



Sviluppi tecnologici short term: Apenet+ su A(E)urora

P. Vicini – INFN Roma

Sviluppo tecnologico da ricerca di base in un ambiente commerciale.

Opportunita' per:

- "forzare" una architettura di calcolo standard verso un sistema dedicato ottimizzato per calcoli di fisica teorica (i.e. 3D torus & GPU)
- attivita' di trasferimento tecnologico
- ottenere co-finanziamenti per le attivita' di ricerca tecnologica (will see...)

I sistemi installati sono un testbed ottimale per architettura APEnet+

- Test di deadlock-livelock legati alla size del sistema
 - $n \cdot 10^{2-3}$ nodi Eurotech vs ~ 10 nodi QUonG
- Risultati dei benchmarks confrontabili con i "competitors"...
- I sistemi nD toroidali sono naturalmente piu' "efficienti" per scale di rete medio-grandi

Alcune ipotesi...

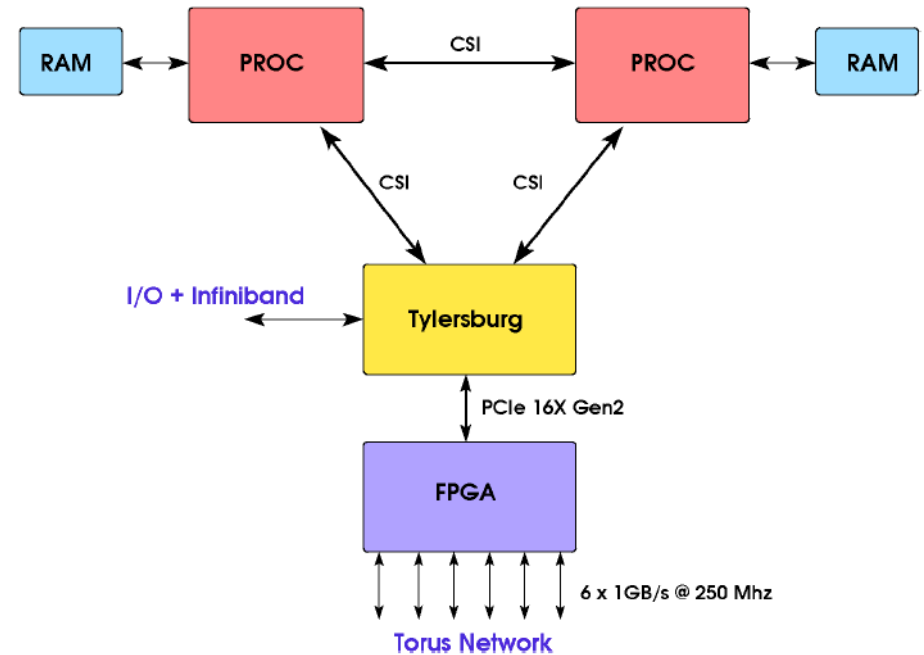
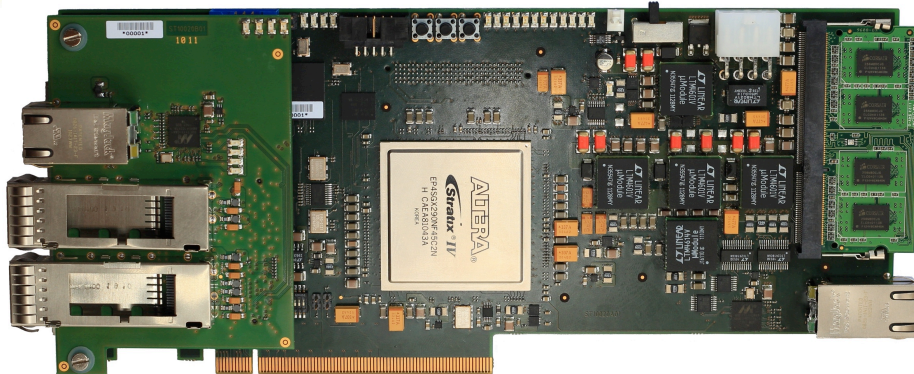
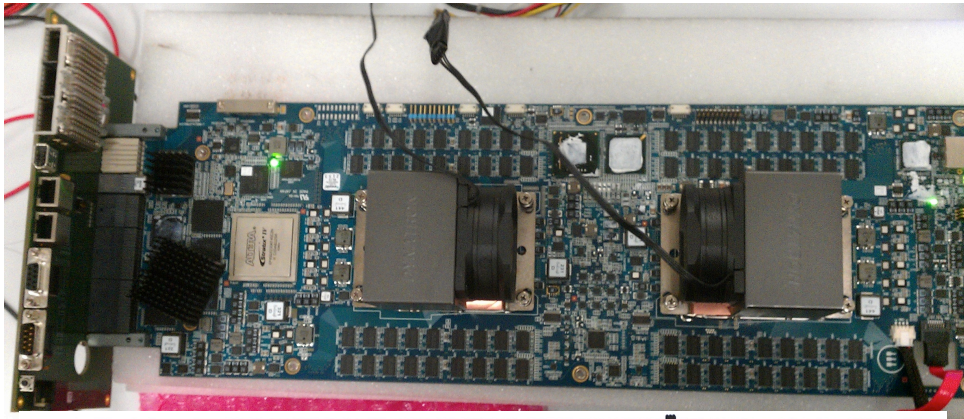
- Interesse nella collaborazione con Ente di Ricerca di eccellenza
- Sviluppo della prima piattaforma di supercalcolo Europea con tecnologia di network scalabile all'Exascale (10^5 - 10^6 nodi)
- Acquisizione di una tecnologia state-of-the-art per integrazione network-GPU (APENet+ peer-to-peer)

... piu' prosaicamente

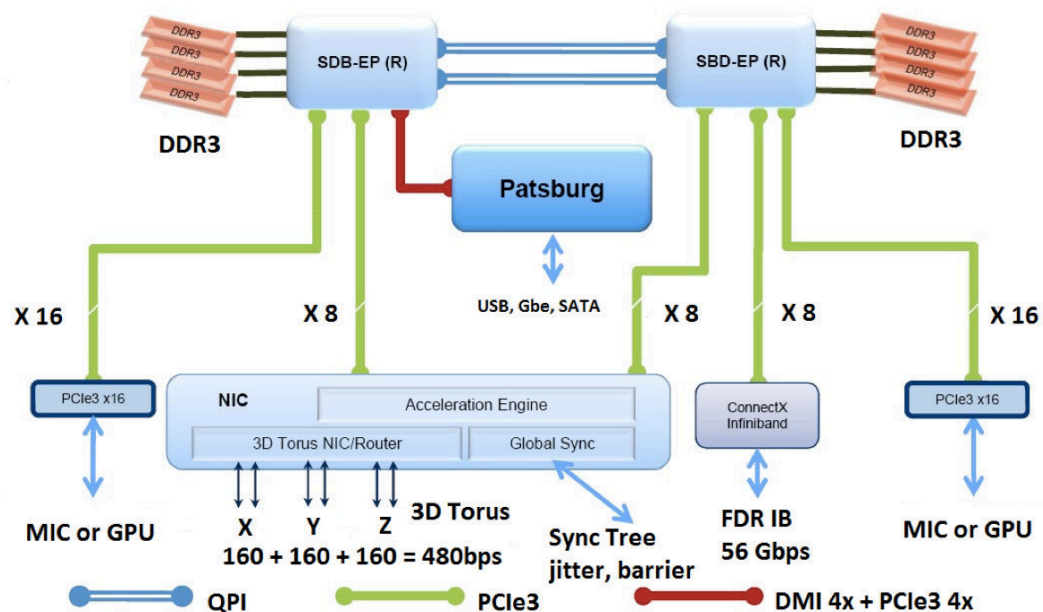
- A(E)URORA enhancements con basso rapporto costi/benefici (...)
- Realizzazione della piattaforma prototipale finale di SUMA
- Realizzazione della piattaforma di sviluppo per sistemi accelerati riconfigurabili come da nuova proposta di progetto in ambito EU

E' tecnologicamente possibile?

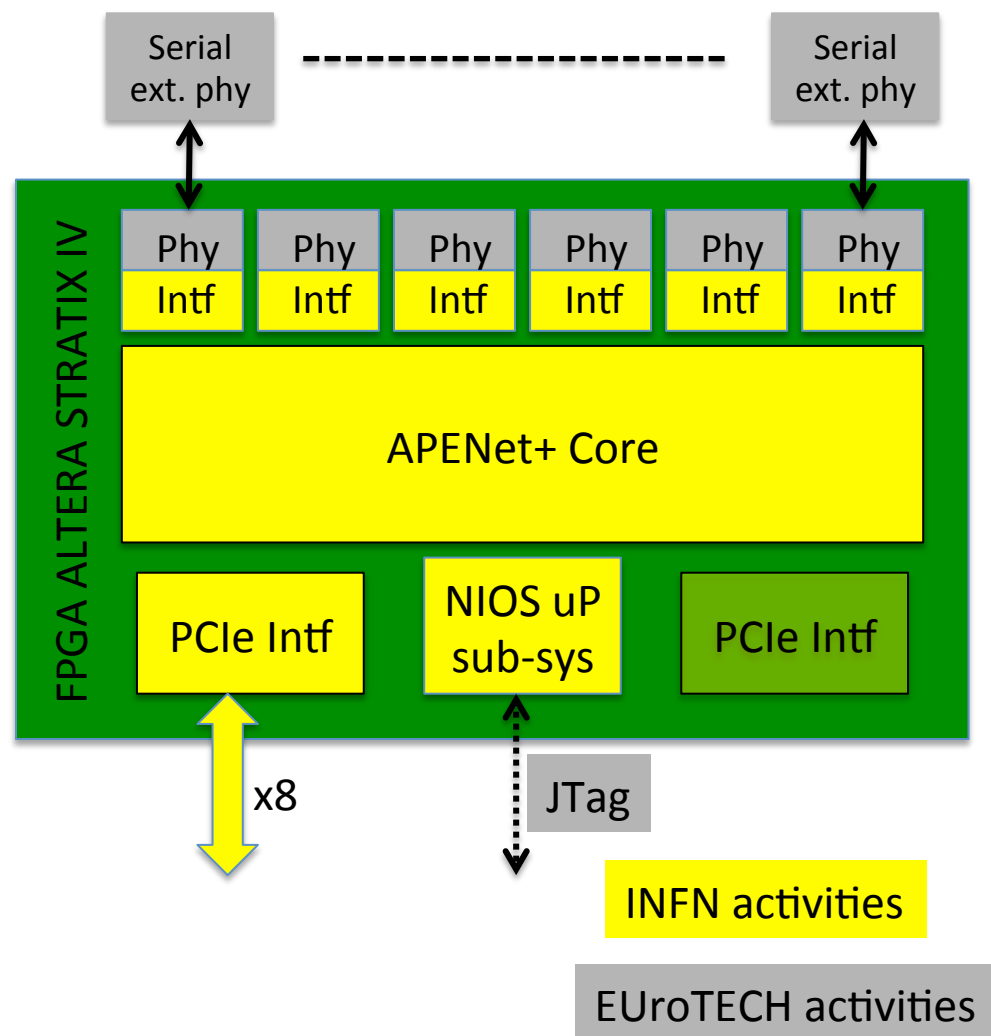
- Componentistica ed interfaccia di I/O simili ad APENet+
 - FPGA ALTERA Stratix IV (stessa famiglia)
 - Host I/O PCI Express x8 Gen2 (stessa tipologia)
 - 3D Torus links meno veloci (10 Gbps vs 34 Gbps) e implementati con PHY esterni su backplane i.e. gestione semplificata



- Sandy Bridge on board....
- FPGA 28nm Stratix V on-board
- Doppia interfaccia PCIe x8
 - Dobbiamo imparare ad usarla ma sulla carta raddoppia la banda
- Integrazione nei sistemi EURORA di GPU NVidia K20
 - Beneficio congiunto grazie al nostro hardware controller del peer-to-peer
- Utilizzo di link embedded per realizzazione dei canali seriali



- L'attività e' appena partita (settembre 2012)
- Scrittura del MOU in progress
- A Roma abbiamo una piattaforma prototipale AURORA dual-proc Sandy Bridge
- INFN focus:
 - Interfaccia PCIe
 - NIOS sub system
 - APENet+ Core
- Eurotech focus:
 - JTAG
 - External Phy Interface
- Sviluppo congiunti di software di sistema



- 1) Porting dell'interfaccia PCIe da APEnet+ ad Aurora
 - HW ok, SW driver ok per entrambi i canali. Test di lettura/scrittura registri e dati
 - Stiamo indagando sulla possibilità di integrare entrambi i canali PCIe nello stesso design (x2 data rate)
- 2) Test PHY esterni: ok in System Diagnostic Loopback ma problemi (sotto analisi) nella trasmissione dati verso FPGA. Competenza Eurotech
- 3) NIOS sub-system integrato ok. I/O errors quando data rate cresce. Ipotesi di jtag channel rumoroso (?)
- 4) Nuovi sviluppi (necessari data l'architettura)
 - l'interfaccia verso i PHY esterni che potrebbe necessitare di nuovi registri e di una procedura di reset/allineamento diversa
 - NIOS I/O su PCIe (no jtag) con nuovi registri e relativo supporto nel driver

Tempi e risultati dipendono fortemente dall'impegno Eurotech

- Serve accesso completo alla documentazione (sotto NDA...)
- Serve almeno un FTE integrato in INFN per porting e codifica dei blocchi hw mancanti e sviluppo del firmware aggiuntivo.
- Serve qualche FTE in Eurotech per test e supporto al debug e integrazione PHY esterni con core APENet
- Sviluppo API RDMA ottimizzate e MPI programming model

In questa ipotesi di "massimo" supporto:

- HW Coding completato: Giugno 2013
- APENet integrato con supporto API RDMA native: Settembre 2013
- APENet MPI up and running: 4Q13

Complessita' dipende da:

- nuova tecnologia FPGA (Stratix V)
- utilizzo di link seriali embedded
- PCI Gen3 x8
- Integrazione hw ottimizzato per accesso a bassa latenza per le nuove architetture GPU Kepler

Esperimenti di porting su EURORA a partire dall'inizio del 2013 sfruttando la sinergia con le attività APE/SUMA

- APENet+ upgrade a StratixV
- Incremento del data rate sui canali 3D di APENet++
- Integrazione con Kepler (e oltre)

No release date (ad oggi).

Probabilmente "primi vagiti" (...) entro la fine del 2013