H2020

LEIT-ICT-4 2015

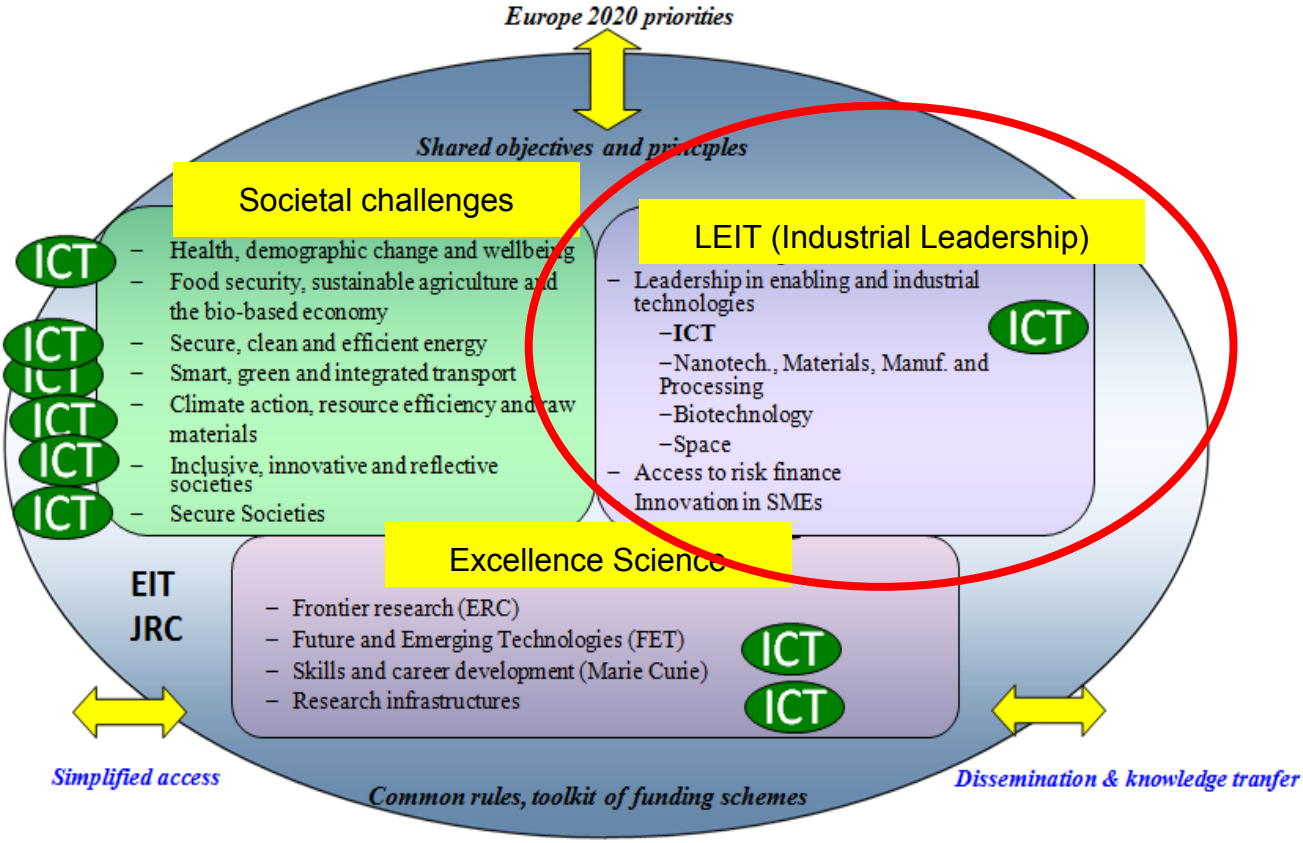
Customised and low power computing

An INFN opportunity for new HPC challenges

A.Ferraro, D.Salomoni, C.Vistoli, D.Cesini

INFN-CNAF

H2020 Pillars



LEIT (Leadership in enabling and industrial technologies)

LEIT facts:

- Focused on new opportunities for industrial leadership in ICT, Key Enabling Technologies (nanotechnology, materials, etc.) and space
- Emphasis on areas of R&D and innovation where mastering new technological opportunities
- The involvement of industrial participants, and of SMEs in particular, is crucial in maximising the expected impact of the actions.
- The funded projects will be outcome oriented, developing key technology building blocks, bringing solutions closer to the market, and paving the way for industrial and commercial implementation

R&D

TRASFERIMENTO TECNOLOGICO

LEIT-ICT topics:

- New generation of components and systems
- Advanced computing:
  - Customized and low-power computing - ICT4
- Future internet
- Content technologies and information management
- Robotics, Micro- and nano-electronic technologies, Photonics, ICT for manufacturing and Factory of the Future

LEIT -> ICT -> Advanced Computing -> ICT-4

(Customised and low power computing)

ICT4.a	ICT4.b	ICT4.c
Research&Innovation Actions	**Stimulate broad adoption Actions**	**Support Actions**
37M€	17M€	3M€
One or ~~both the following themes:~~ • Integration of HW&SW components into fully working prototypes of low-power micro-servers. Specific emphasis is given on low-power, low-cost, high-density, secure, reliable, scalable small form-factor datacentres ("datacentre-in-a-box") **[5-8M€]** • New programming approaches to exploit the full potential of the next generations of computing systems based on heterogeneous parallel architectures (beyond performance, optimisation should include energy efficiency). Proposals requesting a small contribution are expected. **[2-4M€]**	One or both the following themes: • Establish reference architectures and platforms for customised low-power heterogeneous computing systems delivering high performance functionality **[2-4M€]** • Bring together all actors along the value chain to customise and use advanced low power computing systems in cyber-physical systems. Clustered in large scale projects driven by networks of European centres of excellence to achieve critical mass and to better exploit EU-added value. **[5-8M€]**	Support actions for cross-sectorial platform-building, for clustering of related research projects, for structuring the European academic and industrial research communities, for dissemination of programme achievements and impact analysis **[2-4M €]**

Expected impact:

- Strengthening the technology competences of European suppliers and the academic community.
- Reduction of energy consumption of servers by 2 orders of magnitude as compared to state of the art in 2013.
- Double the productivity in efficiently programming and maintaining advanced computing systems powering cyber-physical systems as compared to state of the art in programming embedded systems in 2013.
- Increase the adoption of form-factor data-centres and heterogeneous highly parallel computing systems.
- Higher involvement of SMEs, both on the supply and the demand-side.
- Increased adoption of concurrency in applications across all sectors; higher degree of parallelism in applications

DRAFT

ICT4 - Customised and low power computing

Yes, but **what's new?**

Embedded systems have evolved into cyber-physical systems (connected)

Wide availability of multi-core and the end of Moore's law

Optimisation is now multi-dimensional: energy, performance, response time, security... (more non-functional requirements)

HPC remains a different market, but now "embedded HPC" is emerging

Programmers don't have to be specialists in parallel algorithms

We do NOT want:

Proposals with no industrial or innovation aspects.

Proposals without prototypes, applications or experiments.

Who are the leading players?

System integrators (Bull, Eurotech, Ericsson, IBM, Thales)

Hardware (ST-Microelectronics, ARM, Infineon, Intel)

Research (Barcelona Supercomputing Center, FORTH, Fraunhofer, INRIA, POLIMI, UNI-Stuttgart, TU-Wien)

Industrial users (Airbus, Comau, EDF, Philips)

... and many others (including many innovative SMEs)

Call scheduling

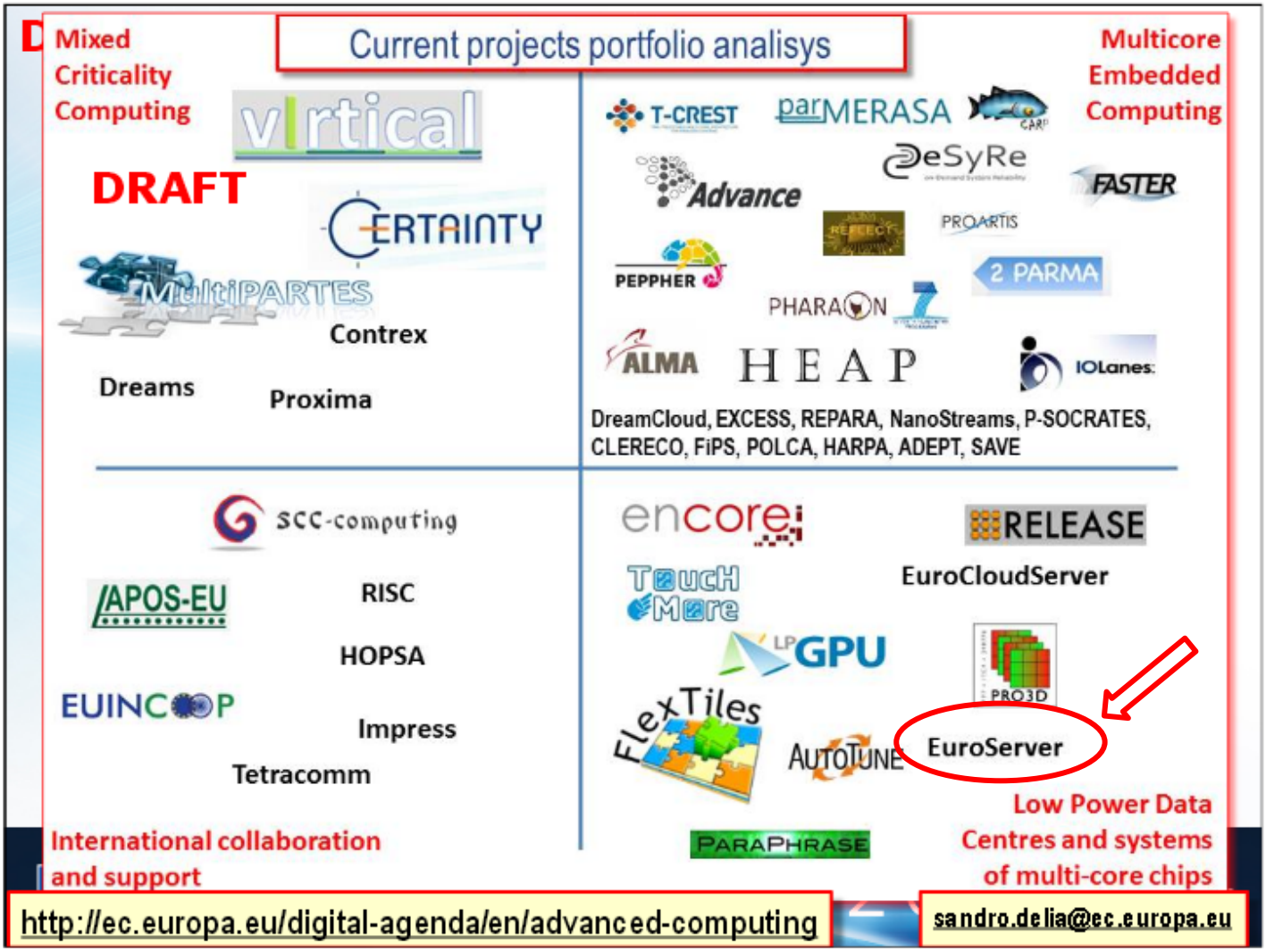
Call	H2020-ICT-2015
Publication date	15/10/2014
Deadline	21/4/2015

Deadline: 2015 !!!

Time to: improve 32bit low-power computing competencies
ready for the 64bit low-power era (2014/2015)
create INFN low-power HPC consciousness
include other INFN interested groups
make agreements with SMEs
attend European events (Open days, etc.)
acquire results of already completed projects/initiatives
define a well defined proposal

Il Servizio Fondi Esterni evidenzia che tra le call di H2020 con potenziale rilevanza per la partecipazione dell'INFN c'è anche LEIT-ICT-4-2015

B	C	D	E	F	G
TOPIC	sub-topic	INFN best-suited	INFN can be suited	INFN suited ?	NOTE
LEIT					
ICT					
	Advanced Computing				look@ draft: work complementary to work in other group (FET,ECSEL,etc..)
	ICT4-2015		**High performing low-power, low-cost; platform building**		R&I, innovation, support activities


Current projects portfolio analisys
Mixed Criticality Computing
Multicore Embedded Computing
DRAFT
vIrtical
CERTAINTY
MultiPARTES
Contrex
Dreams
Proxima
T-CREST
parMERASA
CAP
eSyRe
Advance
FASTER
REFLECT
PROARTIS
PEPPHER
2 PARMA
PHARAON
ALMA
HEAP
IOLanes
DreamCloud, EXCESS, REPARA, NanoStreams, P-SOCRATES, CLERECO, FiPS, POLCA, HARPA, ADEPT, SAVE
SCC-computing
encore
RELEASE
TOUCH MORE
EuroCloudServer
APOS-EU
RISC
HOPSA
LPGPU
PRO3D
EUINCOOP
Impress
FlexTiles
AUTOTUNE
EuroServer
Tetracomm
PARAPHRASE
International collaboration and support
Low Power Data Centres and systems of multi-core chips
http://ec.europa.eu/digital-agenda/en/advanced-computing
sandro.delia@ec.europa.eu

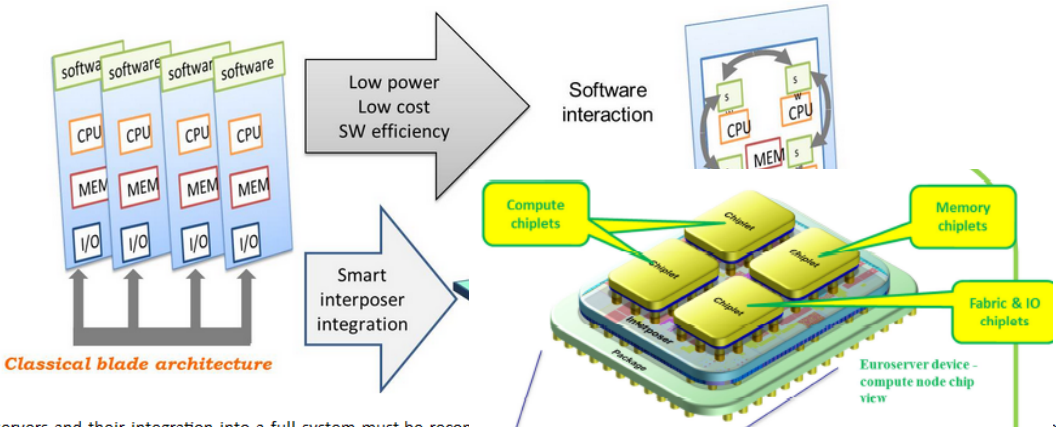
CURRENT EUROPEAN INITIATIVES: Euroserver Project

Green Computing Node for European Micro-servers

SEVENTH FRAMEWORK PROGRAMME

Home | Partners | Innovations | Applications | Contacts

Data Centres (DCs) are a key resource for innovation and leadership of industry in Europe. They drive the Information Society through hosted cloud applications. To sustain the ever-increasing demand of storing and processing data, DCs need to improve their capabilities and scale in size. With current server technology, however, DC scaling is limited by the IT equipment's density and energy consumption. To keep up with data growth, in the face of power distribution constraints and steadily increasing energy costs, the IT equipment must become dramatically smaller and more efficient in power and energy. This moves the focus of server design and the interests of industry from performance to power/energy efficiency and total cost of ownership (TCO).

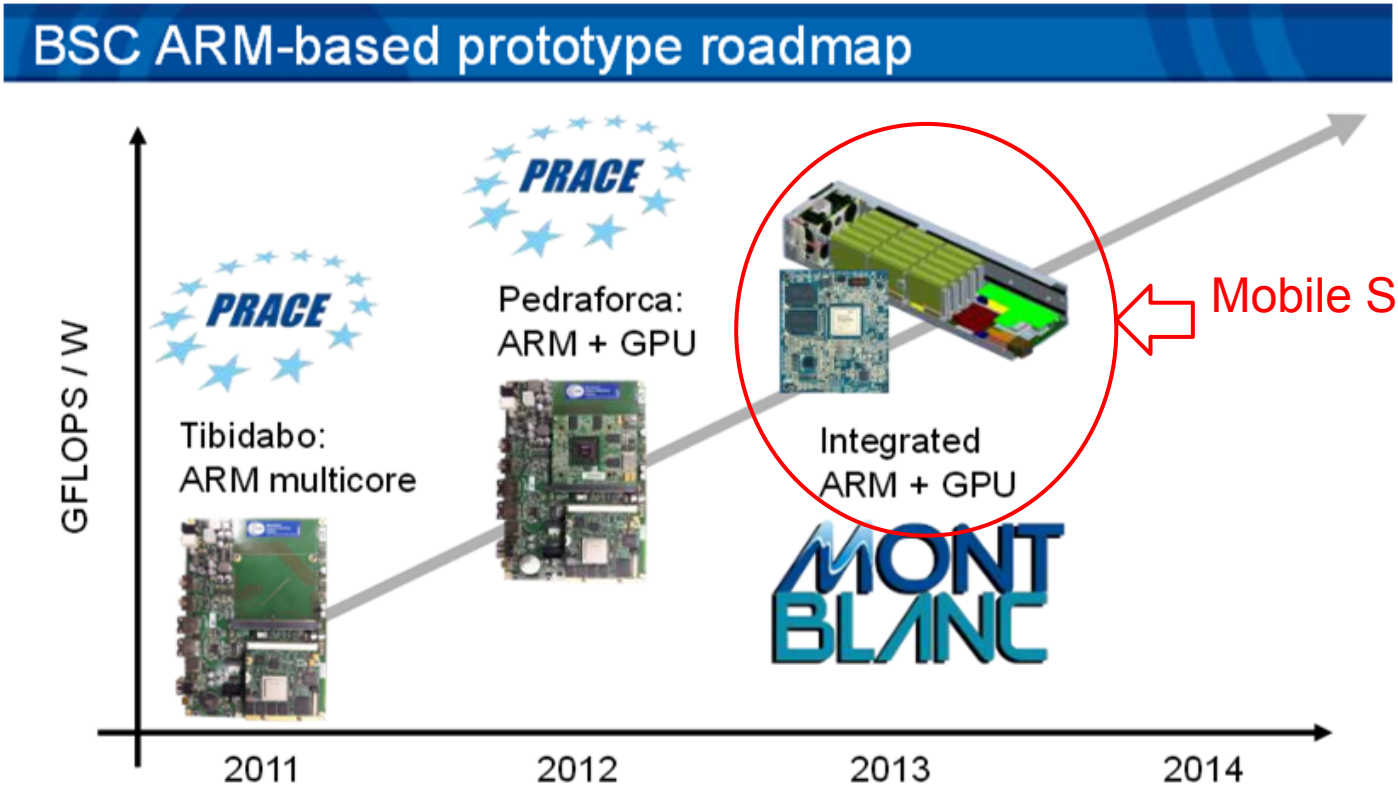


The basic components of future servers and their integration into a full system must be reconsidered from the ground up. The processor, the memory hierarchy, I/O, the system interconnects and the systems' software all require fundamental changes to match the application's performance in less space and with drastically lower energy costs.

EUROSERVER is addressing these challenges in a holistic manner; we advocate the use of state-of-the-art low-power ARM processors in a new server system-architecture that uses 3D integration to scale with both the numbers of cores, and the memory and I/O, all managed by new systems software providing transparent system-wide virtualization and efficient resource use by cloud applications. The EUROSERVER prototype will demonstrate how the proposed approach can lead to 10x DC Energy Efficiency by 2020.

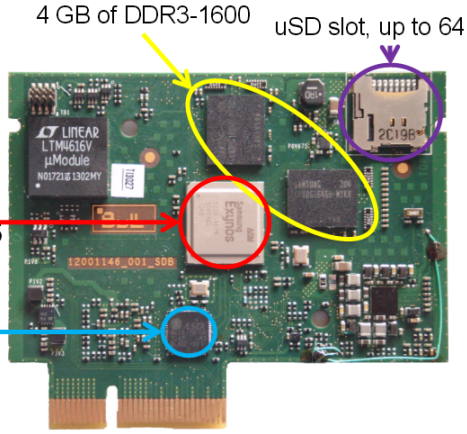
ARM · Barcelona Supercomputing Center · cea · CHALMERS · EUROTECH Imagine. Build. Succeed. · FORTH · onapp · life.augmented · TECHNISCHE UNIVERSITÄT DRESDEN

CURRENT EUROPEAN INITIATIVES: MontBlanc Project



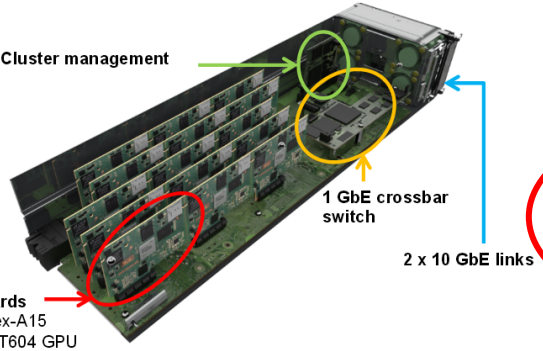
- Prototypes are critical to accelerate software development
  - System software stack + applications

CURRENT EUROPEAN INITIATIVES: MontBlanc Project



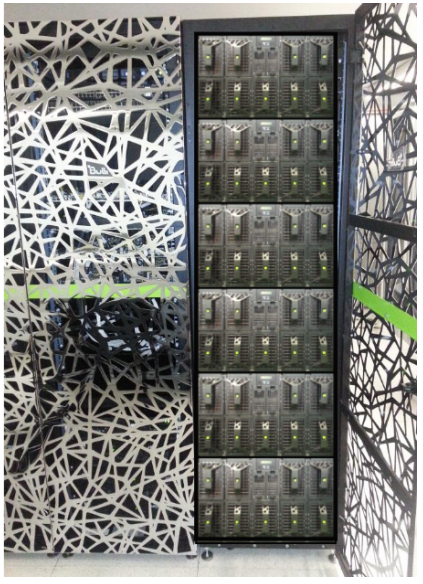
6.8GF CPUs
25.5GF GPU
Double precision

- 15 node-cluster in a standard Bull B505 enclosure



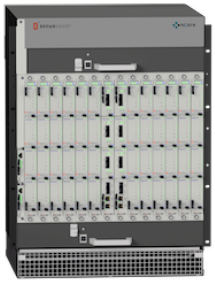
- 6 BullX chassis
- 54 Compute blades
- 810 Compute cards
  - 1620 CPU
  - 810 GPU
  - 3.2 TB of DRAM
  - 52 TB of Flash

- 26 TFLOPS
- 18 kWatt



Cluster commerciali esistenti basati su SoC

BrownDwarf Y-Class Supercomputer



The nCore
unprecedent
with a low la
programmin

nCore

This combination meets users' needs for eco-friendly energy consumption, powerful computational ability and mitigates technical risk by preserving existing investment in developed applications.

Designed to fulfill the demands of high performan applications in wide ranging fields, the Y-class sys allows scientists and engineers to gain insight into complex problems while delivering real-world results.

Texas Instruments

RapidIO
The Embedded Fabric Choice

BrownDwarf Y-Class Brochure (PDF) | Y-Class Power Telemetry Brochure (PDF)
Contact nCore

Features

- Advanced parallel computing architecture
- High Performance, Low Latency Interconnect
- Unified Programming Model using C/C++

HP Moonshot System

SoC Texas Instruments (ARM), AMD (ARM), Intel (X86)

BOSTON
Servers · Storage · Solutions

BUILT-IN ENE

Advanced system, power and fabric management software for energy-proportional computing, hosted on dual-core ARM® Cortex A15 processor. The Calxeda Fleet Engines™ replace separate BMC chips and FPGAs, along with dedicated Ethernet MAC for secure management.

LOWER COST, LOWER POWER & HIGHER SCALABILITY

Quad-core ARM Cortex processors integrated with high bandwidth memory controllers and 4MB L2 Cache. ARM Cortex A15 (up to 1.8 GHz) processors deliver:

- 2x performance improvements
- 3x memory bandwidth, and
- 4x memory capacity

Advanced system, power and fabric management software for energy-proportional computing minimises the number of Ethernet PHYs – saving power, cost, cabling, and top-of-rack Ethernet switch Ports requirements.

It is highly scalable thanks to the integrated 80GB Fleet Fabric™ switch, while the embedded Fleet Engine™ simultaneously provides out-of-band control and intelligence for autonomic operation and power optimization.

IDEAL FOR COMPUTE CLOUDS, DISTRIBUTED STORAGE, & CONTENT DELIVERY

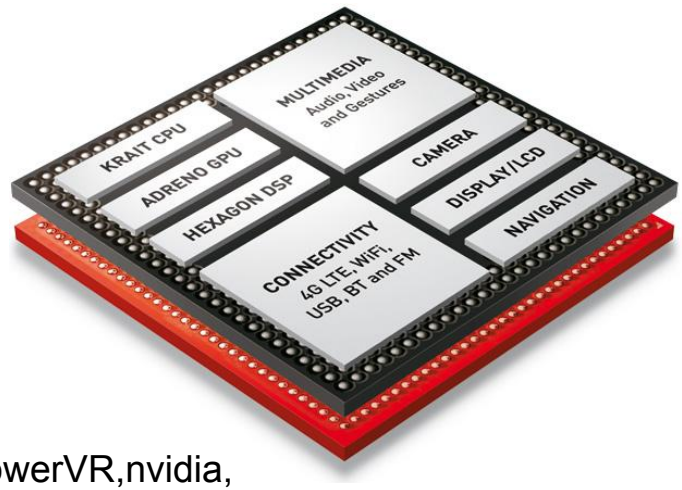
Tra progetti di ricerca e soluzioni commerciali il mondo HPC low power è variegato e complesso

SoC

1. ARMv7/v8 based: Nvidia Tegra, Qualcomm Snapdragon, Samsung Exynos, Freescale imx6, TI OMAP/Keystone II, Mediatek, Allwinner
2. Intel Atom based (Avoton, Broadwell, etc..)
3. HSA based (AMD Kaveri)

Interfacce I/O

1. PCIe, MIPI, M-PCIe
2. Infiniband , RapidIO , 1/10 Gbit
3. NVLINK?



Acceleratori

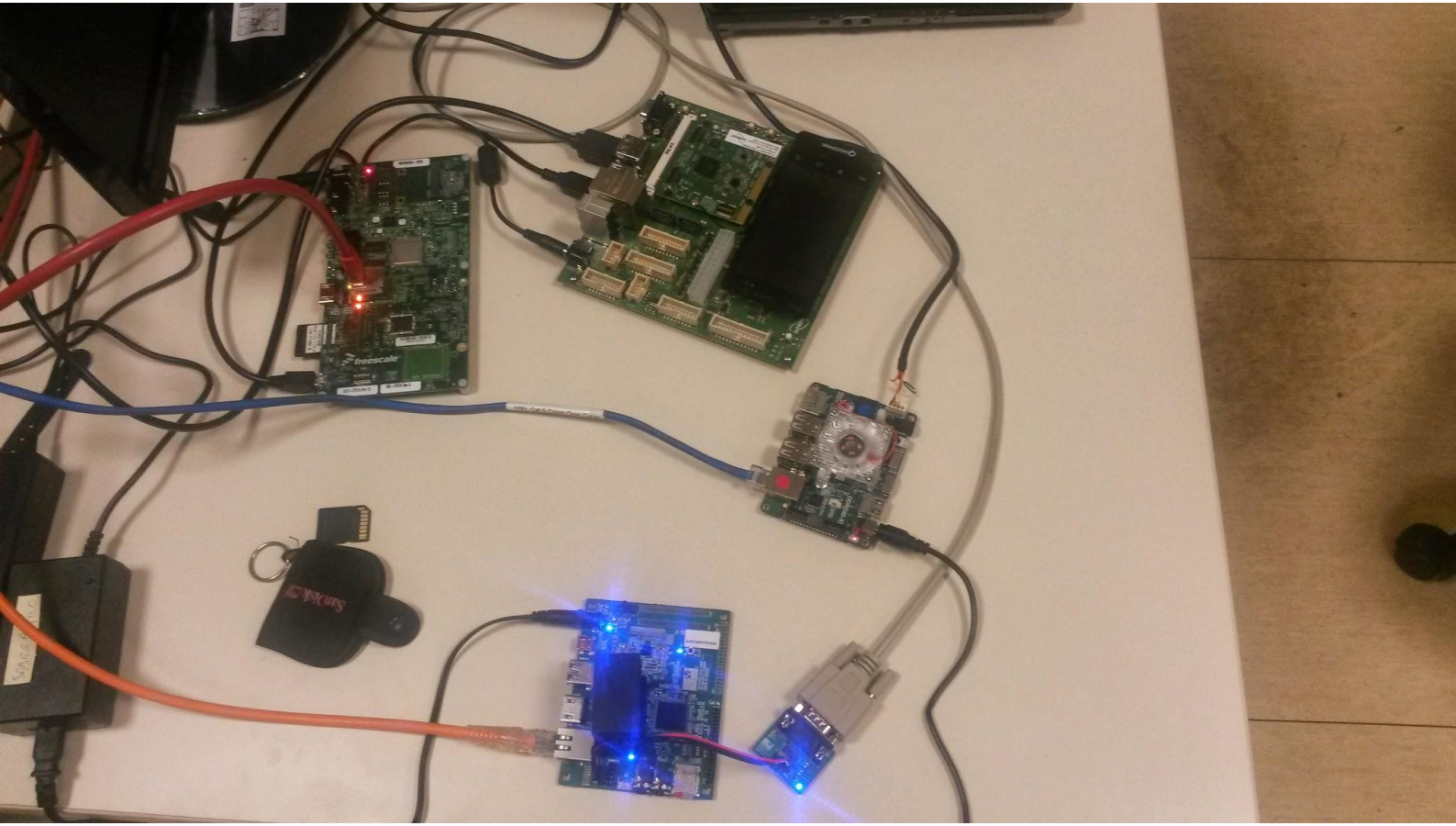
1. GPU (ARM, Imagination Tech., PowerVR,nvidia, Qualcom, etc..)
2. DSP
3. FPGA

Testing at CNAF: development Board

BOARD	SOC				GFLOPS (CPU+GPU)	Eth
	Model	ARM IP	GPU IP	DSP IP		
FREESCALE (Embedded SoC) SABRE Board	**Freescale** i.MX6Q	**ARM** A9(4)	**Vivante** GC2100 (19.2GFlops)		25	1Gb
ARNDALE (Mobile SoC) Octa Board	**Samsung** Exynos 5420	**ARM** A15(4) A7(4)	**ARM** Mali-T628 MP6 (129GFlops)		135	10/100
HARDKERNEL (Mobile SoC) Odroid-XU-E	**Samsung** Exynos 5410	**ARM** A15(4) A7(4)	**Imagination Technologies** PowerVR SGX544MP3 (51.1 GFlops)		65	10/100
INTRINSIC (Mobile SoC) DragonBoard	**Qualcomm** Snapdragon 800	**Qualcomm** Krait(4)	**Qualcomm** Adreno 330 (130GFlops)		145	1Gb
TI (Embedded SoC) EVMK2H	**TI Keystone** 66AK2H14	**ARM** A15(2)		**TI** MS320C66x (189Gflops)	210	1Gb (10Gb)

TDP tra 5W e 10W

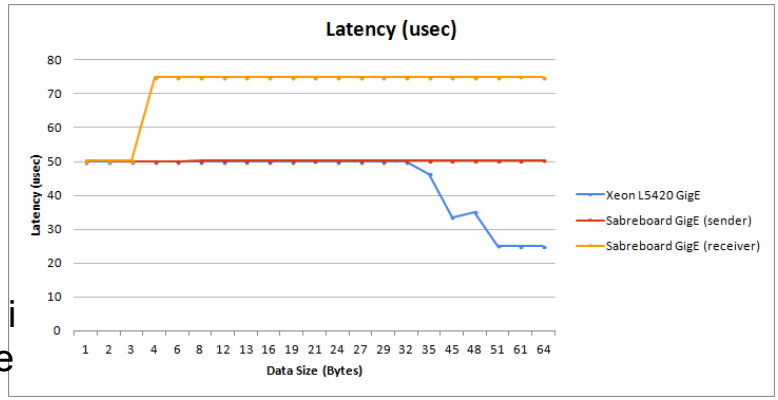
“HPC Cluster” at CNAF

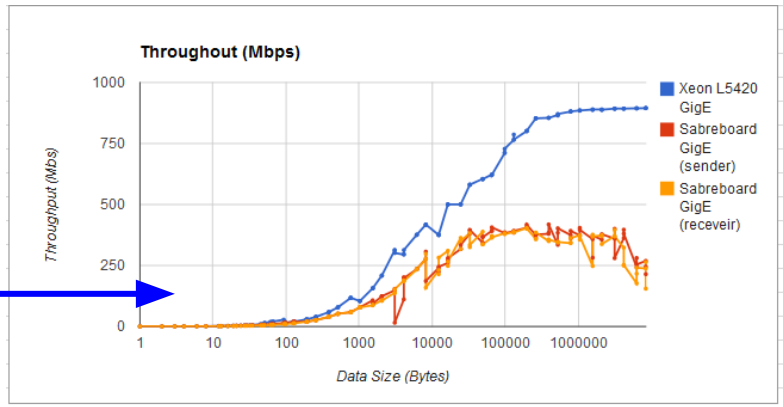


Criticità già emerse nei SoC ARM

- Indirizzamento 32 bit per processo
- ECC quasi sempre assente
- Supporto Linux limitato
- Supporto OpenCL limitato/assente su Linux
- CUDA assente
- RAM GPU limitata
- 8 core solo teorici (bigLittle)
- Assenza di interfacce I/O standard per i server (eth, PCIe) nei SoC ARM mobile
  - presenti nei SoC embedded
- Affidabilità 24x7 da verificare
  - Power management
  - Cooling
- Molte librerie solo per x86
  - binari non disponibili o immaturi
- Bachi HW

ETH SABREBOARD





Possibili miglioramenti con ARMv8 64bit, M-PCIe/NVLINK, NVIDIA K1, etc

COSA: COmputing on SoC Architectures

Obiettivi:

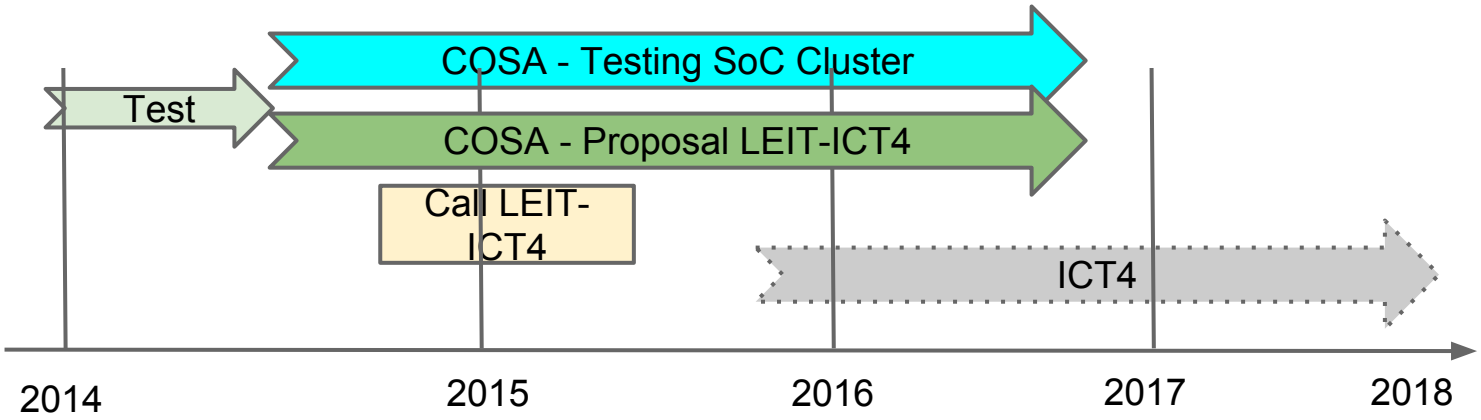
1) Acquisizione esperienza HPC su architetture SoC

- Attività focalizzata sull'uso della GPU nel SoC (finora focus su CPU ARM)
- Testing dei nuovi SoC su evaluation board - technology tracking
- Testing dei cluster esistenti basati su SoC
  - collaborazione con partner commerciali e progetti europei
- Creazione di un cluster con development board per SoC
  - Porting di applicazioni di interesse per l'INFN
  - Paradigmi di programmazione (MPI, OPENMP, CUDA, OpenCl, OpenACC)

2) Creazione proposal e consorzio per LEIT-ICT4 (Customised and low power computing)

- realizzazione di un server "HPC" portatile con partner commerciali (datacenter in a box)

Possibili sedi coinvolte: **ROMA1, CNAF, PR, PI, PD, FE**



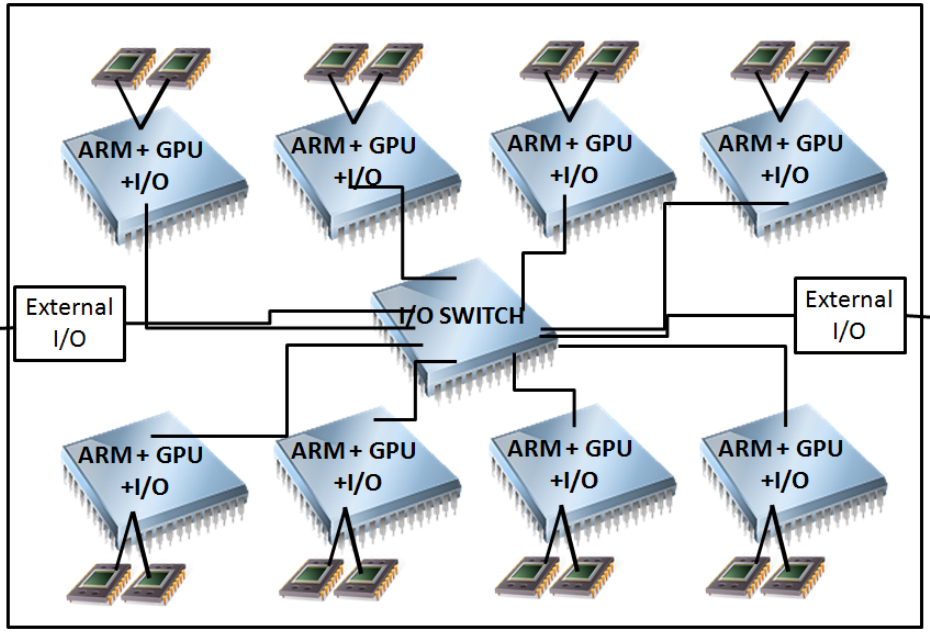
COSA: Applicazioni

- Applicazioni non parallele
  - job degli esperimenti LHC
  - esperienza consolidata nel porting di queste applicazioni su processori ARM (cfr. Boccali at CCR workshop a Genova)
- Applicazioni parallele su sistemi ibridi “small HPC”
  - ad esempio quelle in fase di studio per i trigger degli esperimenti LHC che utilizzano sia CPU che GPU (cfr. Silvia Amerio al ws di Trieste: HLT for HL-LHC, Technology and architecture for next decade TDAQ)
- Debugging e test di applicazioni high-end su architetture ibride
  - ad es. applicazioni Parma e Ferrara con OpenCL e OpenACC
- Applicazioni basso parallelismo che necessitano di una sistema portatile e a basso consumo
  - attività che si svolgono fuori dal laboratorio e spesso in condizioni disagiate
  - ad esempio il post processing delle tomografie assiali computerizzate di reperti archeologici e museali del gruppo infn-bologna/unibo (Brancaccio et al.)
- Simulazioni di reti neurali - Roma1 (Vicini)

COSA potrebbe partire dall’esperienza di COKA sulle applicazioni INFN

SoC Architecture

Knights Architecture



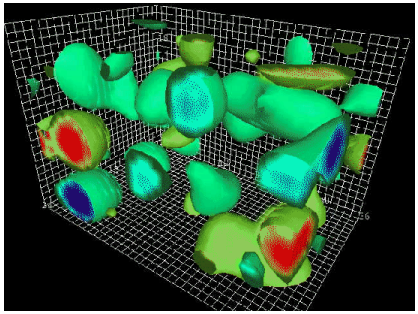
Mobile or embedded SoC based board
"MontBlanc like"

+

Embedded Portable Low Power Server

+

HW/SW Integration with Scientific Applications



Key contribution 4/5:
Demonstration on an Embedded Server

EURO SERVER

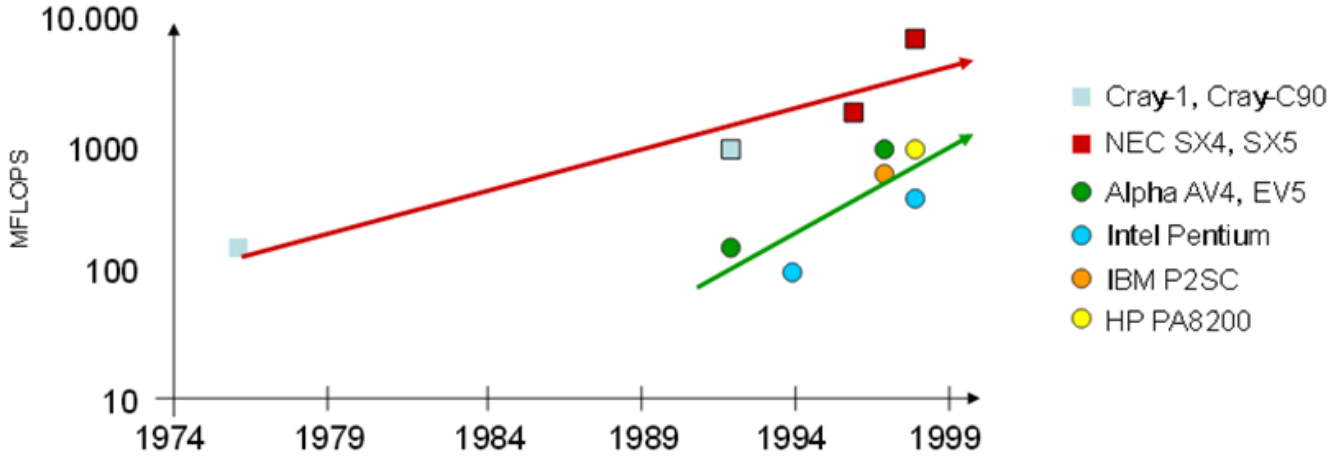
Tentative specs

- Target form factor: Eurotech Mounted Mobile Computer
- Size: 13 x 25 x 8 cm
- Prototype targeted for rugged sealed enclosure and extended temperature range
- Modular system design
  - Shared design for micro-server board and carrier board with Enterprise Server version
  - Up to 2 micro-server slots
  - Support for optional general purpose card (using 1 slot) for specific function or peripheral
- 9 to 36 Vdc in; 30W max;
- Passive conduction cooling
- Support for Backup battery pack to investigate

BACKUP

HPC trends (1)

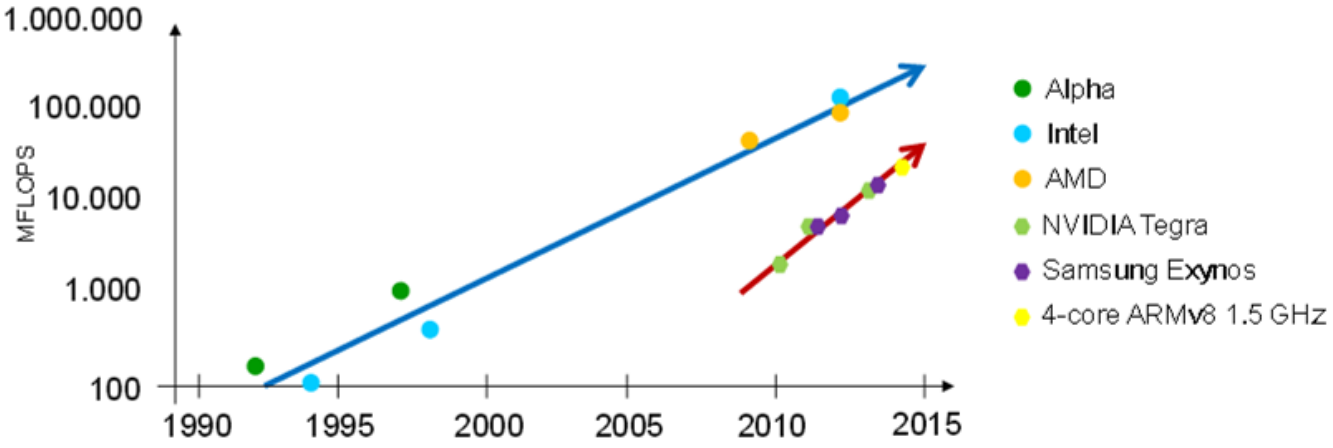
Commodity CPU vs Vector supercomputer



- Microprocessors killed the Vector supercomputers
  - They were not faster ...
  - ... but they were significantly cheaper and greener
- Need 10 microprocessors to achieve the performance of 1 Vector CPU
  - SIMD vs. MIMD programming paradigms

HPC trends (2)

Mobile CPU vs Commodity CPU



- Microprocessors killed the Vector supercomputers
  - They were not faster ...
  - ... but they were significantly cheaper and greener
- History may be about to repeat itself …
  - Mobile processor are not faster …
  - … but they are significantly cheaper

MONT-BLANC

Porting applications to ARM

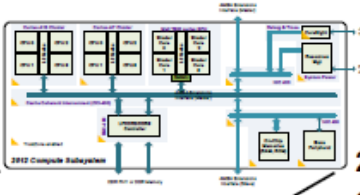
Application	Domain	Institution	Prog. Model			Scalability	ARM port
			MPI	OpenMP	Other		
YALES2	Combustion	CNRS/CORIA	Y			>32K	✓
EUTERPE	Fusion	BSC	Y	Y		>60K	✓
SPECFEM3D	Wave propagation	CNRS	Y		CUDA, SMPSs	>150K, >1K GPU	✓
MP2C	Multi-particle collision	JSC	Y			>65K	✓
BigDFT	Elect. Structure	CEA	Y	Y	CUDA, OpenCL	>2K, >300 GPU	✓
Quantum Expresso	Elect. Strcuture	CINECA	Y	Y	CUDA	Good	✓
PEPC	Coulomg + gravitational forces	JSC	Y		Pthreads, SMPSs	>300K	✓
SMMP	Protein folding	JSC	Y		OpenCL	16K	✓
ProFASI	Protein folding	JSC	Y			Good	✓
COSMO	Weather forecast	CINECA	Y	Y			✓
BQCD	Particle physics	LRZ	Y	Y		~300K	✓

« Porting full-scale HPC applications to ARM cluster requires minimal effort

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EUROSERVER Vision: Started Sept'13

EURO SERVER



1) Start with a standard scalable compute unit "die"

SEVENTH FRAMEWORK PROGRAMME

2) Connect a few together along with its local memory partition

Compute Node Multi-Chip-Module

Compute Chiplet | Compute Chiplet

Active Interposer (2D SoC + TSV)

Bumps

Memories

Package Substrate

3) Put a few on a board With shared virtualized IO's

Non-Volatile Memory

Configurable Multi-Protocol I/O Interfaces

Compute Node (64xCPU)

Compute Node

Compute Node

Compute Node

Local I/O and Remote Routing

Consumer supply

256xCPU

4) In future, integrate into a single MCM

Delivering a result 100x "better" than today

The Architecture for the Digital World® ARM®

Key contribution 4/5: Demonstration on an Embedded Server

EURO SERVER

- Tentative specs
  - Target form factor: Eurotech Mounted Mobile Computer
  - Size: 13 x 25 x 8 cm
  - Prototype targeted for rugged sealed enclosure and extended temperature range
  - Modular system design
    - Shared design for micro-server board and carrier board with Enterprise Server version
    - Up to 2 micro-server slots
    - Support for optional general purpose card (using 1 slot) for specific function or peripheral
  - 9 to 36 Vdc in; 30W max;
  - Passive conduction cooling
  - Support for Backup battery pack to investigate

Copyright EUROSERVER Consortium

HPlLC

High Performance lowpower lowcost Computing

DESCRIZIONE

- Implementazione di cluster HPC low-power/low-cost per applicazioni scientifiche
  - Nodi GPGPU mobile/embedded
    - cores CPU e cores GPU/DSP nello stesso SOC
    - technologia IP europea (a differenza delle attuali installazioni HPC)
  - Small is good(“data centre in a box”). Nicchia tutta da esplorare.
  - Non sovrapposto a progetti exascale (PRACE/MontBlanc)
  - Collaborazione con SME interessate
    - Italiane: SECO, E4, EUROTECH, etc.
    - Europee: ARM(?), Imagination Technologies (?), etc.

- Application Porting per selected use cases di interesse INFN e/o altre comunità scientifiche
  - misurazioni sul campo, strumentazione remota (tutti quelli che al momento girano con server 1U al seguito)

- Valutazione performance/watt e nuove metriche di billing:
  - cpu consumption vs cpu time

HPllC

High Performance lowpower lowcost Computing

MOTIVAZIONI

- Consolida collaborazioni con SME ⇐ è il mandato di LEIT
- Permette acquisizione di know-how su tecnologie all'avanguardia e con inevitabili potenzialità di crescita anche in campo scientifico:
  - già dal 2014/2015 prime CPU low-power a 64bit

Trasferimento tecnologico

- Abbiamo già esperienze sparse nell'INFN da capitalizzare
  - CMS, R&D del CNAF, COKA, SUMA e forse altre fuori dai gruppi ICT
- Use cases INFN probabilmente già implementabili:
  - **CMS su ARM (vd. presentazione Boccali al CCR WS di Genova)**
    - "Il porting di uno stack software complesso come quello di CMS su ARM è stato in gran parte effettuato"
    - https://agenda.infn.it/getFile.py/access?contribId=5&sessionId=3&resId=1&materialId=slides&confId=6179
  - **Tomografia computerizzata e software parallelo di ricostruzione (R.Brancaccio et al.)**
    - Attualmente il cluster HPC è ingombrante. Manca una soluzione HPC "portatile"
    - https://agenda.cnaf.infn.it/conferenceDisplay.py?confId=590

CMS su ARM

(Da presentazione T.Boccali CCR Ge)

Conclusioni

- Il porting di uno stack software complesso come quello di CMS su ARM e' stato in gran parte effettuato
- Sono stati trovati problemi piu' o meno attesi, ma in generale il fatto di avere lo stesso compilatore e' un grande vantaggio che scherma da quasi tutti i problemi
- Primi benchmarks sul sw portato (e altri benchamark HEP) dicono che consistentemente ci si possono aspettare performance pari a 0.25x per core, ma circa 4x se scalate anche per il consumo
- Avremo questi in futuro?

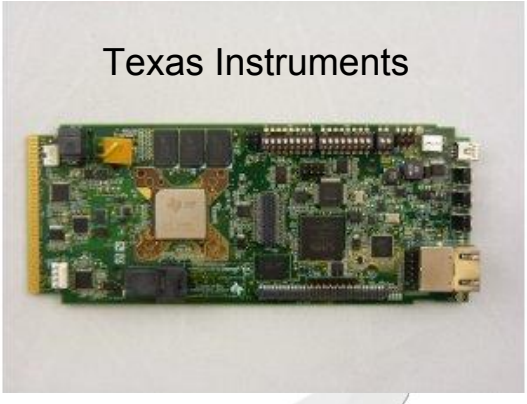
1 rack (42U):

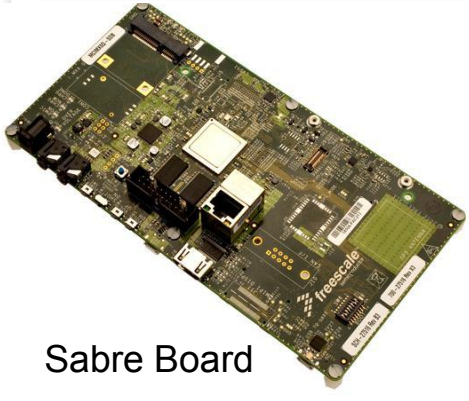
- **1000 CPUs, 4000 cores**
- **4 TB RAM**
- **~5 kW**

Macchina	Prezzo (EUR)	Numero cores	HS06 /core	HS06/2U	RAM (GB)	Potenza dissipata
Boston Viridis (2U)	10000 (?)	192	3.25	624	192	300 W
Olidata come da AQ (2U, 4 schede madri)	7000	64	11	690	256	1.2 kW (?)

HPlLC

High Performance lowpower lowcost Computing





Sabre Board

DragonBoard

Odroid-XU-E