Green Flash

High performance computing for real-time science

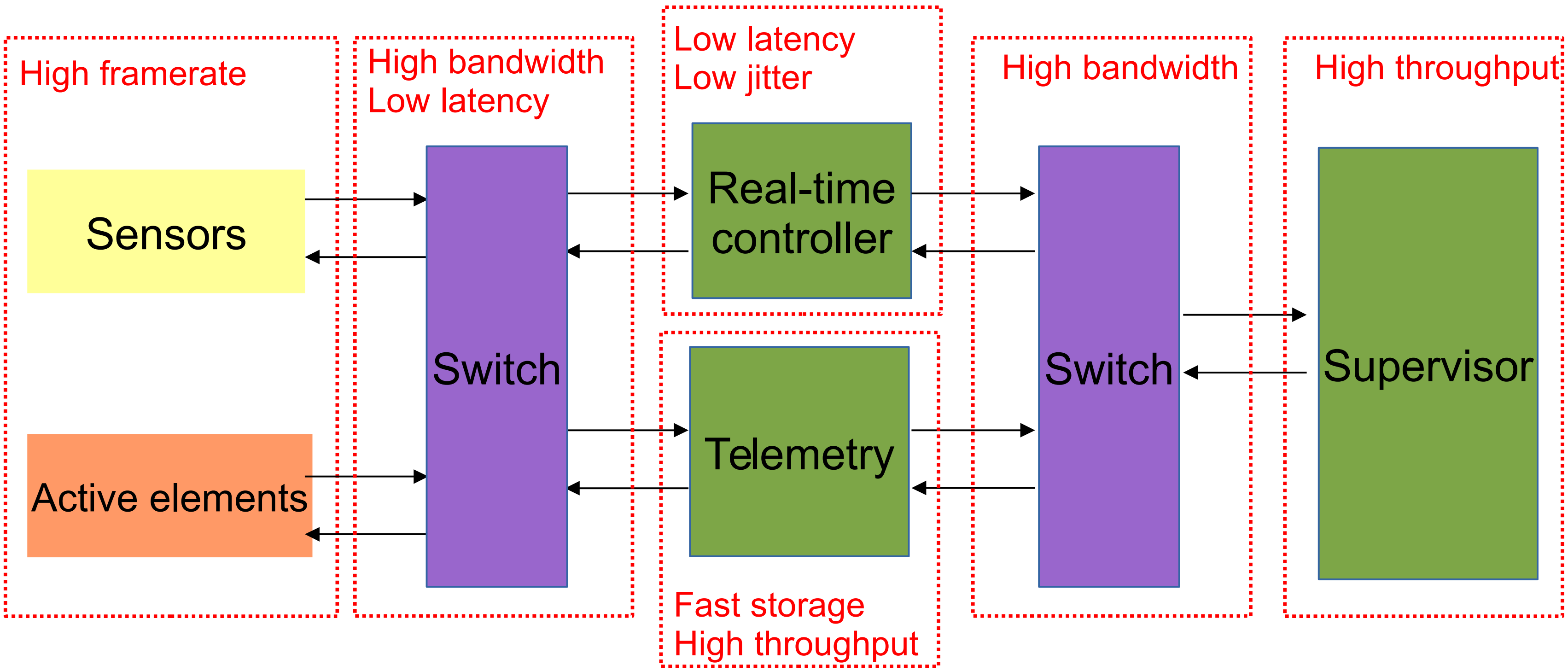
Contribution from Observatoire de Paris on WP 4 and 6

Project #671662 funded by European Commission under program H2020-EU.1.2.2 coordinated in H2020-FETHPC-2014

WP 4 : Accelerators for real-time HPC

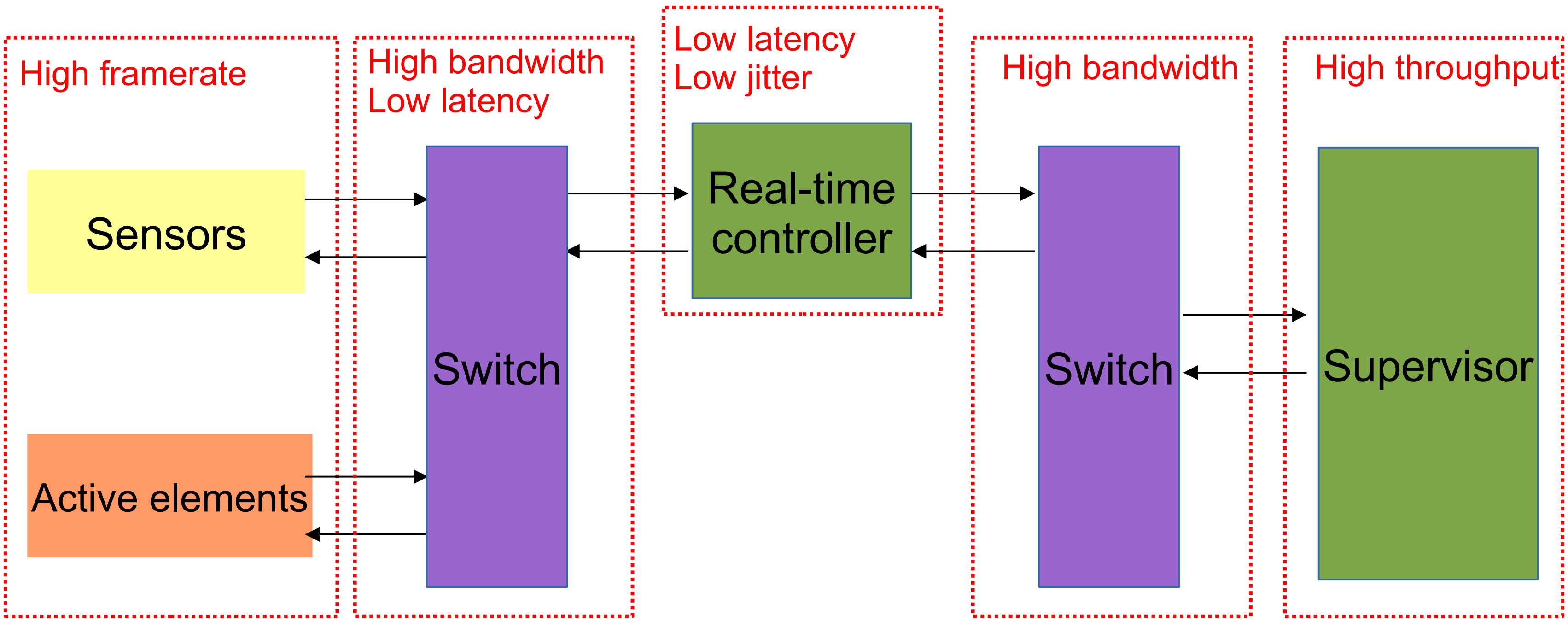
- Assess various HW accelerator options on a real-time application
  - GPU : lead by OdP with contribution from UoD
  - Xeon Phi : lead by UoD
  - FPGA : lead by UoD with contribution from OdP
- Assess performance of same hardware on complex data pipeline
  - Supervisor module for AO : lead by OdP
  - Criterion optimization and large matrix inversion

WP 4



High framerate
Sensors
Active elements
High bandwidth
Low latency
Switch
Low latency
Low jitter
Real-time controller
Telemetry
Fast storage
High throughput
High bandwidth
Switch
High throughput
Supervisor

WP 4


High framerate
Sensors
Active elements
High bandwidth
Low latency
Switch
Low latency
Low jitter
Real-time controller
High bandwidth
Switch
High throughput
Supervisor

WP 4

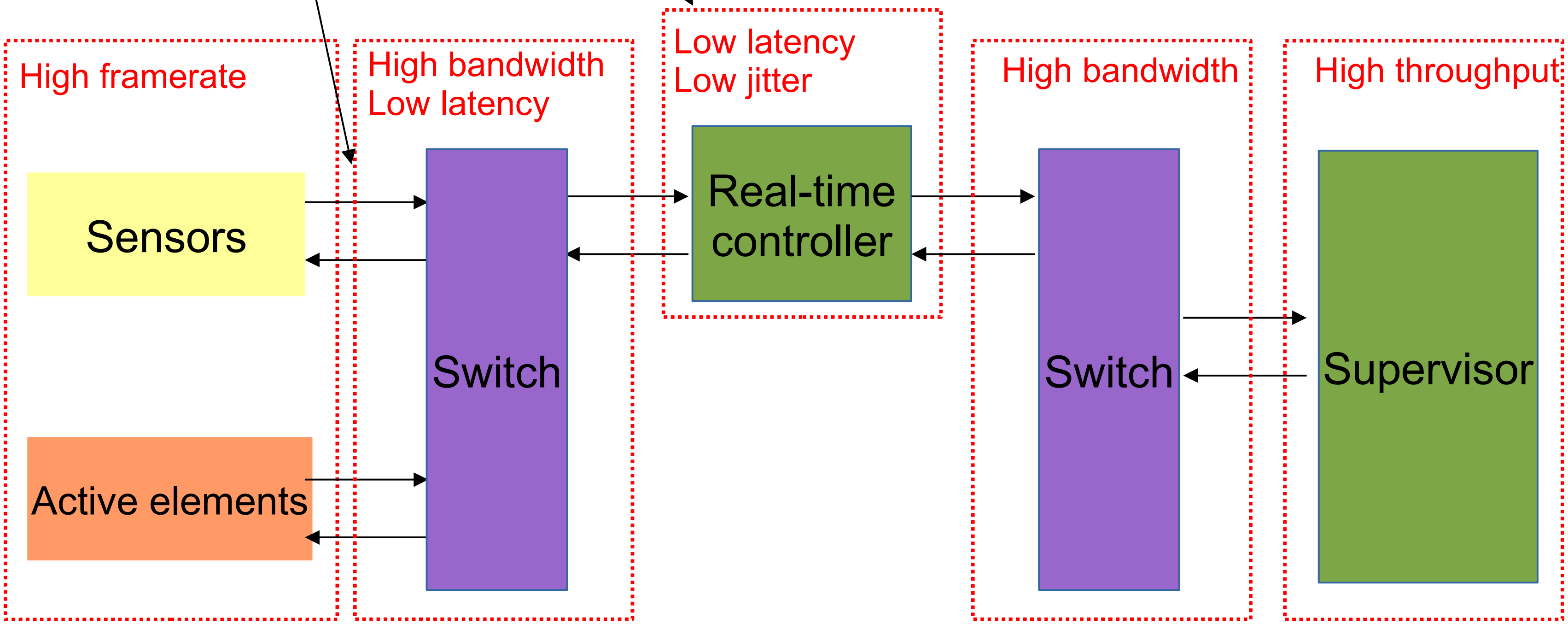
Real-time pipeline.
Includes sensors pixels streams processing and MVM for control of active elements

Pixel Streams processing, 2 options:
* tens of GFLOPS in simple arithmetics or
* hundreds of GFLOPS in batched Fourier Transform

MVM : up to 5 TFLOP/s (2.5 TMAC/s)

Up to 250 Gb/s of streaming data

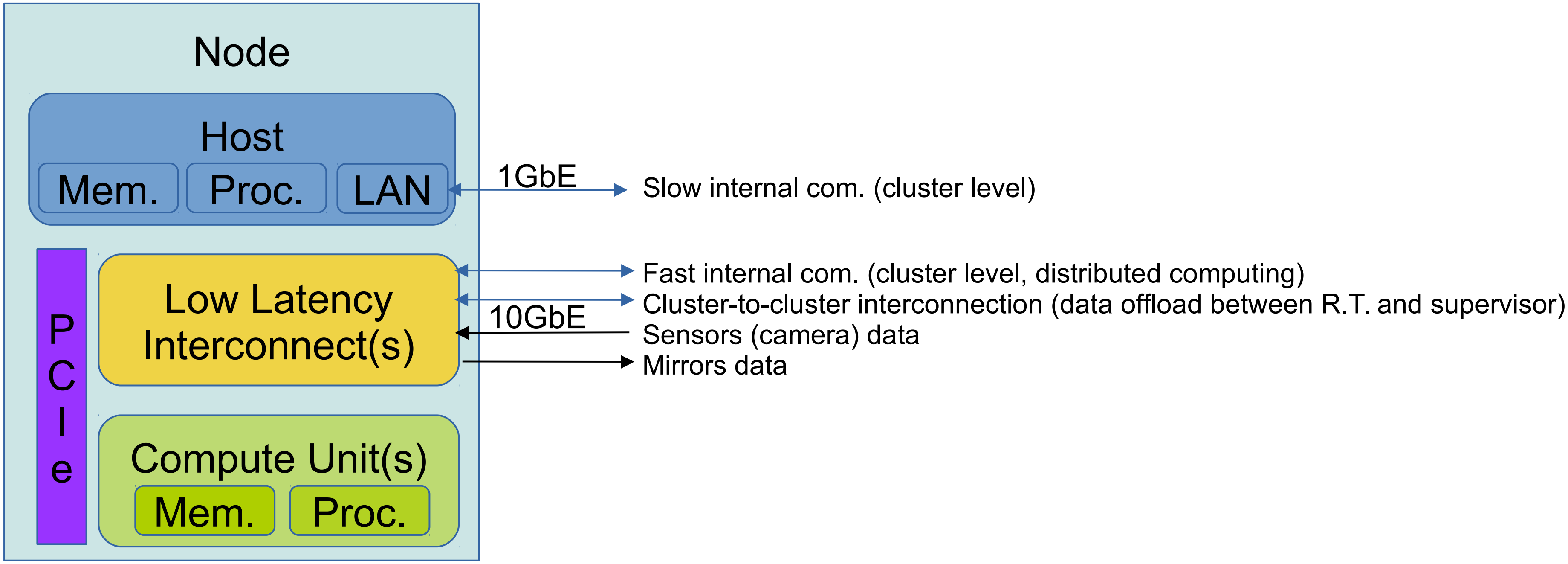
Performance must be deterministic, max latency : 2ms



Generic platform based on accelerators

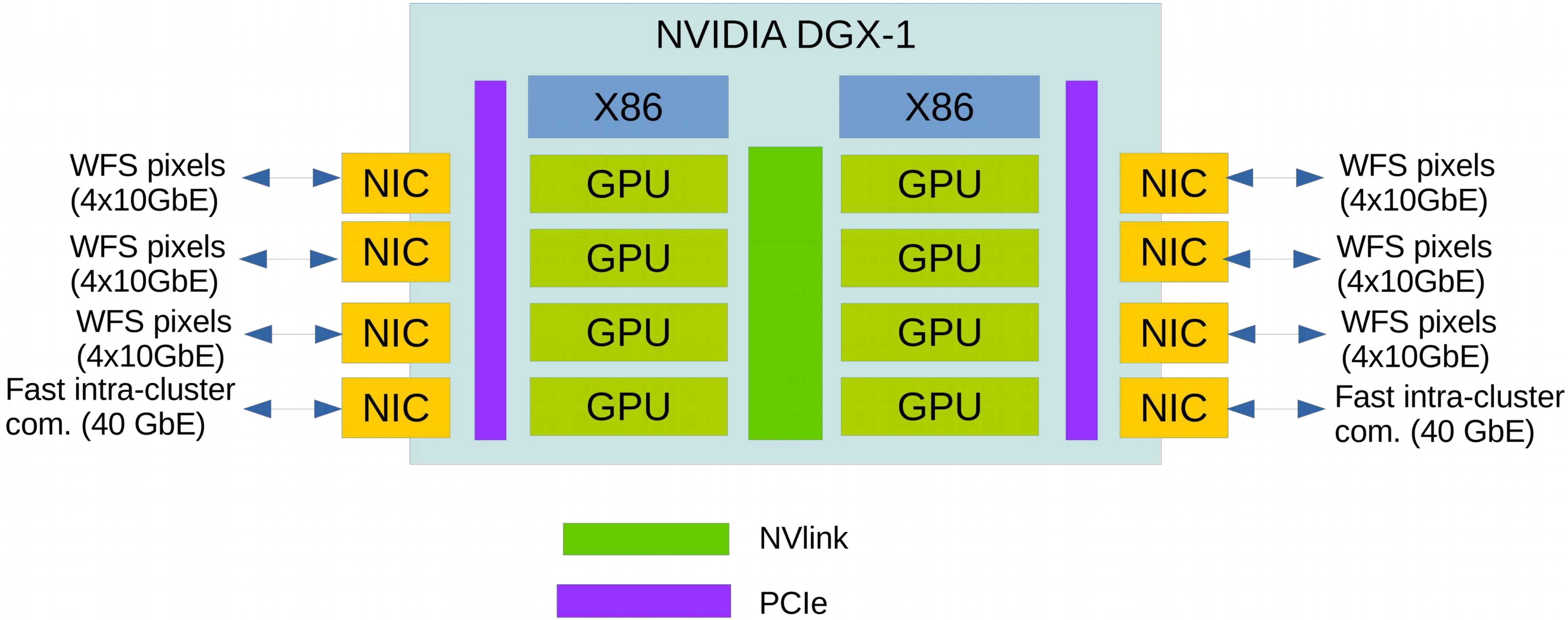
One generic node architecture, two applications :
- Real-time memory bound linear algebra (AO linear control, a.k.a. real-time pipeline)
- High throughput compute bound linear algebra (AO supervisory tasks, a.k.a. supervisor)

For each application, nodes are interconnected into a cluster.
For the full featured prototype, the two clusters are interconnected



RT data pipeline with GPUs

- Prototype using latest generation GPU cluster



- Concept studied at LESIA

System dimensioning

MCAO @ E-ELT scale

- POLC control scheme + LGS WFS : 2.5 TMAC/s with 250 Gb/s of streaming data
- Upper limit from instruments specification capture during PDR (actual first light instruments may require less)

Memory bandwidth

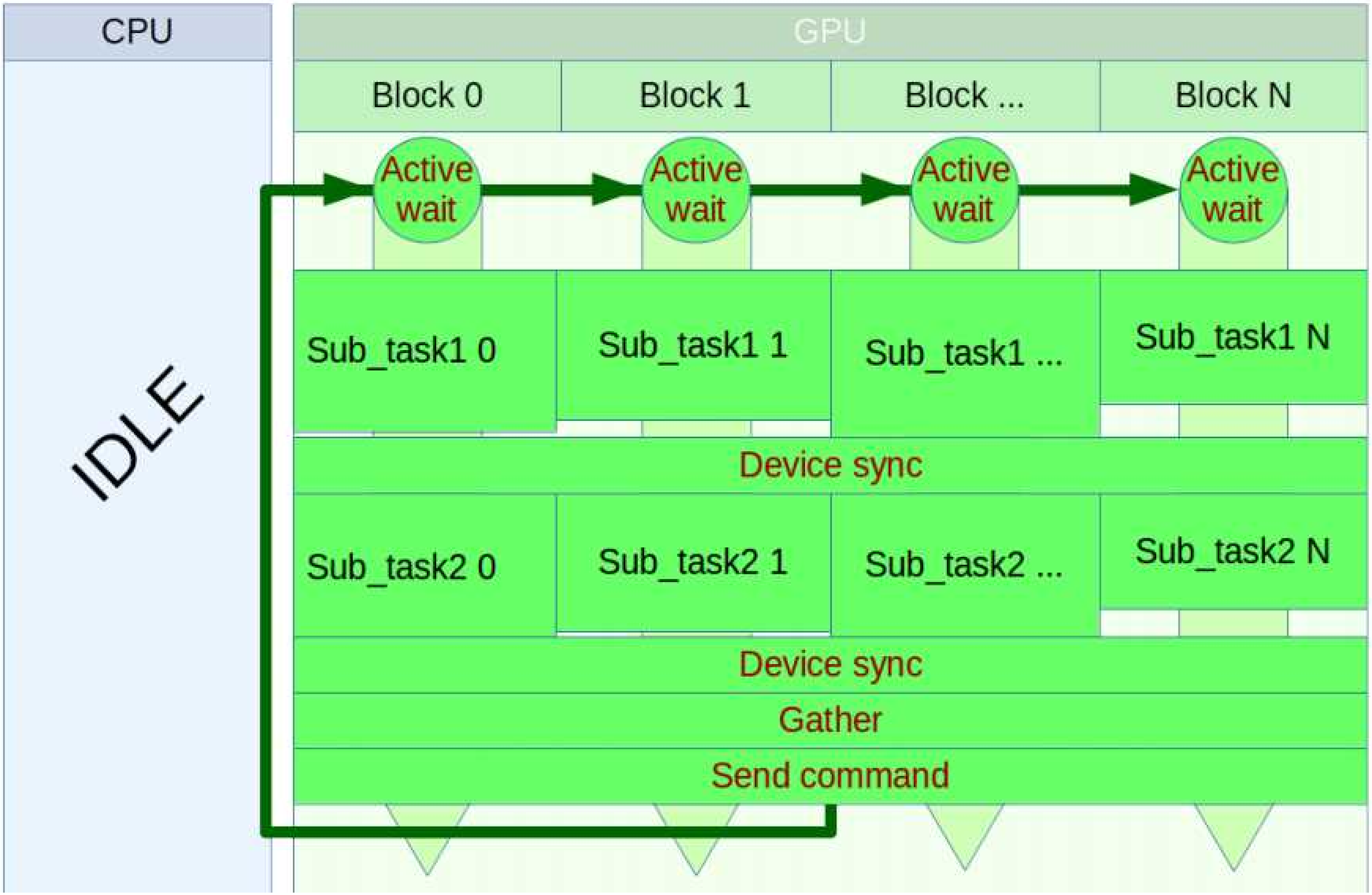
	K20C	K40	K80	P100
B_{theo}	208	288	240 (x2)	732
$B_{no\ ECC}$	175 (84%)	236 (82%)	200 (x2, 83%)	460 (62%)
B_{ECC}	150 (72%)	208 (72%)	173 (x2, 72%)	460 (62%)

Number of GPUs required

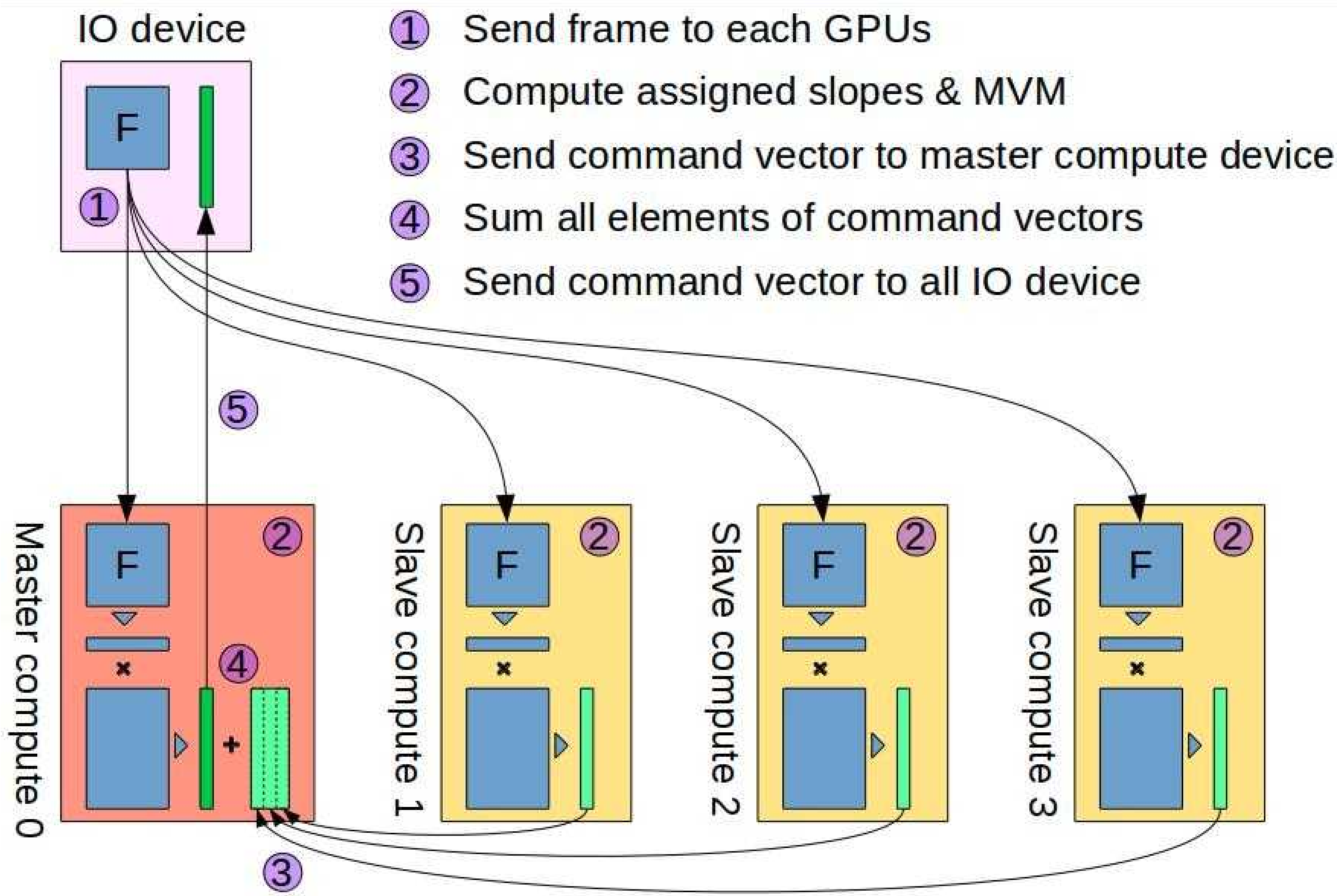
ECC	K20C	K40	K80	P100
Off	12	9	6	5
On	14	10	6	5

l'Observatoire de Paris — LESIA
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique
Durham University
MICROGATE
PLDA

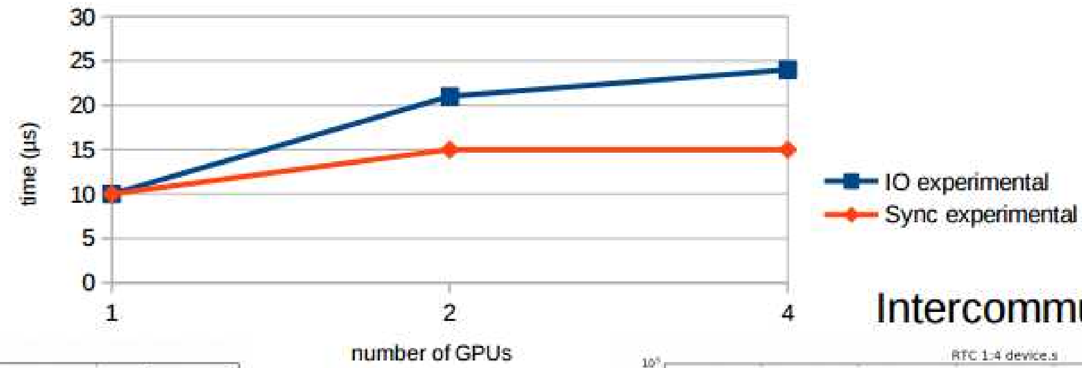
Persistent kernel implementation



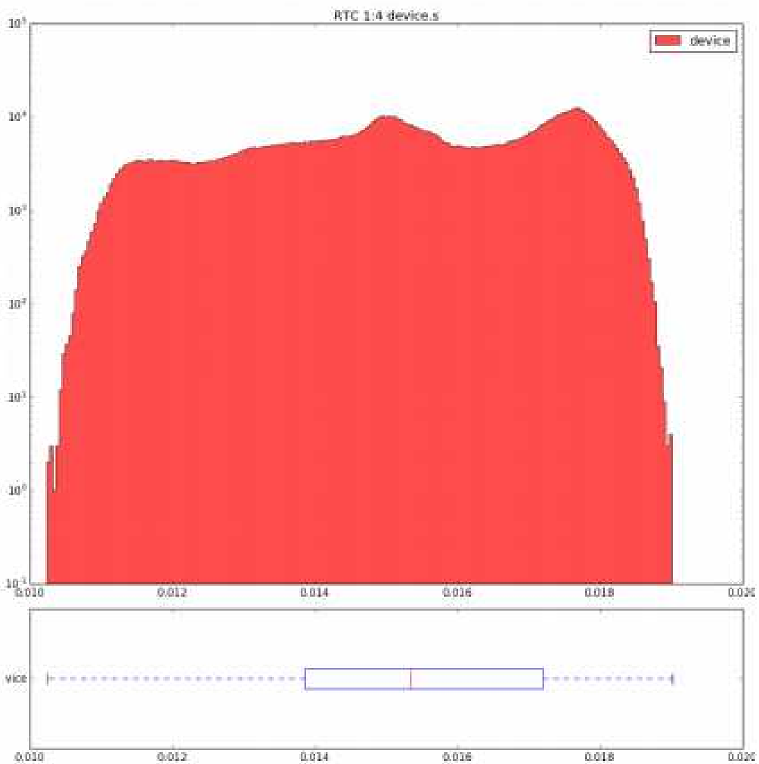
Multi-GPU prototype



Persistent kernel implementation

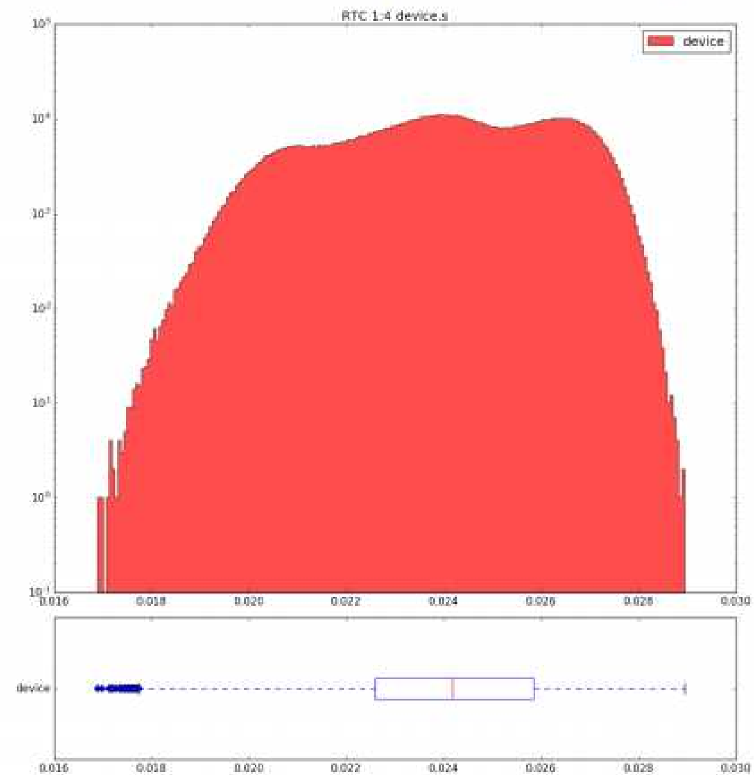


Synchronize jitter



Average : 15µs Jitter : 8.8µs

Intercommunication jitter

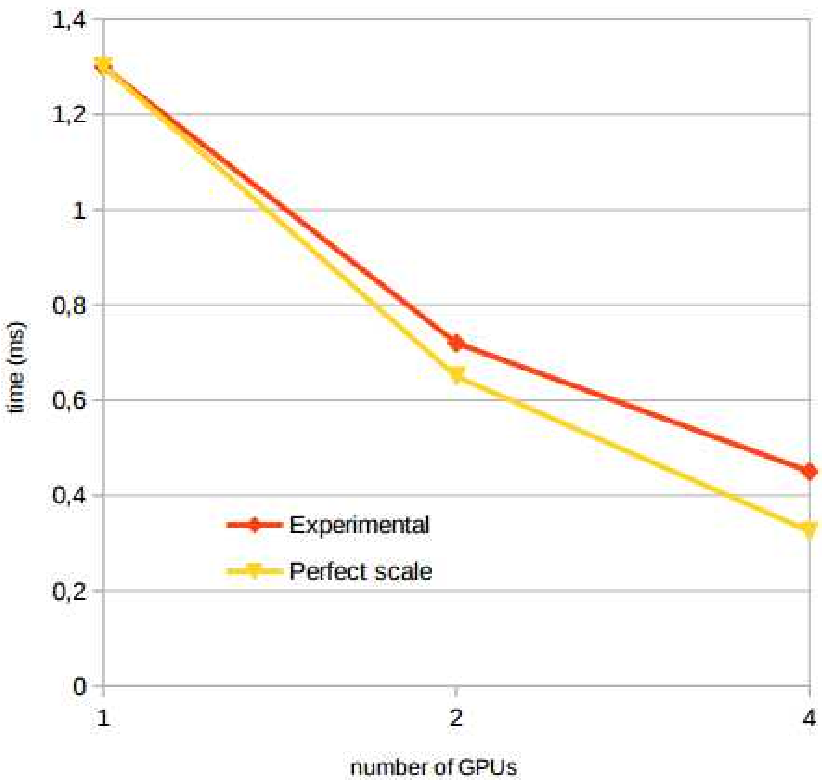


Average : 24µs Jitter : 12µs

Persistent kernel implementation

Strong scalability

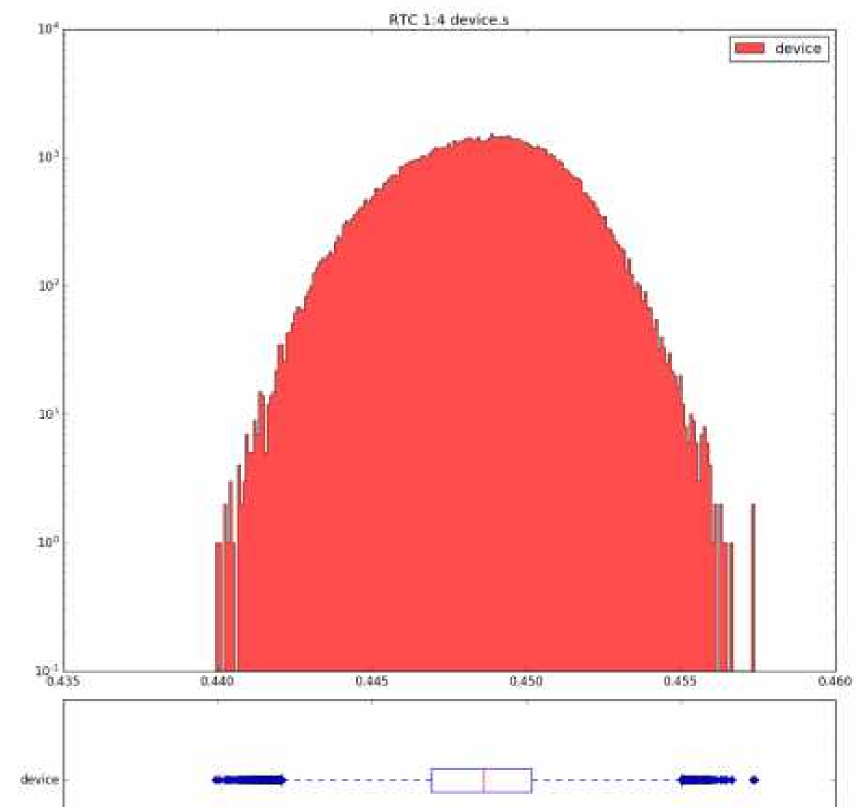
Constant case with **10,048 slopes x 15,000 commands**



Histogram

Case with **10,048 slopes x 15,000 commands** on 4 devices

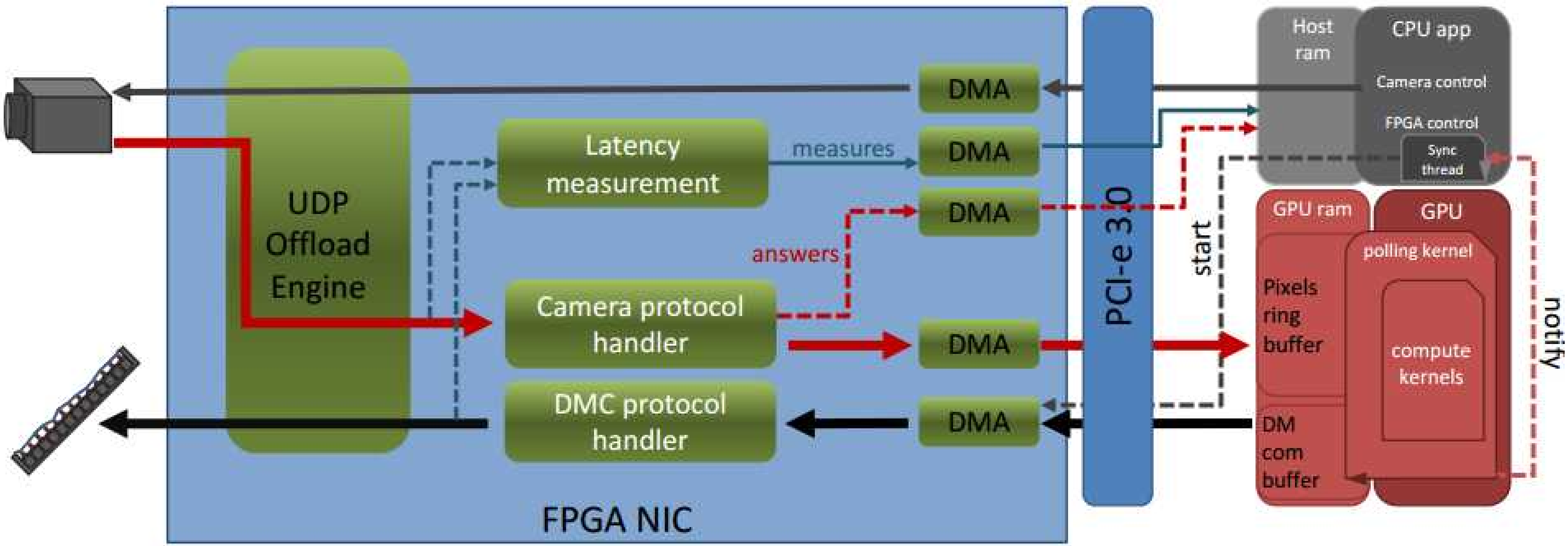
Average : 0.45ms   Jitter : 17µs



Data acquisition

FPGA writes/reads directly to/from GPU memory

Using only writes would be better though

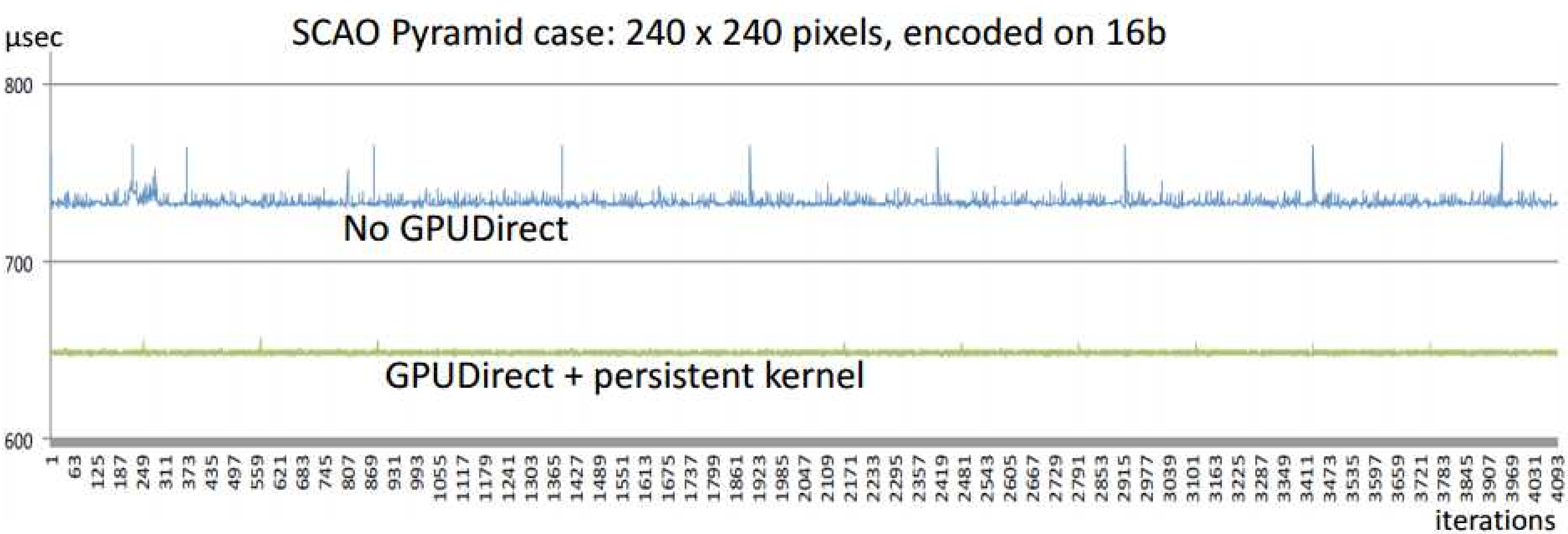


Green Flash

Data acquisition + persistent kernels

FPGA PLDA XPressG5
GPU Tesla C2070
OS Debian wheezy

Camera EVT HS-2000M
10GbE network

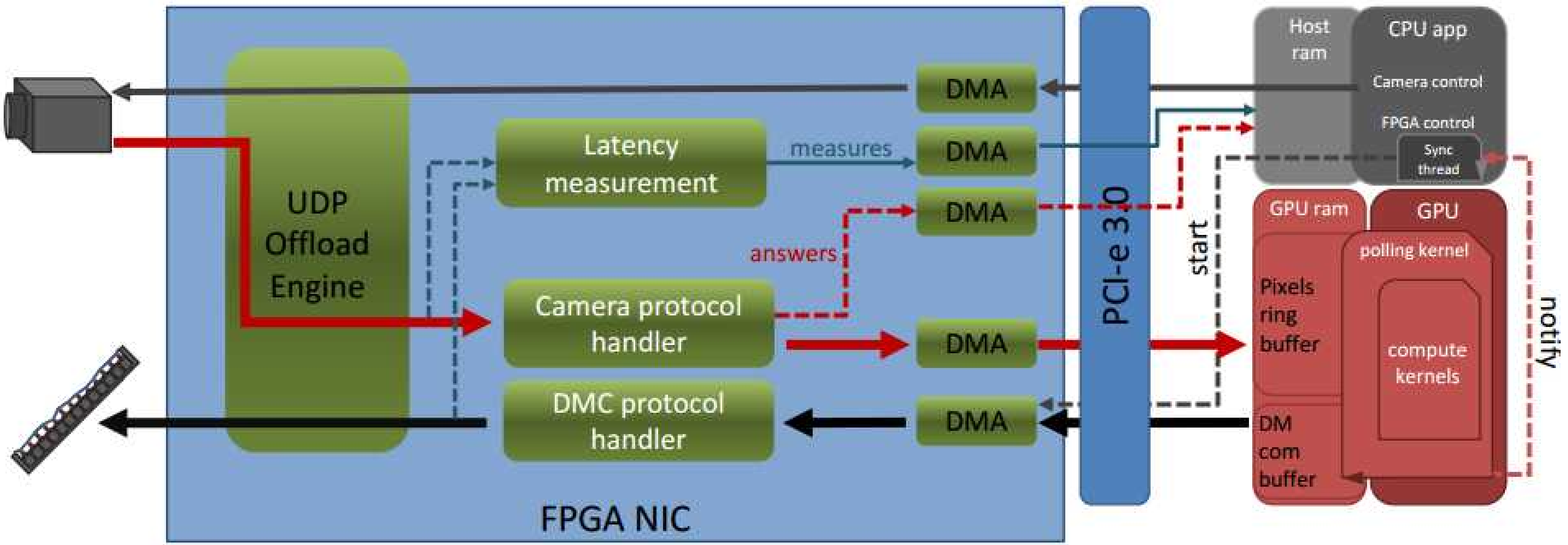


l'Observatoire de Paris — LESIA
Durham University
MICROGATE
PLDA

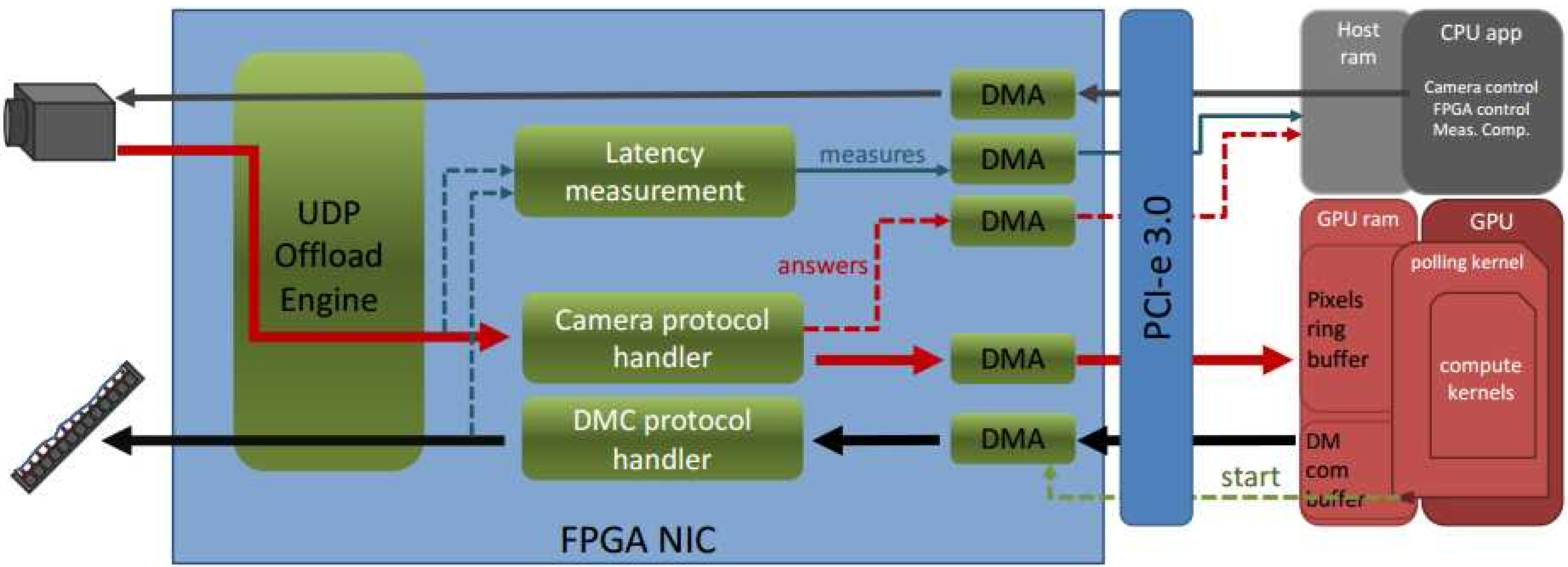
Data acquisition

FPGA writes/reads directly to/from GPU memory

Using only writes would be better though



FPGA/GPU optimized sync.



Little to no improvements, but CPU free for other kind of computations

WP 4

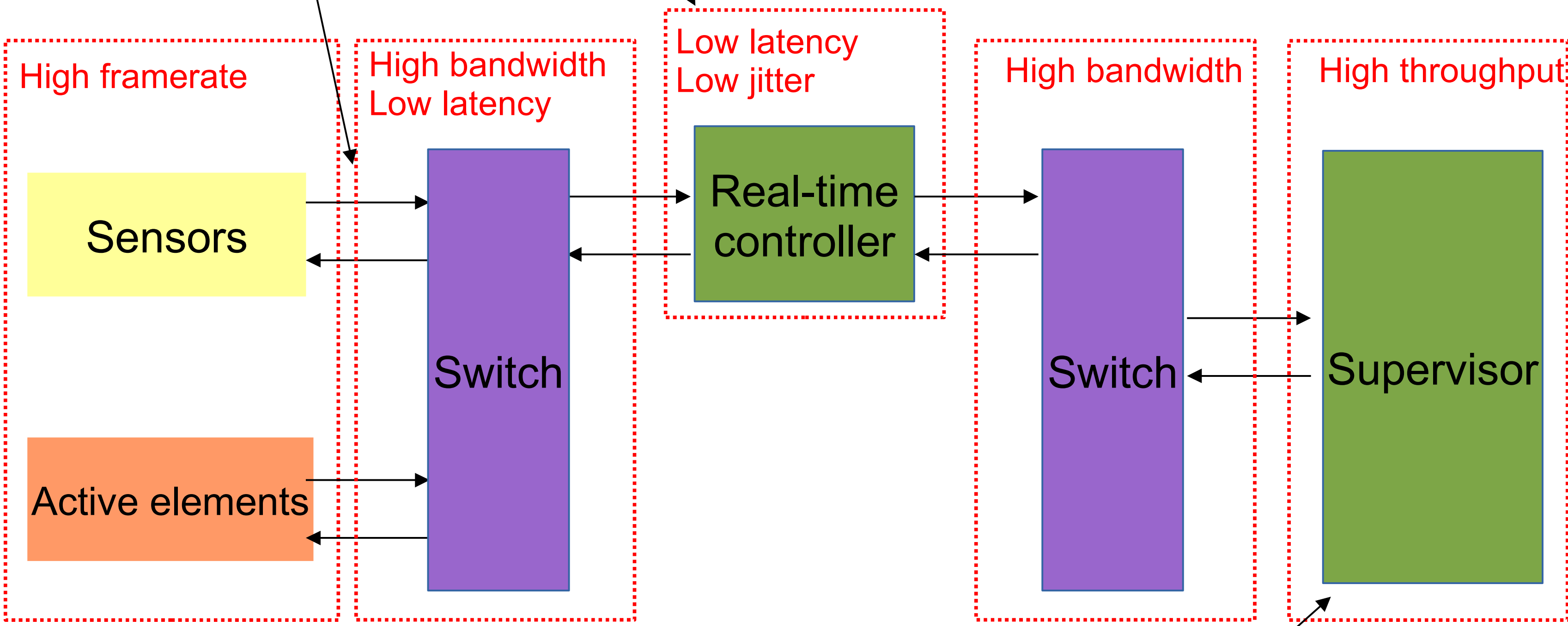
Real-time pipeline.
Includes sensors pixels streams processing and MVM for control of active elements

Pixel Streams processing, 2 options:
* tens of GFLOPS in simple arithmetics or
* hundreds of GFLOPS in batched Fourier Transform

MVM : up to 5 TFLOP/s (2.5 TMAC/s)

Up to 250 Gb/s of streaming data

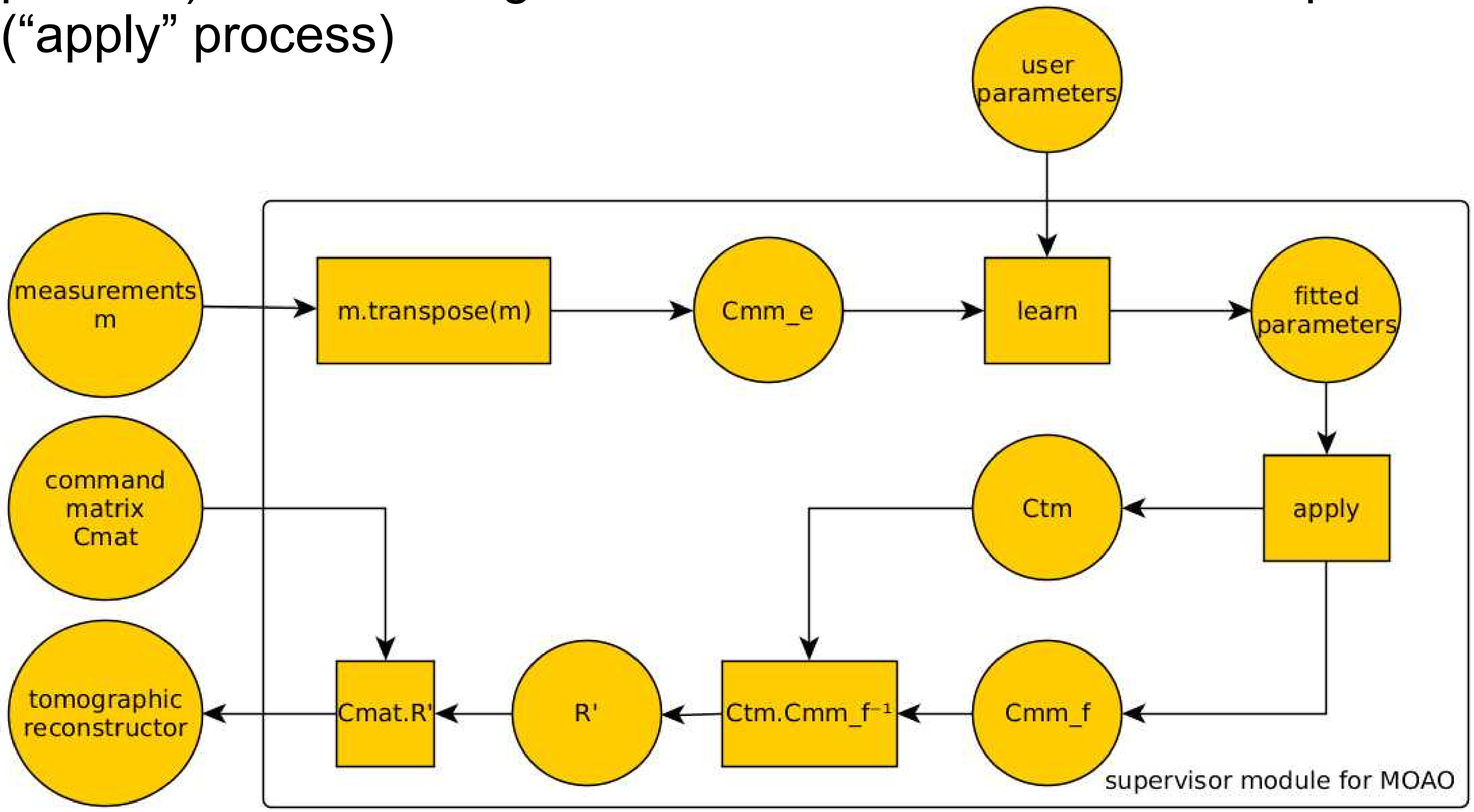
Performance must be deterministic, max latency : 2ms



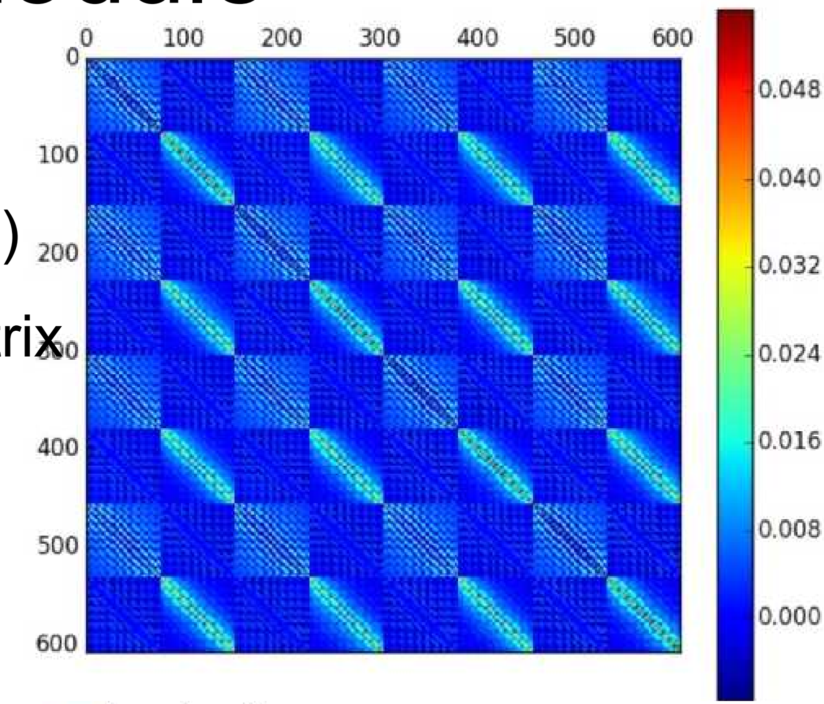
Supervisory module. Use the output data stream from RT pipeline to re-optimize the control matrix
2 stages : function optimization (gradient descent) and Choleski inversion : up to 100 TFLOP/s

Loop supervision module

Mix of cost function optimization for parameters identification (“Learn” process) and linear algebra for reconstructor matrix computation (“apply” process)



Loop supervision module

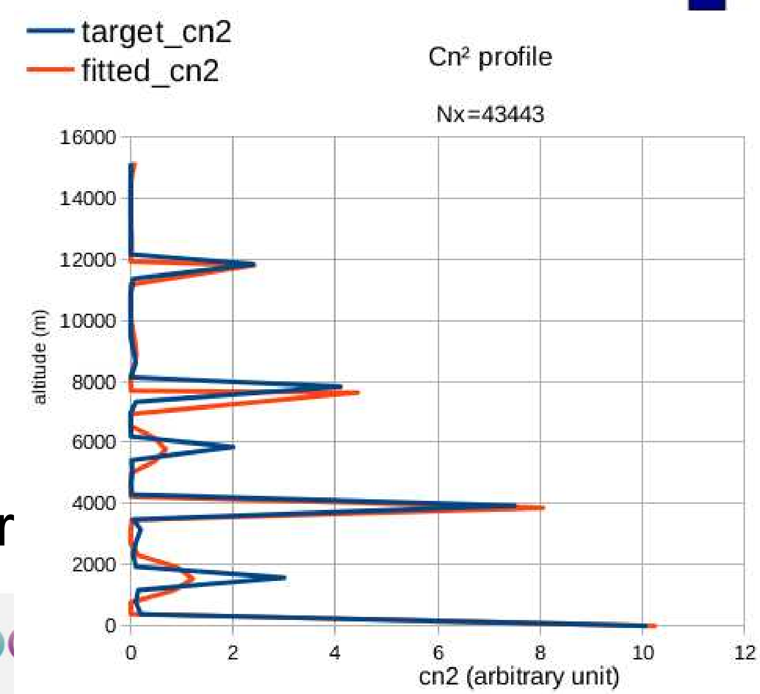


Parameters identification ("Learn" process)

- Fitting measurements covariance matrix on a model including system and turbulence parameters
- Using a score function

$$F(x)=\sum_{k=1}^{N^2}[Cmm_k-f_k(x)]^2$$

- Levenberg-Marquardt algorithm for function optimization
- Exemple of turbulence profile reconstruction
- Dual stage process (5 layers + 40 layer



l'Observatoire de Paris — LESIA
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique
Durham University
MICRO

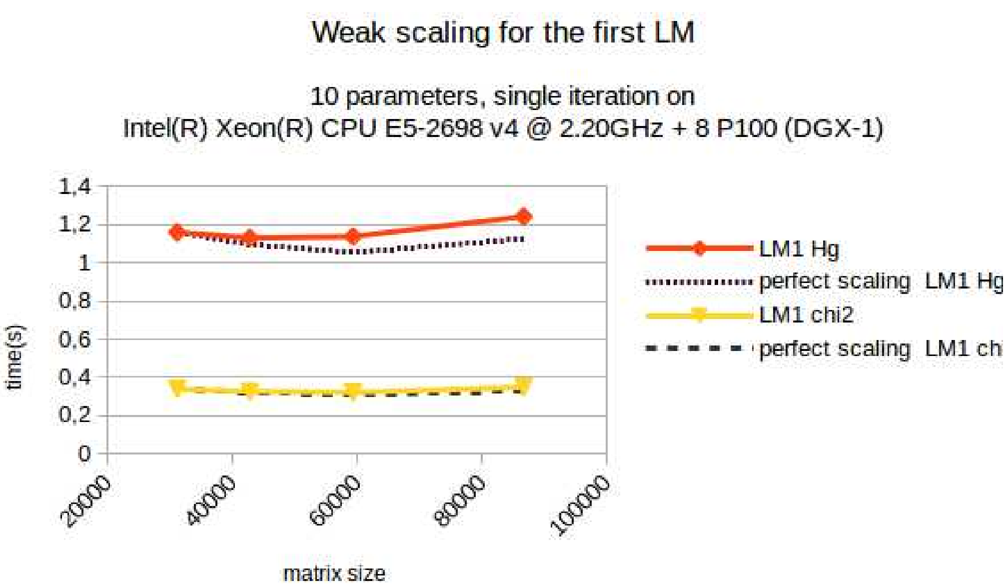
Loop supervision module

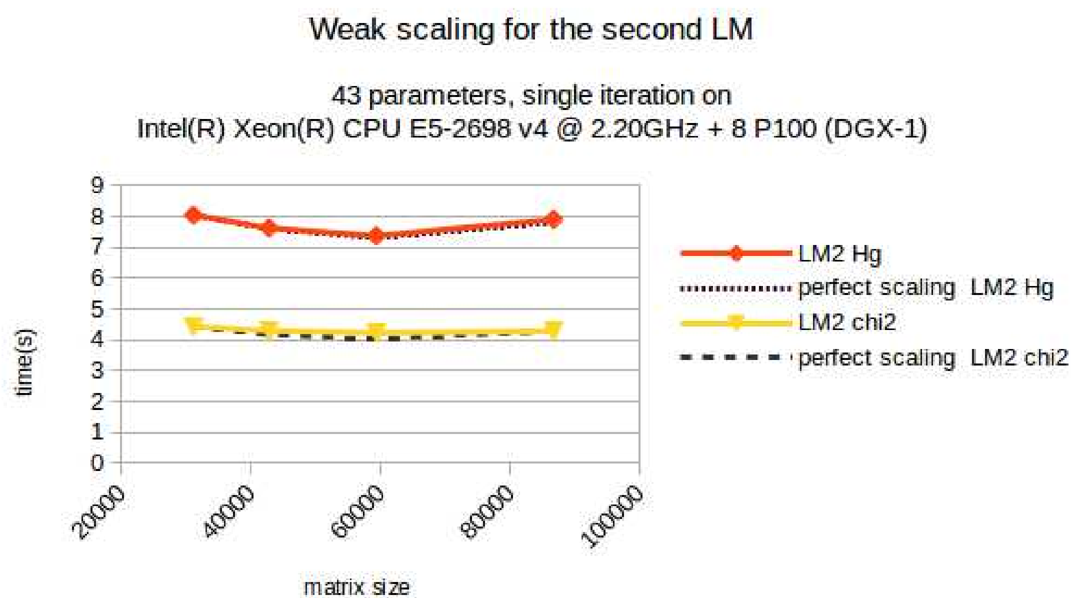
Performance for parameters identification ("Learn" process)

Multi-GPU process, including matrix generation and LM fit

Time to solution for a matrix size of 86k :240s (4 minutes)

- first pass (5 layers) : 25s
- Second pass (40 layers) : 213s





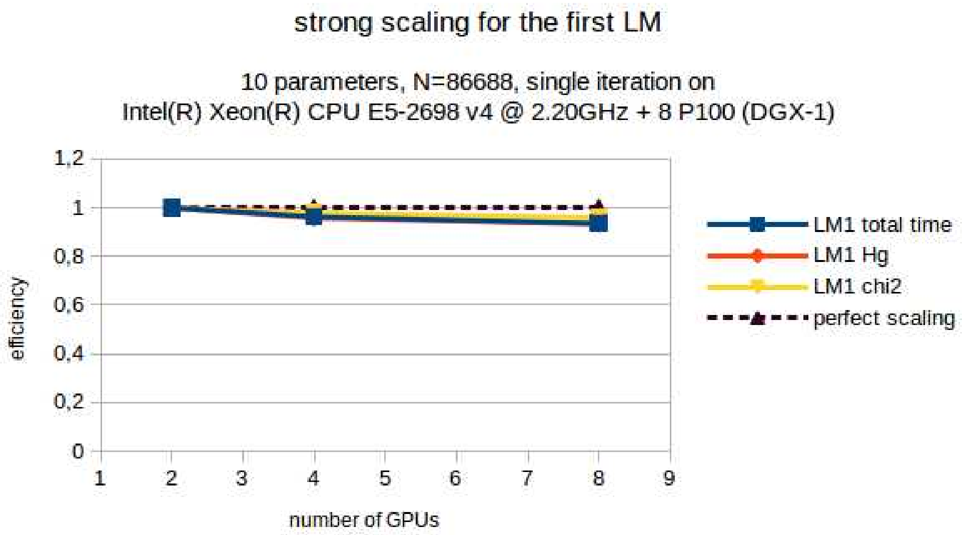
Loop supervision module

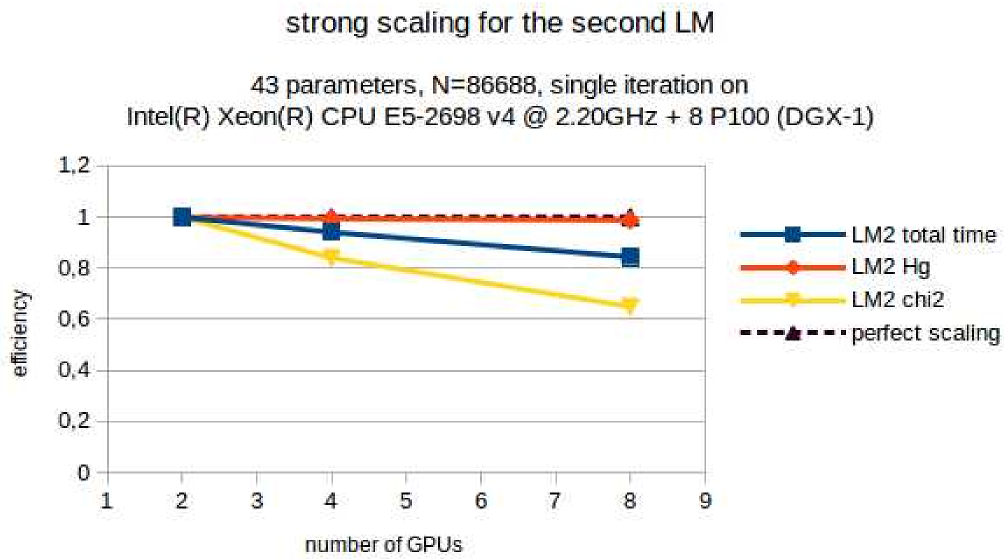
Performance for parameters identification ("Learn" process)

Multi-GPU process, including matrix generation and LM fit

Time to solution for a matrix size of 86k :

- first pass (5 layers) : 25sec
- Second pass (40 layers) : 213sec





Loop supervision module

Reconstructor matrix computation ("apply" process)

- Compute the tomographic reconstructor matrix using covarince matrix between "truth" sensor and other WFS and invert of measurements covariance matrix

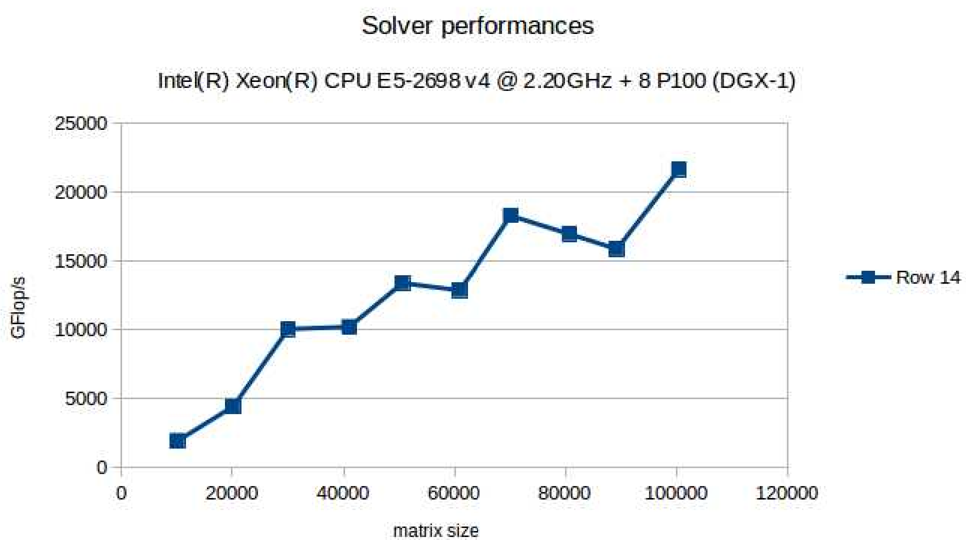
$$R' = Ctm \cdot Cmm_f^{-1}$$

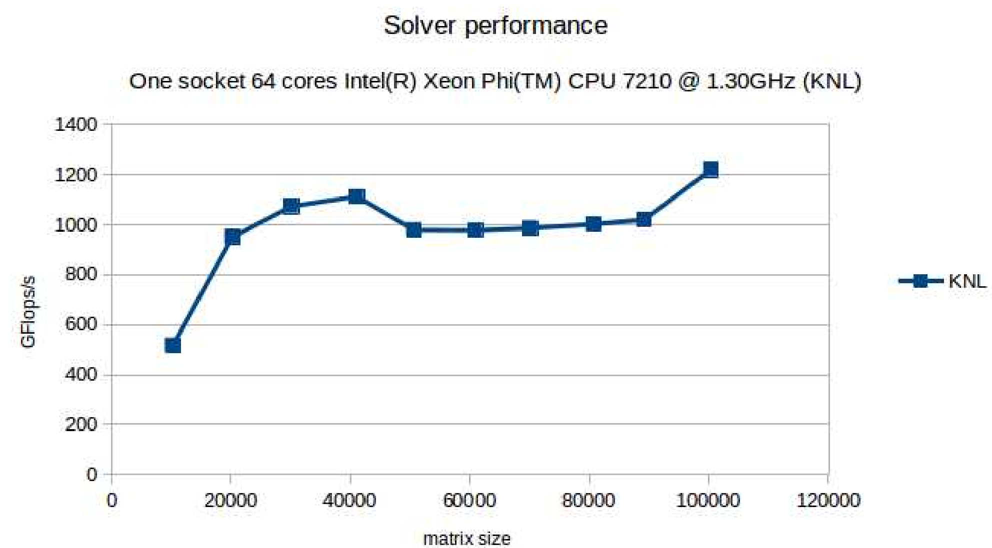
- Can use various methods : LU or Cholesky factorisation. "Brute" force : direct solver
- Standard Lapack routine : "posv" : mostly compute-bound, high level of scalability
- Highly portable code : explore various architectures by using standard vendor provided maths libraries

Loop supervision module

Performance for reconstructor matrix computation (“apply” process)

Comparing last generation of GPU (NVIDIA P100) and last generation of Intel Xeon Phi (KNL)



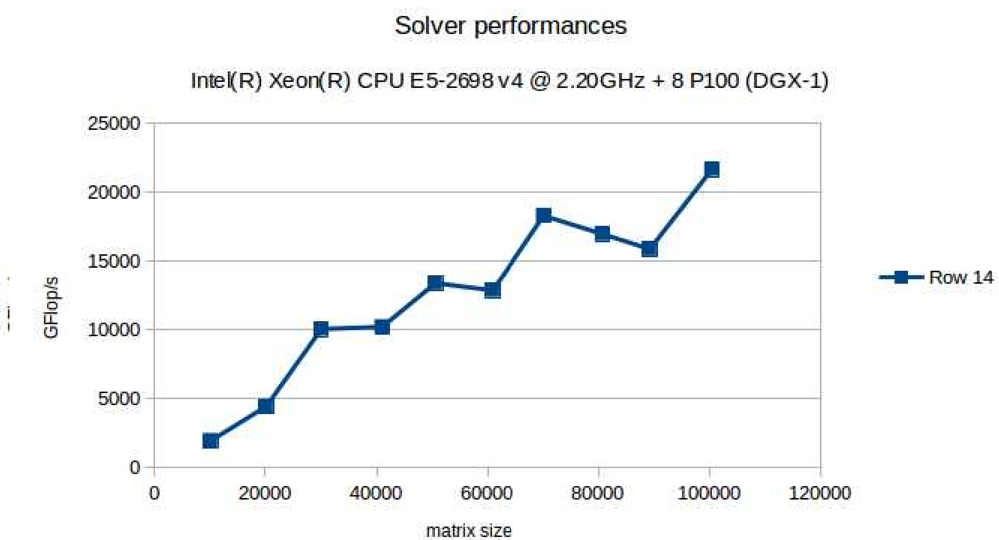


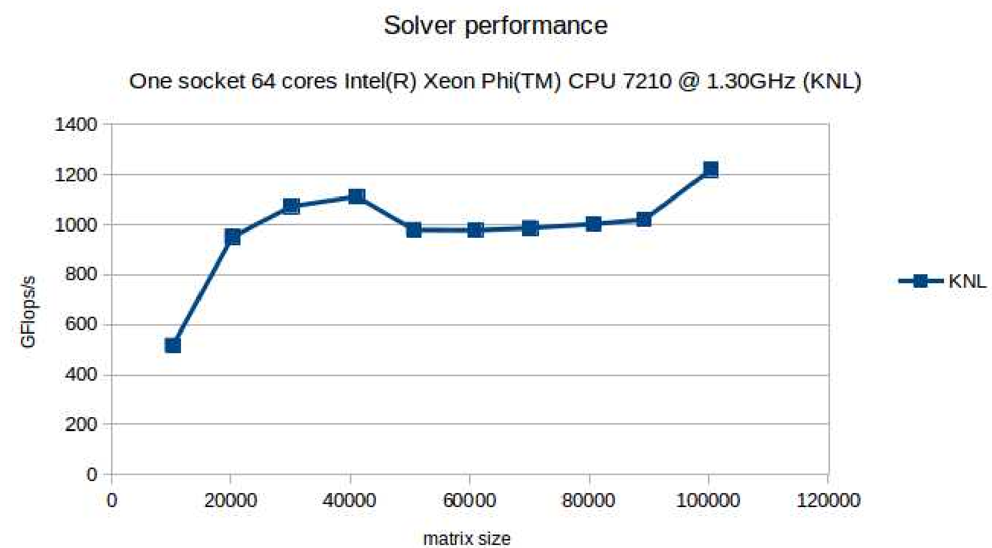
8 GPUs together reach more than 21 TFLOP/s while a single KNL can only reach about 1.2 TFLOP/s in peak performance

Loop supervision module

Performance for reconstructor matrix computation ("apply" process)

Comparing last generation of GPU (NVIDIA P100) and last generation of Intel Xeon Phi (KNL)





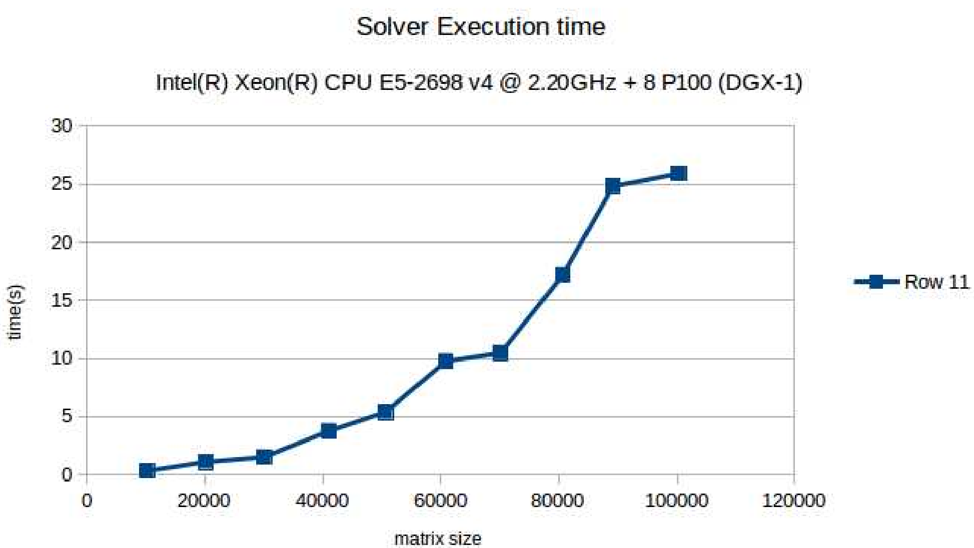
GPUs can deliver better peak perf. (saturation not reached, expect >2.5 or more) and the NVlink interconnect seems to perform very well

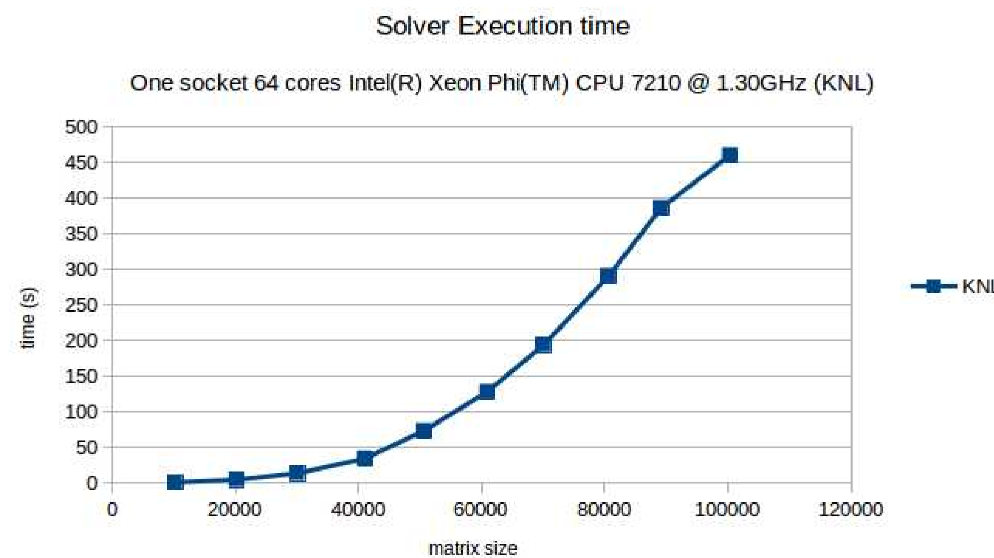
l'Observatoire de Paris — LESIA
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique
Durham University
MICROGATE
PLDA

Loop supervision module

Performance for reconstructor matrix computation ("apply" process)

- Comparing last generation of GPU (NVIDIA P100) and last generation of Intel Xeon Phi (KNL)





- Record time-to-solution on DGX-1 : MAORY / HARMONI full scale (100k x 100k matrix) : 25sec to compute tomographic reconstructor

WP4: deliverables

Task 4.1 (OdP):

- D4.1: GPU cluster for RT-box design and test report (OdP – M6 – delivered)
- D4.2: GPU cluster for RT-box prototype (OdP – M24)

Task 4.2 (OdP):

- D4.3: GPU cluster for supervisor design and test report (OdP – M6 – delivered)

Task 4.3 (UoD):

- D4.4: Intel Xeon Phi cluster for RT-box prototype design and test report (UoD – delivered)
- D4.5: Intel Xeon Phi cluster for RT-box prototype (UoD – M24)

Task 4.4 (UoD):

- D4.6 FPGA cluster for RT-box prototype design and test report (UoD – M24)
- D4.7: FPGA cluster for RT-box prototype (UoD – M36)

Contribution to WP 6: simulator

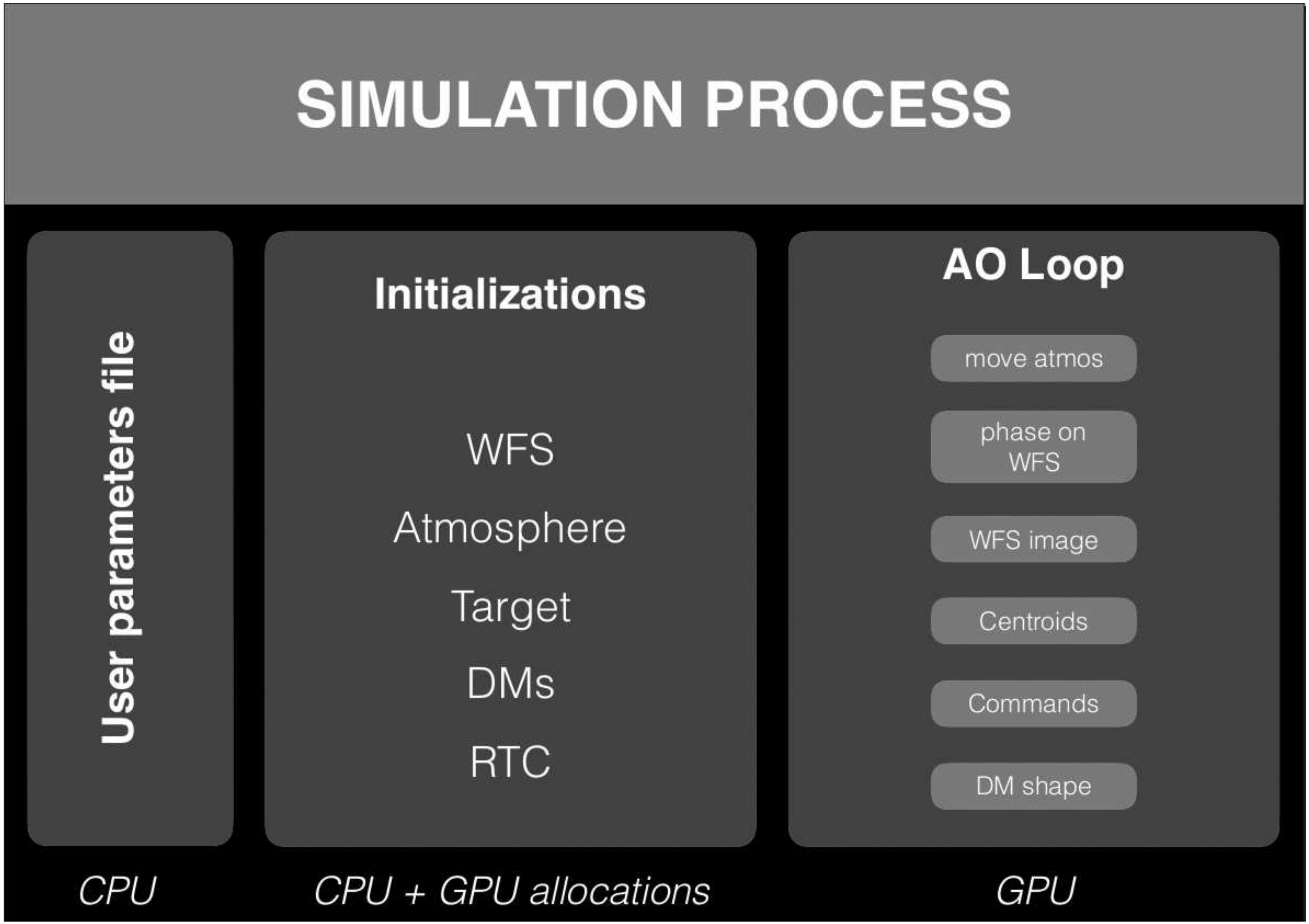
OdP team responsible for simulator SW development, integration and maintenance (task 6.2)

- efficient simulation platform based on GPU
- optimized interface with real-time core pipeline
- inherit from large scale project funded by French National Research Agency (ANR): COMPASS, 1M€ budget, 6 partners in France, 36 months (ended in Feb. 2016)

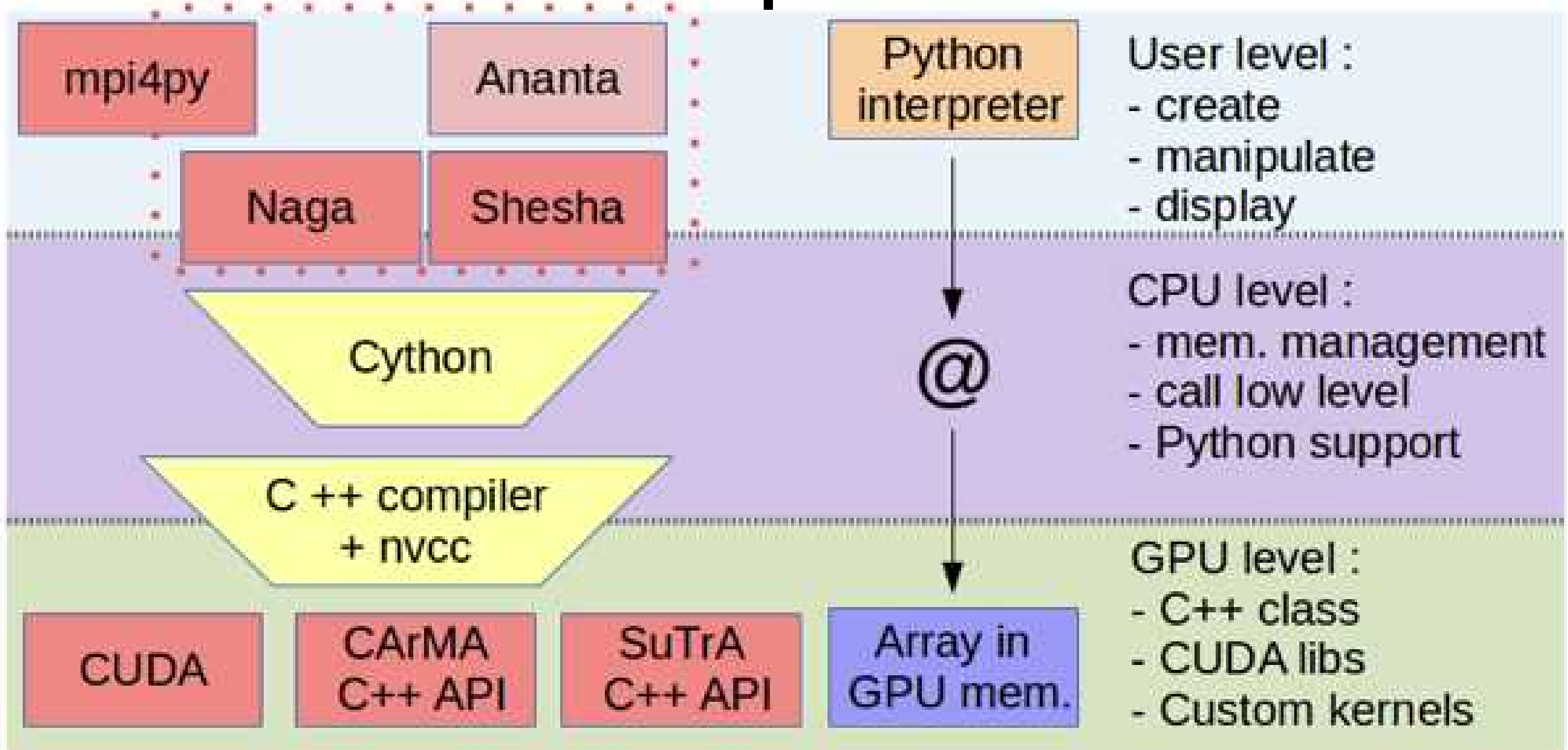
Coordination with UoD for deployment on selected HW and interfaces

- several levels of simulation with various accuracy
- critical component for final prototype performance assessment

The COMPASS platform



The COMPASS platform

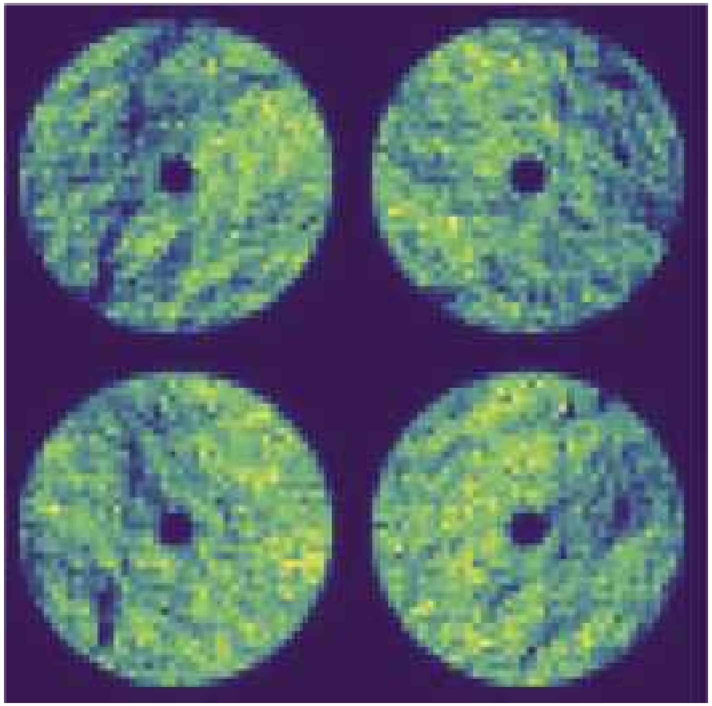


User interface coded in Python for long tem maintenance

Main computations relies on GPU:

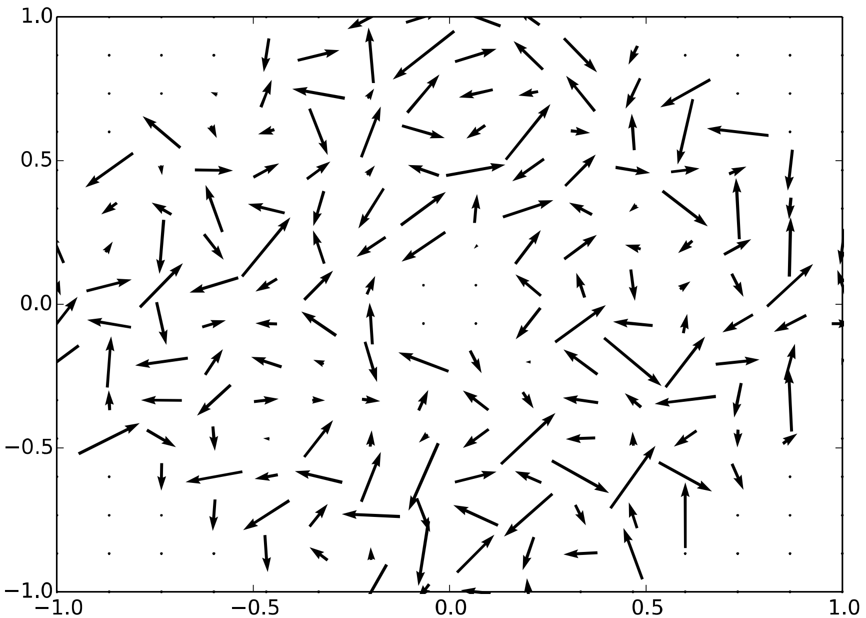
- **CArMA**: C++ Api for Massively parallel Applications
- **SuTrA**: Simulation Tool for Adaptive optics
- Use optimized libraries such as CUBLAS, CUFFT, MAGMA…

Features



Wavefront Sensor models:

- Shack-Hartmann
- Pyramid
- Laser Guide Star



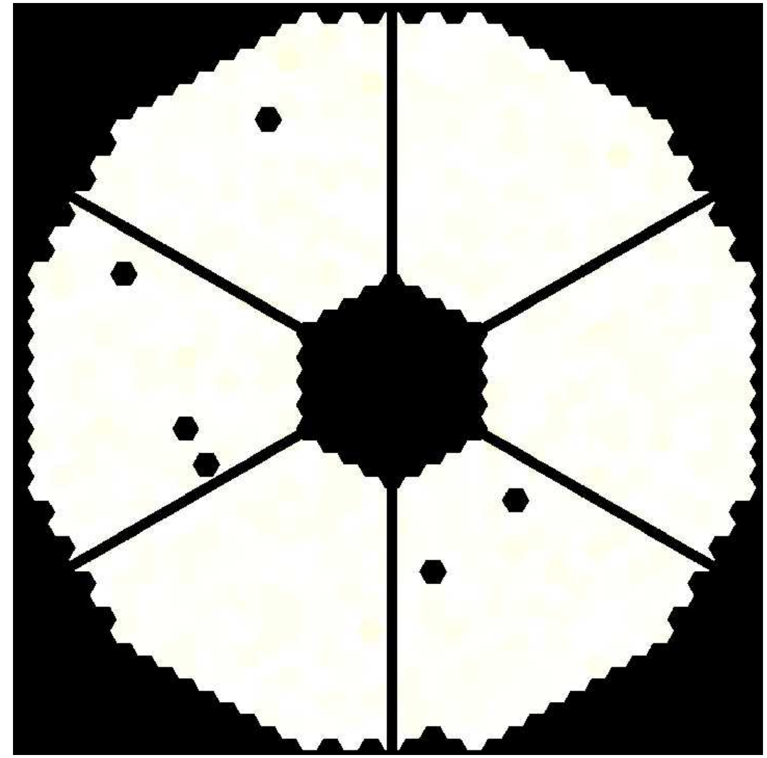
Centroiding methods:

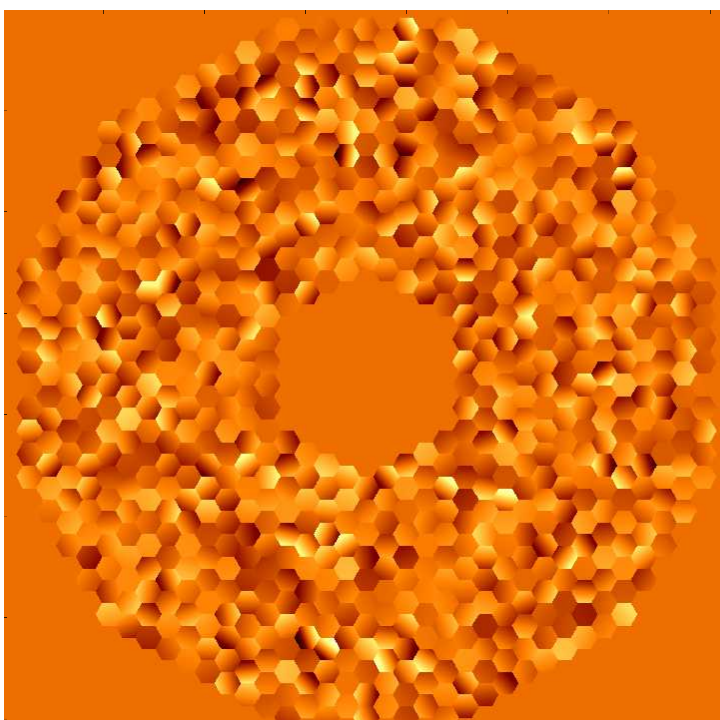
- Center of gravity (cog)
- Thresholded cog
- Weighted cog
- Brightest pixels
- Correlation

Features

Controllers:

- Least square
- Modal optimization
- Minimum variance
- CuReD
- Projection



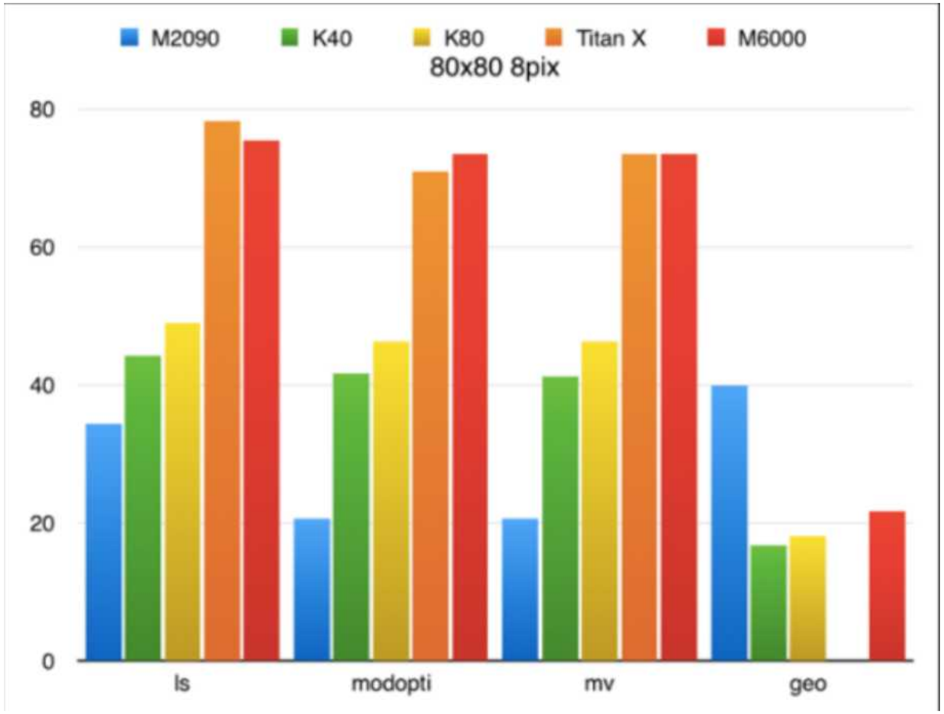


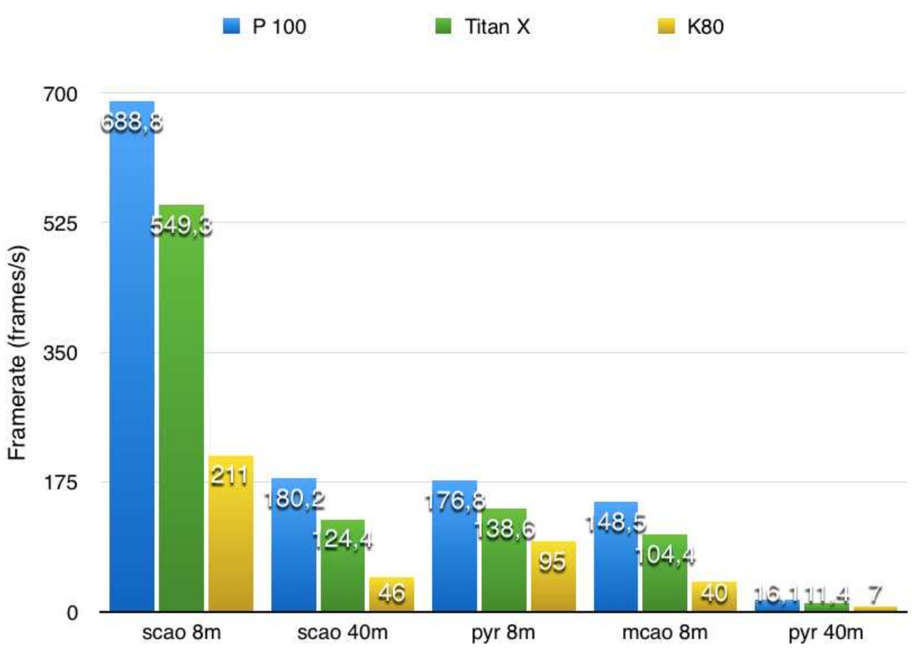
E-ELT:

- Hexagonal pupil
- Spiders
- Phase aberration
- M4 influence functions

Real-time simulator

- Using COMPASS for E2E should provide a scalable solution over the long term
  - Execution times from F. Ferreira





Contribution to WP 6: simulator

Current development on interface with real-time pipeline

- Based on previous experience with instruments control SW
- Using DDS as middleware
- Sustained simulation framerate is 90% on standalone simulation speed
- In collaboration with UoD

Implementing key components in the simulation

- Accurate error budget for the AO loop : prototype output validation (PhD thesis)
- Critical E-ELT features : deformable mirror and segmented pupil (SW engineer + instrument scientist)

WP6: deliverables

Task 6.1 (UoD):

- D6.1: Simulator interface definition document (UoD – M12)
- D2.2: Simulator HW final design report (UoD – M24)

Task 6.2 (OdP):

- D6.4: Simulator SW final design report (OdP – M24)

Task 6.3 (UoD):

- D6.3: Simulator performance report (UoD – M24)