COSA Project

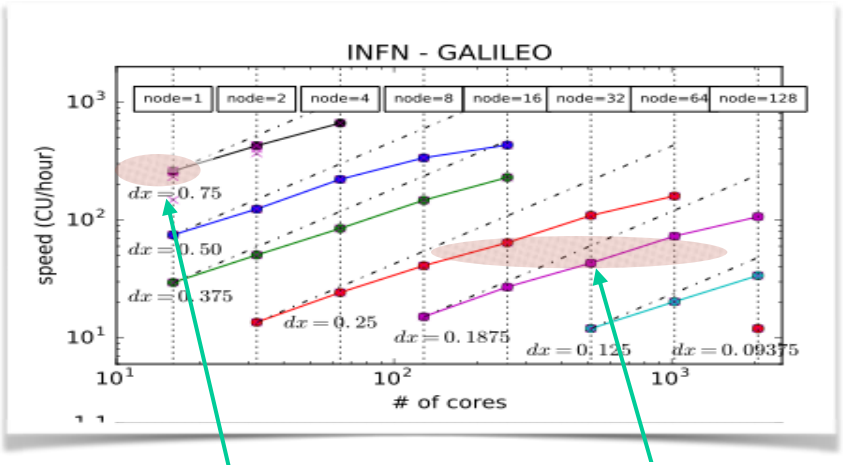
Attività a Parma

Roberto Alfieri

3 Novembre 1016, CNAF

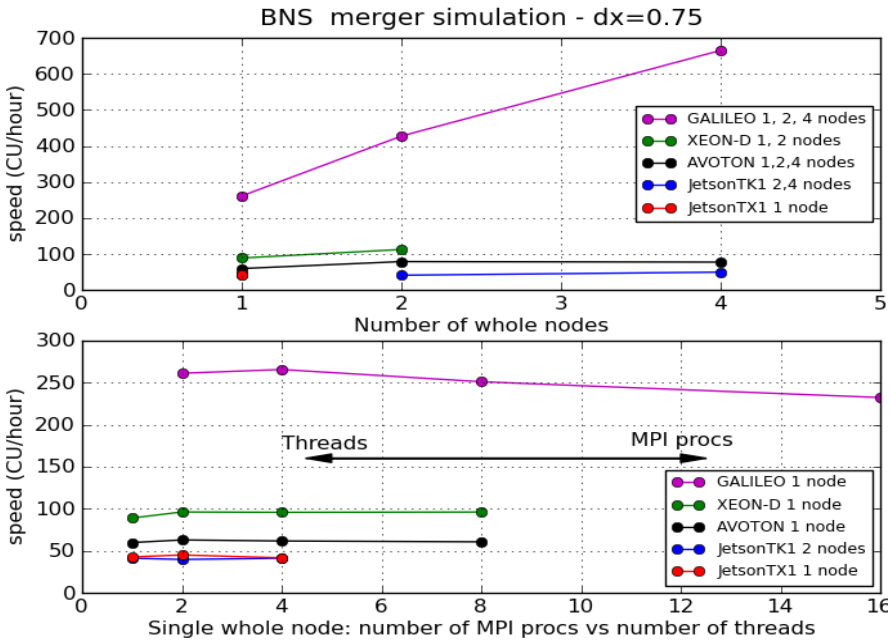
Einstein ToolKit on low-power systems

Scientific problem: High resolution simulation of inspiral and merger phase of binary neutron stars system (most likely source of gravitational waves)
Computational challenge: Cartesian grid with at-least 6 refinement levels.
Standard resolution in the finest grid 0.25 CU and up to 0.125 CU.
MPI+openMP code parallelization already in place using the http://einsteintoolkit.org
Very complex code (more than 100 developers over 20 years using F90,F77,C++,C)
Computational cost: Typical run (dx=0.25) requires al least 108 GB ram. Coarser resolution (dx=0.75) requires 4 GB. 50 ms simulated in a week on HPC systems.
Test environment: COSA low power systems (Xeond, Avoton, Jetson-TK1, Jetson-TX1)
Performance compared with GALILEO (CINECA HPC production system)



Test environment on COSA systems

Production environment on GALILEO



Valutazione performance e problematiche di porting su JETSON TX1 : progetto desMPI

Obiettivo: implementazione su diversi ambienti di calcolo della violazione di DES (2^55 chiavi) per confrontarne le prestazioni.

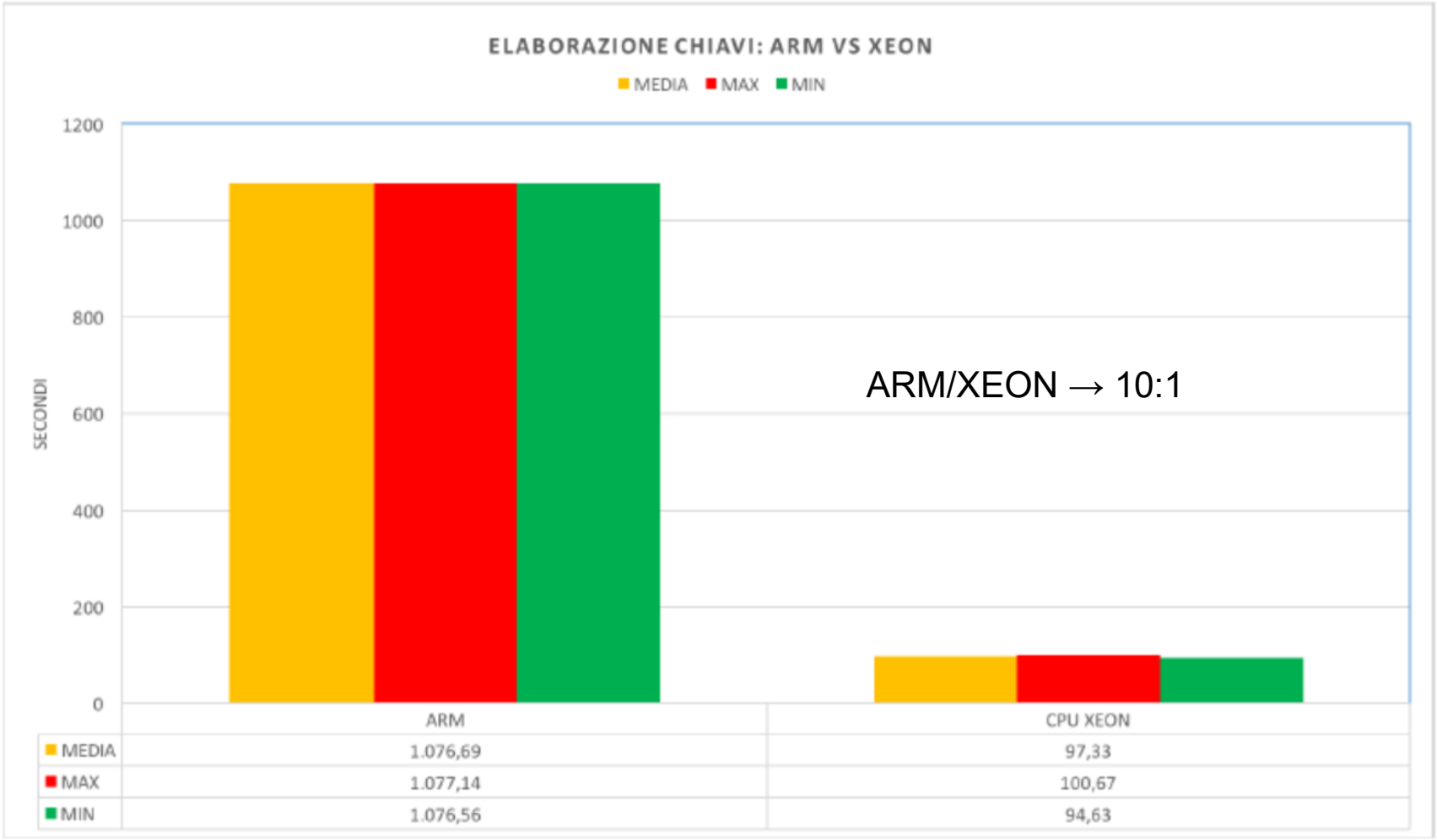
HW:
Nodo 4K40
- CPU XEON E5-2603 v3 1.6GHz, 12 cores
- GPU TESLA K40   (2280 core, 12 GB GDDR5, 288 GB/s, 1.4 Tflops dp, 4.3 sp)

Nodo JETSON TX1
- CPU ARM 64 bit quad A57
- GPU MAXWELL   (256 cores, 4GB LPDDR4, 25GB/s,   0.5 Tflops sp, 1 fp16)

Strumenti SW:
openMPI : architettura manager worker per distribuzione di blocchi di chiavi
openMP : worker su CPU
CUDA-C: worker su GPU

Progetto realizzato in collaborazione con il corso di Crittografia e sviluppato da 4 studenti del corso di laurea in Informatica a UniPR

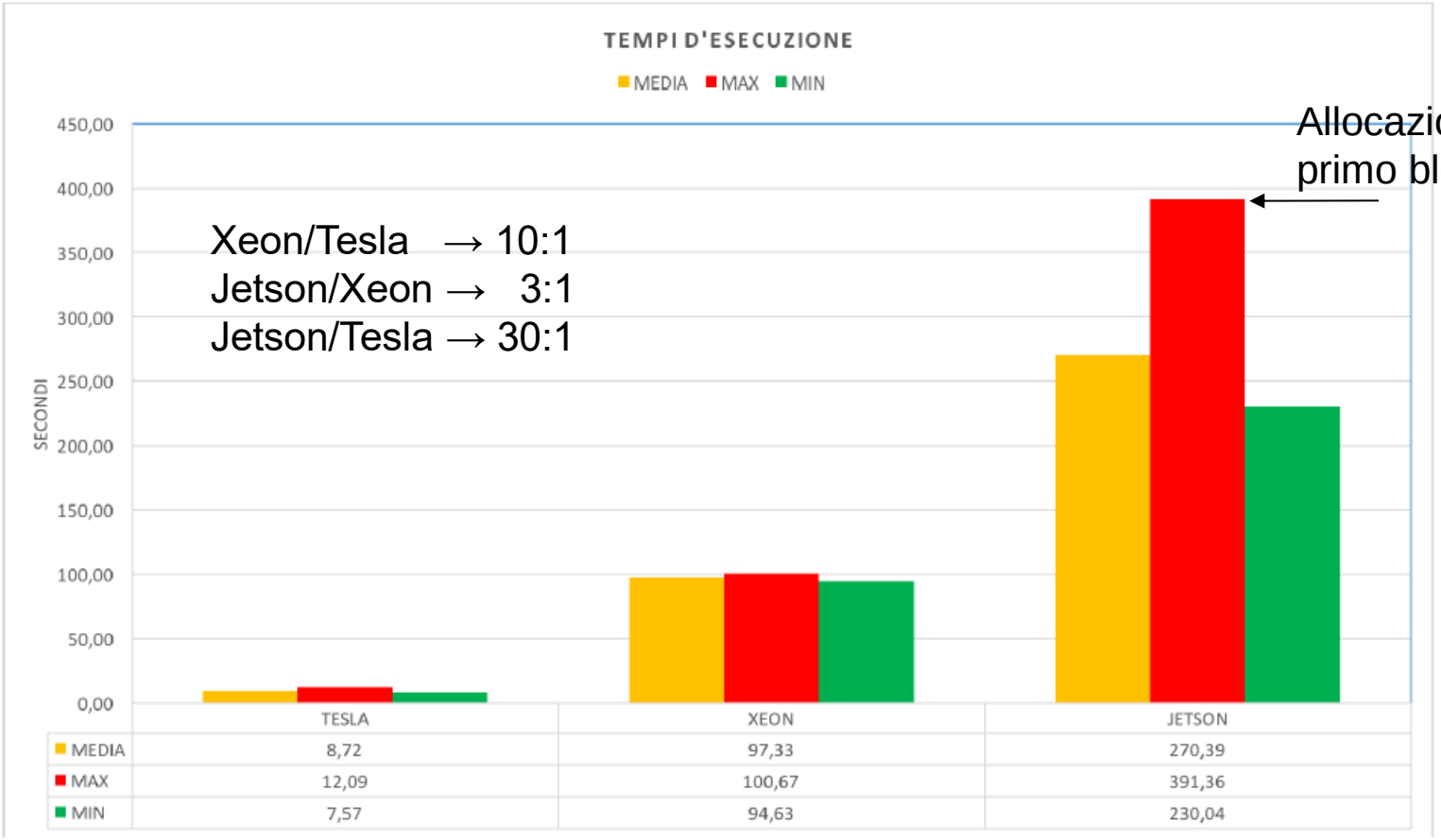
Confronto CPU



	ARM	CPU XEON
MEDIA	1.076,69	97,33
MAX	1.077,14	100,67
MIN	1.076,56	94,63

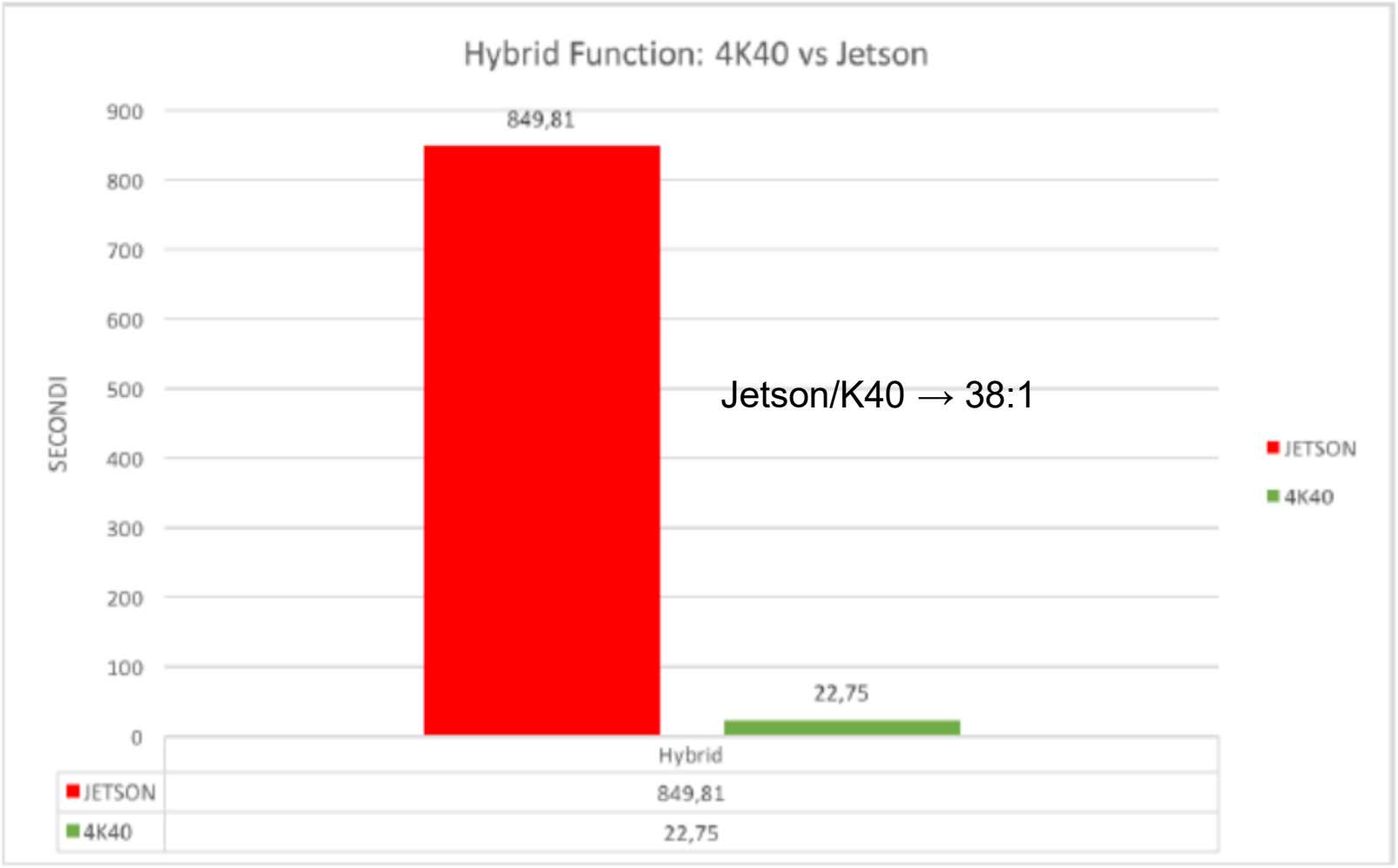
Tempo impiegato per determinare la chiave 512 (1000/100)

Confronto
GPU Tesla K40 / XEON / GPU JETSON



	TESLA	XEON	JETSON
MEDIA	8,72	97,33	270,39
MAX	12,09	100,67	391,36
MIN	7,57	94,63	230,04

Esecuzione ibrida CPU + GPU



Problematiche di porting su TEGRA TX1

4 GB di ram condivisi tra CPU e GPU, 25 GB/s.

Diversa allocazione della memoria, basata su **Zero Copy Memory**

La memoria del device viene semplicemente copiata in una nuova allocazione della stessa memoria fisica permettendone l'accesso tramite puntatore, questo **velocizza tutte le procedure di allocazione**, **deallocazione e copia** della memoria tra device e host, ma richiede la **modifica del codice CUDA**:

- aggiungere nuove dichiarazioni di variabili, allocandole sull'host tramite la **cudaHostAlloc()**
- Passare il puntatore al device con **cudaHostGetDevicePointer()**

Nuovo servizio di Calcolo
UniPR / INFN Parma

Comunità di riferimento

Fisica Teorica / INFN: QCD su reticolo, fisica gravitazionale , Einstein Toolkit.
Fisica della materia e Biofisica, Chimica strutturale: studio sistemi molecolari, Quantum Espresso, Gromacs, Gaussian.
Bioscienze: genomica, applicazioni memory intensive, big data (progetto Elixir)
Farmacia: modellizzazione dei farmaci, GPU
Scienze degli alimenti: dinamica molecolare, collaborazione con Efsa.
Ingegneria Civile: simulazione di fenomeni di allagamento, GPU.
Ingegneria dell'informazione: Simulazioni di Reti e Sistemi Distribuiti, Simulazioni con solutori modali FEM, simulazioni numeriche per il calcolo di probabilità d'errore ed efficienza spettrale per trasmissioni satellitari
Ingegneria Meccanica: dinamica multi-body, GPU.
Medicina: Approcci terapeutici personalizzati integrando molte sorgenti esterne con capacità di analisi semantica, Big Data.

Survey in giugno 2015 -> Nomina di un comitato scientifico

Cluster in fase di acquisizione (UniPR)

XEON E5-2683 (2.1GHz, 16 cores)
1 Tflops dp (2*16*2,1*16 (avx2)), 120w, 8,3 GFlops/w
8 nodi dual XEON E5-2683 , 128 GB, 8 Tflops dp
1 nodo dual XEON E5-2683 , 1 TB, 1 Tflops dp

GPU Pascal P100 (3584 Cuda cores)
4,7 Tflops dp, 250w, 18 GFlops/w
2 nodi dual XEON E5-2683 con 4 P100, 40 TFlops dp

XEON PHI 7250 (1.4 GHz, 68 cores)
3 Tflops dp (68* 1.4 * 32 (avx512)), 215w, 14 GFlops/w
4 nodi PHI 7250, 96 GB, 12 TFlops dp

Rete OmniPath 100Gbps

Installazione prevista per gennaio 2017

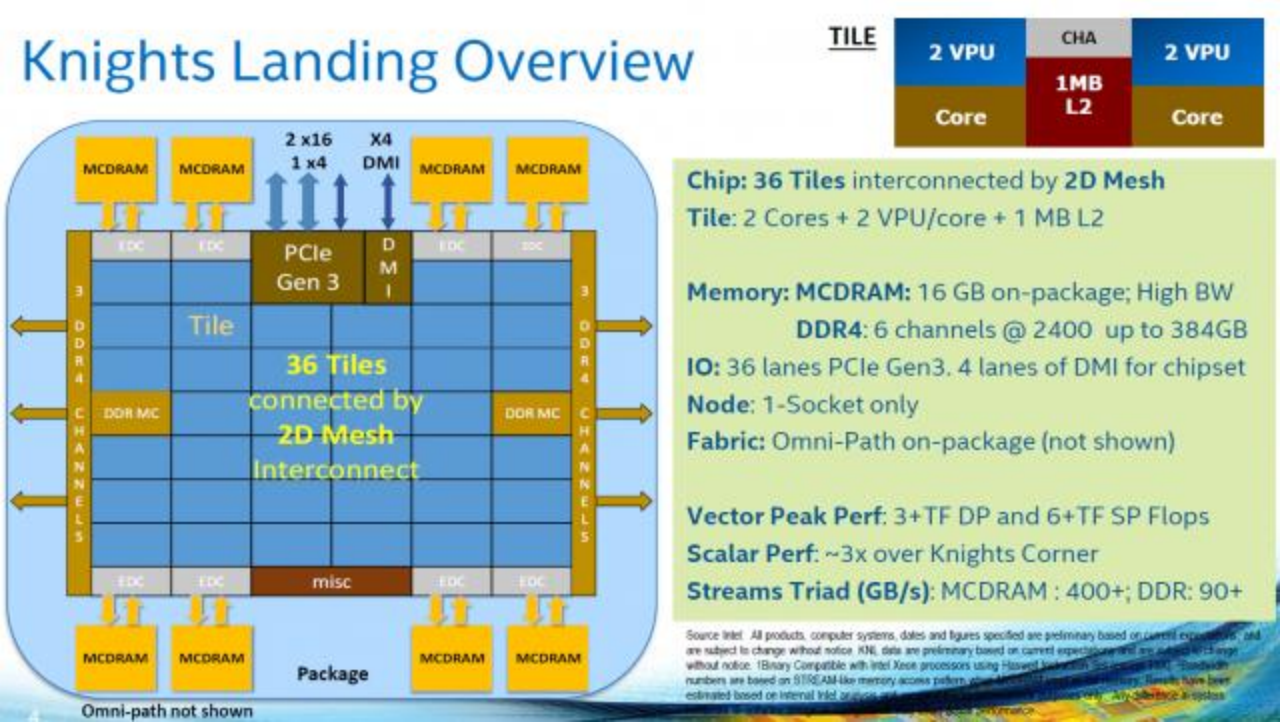
NVIDIA P100

	Tflops	**Mem Bandwidth**
Maxwell Jetson TX1	0.5 sp	25 GB/s
Tesla K80	5.6 sp, 1.8 dp	240/480 GB/s
Pascal P100 PCI	9.3 sp, 4.7 dp	549/732 GB/s
Pascal P100 NVlink	10.6 sp, 5.3 dp	732 GB/s

Knights Landing

4 nodi XEON PHI 7250 , 96 GB DDR4
4x (68* 1.4 * 32 (avx512)) = 4x 3.046 = 12 Tflops dp peak
16GB MCDRAM 500GB/s, max 384 GB DDR4 100GB/s

CINECA: 3600 nodi x 3.046 = 11 Pflops



Mod.	GHz	cores	Opath
7210	1.3	64	
7210F	1.3	64	si
7250	1.4	68	
7250F	1.4	68	si
7290	1.5	72	
7290F	1.5	72	si

Knights Landing

Improvements	What/Why
Self Boot Processor	No PCIe bottleneck
Binary Compatibility with Xeon	Runs all legacy software. No recompilation.
New Core: SLM based	~3x higher ST performance over KNC
Improved Vector density	3+ TFLOPS (DP) peak per chip
AVX 512 ISA	New 512-bit Vector ISA with Masks
Scatter/Gather Engine	Hardware support for gather and scatter
New memory technology: MCDRAM + DDR	Large High Bandwidth Memory → MCDRAM Huge bulk memory → DDR
New on-die interconnect: Mesh	High BW connection between cores and memory

Alcune tematiche

AVX-512 Tflops 32x verifica efficienza e usabilità
(autovettorizzazione vs openMP vs intrinsics)

Compilatore Intel vs Gnu

MCDRAM modelli di utilizzo: flat, cache e ibrida

MESH throughput (tra 2 core e aggregato)

OmniPath Performance OPA (100Gbps peak) , PHI 7250 vs 7250F

….