Application of Cloud Computing to fast and heavy scientific simulations

Marco Landoni

INAF – Osservatorio Astronomico di Brera

marco.landoni@brera.inaf.it

Matteo Genoni, Marco Riva, Andrea Bianco
INAF – Osservatorio Astronomico di Brera

Outline

- Introduction
- The cloud computing paradigm
- Application to ESA-IXPE (high energy) and ESO-HIRES (optical)
- Outlook and conclusion

Warning!

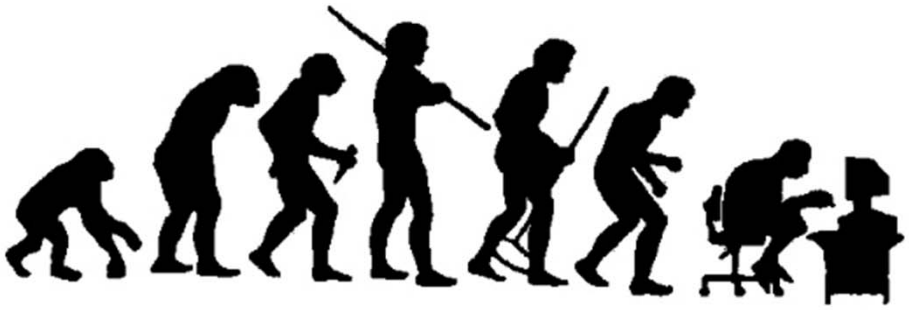
- **One size does not fit all.**

- Each problem has its **proper computational model** and architecture that maximise the return, the **cost efficiency** and the proper **exploitation of common resources**.

- A rule of thumb:
  - Long N-Body cosmological simulation, high energy gamma-ray showers track: **HPC**
  - Deep simulation (unfrequently) for instrumentation that needs answer **now** with **unpredictable schedule** or moderate CPU demanding scientific simulation: **Cloud Computing**

Cloud computing paradigm

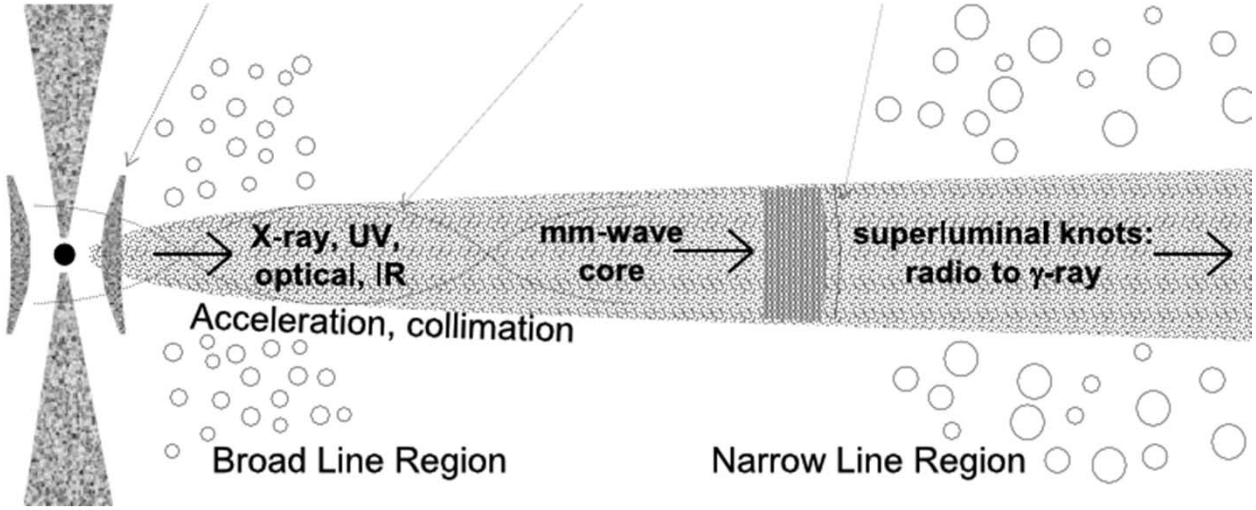
- **Cloud Computing** is a style of **computing paradigm** in which typically real-time scalable resources can be accessible via Internet to users. Pay as-you-go for resource utilisation. (Wikipedia)

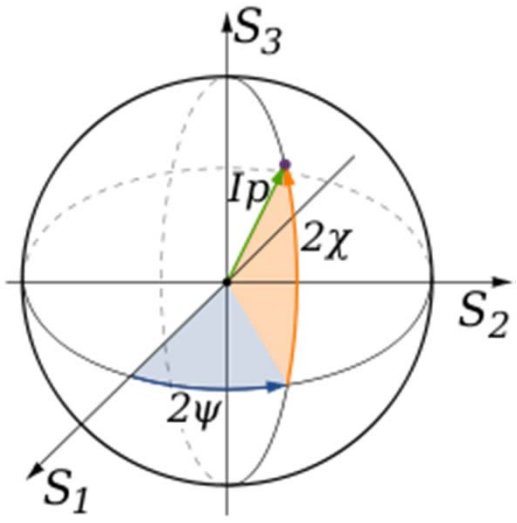


Various providers ....

Google cloud platform applied to International X-Ray Polarimeter Explorer (IXPE)

- **The idea:** Model the jet with “hyperfine cells” that emits synchrotron radiation (polarised) to test physical state of the relativist jet in blazar. Make use of Stokes parameter for each “cell” and, a posteriori, sum up to detect the degree of polarisation.





Tavecchio & Landoni 2017

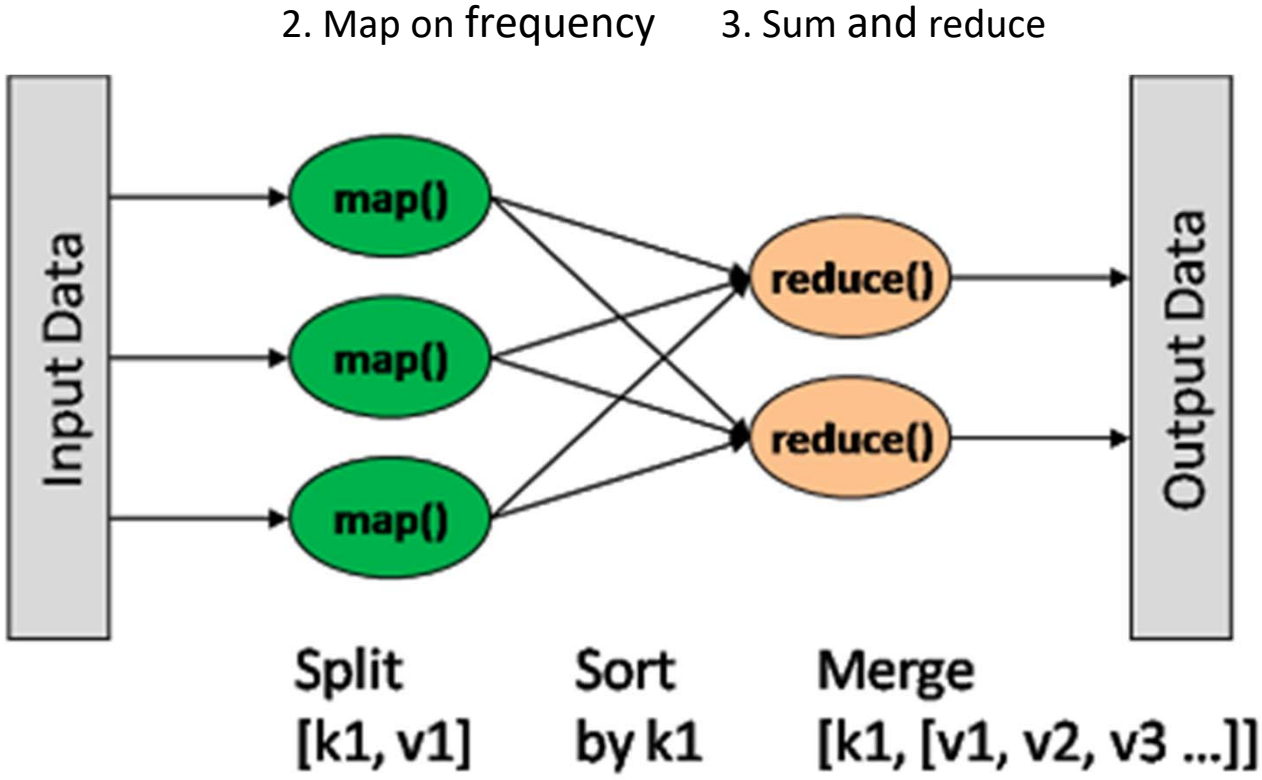
The numbers….

- TB of raw data (binary) processed to simulate the whole jet.
- Use of Google resources for a total of dex kCPU/hours
- **Why** Google Cloud Platform: off-the-shelf Map Reduce database, superior analytic suite for Big Data with access time of minutes.

Tavecchio & Landoni 2017

The Map Reduce on a noSQL database

1. Single cell Stokes parameter



4. Polarisation degree for each frequency

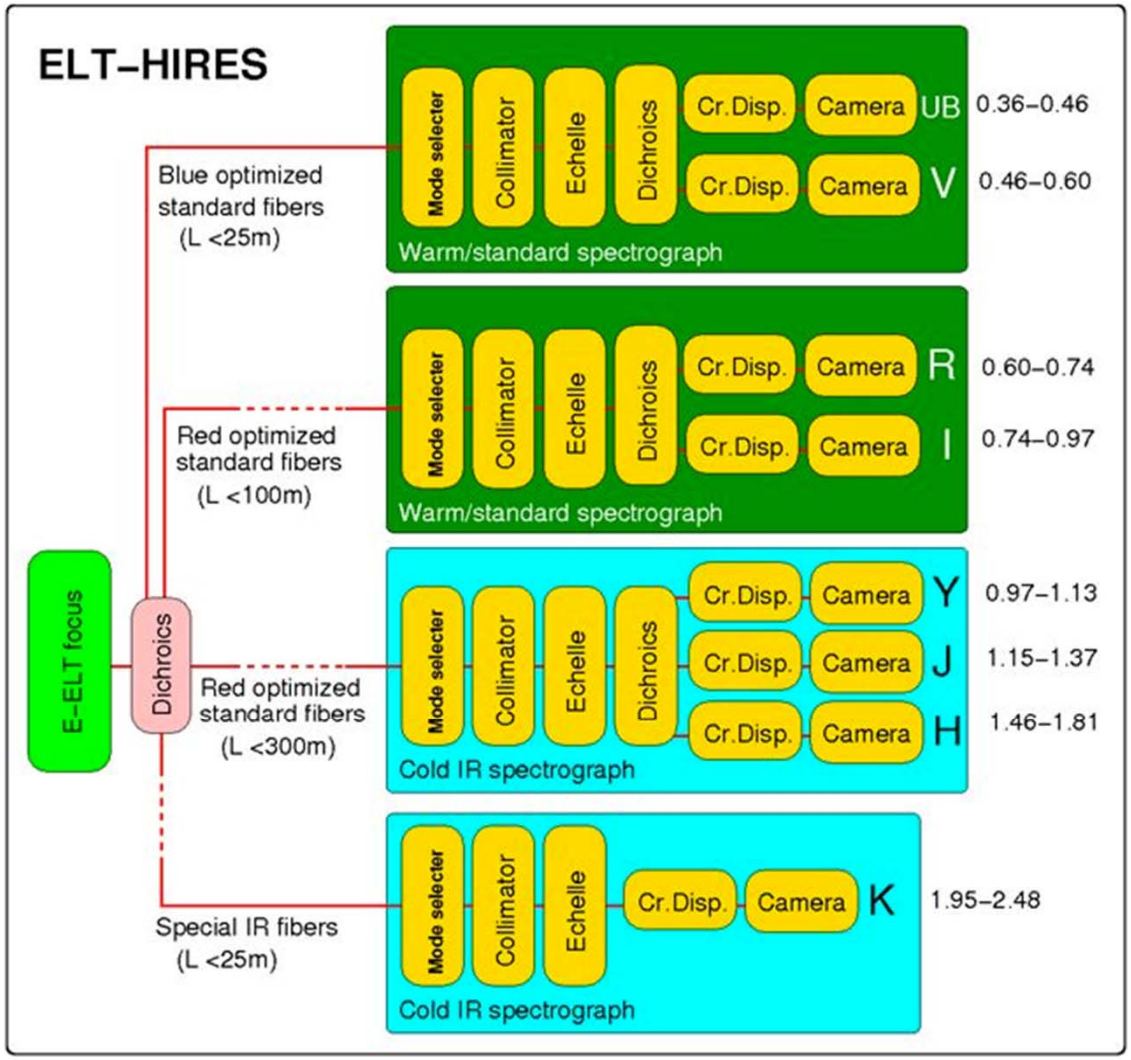
Google Compute engine + Big Query

- Data is produced with Google shared tenancy.
- When simulation data are available a cloud function starts a Big Query (MapReduce) that compute the final degree of polarisation by processing various TB of data.
- Total cost of about **300 EUR** for the whole simulation. Paper in preparation (submission expected this month).

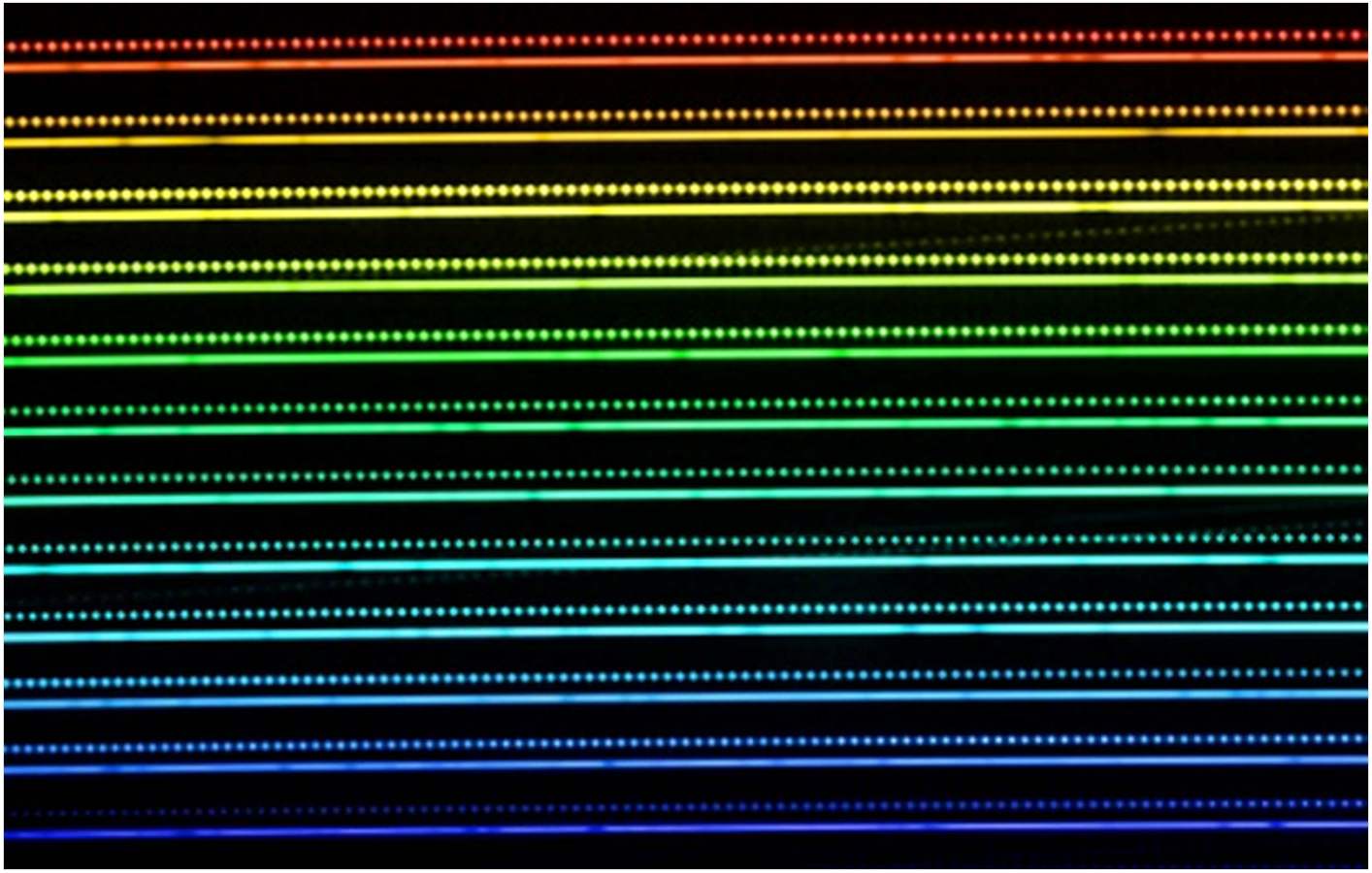
Google BigQuery

Tavecchio & Landoni 2017

E-ELT-HIRES possible scheme of instrument modules

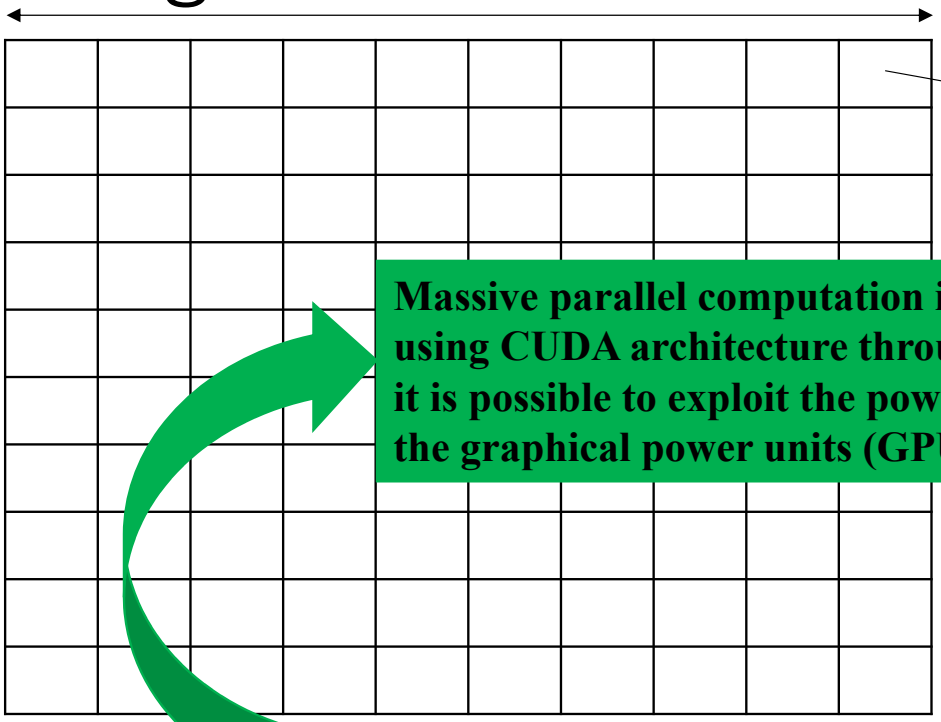


The problem: render echellograms with 1 cm/s precision (fraction of nm)



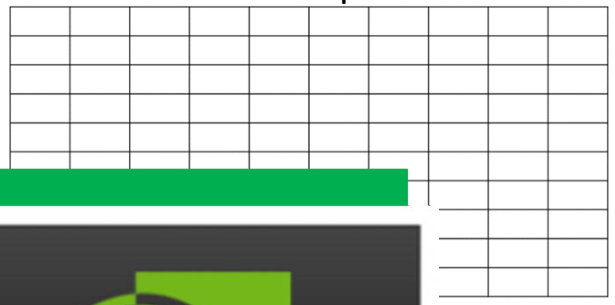
ESO-HARPS + LFC
Raw frame data

Single resolution element…



Single pixel oversampling

10 Sub-pixels

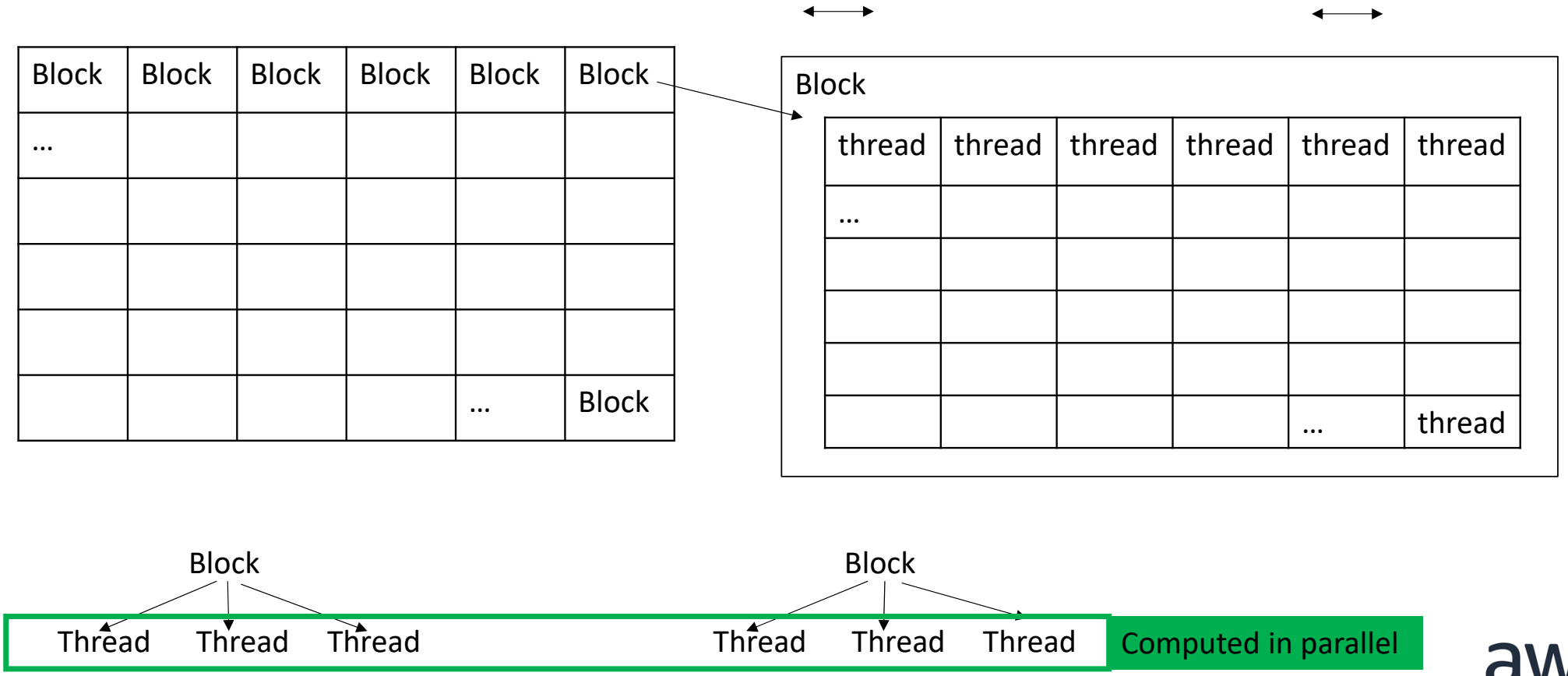


Massive parallel computation implementation using CUDA architecture through which it is possible to exploit the power of the graphical power units (GPU)

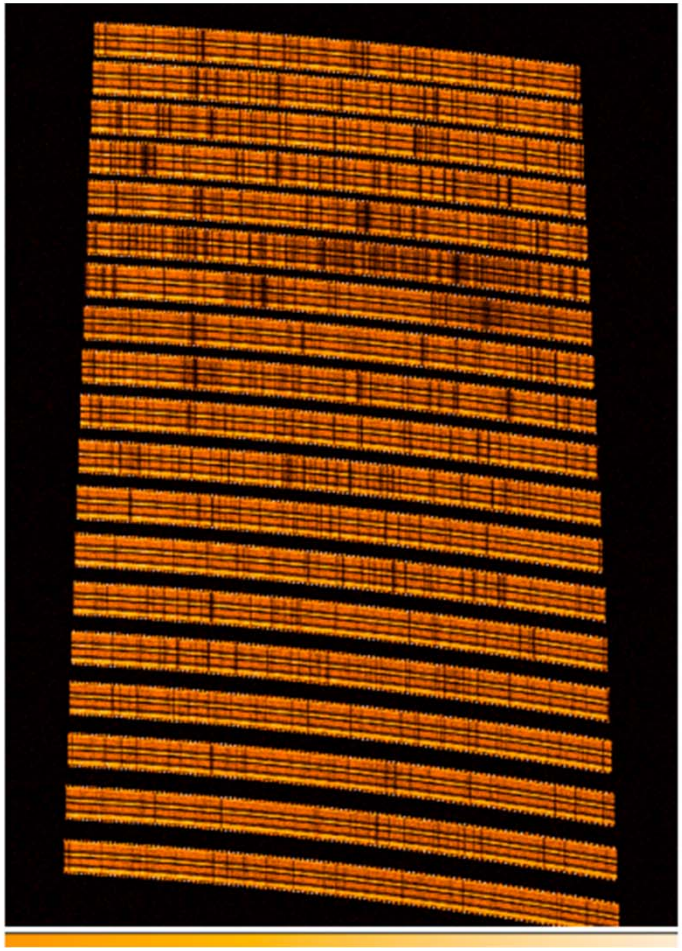
The required co…on time for a single fiber SRE is ≅ 60 s (on a Xeon
This means that for 70 fiber per SRE, about 1000 SRE per each order and 30 order in an echellogram the total time
Required to generate a systhetic echellogram would be = 1.3 10^8 s = 1450 days

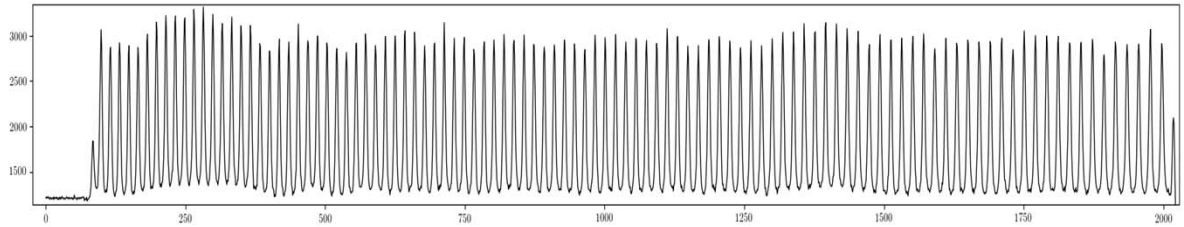
Genoni & Landoni 2016
Genoni & Landoni 2017
Genoni et al 2018 in preparation

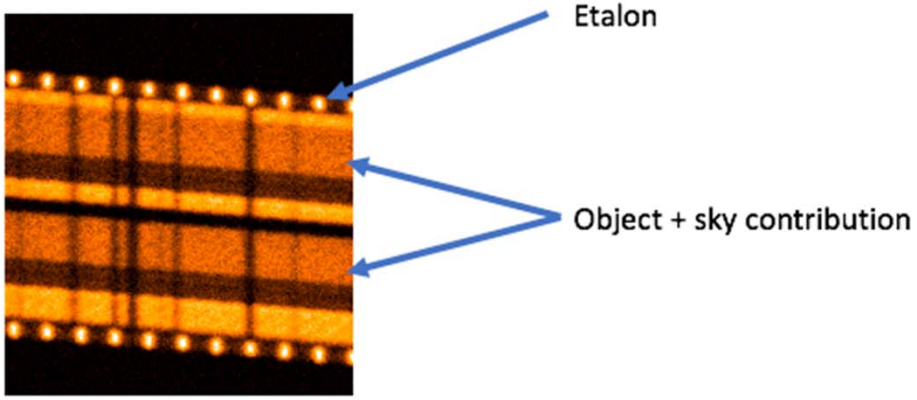
Use of MASSIVE parallelism with very simple CPU

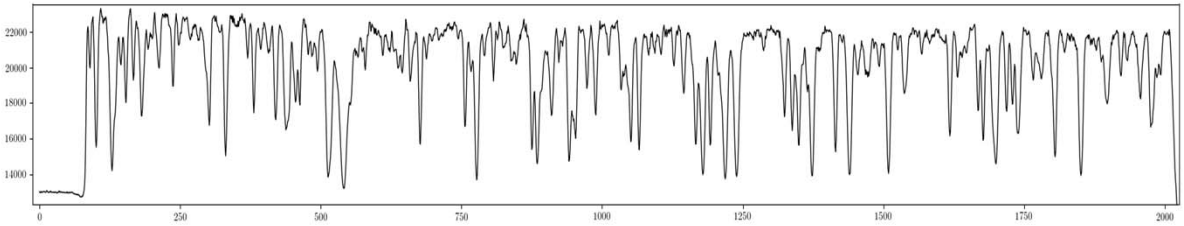


Result




3000
2500
2000
1500
0
250
500
750
1000
1250
1500
1750
2000


Etalon
Object + sky contribution


22000
20000
18000
16000
14000
0
250
500
750
1000
1250
1500
1750
2000

Few numbers

- Various TB of data produced and reduced many times (for different spectral distribution / experiment) to obtain the final echellogram.
- We used GPU-CUDA powered by AWS that provides about 62.000 cores (GPU) used for about 1 days of simulation (1.500.000 GPUcore/hour).
- Data used for a large number of simulations (re-usable).
- Used of shared tenancy. We spent for the whole project about 400 EUR.

Conclusion

- New paradigms based on cheap solution **are available** for common scientific simulation that could not be addressed with on site facility and are not complex enough / designed for / compliance with large cluster and HPC facilities.

- **Remember**: HPC, GRID, Cloud Computing, CUDA, ecc. are computation paradigm exactly as programming paradigm. Each problem has its own solution. **One size does not fit all.**

- **It is fun !**

Thank you!