Introduction to EuroEXA

EuroEXA: Co-designed Innovation and System for Resilient Exascale Computing in Europe: From Applications to Silicon

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 754337

Enrico Calore

INFN Ferrara, Italy

enrico.calore@fe.infn.it

Exascale:
What does it really mean?

1. 1 Billion Billion (10^18) FLOPs or equivalent
2. €100M - €500M per system
3. 20MW - 60MW of power

EuroEXA response: concepts

- Optimizing the components of existing systems is not effective
- We need a holistic approach focusing on the entire stack:
  - Technology
  - Components
  - Architecture
  - Infrastructure
  - System software
  - Applications

EuroEXA in a nutshell

- **Energy efficiency**
  - Tight integration
  - Customized ARM processing chiplet and FPGA acceleration
  - Advanced cooling
  - Reduced Joules/bit transfer (memory compression, UNIMEM)
- **Scalability**
  - Mitigation of transfer cost (memory compression, UNIMEM, network geographic addressing)
  - Application co-design
  - Optimized programming environment (runtime systems, libraries, etc.)
  - Distributed storage on BeeGFS
- **Resilience**
  - Whole system resiliency cost–benefit analysis
  - ARM microarchitecture extensions
  - System software extensions

From application optimization to datacenter design

EuroEXA partnership

Commercial Partners

ICEOTOPE

SYNELIXIS

arm

MAXELER Technologies
Maximum Performance Computing

ZeroPoint Technologies

Academic/Gov. Partners

ICCS

MANCHESTER 1824
The University of Manchester

INAF ISTITUTO NAZIONALE DI ASTROFISICA

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

imec

FORTH
FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS

Fraunhofer

Hartree Centre
Science & Technology Facilities Council

ECMWF

NEURASMUS
A European Master in Neuroscience

INFN
Istituto Nazionale di Fisica Nucleare

Supporters

XILINX

SYNOPSYS
Silicon to Software

SKA
SQUARE KILOMETRE ARRAY

Met Office

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

janssen

CRS4
IDEAS BECOME LIFE

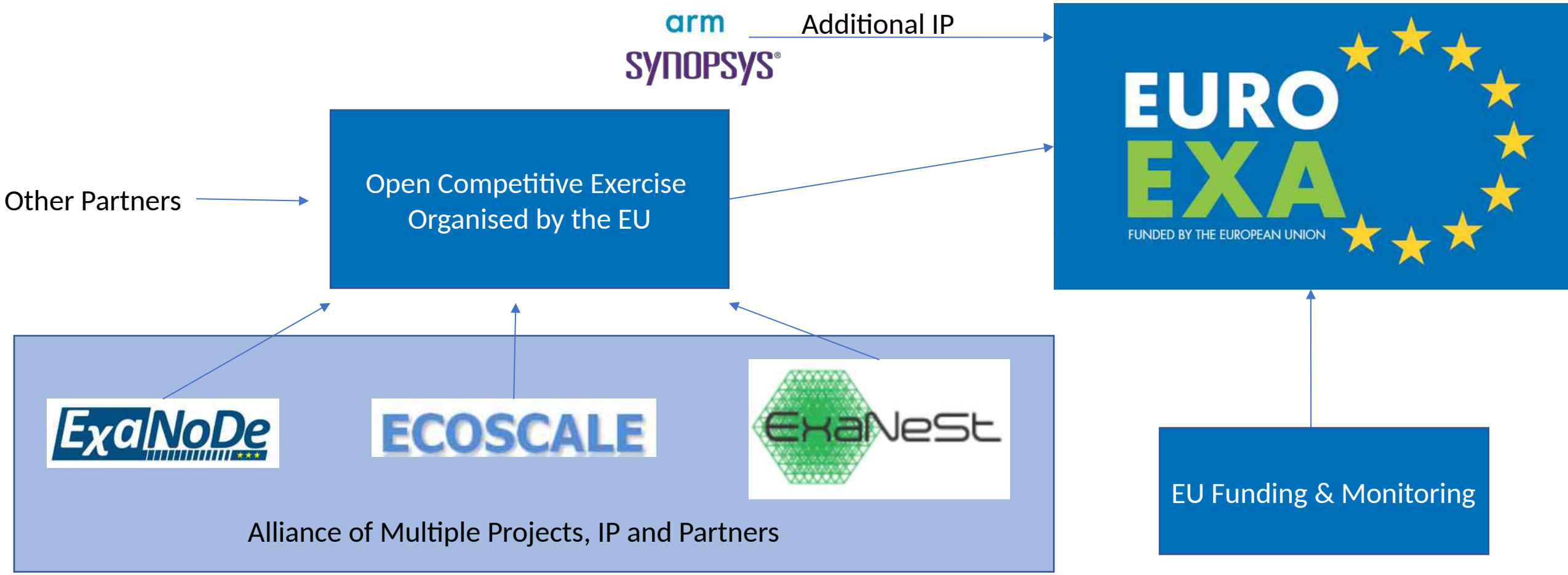
NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Schneider Electric

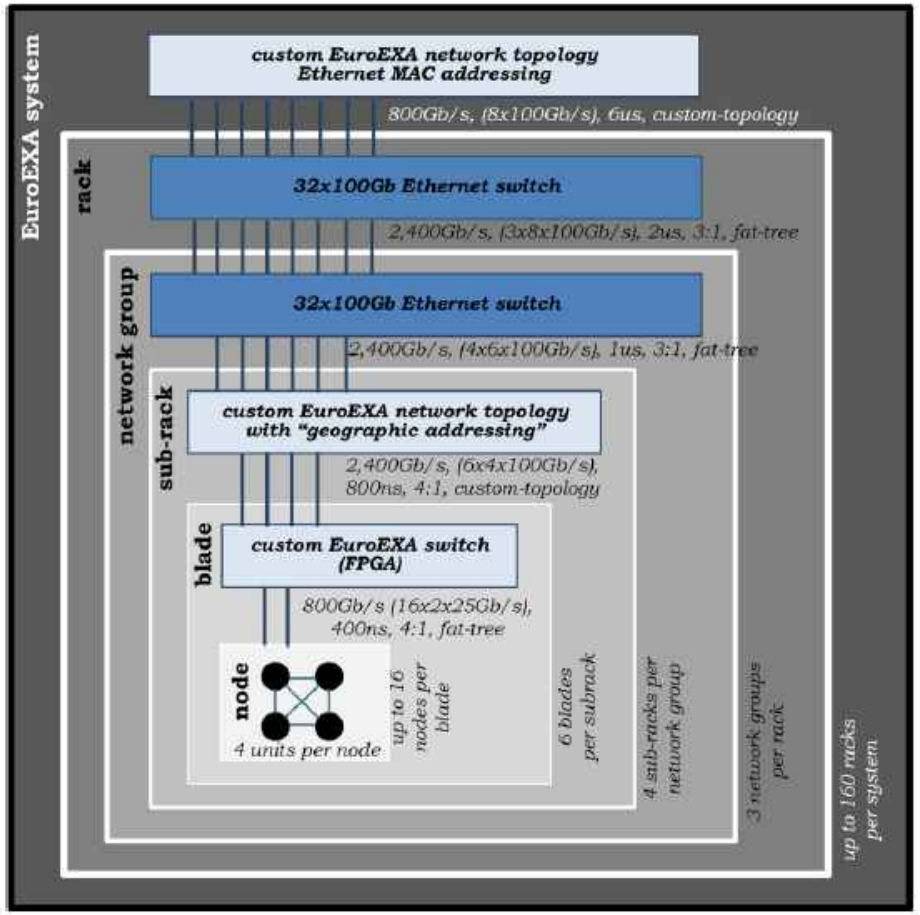
Noldus
Information Technology

SOLVAY
asking more from chemistry®

What is EuroEXA?



System architecture and technology



400 Pflops per system

2.7 Pflops per rack

920 Tflops per netgroup

230 Tflops per subrack

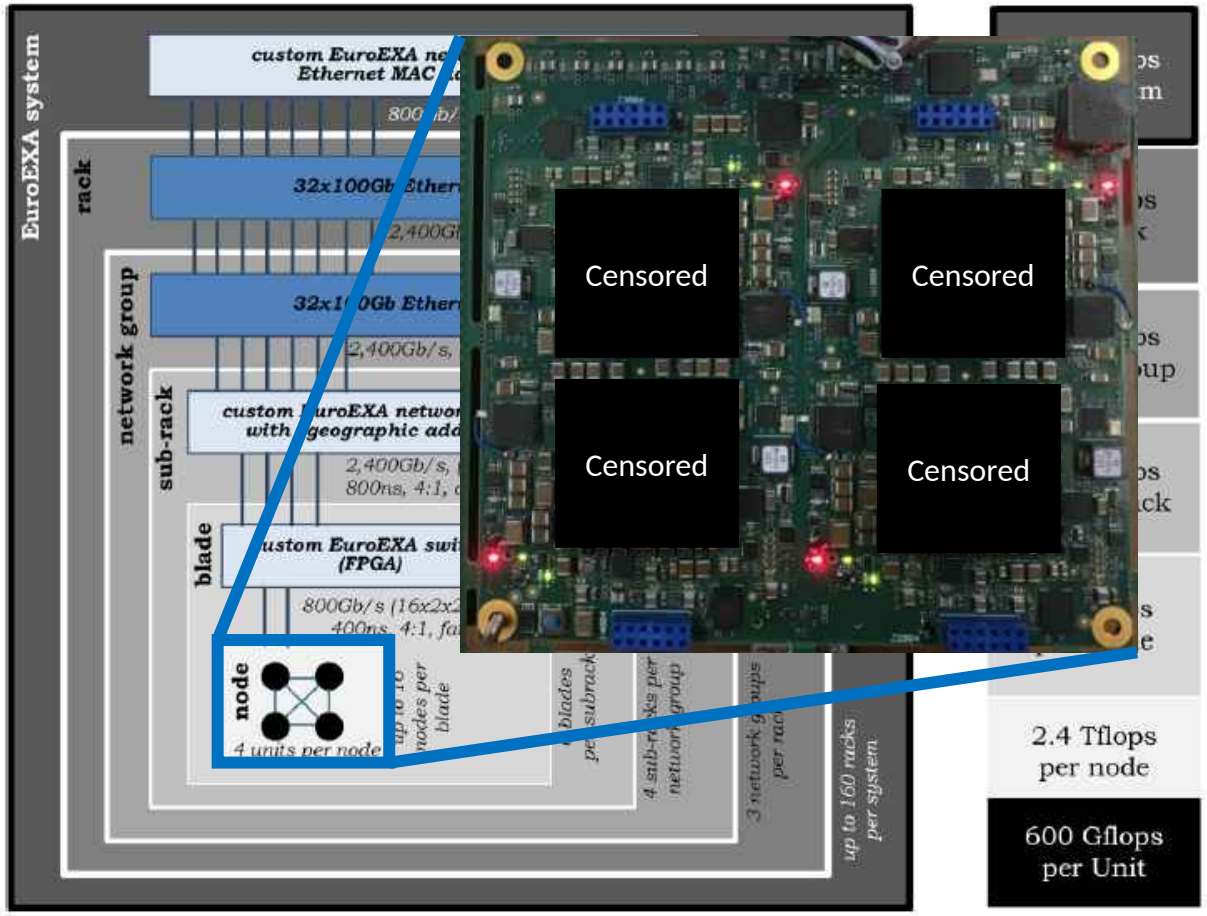
38 Tflops per blade

2.4 Tflops per node

600 Gflops per Unit

- ARM Processing and FPGA DataFlow
- UNIMEM Architecture with PGAS
- Distributed Storage on BeeGFS
- Memory Compression Technologies
- Unique Hybrid Geographically-Addressed, Switching and Topology Interconnect

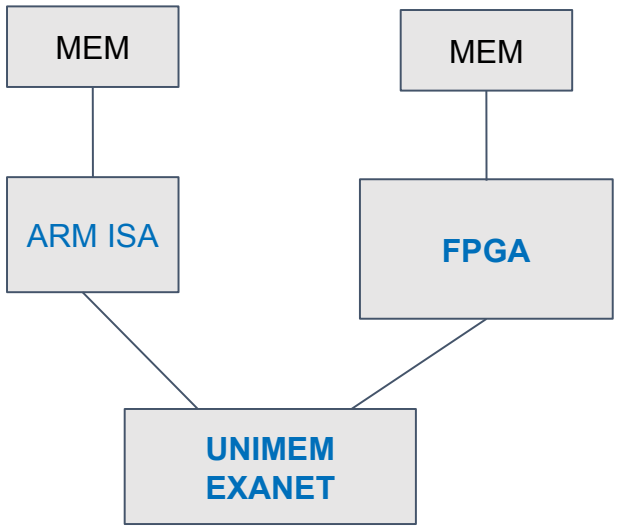
System architecture and



Technology from FORTH:

- 12cm x 13cm
- 4 ARM Processors and 4 FPGA Accelerators
- M.2 SSD
- 4 x SODIMMS + Onboard RAM
- Daughterboard style
- 160Gb/s of I/O

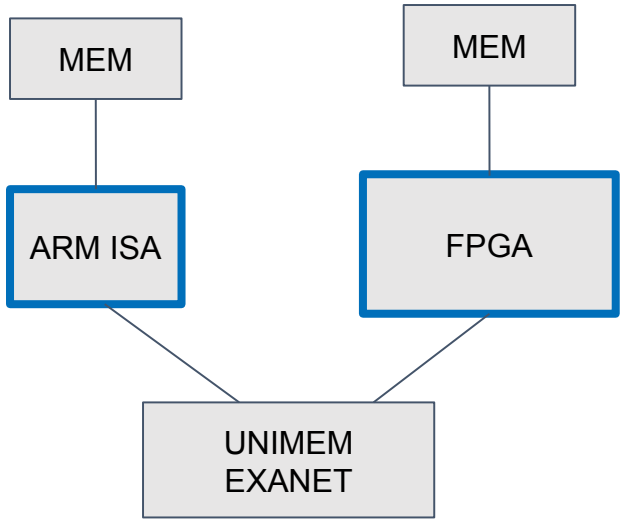
EuroEXA node architecture: some details



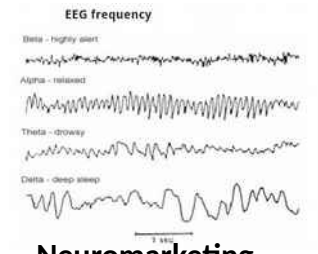
EuroEXA node architecture

- We employ **FPGAs** as our compute accelerator
- We innovate around the **ARM** ISA HW and SW ecosystem
- We scale-up with **EXANET** a low-latency, HPC network
- We support Global Shared Address Space (GSAS) with **UNIMEM**

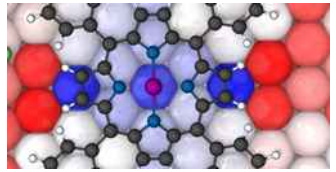
Application porting and optimization

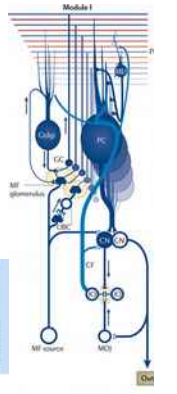


14 applications being ported and optimized for ARM + FPGA

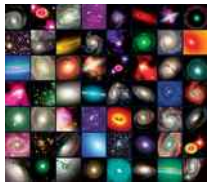


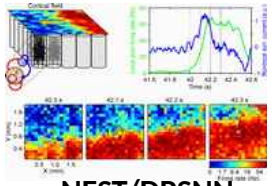
Neuromarketing

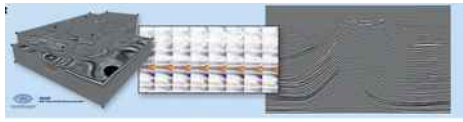

Quantum Espresso

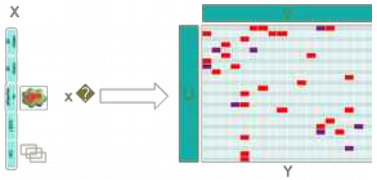

InfOli

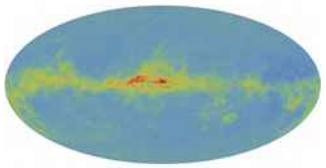

NEMO

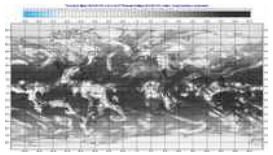

Astronomy image classification

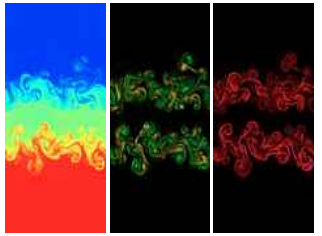

NEST/DPSNN

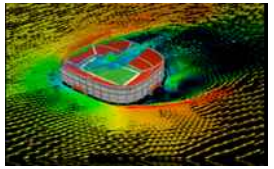

FRTM

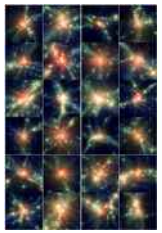

SMURFF

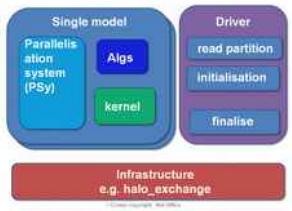

AVU-GSR


IFS


LBM

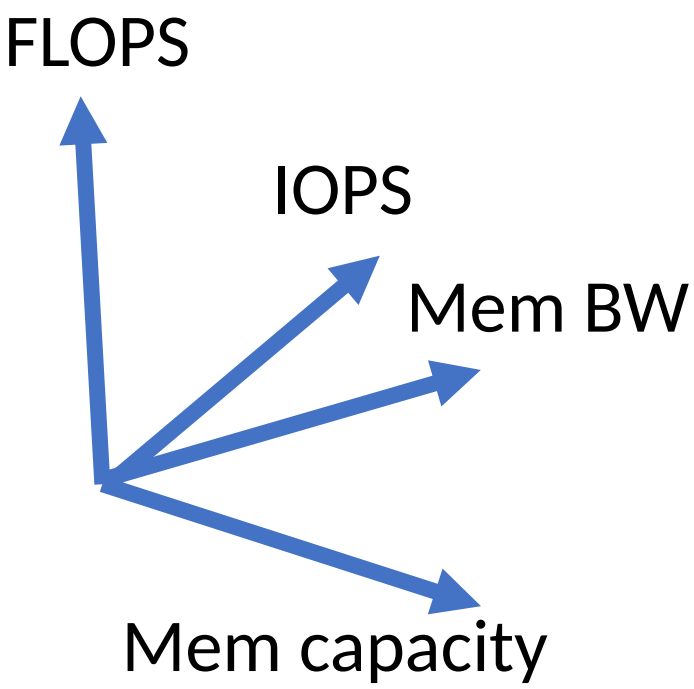

Alya

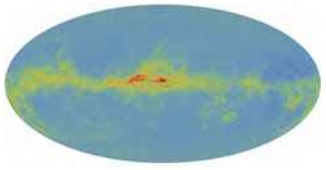

GADGET

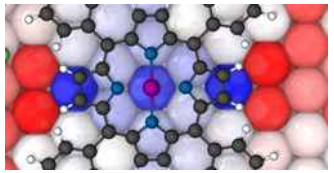


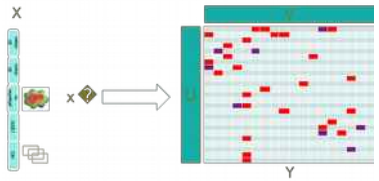
LFRic

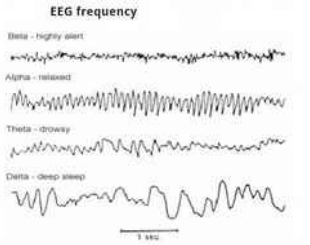
Co-design, demonstration and evaluation using exascale-class apps



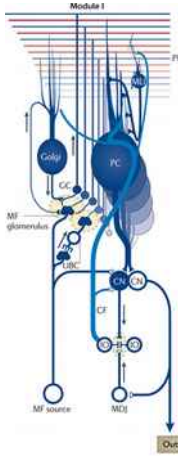

AVU-GSR


Quantum Espresso

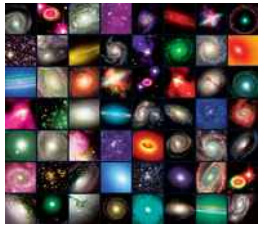

SMURFF

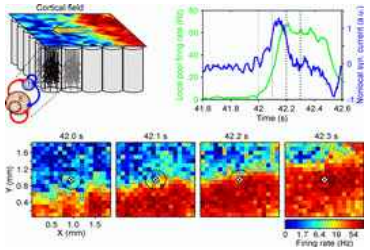


Neuromarketing

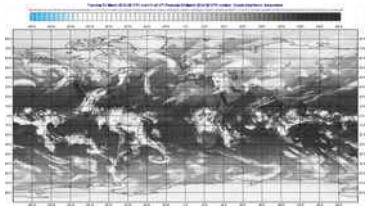

InfOli

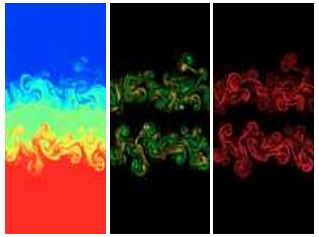

NEMO

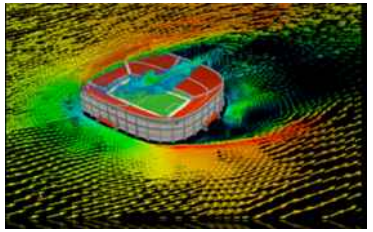

Astronomy image classification

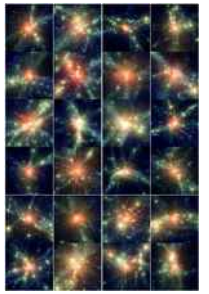

NEST/DPSNN

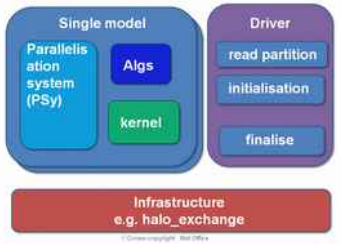

FRTM


IFS


LBM

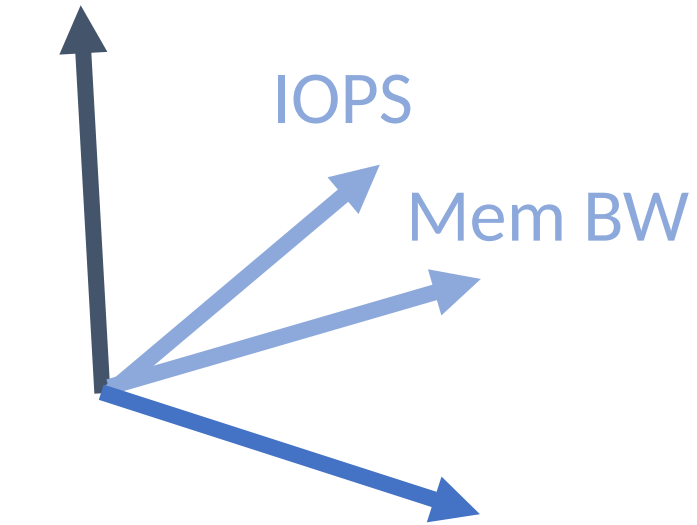

Alya

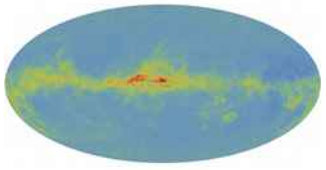

GADGET

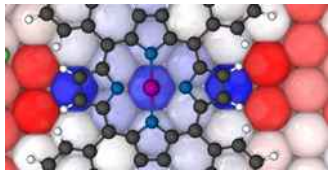


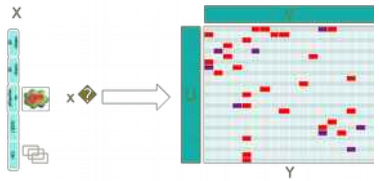
LFRic

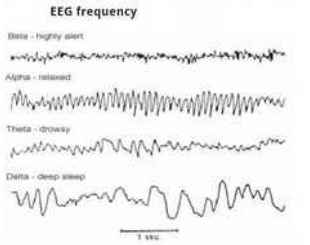
Co-design, demonstration and evaluation using exascale-class apps



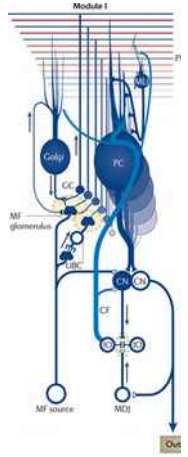

AVU-GSR


Quantum Espresso

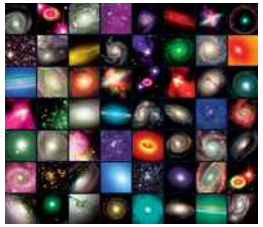

SMURFF

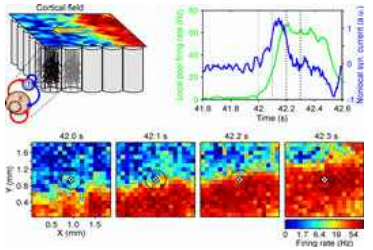


Neuromarketing

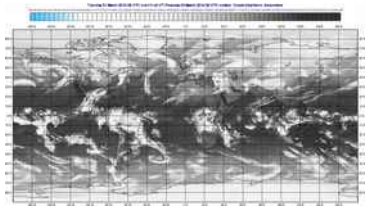

InfOli

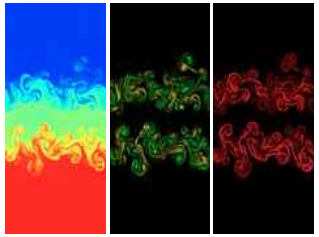

NEMO

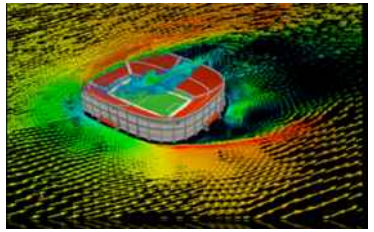

Astronomy image classification


NEST/DPSNN

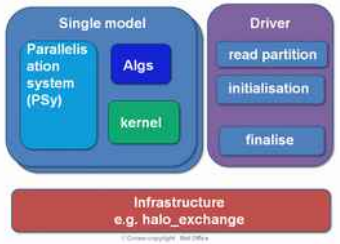

FRTM


IFS


LBM (collide)

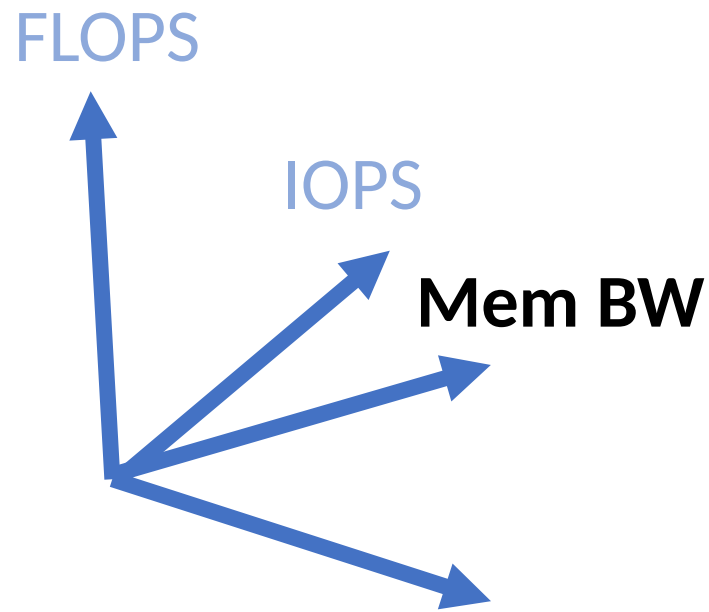

Alya

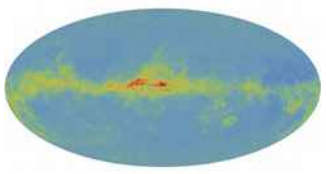

GADGET



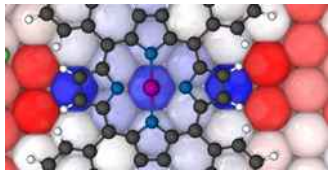
LFRic

Co-design, demonstration and evaluation using exascale-class apps

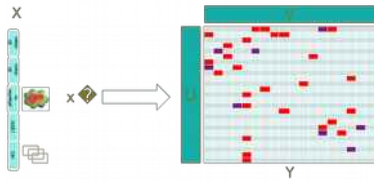




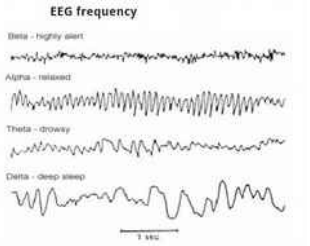
AVU-GSR



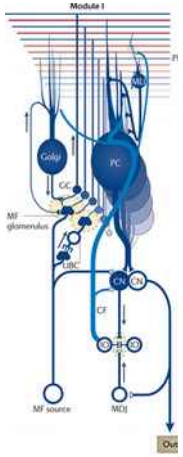
Quantum Espresso



SMURFF

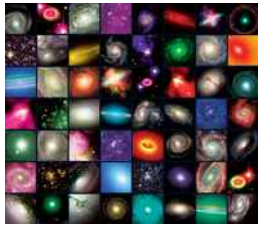


Neuromarketing

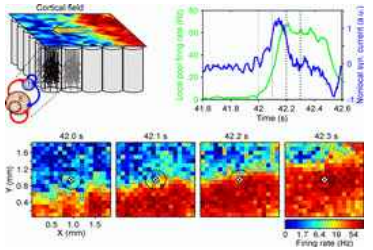


InfOli

NEMO

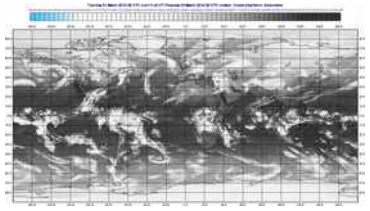


Astronomy image classification

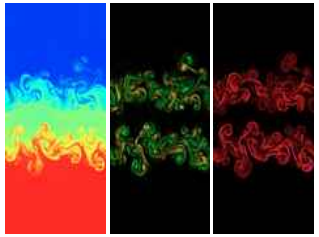


NEST/DPSNN

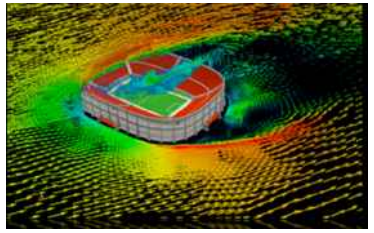
FRTM



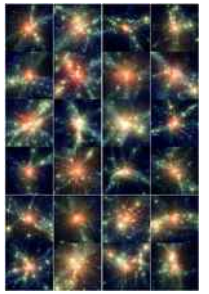
IFS



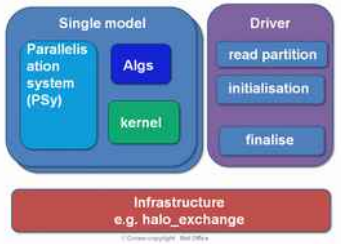
LBM (propagate)



Alya

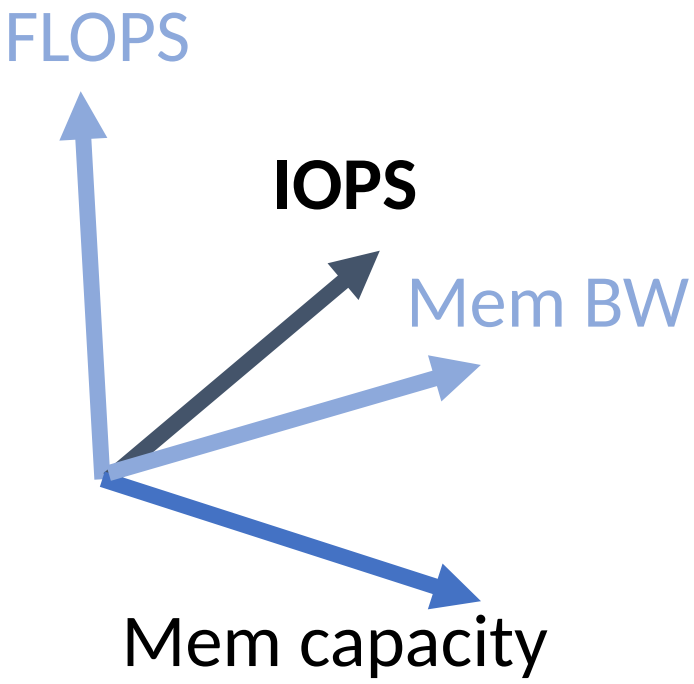


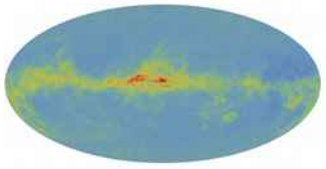
GADGET

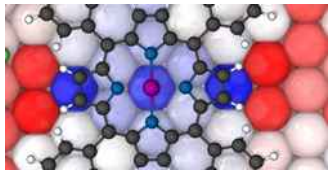


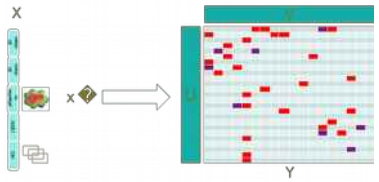
LFRic

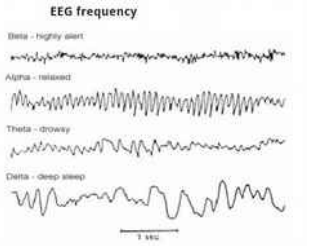
Co-design, demonstration and evaluation using exascale-class apps



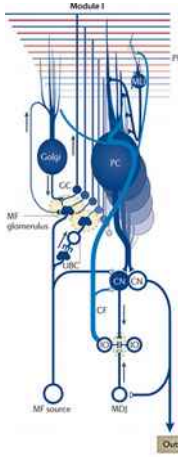

AVU-GSR


Quantum Espresso


SMURFF

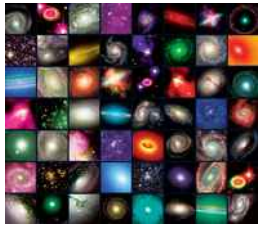


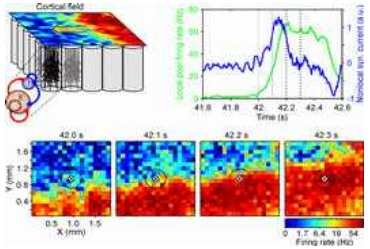
Neuromarketing



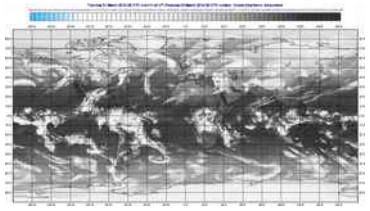
InfOli

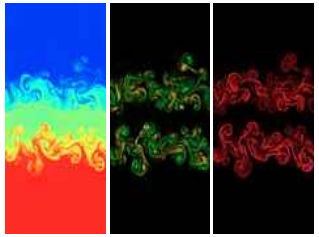

NEMO

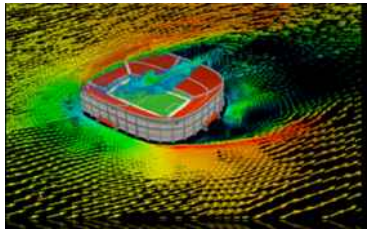

Astronomy image classification

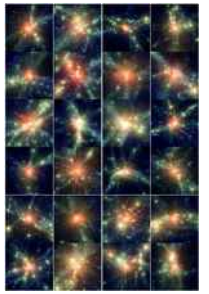

NEST/DPSNN

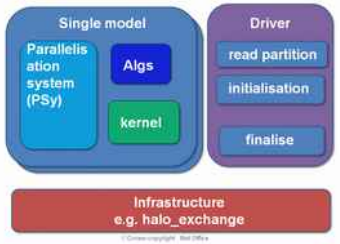

FRTM


IFS


LBM


Alya


GADGET



LFRic

Applications

Working together with a rich mix of key HPC applications from across:

- climate/weather, (ECMWF/STFC/UoM)
- physics/energy (INAF/INFN/Fraunhofer)
- life-science/bioinformatics (Neurasmus/Synelixis/BSC/IMEC)

System software

- OS and system SW adaptations
  - Device drivers, hyperconverged storage
- Programming runtimes extensions
  - MPI, task-based distributed-memory programming, FPGA programming
- Resource allocation optimization

How to program FPGAs?

Traditional low level approaches are difficult to embrace for Scientific HPC applications (e.g. using Hardware Definition Languages)

HPC scientific applications has to adapt to specific characteristics:

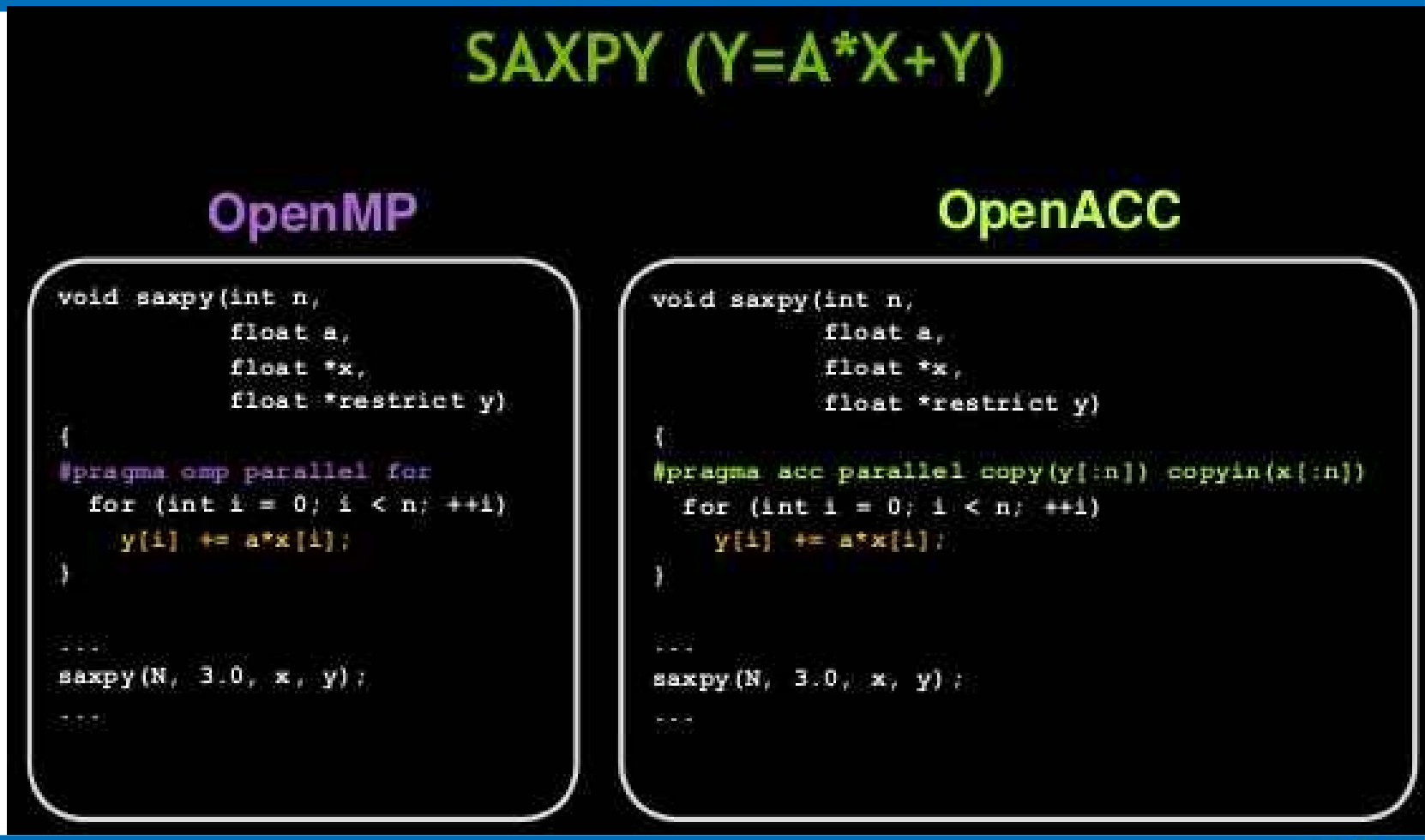
- Software lifetime may be very long; even tens of years.
- Software must be portable across current and future HPC hardware architectures, which are very heterogeneous.
- Software has to be strongly optimized to exploit the available hardware for better performances.

Directive based approach

- Code modifications could be minimal thanks to the annotation of pre-existing C code using #pragma directives.
- Programming efforts needed mainly to re-organize the data structures and to efficiently design data movements.
- If proting is needed, programming efforts would not be lost:
  - Also other directive based languages would benefit from data re-organization and efficiently designed data movements.
  - Switching between directive based languages should be just a matter of changing #pragma directives.

For CPUs and GPUs

- **OpenMP**
  Widely used for CPU multi-threading (lately supports also GPUs/accelerators)
- **OpenACC**
  Introduced for GPUs/accelerators

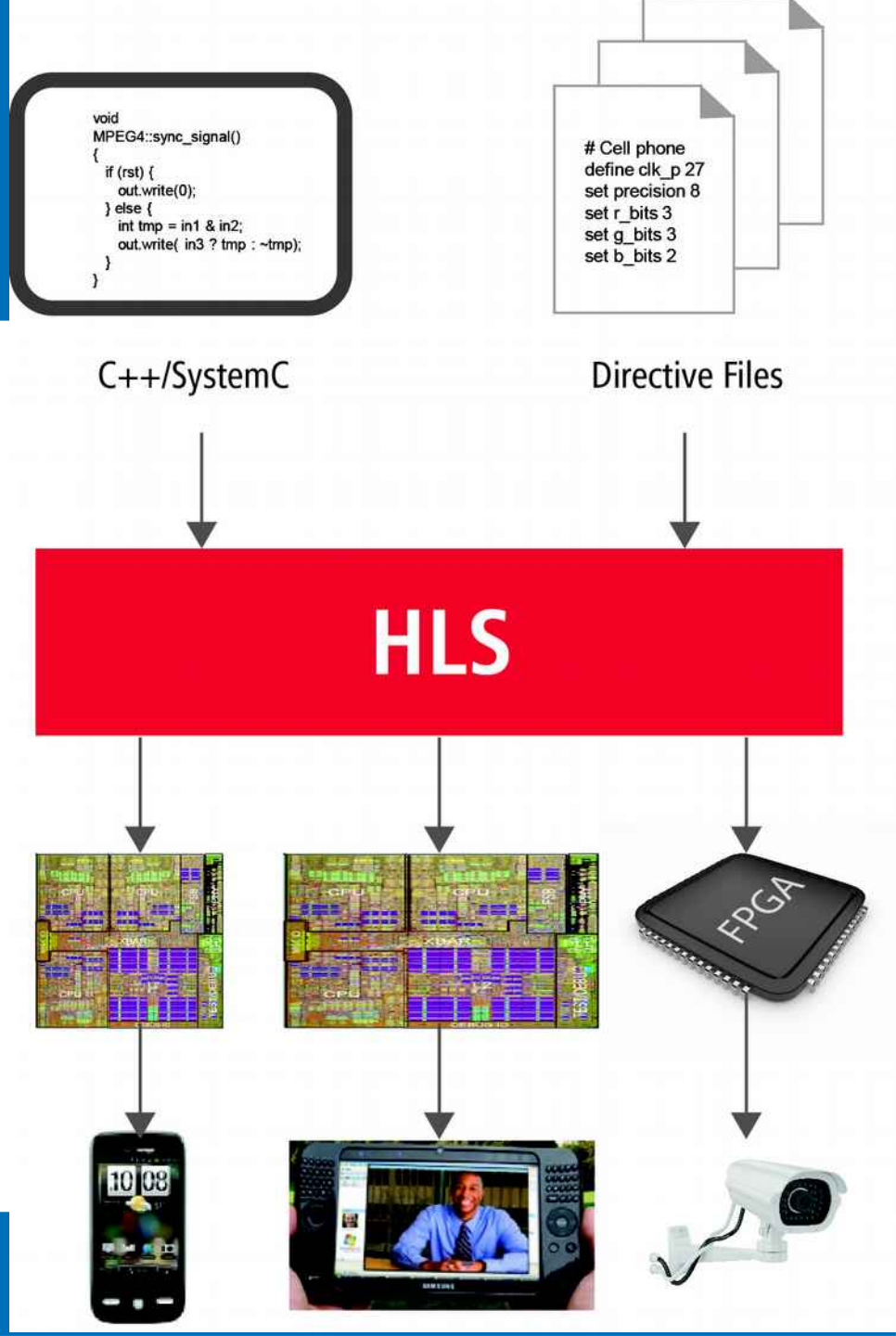


EUROEXA
FUNDED BY THE EUROPEAN UNION

For FPGAs…

High Level Synthesis approachs (aka Algorithmic-based) are getting popular due to accelerated design time and time to market:

HLS (High Level Synthesis)



One model for all: OmpSs

The **OmpSs** Programming Model

Programming model developed at BSC to extend OpenMP with new directives to support asynchronous parallelism and heterogeneity, including devices like GPUs and FPGAs.

For Xilinx FPGAs, it relies on the Vivado HLS compiler.