Green Flash

High performance computing for real-time science

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

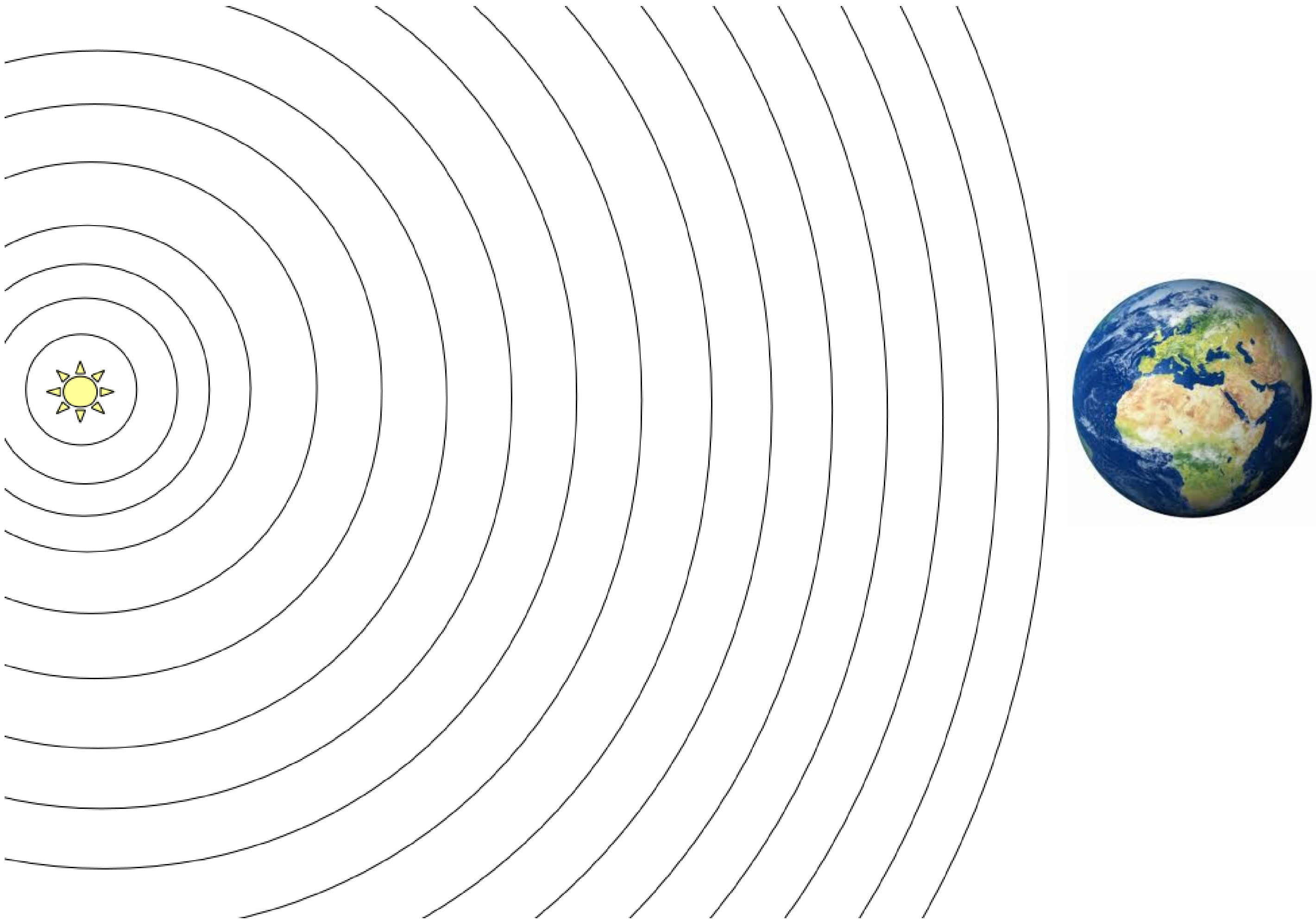
Observing stars from the ground

- Atmospheric turbulence reduces image quality

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

Observing stars from the ground

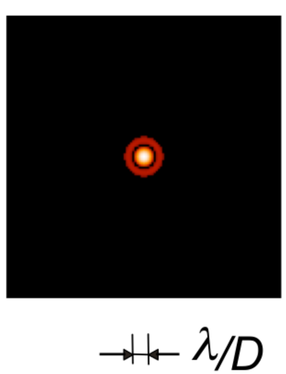
- From a spherical wave to a flat wavefront

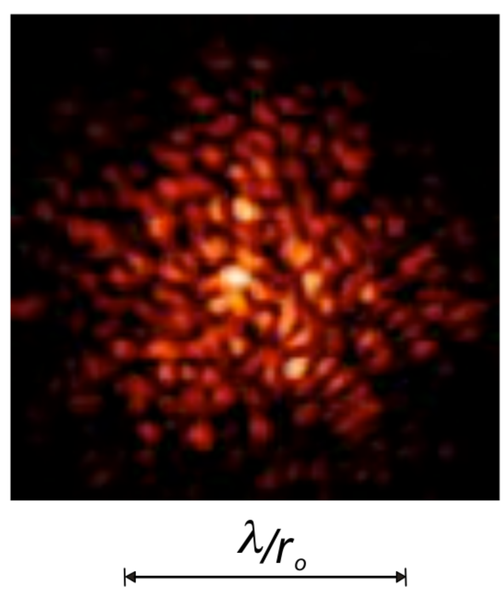


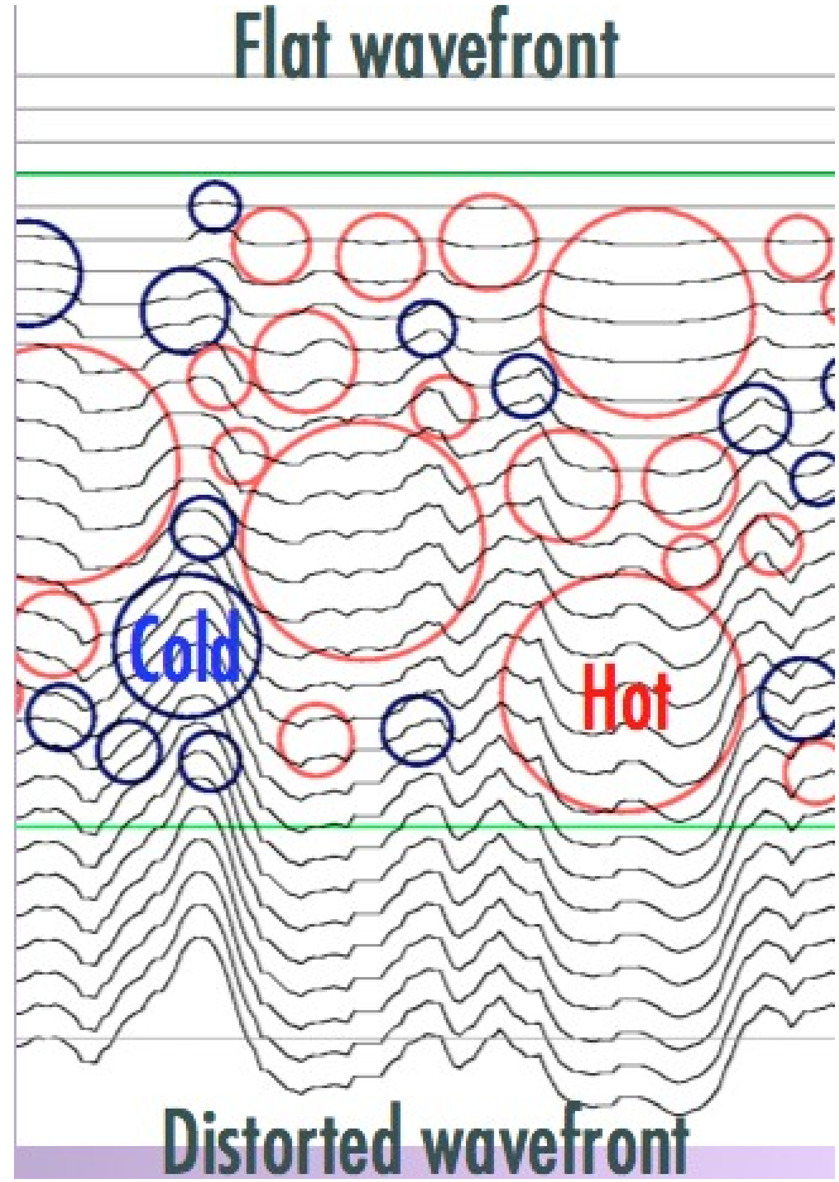
l'Observatoire de Paris — LESIA — Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique · Durham University · MICROGATE · PLDA

Observing stars from the ground

- Mixture of hot and cold air disturbs the wavefront
- Building a larger telescope does not help to get sharper images



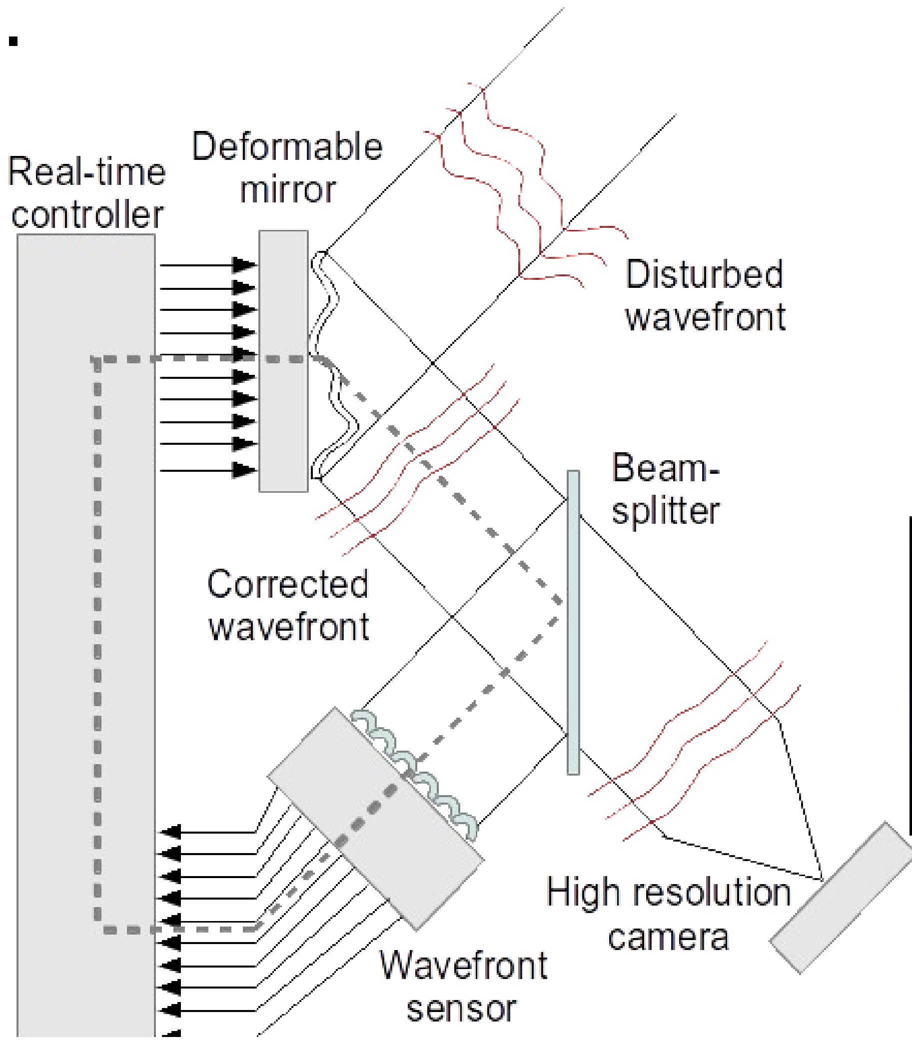


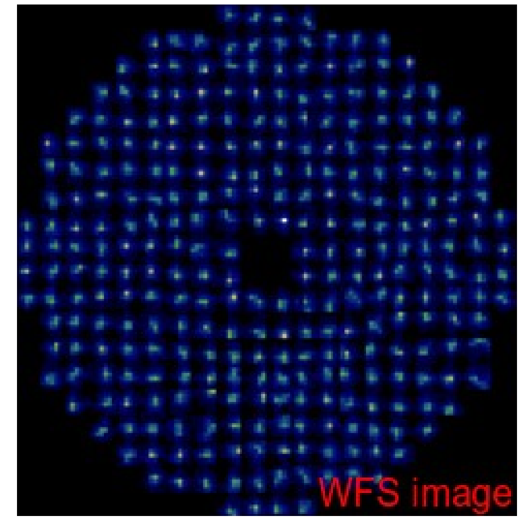


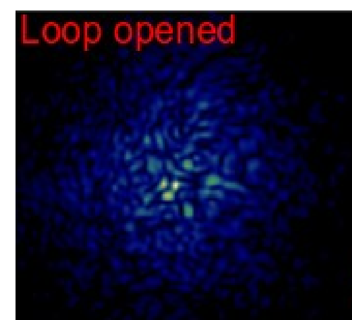
l'Observatoire de Paris — LESIA
Durham University
MICROGATE
PLDA

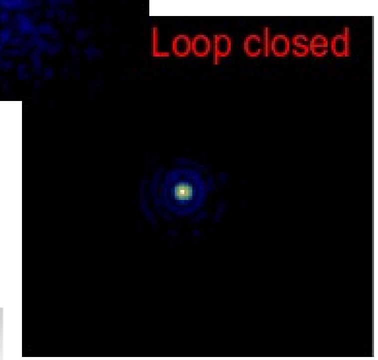
Adaptive optics

- Compensate in real-time for the wavefront errors ...





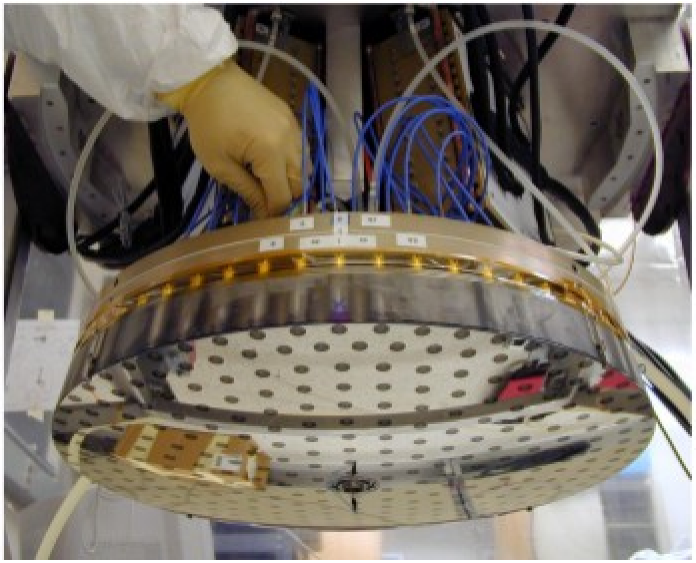


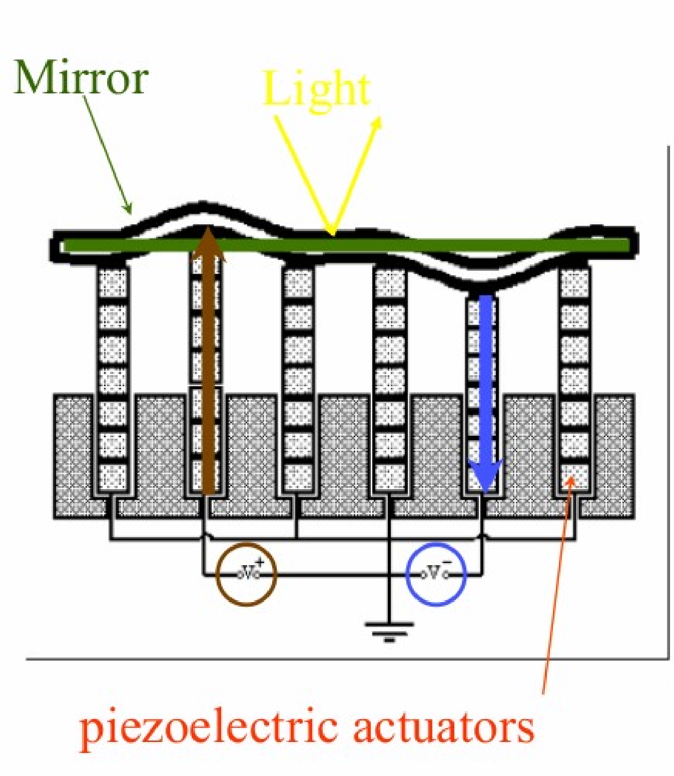


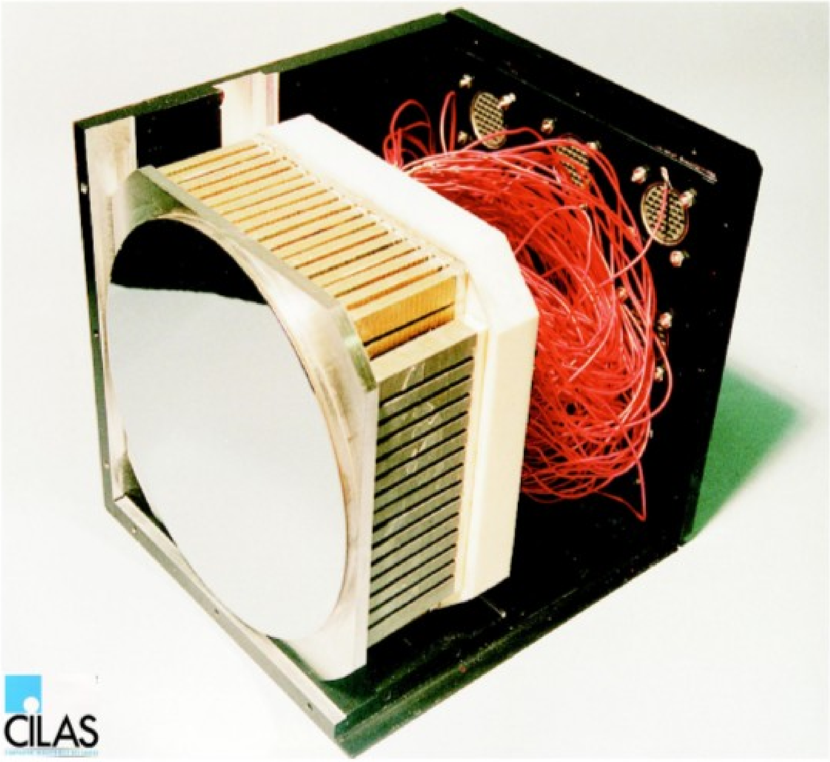
l'Observatoire de Paris LESIA
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique
Durham University
MICROGATE
PLDA

Adaptive optics

- … using one or several deformable mirror(s) …

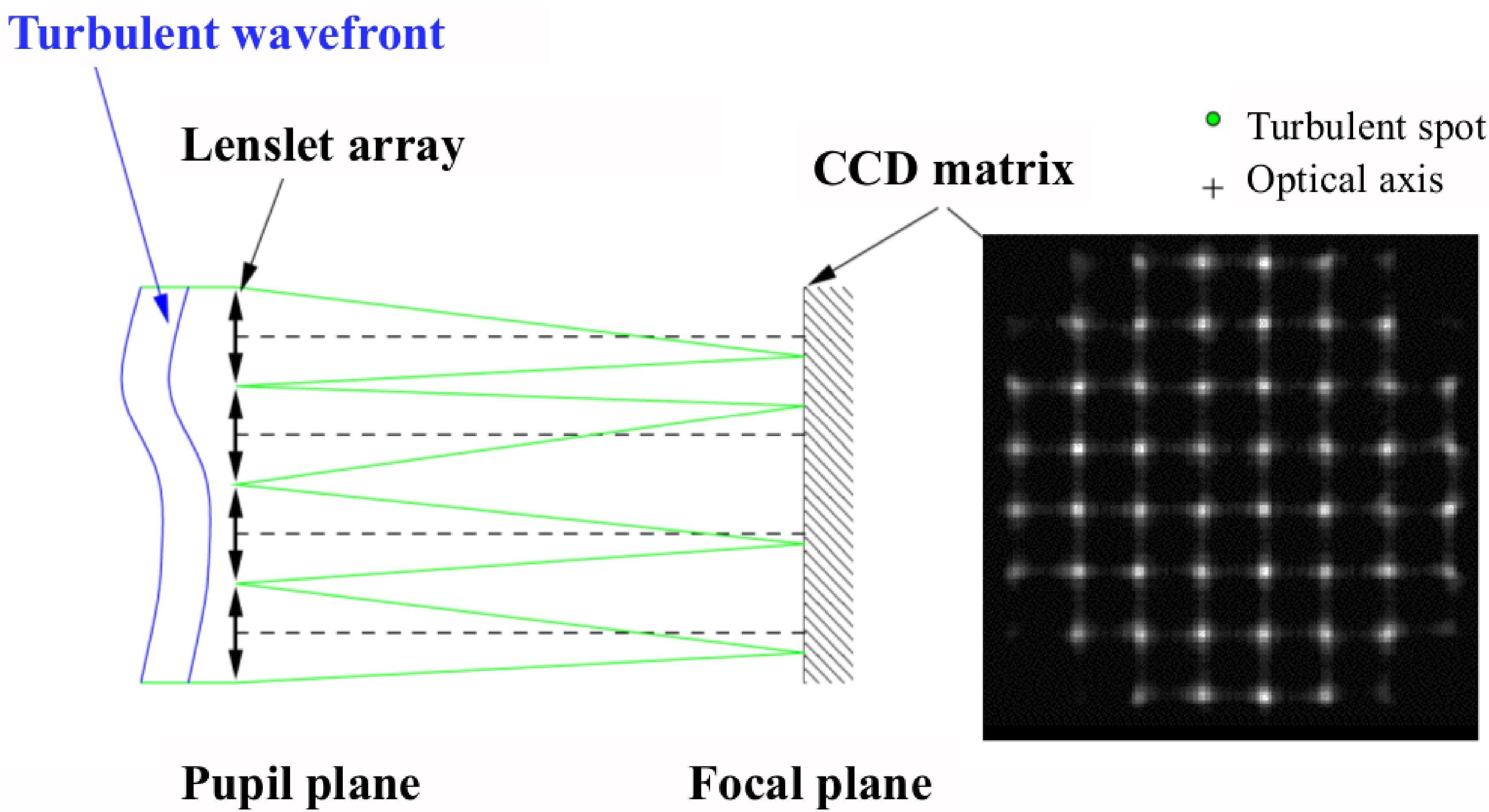






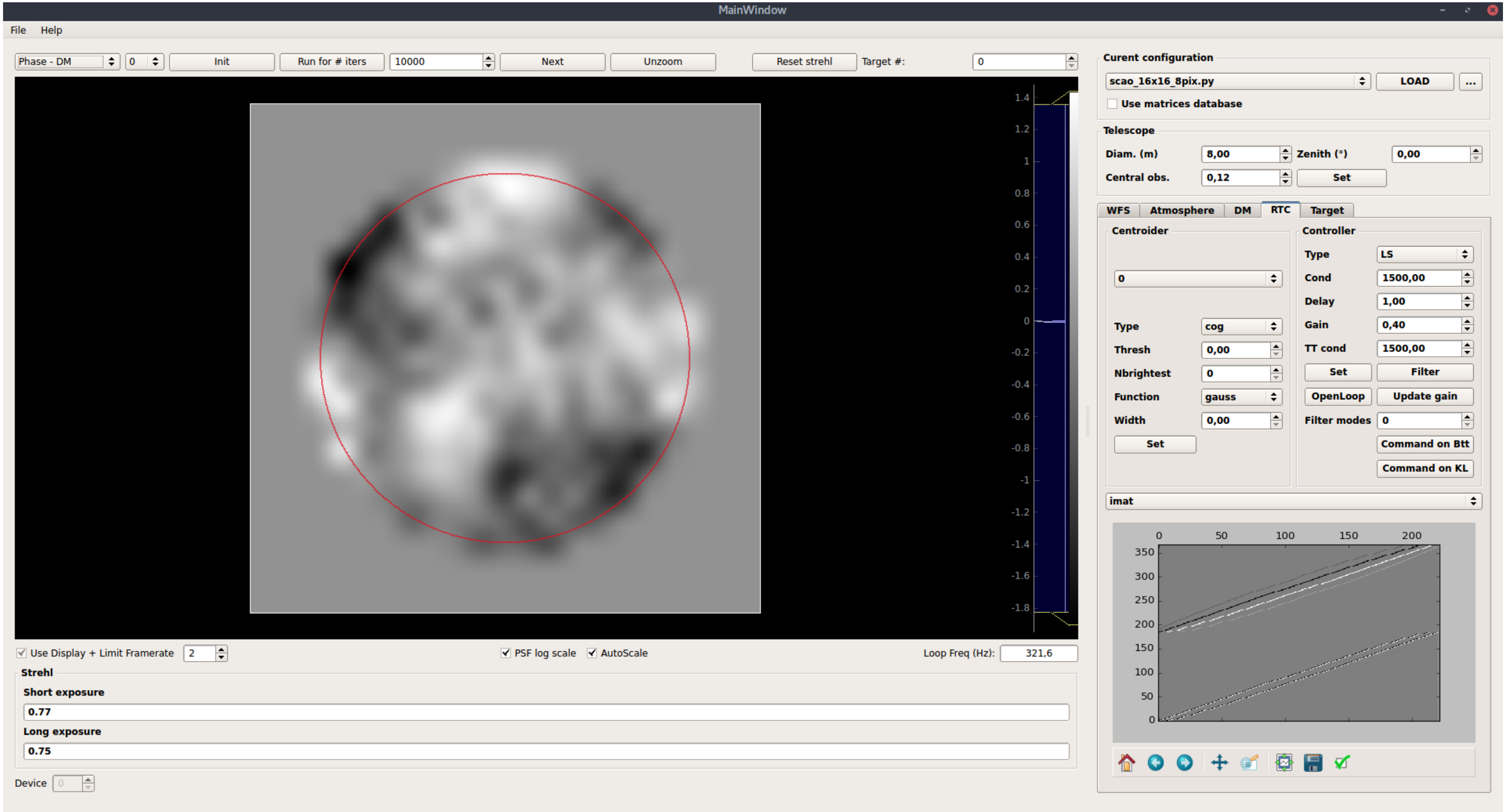
Adaptive optics

- … while wavefront errors are measured using one or several wavefront sensors



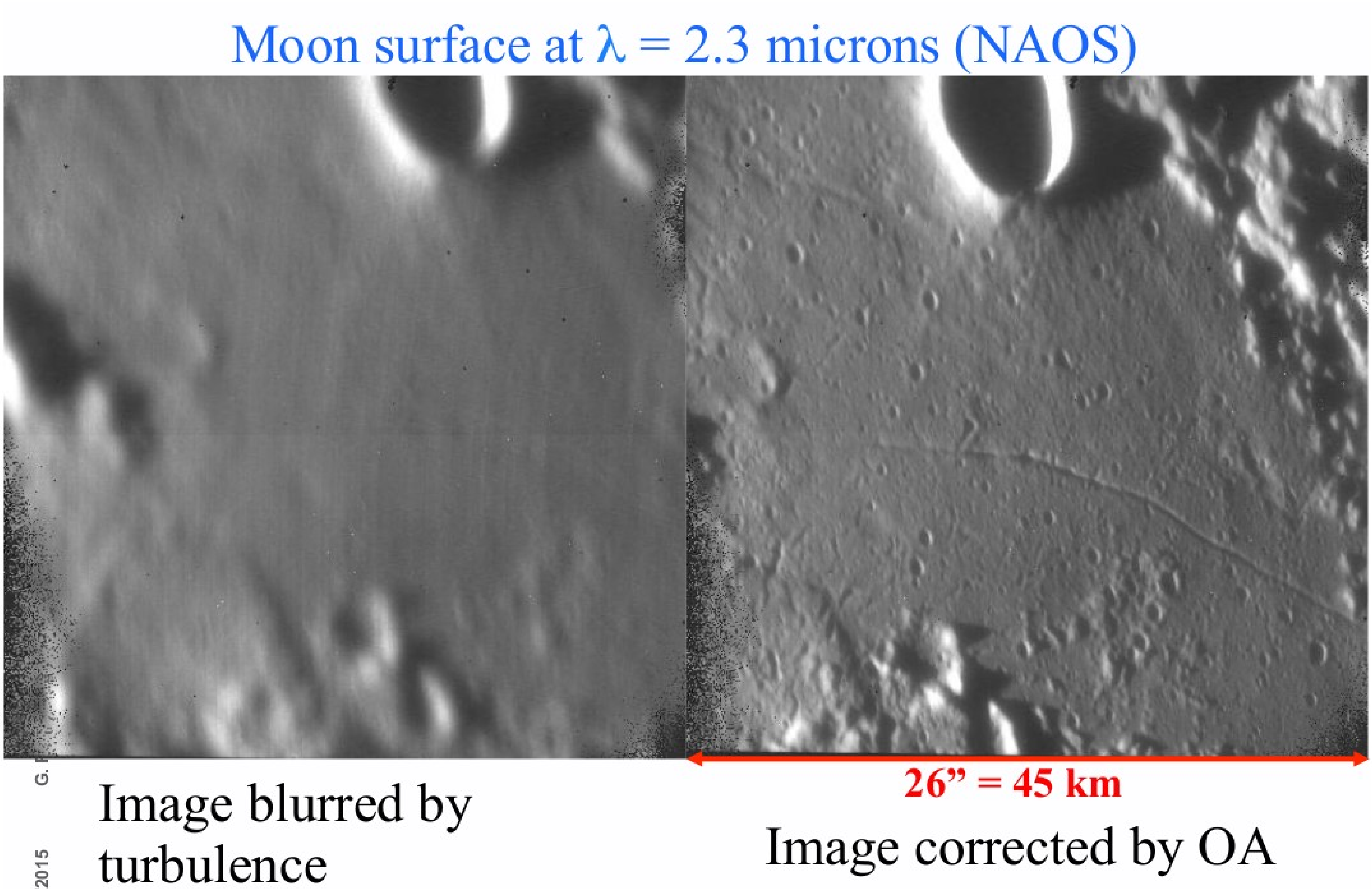
Adaptive optics

- Quick demo using end-to-end simulation tool



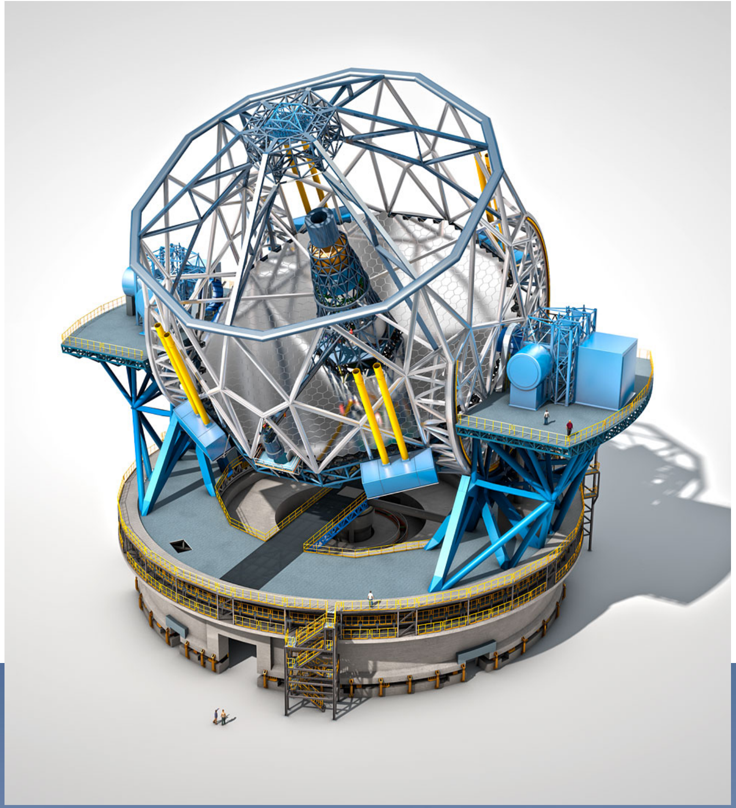
Adaptive optics

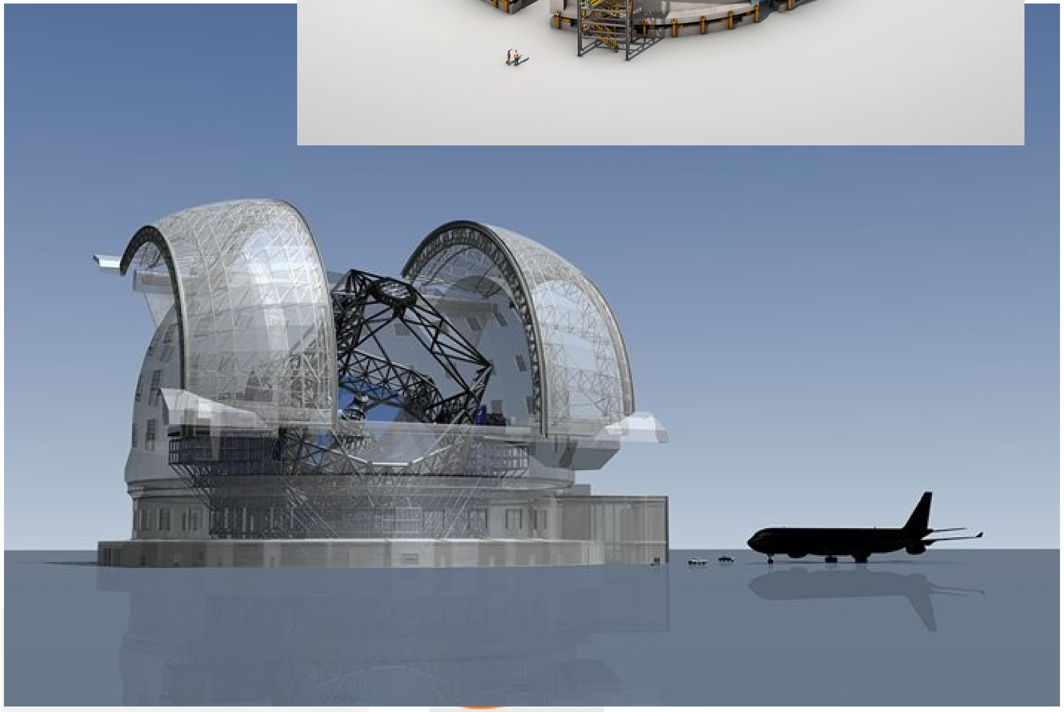
- Astronomers revenge: full resolution of a 8m telescope



European Extremely Large Telescope

- 39m diameter telescope : x5 in diameter => x25 in system complexity
  - 100m dome, 2800 tones structure rotating @ 360°, seismic safe (Chile)
  - 1.2 G€ project, first light foreseen in 2024
  - Construction led by ESO (European Southern Observatory), international organization funded by 15 European countries
  - Telescope components + science instruments built by European research labs + industrial partners



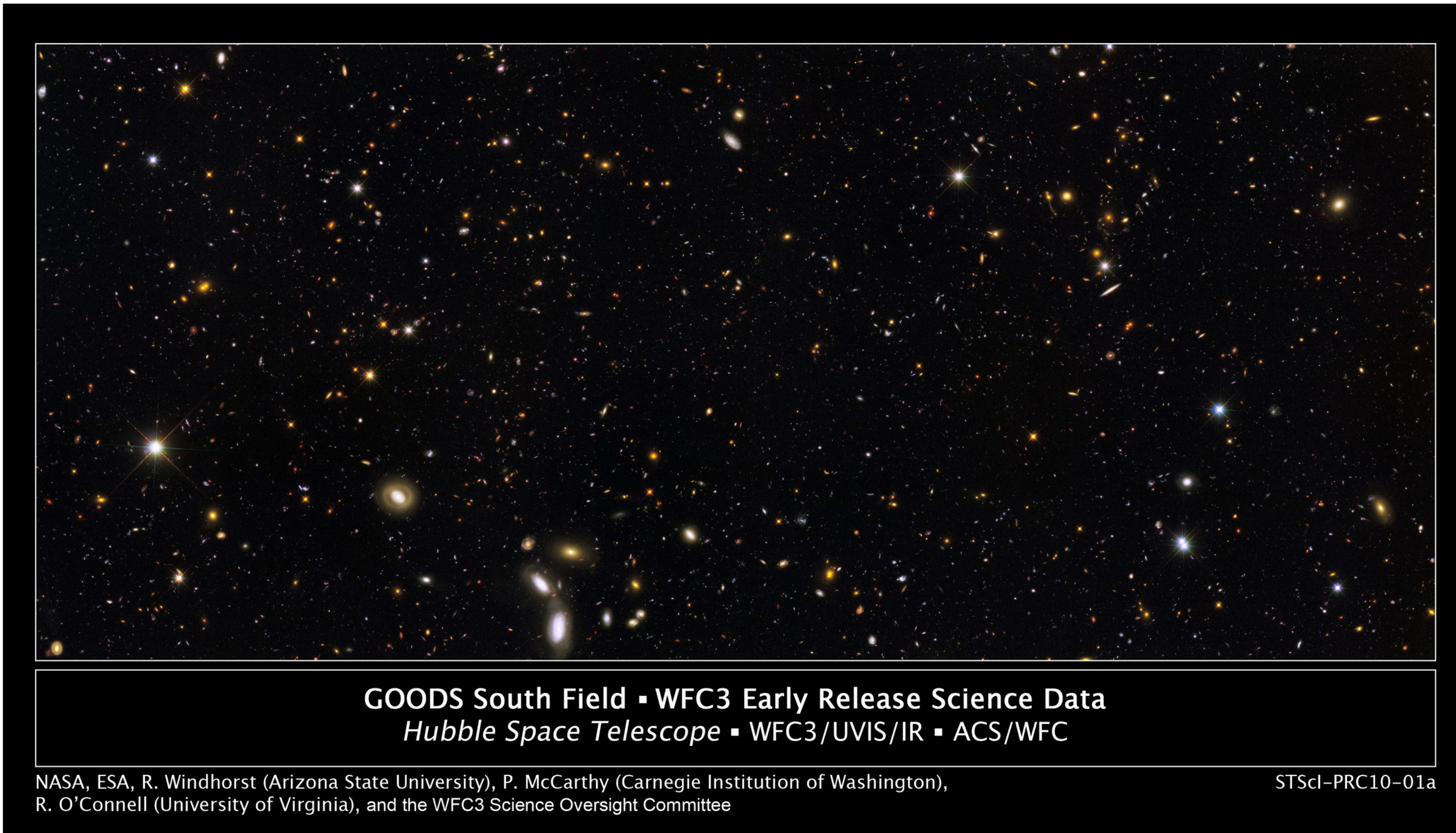


l'Observatoire de Paris | LESIA
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique

Durham University

Addressing grand challenges

- First stars, first galaxies, birth and evolution of the Universe

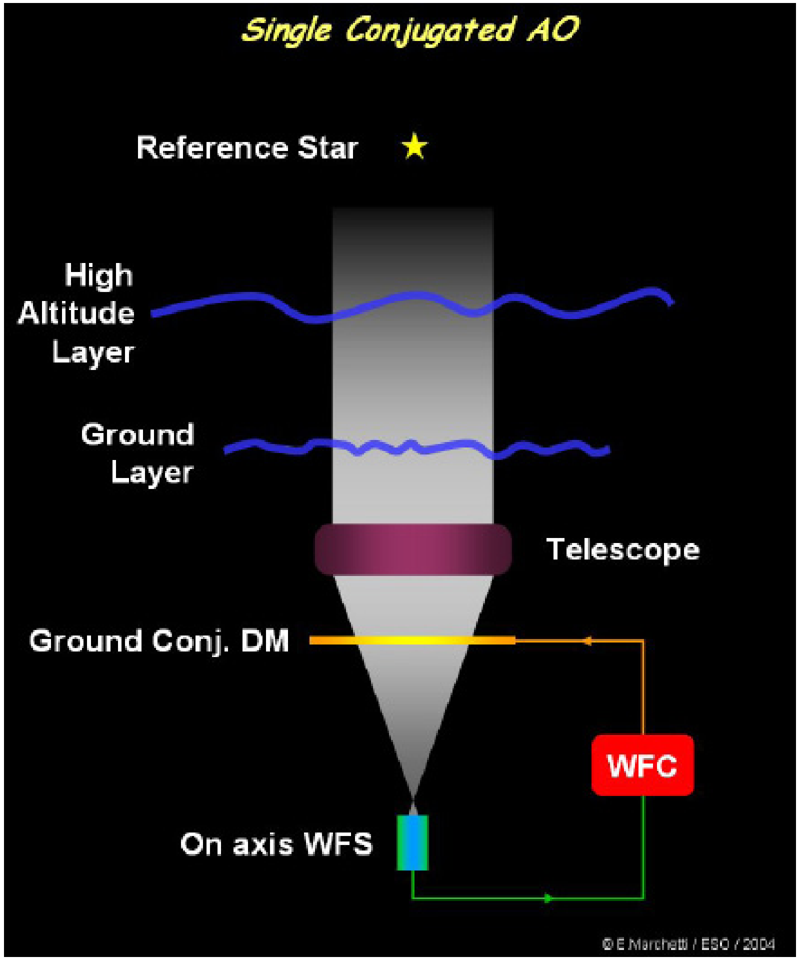


Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique

University

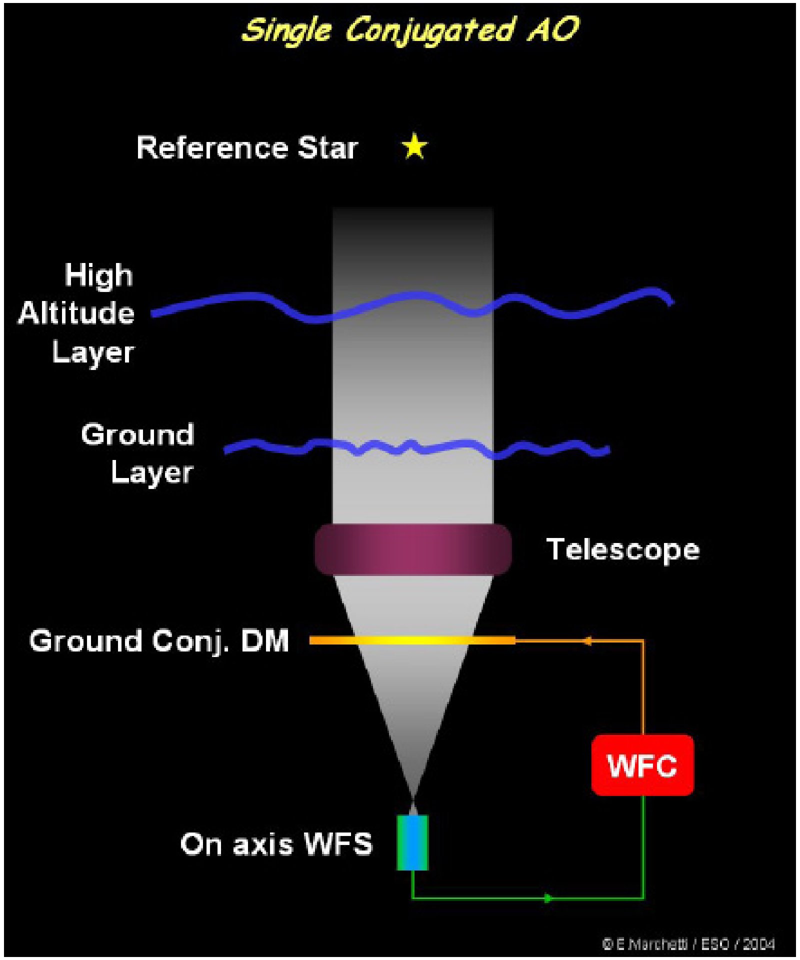
Adaptive optics flavors

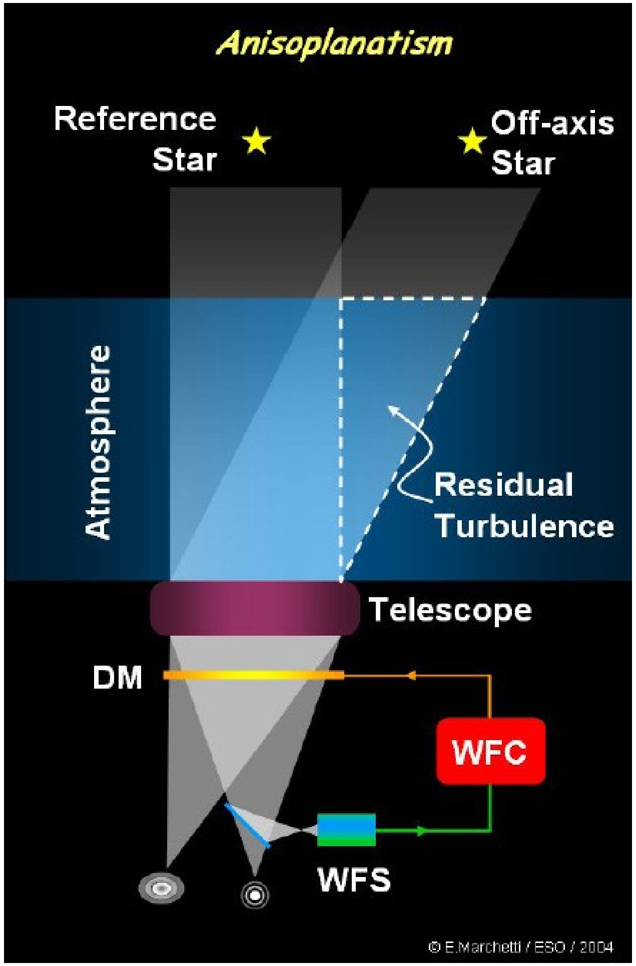
- From classical AO ...



Adaptive optics flavors

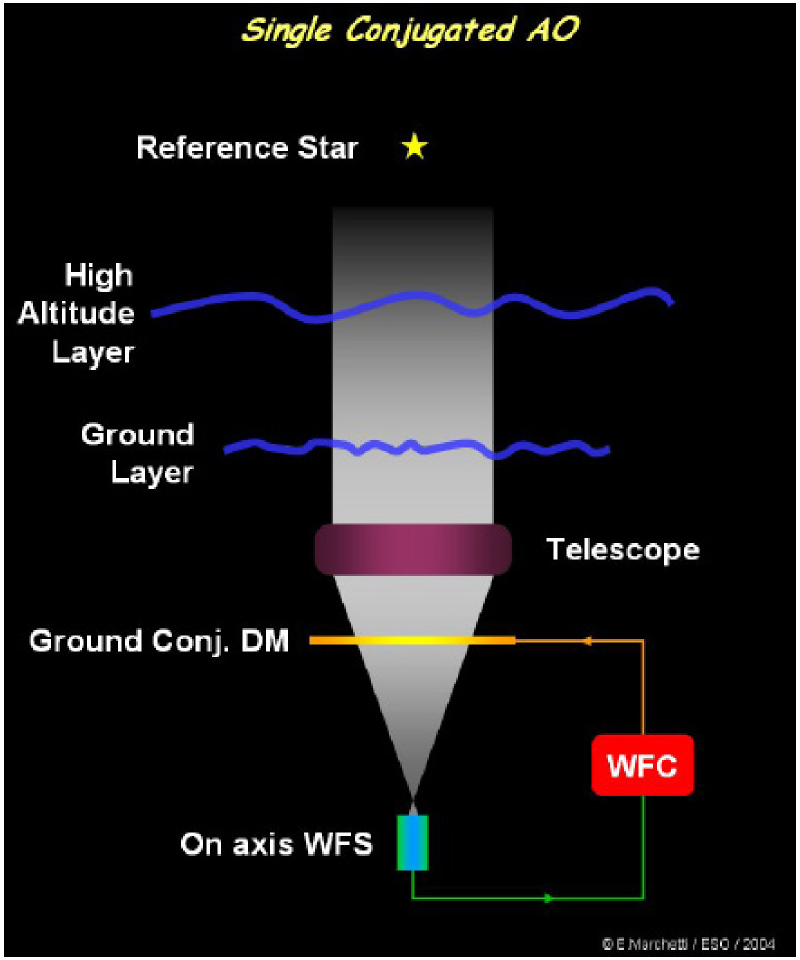
- From classical AO ...

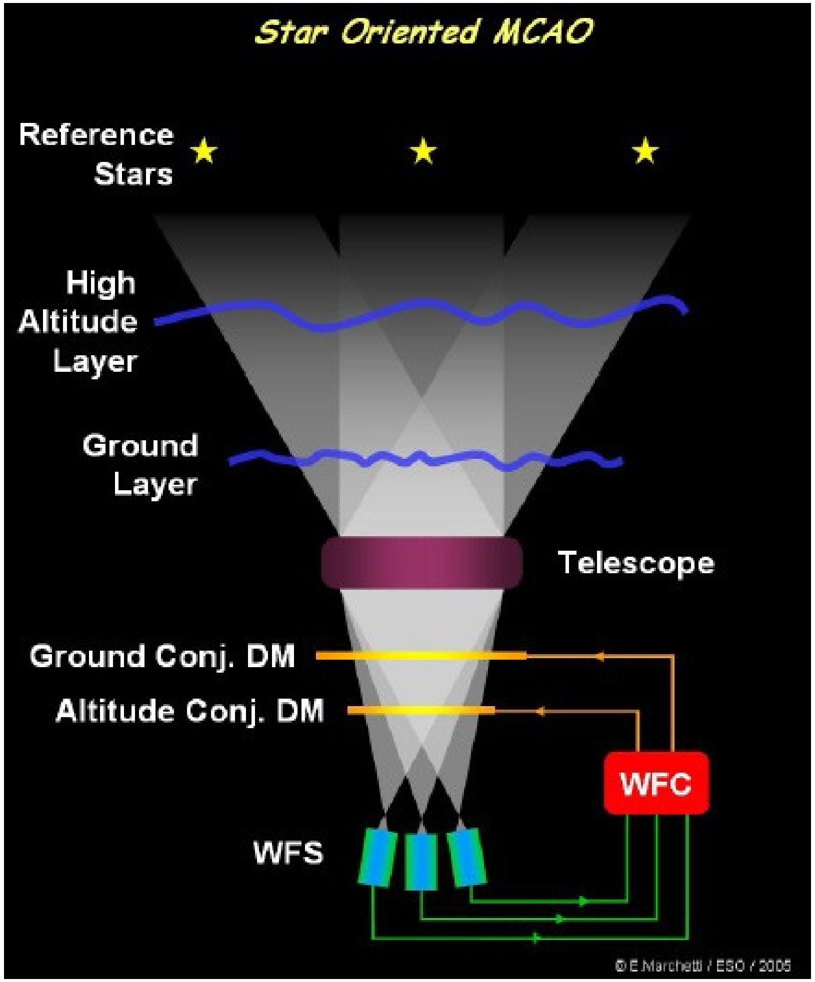




Adaptive optics flavors

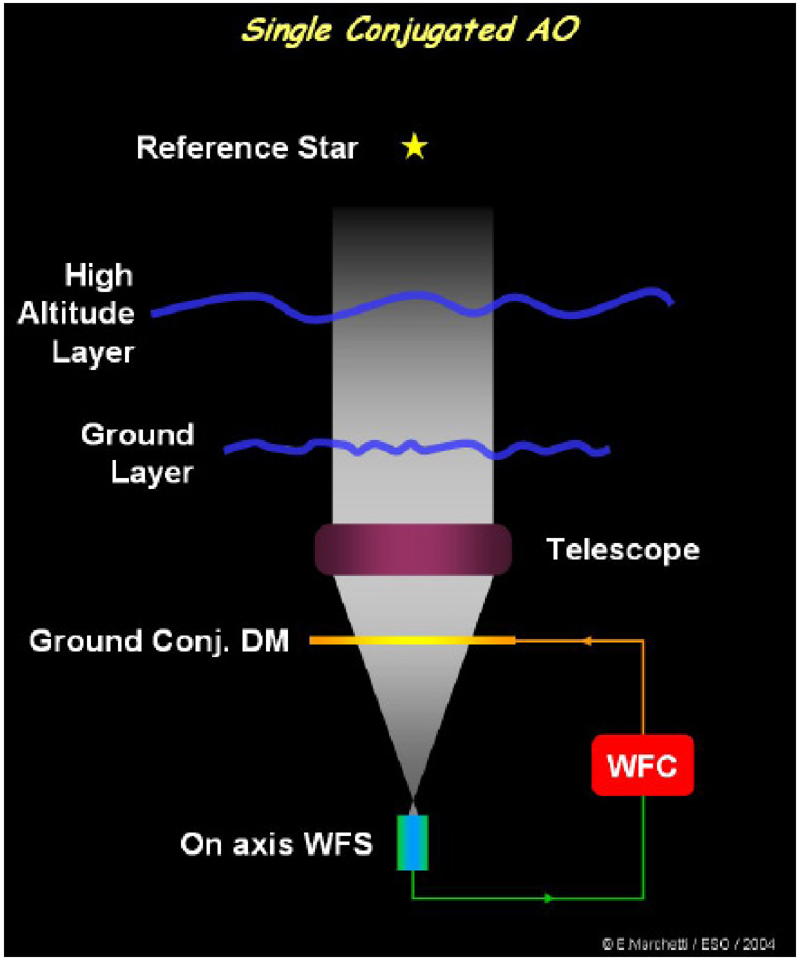
- From classical AO … to multi-conjugate AO (MCAO) ...

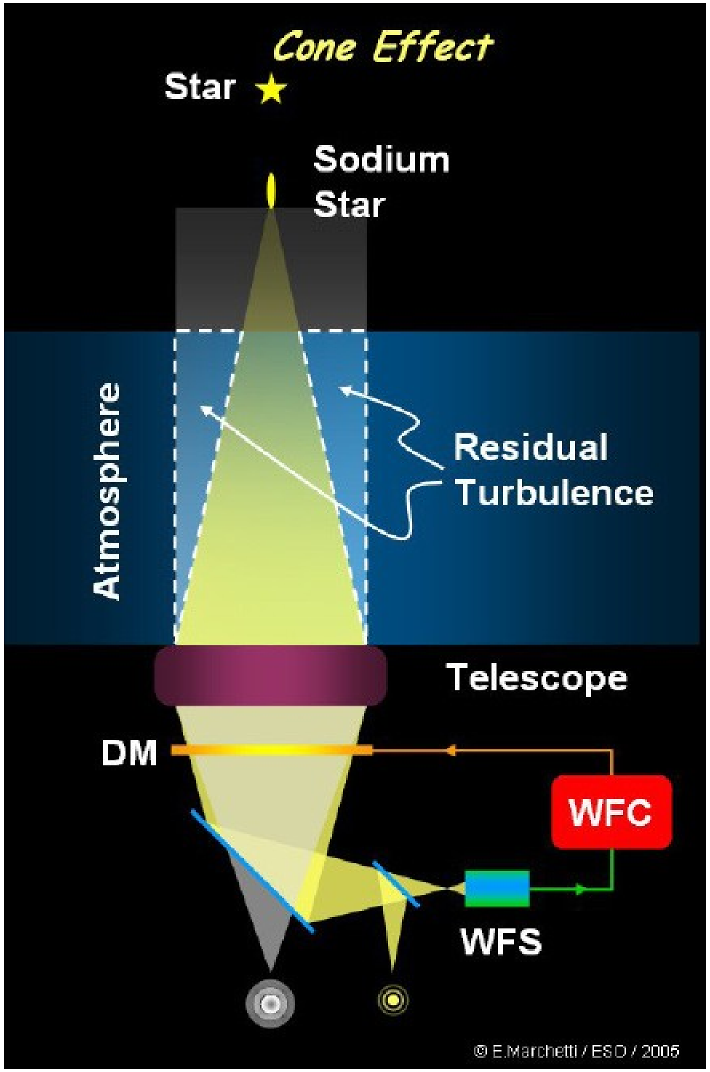




Adaptive optics flavors

- WFS needs a bright star … when not available => create your own !

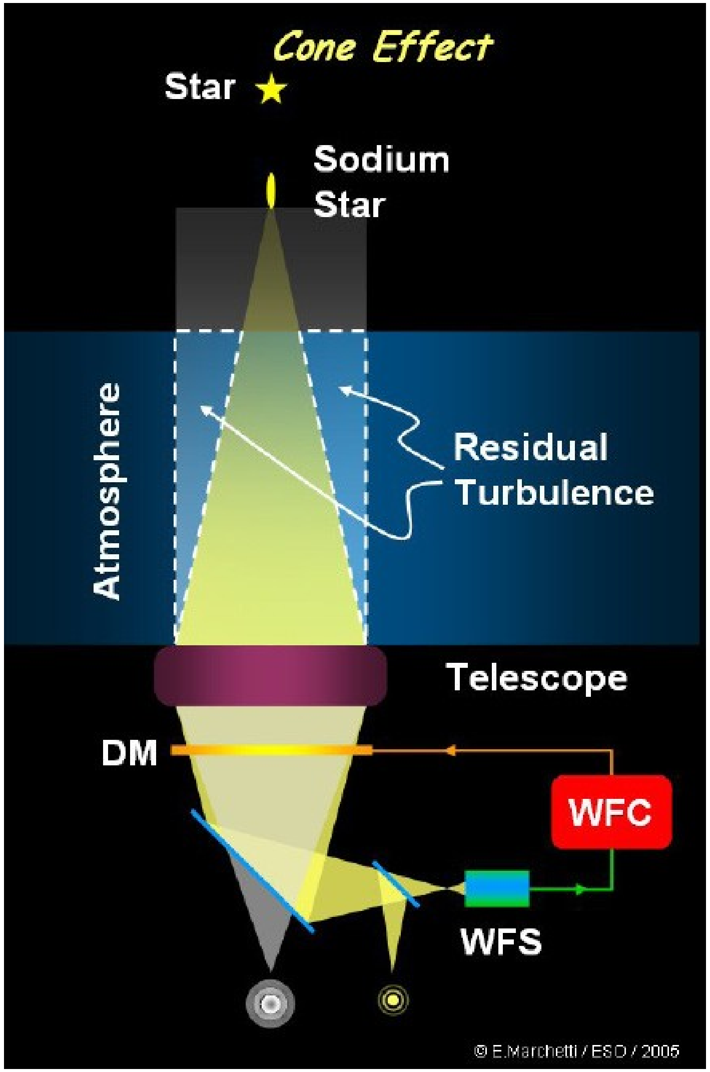


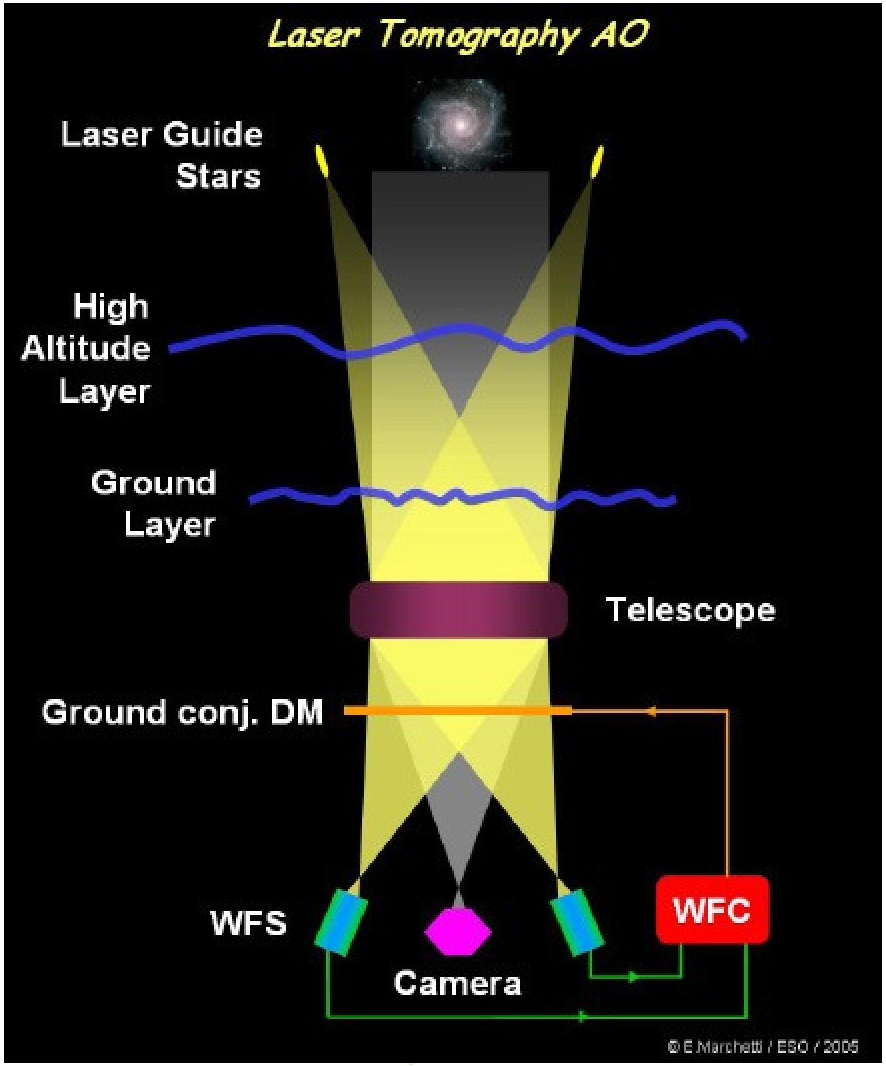


l'Observatoire de Paris — LESIA
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique
Durham University
MICROGATE
PLDA

Adaptive optics flavors

- From Laser guided AO… to Laser-tomography AO ...

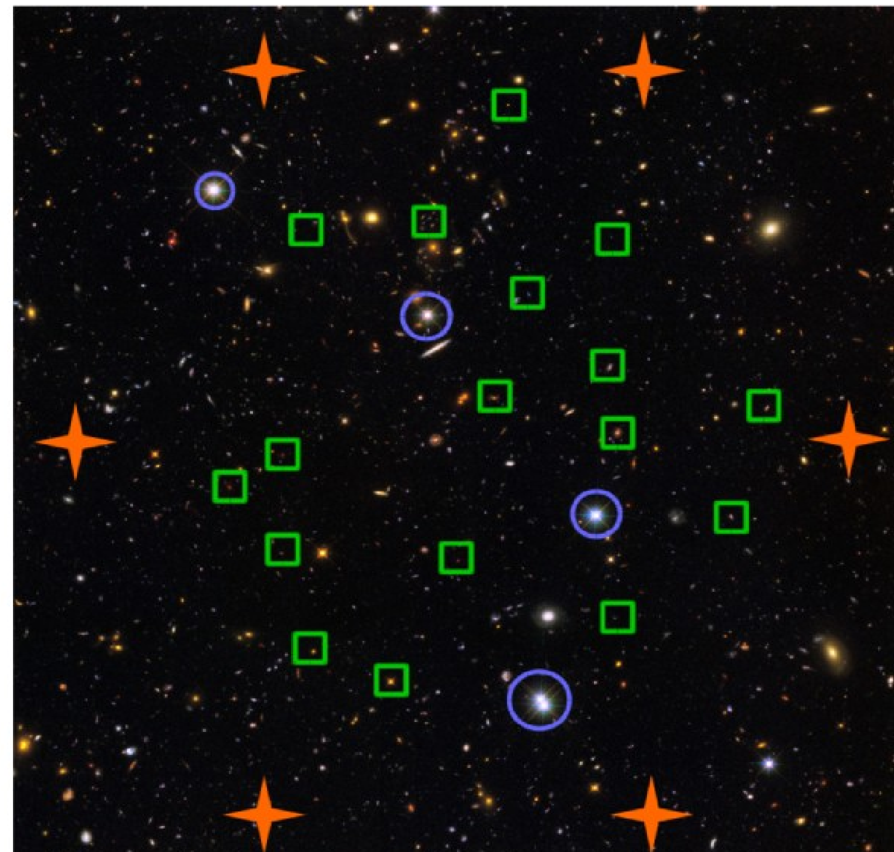




l'Observatoire de Paris — LESIA
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique
Durham University
MICROGATE
PLDA

Adaptive optics flavors

- multiple guide stars and multiple WF correctors

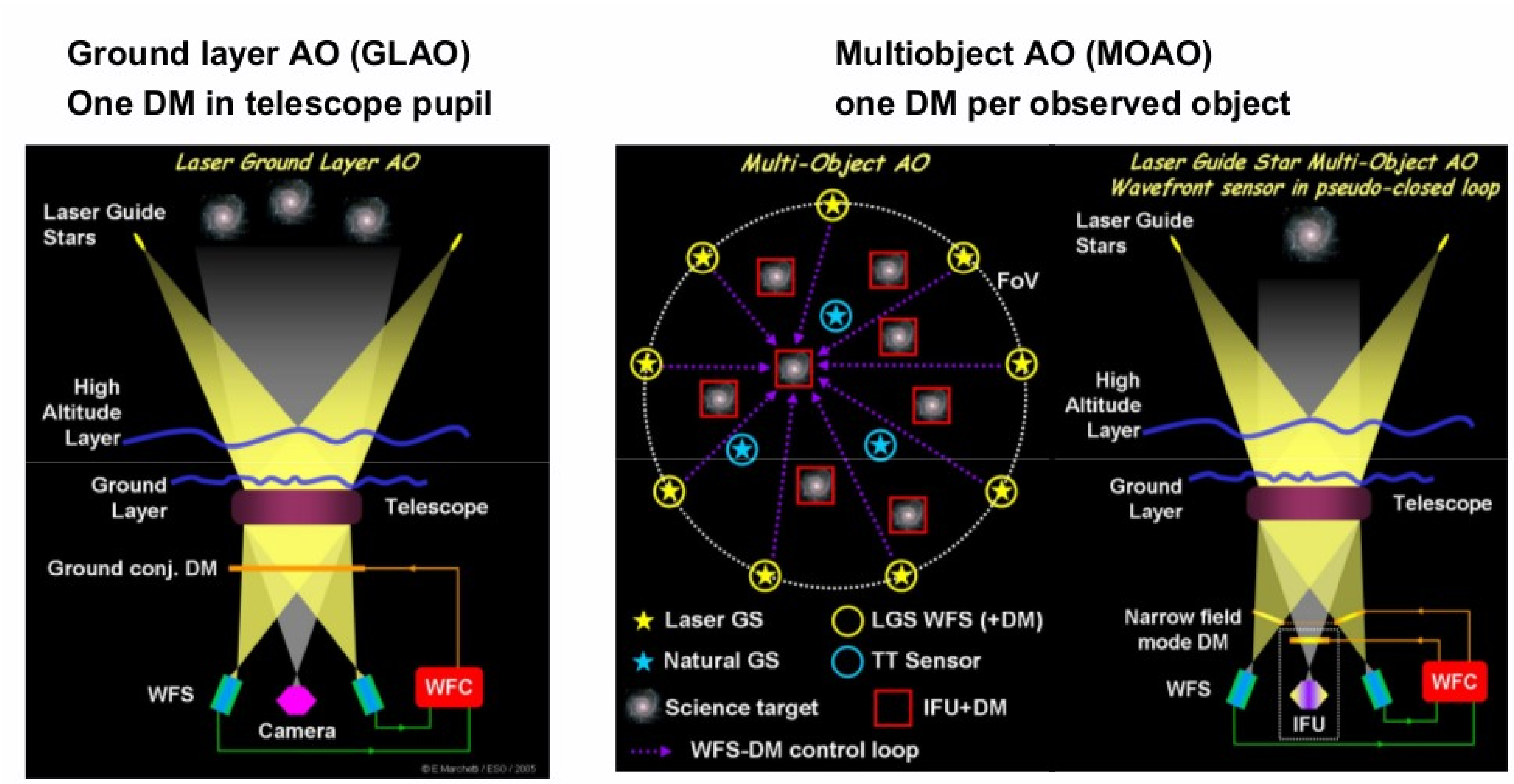


(c) Observing the GOODS South cosmological field with MOAO.

de Paris
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique
University
MICROGATE
PLDA

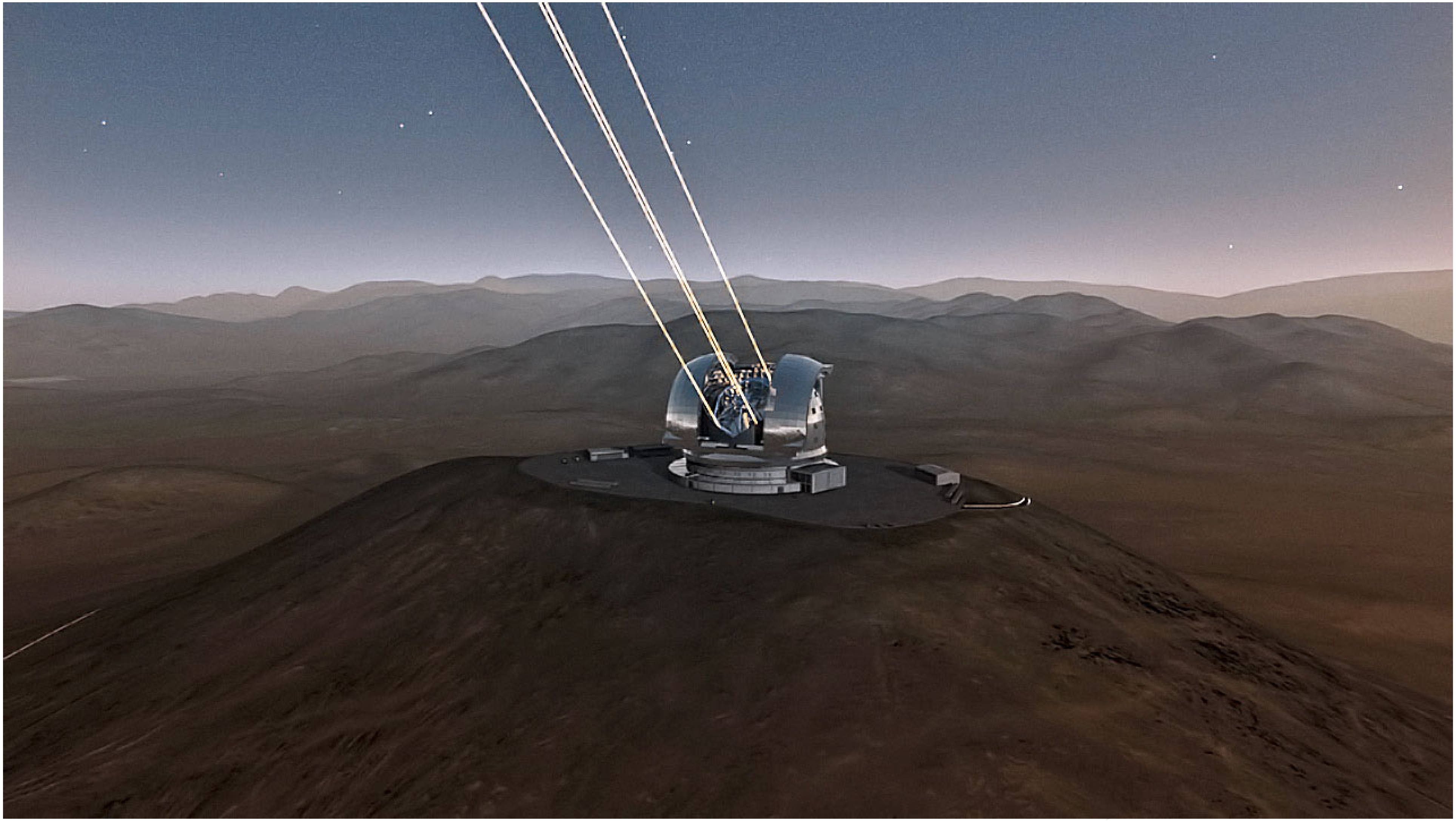
Adaptive optics flavors

- … to ground-layer AO … to multi-object AO (MOAO) !



E-ELT: an adaptive telescope

- … with multiple Laser guide stars ...

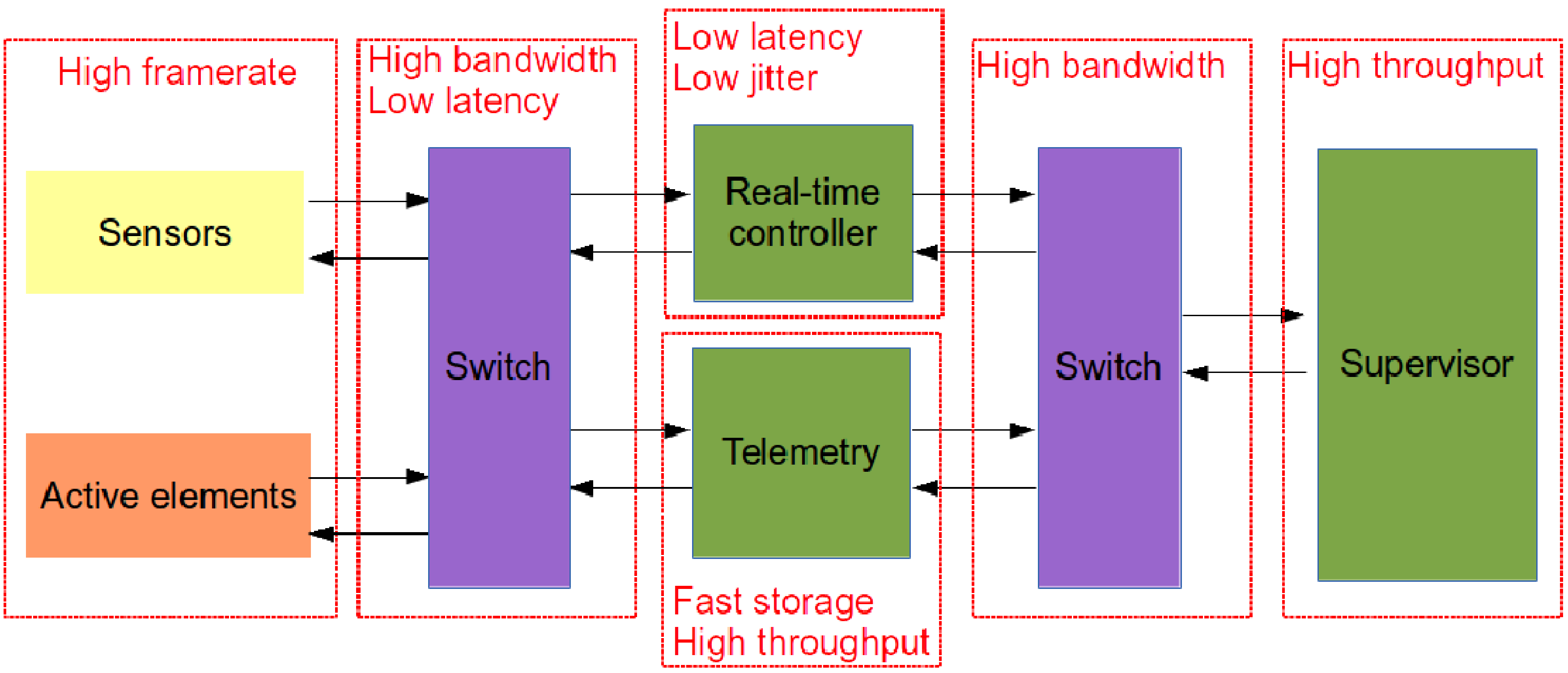


l'Observatoire de Paris — LESIA
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique
Durham University
MICROGATE
PLDA

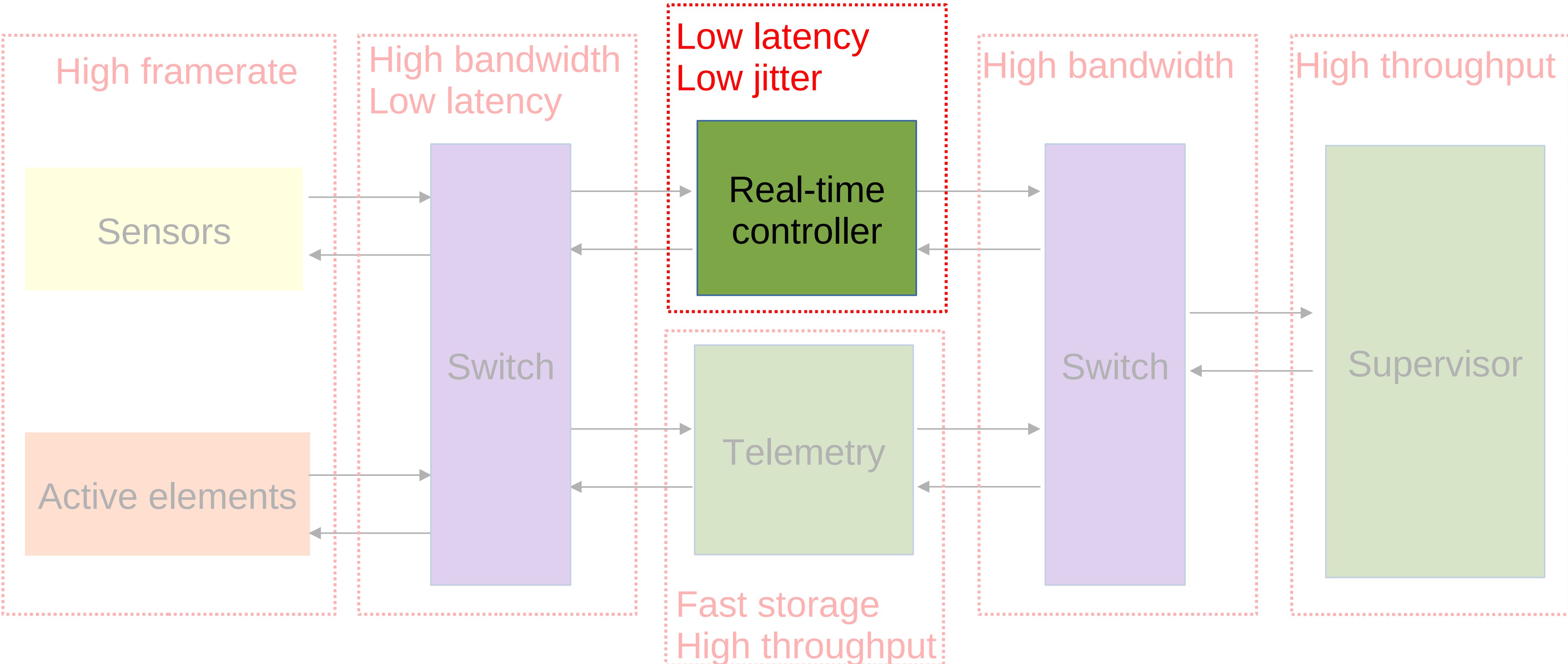
E-ELT: an adaptive telescope

- … and multiple technical challenges !
  - Telescope components (dome, mirror segments, etc ...)
  - Fast and low noise detectors
    - for science channel (in the IR domain) but also for AO (in the visible domain, for WFS)
  - New optical components
    - science instrument but also AO WFS
  - Compute technologies
    - for the **real-time control** (RTC) of the telescope and AO instruments
    - for the **simulation** (instrument design trade-off studies)
- Limited budget for overall instruments set : ~100M€ including staff effort for operations

AO RTC concept

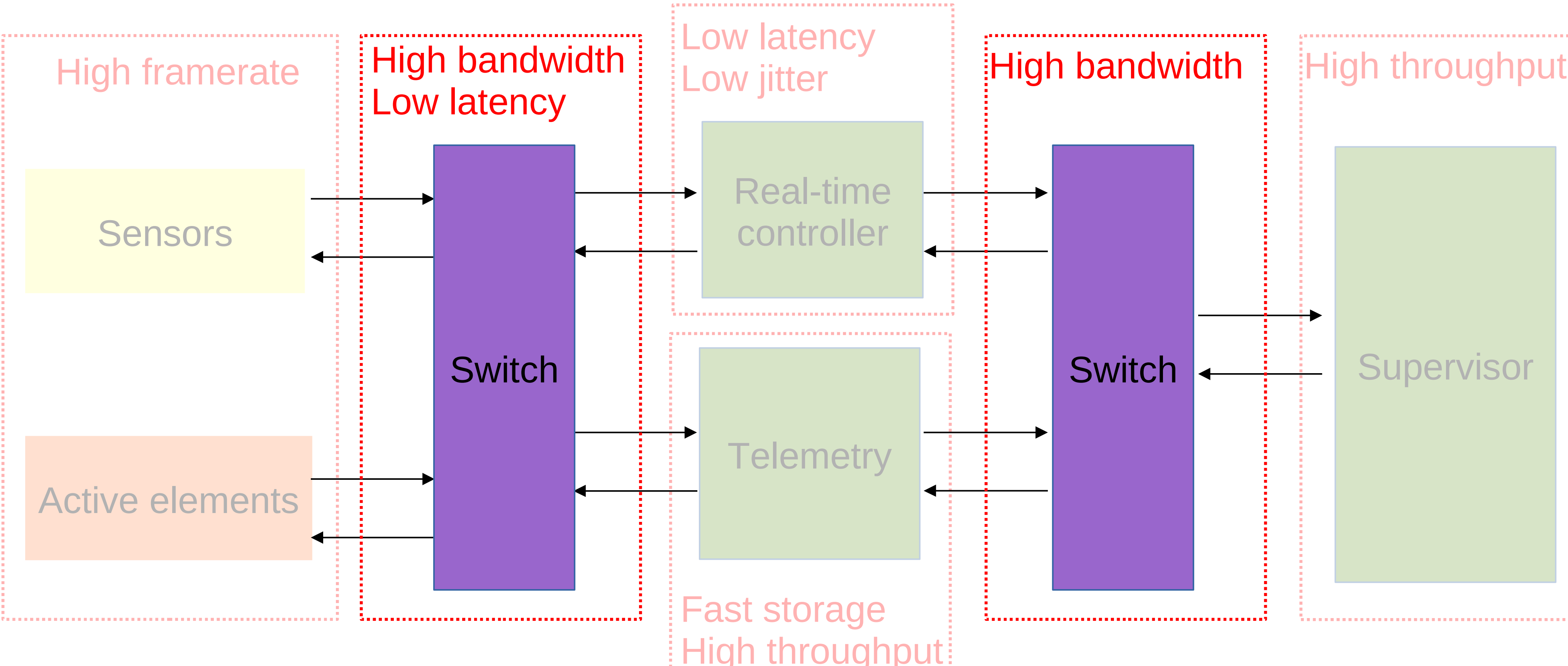


AO RTC concept : data pipeline



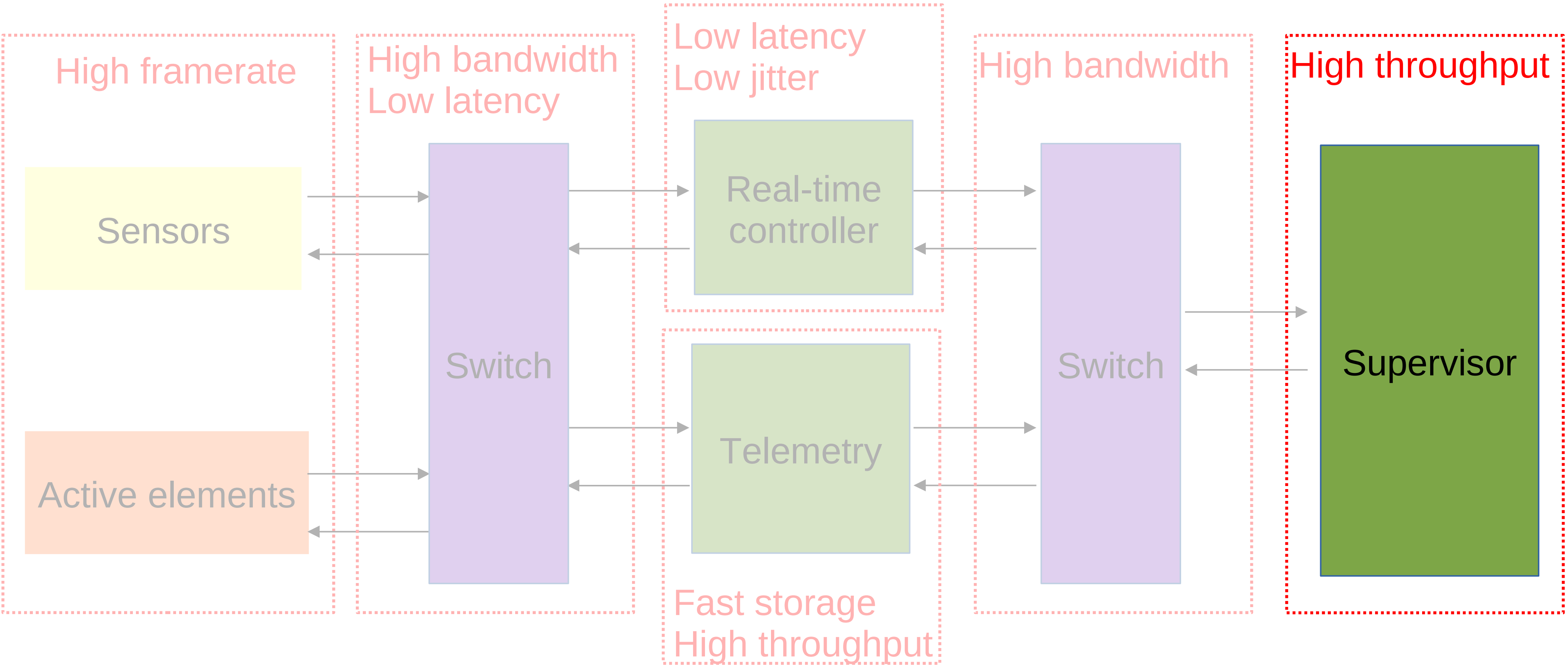
01/22/2016

AO RTC concept : smart interconnect



01/22/2016

AO RTC concept : supervisor



01/22/2016

Introduction to Green Flash

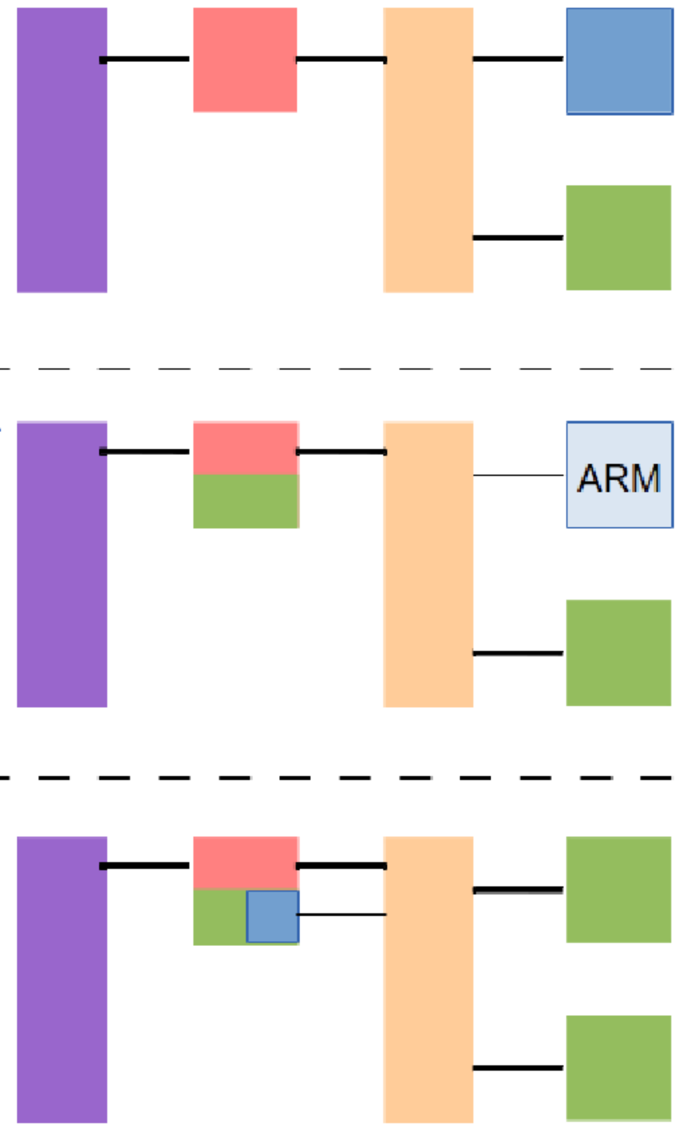
- Program objectives: 3 research axes
  - 2 technological developments and 1 validation study
- Real-time HPC using accelerators and smart interconnects
  - Assess the determinism of accelerators performance
  - Develop a smart interconnect strategy to cope for strong data transfer bandwidth constraints
- Energy efficient platform based on FPGA for HPC
  - Prototype a main board, based on FPGA SoC and PCIe Gen3
  - Cluster such boards and assess performance in terms of energy efficiency and determinism
- AO RTC prototyping and performance assessment
  - Assemble a full functionality prototype for a scalable AO RTC targeting the MAORY system
  - Compare off-the-shelf solutions based on accelerators and new FPGA-based concept

Assessing new HPC concepts

Current trend in heterogeneous HPC

Accelerators, ARM64 & smart interconnect

FPGA-based microserver



Network fabric
Network controller
Local interconnect (PCIe)
Host CPU
Accelerator

Dynamically reconfigurable NIC
Expandable programming environment on FPGA
Features accessible in middleware

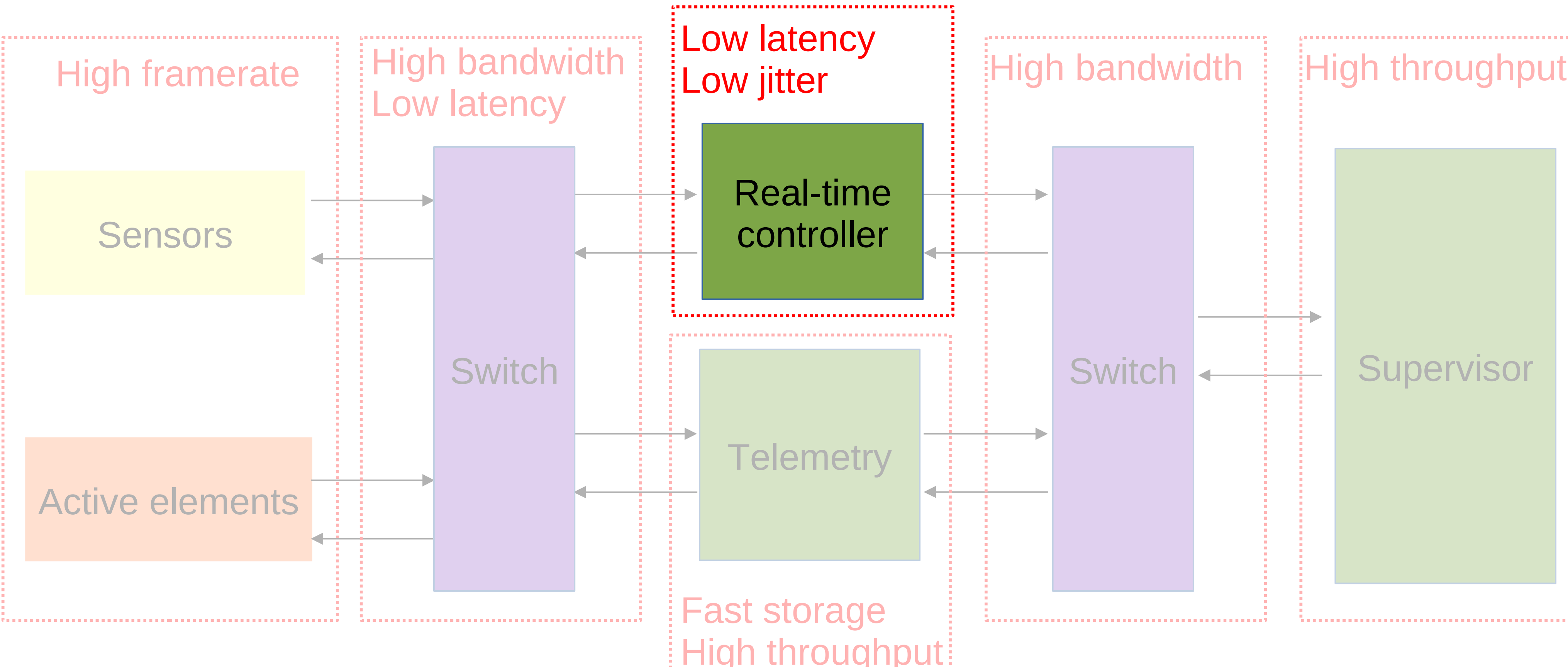
Main board with FPGA + ARM HPS
Integrated ecosystem for interconnect handling and accelerators support
Scalable environment

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

Green Flash project

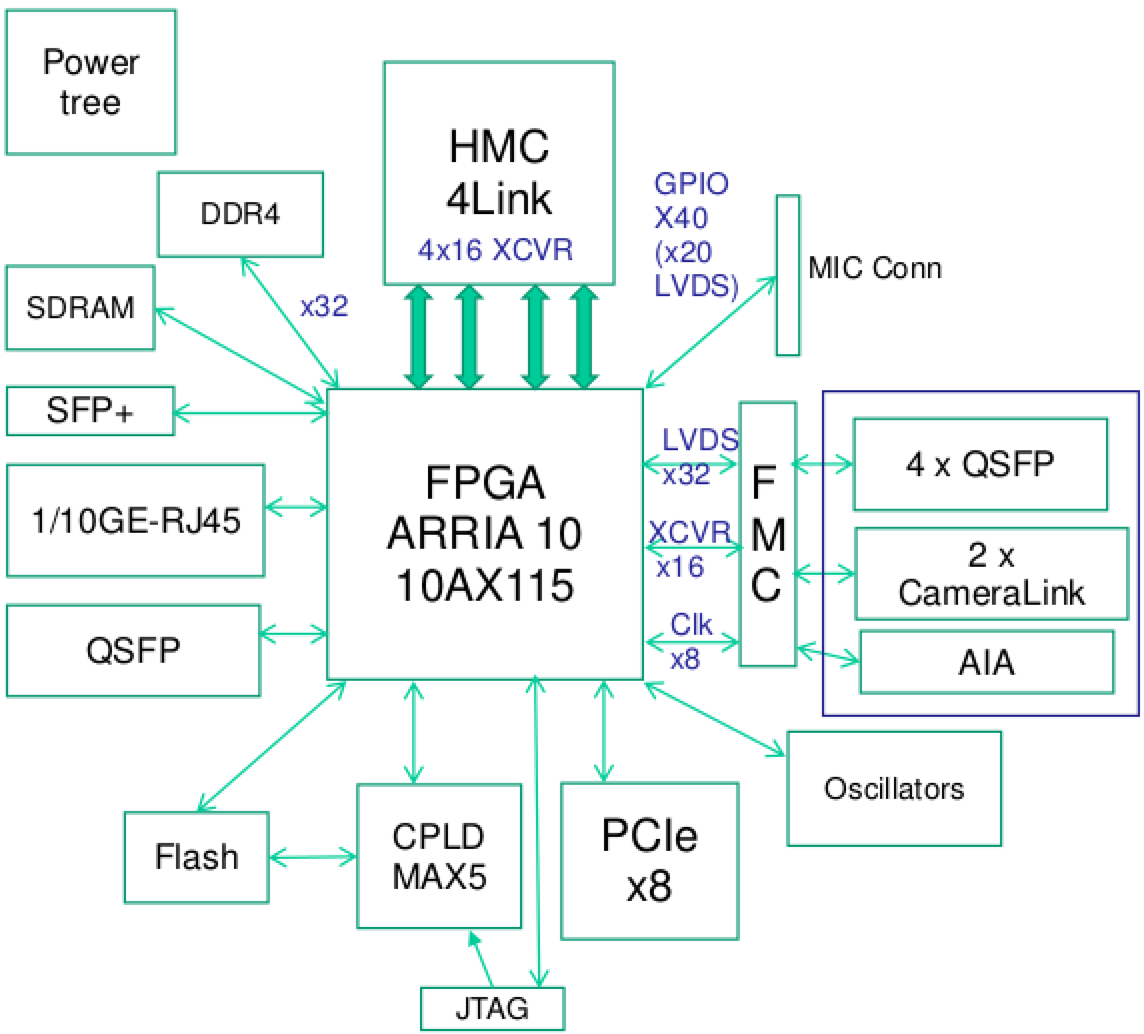
- EU funded Partners
  - 2 academic partners
    - LESIA, Observatoire de Paris, P.I. Damien G.
    - CfAI, University of Durham
  - 2 industrial partners
    - Microgate : Italian SME designing FPGA solutions for various applications (including astronomical AO)
    - PLDA: French SME developing FPGA solutions (mostly IP cores, world leader in PCIe IPs
- KAUST – ECRC as external partner

AO RTC concept : data pipeline



01/22/2016

FPGA solution : µXcomp



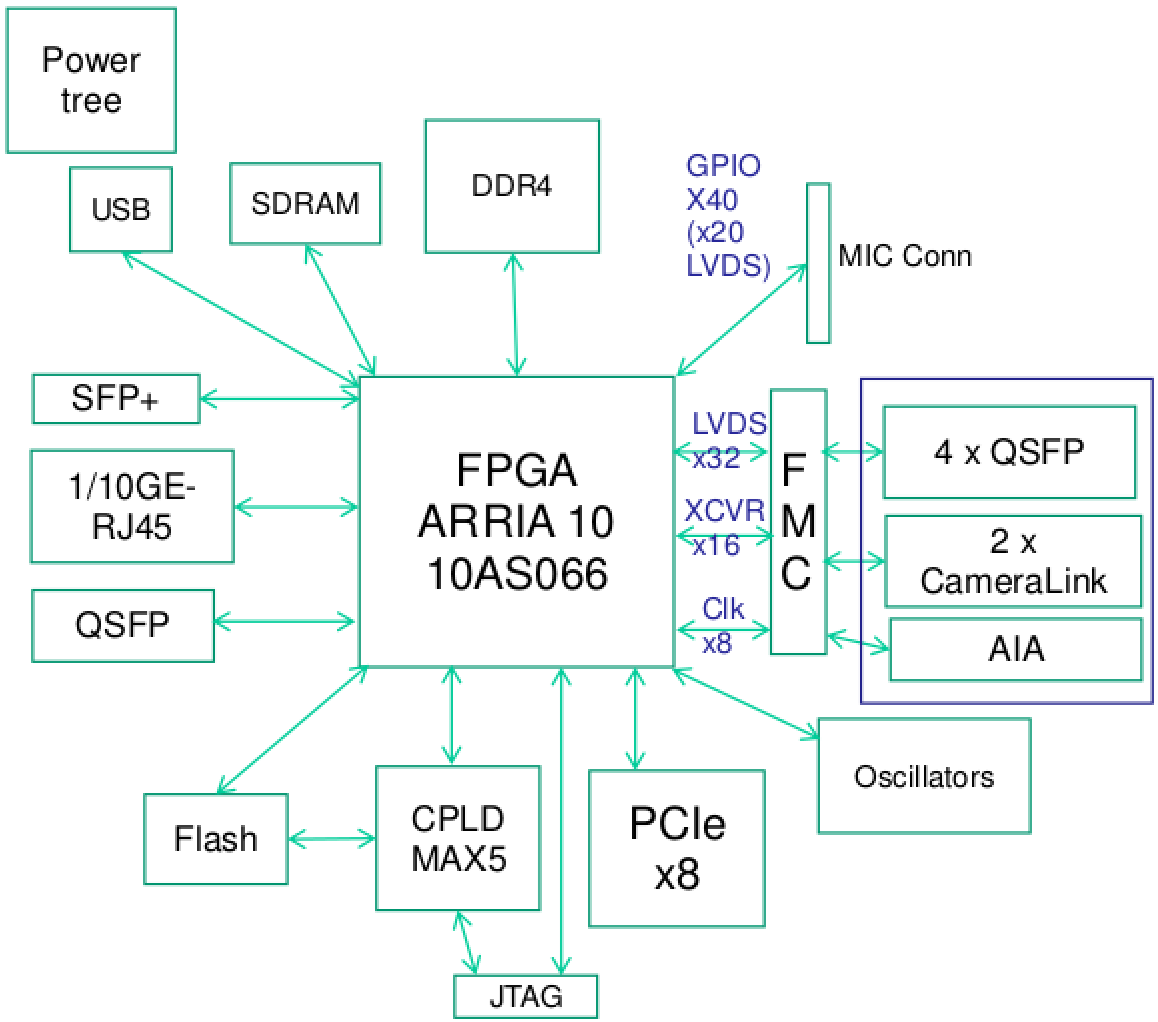
Based on ARRIA 10AX115:

- 1518 DSP blocks
- 6.6MB int. RAM
- 96 XCVR

Board features:

- Optimized for **heavy deterministic computation** in floating-point
- **Large Bandwidth between HMC and FPGA** - 4 links 16 lanes/link up to 15Gbps/lane = 120GB/s bidirectional
- Extremely **low jitter**
- More **power efficient** compared to GPUs
- Offers a lot of different interfaces on board or via the FMC connector and extension cards

FPGA solution : µXlink



ARRIA 10AS066 SoC:

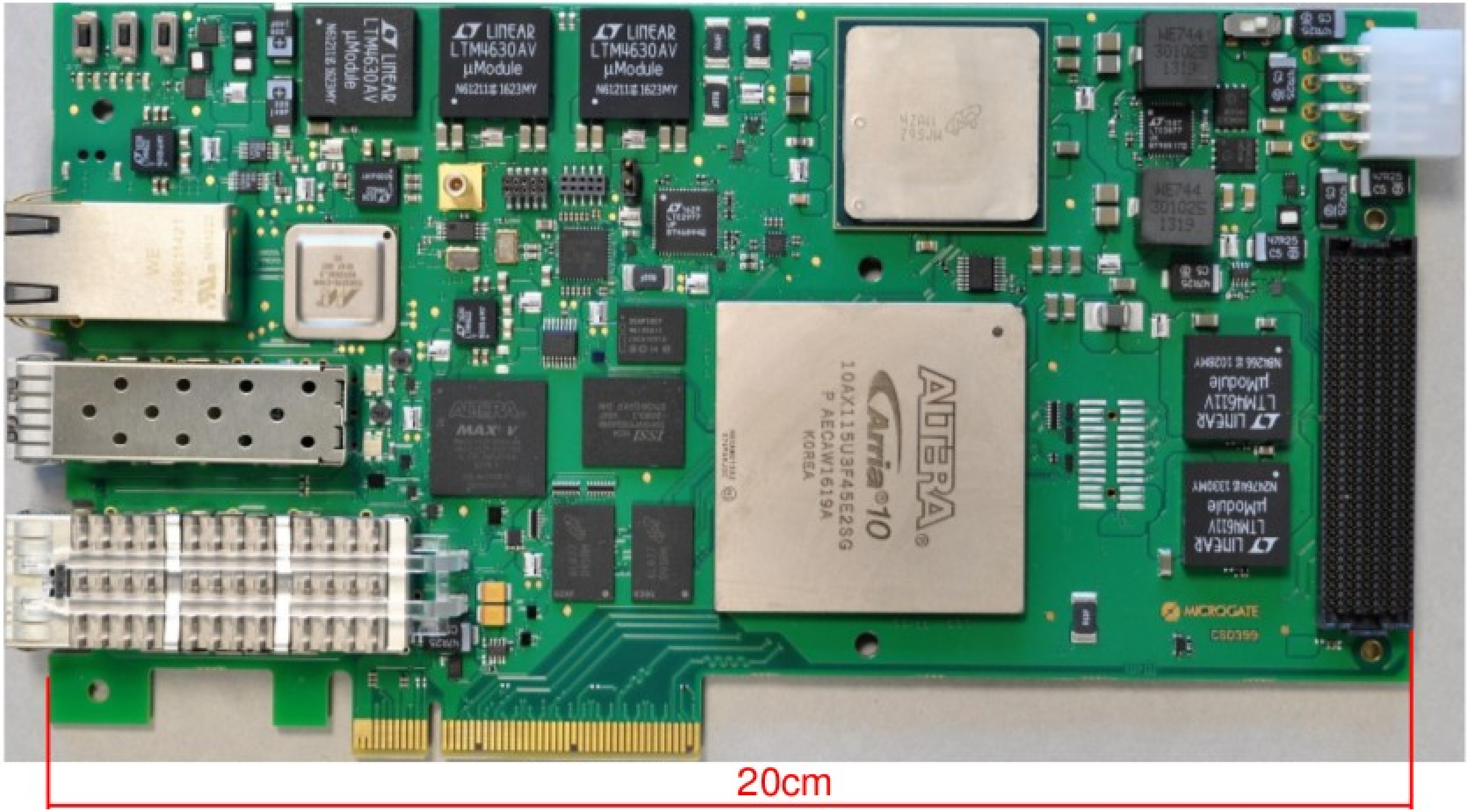
- 1.5GHz ARM dual-core Cortex-A9 on-chip processor
- 1855 DSP blocks
- 5.2MB int. RAM
- max. 48 XCVR

Board features:

- ARM **embedded processor** for **stand-alone** real-time box
- Powerful **PCIe root port** because of ARM and OS
- Management of accelerator cards on the PCIe interface
- Running control software using a full OS (e.g. Linux)
- Easy implementation of different communication protocols
- Offers a lot of different interfaces on board or via the FMC connector and extension cards

FPGA solutions: status

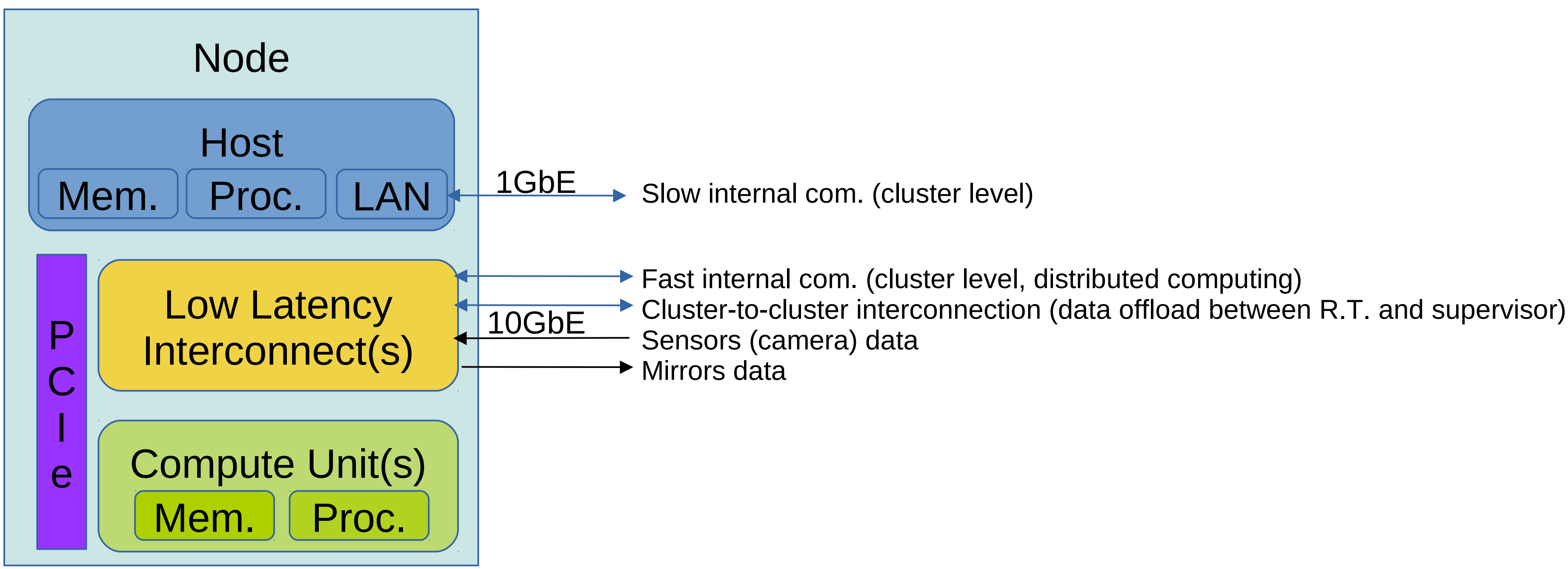
The first prototype of the two FPGA boards, the µXComp board is manufactured and is currently under test. After the validation of the interfaces and the communication between FPGA and HMC some more boards of this type will be produced and made available to the team.



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)

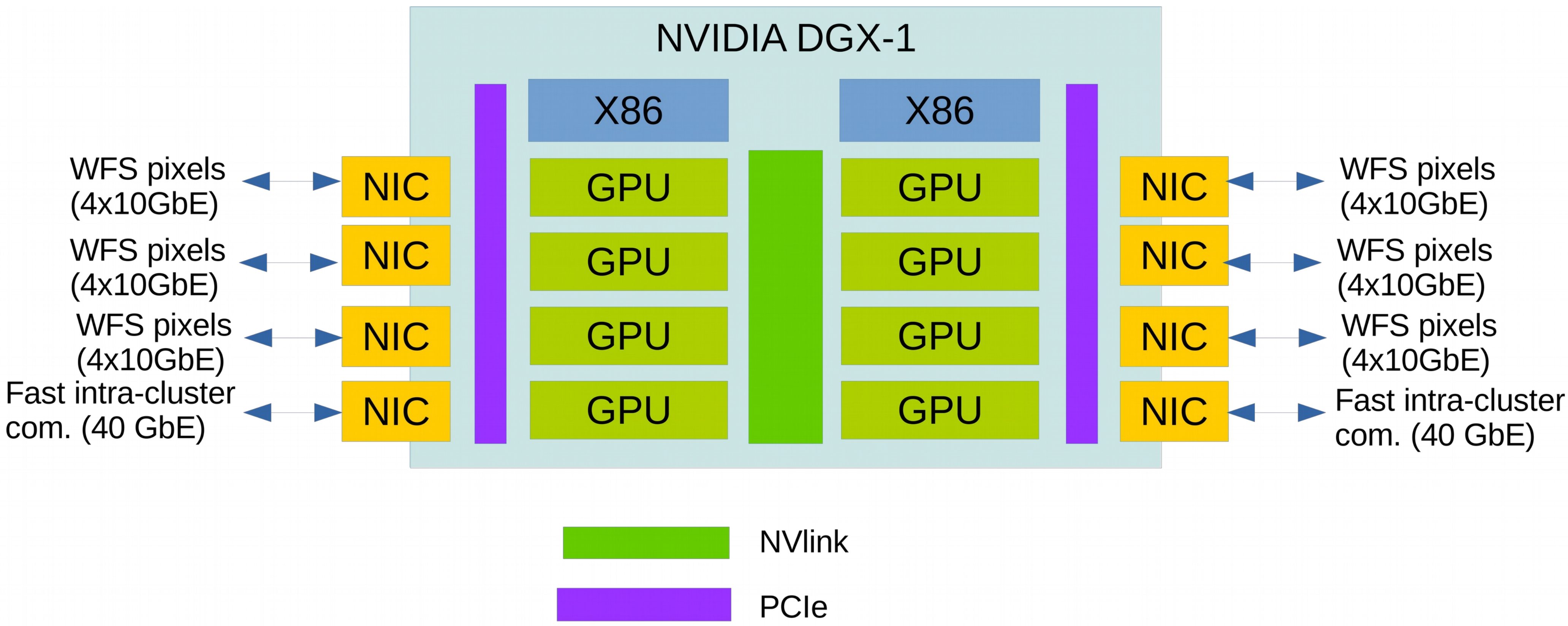
For each application, nodes are interconnected into a cluster.
For the full featured prototype, the two clusters are interconnected



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

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

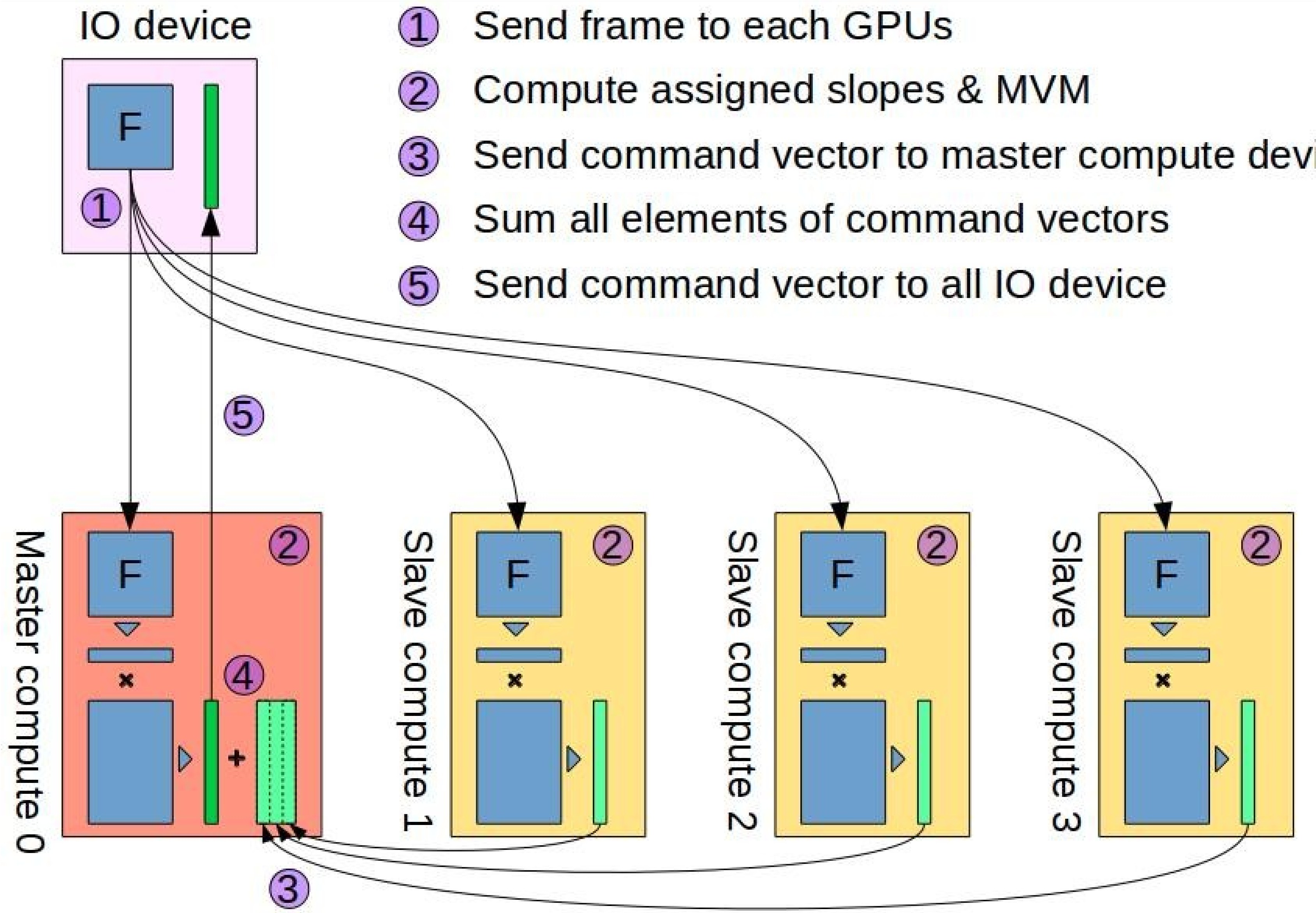
	K20C	K40	K80	P100
B_{theo}	208	288	240 (x2)	732
$B_{no\ ECC}$	175 (84%)	236 (82%)	200 (x2, 83%)	460 (62%)
B_{ECC}	150 (72%)	208 (72%)	173 (x2, 72%)	460 (62%)

Number of GPUs required

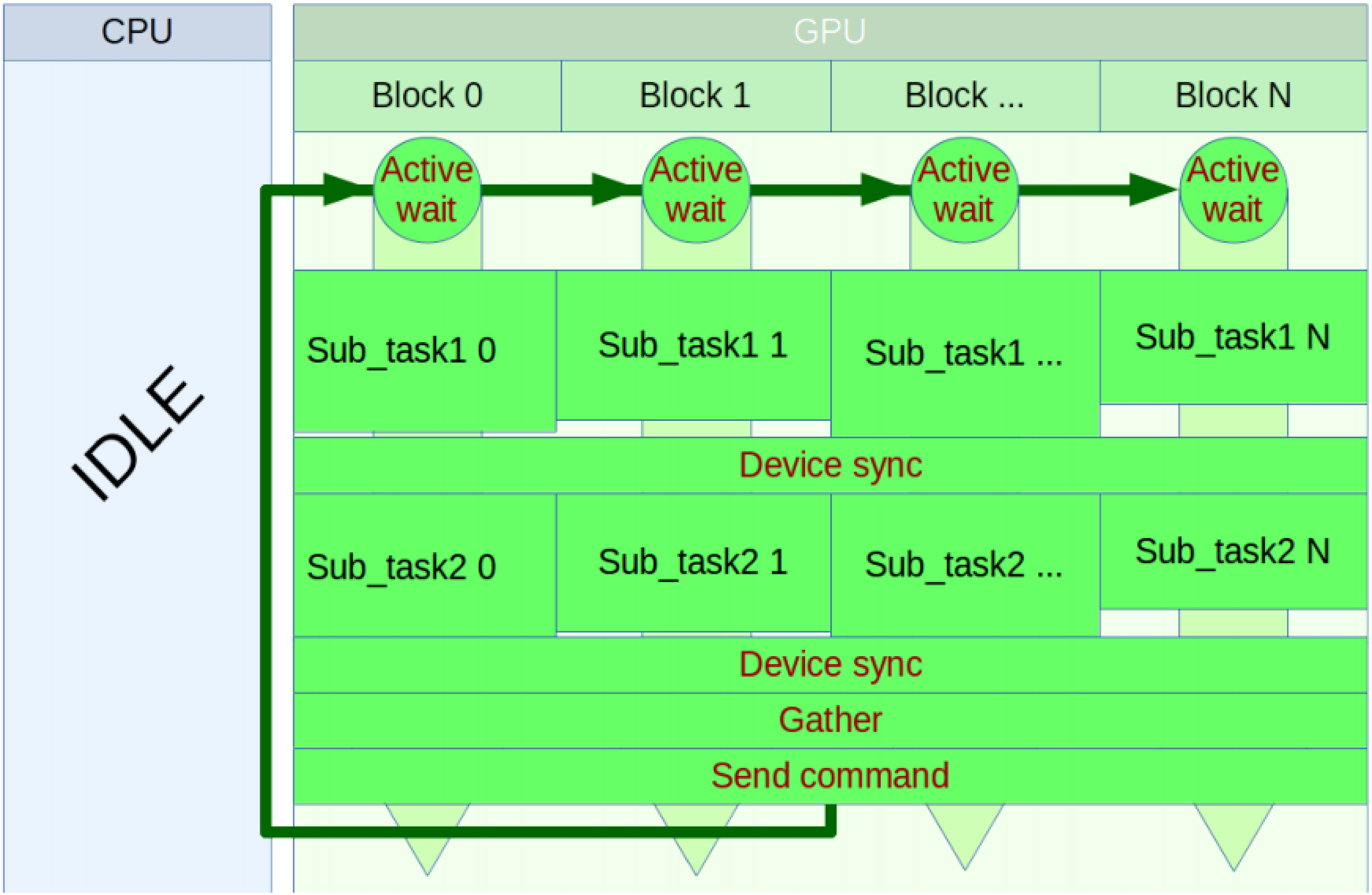
ECC	K20C	K40	K80	P100
Off	12	9	6	5
On	14	10	6	5

Multi-GPU prototype

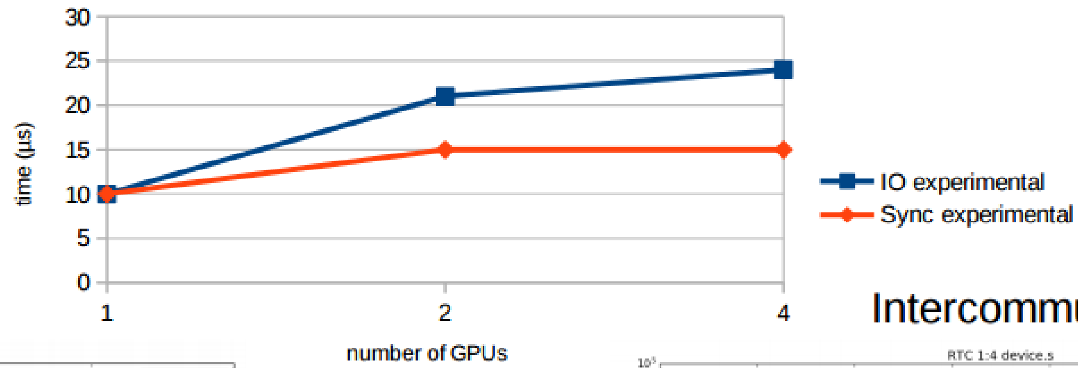
1. Send frame to each GPUs
2. Compute assigned slopes & MVM
3. Send command vector to master compute device
4. Sum all elements of command vectors
5. Send command vector to all IO device



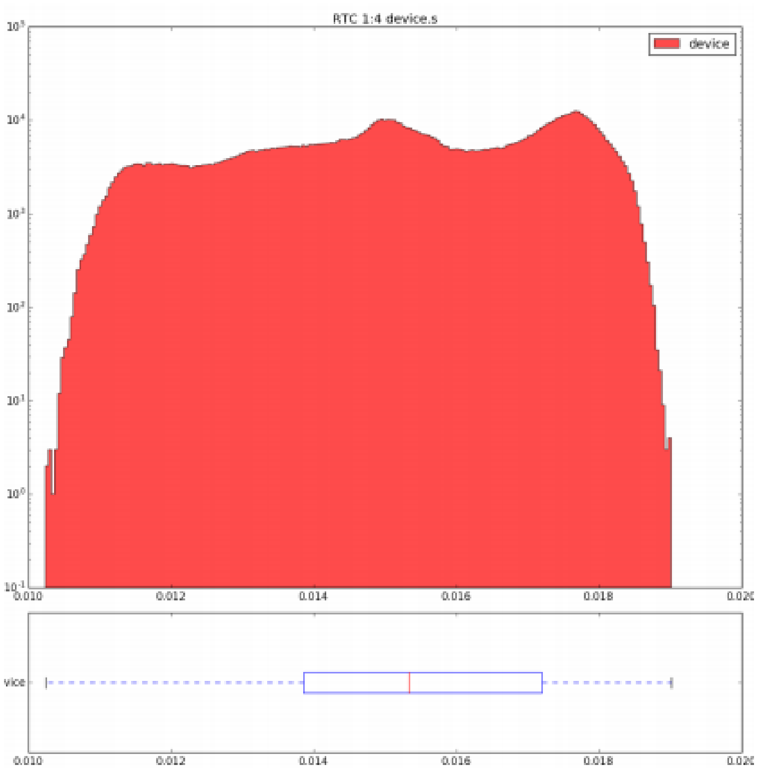
Persistent kernel implementation



Persistent kernel implementation

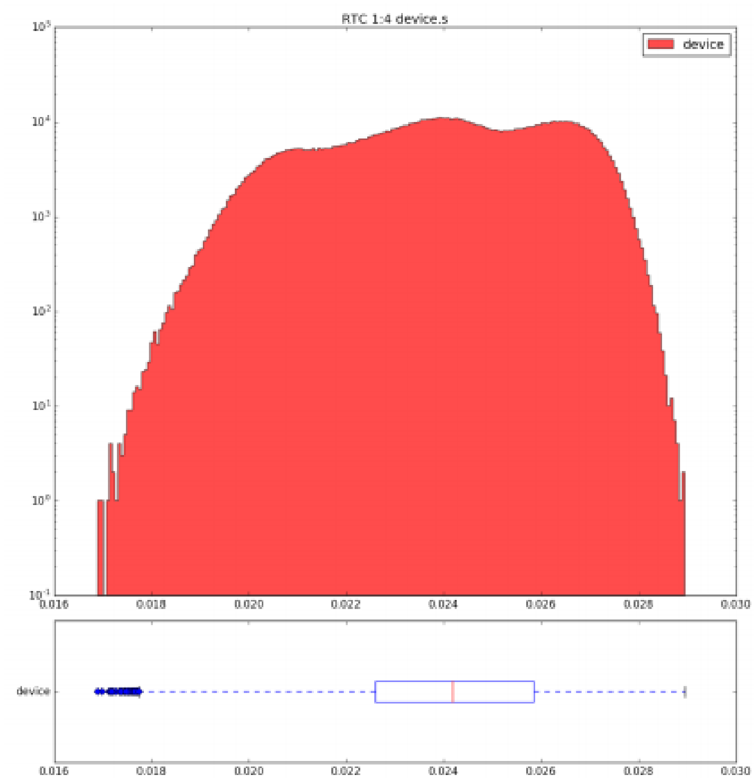


Synchronize jitter



Average : 15µs Jitter : 8.8µs

Intercommunication jitter



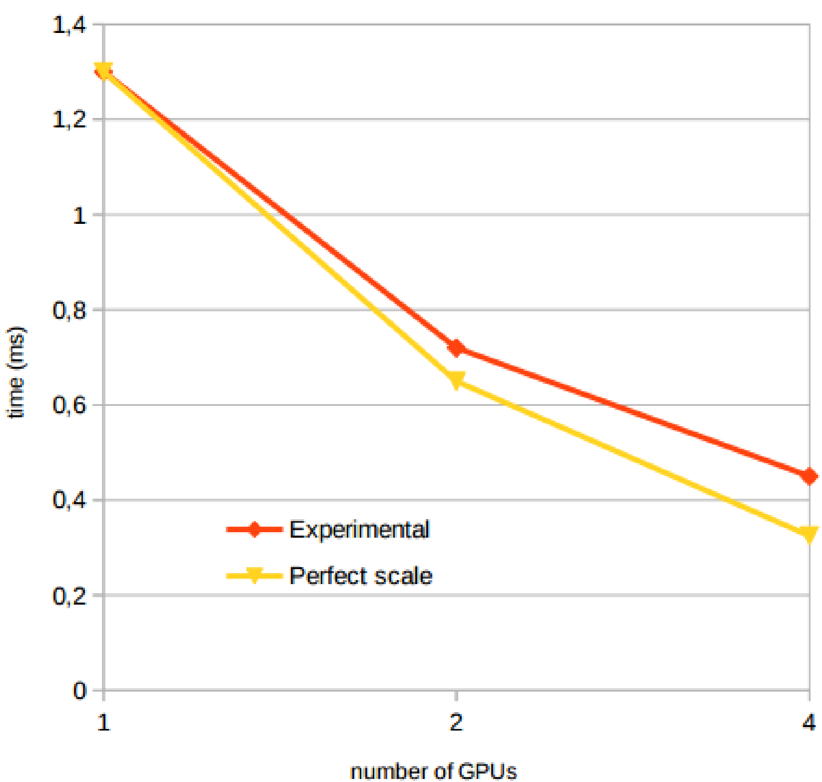
Average : 24µs Jitter : 12µs

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

Persistent kernel implementation

Strong scalability

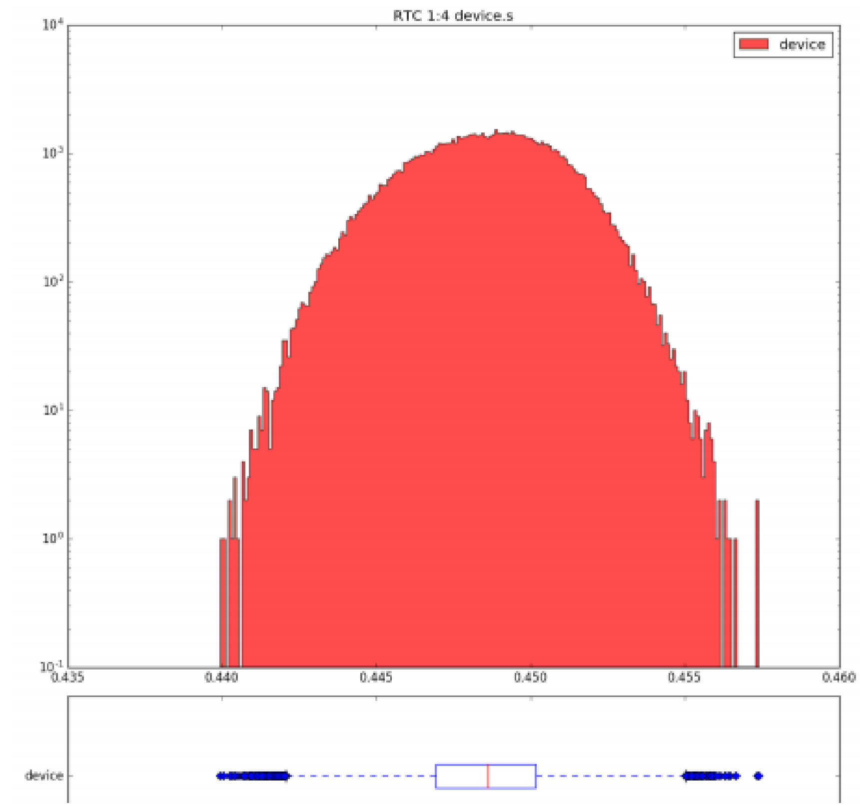
Constant case with **10,048 slopes x 15,000 commands**



Histogram

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

Average : 0.45ms   Jitter : 17µs

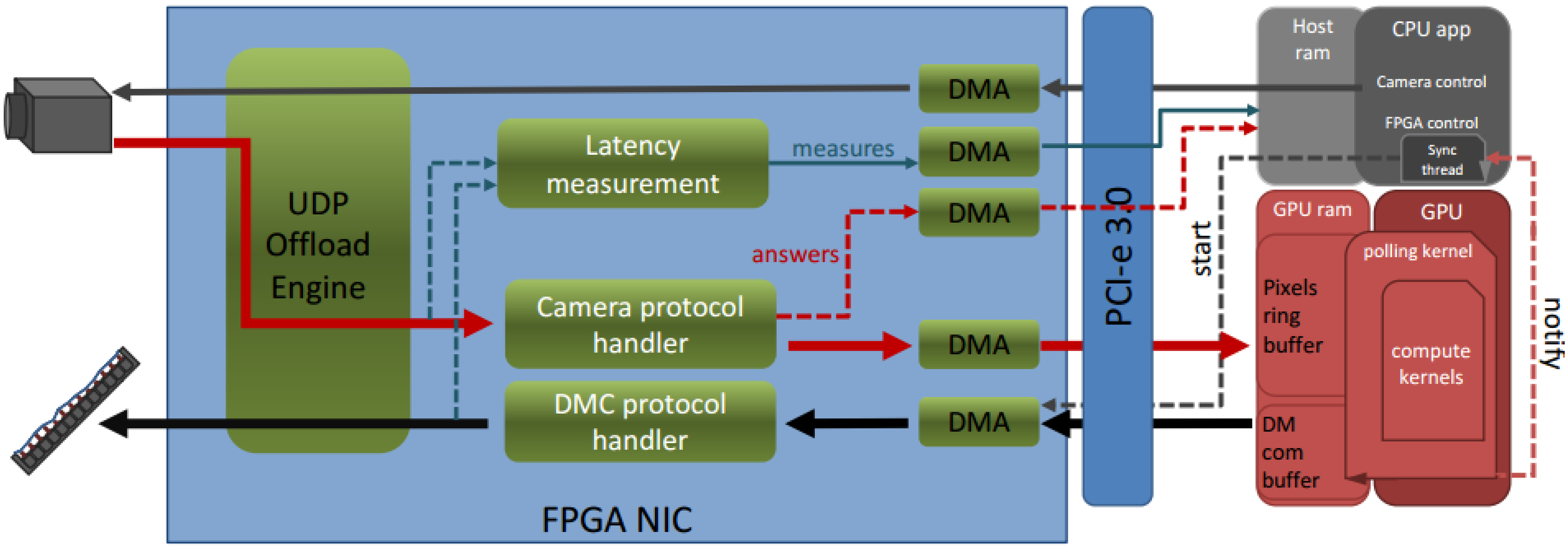


l'Observatoire de Paris — LESIA
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique
Durham University
MICROGATE
PLDA

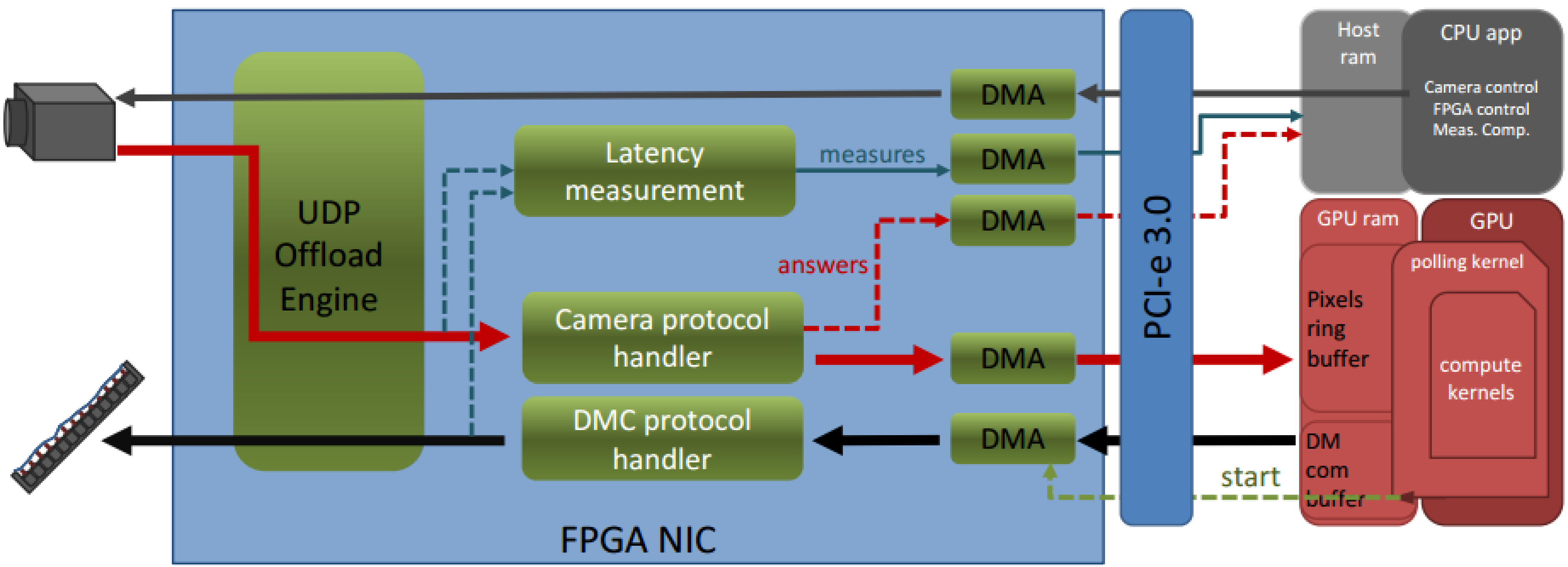
Data acquisition

FPGA writes/reads directly to/from GPU memory

Using only writes would be better though



FPGA/GPU optimized sync.

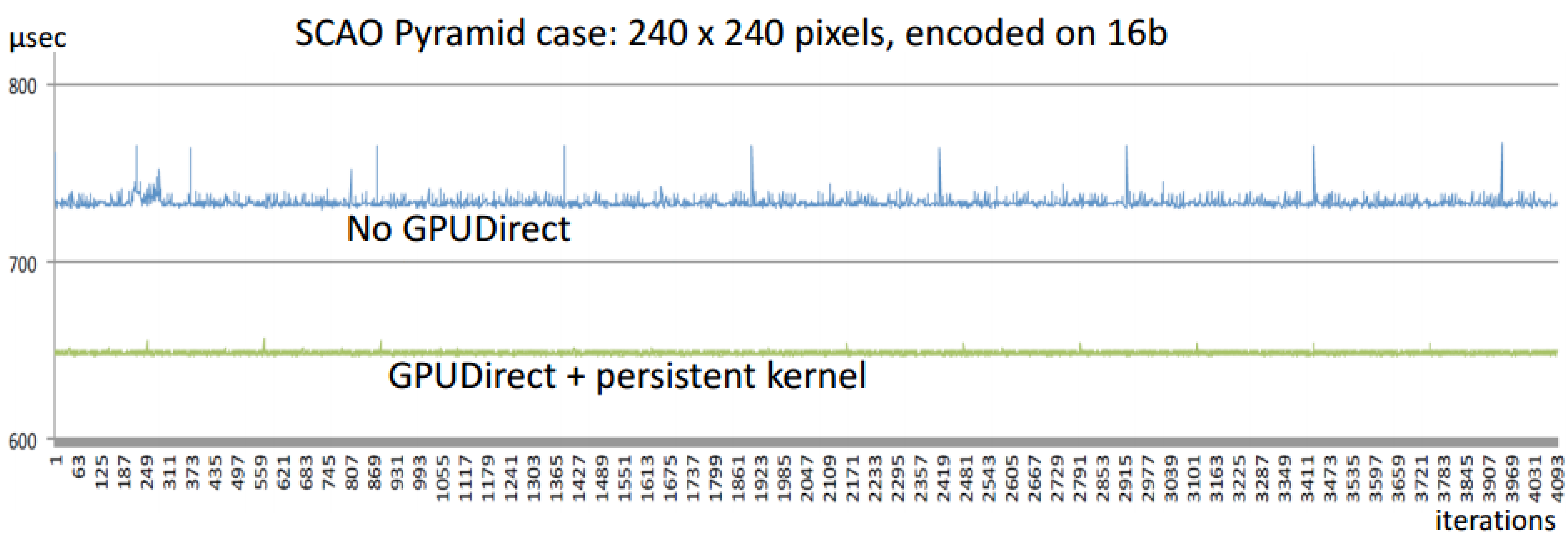


Little to no improvements, but CPU free for other kind of computations

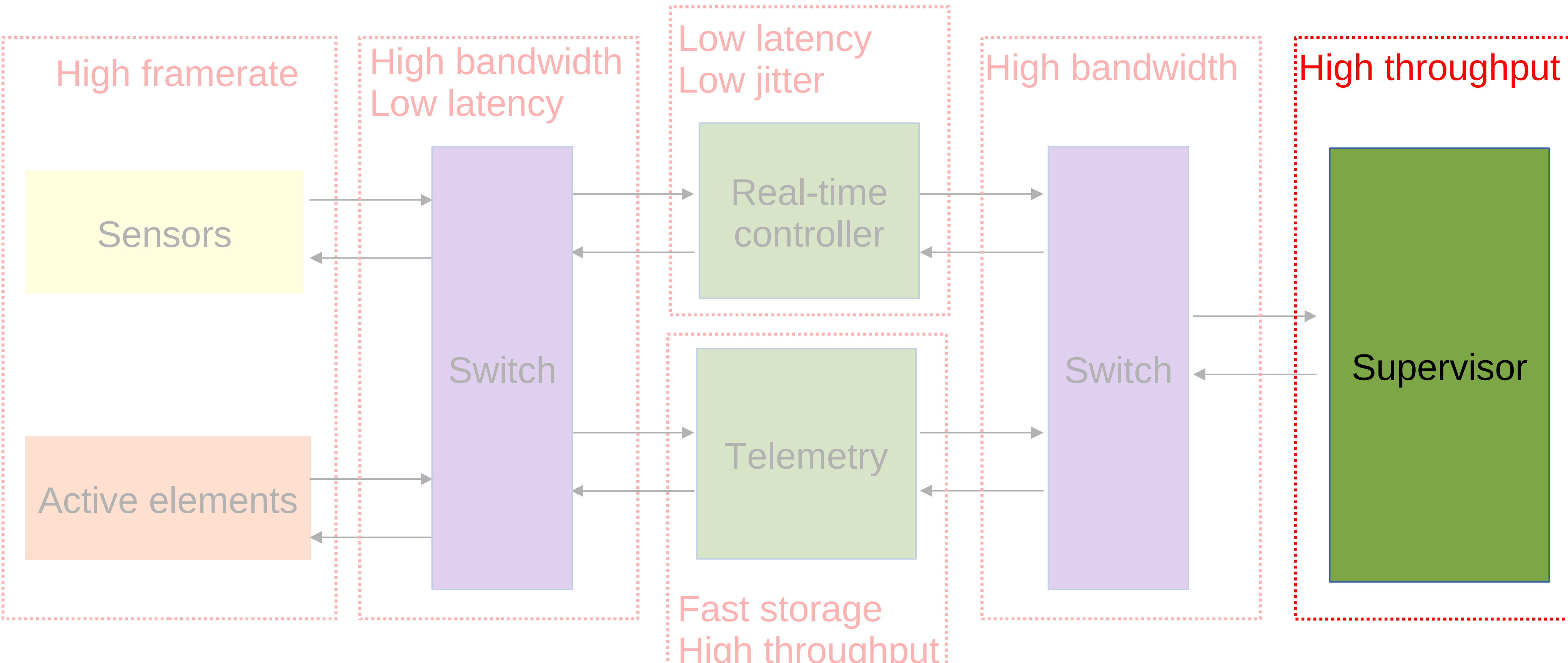
Data acquisition + persistent kernels

FPGA PLDA XPressG5
GPU Tesla C2070
OS Debian wheezy

Camera EVT HS-2000M
10GbE network

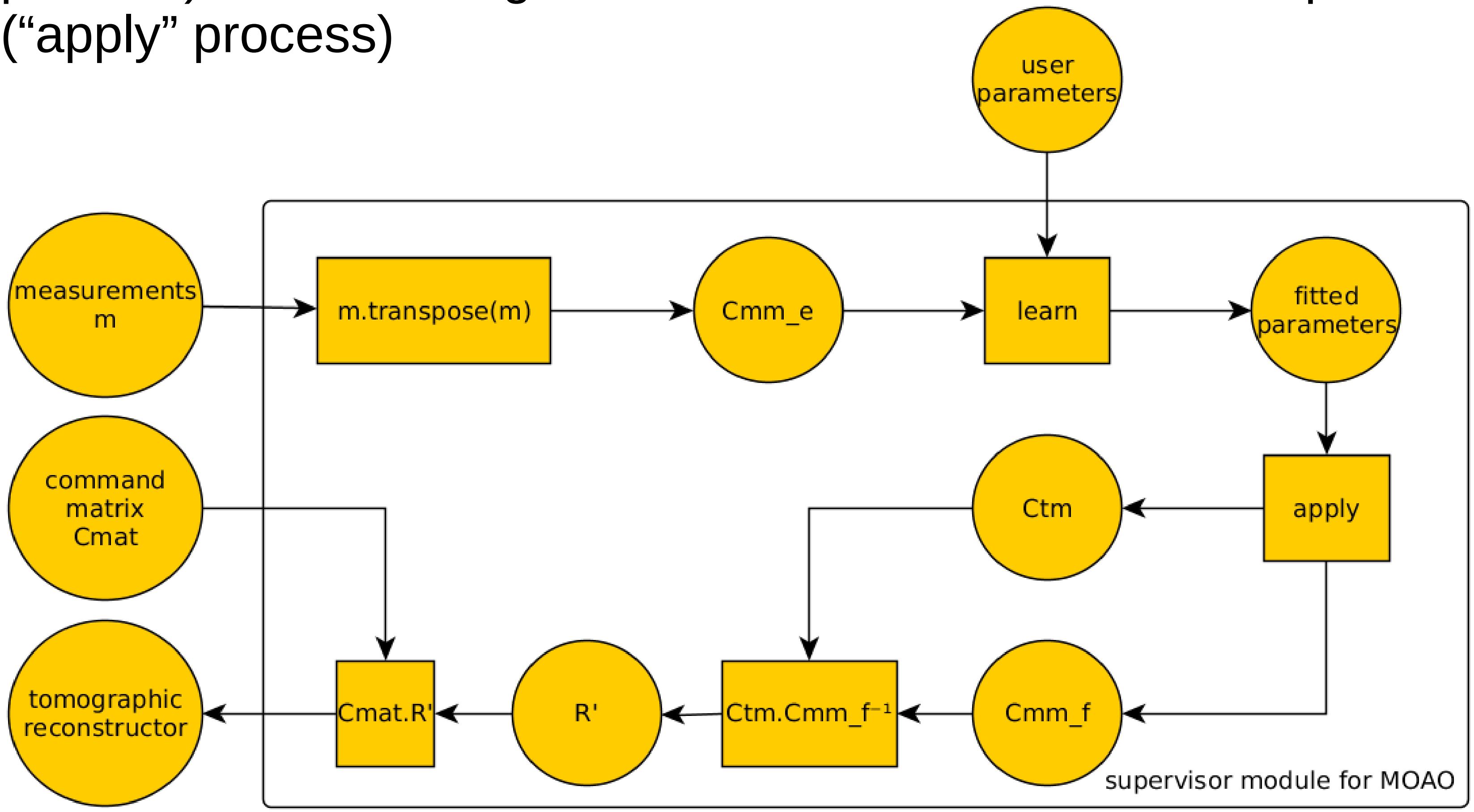


AO RTC concept : supervisor

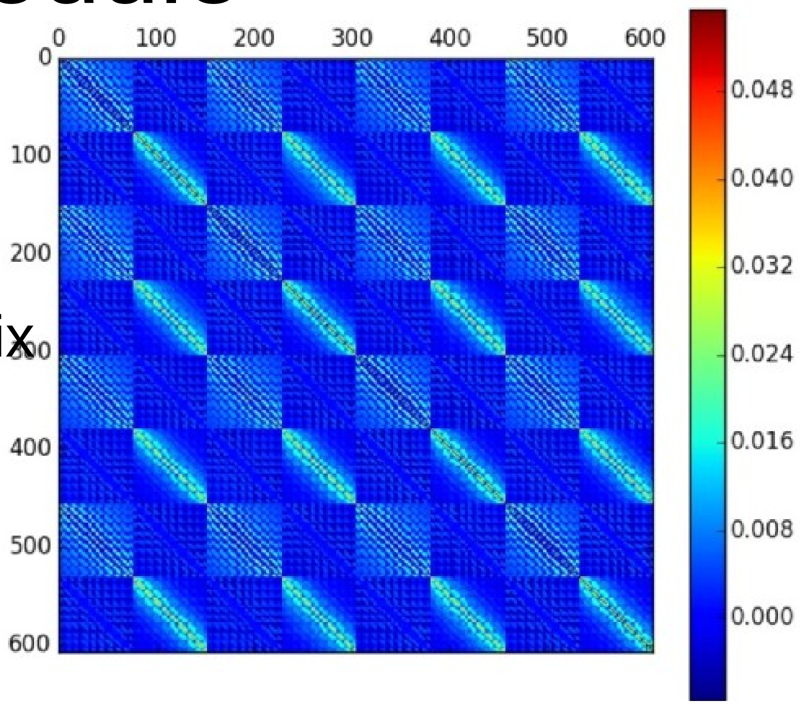


Loop supervision module

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



Loop supervision module

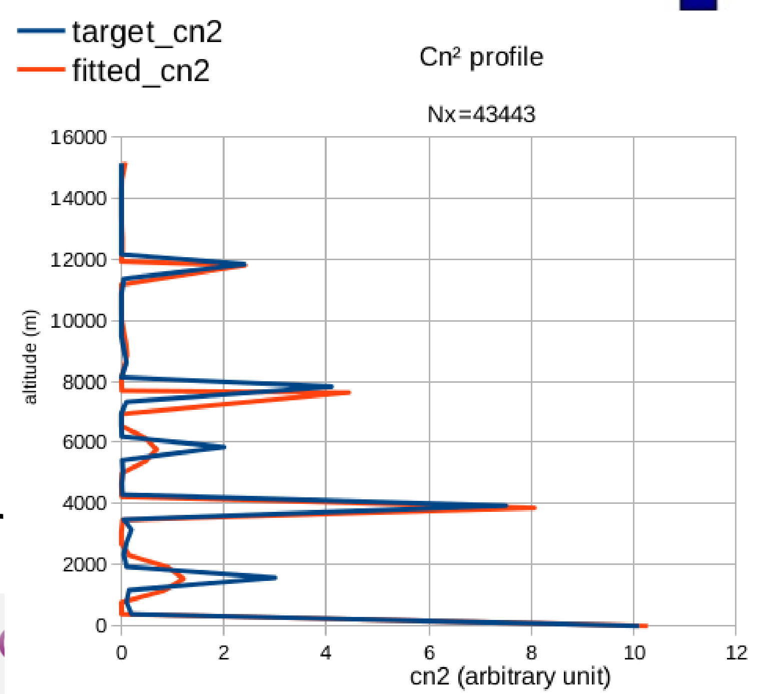


Parameters identification ("Learn" process)

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

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

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



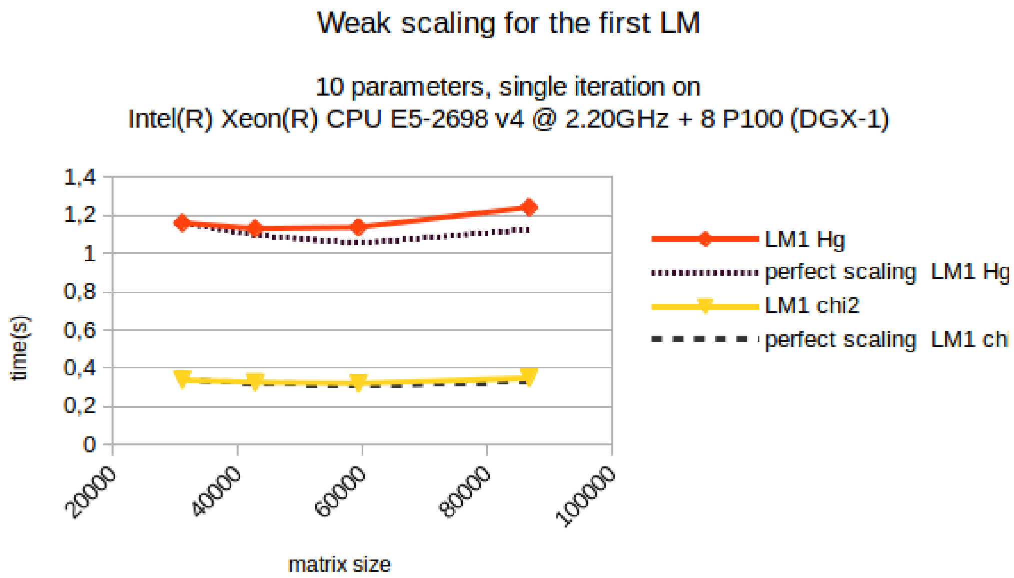
Loop supervision module

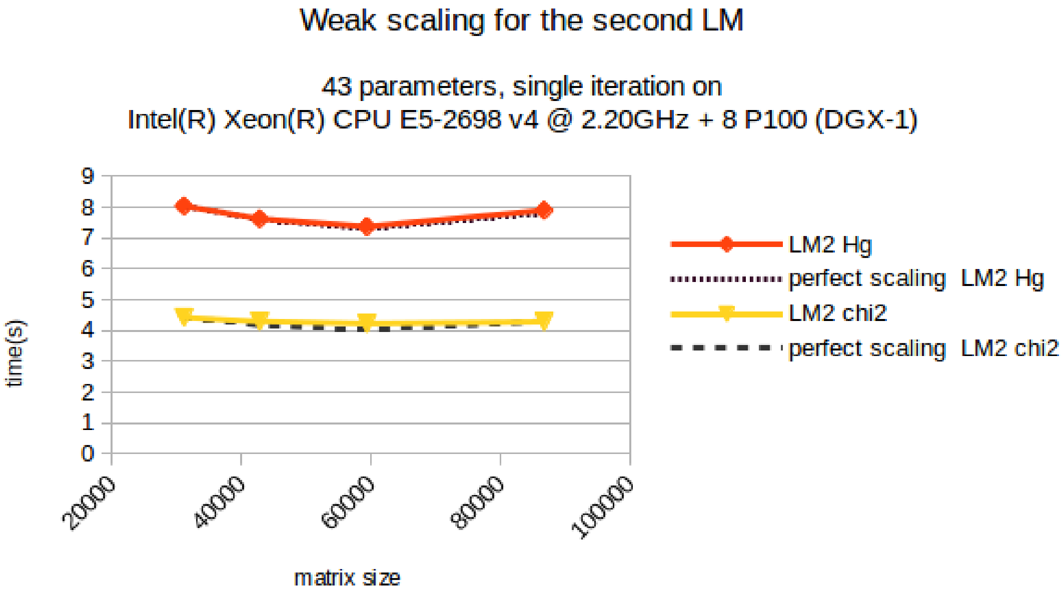
Performance for parameters identification ("Learn" process)

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





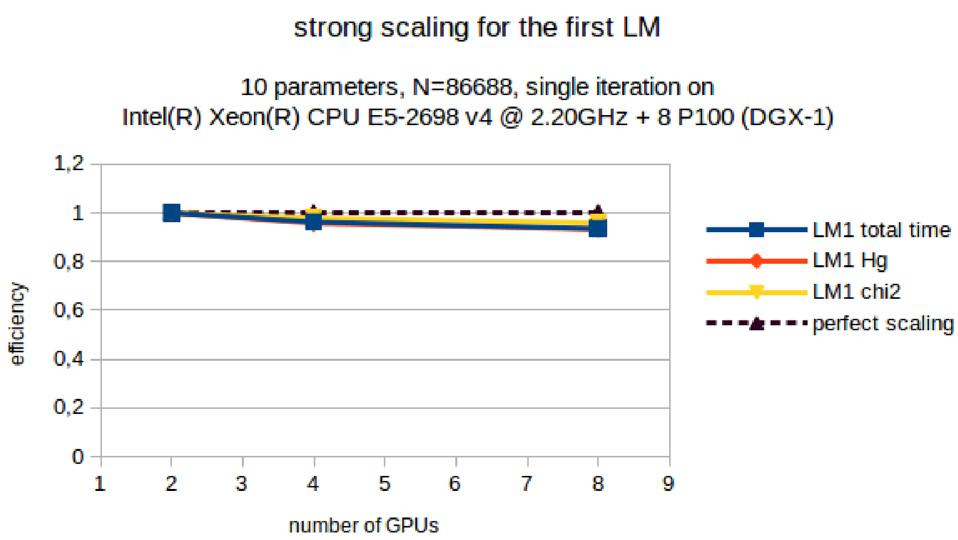
Loop supervision module

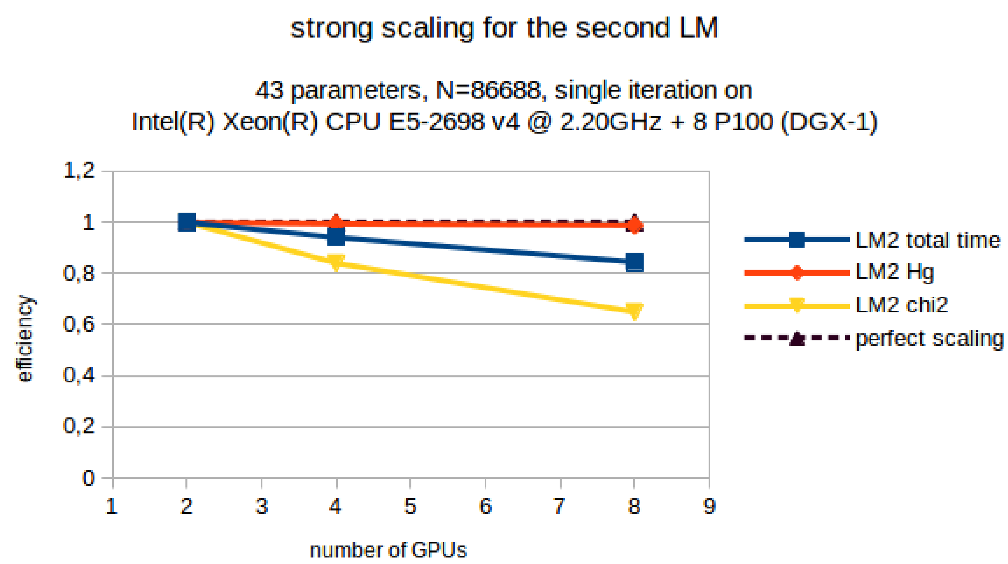
Performance for parameters identification (“Learn” process)

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

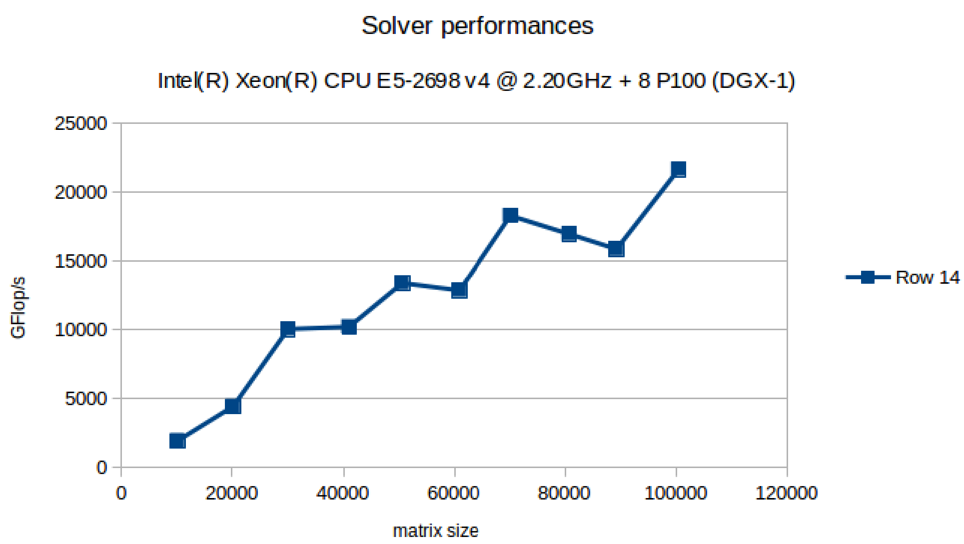
$$R' = Ctm \cdot Cmm_f^{-1}$$

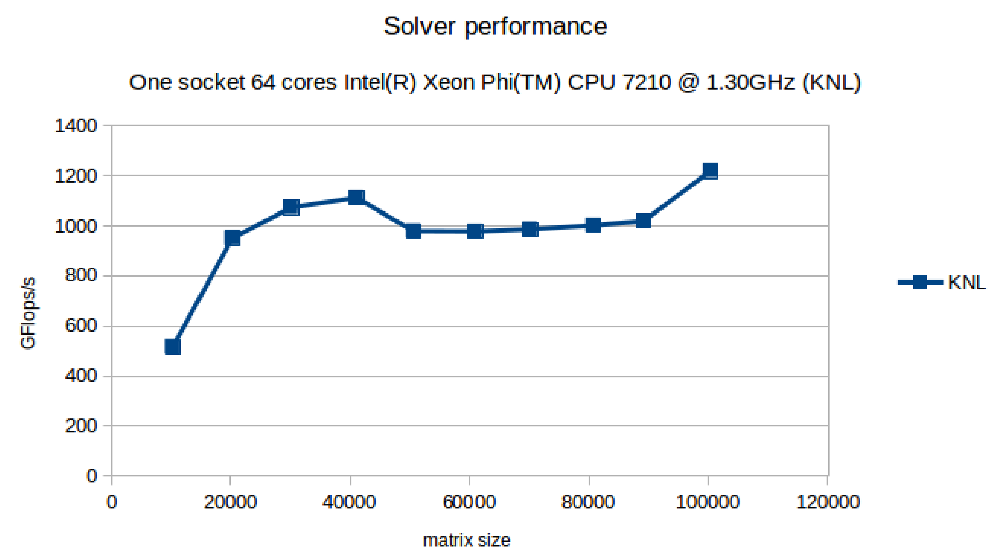
- Can use various methods. “Brute” force : direct solver
- Standard Lapack routine : “posv” : mostly compute-bound, high level of scalability
- Highly portable code : explore various architectures by using standard vendor provided maths libraries

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)



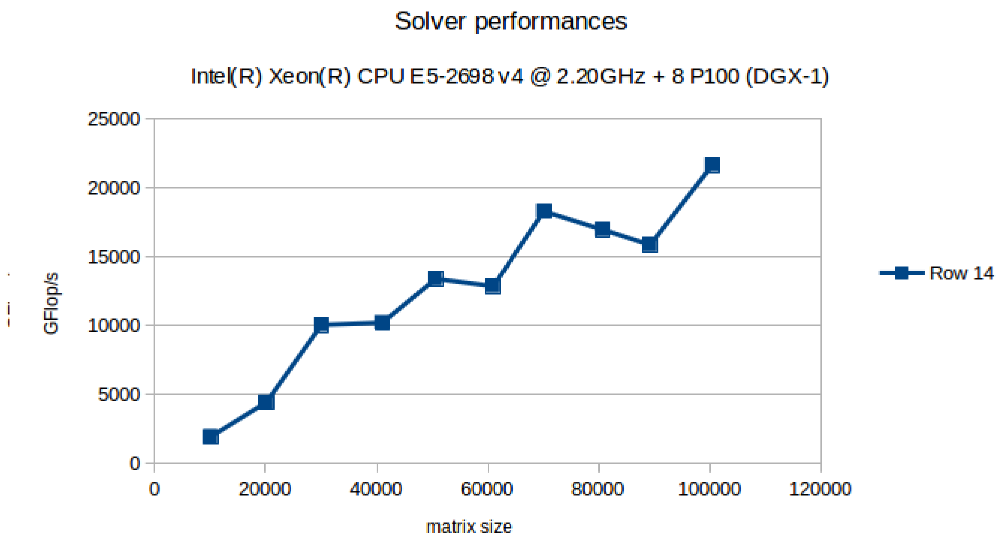


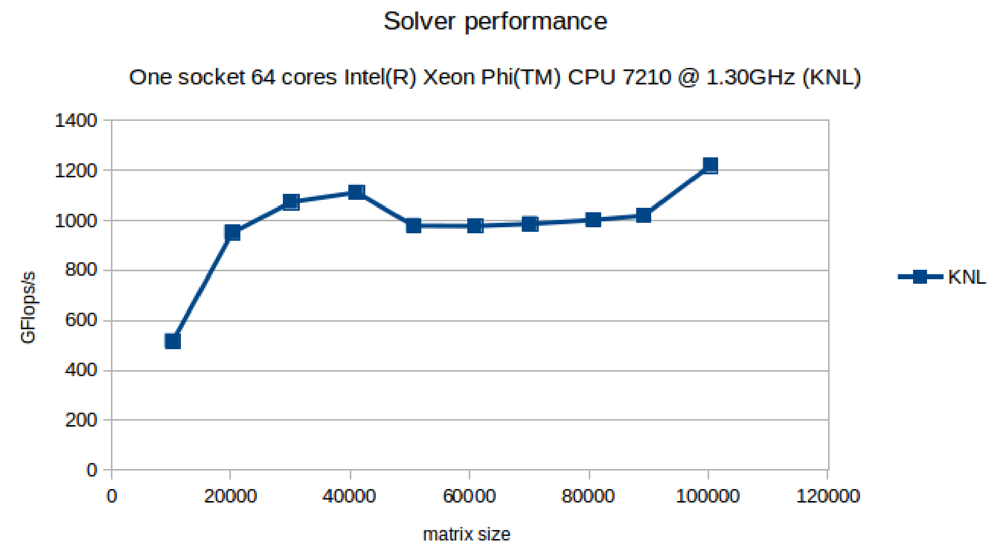
8 GPUs together reach more than 21 TFLOP/s while a single KNL can only reach about 1.2 TFLOP/s in peak performance

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)



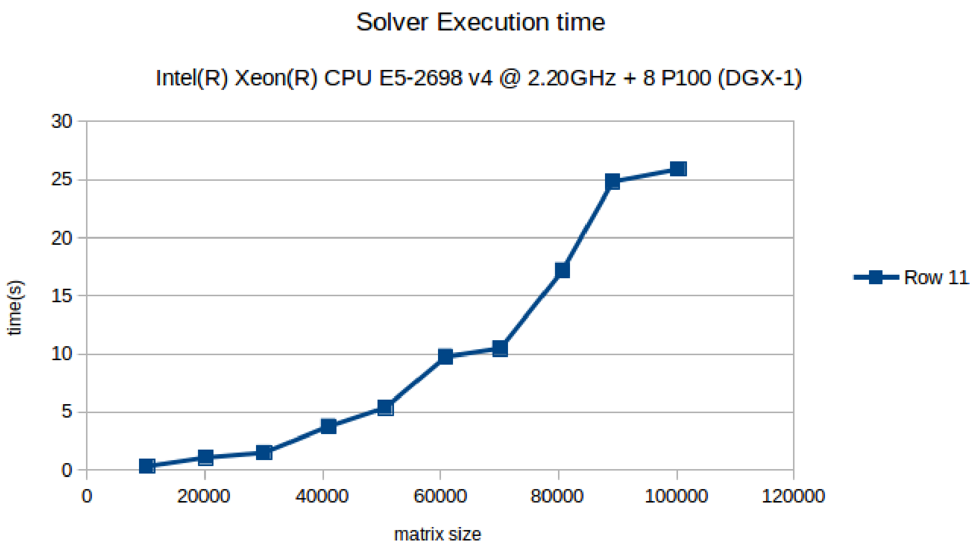


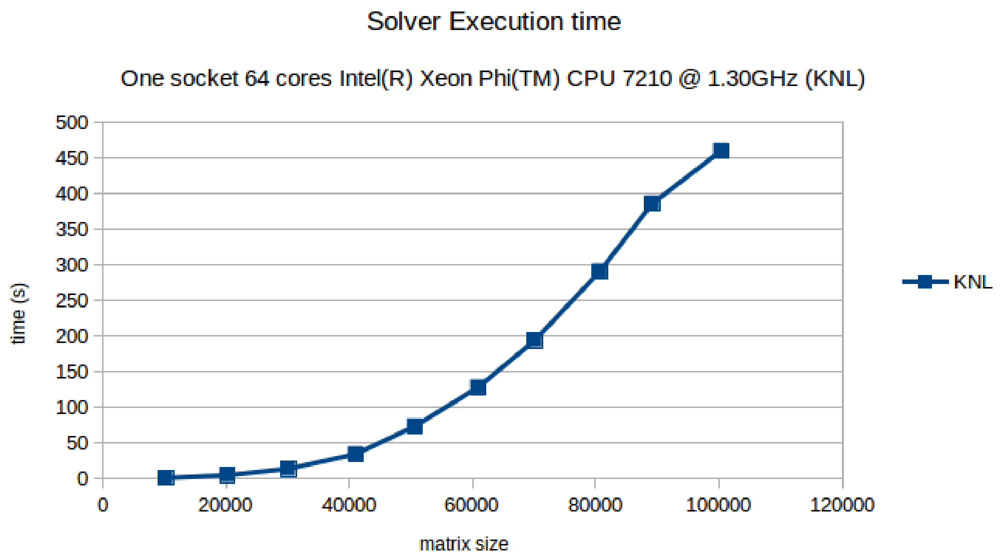
GPUs can deliver better peak perf. (saturation not reached, expect >2.5 or more) and the NVlink interconnect seems to perform very well

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)

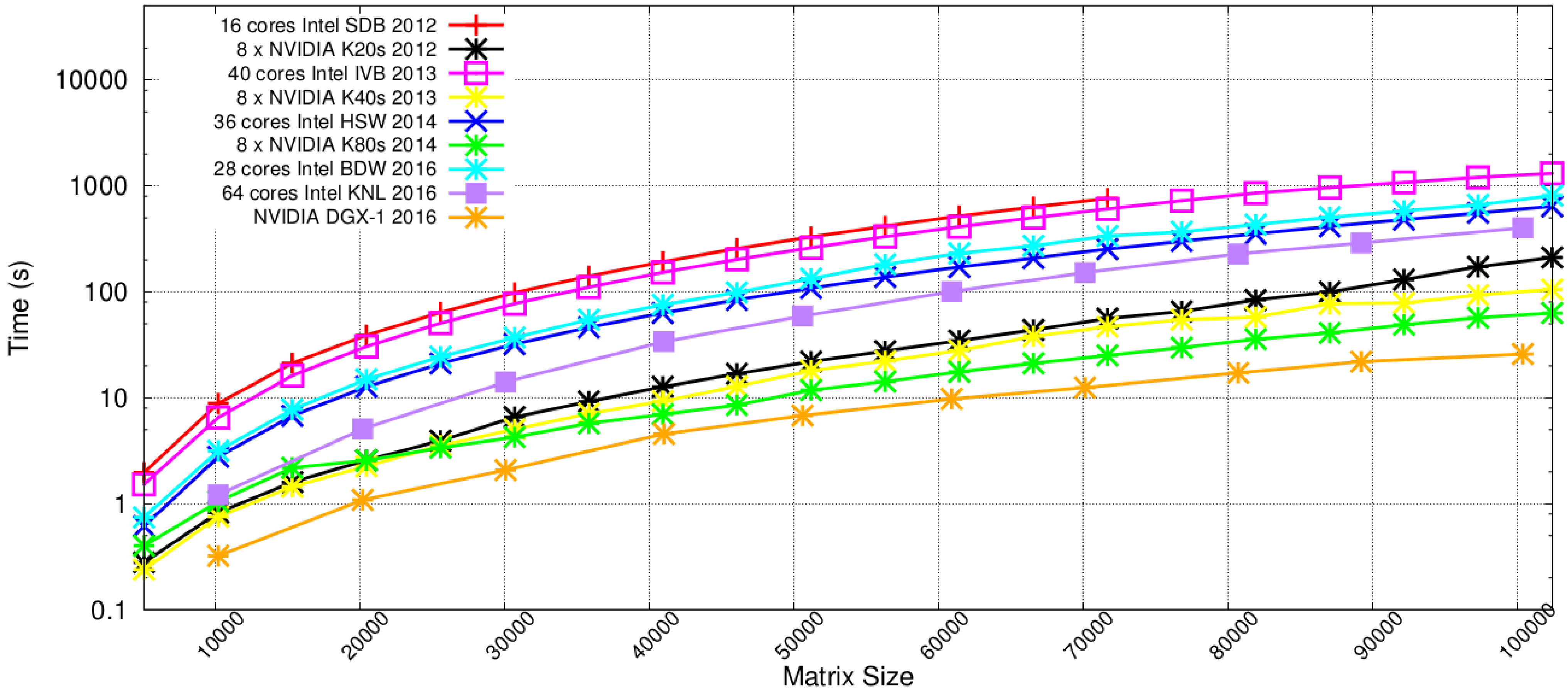




- Record time-to-solution on DGX-1 : MAORY / HARMONI full scale (100k x 100k matrix) : 25sec to compute tomographic reconstructor

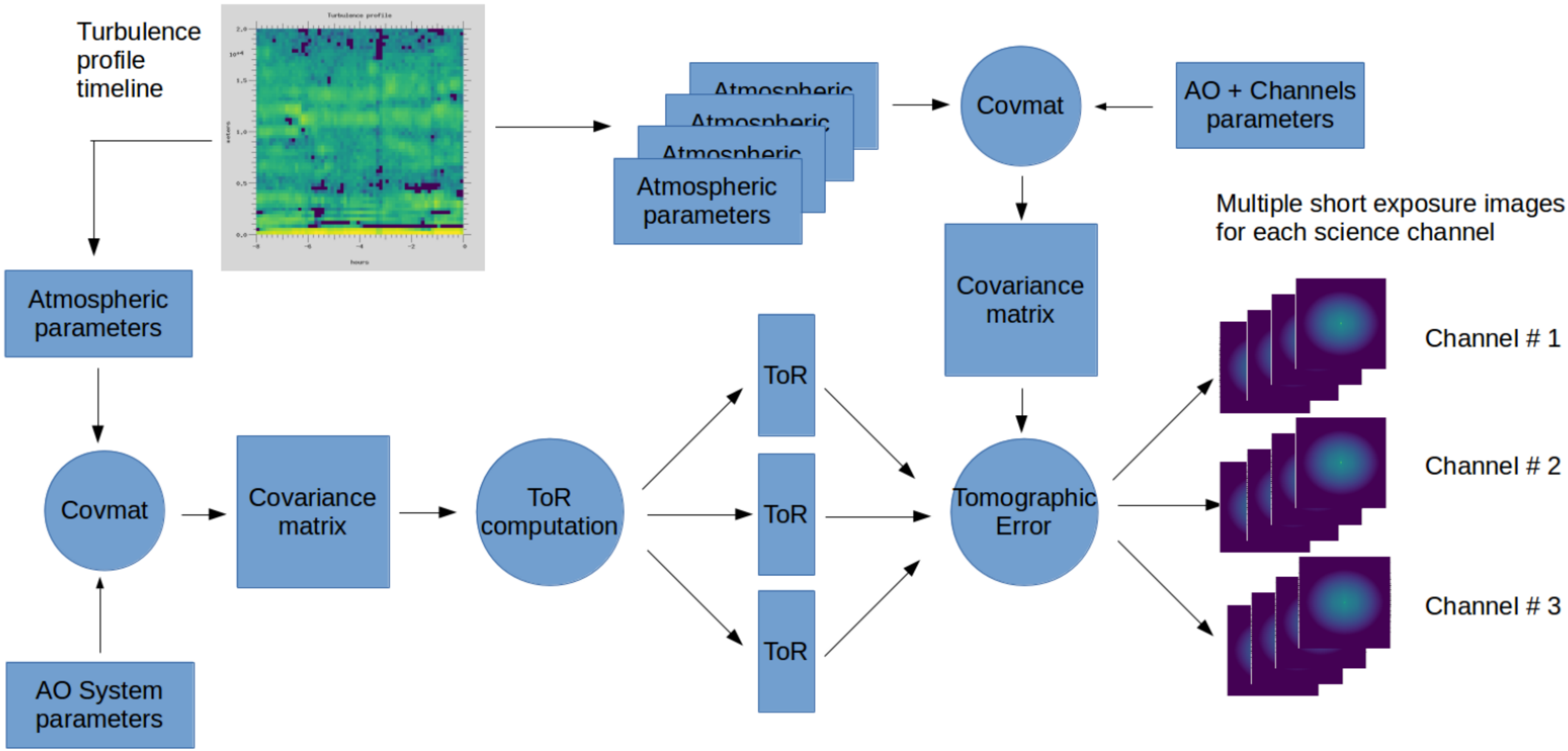
Loop supervision module

Performance evolution of ToR computation over the past years



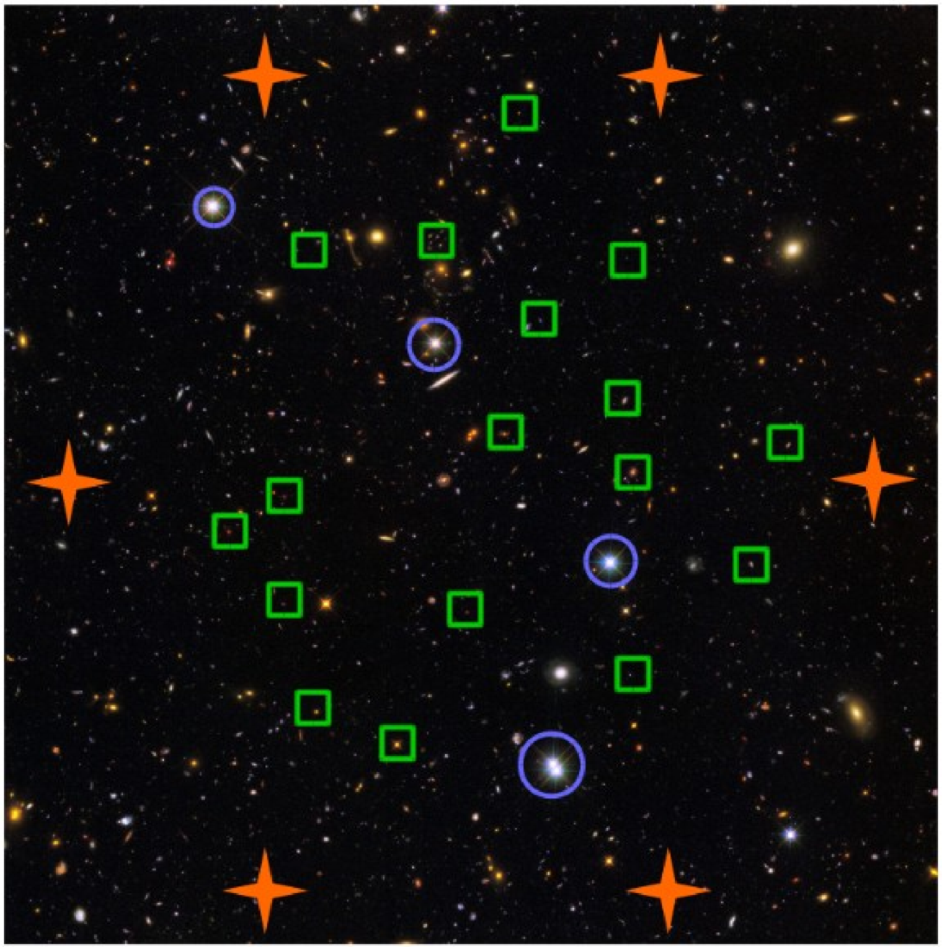
AO simulation pipeline

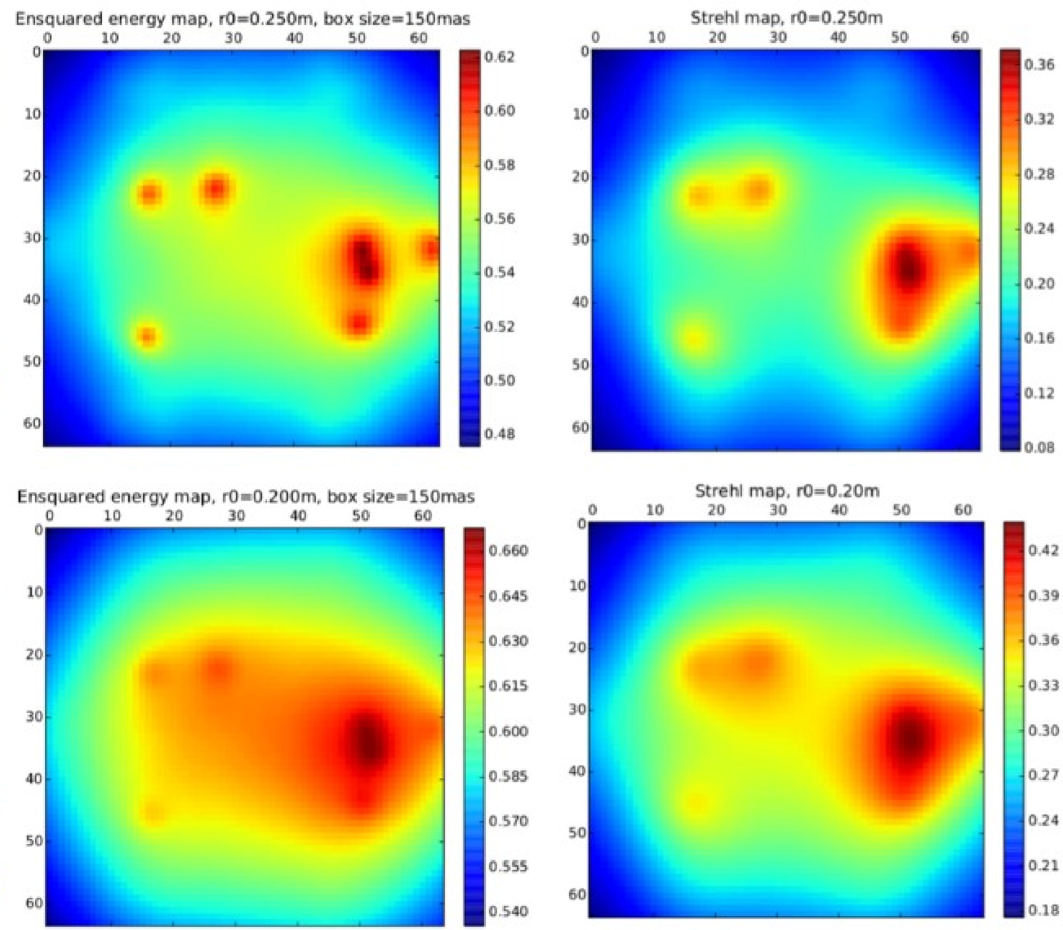
- Critical for AO instrument design studies



AO simulation pipeline

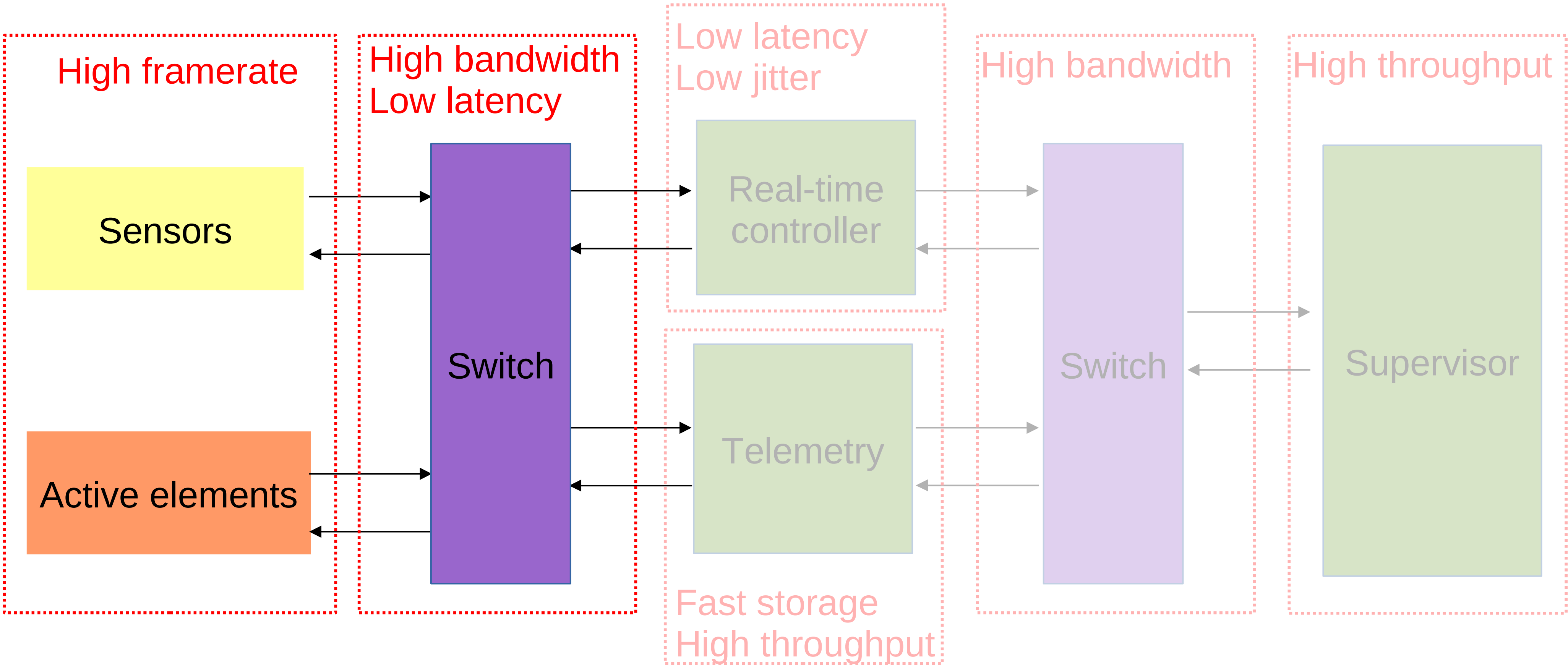
- Goal is to produce an observation forecast





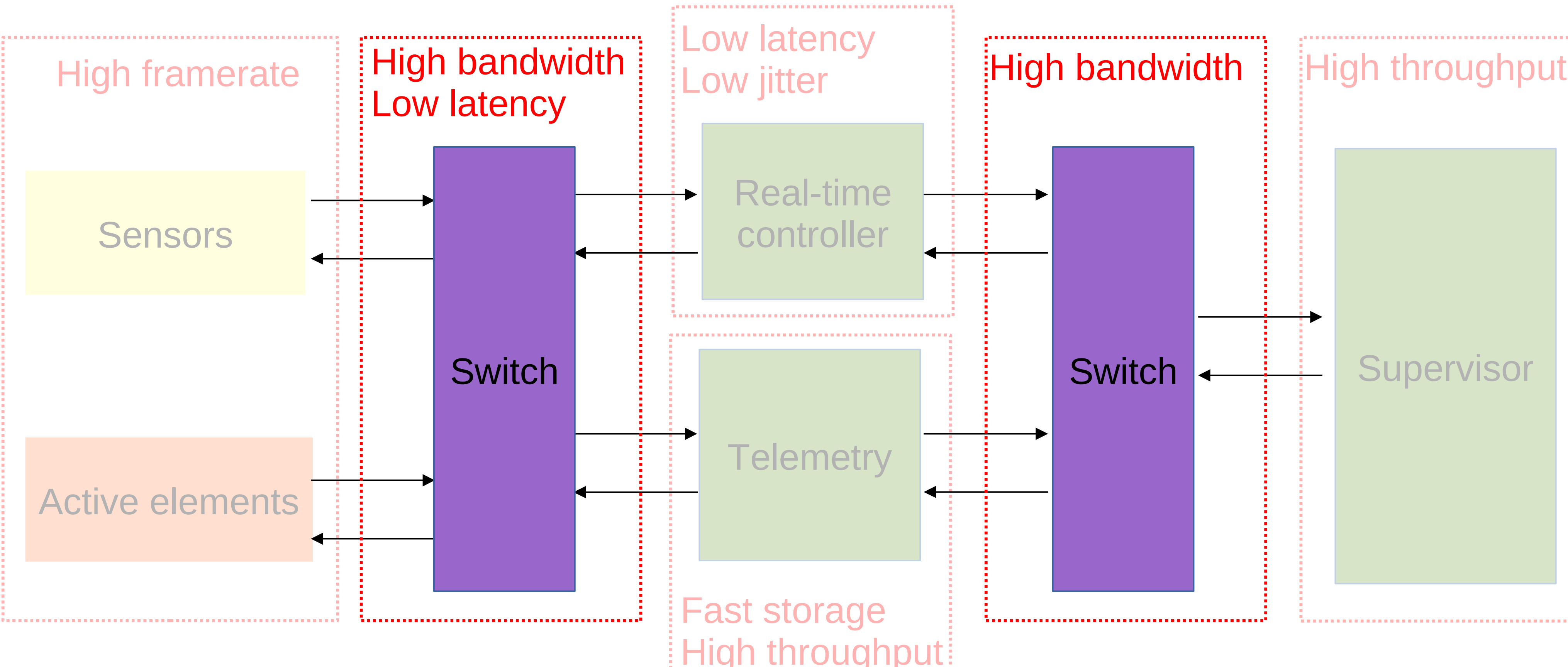
l'Observatoire de Paris — LESIA
Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique
Durham University
MICROGATE
PLDA

AO RTC concept : data streams

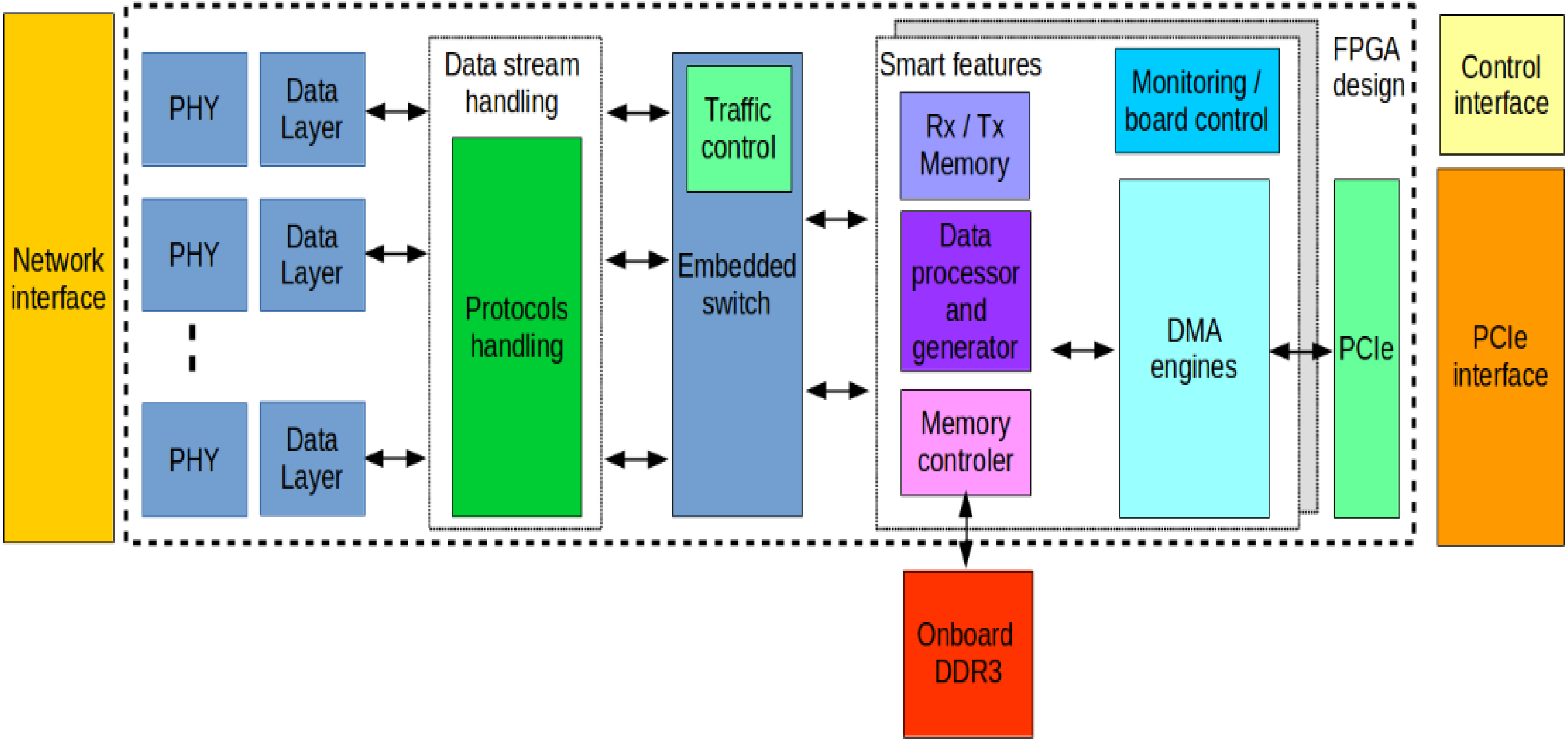


01/22/2016

AO RTC concept : local / global interco.

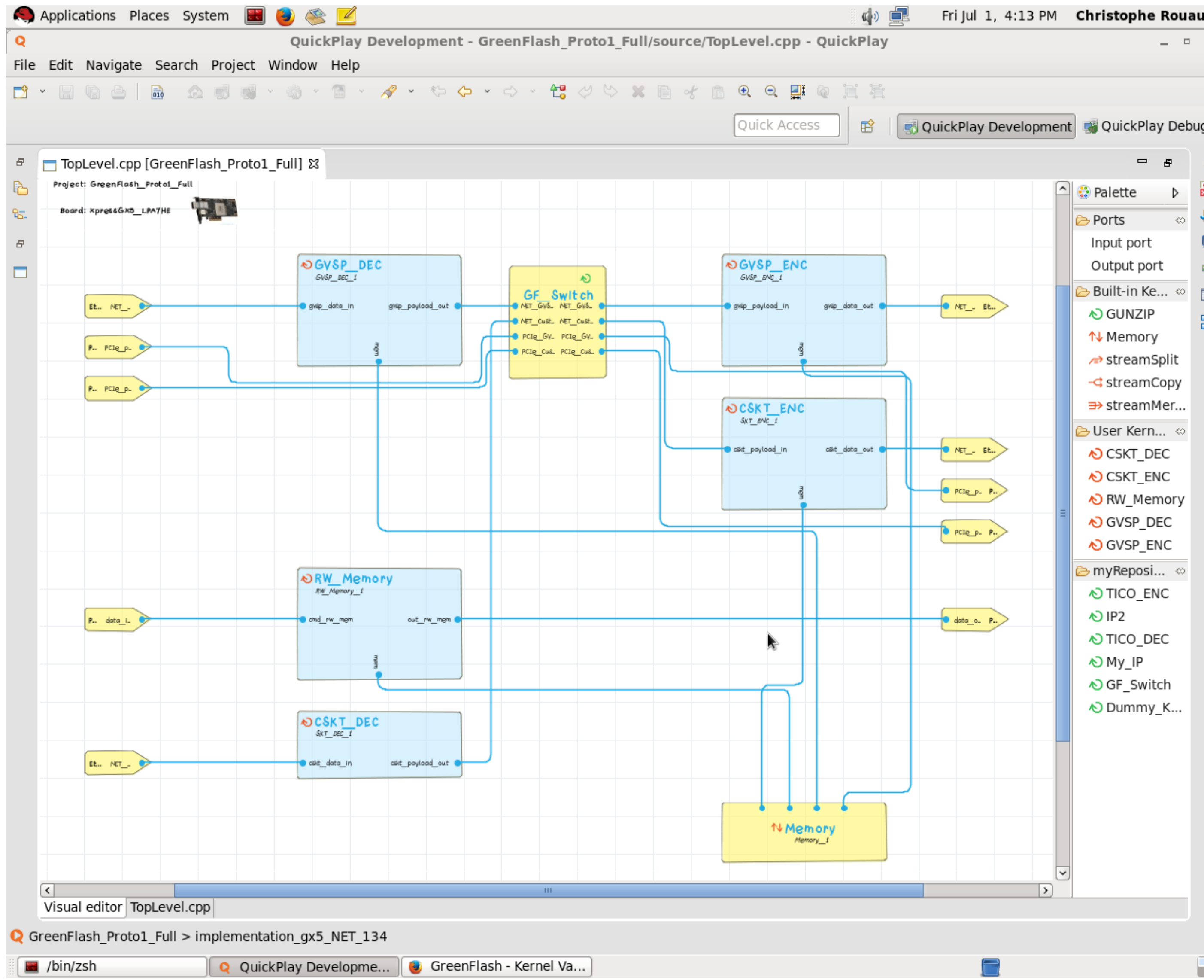


Smart interconnect architecture



Smart interconnect concept

- Eased devel. process using the QuickPlay tool from PLDA



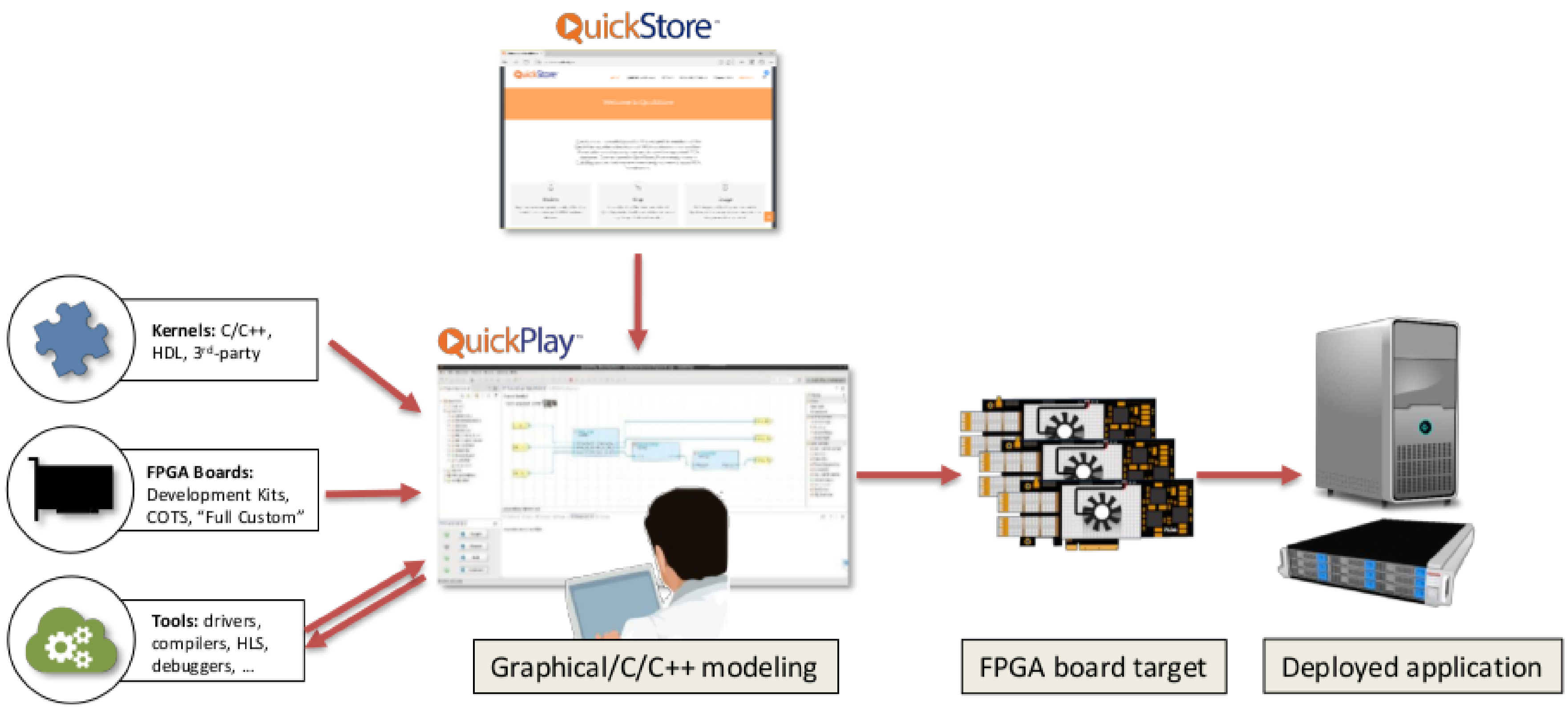
QuickPlay™

QuickStore™

QuickPlay

QuickPlay™

Introducing QuickPlay

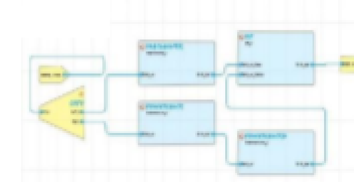


QuickPlay

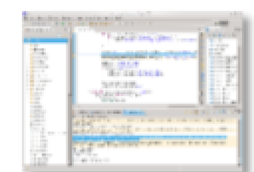
QuickPlay™

FPGA Design with QuickPlay IDE

1. **MODEL**
   – C/C++ functional modeling

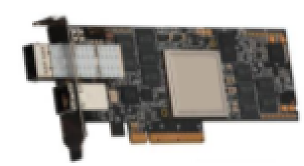


2. **VERIFY & VALIDATE**
   – Desktop execution of system functional model



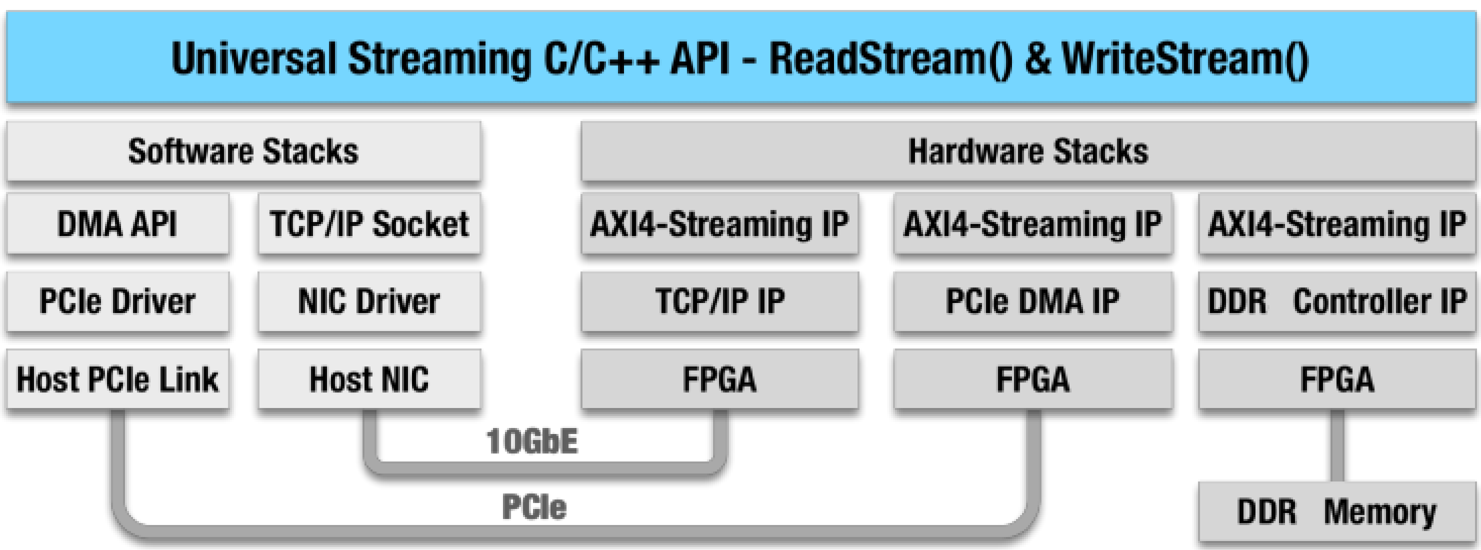
3. **BUILD**
   – Hardware implementation: HLS, Logic Synthesis, P&R

4. **EXECUTE**
   – FPGA based system hardware execution



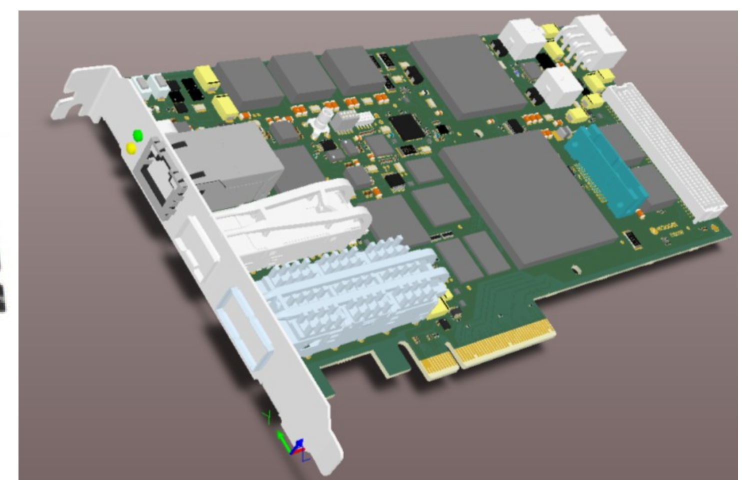
QuickPlay

QuickPlay™ Hardware Accelerator Abstraction Layer

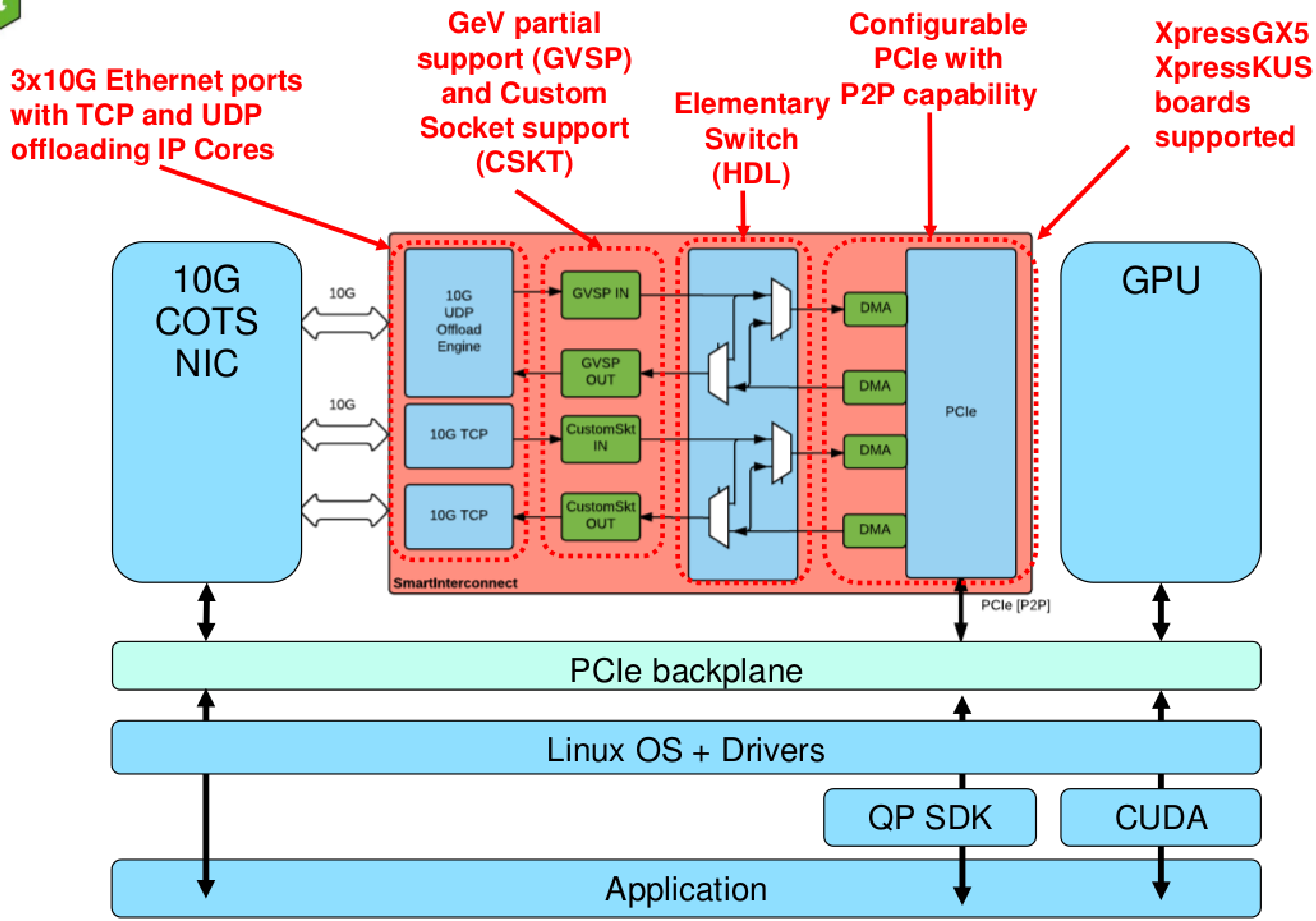


Smart interconnect prototyping

- Single generic design / multiple target boards
  - ExpressK-US board (hosting a Kintex UltraScale from Xilinx)
  - ExpressGX V board (hosting a Stratix V from Altera)
  - µXlink board from microgate (hosting a Arria 10 board from Altera)



Smart interconnect prototyping



01/22/2016

Summary

International joint academic-industrial project funded by EU

- Key contributor to European Extremely Large Telescope project
- Exploiting emerging computing technologies (both HW and SW)
- KAUST-ECRC as external partner

Energy efficiency plays a critical role (remote location, limited operational budget)

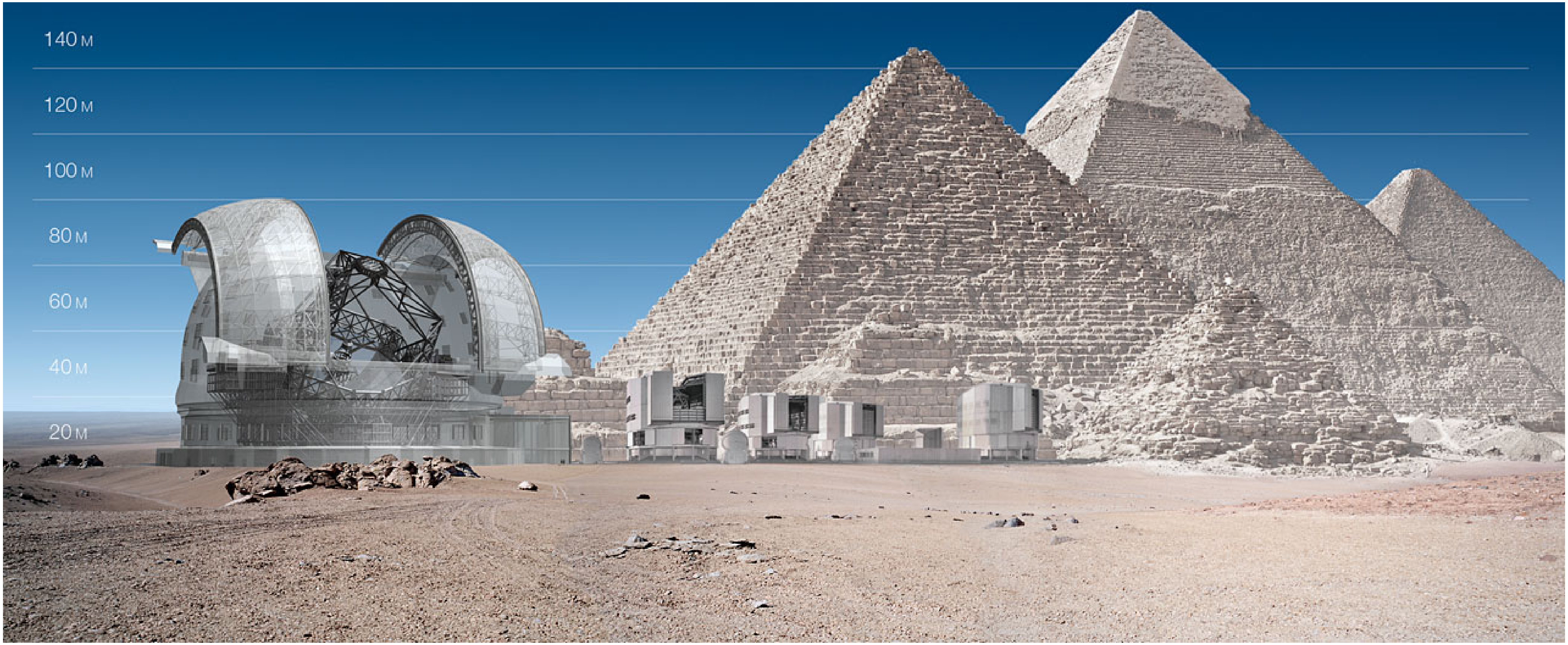
- Accelerator-based system architecture
- Optimized data streaming using FPGA technology

Optimized Linear algebra is key to the various sub-systems

- Not only for system operations but also simulations for the system design
- Original approach designed in collaboration with the team at KAUST

Already enhancing the readiness level of commercial solutions

- Contribution to QuickPlay development environment from PLDA
- Design of innovative FPGA boards at Microgate


140 M
120 M
100 M
80 M
60 M
40 M
20 M