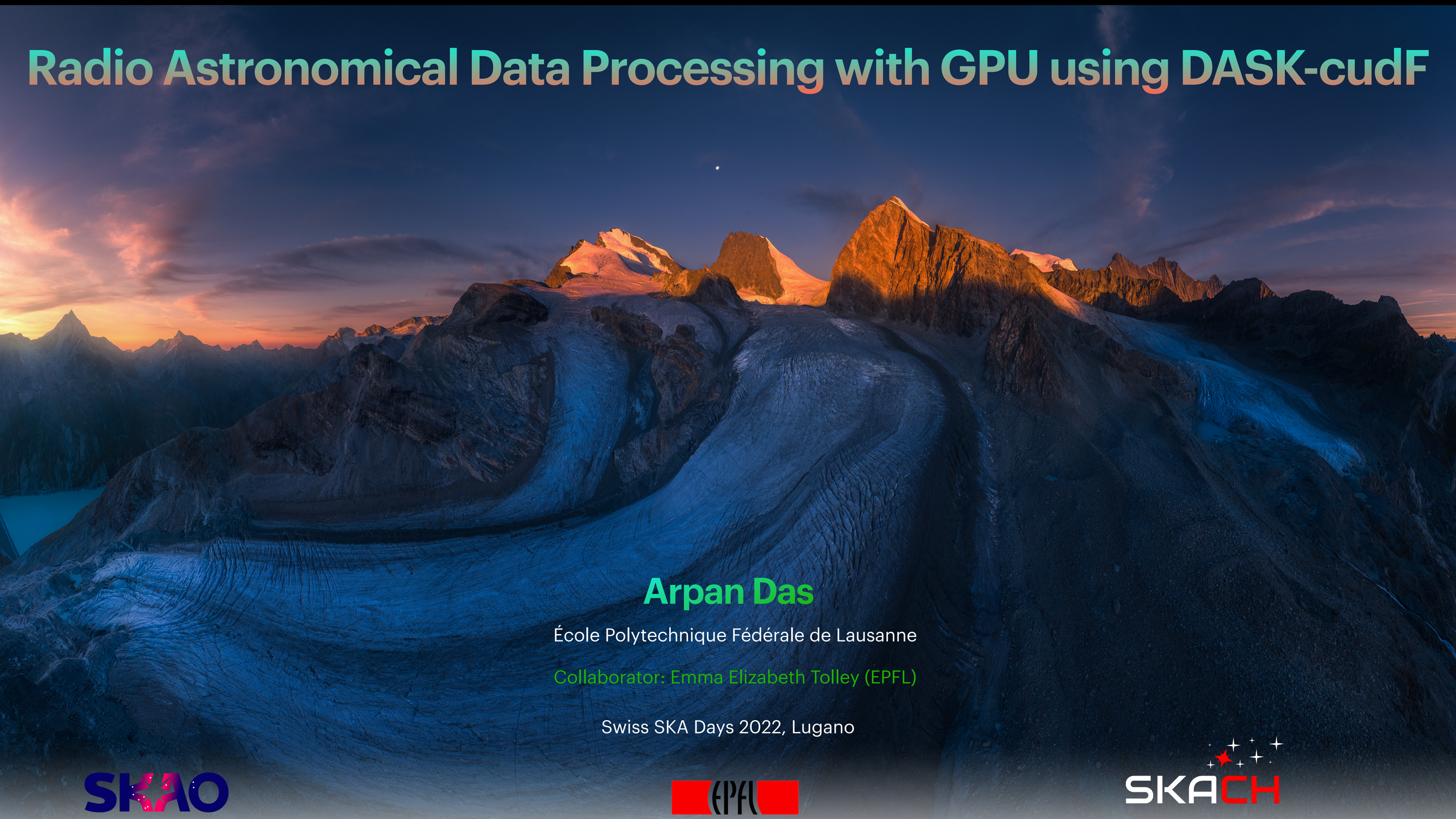
Radio Astronomical Data Processing with GPU using DASK-cudF



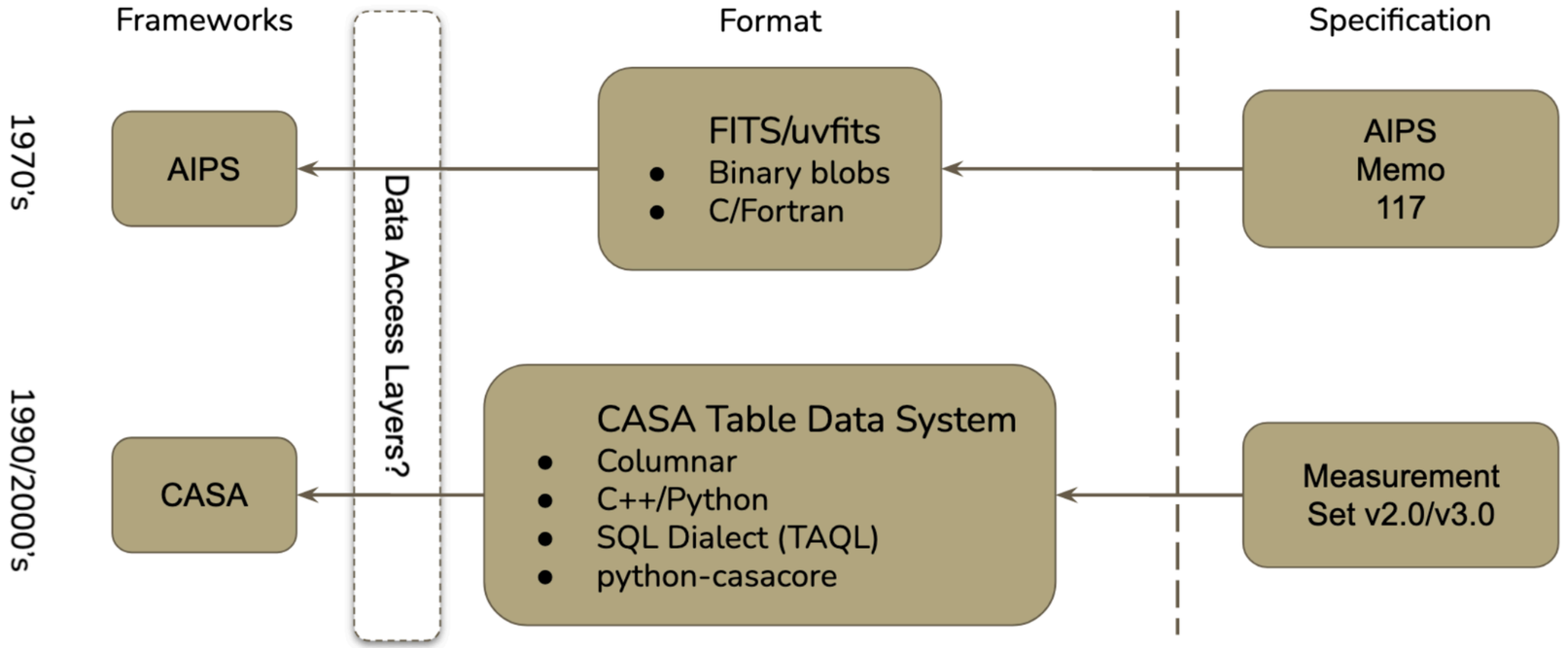
Arpan Das

École Polytechnique Fédérale de Lausanne

Collaborator: Emma Elizabeth Tolley (EPFL)

Swiss SKA Days 2022, Lugano

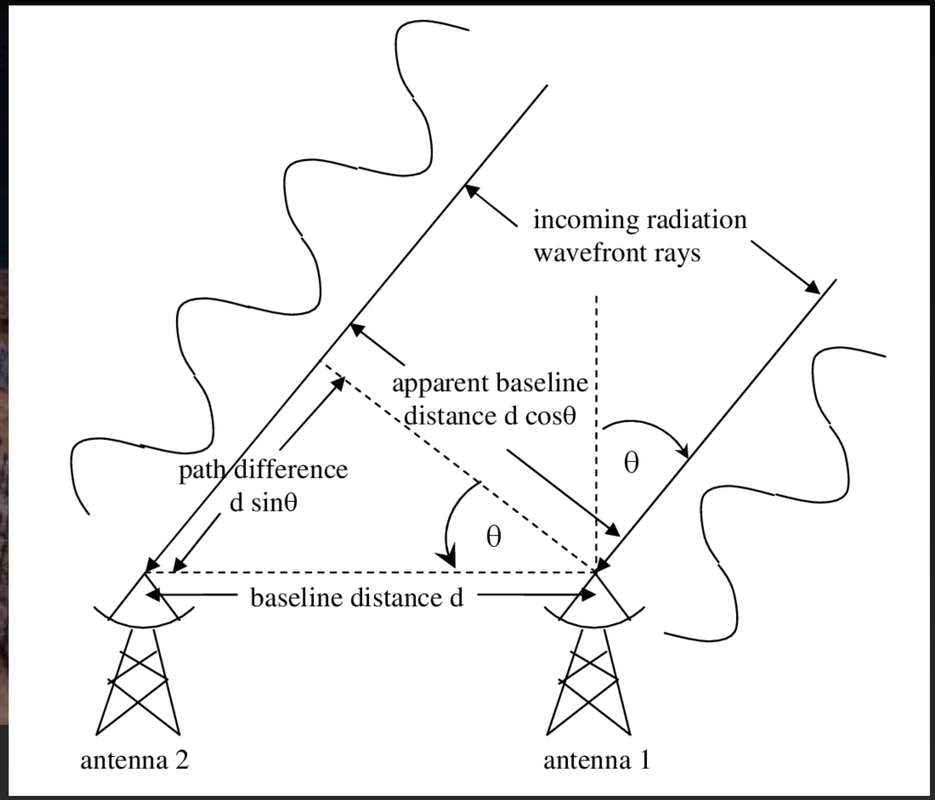
Some Background



=> All data for a single column is stored in one file, access to column data is controlled by a per-process file lock

=> Corruption to part of the MS leads to loss of entire dataset

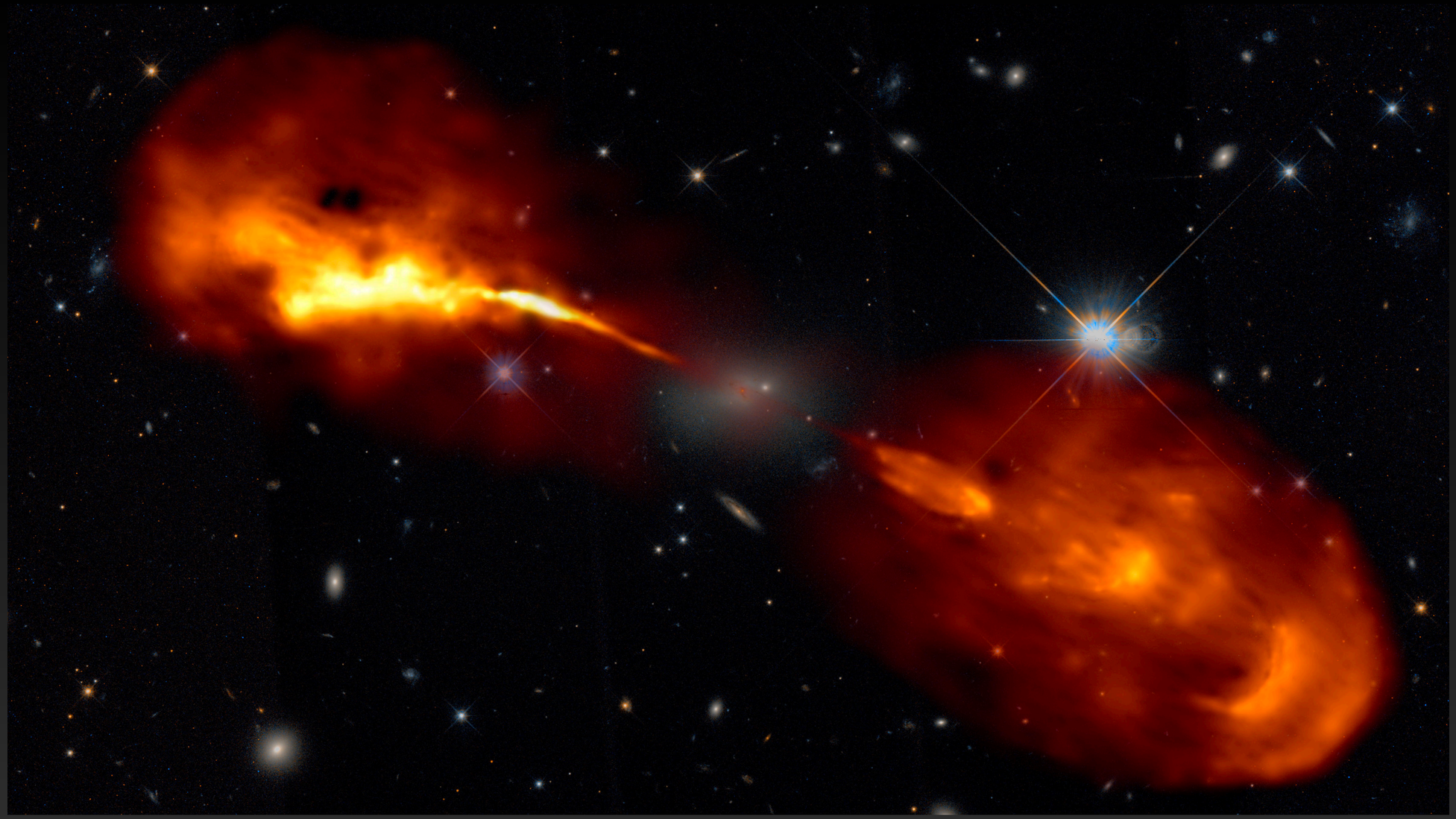
Radio Telescope Data (Frequency Domain)



Space Complexity $O(T \times A^2/2 \times \lambda)$

$T = \text{Timesteps}\,, A = \text{Antenna}\,, \lambda = \text{Channels}$

Email: arpan.das@epfl.ch



The Hercules A black hole jets captured in a high-resolution image captured by the LOFAR radio telescope.

More Data (time steps, antennas, channels) > Better image

Email: arpan.das@epfl.ch

CASA MS table :

- The MeasurementSet (MS) defines a format in which interferometer visibilities and single-dish data are stored.
- It is implemented in software packages using casa/casacore code.

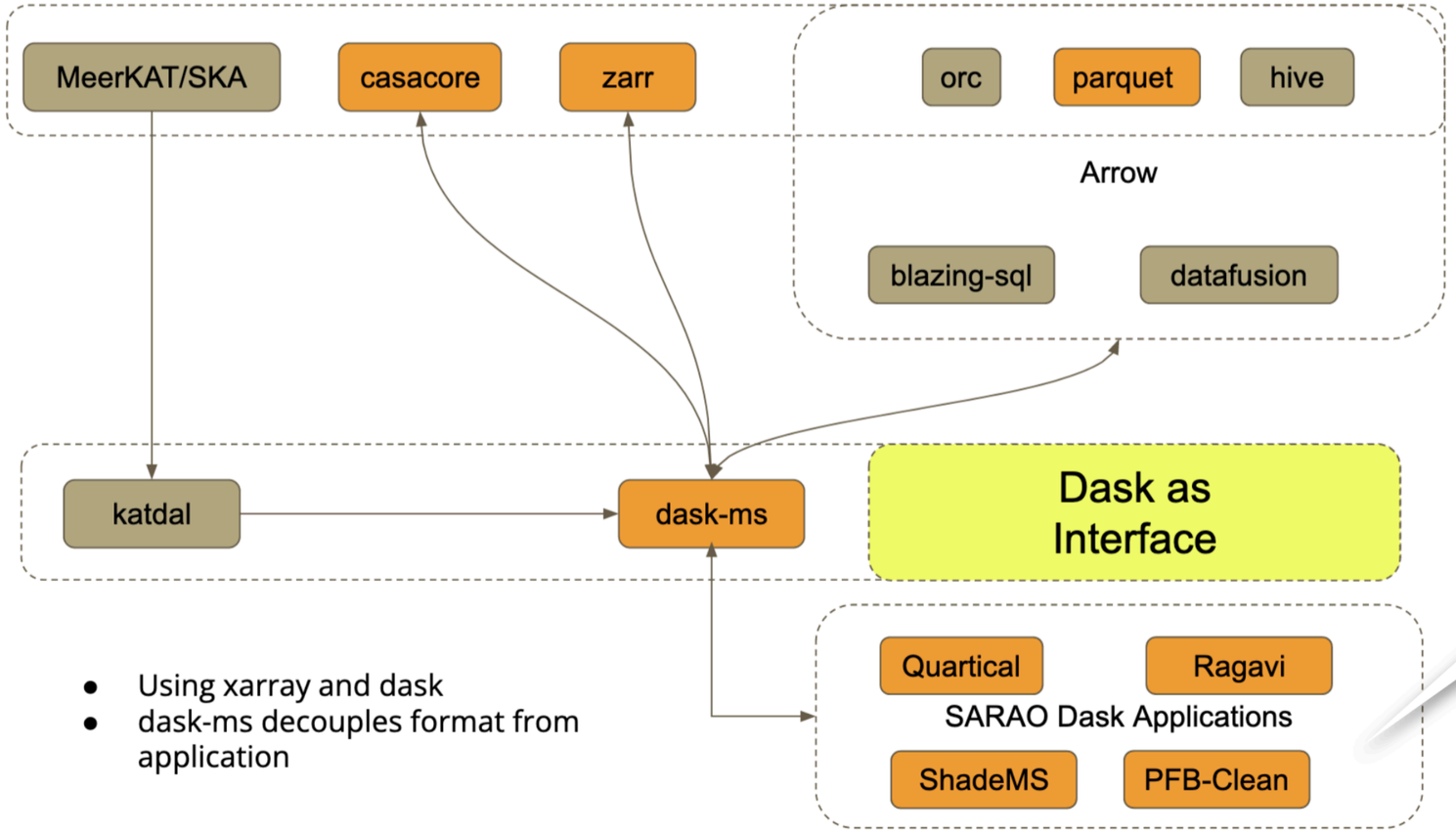
Out[4]:

UVW	FLAG	FLAG_CATEGORY	WEIGHT	SIGMA	ANTENNA1	ANTENNA2	AI
[-0 m, -0 m, -0 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	0	
[197.165 m, -71.1665 m, 239.711 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	1	
[64.0904 m, -45.403 m, 77.7948 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	2	
[181.196 m, 92.4503 m, 221.188 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	3	
[115.505 m, -71.0907 m, 140.264 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	4	
[212.315 m, -170.545 m, 257.6 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	5	
[166.535 m, -82.5291 m, 202.345 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	6	
[116.99 m, -2.04496 m, 142.462 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	7	
[93.6017 m, -167.846 m, 113.043 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	8	
[232.82 m, 18.2794 m, 283.639 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	9	
[205.18 m, -17.5161 m, 249.775 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	10	
[93.7238 m, -258.548 m, 112.68 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	11	
[89.6506 m, 41.9724 m, 109.416 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	12	
[78.9802 m, -5.47738 m, 96.1536 m]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	0	13	

Email: arpan.das@epfl.ch

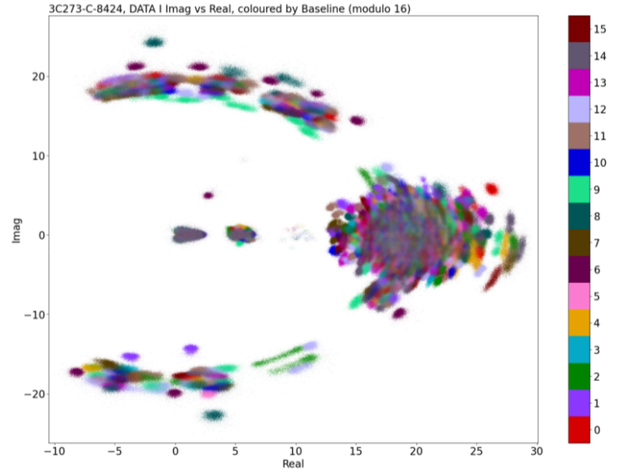
Dask-MS

Creates xarray Datasets from CASA Tables
https://github.com/ratt-ru/dask-ms



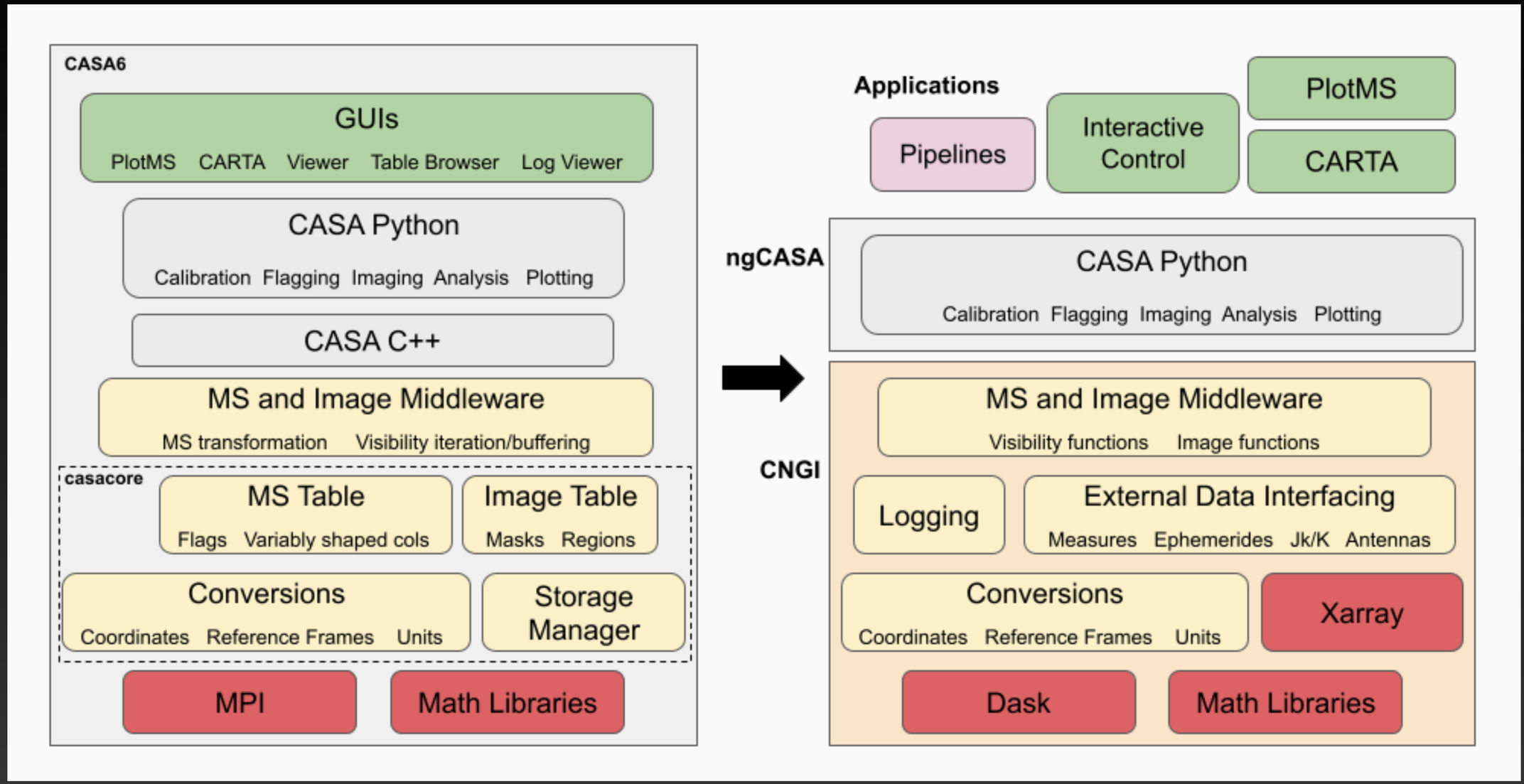
- Using xarray and dask
- dask-ms decouples format from application

ShadeMS plotting 25 million points in 3.5s



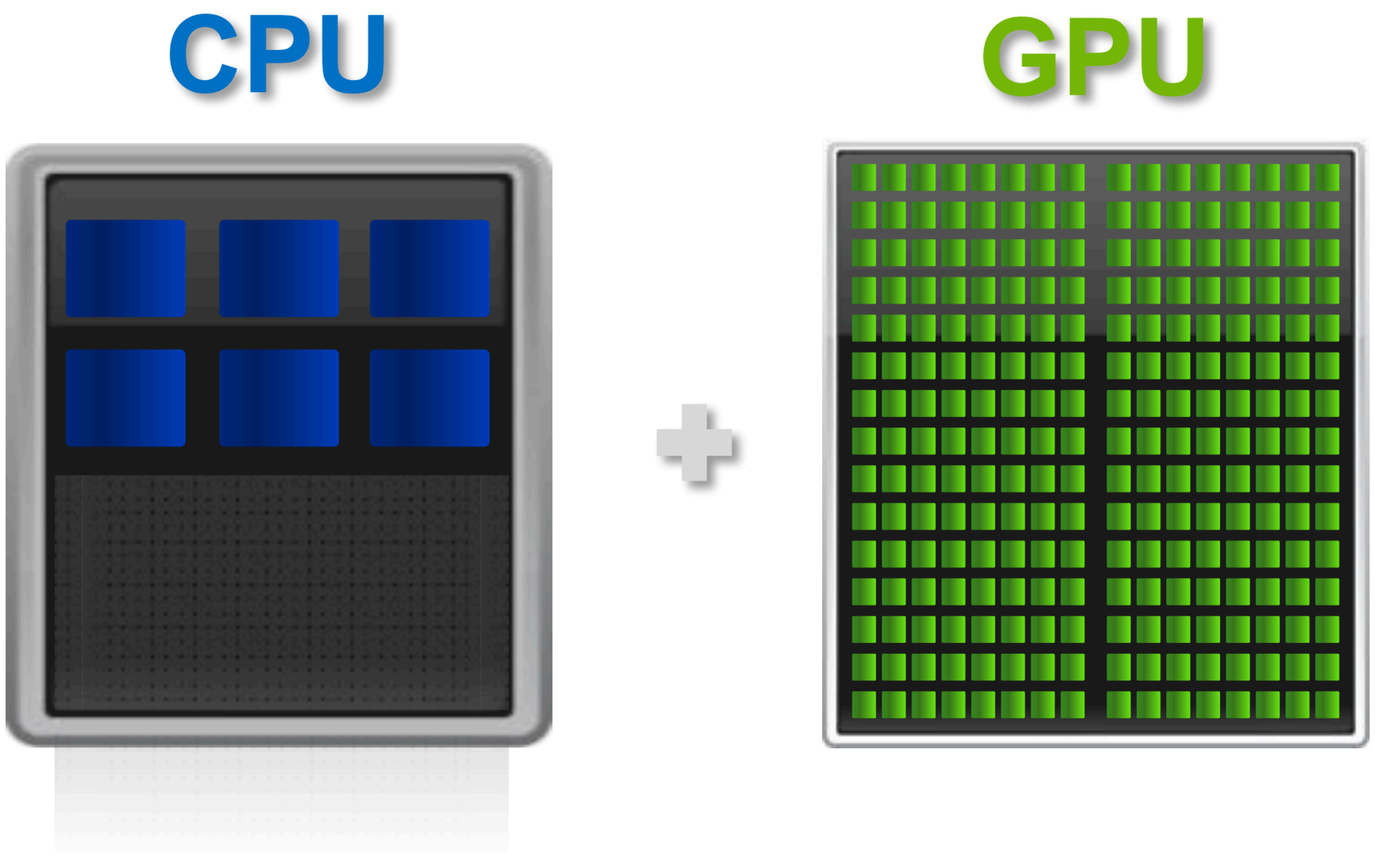
Email: arpan.das@epfl.ch

CASA Next Generation Infrastructure (CNGI)



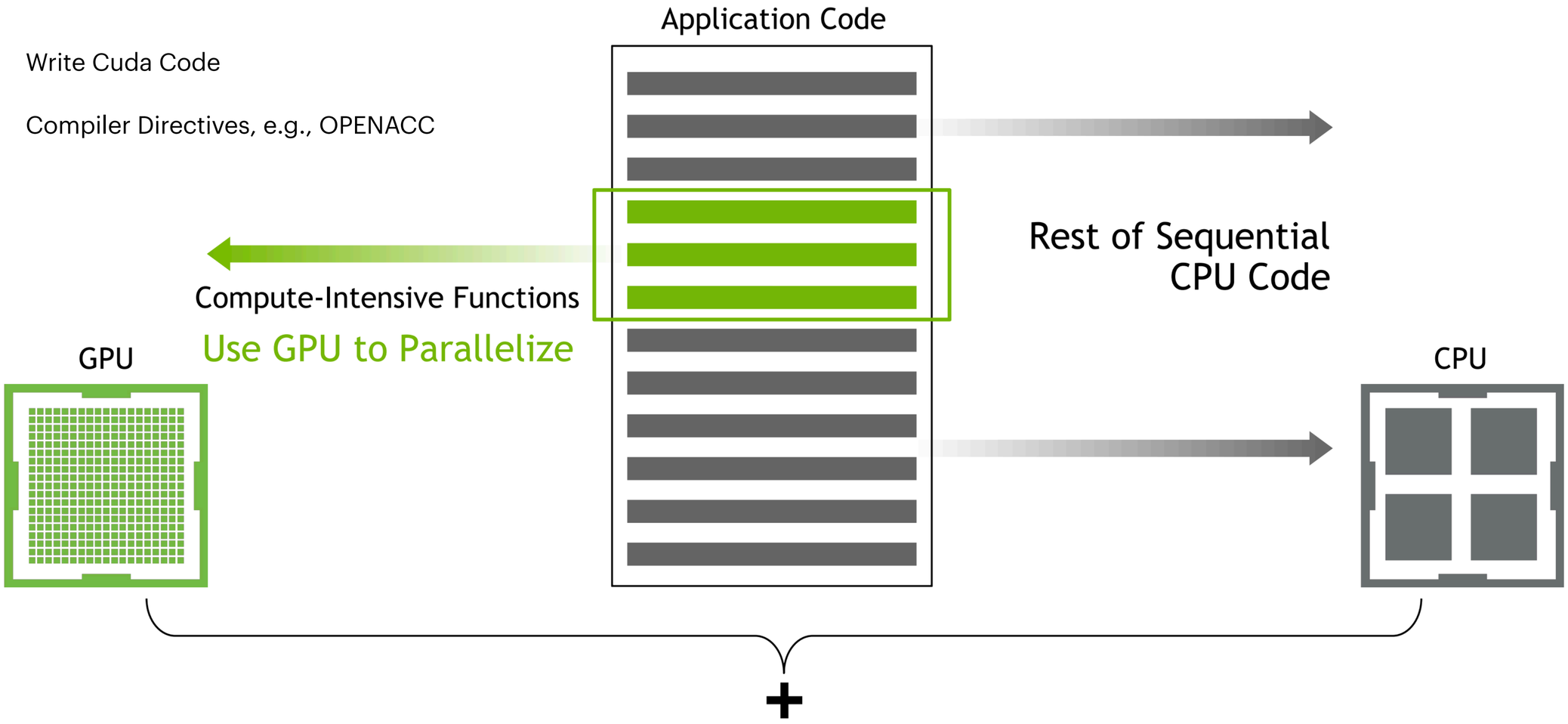
Email: arpan.das@epfl.ch

Add GPUs: Accelerate Science Applications



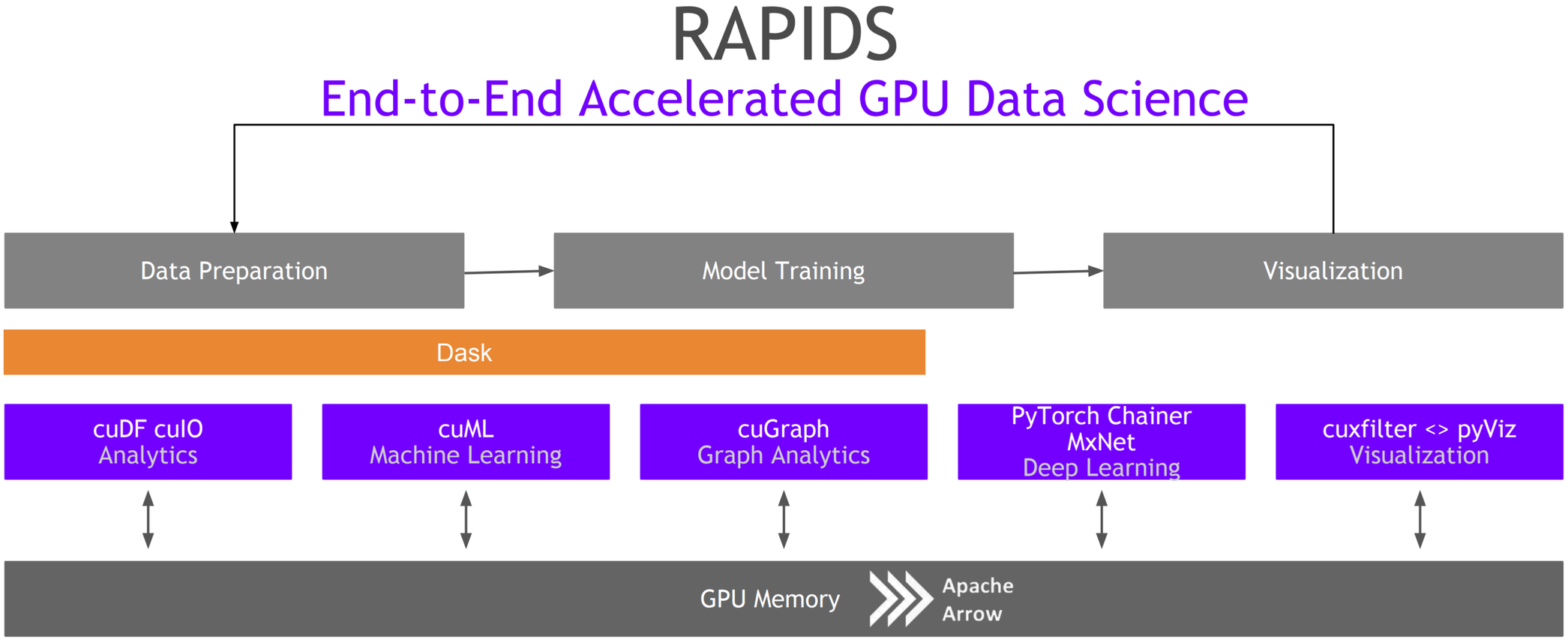
Email: arpan.das@epfl.ch

SMALL CHANGES, BIG SPEED-UP



Email: arpan.das@epfl.ch

Rapids.ai

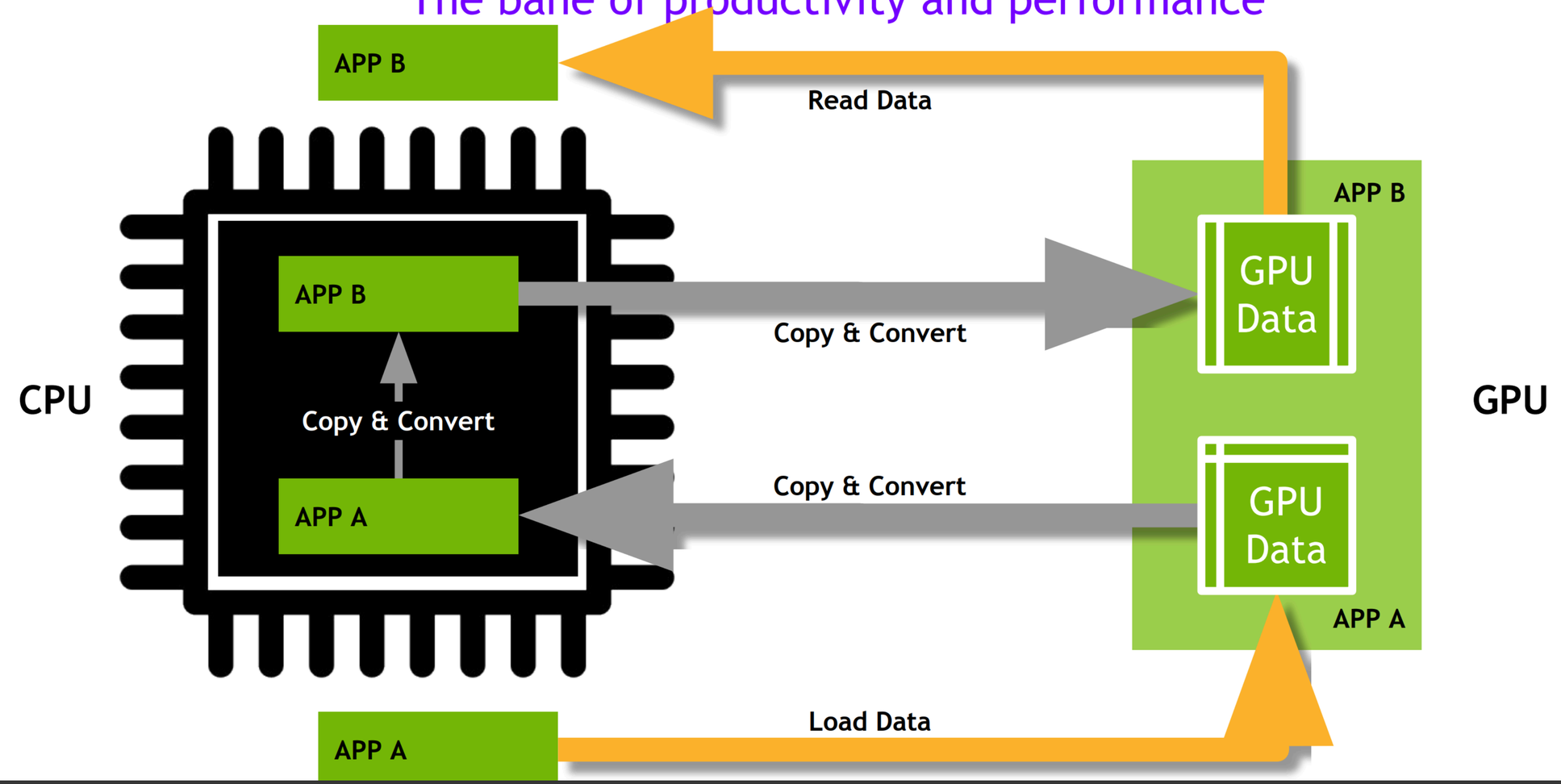


Even parts of the code which is not compute intensive can be benefitted significantly from parallelism

Email: arpan.das@epfl.ch

Data Movement and Transformation

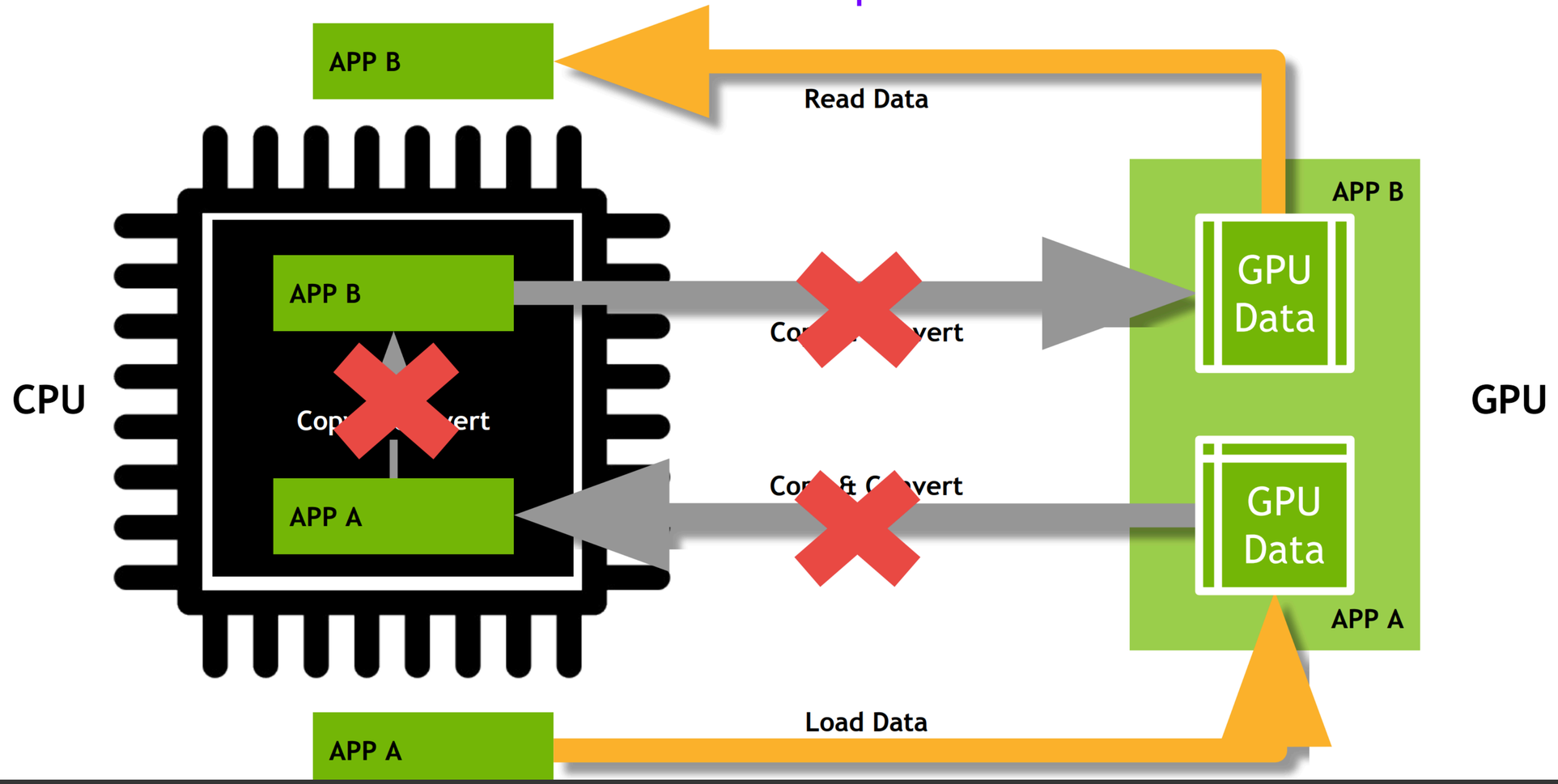
The bane of productivity and performance



Email: arpan.das@epfl.ch

Data Movement and Transformation

What if we could keep data on the GPU?



Email: arpan.das@epfl.ch

CASA MS > Cudf Parquet

UVW	FLAG	FLAG_CATEGORY	WEIGHT	SIGMA	ANTENNA1	ANTENNA2	ARRAY_ID	DATA_DESC_ID	EXPOSURE	...	STATE_ID	TIME
[111.25252215124442, 257.2411317946735, 136.93...	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	39	236	0	0	587.755102	...	-1	51544.275153
[113.42310656694917, 317.90052626415854, 139.9...	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	39	237	0	0	587.755102	...	-1	51544.275153
[86.28487461816599, 302.7320017783404, 106.789...	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	39	238	0	0	587.755102	...	-1	51544.275153
[60.85460025515545, 254.88031233688253, 75.549...	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	39	239	0	0	587.755102	...	-1	51544.275153
[93.33554936019846, 366.7074918878279, 115.737...	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	39	240	0	0	587.755102	...	-1	51544.275153
...	...	...	...	...	...	...	...	...	...	...	...	...
[867.8136446410942, -4974.731727750057, -1084....	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	366	376	0	0	587.755102	...	-1	51544.608486
[10626.71094613529, -45558.48213531486, -13198...	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	365	510	0	0	587.755102	...	-1	51544.608486
[-0.0, -0.0, -0.0]	[[False, False, False, False]]	[[[False, False, False, False]]]	[1.0, 1.0, 1.0, 1.0]	[9999.0, 9999.0, 9999.0, 9999.0]	366	366	0	0	587.755102	...	-1	51544.608486

Email: arpan.das@epfl.ch

Future goals:

- Develop unit tests, extensive validation, and release of a library that performs the conversion
- Write documentation and release code library
- Integrate to existing simulations and software packages for data analysis

Email: arpan.das@epfl.ch

Questions?