

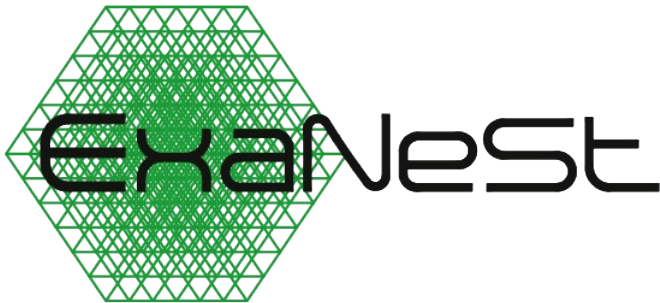
ExaNeSt@FPGA

INAF ICT Workshop 2019

David Goz

With

L. Tornatore, G. Taffoni, S. Bertocco,
A. Ragagnin, G. Murante and I. Coretti

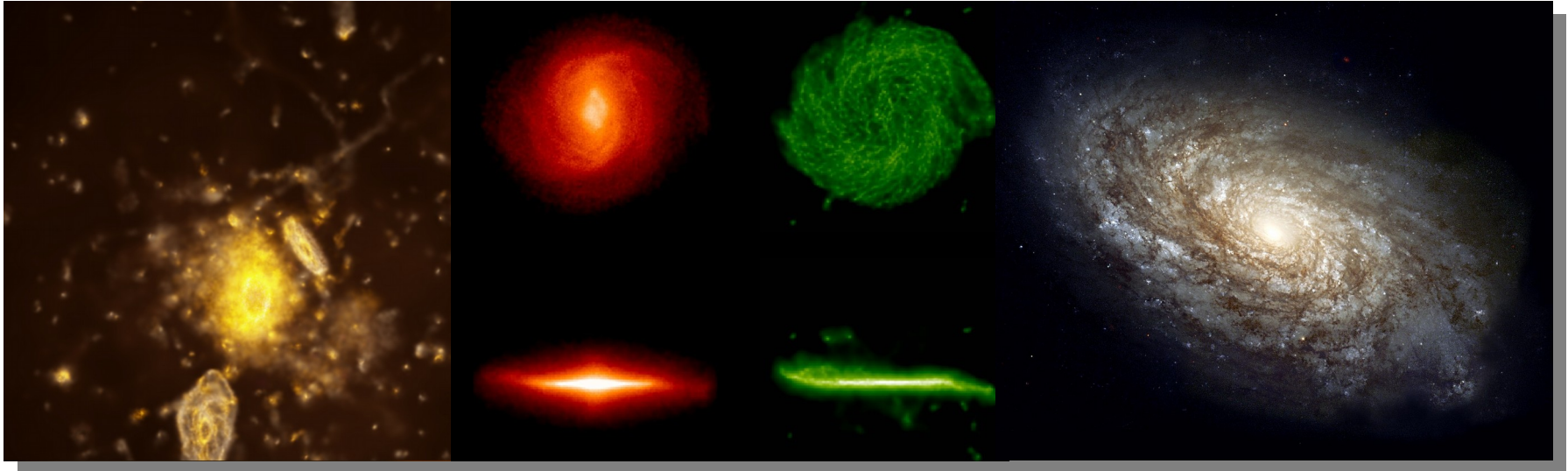


Horizon 2020



Next generation computing roadmap

Most of us rely on numerical codes to perform calculations.



Cosmological simulation of galaxy formation using GADGET code (Springel 2005).

Simulated disk galaxy in cosmological environment at present epoch (Goz et al. 2015).

NGC 4414, a typical spiral galaxy in the constellation of Coma Berenice, is about 60 million light-years away from Earth (Credit HST).

- **HPC** numerical simulations are one of the more effective instrument to compare observations with theoretical models;
- the new generation of observational facilities also implies **high performance data reduction** and **analysis tools**.

Why Exa-scale?

"Crucial problems that we can only hope to address computationally require us to deliver effective computing power orders-of-magnitude greater than we can deploy today".

DOE's Office of Science, 2012

"EXA-scale" is the necessary upscale step that HPC needs to achieve in the next years.

It is defined as the frontier of a sustained performance around 10^{18} flop/s with an energy performance of ~ 50 Gflop/W

There are deep consequences in the way we design, write, and optimize scientific codes.

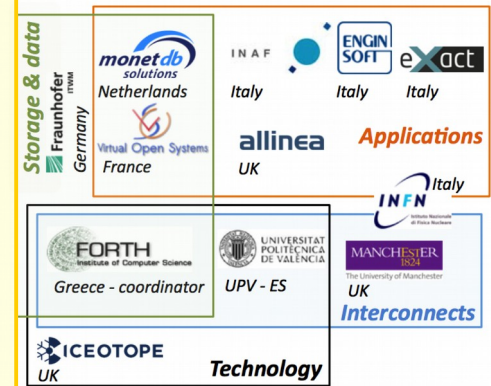
ExaNeSt European project

The **Horizon2020 ExaNeSt project** aims to demonstrate the feasibility of a European technology based ExaScale HPC system.

Who we are: the ExaNeSt consortium combines industrial and academic research expertise.

How we do it: following a **co-design** approach,

- applications drive the HW development and test it;
- applications are **re-designed to develop new HPC SW** able to exploit exascale HW.

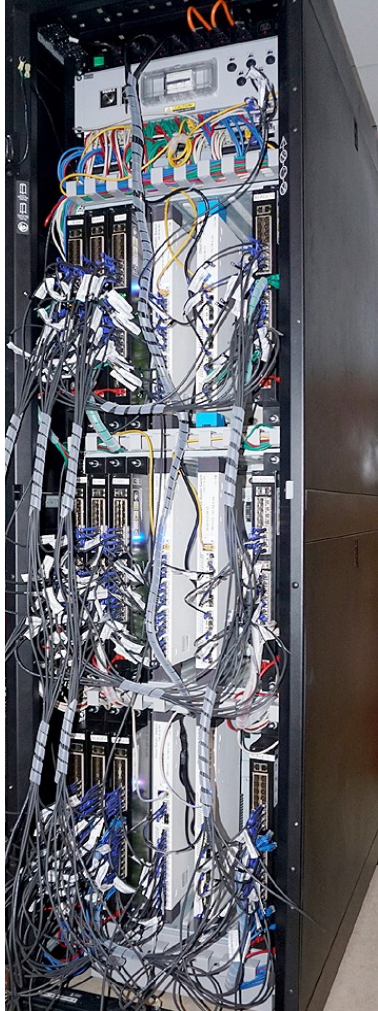


The ExaNeSt Quad-FPGA Daughter Board (QFDB).

ExaNeSt compute unit (QFDB):

- 4 Xilinx Zynq UltraScale+ FPGAs;
- 4 ARMv8 cores @1.5GHz per FPGA;
- 16 GB of DDR4 memory per FPGA;
- one NVM SSD storage device.

ExaNest European project



Testbed of ExaNest
in liquid cooling rack

Project information:

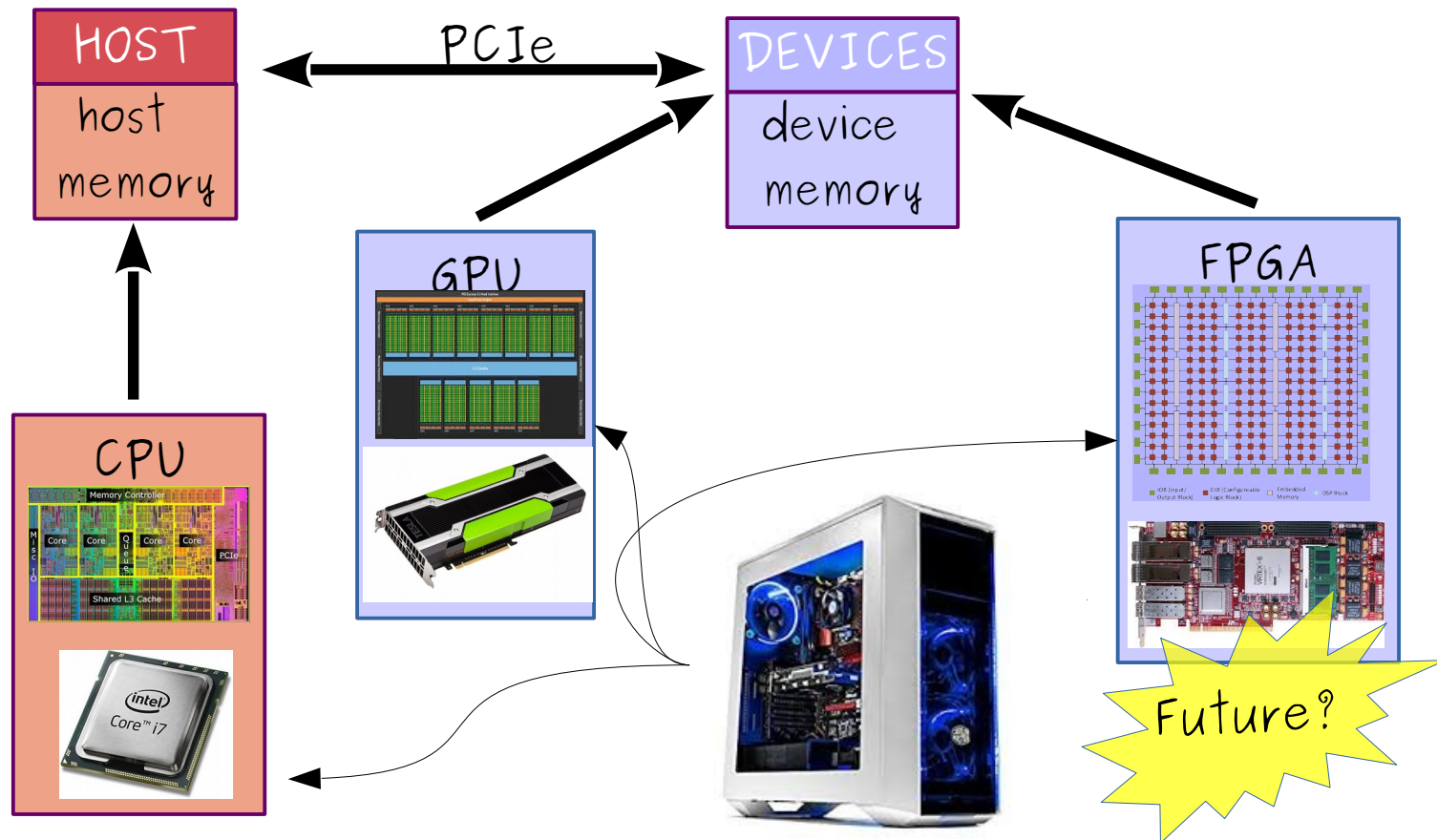
- web site: <http://www.exanest.eu>;
- duration 01 Dec. 2015 – 31 May 2019;
- budget: cost €8.44 M, EU contribution €8.44 M.

ExaNest testbed prototype:

- the HPC Testbed consists now of 6 liquid-cooled blades, which contain a total of 24 "QFDB" boards;
- 96 nodes (FPGA), 384 ARM A53 CPUs (1536 64-bit A53 cores), 1.5 TeraBytes of DRAM memory, and 6 TeraBytes of SSD storage;
- the Interconnection Topology is a 3D Torus, and the network interfaces feature ExaNest-own remote DMA engines with 1024 channels each, multiple mailbox queues, and resilience features;
- more blades will be added.

Heterogeneous hardware

Node level heterogeneous architectures compared to traditional CPUs offer **high peak performance**.



Embedded & mobile hardware

System-on-Chip (SoC) heterogeneous hardware compared to traditional hardware is more **energy and cost efficient**.

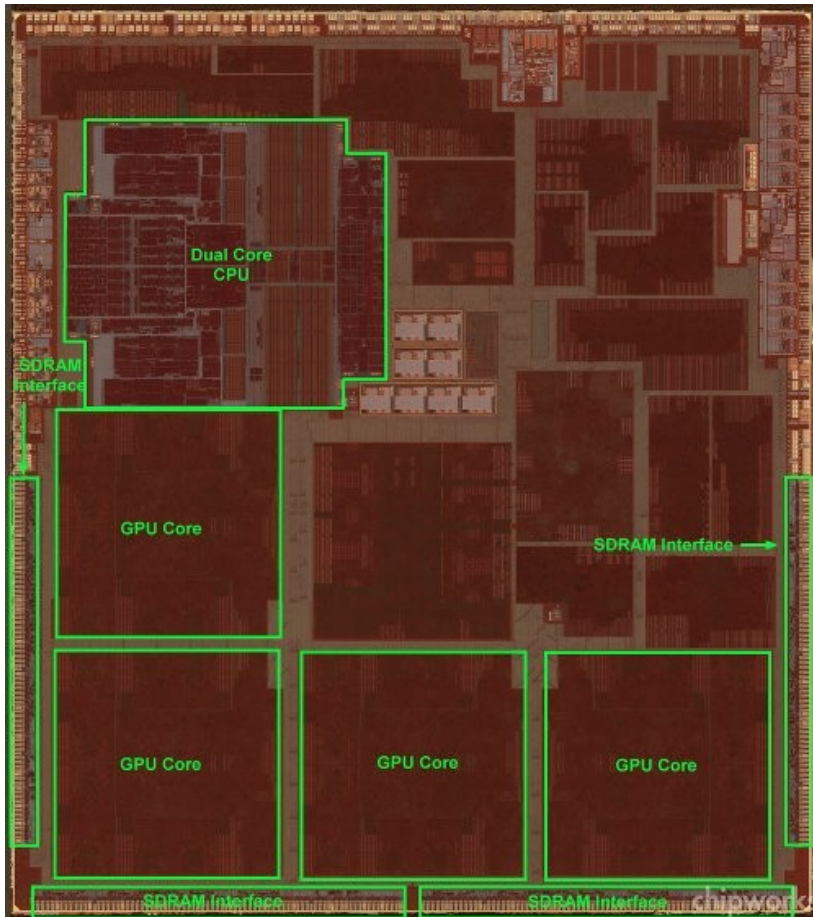
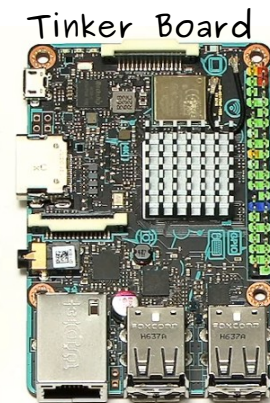
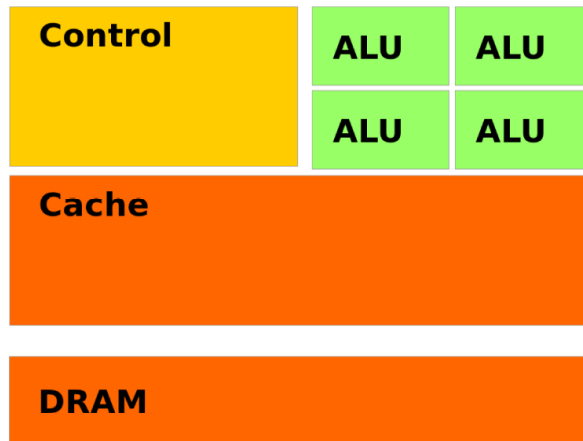


Photo from ChipWorks

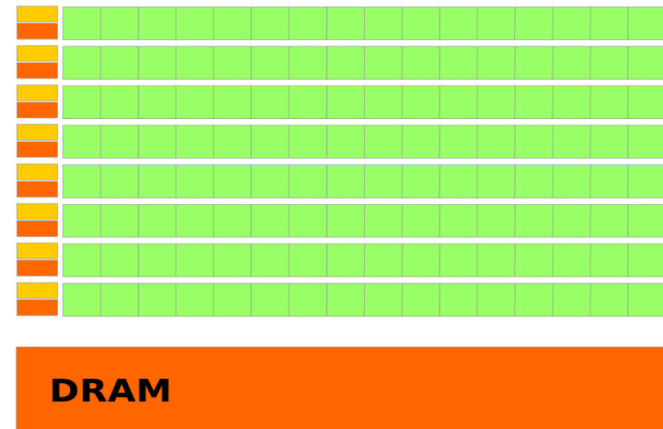
- SoC are in contrast to the motherboard-based PC architecture;
- SoC integrates CPU/GPU/memory interfaces into a single chip;
- SoC has reduced modularity and replaceability of components;
- energy-efficiency is the main concern;
- ARM is the *de facto* SoC technology.



CPU vs GPU



CPU



GPU

- CPU is latency-optimized (each thread runs as fast as possible, but only few threads);
- CPU has few cores (≤ 16);
- CPU excels at irregular control-intensive work (lots of hardware of control, few ALUs);
- Programming languages: C/C++, Fortran, Python, IDL, ...
- Parallel libraries/directives: MPI/OpenMP.

- GPU has highly data-parallel fixed architecture (SIMD);
- GPU is throughput-optimized (thousands of threads, hundreds of cores);
- GPU excels at regular math-intensive work (lots of ALUs for math, little hardware control);
- Very high memory bandwidth (drawback for power consumption);
- Parallel programming: OpenACC (directives), CUDA, OpenCL (low level programming for high performance).

A high-performance problem solved in parallel



Two types of parallelism

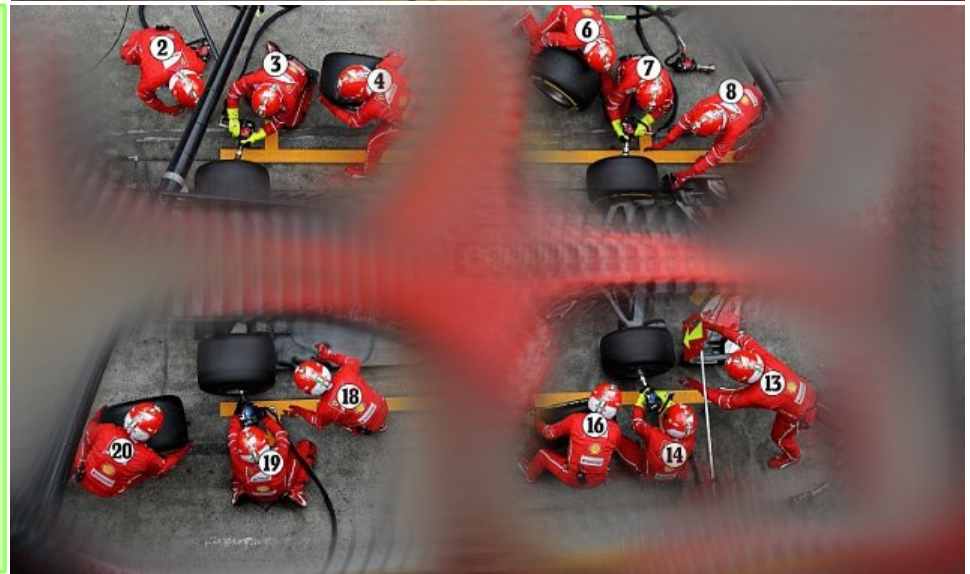
- Task parallelism: different people are performing different tasks at the same time.

Work suitable for CPU!

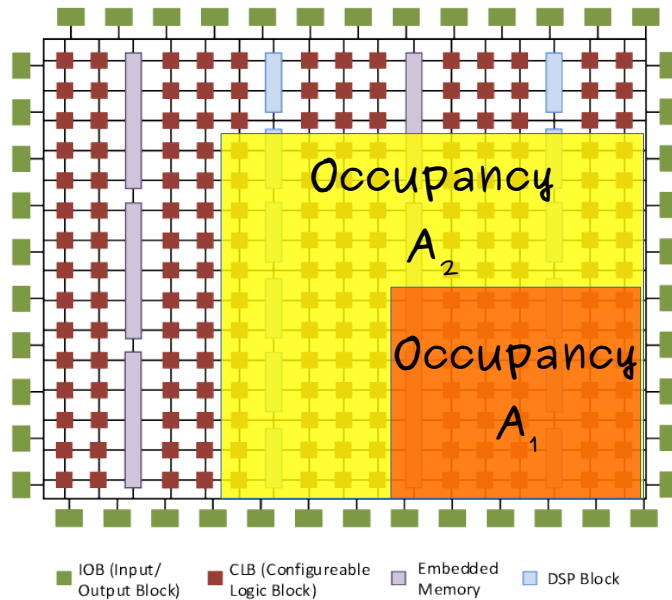


- Data parallelism: different people are performing the same task at the same time, but on different equivalent and independent data.

Work suitable for GPU!

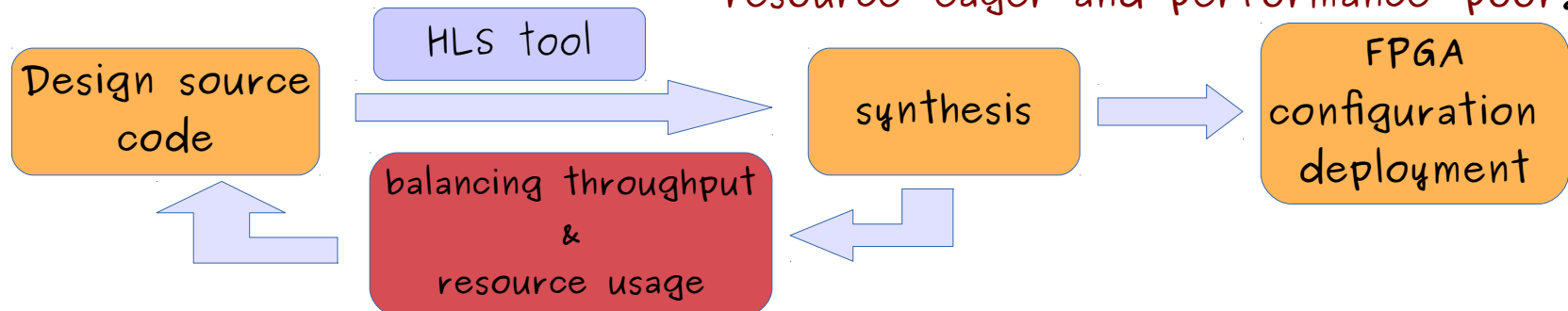


What is a Field Programmable Gate Array (FPGA)?



FPGA is a semiconductor device that can be programmed (i.e. no fixed architecture):

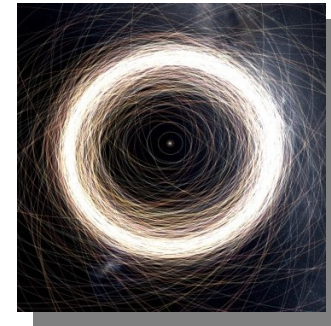
- > desired functionality of the FPGA can be (re)-programmed by downloading a configuration into the device;
- > flexible interconnect, highly parallel customizable architecture (both data-parallelism and task-parallelism);
- > **optimal power efficiency** (3-4x than of GPU);
- > **low level programming required for high performance;**
- > **currently double-precision arithmetic is resource-eager and performance-poor.**



The INAF astrophysical codes in ExaNest

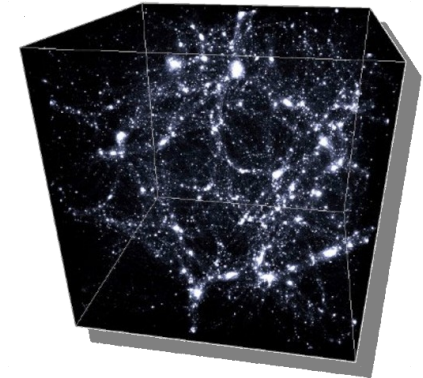
Hy-Nbody (D. Goz et al. 2019):

direct N -Body code to simulate cluster dynamics and close encounters.



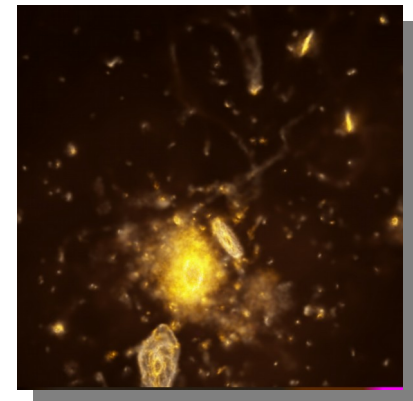
PINOCCHIO (P. Monaco, T. Theuns & G. Taffoni, 2002):

a fast code, based on Lagrangian perturbation theory, to generate catalogues of cosmological dark matter halos and their merger history.


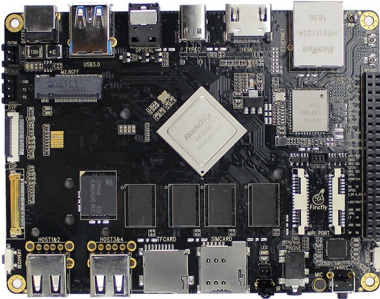
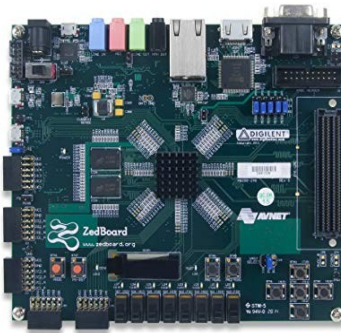



GADGET (V. Springel 2005):

is an N -body and hydrodynamical code for large-scale, high-resolution numerical simulations of cosmic structure formation and evolution.



Computing Platforms

Platform	Desktop	Firefly-RK3399	ZedBoard	QFDB (ExaNeSt)
	ASUS P8B75-M LX	Rockchip RK3399	Xilinx Zynq-7000 MPSoC	Xilinx Zynq- UltraScale+ MPSoC
				
CPU	Intel <u>i7-3770x4</u>	ARM <u>A72x2+A53x4</u>	ARM A9x2	ARM (A53x4)x4
GPU	Nvidia GeForce <u>GTX-1080</u>	ARM <u>Mali-T864</u>	None	None
FPGA	None	None	<u>Zynq-7000</u>	<u>(Zynq-US+)x4</u>
RAM	16GB DDR3	4GB DDR3	512MB DDR3	16x4GB DDR4

INCAS (INTensive Clustered Arm-SoC)

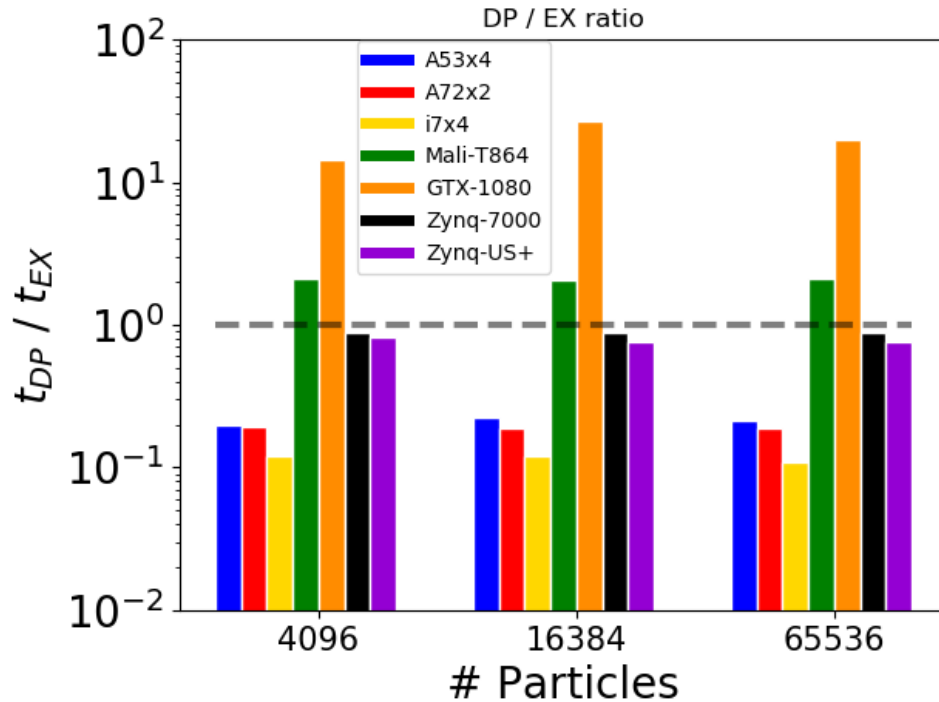
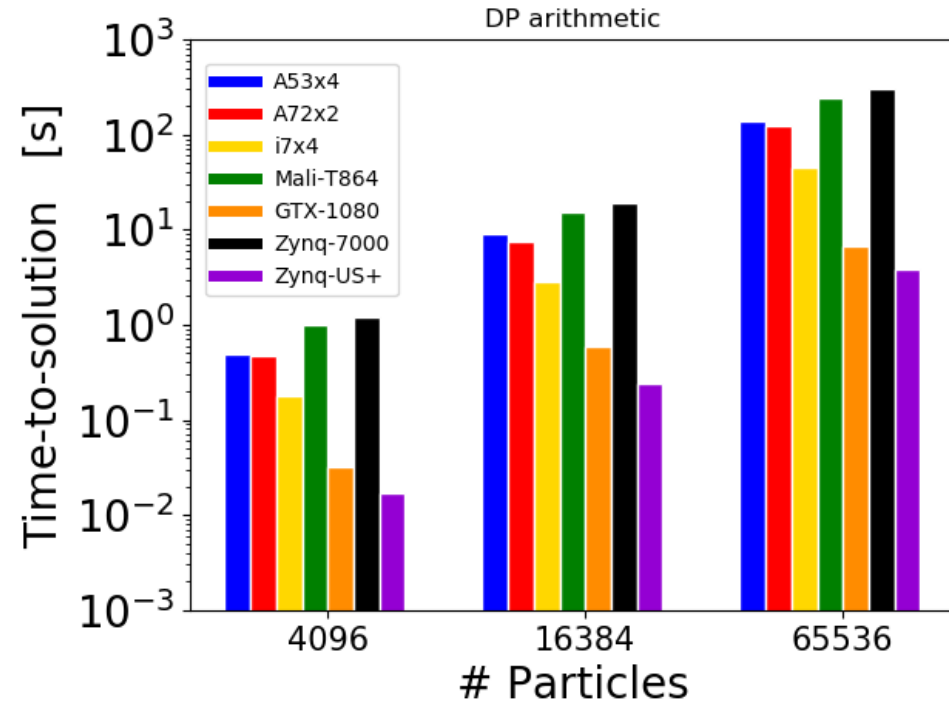
(INAF technical report: S. Bertocco DOI: 10.20371/INAF/PUB/2018_00004)



Cluster components.

Nodes available	8
SoC	Rockchip – RK3399
CPU/node	Six-Core ARM 64-bit (Dual-Core Cortex-A72 and Quad-Core Cortex-A53)
GPU/node	ARM Mali-T864 MP4 Quad-Core
Ram memory/node	4GB dual-channel DDR3
Network	1 Gbps Ethernet
OS	Ubuntu 16.04 LTS
Compiler	gcc version 7.3.0
MPI	OpenMPI version 3.0.1
OpenCL	OpenCL version 2.2
Job scheduler	SLURM version 17.11

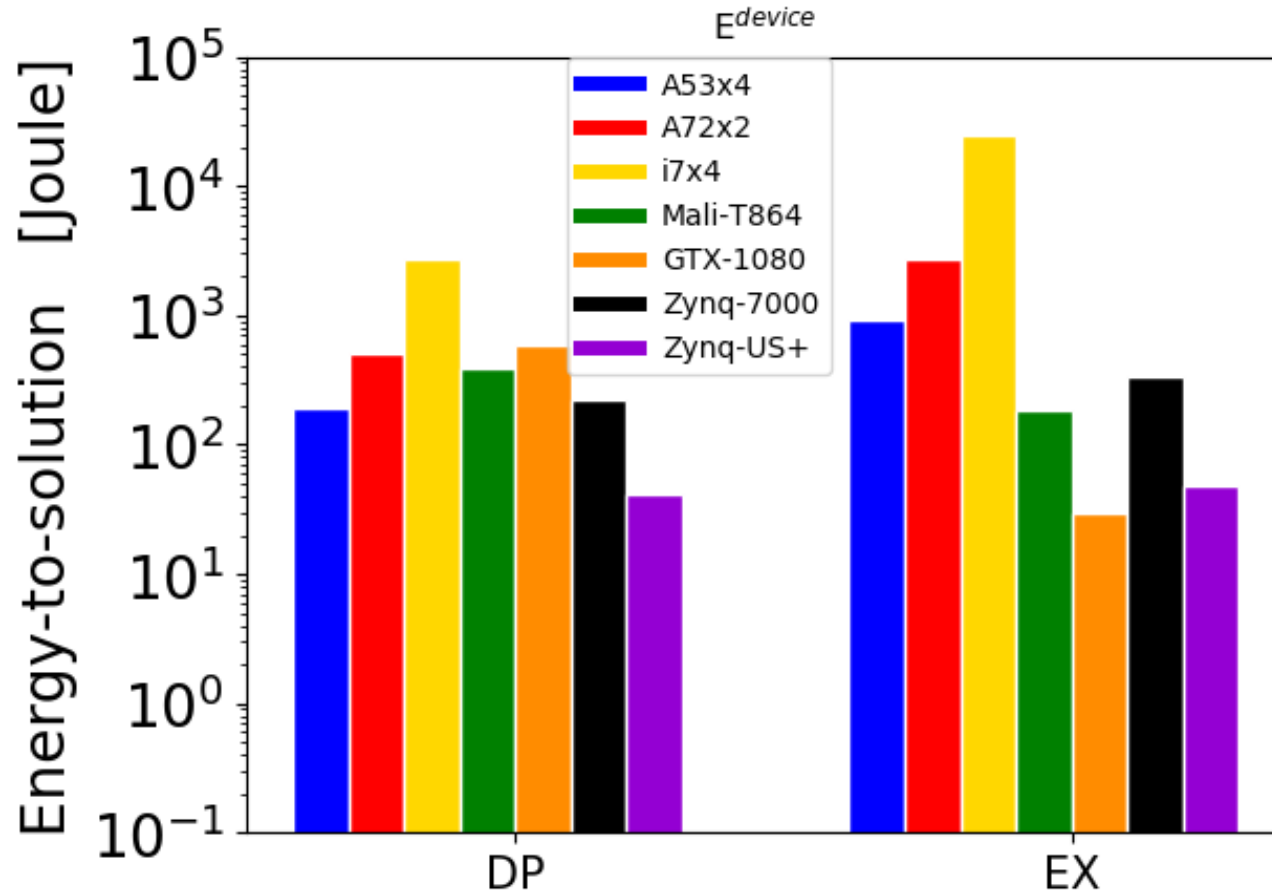
Computational performances [Hy-Mbody]



- GTX and UltraScale+ outperform others;
- UltraScale+ $\approx 100 \text{ GFLOPS FP64}$.

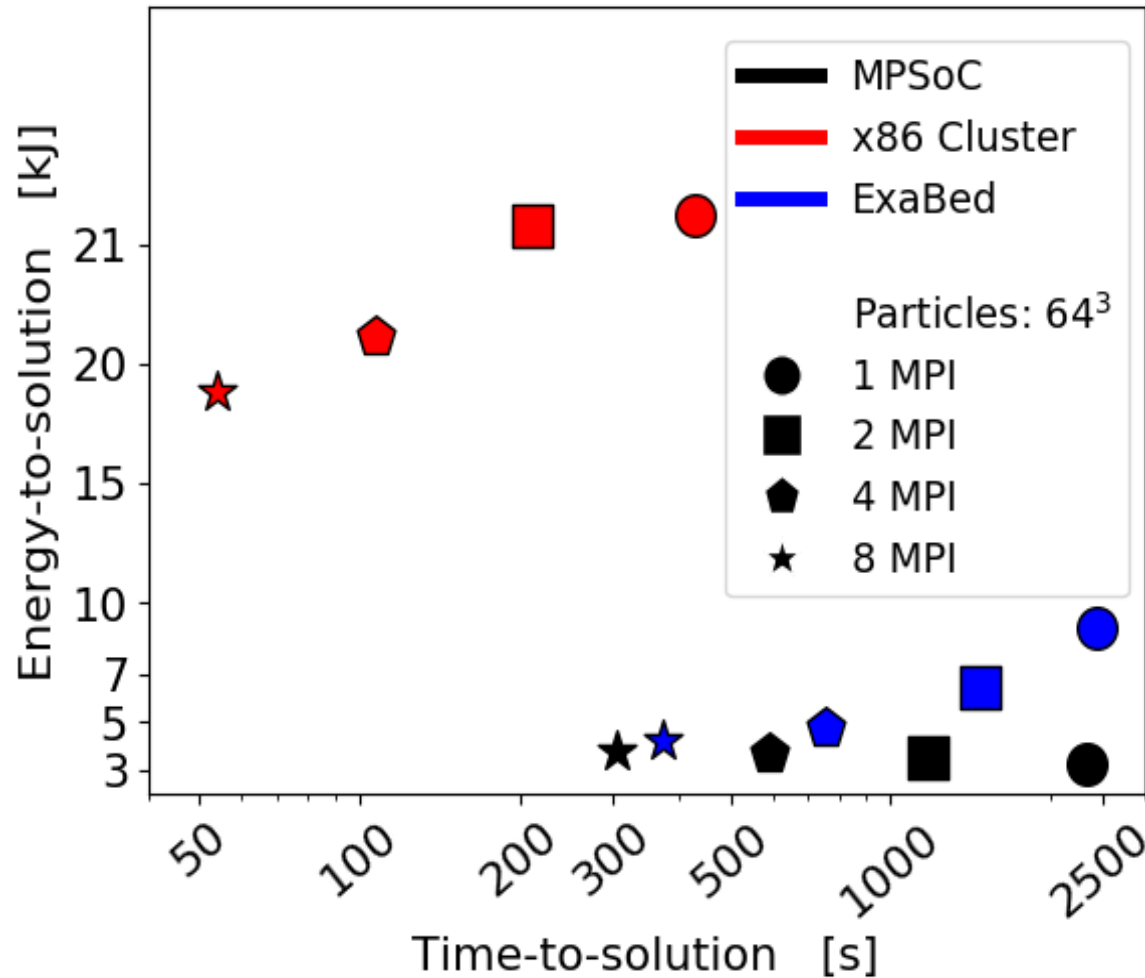
- Only GPUs benefit from EX-arithmetic;
- for GTX $t_{DP} / t_{EX} \approx 20$;
- for Mali $t_{DP} / t_{EX} \approx 2$.

Energy consumption [Hy-Nbody]



$$E_{device} = E_{device}^{baseline} - E_{baseline}$$

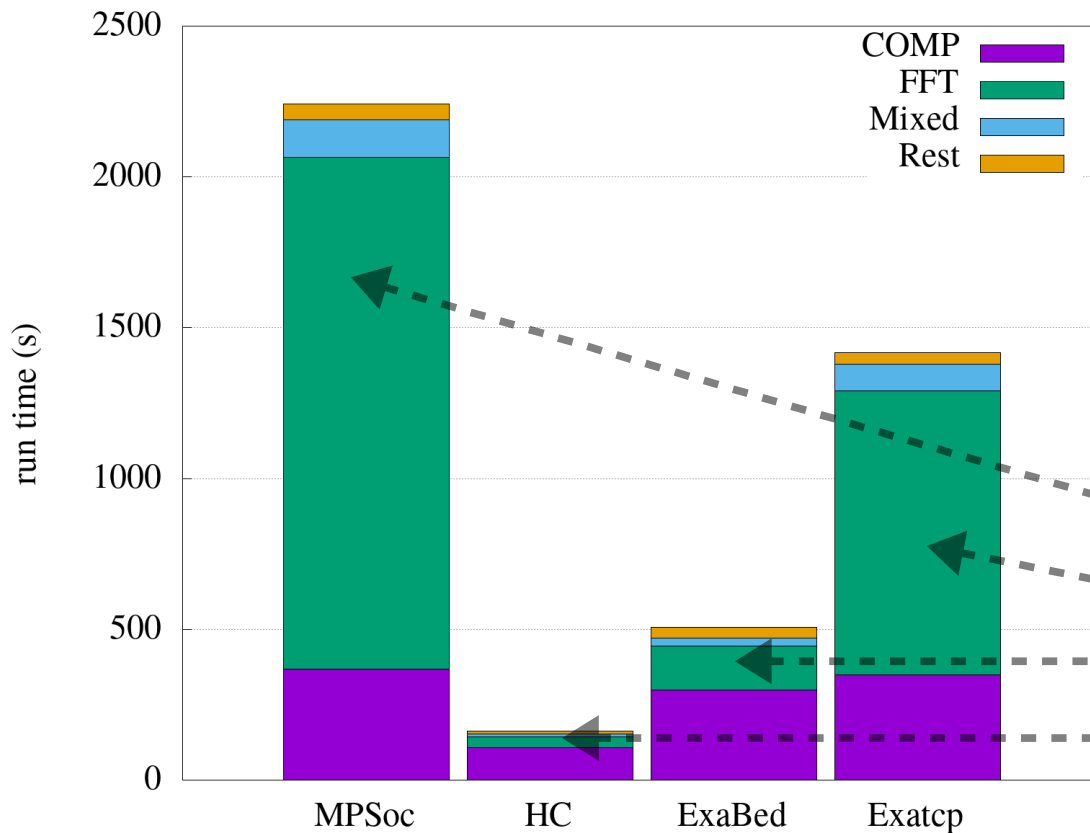
Multi-node performances [Hy-Mbody]



- MPSoC == INCAS;
- x86 Cluster == HOTCAT (CHIPP-INAF cluster);
- ExaBed == ExaNest.

- ARM A53x4/MPI;
- INTEL Xeon E5x4/MPI;
- ARM A53x4/MPI.

Multi-node performances [PINOCCHIO]



- **Violet**: pure comp. fraction of the code;
 - **Green**: FFT-related fraction of the code;
 - **Blue**: comp. + comm.
- INCAS with 1Gb ethernet;
ExaBed with 10 Gb ethernet;
ExaBed with Exanet;
HOTCAT with Infiniband.

Conclusions

- The usage of heterogeneous computing in scientific research (not only HPC) appears to be inevitable;
- SoC technology is emerging as a promising alternative to “traditional” technologies for HPC;
- we will be forced to re-engineer our applications in order to exploit new exascale computing facilities (different devices, complex memory hierarchies);
- we will be forced to devise high performance-per-watt algorithms.

Future developments

- To assess the energy footprint of our applications using the whole ExaNest prototype and compare with HPC resources.

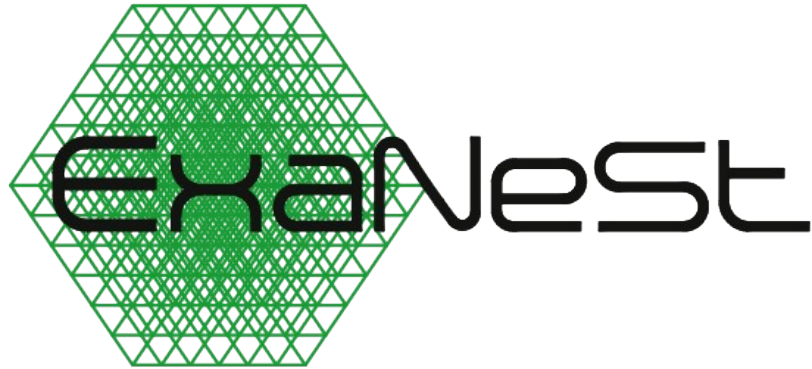
References

INAF Technical Reports:

- Goz D., Tornatore L., Bertocco S. and Taffoni G. DOI: 10.20371/INAF/PUB/2018_00002;
- Goz D., Tornatore L., Bertocco S. and Taffoni G. DOI: 10.20371/INAF/PUB/2018_00005;
- Goz D., Tornatore L., Bertocco S. and Taffoni G. DOI: 10.20371/INAF/PUB/2018_00006;
- Bertocco S., Goz D., Tornatore L. and Taffoni G. DOI: 10.20371/INAF/PUB/2018_00004.

Papers:

- *Direct N-body code on low-power embedded ARM GPUs*, Goz D., Bertocco S., Tornatore L., Taffoni G., Computing Conference 2019 proceedings;
- *Low power high performance computing on Arm system-on-chip in Astrophysics*, Taffoni G., Bertocco S., Coretti I., Goz D., Ragagnin A., Tornatore L., Springer series in "Advances in Intelligent Systems and Computing" 2019.



Horizon 2020

This work was carried out within the ExaNest (FET-HPC) project (grant no. 671553) funded by the European Unions Horizon 2020 research and innovation programme.

