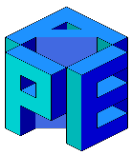


COSA WP4 status e correlazioni con progetto LEIT-ICT *picoLO*

Piero Vicini & Alessandro Lonardo
INFN – Sezione di Roma



COSA WP4 (proposal 09/2014)

WP4



- Implementazione del prototipo di cluster a ROMA1
 - Chiuso agli utenti
- Studio delle architetture di rete per sistemi SoC tramite ARM + FPGA
- Primo anno
 - 4 kit di sviluppo + un server
 - prestazioni del sistema di interconnessione fornito sui sistemi di sviluppo, i.e. Gigabit Ethernet
 - test sintetici sia a livello socket TCP/IP sia a livello di libreria message passing MPI
 - Con WP5 studio delle prestazioni della applicazione DPSNN multi-nodo
 - indicazioni utili per lo sviluppo di una rete di comunicazione dedicata
- Dal secondo semestre del primo anno
 - Espansione a 16 nodi per scalabilità
 - Progettazione e realizzazione architettura di interconnessione dedicata a bassa latenza
- Sedi coinvolte: ROMA1

30/09/2014

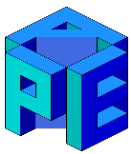
COSA Project – D. Cesini – Ferrara CSNV

ROMA1



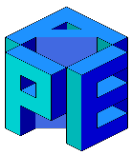
4 board ARM+FPGA based
+ 1 server } Anno I

16 board ARM+FPGA based
+ 4 server } Anno II



FPGA (Embedded ARM) + Custom network

- Sviluppo di network custom “significative” richiede l’utilizzo di componenti basati su tecnologia alla stato dell’arte
 - Linee seriali con frequenza di switching del singolo transceiver > 10 gbs (preferibilmente 14-20 gbps)
 - Supporto per host interface basato su protocolli high-end
- >10 anni fa...
 - Realizzazione di ASIC dedicati (apeNExt)
 - Minimal power consumption perche’ fortemente orientati all’applicazione,
 - altamente efficienti in termini di dimensioni del silicio (i.e costo),
 - “relative-high speed” per il livello fisico usabile -> disegno, caratterizzazione elettrica e test del sistema relativamente complesso ma fattibile
- Oggi...
 - ASIC NRE costs non compatibili: $O(N*10ME)$ per chip
 - Switching frequency dei canali seriali > 20gbps (40 gbps e oltre)
 - Complessita’ del disegno, alti costi per procurement del testbed per la misura e caratterizzazione del livello fisico



FPGA+Custom network: componenti

Altera Generation 10 FPGA:

- ❑ **Arria10** mid range (20nm) (from 2014)
- ❑ **Stratix10** high-end
 - ❑ **Introduction 2015**
 - ❑ **INTEL TriGate 14nm** -> 70% of StratixV po consumption
 - ❑ **96 transceivers @32Gbps** (56Gbps?) for chi to-chip interconnection and @28Gbps for backplane/cable interconnection
 - ❑ **Many industrial standards supported**
 - ❑ included CAUI-x (Nvlink compatible)
 - ❑ PCIe Gen3(4)
 - ❑ 40/100GEth
 - ❑ tons of programmable **logic @1GHz**
 - ❑ ...and "for free"...
 - ❑ Support to **HMC**
 - ❑ **10 Tflops of DSP** single precision FP performances
 - ❑ Multiple (4->8) **ARM Cores (a53) @1.5GHz**

Altera's 10th Generation Portfolio – 2x More

2x More Bandwidth
Arria 10: 28.05 Gbps
Stratix 10: 56 Gbps

Up to 1 GHz Core Performance
Arria 10: 1.6x more than Stratix 10: 2x more than !

SoC
"All-In" with ARM for SoC

Lower Power
60-70% Lower Power

Power Saving Innovations

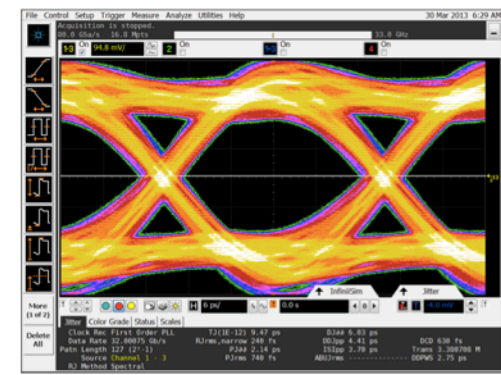
SoC across Density Range
>2x more Fabric Density

14 nm
4th Generation HK+MG
2nd Generation Tri-Gate

Altera
MEASURABLE ADVANTAGE™

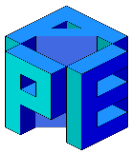
© 2013 Altera Corporation—Public
* Core performance versus previous generation

Industry's First Transceivers @20 nm
Transmit Eye at 32 Gbps



First Transceivers @ 20 nm Validate 28G Operation for Arria 10

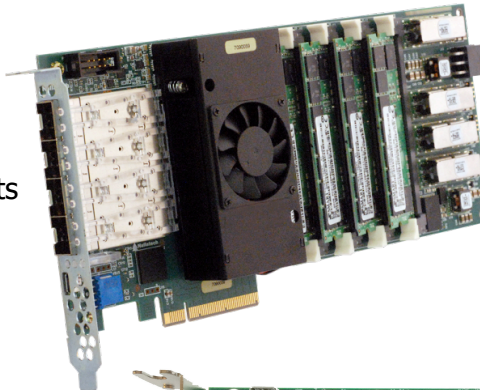
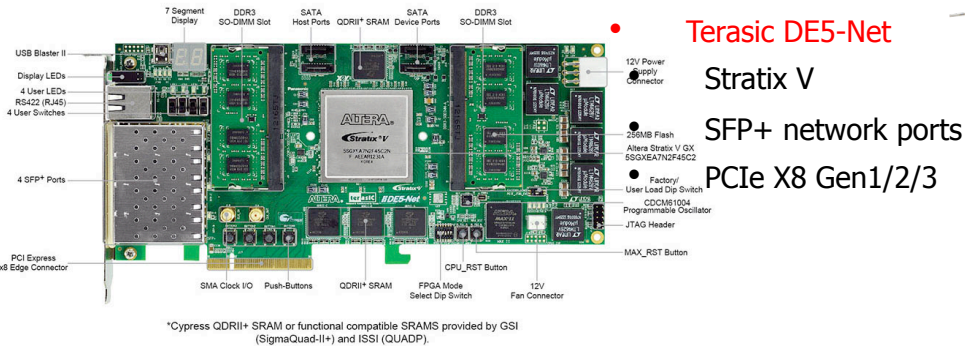




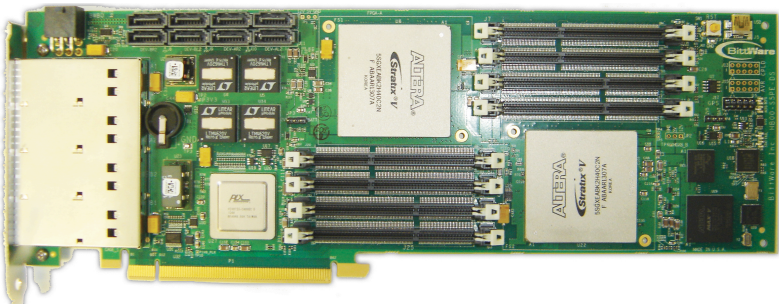
FPGA+Custom network: sistemi di sviluppo

High-end o SoC con multicore CPU integrati

- APEnet+, Terasic, Nallatech, Bittware,...



- **Nallatech 395-AB**
- Stratix V AB
- Highest density memory: 4x 32GB of DDR3
- (4) SFP+ network ports supporting a range of network protocols and speeds

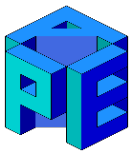


- **Bittware S5-PCI-DS**
- Dual Stratix V
- OpenCl enabled, PCI x16, huge memory banks..., SFP+ conn.



- **Bittware A10PL4 (Arria10 based)**
- Altera Arria 10 GT/GX FPGA
- PCIe x8 Gen1, Gen2, or Gen3
- QSFP for 2x 100GigE, 2x 40GigE, or 8x 10GigE
- Memory: up to 32 GBytes of DDR4

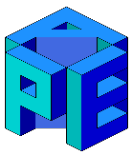




Perche' FPGA SoC in COSA?

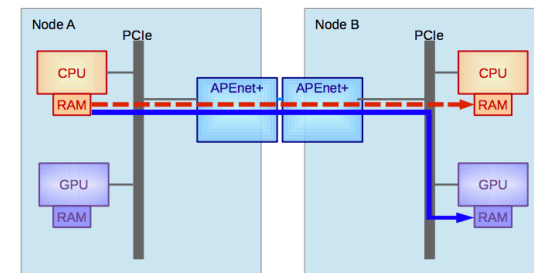
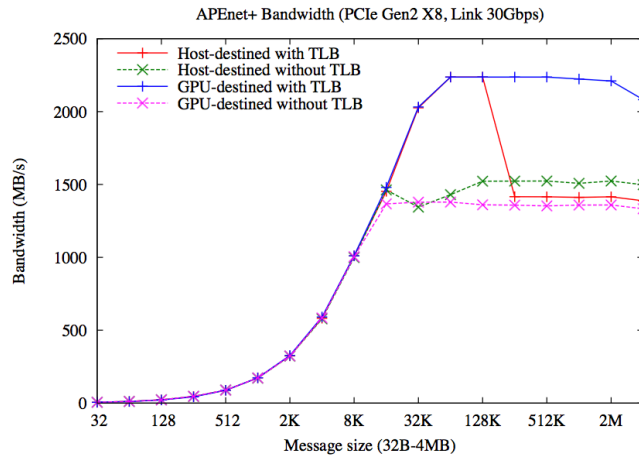
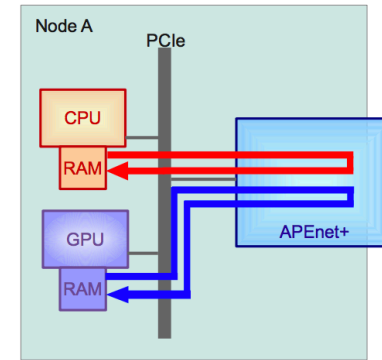
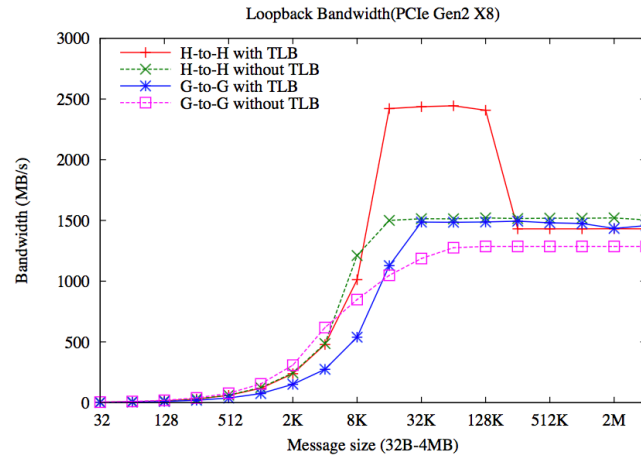
FPGA SoC (System on Chip) e' un componente ibrido che integra una sezione hardware programmabile e configurabile dall'utente (FPGA) e core(s) di processori low power, high performance.

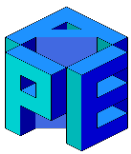
- Due motivazioni principali per l'adozione di FPGA SoC:
 - Accelerazione di task computazionali eseguiti dal uP embedded nella FPGA correlati al protocollo RDMA implementato nell'architettura di rete custom APEnet
 - APEnet V5, NaNet e derivati, sistemi picoLO
 - FPGA Embedded hardened multiple ARM cores (anche a 64 bit) come processore di calcolo ("accettabile" in termini di prestazioni) caratterizzato da basso consumo di potenza e integrazione diretta con la network.
 - Esplorazione di nuove architetture dedicate (es. DPSNN)
 - ExaNeSt e derivati...



FPGA Embedded uP + RDMA(2)

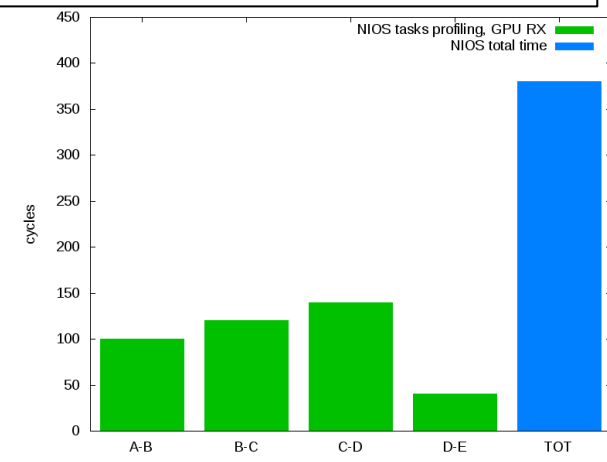
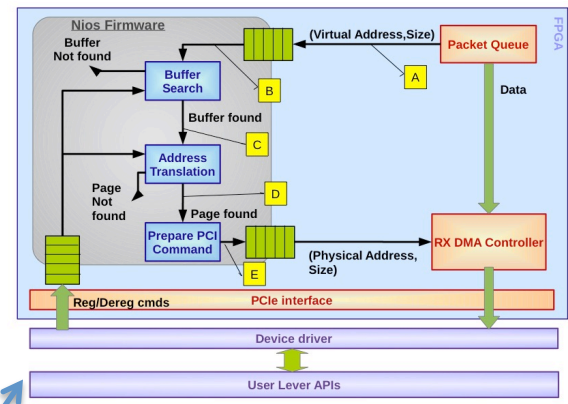
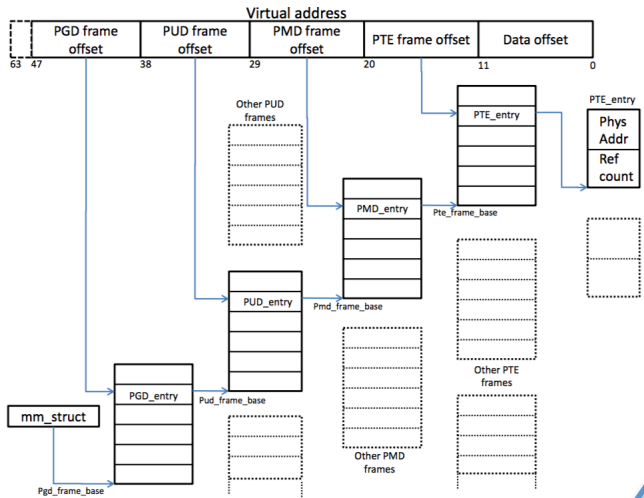
How TLB impacted on performances





FPGA Embedded uP + RDMA(3)

"ASIP acceleration for virtual-to-physical address translation on RDMA-enabled FPGA-based network interfaces"; Ammendola et al, *Future Generation Computer Systems*, 2015



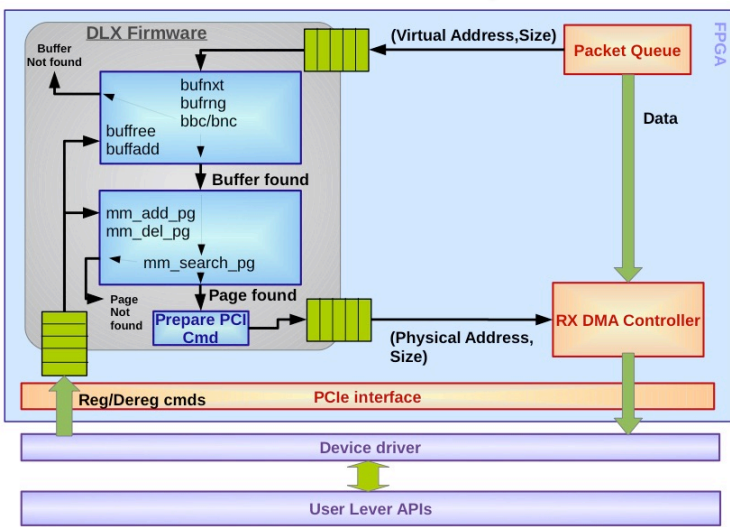
Logic Block	Comb. ALUT	ALMs	Register	Memory [KB]	Freq [MHz]
APENet+	89274 (38%)	74862	75552 (32%)	1110.6 (64%)	
Nios II	1975	1702	1528	140.1	200
Nios II subsystem	13188	10470	13026	306.0	200
V2P	1168	932	865	8.8	250
BSRC	837	968	1009	0.2	250
TLB total	11450	9026	8303	70.8	250
D64OPT	8138	14083	9479	49.2	165

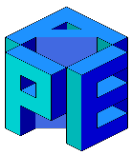
Table 1: Altera Stratix IV EP4SGX290 resource utilization by entity.

Operations	Nios II	DLX	D64	D64AC	D64SB	D64OPT
Append 0...31	6839	4867	4836	4836	418	419
Search 0, 16, 31	1802	2454	1938	1644	794	304
Remove 0, 16, 31	1192	3523	2732	2438	1504	560
Search 16	596	787	517	455	483	166
Total	10433	11631	10023	9373	3199	1449

Operations	Nios II	v0	v1	v2	v3
mm_del_pg	165	125	83	78	78
mm_add_pg	1299	85	77	77	77
mm_search_pg	29	44	38	13	12

nt proces-





FPGA Embedded ARMs come nodo di calcolo (1)

- Embedded ARM cores, presenti nella corrente (e prossima) generazione di FPGA, hanno *performance molto interessanti* e sono accoppiati a strutture di memoria state-of-the-art (HMC)
- Integrazione "stretta" tra processore embedded e rete dovrebbe garantire
 - alti ratio flops/watt e flops/volume
 - esplorazione di architetture di calcolo esotico efficienti su applicazioni scientifiche di nostro interesse (ad es. DPSNN)
- Si parla di ALTERA ma la componentistica XILINX ha simili architettura e performances...

SoC FPGA – Benefits of Integration

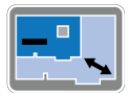
Altera – INFN Confidential



Increased system performance



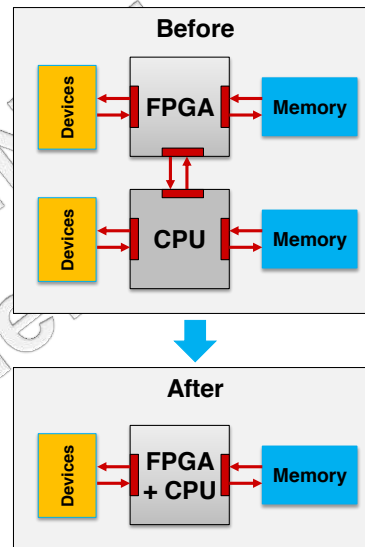
Reduced power consumption



Reduce board size



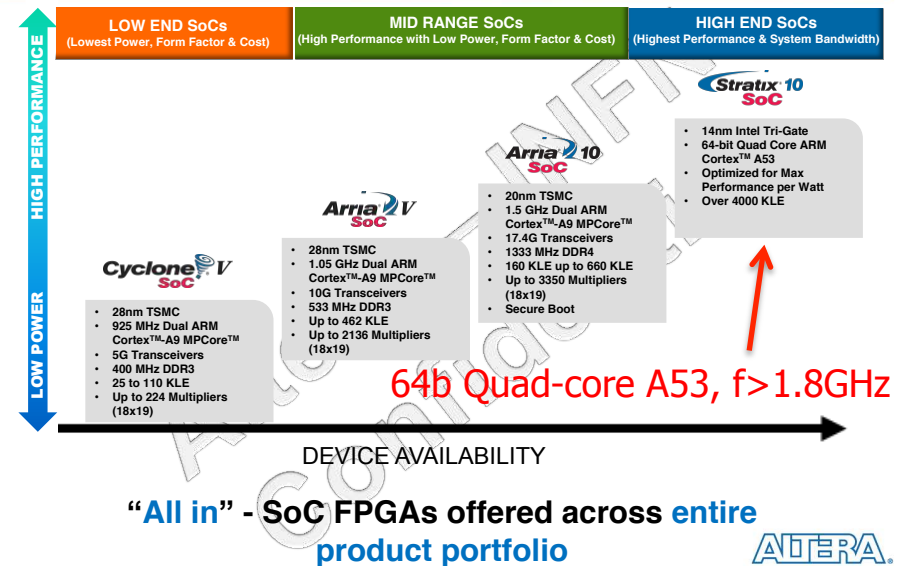
Reduced system cost



ALTERA
MEASURABLE ADVANTAGE™

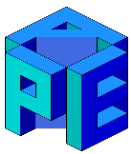
Altera SoC Product Portfolio

Altera – INFN Confidential

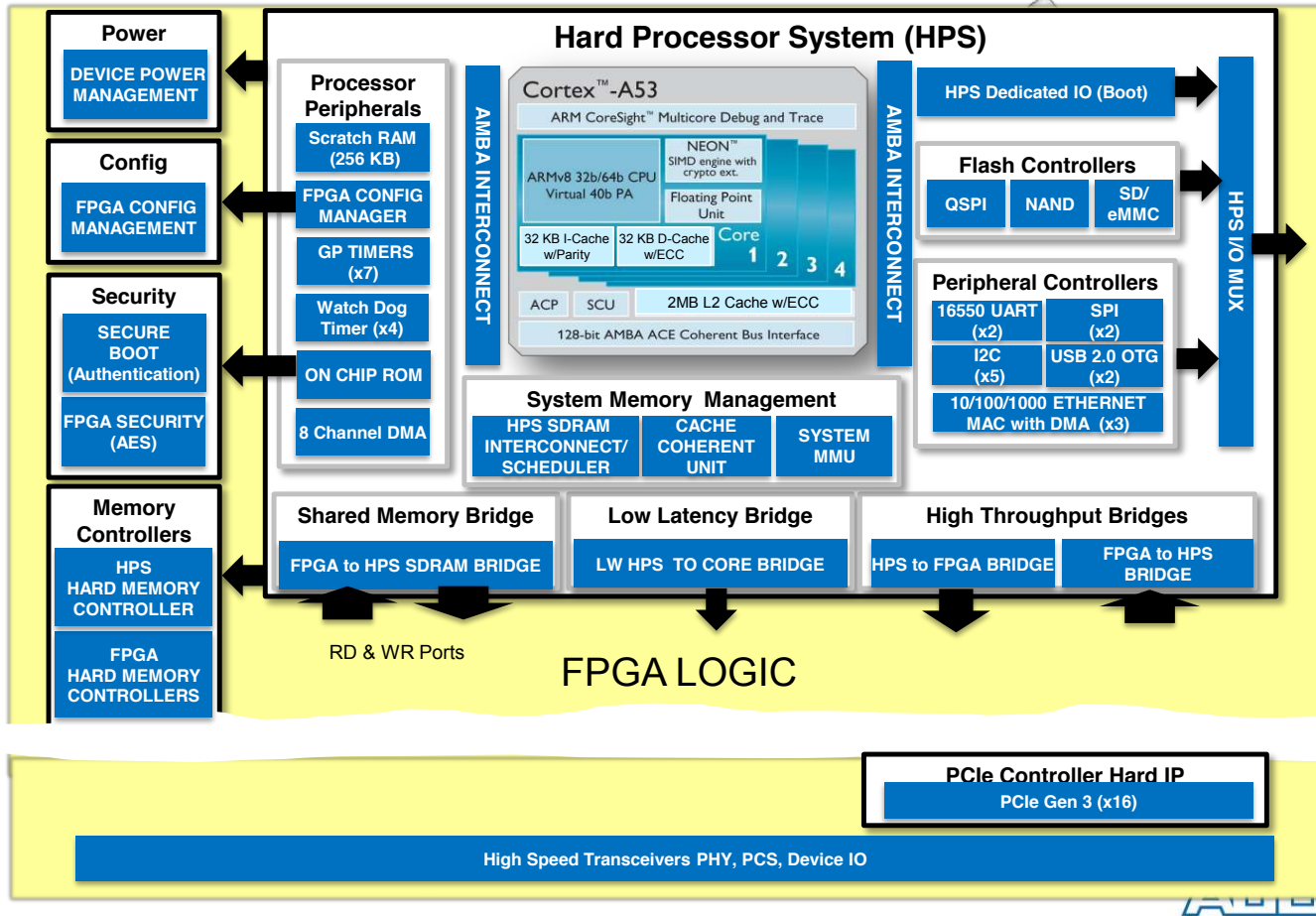


ALTERA
MEASURABLE ADVANTAGE™





FPGA Embedded ARMs come nodo di calcolo (2)



High performance vs. ARM Cortex™ A9

- Quad Core (8 stage, dual issue in-order pipeline)
- Improved Integer Performance
- Improved NEON and FPU Co-processor performance
 - 16 -> 32 registers makes SIMD more usable

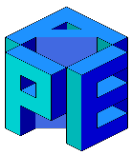
64-bit ARM v8 Architecture

- Software compatible with 32b w/AArch32
- 64b registers, 48b virtual address, 32b instructions
- Large application address space and wider data movement

Improved Cache with Data Integrity

- 2M L2 Cache with ECC (16-way set associative)
- L1 D-Cache now with ECC (4-way set associative)





FPGA Embedded ARMs come nodo di calcolo (3)

High Performance Interconnect

Application Usage

- Co-processing with FPGA based processors, real time, packet & search engines etc.

High Throughput Bridge

- Order of magnitude lower latency vs. chip to chip interfaces

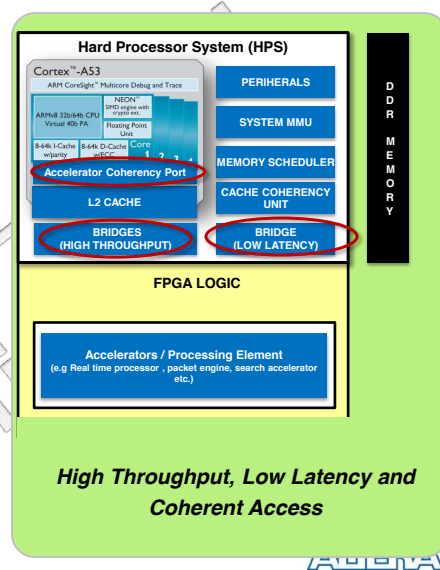
Low latency Bridge (non blocking)

- Allows simple accesses to FPGA logic without blocking high throughput traffic

Accelerator Coherency Port

- Increased performance, coherency managed in hardware eliminates need for L2 Cache Flush
- Supports wide range of user defined masters (FPGA accelerators e.g.) using ACP mapper IP

Common AMBA interface as previous generation SoCs allows IP reuse



Verso completa integrazione in OpenPower (CAPI protocol...)

High Performance HPS Hard Memory Controller

Application Usage

- Can be used as dedicated memory controller for HPS or
- Shared memory between FPGA and HPS (co-processing)

Efficient Memory Scheduling

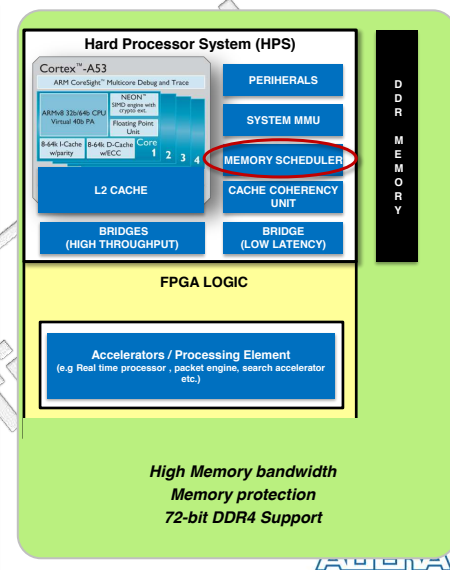
- Built in scheduler employs intelligent algorithm (deficit weight round-robin) ensure efficient access to memory

Built in Memory Protection Unit

- Prevents FPGA from accesses restricted OS space in memory and crashing the system

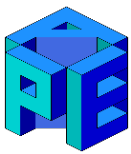
Support for 72-bit DDR4 Memory

- 3200 Mbps



Supporto per tecnologia di memoria state-of-the-art



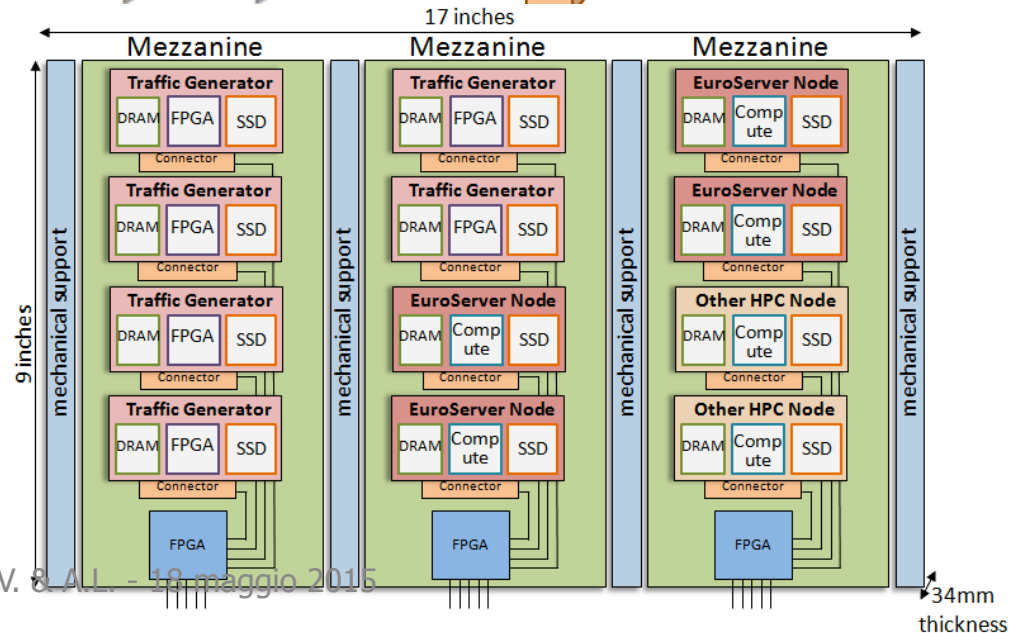
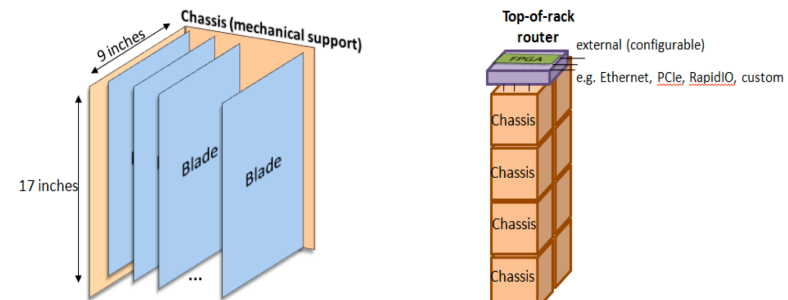


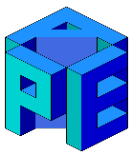
ExaNeSt: un caso di studio immediato

- "...ExaNeSt will develop, evaluate, and prototype the **physical platform and architectural solution for a unified Communication and Storage Interconnect**, plus the physical rack and environmental structures required to deliver European Exascale Systems..."
- INFN nel progetto con responsabilita' e tasks legati alla network, allo storage ed al benchmarking della piattaforma attraverso applicazioni (DPSNN e LQCD)
- Possibilita' di esplorare architetture scalabili non convenzionali FPGA-based per le applicazioni scientifiche di nostro interesse

ExaNeSt assembla un rack completo interamente popolato e "fully functional"

- Blade: N(3) mezzanine card che ospitano
 - "Traffic Generator Module" e "medium" performance processors ARM-based integrati in FPGA high-end
 - ARM
 - High-end FPGA per network di comunicazione
- 9 blades/chassis 8 chassis/rack
- ToR configurabile per scalabilita' all'ExaFlops e high performance I/O
 - Testbed per link ottici





Cluster FPGA-based di Roma: stato

Componenti:

- Server 3U multi PCIe slot per alloggiamento schede e sviluppo firmware/software
 - In attesa della selezione finale dei FPGA dev kit
- 2+2 development kit per FPGA SoC
 - Nel timeframe del primo anno di Cosa disponibili CycloneV o ArriaV, Arria10
 - Target system Arria10 Soc Dev Kit
 - All'inizio del progetto erano previste arrivare per 1Q15.
 - Ora previste sul mercato per 3Q15 ritardo dovuto al design della board non adeguato in termini di power e signal integrity....
 - Form factor non ottimale: serve jumper cable per connessione allo slot del server
 - Piccolo problema da risolvere: costo reale ~2x il costo stimato a Luglio 2014...

V.le Palmiro Togliatti, 1639 IT - 00155 Roma
tel + 39 06 4063665 - 789
email: marco.passeri@ebv.com

I.N.F.N. SEZIONE DI ROMA 713836

84001850589
PIAZZALE ALDO MORO, 2
IT 00185 ROMA

QuoteRef.	Your Reference	Your Date	Contact	Page	Date
1284720			Marco Passeri	1	06/02/2015

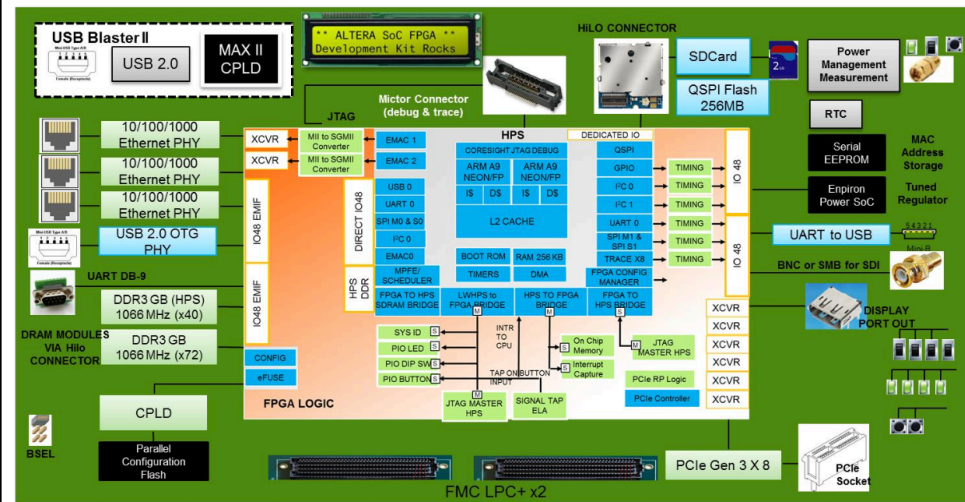
C.A. PIERO VICINI

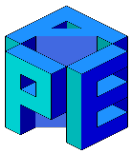
Facendo riferimento alla Vostra gradita richiesta, Vi rimettiamo la nostra migliore offerta con validità 30 giorni per quanto di seguito specificato.
la presente offerta / fornitura è regolata dalle "condizioni di vendita" disponibili sul sito "www.ebv.com" alla sezione "term & condition".

Line	Device	Manuf.	Qty.	MPQ	Price/Unit	Stock	Leadtime Weeks
1	DK-SOC-10AS066S-A RoHS-compliant* SoC Development Kit - 4995 usd	ALT	1	1	4.350,0000 EUR	MAGGIO 2015	99
2	DK-DEV-10AX115S-A RoHS-compliant* FPGA Development Kit - 4495 usd	ALT	1	1	3.950,0000 EUR	MAGGIO 2015	99

- merce resa f.co nostro magazzino
- legame valutario fisso no cambio
- pagamenti come in uso

Cordiali Saluti
Marco Passeri

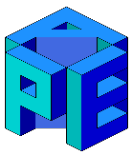




APEnet+ V5 status

In attesa delle nuove FPGA con ARM cores embedded prosegue lo sviluppo e l'ottimizzazione della versione V5 di APEnet+. In particolare:

- ❑ Finalizzazione dell'interfaccia con PCIe gen3 x8 realizzata in due flavors distinti
 - PLDA PCIe core (black box, high cost...)
 - Altera core nativo:
 - richiesto sviluppo e ottimizzazione del backend
 - necessario in prospettiva per il controllo completo della tecnologia di interfaccia con l'host
- ❑ Ottimizzazione dell'interfaccia con la nuova architettura del link seriale
- ❑ Ottimizzazione del motore RDMA GPU/CPU
 - Dual TX RDMA
- ❑ Sviluppo driver in user space



APEnet+ V5 status: piattaforma di prototipazione

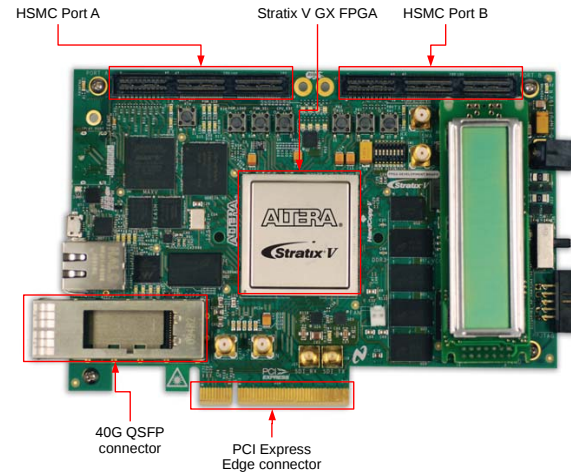
APEnet+ V5

Based on DK-DEV-5SGXEA7N dev kit:

- **New 28nm Stratix V FPGA**
- 40Gb QSFP+ standard interconnect fabric
- HSMC expansion ports
- PCIe connector
- 1Gbit PHY

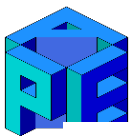
- **PCIe Gen3 X8**
 - Featuring 8.0 GT/s
 - Encoding scheme
 - 8b10b (20%) -> 128b130b (2%)
 - PLDA PCIe CORE IP uses axi4 interface protocol

- **Enhanced embedded transceiver**
 - up to 14.1 Gbps
 - X channel implemented using 40Gbps QSPF+ connector
 - measured at 45Gbps
 - Y/Z channels implemented on the HSMC interfaces
 - measured at 31.2Gbps/channel (to be improved)

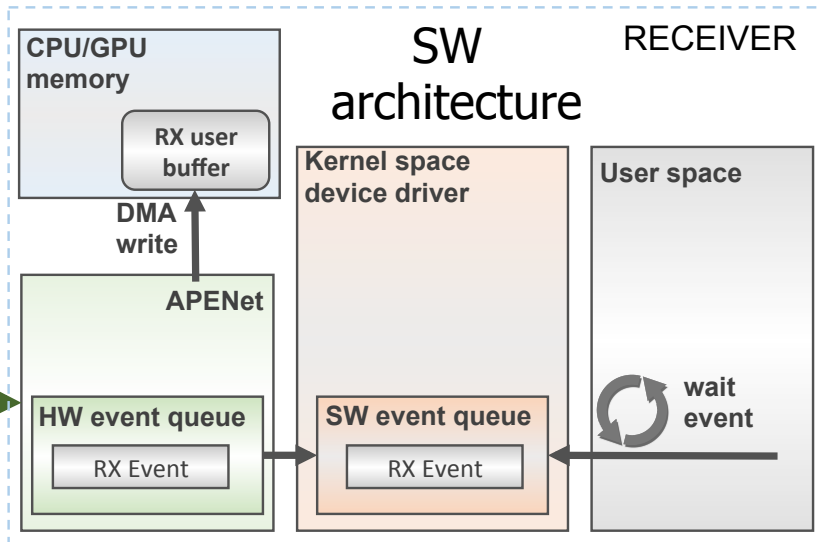
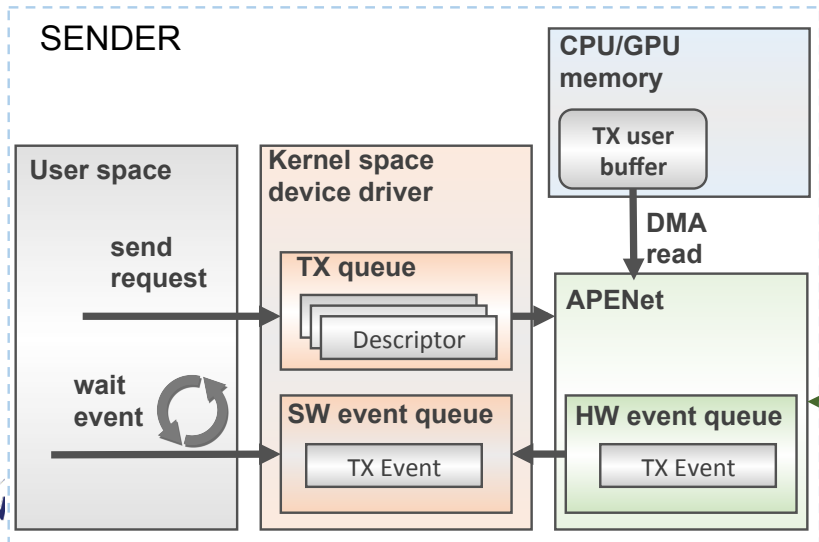
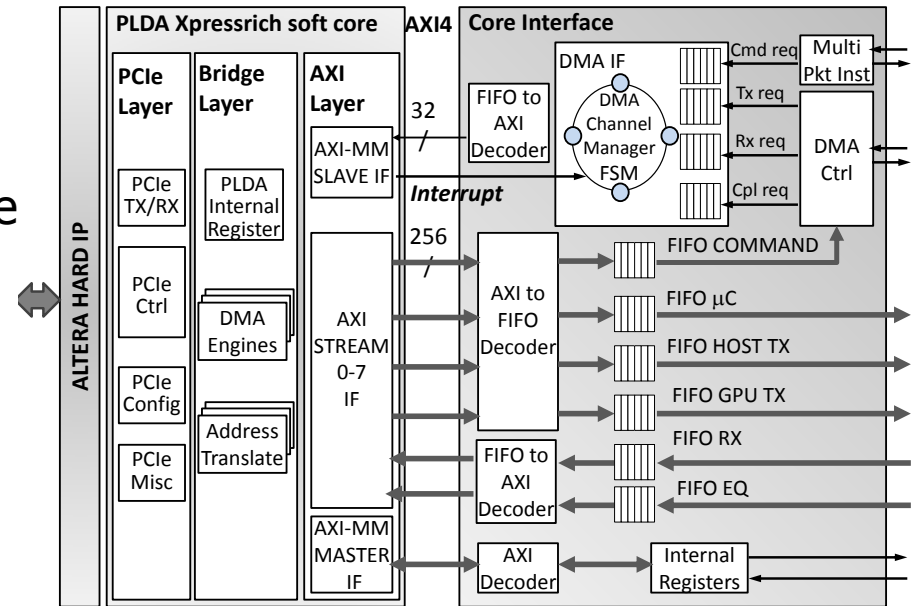
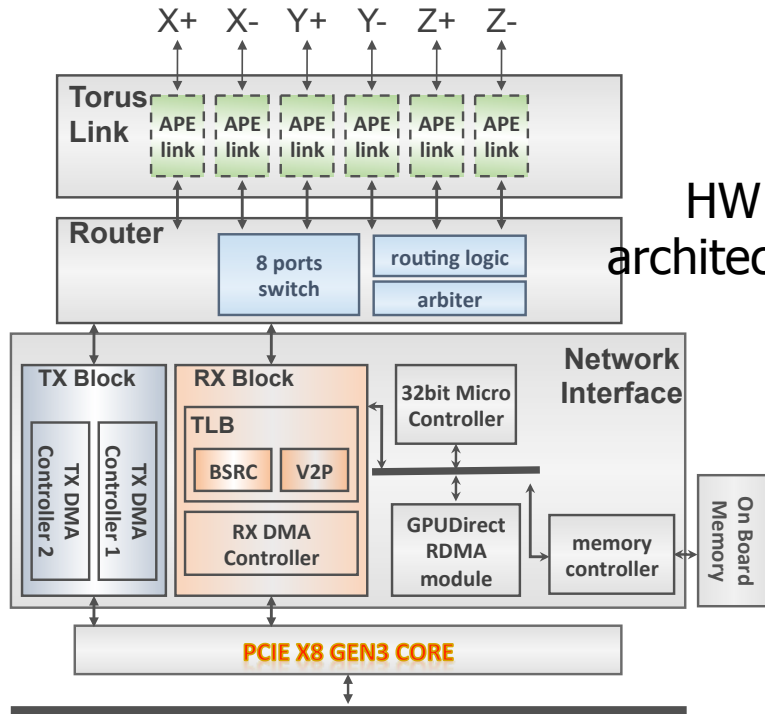


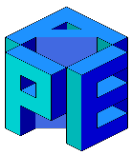
Preliminary BER measurement

Cable	BER	Data Rate
10m Mellanox optical cable	< 2.36 E-14	11.3 Gbps
1m Mellanox copper cable	< 1.10 E-13	10.0 Gbps



APENet+ V5 status: architettura HW e SW

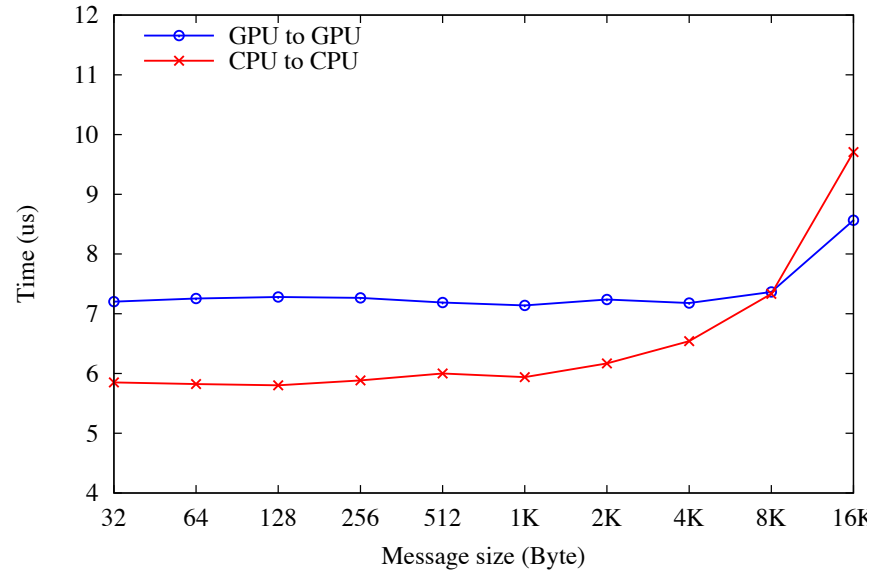
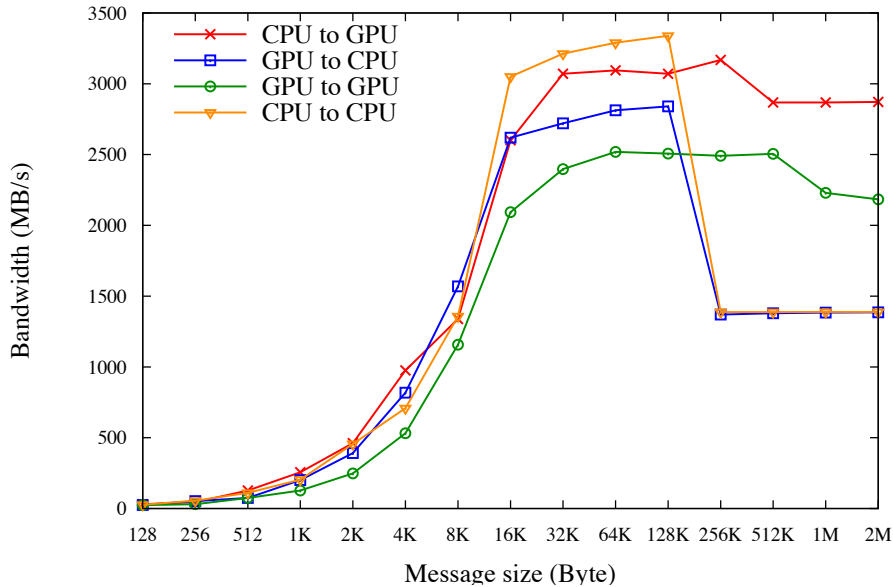




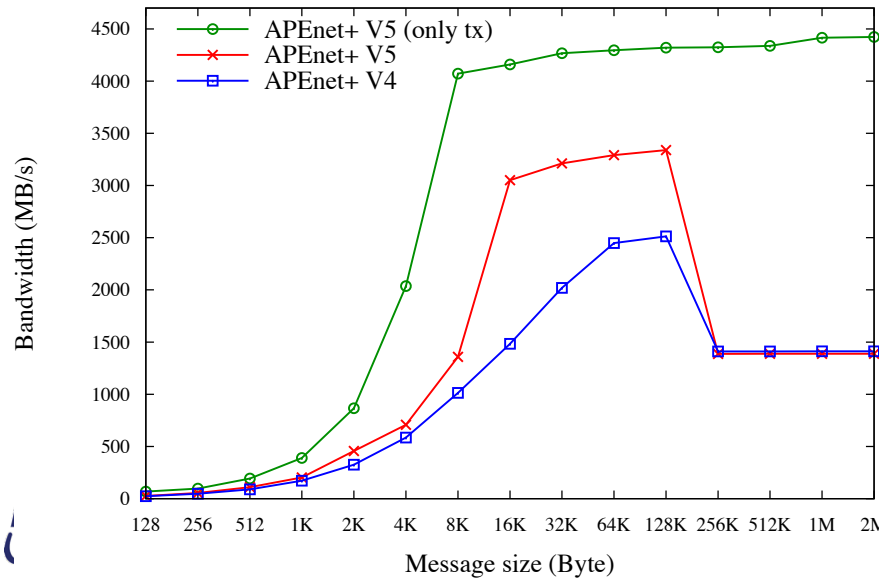
APEnet+ V5 status: risultati

APEnet+ V5 (PCIe Gen3 X8) & GPU Tesla K20X (on Ivy Bridge) ...in progress

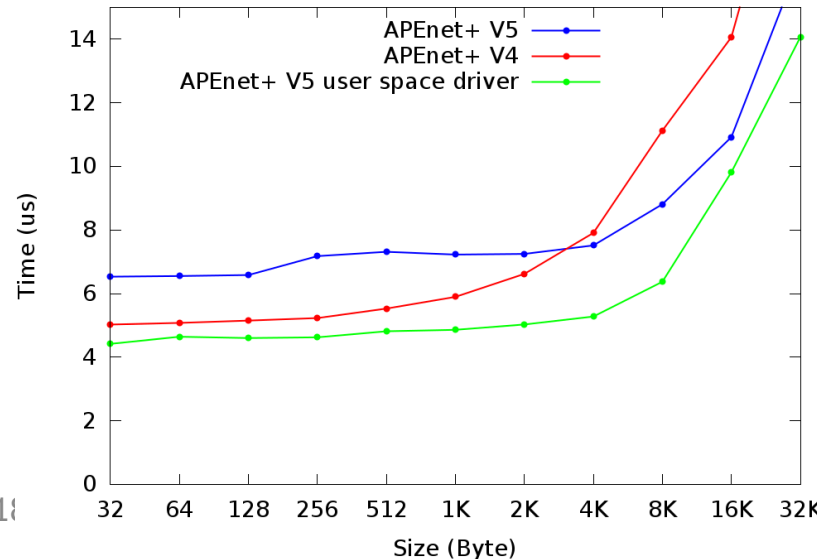
APEnet+ V5 (PCIe Gen3 X8) & GPU Tesla K20X (on Ivy Bridge)

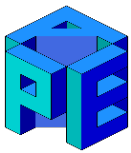


APEnet+ V5 (PCIe Gen3 X8) & APEnet+ V4 (PCIe Gen2 X8)



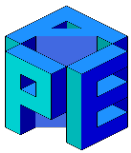
Latency





Proposta picoLO

- LEIT Research and Innovation Action (RIA) H2020 ICT04 – 2015 (a)
 - **Next generation servers, micro-server and highly parallel embedded computing systems based on ultra-low power architectures:** The target is highly performing low-power low-cost micro-servers, using cutting-edge technologies like, for example, optical interconnects, 3D integrated system on chip, innovative power management, which can be deployed across the full spectrum of home, embedded, and business applications. Focus is on integration of hardware and software components into fully working prototypes and including validation under real-life workloads from various application areas. Specific emphasis is given on low-power, low-cost, high-density, secure, reliable, scalable small form-factor datacentres ("datacentre-in-a-box").
 - **New cross-layer programming approaches** empowering developers to effectively master and exploit the full potential of the next generations of computing systems based on heterogeneous parallel architectures and constituting the computing continuum. Beyond performance, optimisation should include energy efficiency, time-criticality, dependability, data movement, security and cost-effectiveness. Research should also aim at radically increasing the productivity in programming and maintaining intrinsically parallel code by marginalising the need for dual expertise - application engineering and computer system engineering. Focus is on holistic approaches hiding the complexity between the computing HW component level and the level of application families.
- Chiusa il 14-04-2015.
- Valutazione entro settembre 2015.



Proposta picoLO

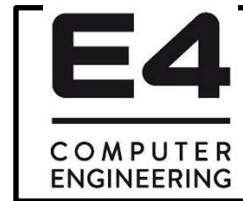


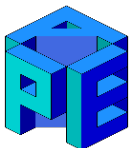
**Barcelona
Supercomputing
Center**
*Centro Nacional
de Supercomputación*

SIEMENS

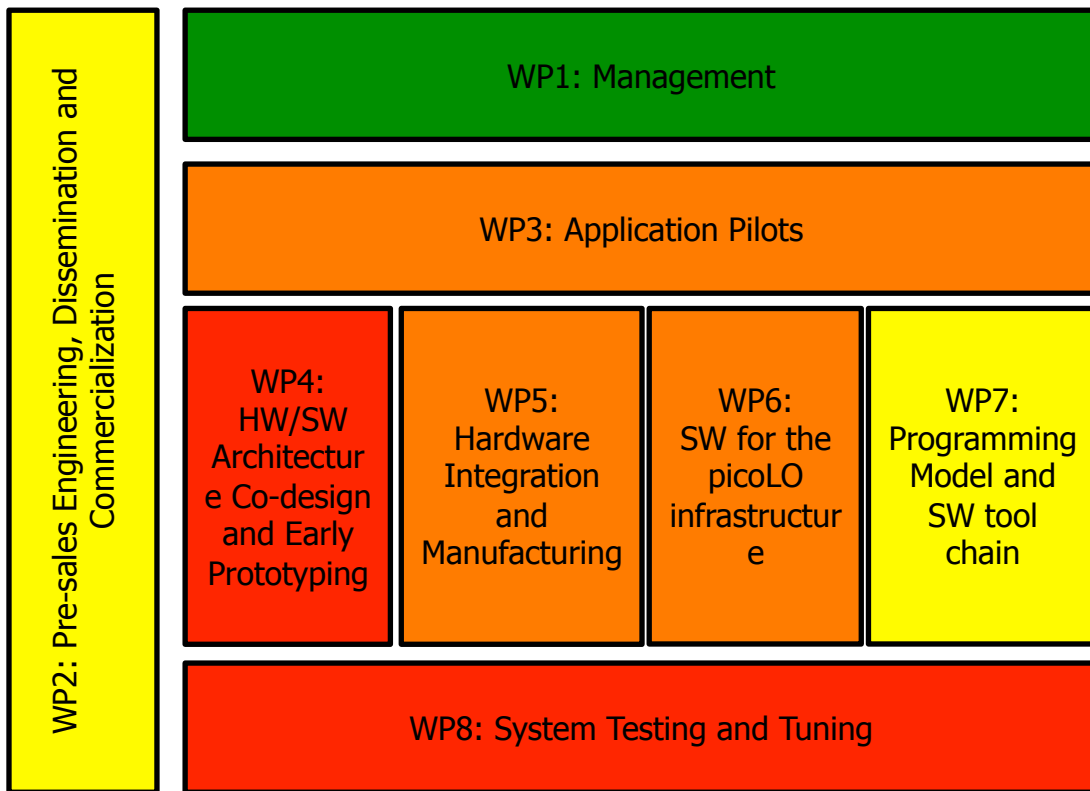


FEV





Proposta picoLO



Durata: 36 Mesi.

Budget totale: 7.8 M€.

Budget INFN: 1.2 M€.

Personale INFN: 5-7 unità.

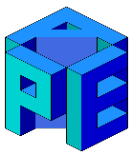
Sezioni coinvolte: Roma, CNAF.

people@Roma:

1. Alessandro Lonardo (contact)
2. Pier Stanislao Paolucci
3. Piero Vicini

people@CNAF:

1. Daniele Cesini
2. Andrea Ferraro



Proposta picoLO

WP1: Management

Leader: BSC, Contributors: All partners

WP2: Pre-sales Engineering, Dissemination and Commercialization

Leader: E4, Contributors: BSC, SIE, INFN, SECO

WP3: Applications

Leader: SIE, Contributors: BSC, INFN, LUH, FEV

WP4: Hardware/Software Architecture Co-design and Early Prototyping

Leader: INFN, Contributors: BSC, UJF-TIMA, SECO, E4

WP5: Hardware Integration and Manufacturing

Leader: SECO, Contributors: INFN, E4

WP6: Software for the picoLO infrastructure

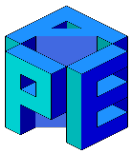
Leader: UJF-TIMA, Contributors: INFN, E4

WP7: Programming model and Software toolchain

Leader: BSC, Contributors: SIE, INFN, UJF-TIMA, JUELICH, LUH

WP8: System Testing and Tuning

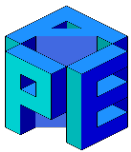
Leader: INFN, Contributors: BSC, SIE, UJF-TIMA, SECO, FEV



Proposta picoLO

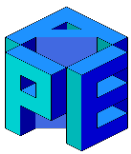
Architettura di calcolo **embedded scalabile** caratterizzata da valori ottimali dei rapporti FLOPS/W and FLOPS/€

- La **picoCARD** rappresenta il modulo elementare attorno a cui è costruita l'architettura.
- Due tipi di picoCARD
 - **picoCOMP**, implementano le capacità di calcolo, storage e I/O locali;
 - **picoXIO**, implementano la connettività a bassa latenza e l'I/O di sistema (vedi oltre).
- La picoCOMP è una carrier board che alloggia un certo numero di (ad es. 2) di Computer on Module (**picoPROC**) (standard da definire, ad es. Qseven).
- Due tipi di picoPROC:
 - picoPROC-1: SoC multi-core ARM con acceleratore many-core (ad es. **nVIDIA Tegra X1**).
 - picoPROC-2: FPGA SoC device (ad es. **ALTERA Arria-10**).



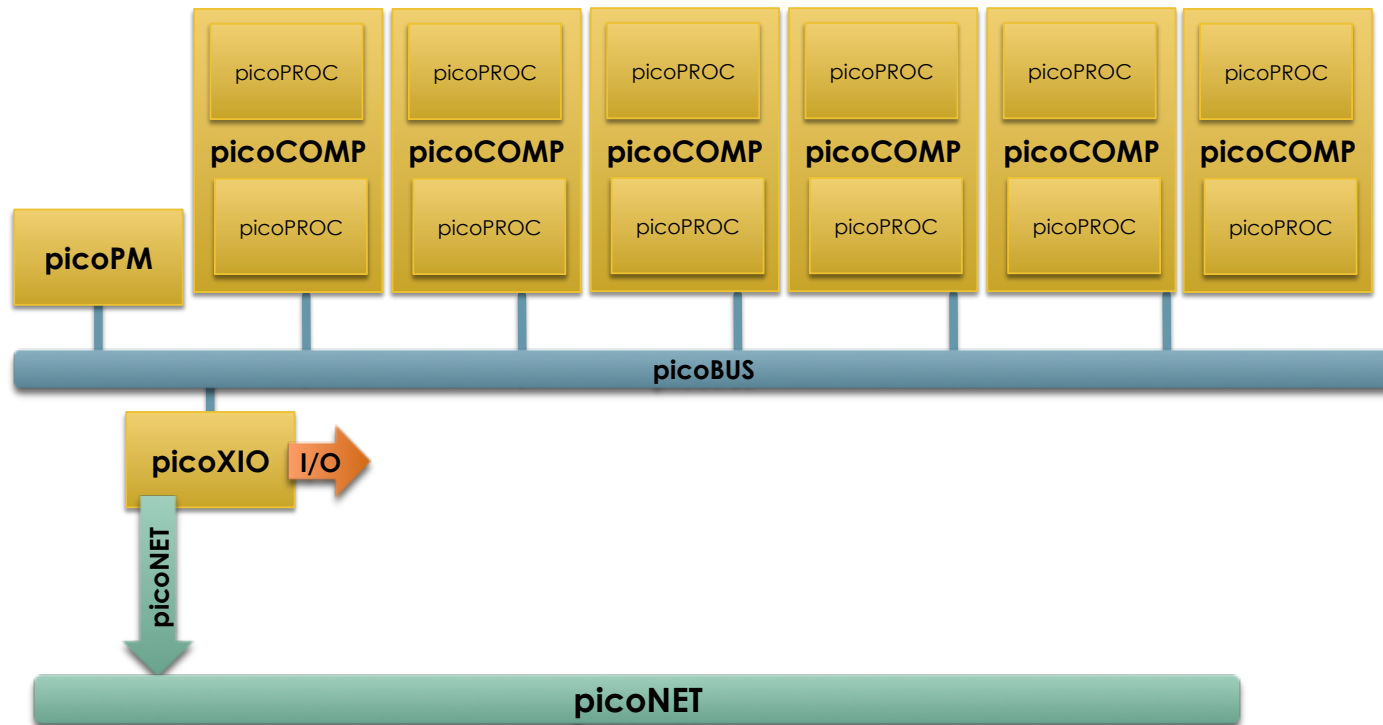
Proposta picoLO



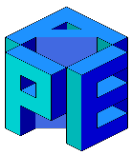


Proposta picoLO

Un certo numero (max 6) di picoCOMP possono essere assemblate assieme ad una board picoXIO in modo da costruire un sistema **picoBOX** ottimale per l'applicazione di interesse.



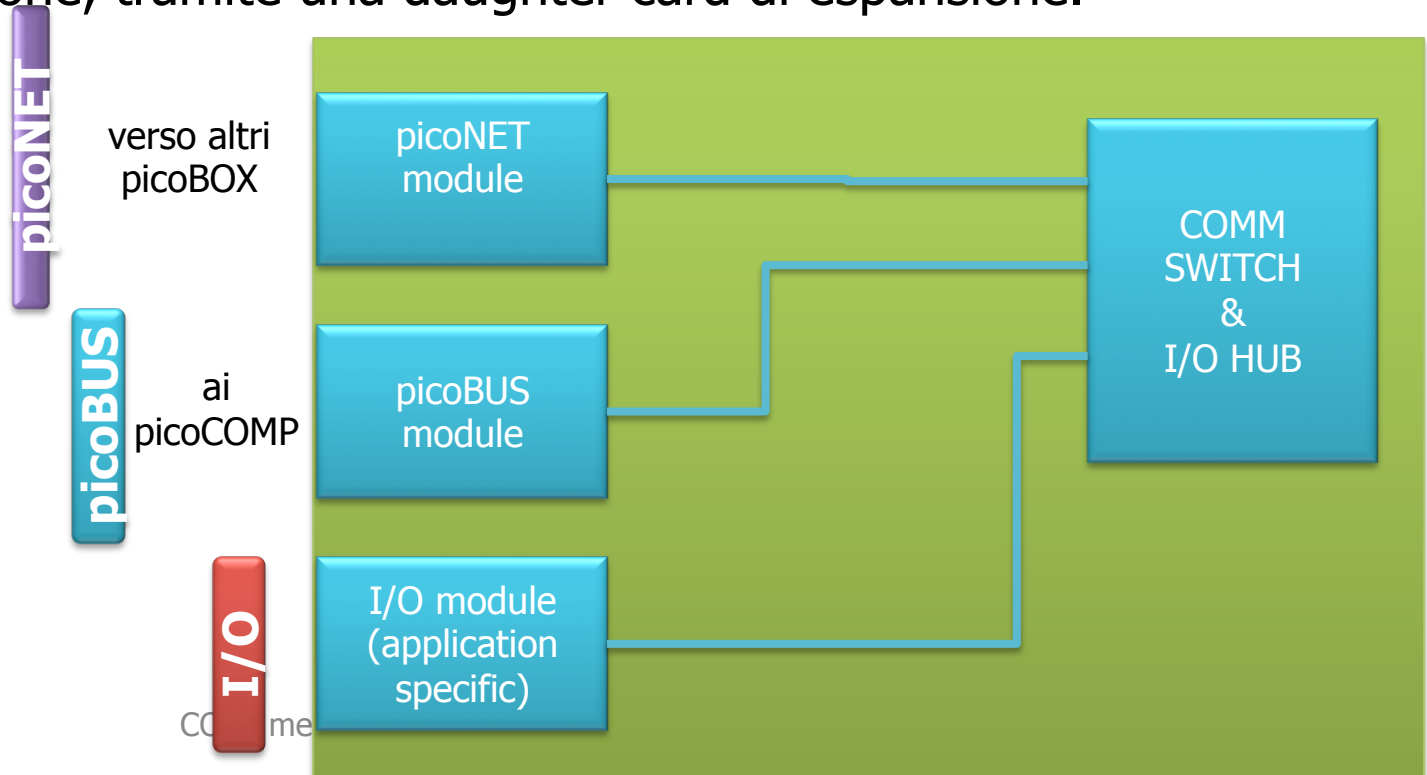
Il modulo **picoPM** implementa il power supply ed i servizi di monitoring e management remoti per i componenti del picoBOX.

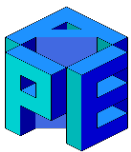


Proposta picoLO

La picoXIO card implementa

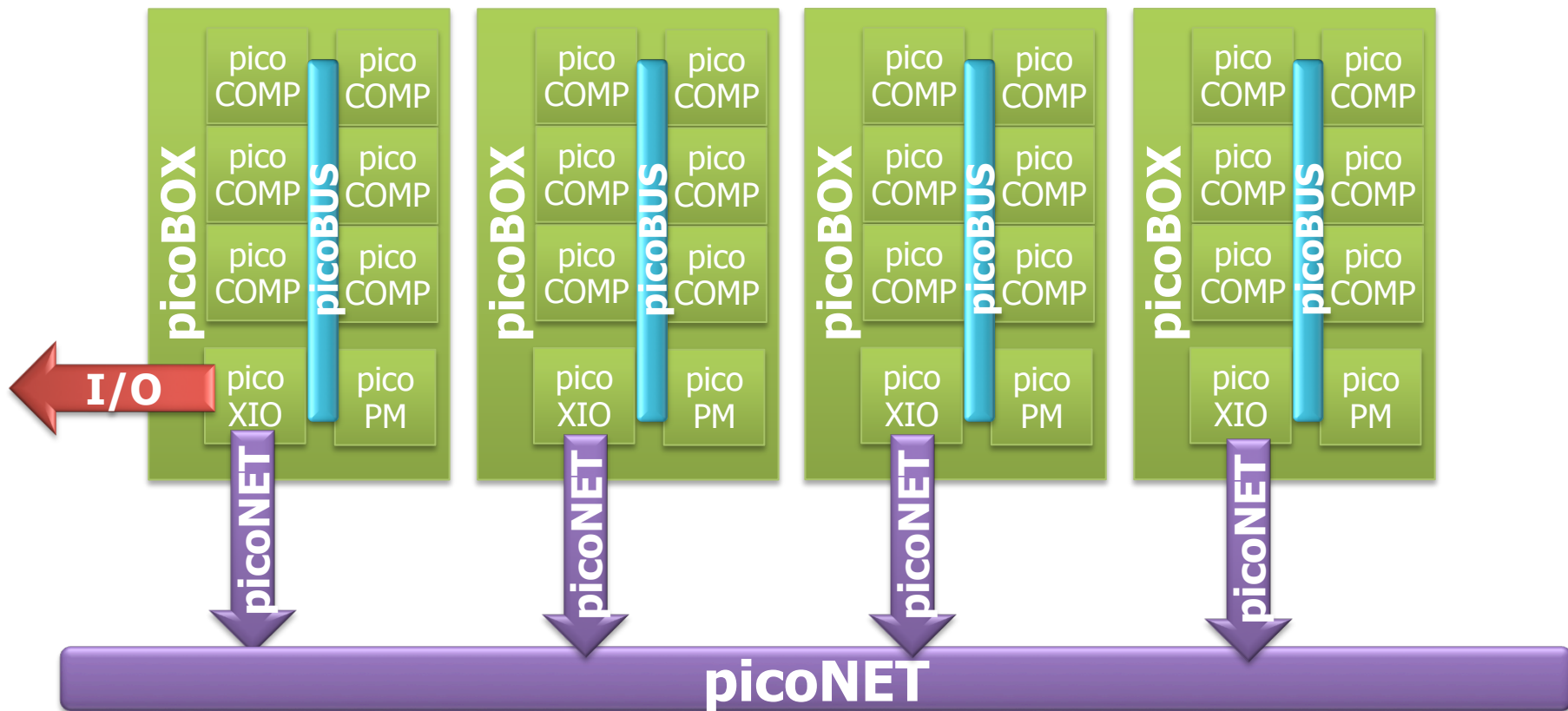
1. le comunicazioni a bassa latenza tra picoPROC che appartengono alla stessa picoBOX e con le omologhe picoXIO che appartengono a picoBOX diverse.
2. Un certo numero di canali di I/O standard (ad es. 10GbE) condivisi dai picoPROC della picoBOX.
3. Uno o più socket per aggiungere funzionalità di I/O specifiche per la applicazione, tramite una daughter card di espansione.

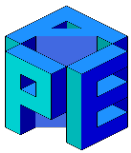




Proposta picoLO

Diversi picoBOX possono essere collegati dalla rete picoNET per costruire un cluster (**picoSTACK**).

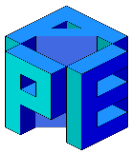




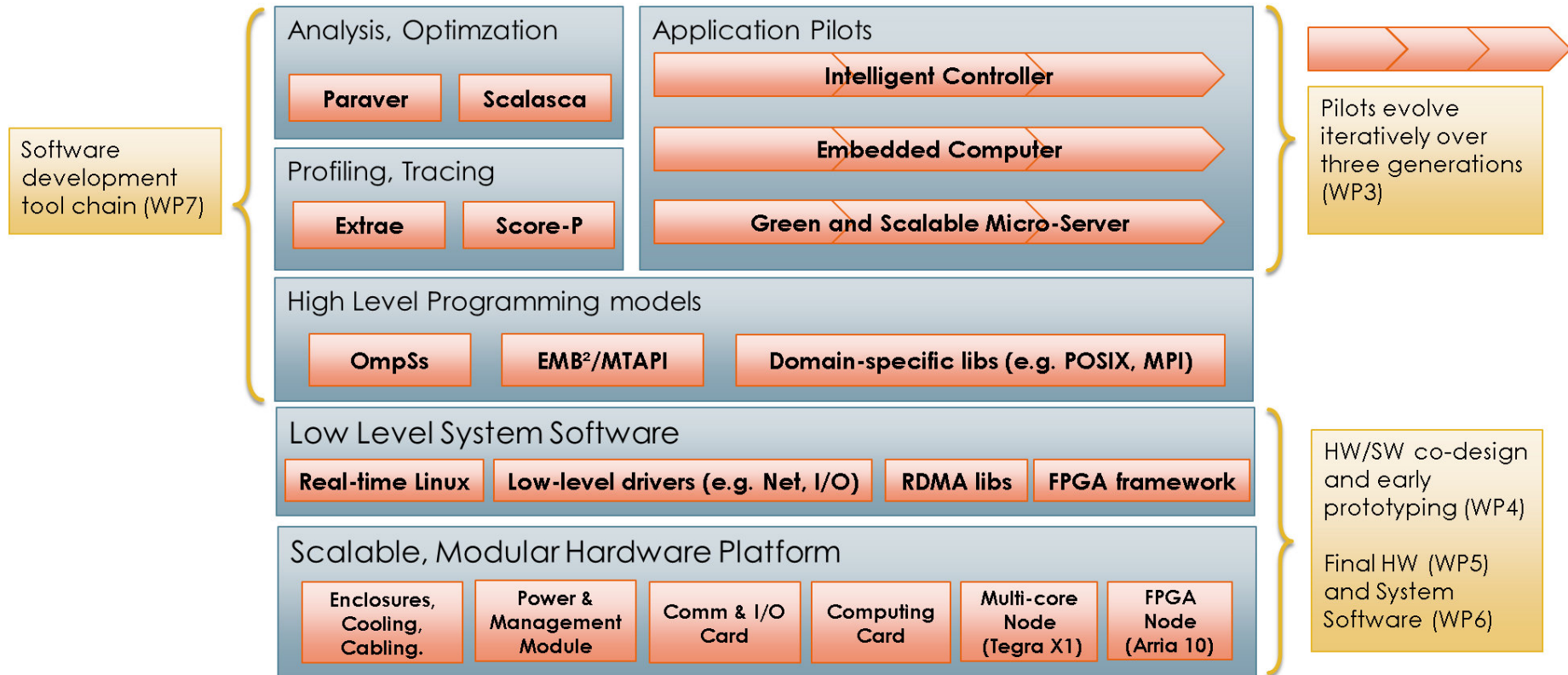
Proposta picoLO

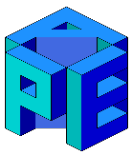
Application pilot che individuano tre diverse tipologie di use case della piattaforma

1. Intelligent CONTROLLER
 - a. Vision-based automatic train operation
 - b. Autonomous longitudinal driving cars
 - c. High Energy Physics low-level trigger.
2. Embedded COMPUTER
 - a. Medical CT image reconstruction
 - b. Automated somatic mutation detection
 - c. Low power computing for X-Ray Tomography to Cultural Heritage diagnostics
3. Green Scalable MICRO-SERVER
 - a. Large scale unsupervised neural network learning (DPSNN-STDP)



Proposta picoLO

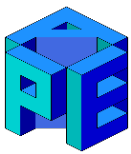




Proposta picoLO

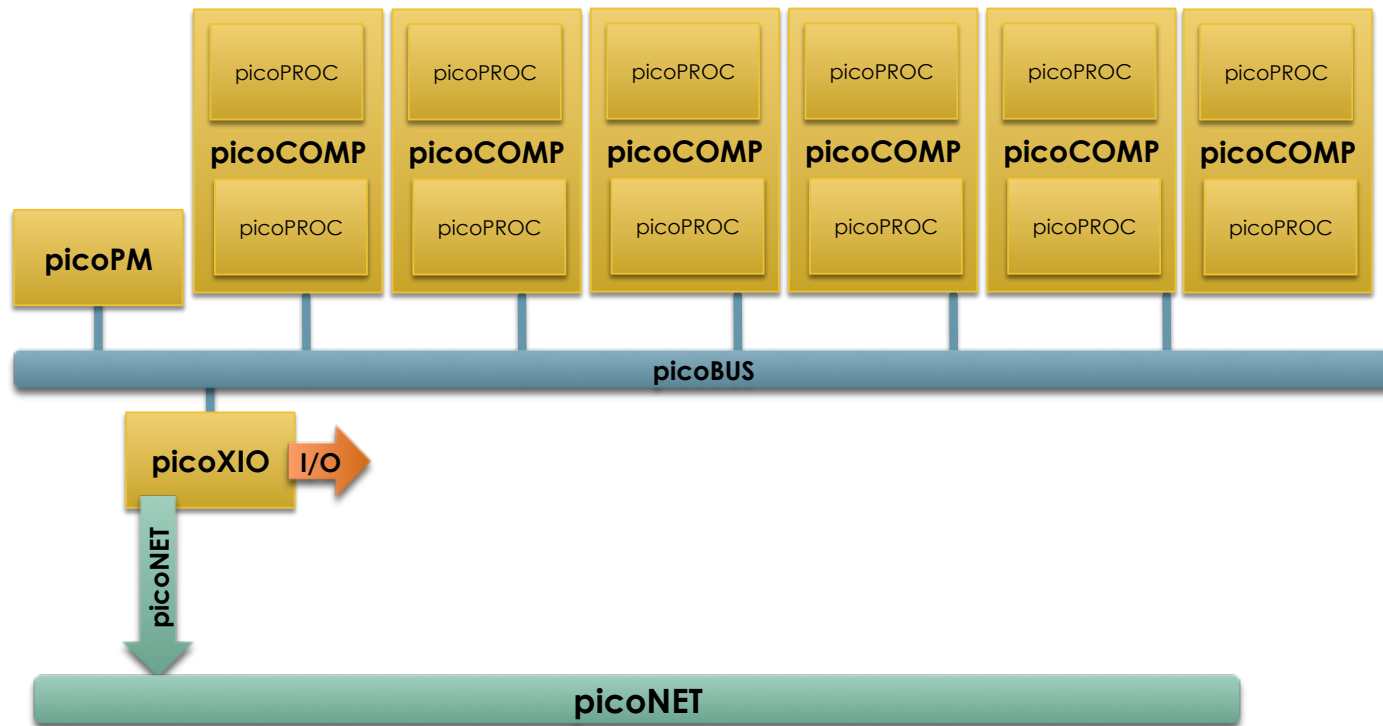
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36			
WP1 Management																																							
T1.1 Internal consortium communication tools																																							
T1.2 Project management definition and quality assurance procedures																																							
T1.3 Project operation management and progress tracking																																							
WP2 Pre-Sales Engineering, Dissemination and Commercialization																																							
T2.1 Dissemination and exploitation strategy																																							
T2.2 Dissemination materials																																							
T2.3 Dissemination towards Research and Data-centre Communities																																							
T2.4 Pre-sale engineering and commercialization																																							
T2.5 User interface for system management and usage																																							
T2.6 In-cloud applications repository																																							
WP3 Applications																																							
T3.1 Demonstrator 1: "Vision-based automatic train operation"																																							
T3.2 Demonstrator 2: "Autonomous longitudinal driving cars"																																							
T3.3 Demonstrator 3: "High Energy Physics lowlevel trigger"																																							
T3.4 Demonstrator 4: "Medical CT image reconstruction"																																							
T3.5 Demonstrator 5: "Automated somatic mutation detection"																																							
T3.6 Demonstrator 6: "Portable CT scanner"																																							
T3.7 Demonstrator 7: "Large scale unsupervised neural network learning"																																							
T3.8 Collection of "lessons learned"																																							
WP4 Hardware/software architecture co-design and early prototyping																																							
T4.1 Collecting of architectural requirements																																							
T4.2 System hardware design																																							
T4.3 Design exploration of picoLO system mechanical integration																																							
T4.4 System software design																																							
T4.5 System early prototyping																																							
WP5 Hardware integration and manufacturing																																							
T5.1 Design, manufacturing, testing of picoPRCC-1																																							
T5.2 Design, manufacturing, testing of picoPRCC-2																																							
T5.3 Design, manufacturing, testing of picoCCMP																																							
T5.4 Design, manufacturing, testing of picoXIO																																							
T5.5 Manufacturing and testing of network expansion card (picoNET)																																							
T5.6 Manufacturing, testing of IO expansion card (picoDAC)																																							
T5.7 Power supply and management/monitoring module																																							
WP6 Software for the picoLO infrastructure																																							
T6.1 Lowlevel communication software and bootstrap																																							
T6.2 Tool infrastructure for the use of FPGA																																							
T6.3 Real-time aspect for multi-core systems																																							
T6.4 Energy saving management																																							
WP7 Programing model and software toolchain																																							
T7.1 port CmpSs to the picoLO platform with SMP support																																							
T7.2 add CmpSs@cluster support																																							
T7.3 Add support for collective offload within CmpSs																																							
T7.4 Add CmpSs support for reshaping techniques in ARM																																							
T7.5 CmpSs + CUDA+ FPGA																																							
T7.6 MIAP1 communication plugin																																							
T7.7 MIAP1 task execution plugins for heterogeneous platforms																																							
T7.8 Dataflow programming on heterogeneous platforms																																							
T7.9 Integration of Tareador tool																																							
T7.10 Profiling and tracing of MIAP1 systems																																							
T7.11 Trace-based performance analysis																																							
T7.12 Tool interoperability																																							
T7.13 Algorithm implementation																																							
T7.14 Parameter extraction																																							
T7.15 Modelling parameter behavior																																							
WP8 System testing and tuning																																							
T8.1 Hardware test																																							
T8.2 System software test																																							
T8.3 picoLO Reference Applications Benchmarking and Tuning																																							

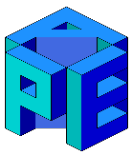




Longer term R&D: COSA & sistemi picoLO-like

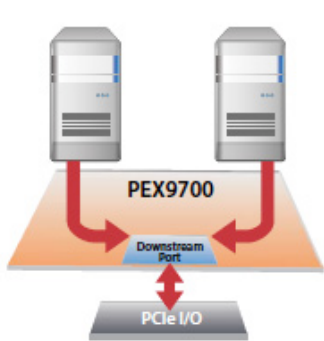
- Proposta LEIT- progetto **picoLO**,
 - sviluppo di sistemi computazionali ad alta' densita', alte prestazioni e basso consumo di potenza orientati alle applicazioni embedded/ industriali *computing demanding*
- Sinergie forti (non solo per la composizione del team...) con sviluppi HW/SW in ambito WP4
 - studio di meccanismi di interconnessione a livello di singola box di calcolo efficienti in termini di banda, latenza e costo



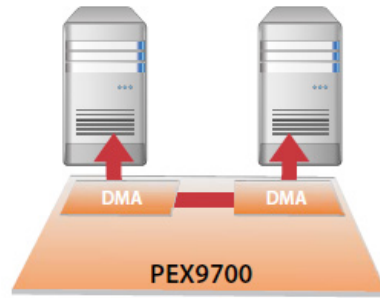


Longer term R&D: COSA & sistemi picoLO-like

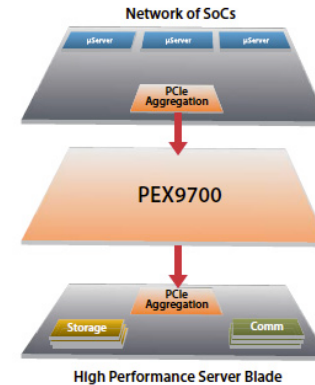
AVAGO/PLX PCI Express Fabric: Multi-Node cluster over PCIe Fabric i.e. no bridging device per multiple CPU network interconnection



Shared I/O



Host-to-host DMA



uServer:
Aggregazione,
isolamento

- ❑ Riteniamo necessario acquisire il testbed PLX PCIe Fabric (PFX55033), a breve sul mercato equipaggiato con switch device **PLX 9797**
 - ❑ 97 lanes, 25 porte PCIe Gen3 switch, numero di lanes per porta configurabile
 - ❑ Porte dedicate per multiple CPU boot e management

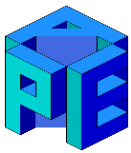
❑ Finanziamento:

- ❑ richiesta di fondi aggiuntivi in CSN5 e/o
- ❑ utilizzando economia di scala derivante dal progetto NaNet di cui Lonardo è responsabile

Line	Device	Manuf.	Qty.	MPQ	Price/Unit	Stock	Leadtime Weeks
1	PXF55033-AA RoHS-compliant* 32-Port PCIe Top-of-Rack Fabric Switch □ Capella 2 ; using PEX9797 - ROHS compliant	AVA	1	0	12.800,0000 EUR	2/3 WKS DRO	
2	PXF51033-AA RoHS-compliant* 2 -Port PCIe Bus Extender (ExpressNIC) with PCIe retimer	AVA	1	0	220,0000 EUR	2/3 WKS DRO	

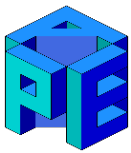
- merce resa f.co nostro magazzino
- legame valutario fisso no cambio
- pagamenti come in uso





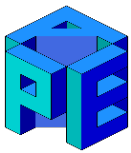
Longer term R&D: COSA & Nvidia TX1

- Come dimostrato dai benchmark preliminari eseguiti (DPSNN et al.) le piattaforme basate Nvidia Tegra K/X offrono alte prestazioni e alti ratio flops/W e flops/\$
- X1 specs (preliminari):
 - GPU Maxwell @1.8+ GHz, 1 TFLOPS (fp16) 512 GFLOPS (fp32)
 - ARMv8 4 ARM Cortex-A57 + 4 ARM Cortex-A53 Octa-Core (64-bit)
 - PCIe Gen2 5 lanes, 2 porte indipendenti (x4,x1)
- Anche in ambito esplorazione dell'interconnessione intra-system basata su PCIe ExpressFabric ed interazione tra nodi computazionali Tegra e network toroidale FPGA-based riteniamo opportuno acquisire qualche piattaforma di sviluppo **Nvidia Tegra X1**
 - Disponibilita' prevista Q415
 - Costo del sistema in linea (forse leggermente maggiore) con il costo del Jetson TK1 (N*100\$)
- Una prima stima ~1.0-1.5KE per 4 sistemi



Conclusioni

- Nell'ambito del WP4 (sviluppo di reti custom FPGA-based) ed in sinergia con il nuovo progetto FET-HPC "ExaNeSt" (starting date 1/12/2015) prevediamo di utilizzare gli ARM cores embedded per realizzare architetture di calcolo integrate rete+CPU
 - Accelerazione di task computazionali RDMA-related per ottimizzazione prestazioni
 - Computing ARM based
- L'acquisizione dei sistemi di sviluppo FPGA-based necessari a tali sviluppi (Arria10 -> Stratix10) ritardata di alcuni mesi (4Q15)
- Nel frattempo procede lo sviluppo della nuova release di APEnet+ V5 con risultati in linea con quanto aspettato e prospettive interessanti
- A medio/lungo termine pensiamo di poter esplorare le idee architettoniche correlate al nuovo proposal LEIT picoLO cominciando ad acquisire i componenti necessari
 - PCI Express Fabric
 - Nvidia Tegra X1
- Servono risorse economiche aggiuntive (~10-15 KE) che dobbiamo capire come reperire



THANK YOU!!



Roberto
Ammendola



Andrea
Biagioni



Ottorino
Frezza



Francesca
Lo Cicero



Alessandro
Lonardo



Michele
Martinelli



Pier Stanislaio
Paolucci



Elena
Pastorelli



Davide
Rossetti



Francesco
Simula



Laura
Tosoratto



Piero
Vicini

This work was partially supported by EU Framework Programme 7 EURETILE project, grant number 247846; Roberto Ammendola and Michele Martinelli were supported by MIUR (Italy) through INFN SUMA project.

