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

Durham University

MICROGATE

PLDA

Green Flash

High performance computing for real-time science

Project overview and management (WP 1 & 2)
Final Design Review, April 6th 2018

European Commission

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

Goal of this review

- Get feedback !
- Share results obtained so far
- Status update of remaining individual prototyping activity
- Discuss the final prototypes design
- After the review:
  - Produce a final performance report, a final prototypes design report and integration and tests plan (to be finalized and submitted to EC by end of May)

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

What this is about … really

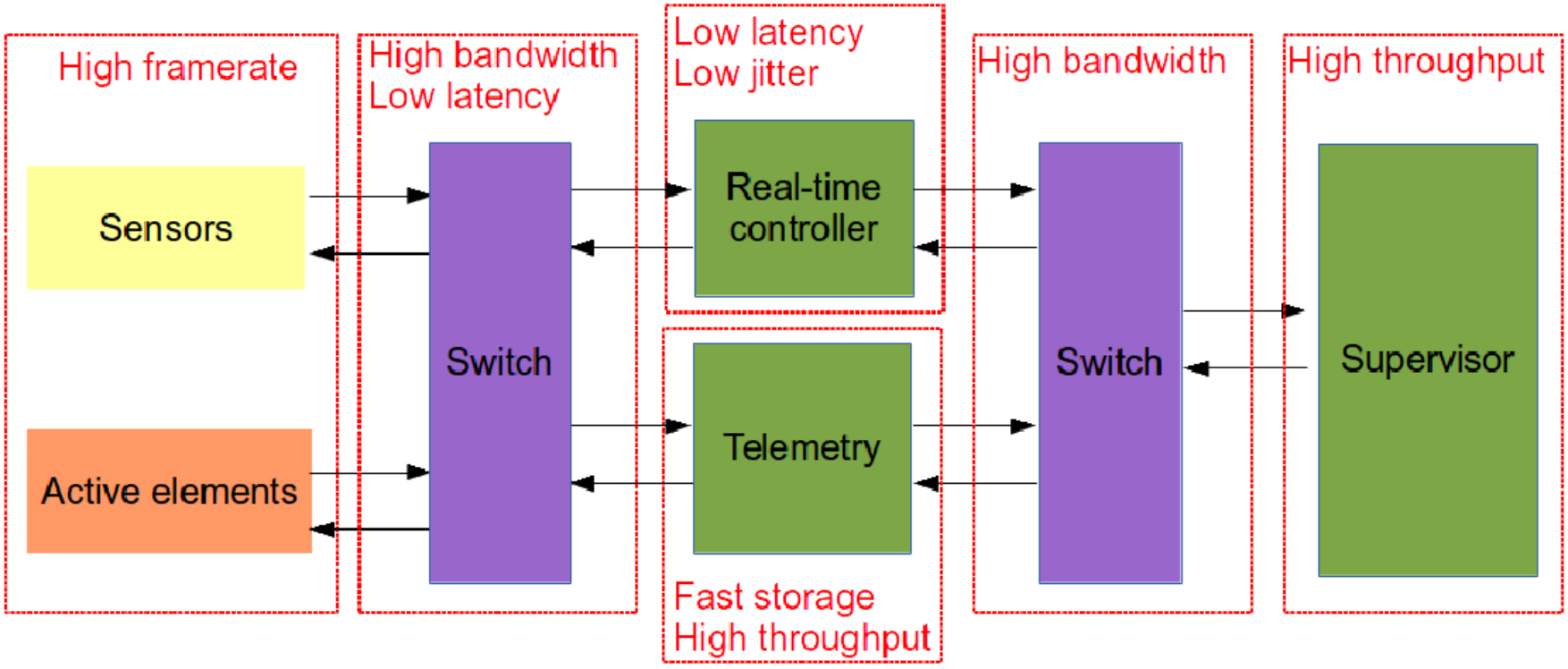
- Find the best trade-off for ELT sized AO systems RTC
  - Comprehensive assessment of existing technologies
  - Development of new custom solutions for comparison
  - Propose new development processes to reduce cost and increase maintainability
- Build a full featured RTC prototype at the largest scale possible
  - Technology down-selection from a number of criteria : performance, cost, compliance to standards, obsolescence, maintainability
  - State of the art systems to be assessed in the lab, with a simulator

Green Flash project

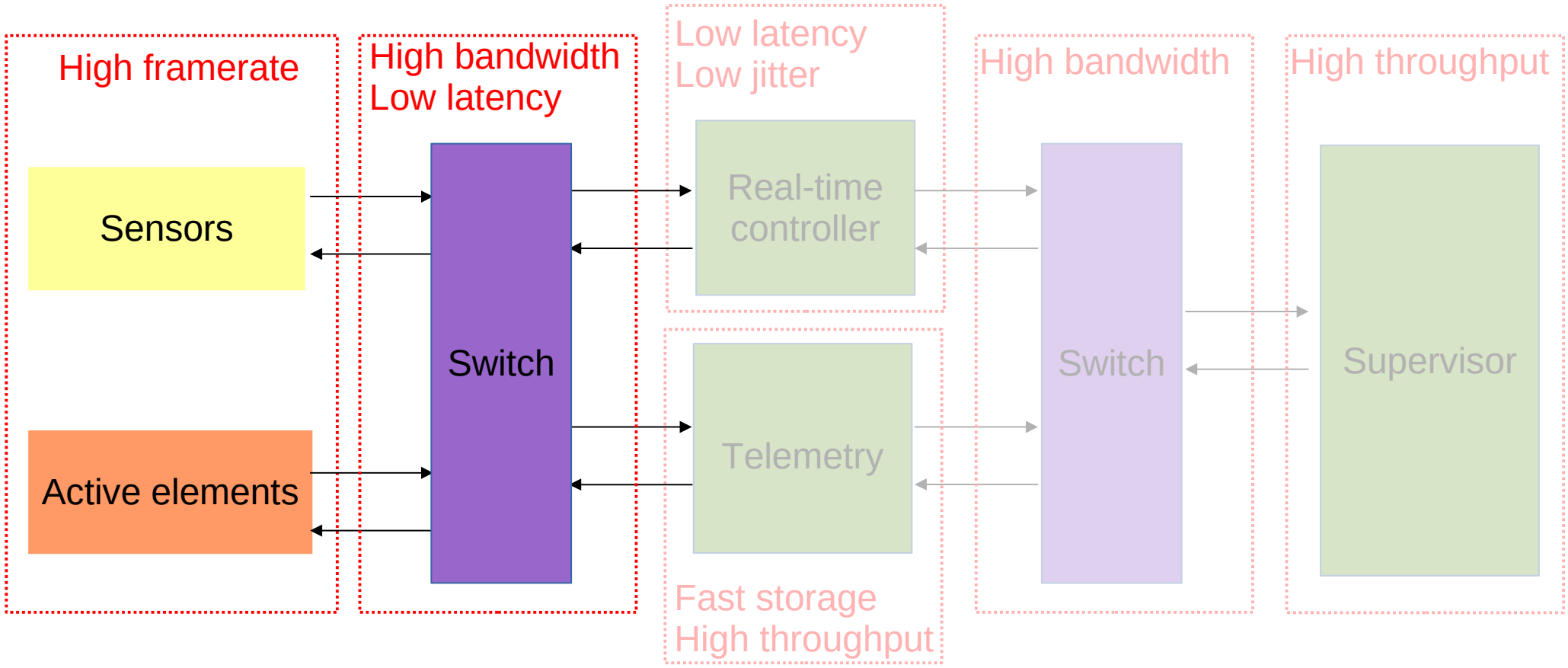
- Partners
  - 2 academic partners
    - LESIA, Observatoire de Paris, P.I. Damien G.
    - CfAI, University of Durham
  - 2 industrial partners
    - Microgate : Italian SME designing FPGA solutions for various applications (including astronomical AO)
    - PLDA / Accelize: French SME developing FPGA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

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

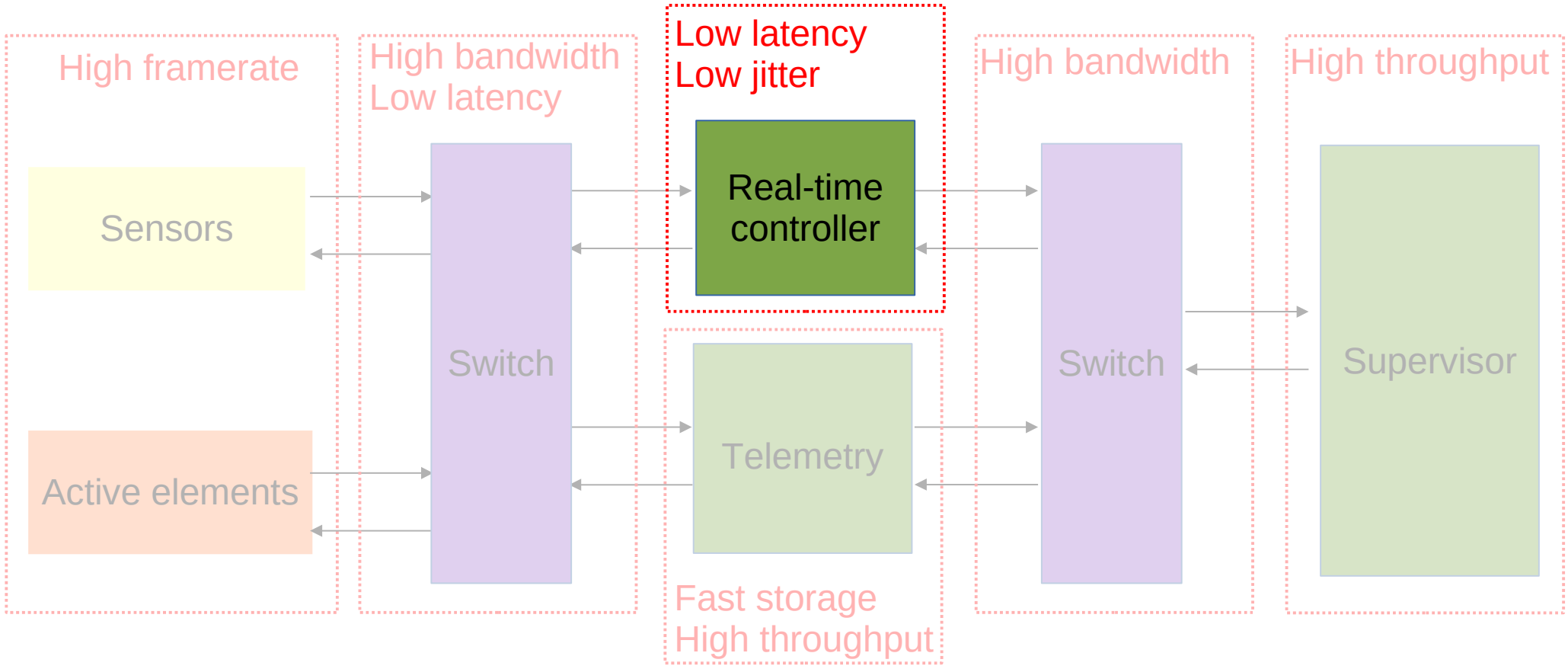
AO RTC concept



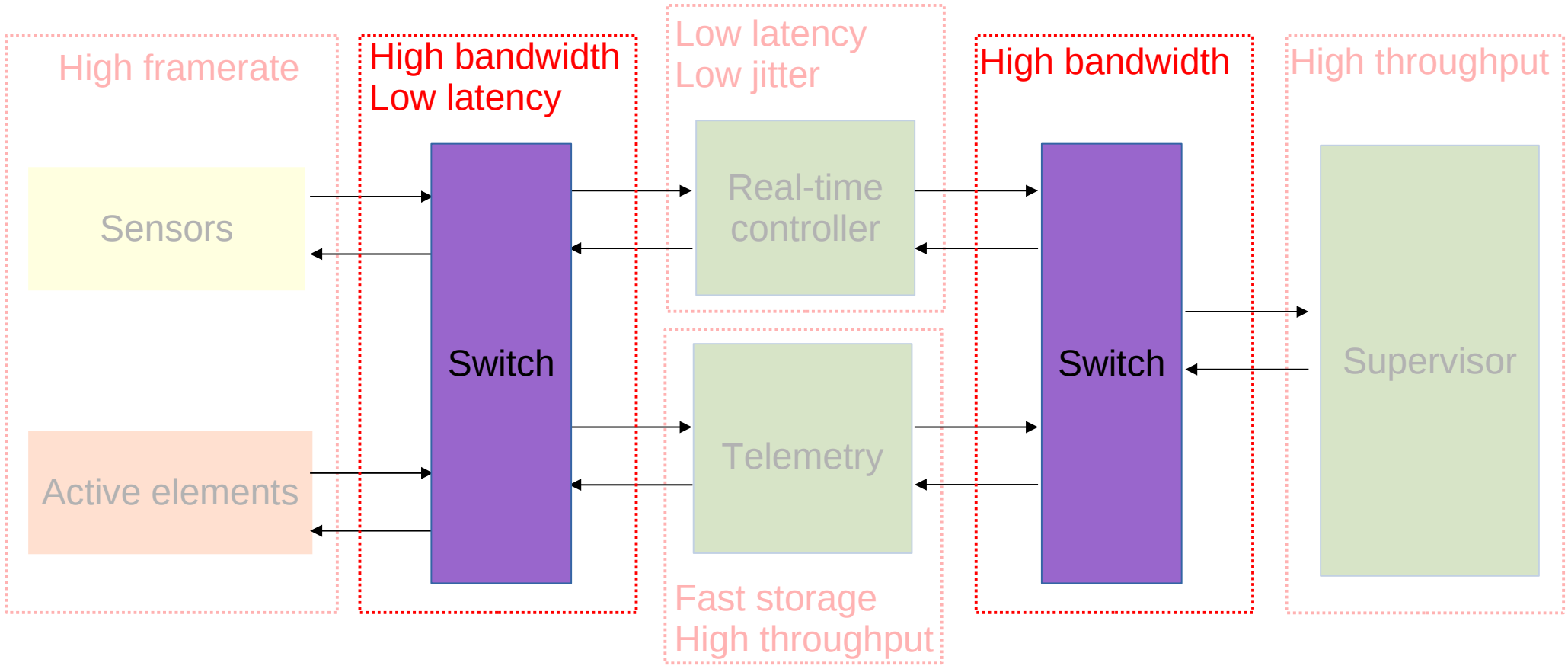
AO RTC concept : RT simulator



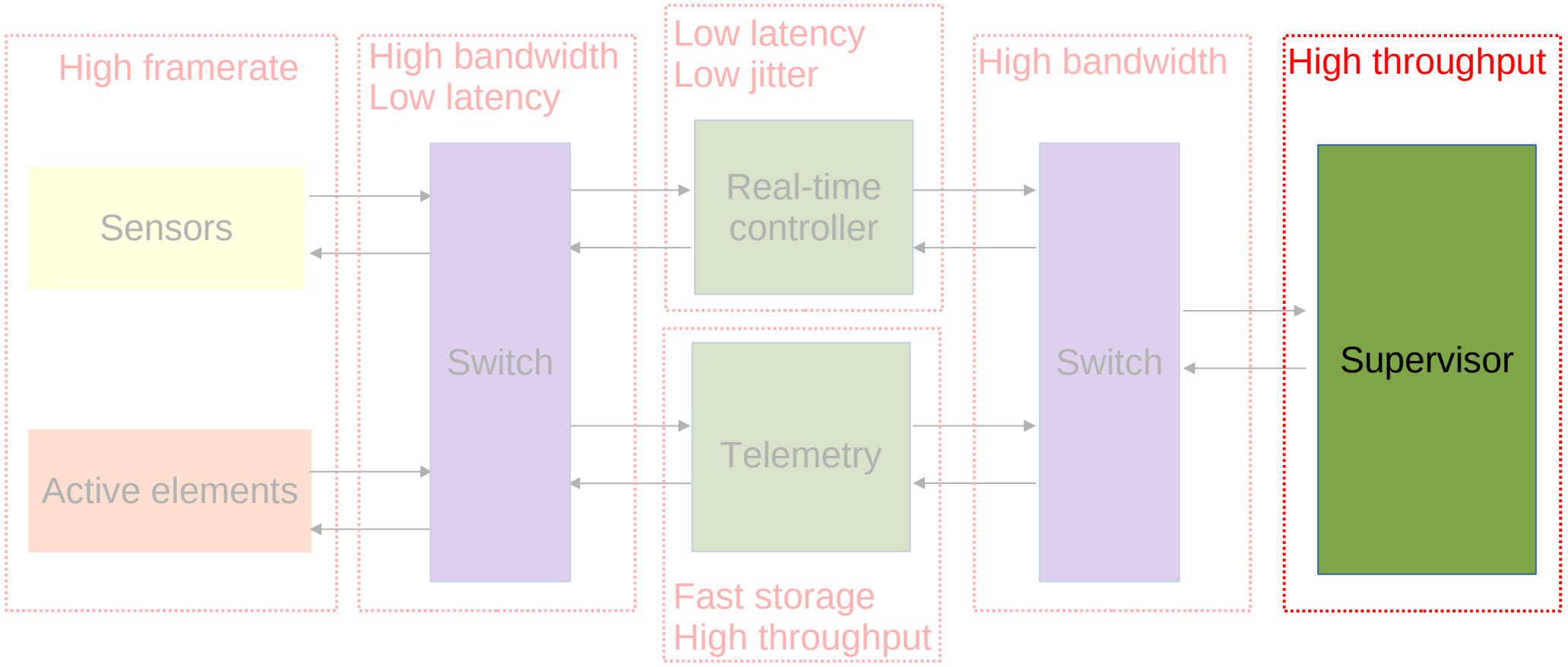
AO RTC concept : data pipeline



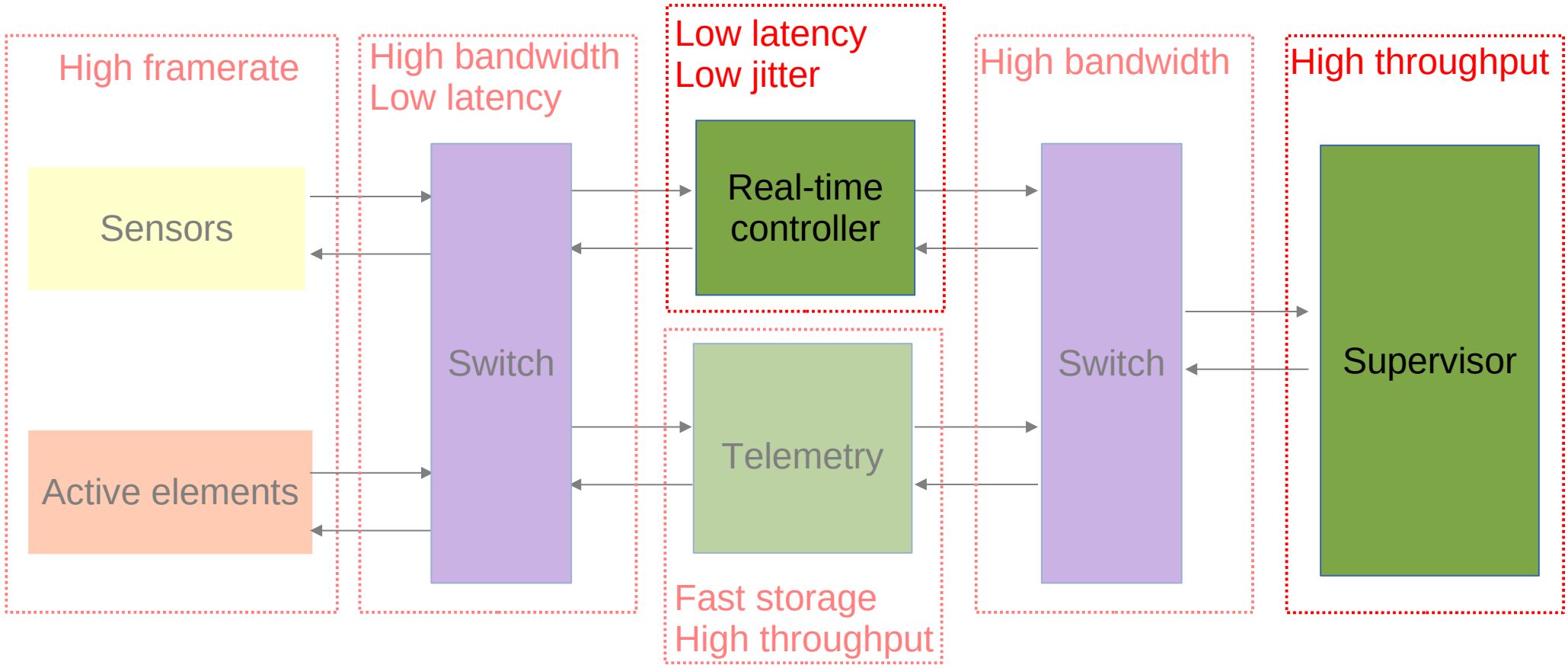
AO RTC concept : smart interconnect



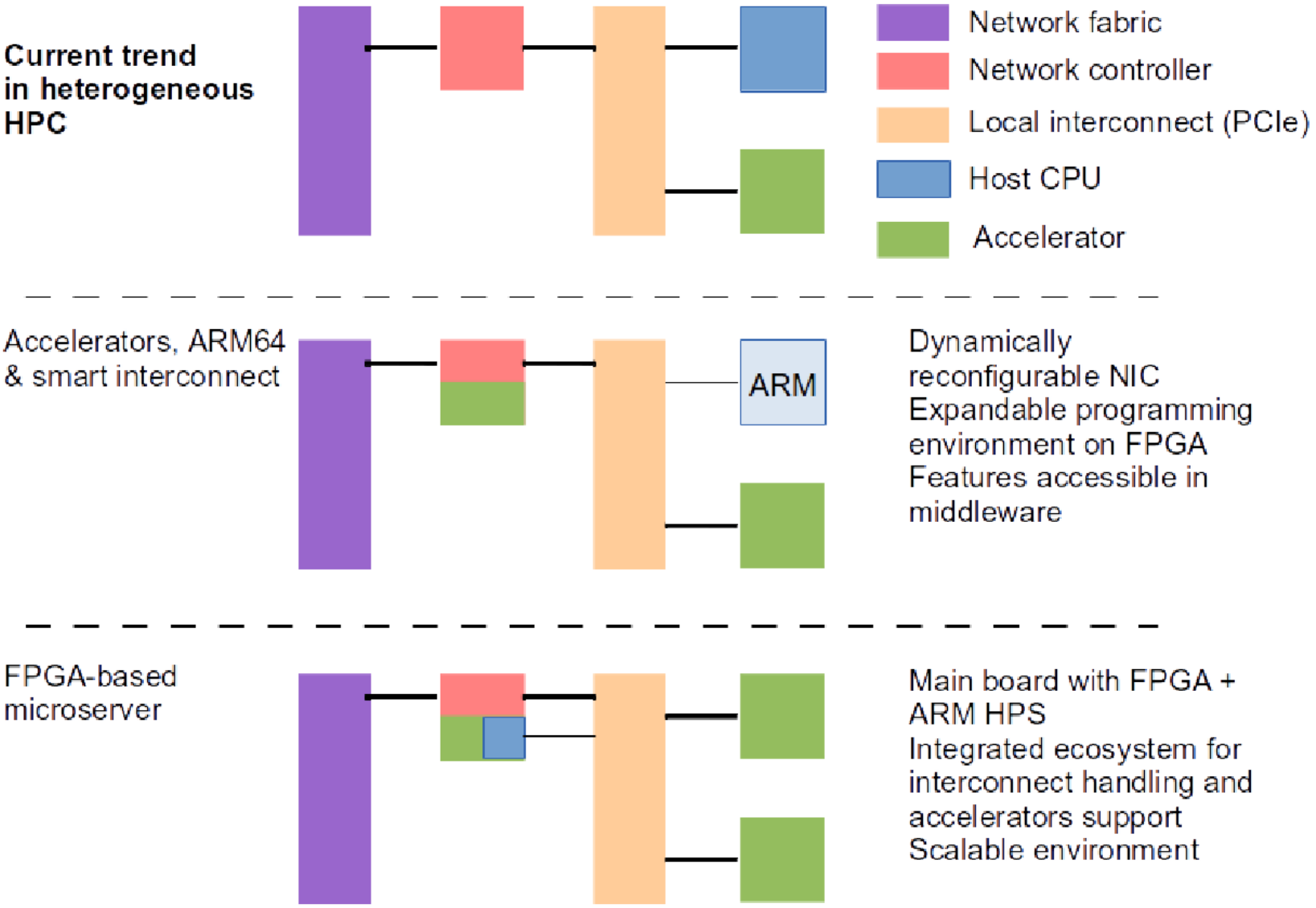
AO RTC concept : supervisor



AO RTC concept : SW & MW



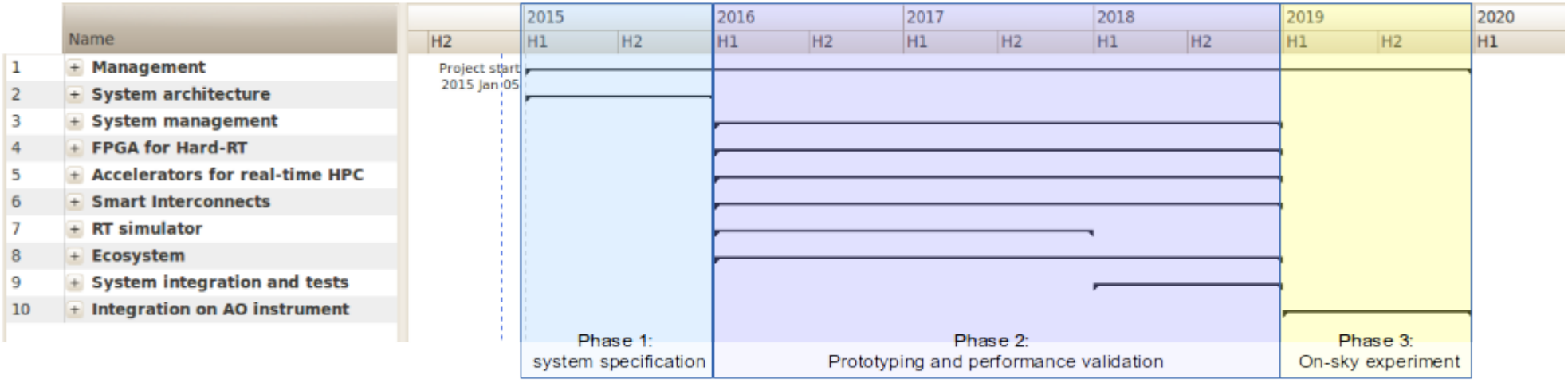
Assessing new HPC concepts



Project Management

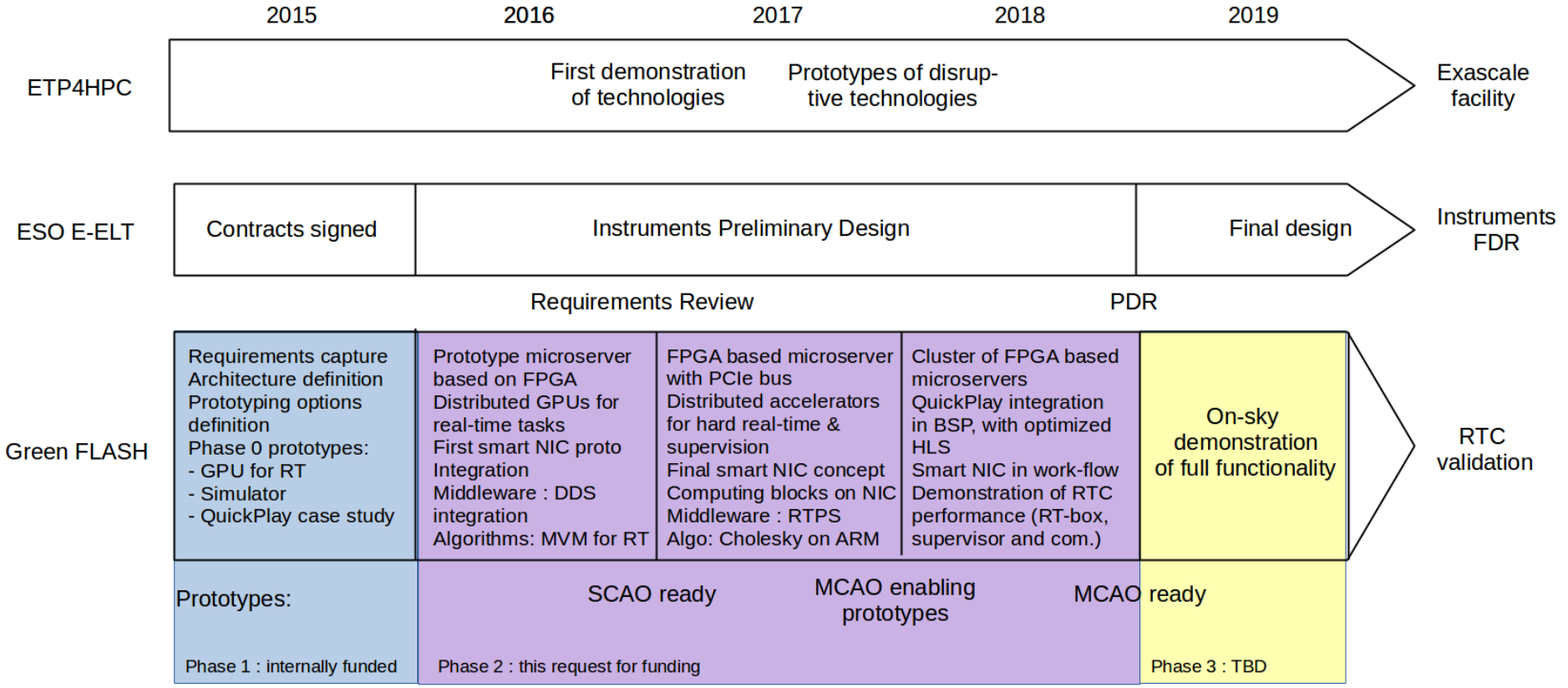
3 phases. Phase 2 funded through H2020 program

- Main prototyping phase from system specifications
- On-sky experiment to be defined along the course of the project
- **Project extension requested until end 2018 (+3 months)**



Project Management

- Good convergence with H2020 ETP4HPC / E-ELT project timeline



Work Packages

WP1: Project management (OdP)

WP2: System management (OdP)

WP3: FPGA for hard-RT (Mic)

WP4: Accelerators for real-time HPC (OdP)

WP5: Smart interconnect (PLDA)

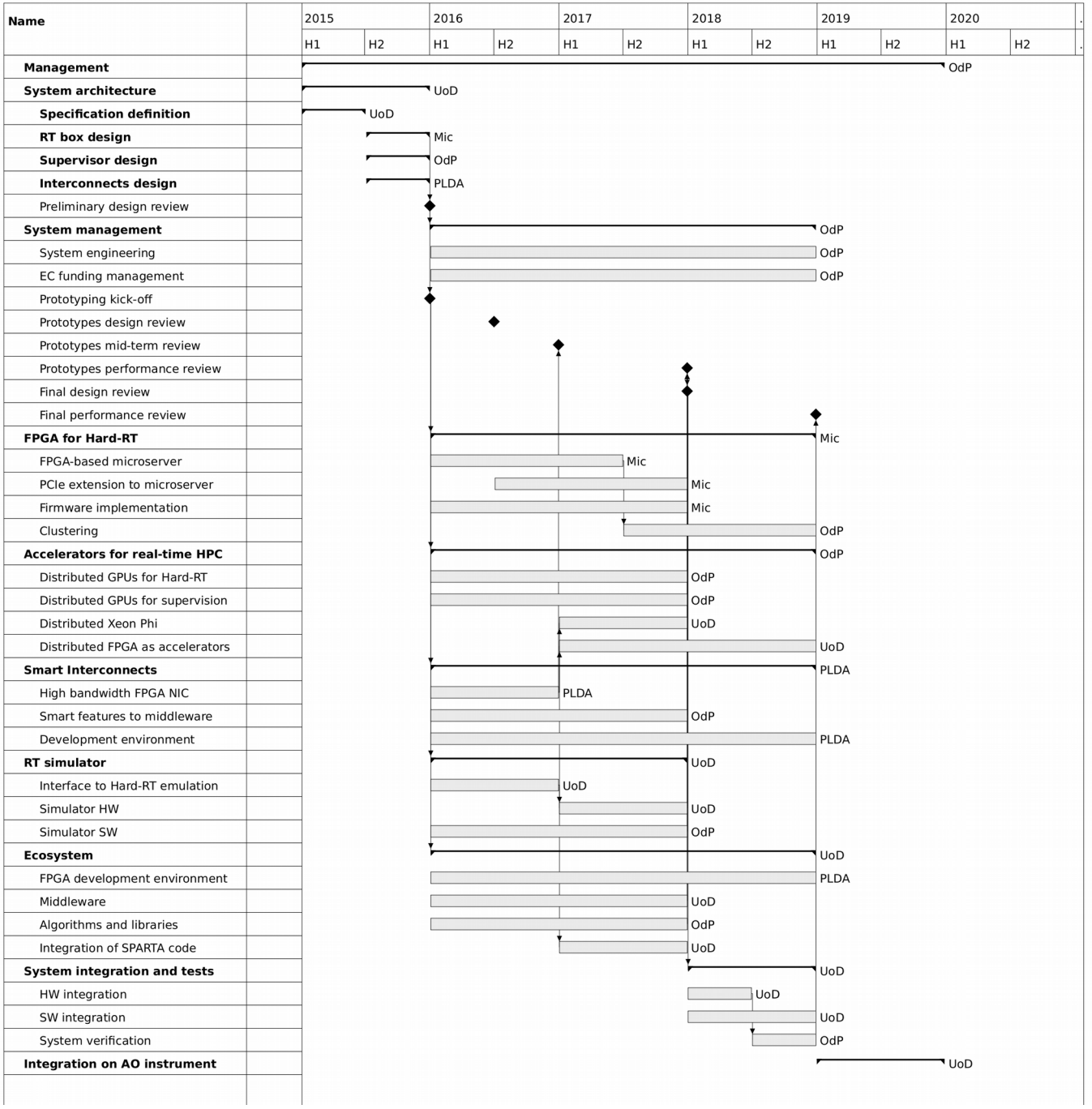
WP6: RT simulator (UoD)

WP7: Ecosystem (UoD)

WP8: System integration and tests (UoD)

WP9: Integration on AO instrument (UoD)

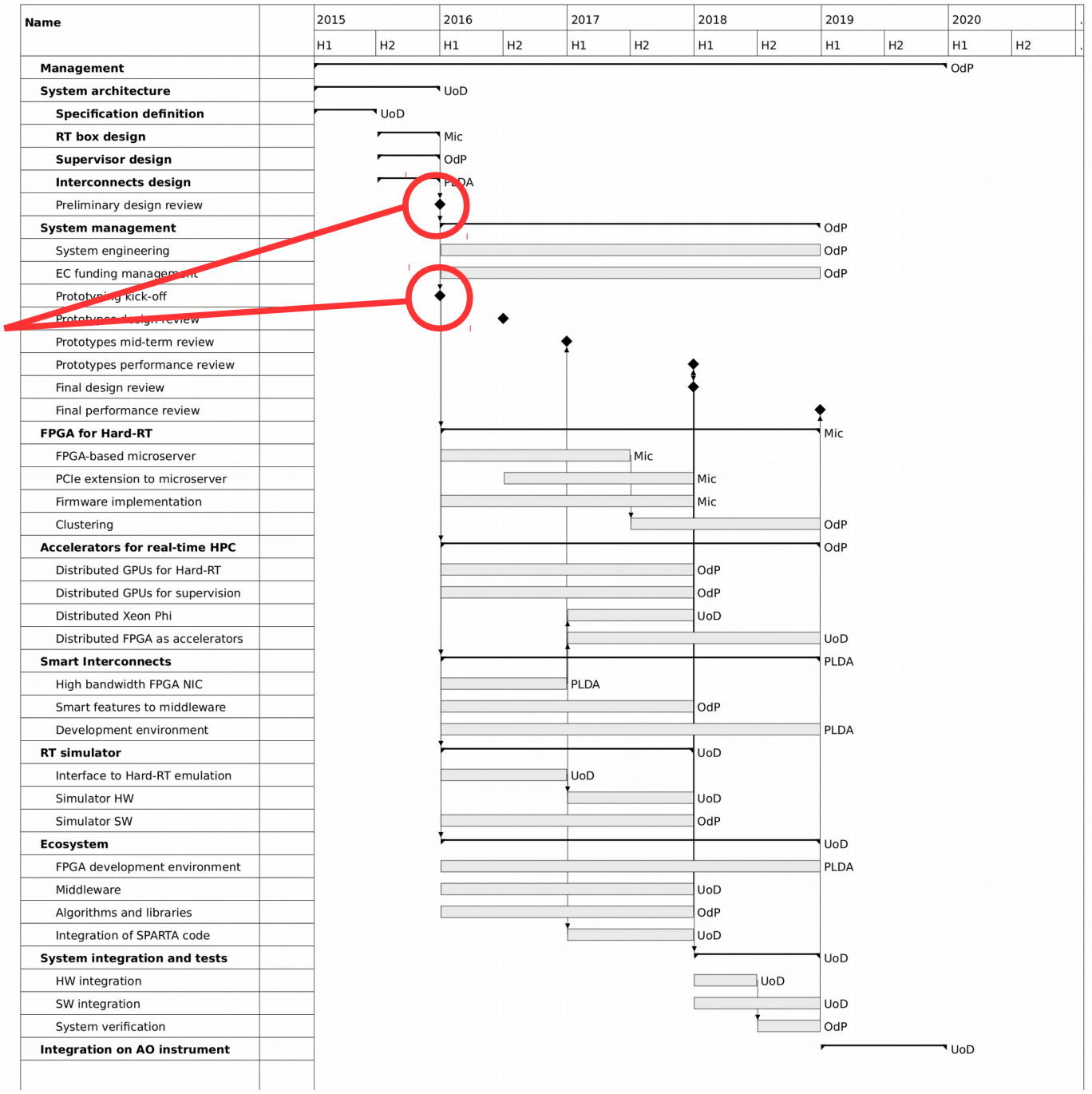
WP1



April 6th 2018

WP1

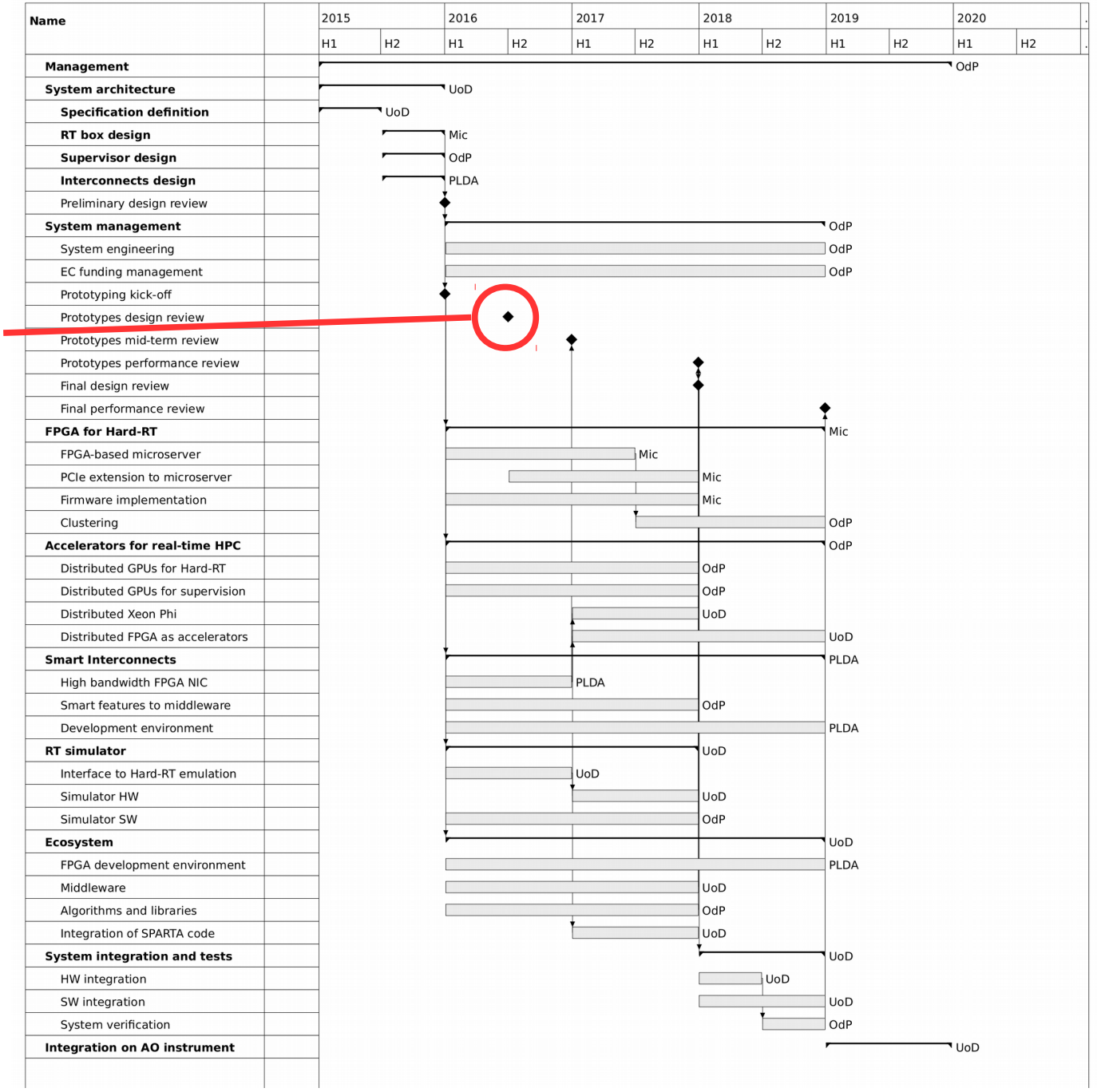
- Preliminary design review
- Held on 22nd Jan. 2016
- Review panel from E-ELT instruments community and AO RTC specialist
- Design docs for all prototyping activities
- Including E-ELT AO RTC specification capture



April 6th 2018

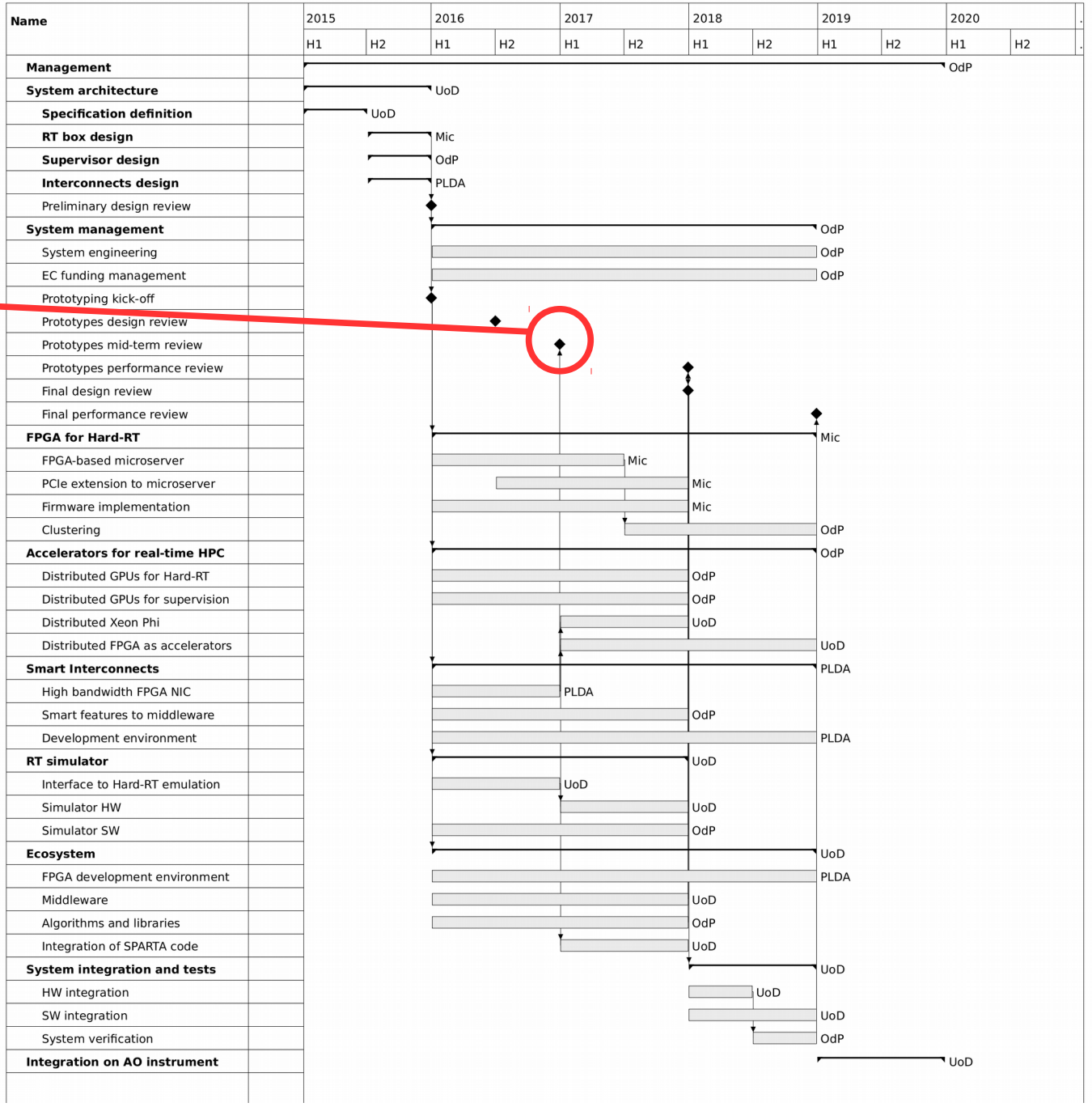
WP1

- Prototypes design review
- Delayed due to unavailability of components (FPGA)
- Could not happen between Oct. and Nov. 2016
- Prototyping activities had started on the basis of PDR results
- 9 months check review in Brussels did the job



WP1

- Prototyping mid-term review
- Held on 1st Feb. 2017
- 2-3 months delay
- Received a lot of feedback from community from panel
- Helped to prepare 18 months check review in Brussels which was a success

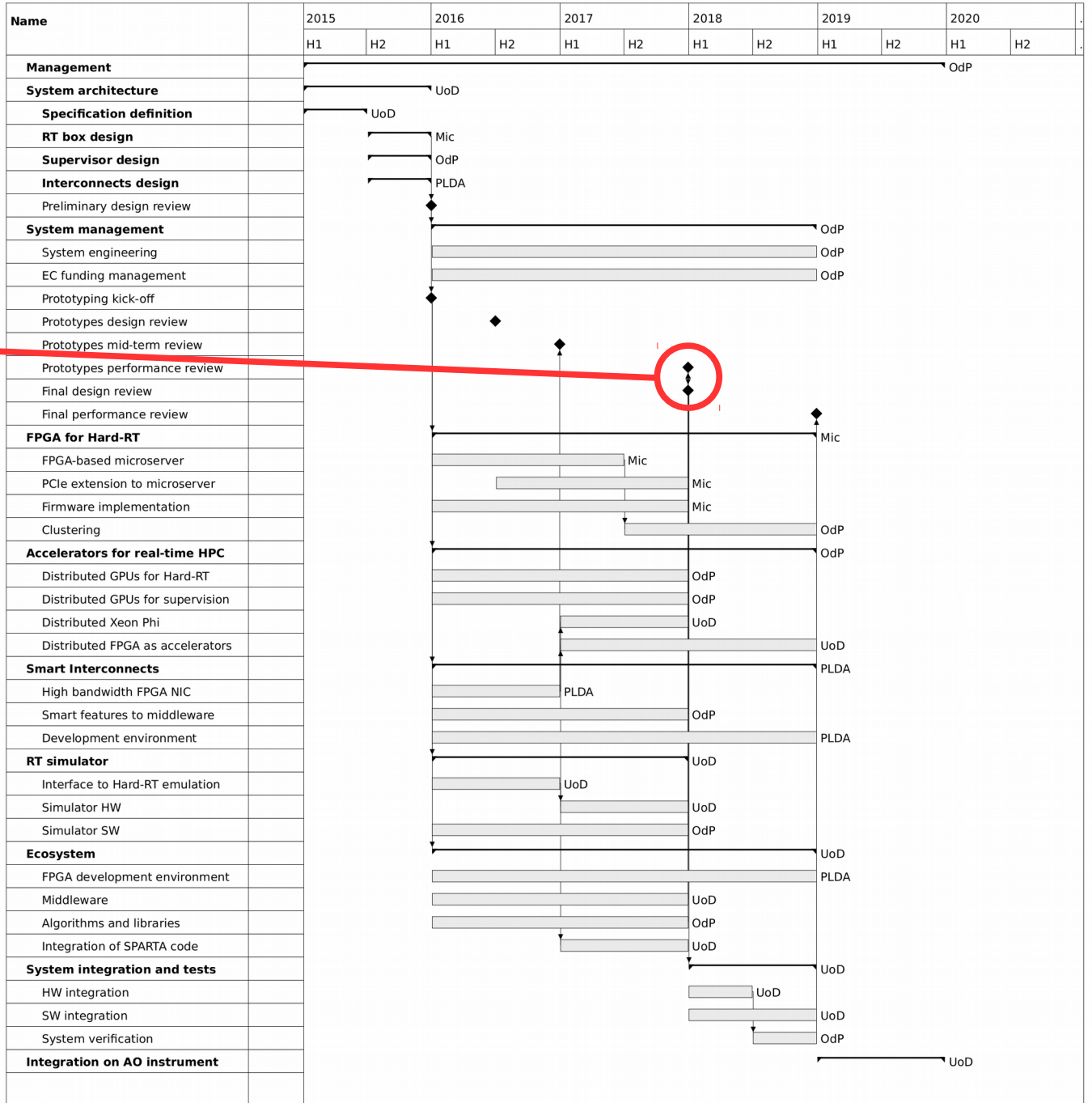


WP1

Prototyping performance review and final design review for full scale prototype

Today !

5-6 months delay

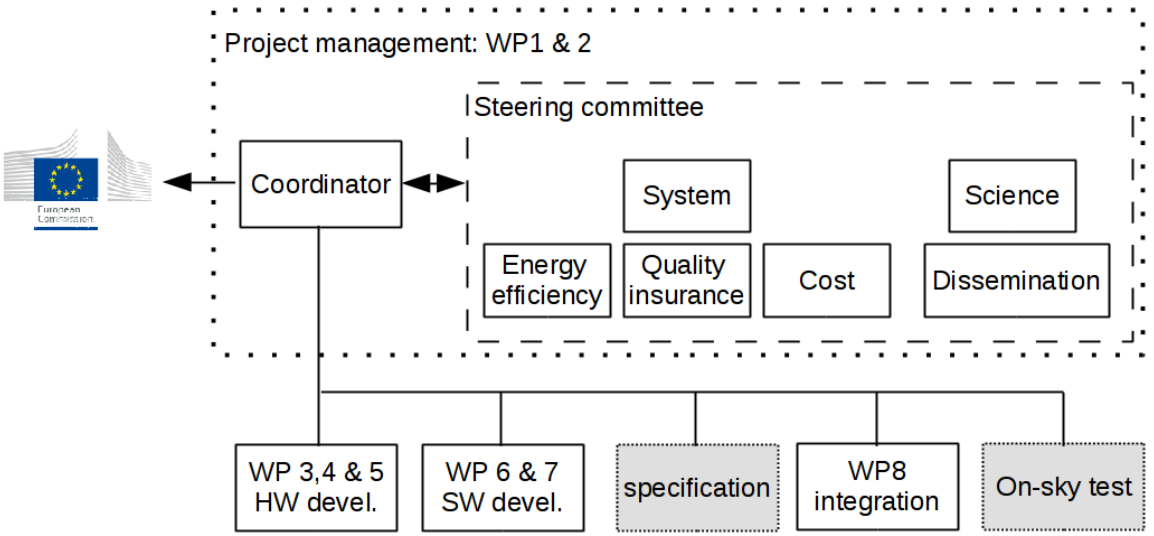


April 6th 2018

WP1: project management

Task 1.1: project management (OdP)

- Coordination of steering committee
- Meeting on a monthly basis



Deliverables

- D1.6: data management plan (OdP – M6) : **descoped**
- D1.7: project periodic report 1 (OdP – M18 moved to M24)
- D1.8: project periodic report 2 (OdP – M36) : moved to M39
- D1.9: project final report (OdP – M36) : moved to M39

WP1: project management

Task 1.2: Dissemination strategy

Science activities

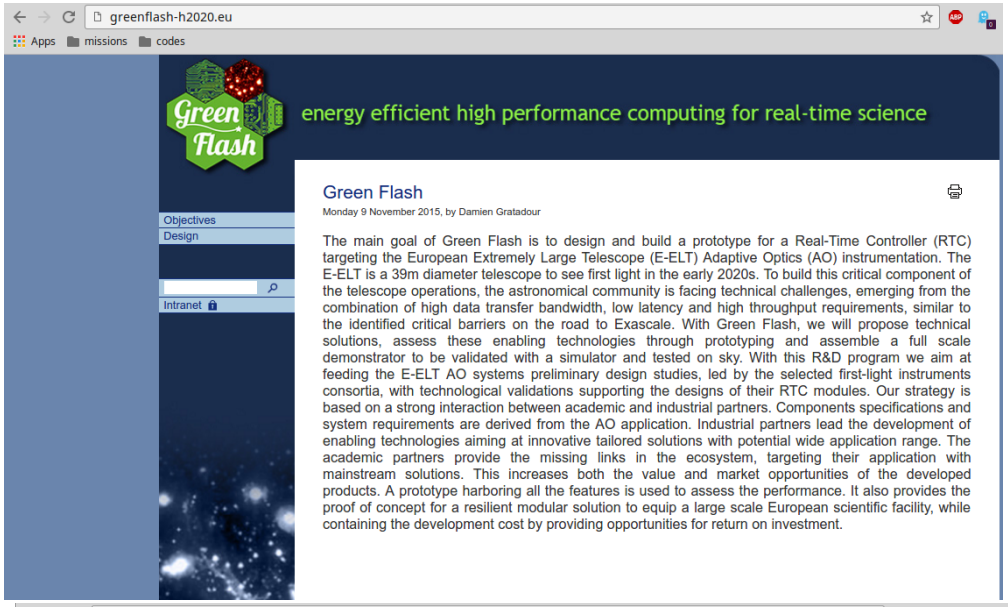
Publication / communication strategies

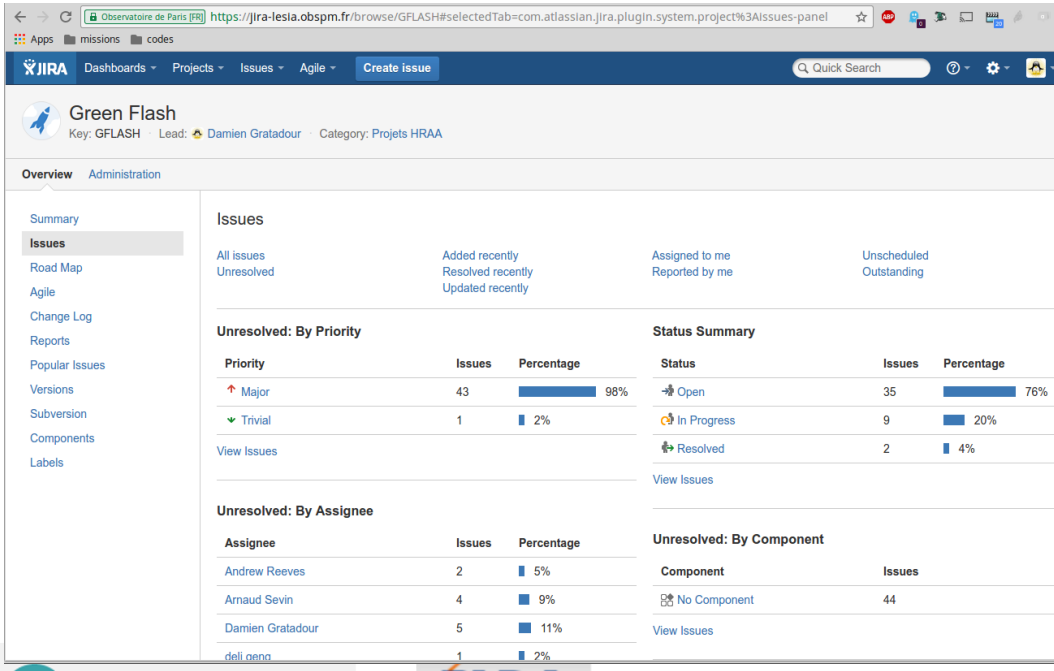
Strategy for on-sky experiment

Deliverables:

- D1.1: project internal website (OdP – M6)
- D1.2: project public website (OdP – M6)

WP will run over the course of the project





WP1: project management

Communications:

- EXDCI workshop during European HPC summit week (May 2016)
- GPU technology conference (April 2016, March 2017)
  - Blog post on NVIDIA website : https://blogs.nvidia.com/blog/2016/05/23/reshaping-light-rays-with-gpus/
- SIAM ALA 2018 (May 2018)

Publications

- 3 contributions to SPIE'16 conference on astronomical telescopes and instrumentation (Edinburgh, June 2016, to be published)
  - Gratadour et al. : overview of the project
  - Perret et al. : interconnect prototyping
  - Ferreira et al. : simulations and error budget
- 9 contributions to AO4ELT5
  - Biasi et al. (FPGA microserver), Reeves et al. (simulator), Perret et al. (smart interface), Doucet et al. (supervisor), Bernard et al. (GPU RT core), Jenkins et al. (Xeon Phi RT core), Ferreira (simulator, PSF reconstruction), Vidal et al. (simulation), Petit et al. (implementation)
- 12 submissions to SPIE 2018 covering all Wps and including several talks
- 1 paper at IPDPS 2018 (Ltaief, Gratadour et al., alternative supervisor strategies)
- Several papers in prep. : Ferreira et al. (A&A: accepted, IEEE: submitted), Doucet et al. (in prep), Bernard et al. (in prep.), etc ..

WP1: project management

Task 1.3: Energy efficiency (Mic)

Definition of criteria to assess energy efficiency

Lead the monitoring of energy efficiency for each sub-system and report

Deliverables :

- D1.3: Energy efficiency report (Mic – M36 moved to M39)

Task 1.4: Quality insurance (PLDA)

Management of the development process to secure quality of deliverables

Description of environments and promoted standards

Deliverables

- D1.4: Development process report (PLDA – M36 moved to M39)

Task 1.5: Cost reduction (Mic)

Lead a market survey to identify trends

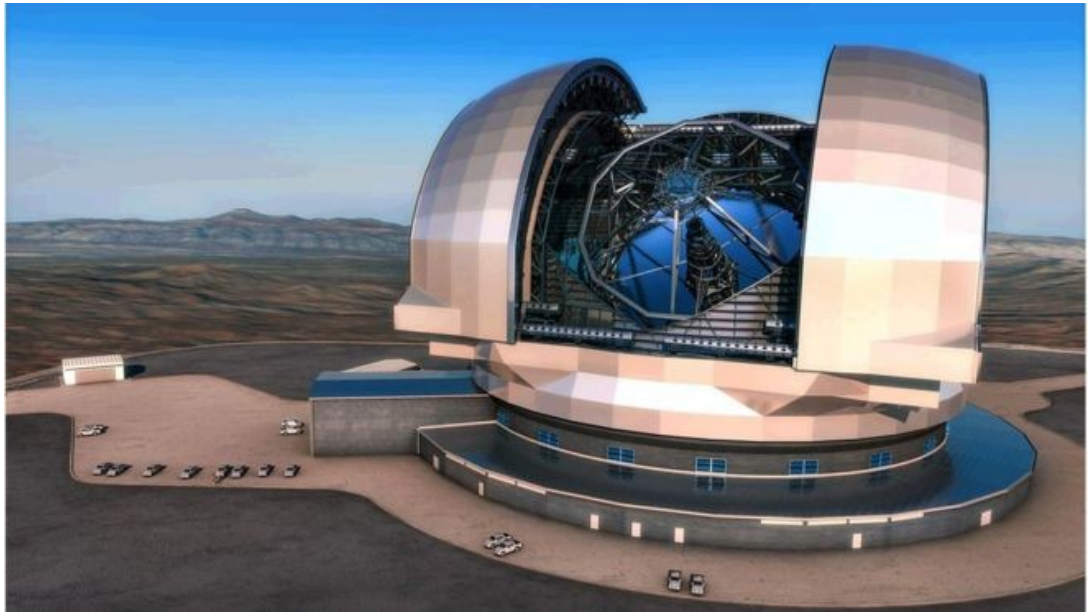
Procedures to monitor production cost and reporting

Deliverables

- D1.5: Cost projection report (Mic – M36 moved to M39)

Interaction with external partners

- Follow up on E-ELT instruments design specifications
  - After PDR and MTR
- Collaboration with ESO
  - ESO representative participating to steering committee meetings as observer
- Collaboration with O. Guyon (Subaru)
  - Enhancing the readiness level of the CACAO framework (multi-platform open source AO RTC SW)



Part of standards has become public by end 2017

- Collaboration with HPC partners
  - Collaboration with Extreme Computing Research Center (ECRC) @KAUST on optimized linear algebra
  - Several papers published and in preparation
  - Co-supervision of PhD thesis

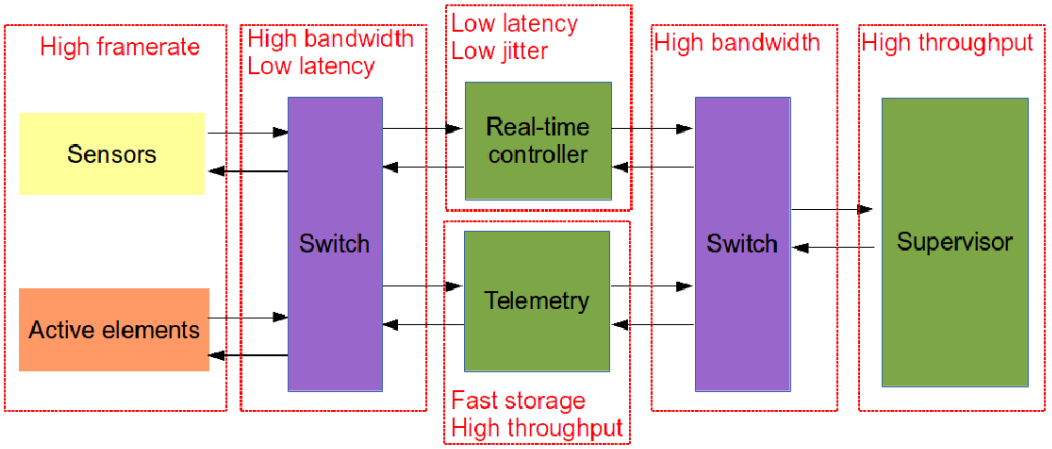
جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology
Extreme Computing Research Center

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

WP2: system management

Task 2.1: System architecture

- Requirements capture and specification definition
- Strategies for RT box, supervisor and interconnect



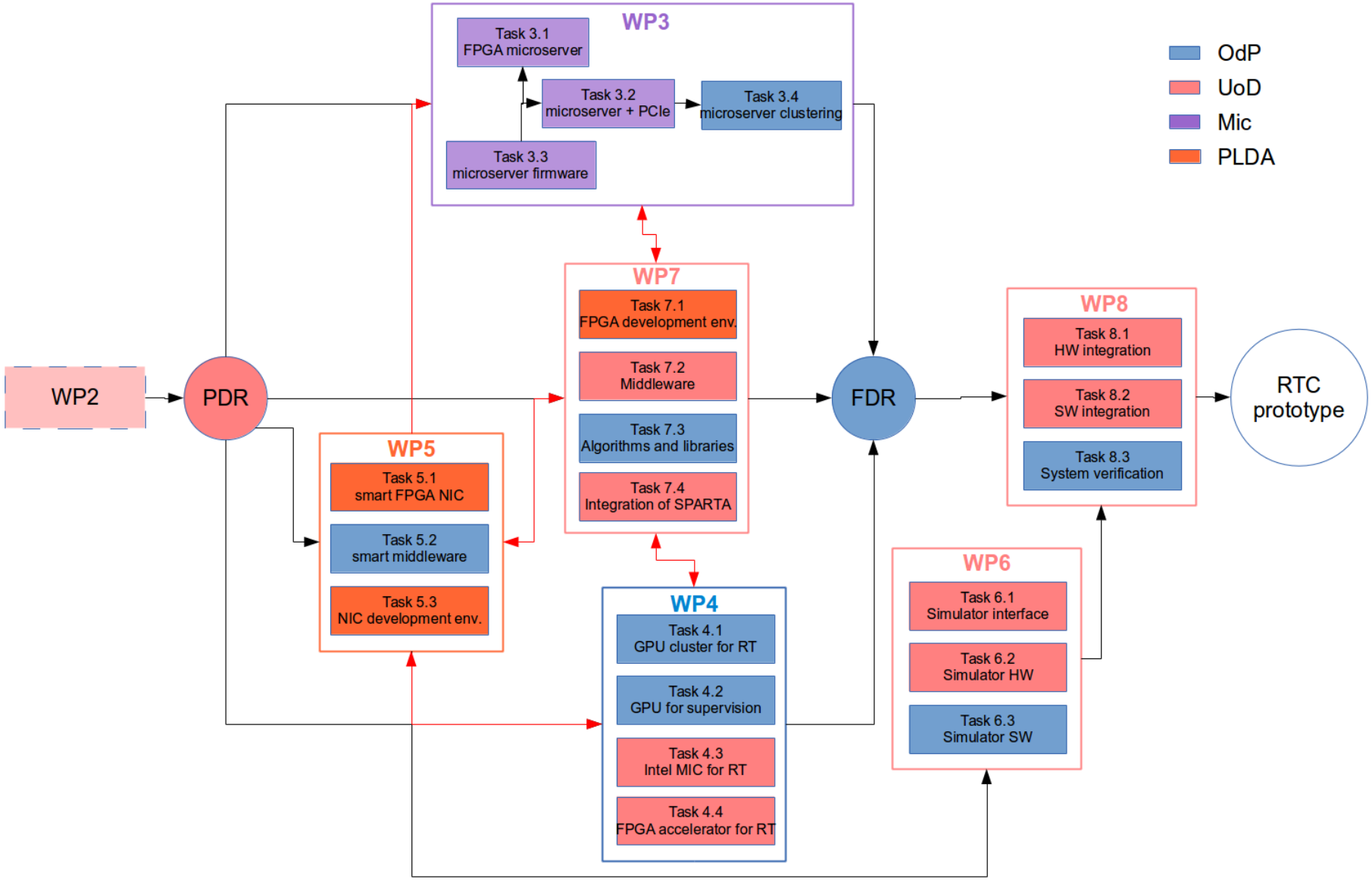
Deliverables (all internal):

- D2.1a: requirements document (UoD - M6)
- D2.1b: preliminary design document (UoD – M12)
- D2.1c: RTC box architectural options definition (Mic – M12)
- D2.1d: Supervisor module architectural options definition (OdP – M12)
- D2.1e: Interconnect strategies definition document (PLDA - M12)

Preliminary Design review held on January 2016

- PDR documentation + feedback from panel

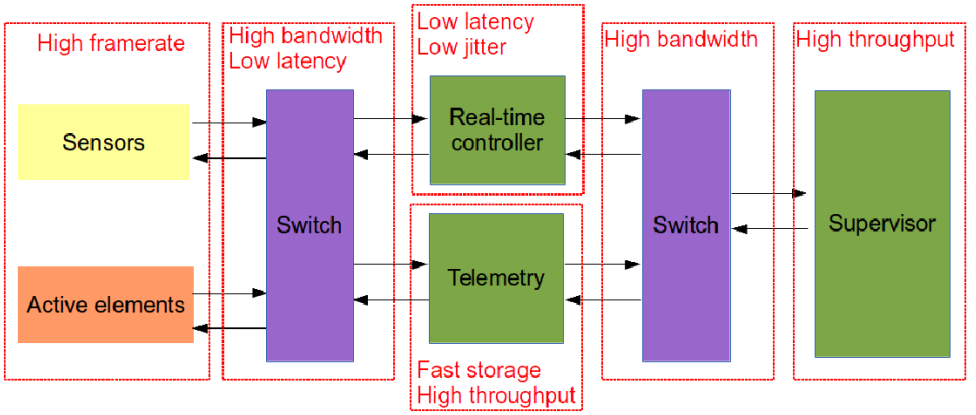
Work Packages interaction



WP2: system management

Task 2.2: System engineering

- General coordination of prototyping activities
- Produce regular progress reports
- Organize mid-term and final reviews



Deliverables

- D2.2a: Prototypes performance mid-term report (OdP – M12 moved to M18)
- D2.2b: Prototypes final performance report (OdP – M24 moved to M30)
- D2.2c: Full scale prototype final design report (OdP – M24 moved to M30)
- D2.2a: Full scale prototype final performance report (OdP – M36 moved to M39)

Staff at OdP and start dates

Damien Gratadour – Project lead – 01/10/15 (project management)

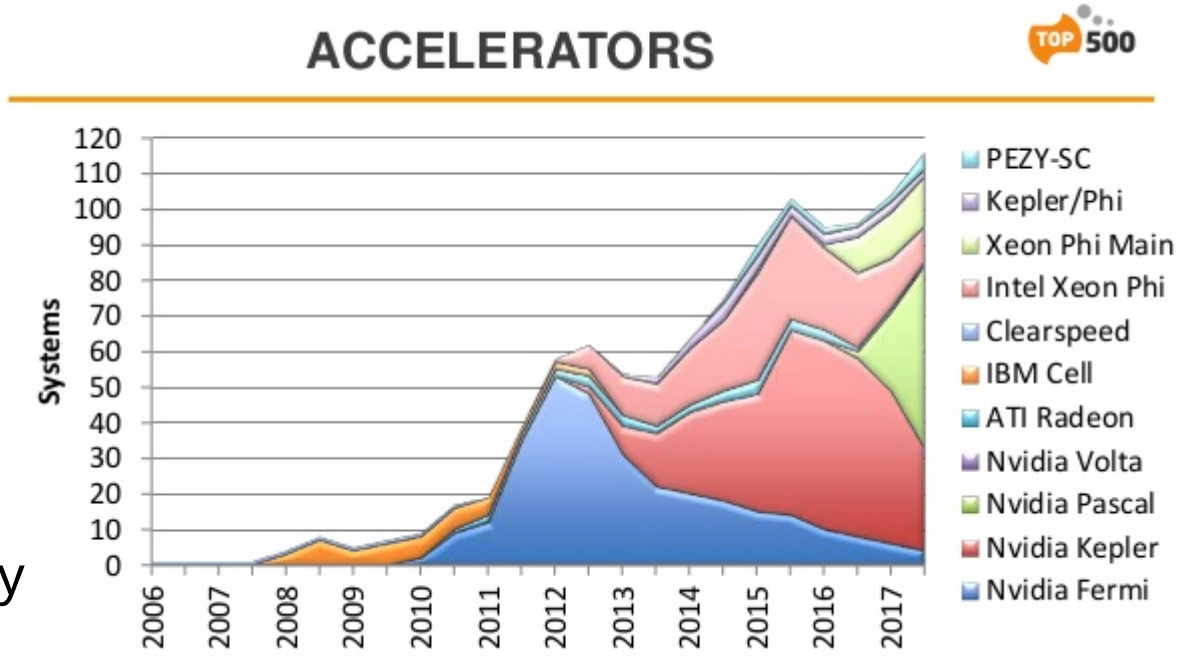
- Eric Gendron – Project scientist – 01/10/15
- Arnaud Sevin – System engineer – 01/10/15 (project coordination)
- Denis Perret – FPGA specialist – 01/10/15 (FPGA development coordinator)
- Roderick Dembet – SW development – 01/01/16 → 01/04/16
- Julien Bernard – research engineer SW – 01/10/15 (WP4 lead)
- Florian Ferreira – PhD student – 01/10/15 (WP6 – Simulator validation)
- Maxime Lainé – SW engineer, FPGA development – 01/10/15 → 31/12/18 (WP4 – WP5 coordination)
- Nicolas Doucet – research engineer SW / PhD student – 01/10/15 (WP4 – algorithms)
- Sébastien Durand – SW engineer – 01/01/16 → 01/09/17 (WP6 – simulator)
- Other support: Tristan Buey (project management), Julien brulé (sys. admin.)

HPC landscape: supercomputers

- US and China are dominating the high end HPC landscape
  - Strategic investment: “computing is the new sinews of war”
  - China is developing its own components and cluster architectures
  - US: 3 Exascale supercomputers projects (2020-2022)
    - IBM + NVIDIA + Mellanox x2, with existing technologies
    - Intel x1: unknown technology (previsously scheduled Xeon Phi based system has been descoped)
- Portable SW stacks are expanding
  - Programming models: OpenMP, OpenACC, StarPU, OmpSs, etc ...
  - Numerical libraries: Chameleon, MAGMA, etc ...
  - Deployment tools: containers
- Efficient / Smart interconnect strategies are critical to many applications
  - Core component of hyperscale computers
  - Integrators effort: Cray Aries, Fudjitsu Tofu, etc …
  - Vendors effort : Intel OmniPath, NVIDIA NVlink, IBM OpenCAPI
- Private investment has become far larger than government funding
  - GAFAA et al. needs are the main drivers for new technology developments from main vendors
  - Hyperscalers are also maintaining the largest and most advanced HPC facilities

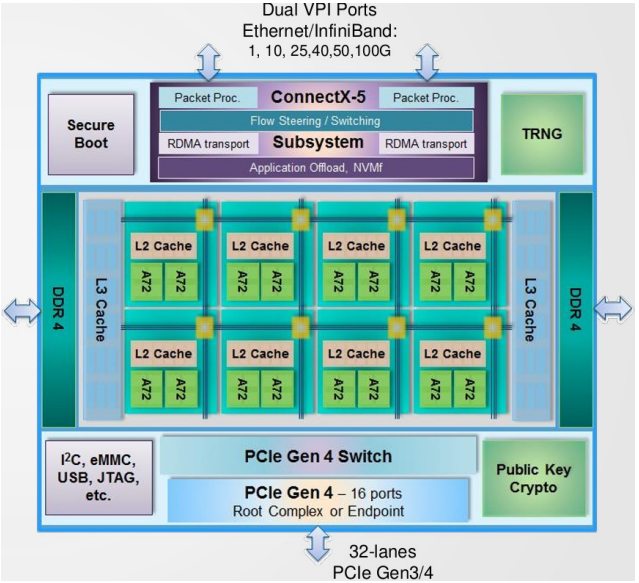
HPC landscape: co-processors

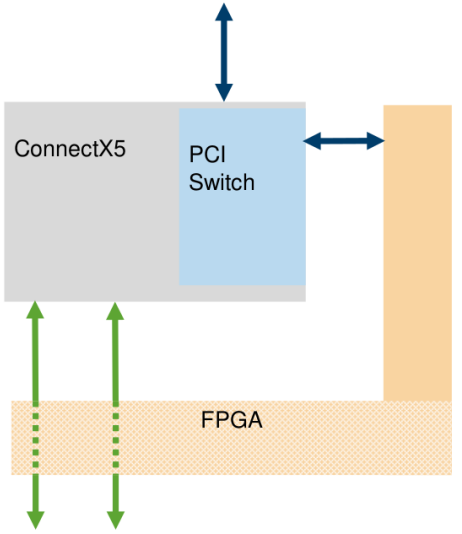
- GPU technology is the leading trend for HPC applications
  - 2 over 3 US exascale computers projects
  - growing user community: 75% of the systems with HW accelerators in the top500
  - Also dominating the Green500 list
- New HPC usage for industrial applications
  - Dedicated vendor developments
  - Long term support, backward compatibility



HPC landscape: interconnects

- Mellanox new general purpose interconnect solutions
  - BlueField (smart NIC) and Innova (flexible NIC) lines of products

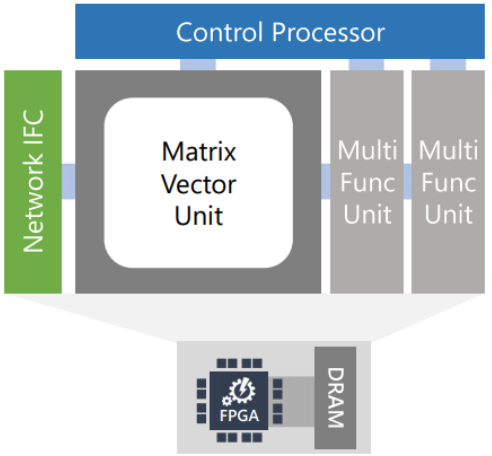


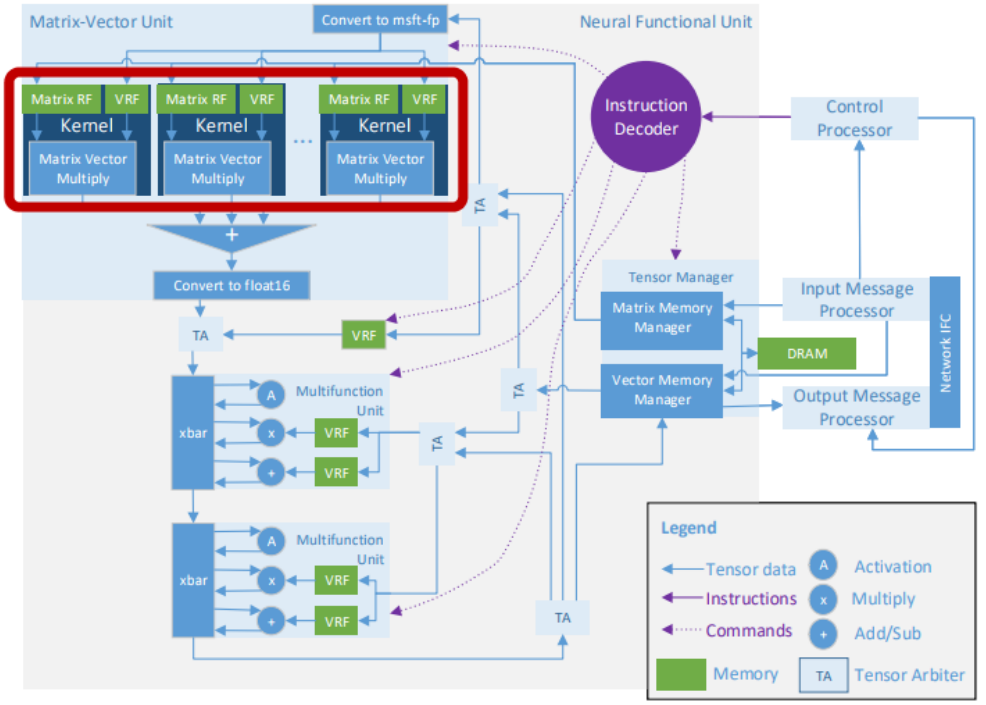


- Well aligned with Green Flash smart interconnect and microserver developments
  - But: Green Flash solutions provide more flexibility / modularity at the cost of lower level of maturity

HPC landscape: real-time computing

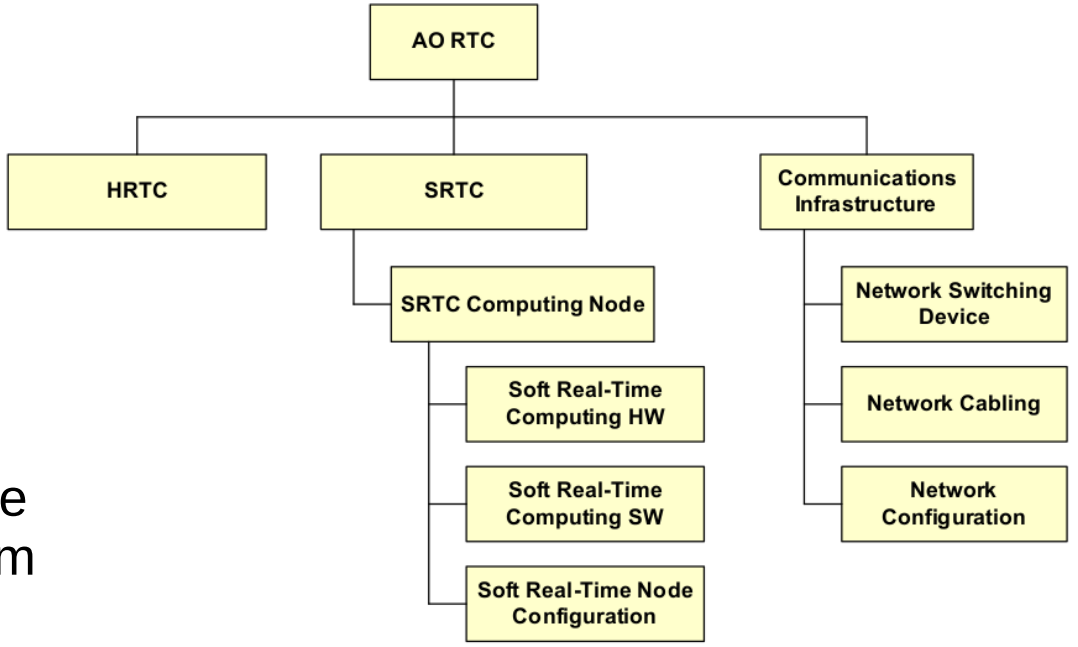
- At the bleeding edge: Microsoft brainwaves project
  - Real-time computing for AI
  - Soft DPU + persistent kernels





- Well aligned with Green Flash solution on GPUs: persistent kernels
  - While a custom approach today, will probably become mainstream in GPU computing for other applications

E-ELT context



- ESO standard for RTC architecture
- **Green Flash concept fits well with current breakdown structure**
  - communication infrastructure as independent subsystem
  - HRTC loosely constrained
  - SRTC SW stack includes ESO RtcToolkit which includes compatibility with Green Flash solutions
- No dedicated subsystem for data storage
- Ongoing prototyping study on x86 technology for HRTC contracted by ESO

Green Flash final prototypes

- Main finding from HPC landscape and AO requirements: there are several solutions to build an AO RTC
  - Trade-off between Performance, Scalability, Maintenance cost, etc ...
- HPC market moving fast, cannot predict mainstream in 5 years
- Our strategy: implement several “state of the art” prototypes instead of a single (meaningless) flagship prototype
  - 1 GPU (+NVlink) based prototype at OdP
    - Will include smart interconnect (data acquisition + HW simulator) from PLDA
    - Will include µXComp boards
  - 1 Xeon Phi (+OmniPath) based prototype at UoD
  - 1 FPGA (+PCIe) based microserver prototype at Mic
    - Including µXLink and µXComp boards
- Follow subsystem breakdown from ESO standards (HRTC, SRTC and ComIF)

Green Flash achievements (so far)

- Demonstrate real-time performance on various multi-coprocessors environment
  - On GPUs and Xeon Phis, including scalability
- Design and prototype a smart interconnect solution from a network feed to the GPU memory
  - With dedicated development environment for FPGA, scalable and portable
  - With performance consistent with real-time operations
- Design and prototype an accelerator board based on high end FPGA coupled to high bandwidth HMC memory
  - Available as COTS product from Microgate
- Design a microserver concept, based on high end FPGA with ARM HPS
  - First prototypes being manufactured
- Implement a scalable, portable and optimized AO supervision loop, leveraging up to 40 TFLOPs on a single GPU server
  - Bringing down to the minute the time-to-solution for the AO tomographic reconstrutor computation at E-ELT scales
- Enable an AO real-time simulator including
  - Real-time data shaper and fakeCam/DMC concept based on FPGA
  - Super fast end-to-end simulation tightly integrated with interconnect system

Green Flash outcomes (so far)

- MG and collaborators have been selected by WMKO to build the next generation RTC for the Keck II telescope and that it is our intention to sign a contract with Keck
  - smart interconnect relying on µXlink board for HW interfaces (versatile low power interface close to the instrument)
  - GPU server as a compute engine (in the computer room)
  - Partnership between Microgate, Swinburne University (Melbourne) and OdP, 2 years project
- µXLink concept will also be used as smart interconnect in other projects:
  - E-ELT M4 (entering MAIT after this FDR), GMT adaptive secondary, TMT and Subaru adaptive M2 (still in feasability phase)
- QuickPlay tool from Accelize has gone from the pre-release level to a mature tool deployed in the cloud (OVH, AWS)
  - Work led in Green Flash enabled new features and helped in increasing reliability
- Academic partners in Green Flash are involved strongly in various E-ELT instrument consortia
  - MICADO, HARMONI, MOSAIC
  - Will be able to apply most of Green Flash “recipes” to build these RTC systems (lower cost, higher reliability)

Summary

Project has progressed consistently for about 30 months

- PDR occurred in Jan. 2016 with feedback from community
- Prototyping activities have been assessed during MTR in Feb. 2017
- Entering final prototyping phase with few months delay
- Requested a few months extension: full scope of project will be covered

Complex project structure

- Strong effort on OdP side for system management

Collaborations initiated

- Good feedback from the community on different aspects (HPC + instrumentation)
- Evaluate the convergence and minimize additional effort

Enhancing the readiness level of commercial solutions

- Contribution to QuickPlay development (see Christophe's' presentation)
- Design of an innovative FPGA boards (see Roberto's presentation)