

HPC 2006: luci e ombre nel futuro del supercalcolo

Claudio Arlandini

CILEA, Segrate

Abstract

L'Advanced Research Workshop on High Performance Computing Technology and Applications (HPC2006) è uno degli appuntamenti tradizionali in cui si discutono gli sviluppi correnti e futuri nel campo dell'architettura dei processori e delle tecnologie software di interesse per il calcolo ad alte prestazioni, nonché le più promettenti applicazioni a nuovi problemi scientifici o dell'industria. Il CILEA è orgoglioso sponsor dell'edizione 2006. Presentiamo qui una selezione degli interventi più interessanti del convegno.

The Advanced Research Workshop on High Performance Computing Technology and Applications (HPC2006) is a traditional location to discuss current status and future developments in microprocessor architectures and software development for High Performance Computing. CILEA is proud sponsor of 2006 edition. The article presents a selection of most relevant talks.

Keywords: Supercalcolo, Conferenze, Processori.

Il workshop [1] (Fig. 1) ha una storia ormai più che decennale e riunisce nella bella cornice del Grand Hotel San Michele di Cetraro (CS) il gotha dei produttori di hardware e software e i più importanti ricercatori internazionali del settore.

Nel 2006, grazie anche alla collocazione temporale strategica, subito dopo l'immane convegno Supercomputing tenuto a Dresda, è stata mantenuta un'alta affluenza di pubblico e soprattutto un'elevata qualità degli interventi e delle discussioni. Il convegno si è svolto sotto l'affabile e attenta regia del Prof. Lucio Grandinetti, dell'Università della Calabria, principale organizzatore dell'evento. Il workshop si è tenuto nel periodo dal 3 al 6 luglio e ha visto, tra gli sponsor, oltre al CILEA, enti del calibro dell'ENEA e del Global Grid Forum, oltre a multinazionali quali Microsoft, Hewlett Packard, IBM e SUN Microsystems.

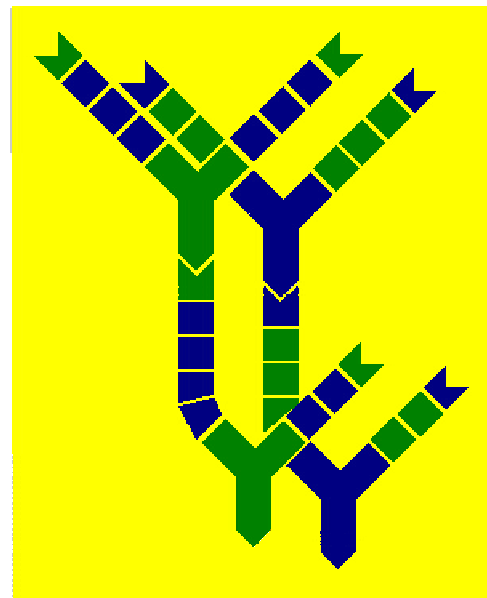


Fig. 1 – Logo del Convegno

I temi

Il calcolo ad alte prestazioni (HPC) è uno dei temi dell'informatica che maggiormente attrae l'attenzione mondiale come strumento fondamentale per la ricerca scientifica, la produzione industriale e il "decision making". Enti pubblici e privati investono significative quantità di energia e risorse finanziarie per costruire infrastrutture adeguate e formare uomini capaci di trarre il massimo da esse.

Le architetture hardware e software per l'HPC sono in costante evoluzione ed emergono varie applicazioni a problemi reali; tuttavia molti potenziali beneficiari non ne sono ancora del tutto consapevoli.

Lo scopo del Workshop, nelle intenzioni del comitato organizzatore, è di contribuire a fare il punto sull'attualità e gli sviluppi futuri delle tecnologie per l'HPC e di discutere promettenti applicazioni a nuovi problemi di scienza, industria e finanza.

Il primo giorno ha visto due sessioni, la prima sullo "Stato dell'arte e argomenti scottanti in HPC", in cui accademici del calibro di Jack Dongarra, dell'Oak Ridge National Laboratory, e Ian Foster, dell'Università di Chicago, (ritenuto insieme a Carl Kesselman "inventore" del Grid Computing) hanno presentato le loro visioni e i loro attuali progetti. La seconda sessione, "Sistemi informatici e soluzioni emergenti", è stata in realtà una vetrina pubblicitaria per le nuove architetture hardware e le tecnologie di rete che stanno per essere lanciate sul mercato.

I giorni successivi sono stati monopolizzati da interventi sul Grid Computing, a testimonianza dell'interesse e della febbrile attività della comunità internazionale sulle tecnologie di griglia computazionale. L'autore di questo articolo ha presentato un intervento sullo stato e sui i futuri sviluppi del progetto SEPAC [2].

I momenti più significativi della conferenza

Jack Dongarra [3], guru di MPI e uomo chiave della classifica TOP500 [4], ha aperto il congresso con una significativa panoramica sullo stato dell'arte del supercalcolo, evidenziando quali saranno, a suo dire, i protagonisti del futuro prossimo.

Significativamente, il suo intervento si intitolava "Le sfide dei processori multicore e degli acceleratori specializzati". Negli ultimi anni la potenza dei processori è aumentata incrementando la densità dei gate logici nel

chip e riducendo il tempo di ciclo. Stiamo ora raggiungendo una barriera che rende questa strada non più ulteriormente percorribile (Fig. 2): essendoci una relazione cubica tra frequenza e potenza rilasciata, il calore generato da un processore sta raggiungendo livelli non più gestibili, con densità di W/cm² prossime a quelle del nocciolo di un reattore nucleare.

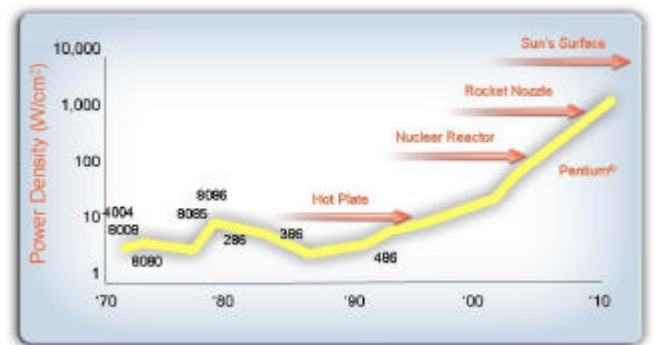


Fig. 2 – Trend della densità di potenza erogata dai processori Intel

La soluzione è di continuare il processo di miniaturizzazione, ma senza ricorrere a ulteriori aumenti vertiginosi della velocità di clock. Si pensa invece di raggruppare più core di calcolo sulla stessa base, aumentando cioè il livello di parallelismo dei sistemi. In questo modo, Dongarra prevede che la legge di Moore verrà rispettata almeno per i prossimi 5 anni, al termine dei quali saranno probabilmente sul mercato processori con 128 core per chip, capaci di processare 512 processi logici per ogni ciclo di clock.

Lavorare con simili processori comporta una serie di sfide tecnologiche non indifferenti. In particolare vi è la necessità di avere un adeguato accesso alla memoria, sia in termini di velocità che di latenza, così come dovranno essere messi a punto adeguati meccanismi per un coordinamento interprocessore, per esempio in termini di sincronizzazione e *context switching*. Soprattutto sarà necessario un effettivo sfruttamento del parallelismo a livello dell'applicazione, con l'introduzione di paradigmi del tutto nuovi. Tutto il software, e in particolare le librerie matematiche, dovrà essere ripensato in una luce completamente nuova.

Contemporaneamente, si sta pensando ai potenziali guadagni che si avrebbero dallo sfruttamento di dispositivi ad alta potenza, ma di più bassa precisione rispetto ai normali processori utilizzati per il calcolo ad alte

prestazioni, come possono essere le Graphical Processing Unit (GPU) delle moderne schede video o alcuni processori specializzati. Un esempio tipico è il processore IBM Cell, di cui è dotata la Sony PlayStation 3 (Fig. 3) [5].

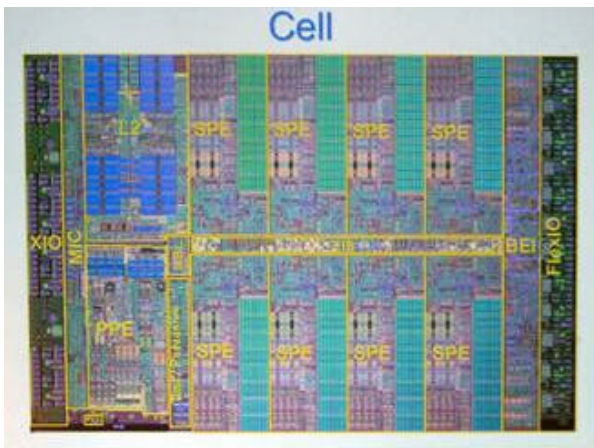


Fig. 3 – Il processore IBM Cell utilizzato dalla PlayStation 3

Ogni processore Cell è composto di 8 Arithmetic Processing Unit (APU), ovvero di 8 processori vettoriali autonomi e indipendenti, ognuno dei quali possiede quattro Floating Point Unit (FPU) da 8 Gflop/s ciascuna. Il risultato è che il processore ha una potenza di picco di ben 256 Gflop/s, per calcoli a virgola mobile a 32 bit, ma a 64 bit, che è quello che interessa al calcolo scientifico, la potenza scende drasticamente a 25 Gflop/s. Analogamente, una delle più performanti schede video oggi in circolazione, la ATI X1900XTX [6], ha una potenza di ben 400 Gflop/s per calcoli a 32 bit, mentre per i calcoli 64 bit è necessaria una costosa (in termini di prestazioni) emulazione software. Dongarra si chiede se sia comunque possibile sfruttare tale potenza per il calcolo scientifico e la risposta è affermativa, se si sanno sfruttare abilmente certe proprietà degli algoritmi, usando un sistema misto, in cui si lavora a 32 bit quanto più possibile, per poi raffinare la soluzione con un'opportuna correzione ottenuta calcolando a 64 bit. Come esempio ha mostrato che la soluzione di sistemi lineari densi è scomponibile in un processo iterativo di complessità $O(n^3)$ a bassa precisione e uno di complessità $O(n^2)$ ad alta. Quando questo algoritmo si implementa su un processore Pentium che, usando il set di istruzioni SSE2 è capace di quattro operazioni a virgola mobile in precisione singola per ciclo e

due in precisione doppia, si ottiene il raddoppio delle prestazioni rispetto all'utilizzo dell'algoritmo standard. Addirittura, per calcoli in quadrupla precisione su processori Xeon si ottengono aumenti di fattori da 10 a 100, quando si utilizza lo stesso algoritmo rispetto all'uso di librerie ottimizzate BLAS, considerate finora l'implementazione di riferimento per il calcolo in quadrupla precisione.

Tra le novità più interessanti presentate al convegno dalle grandi case produttrici due riguardano i processori e una le interconnessioni di rete.

Intel sta cominciando a far trapelare dati sulla nuova generazione del processore Xeon [7] (nome in codice Woodcrest), che sta per uscire sul mercato. Sarà un processore *multicore*, ma completamente rinnovato rispetto alla generazione precedente, tanto da essere capace di quattro operazioni FP per ciclo di clock (il doppio), e capace di prestazioni per *core* che, a seconda delle applicazioni, sono dal 20 al 90% migliori, pur con una frequenza operativa più bassa. Questo ridarà probabilmente a Intel una posizione di leadership, sottrattagli negli ultimi tempi da AMD, grazie ai suoi processori Opteron.

Intel e HP hanno anche annunciato l'uscita della nuova generazione di processori Itanium [8] (nome in codice Montecito), dualcore, con 1 MB di cache di primo livello e ben 12 MB di secondo livello, per *core*. Un processore quindi che, seppur penalizzato da un certo ritardo nel suo sviluppo, va a competere direttamente con i processori IBM Power.

Per quanto riguarda le tecnologie di interconnessione, la Myricom [9] ha annunciato la lungamente attesa nuova generazione delle schede Myrinet, Myri-10G, capace di 10 Gbit/s, con latenze dell'ordine di 2.3 μ s. Per quanto tecnologie concorrenti, come Infiniband, abbiano raggiunto da tempo prestazioni simili, la novità sostanziale è che le schede supportano sia i protocolli proprietari Myrinet che Ethernet a livello Data Link. Questo significa due cose. Innanzitutto, gli switch Myricom manterranno le loro caratteristiche di efficienza e scalabilità, ma potranno avere porte miste 10-Gigabit Ethernet e 10-Gigabit Myrinet. Secondo, sarà possibile interfacciare cluster di nodi con schede 10-Gigabit Myrinet a switch 10-Gigabit Ethernet con una minima perdita di prestazioni, ma notevole risparmio economico.

Per quanto riguarda l'evoluzione del Grid, Ian Foster[10] ne ha dato una interessante

panoramica. L'intento principale del suo intervento è stato soprattutto quello di mostrare l'effetto delle tecnologie di Grid sulla ricerca scientifica. Secondo Foster, siamo alla vigilia di una vera e propria rivoluzione copernicana: l'avvento della "Scienza 2.0", come scherzosamente l'ha definita. Tutti sappiamo quanto significhi l'uso del web per la scienza contemporanea e che esso è stato creato dagli scienziati del CERN. Negli ultimi tempi sono in corso molte discussioni sull'evoluzione del web, ovvero su quello che viene comunemente definito "Web 2.0". Gli elementi costitutivi del nuovo web, secondo Foster, si possono sintetizzare nei seguenti punti:

- software come servizio: verranno costituiti servizi di rete ricchi da un punto di vista della disponibilità di dati e di computazione;
- servizi come piattaforme: i vari servizi potranno essere facilmente combinati assieme, per creare nuove capacità che, a loro volta, potranno essere rese disponibili come nuovi servizi;
- alla base ci saranno massicce infrastrutture: Google progetta di spendere 1.5 miliardi di dollari in hardware e infrastrutture solamente nel 2006; molti altri stanno spendendo in maniera significativa;
- il costo dei servizi verrà pagato quasi interamente dalla pubblicità.

La scienza non farà che adattarsi al nuovo paradigma: qualcuno crea servizi (banche dati o funzioni), che il ricercatore scopre e compone per creare nuove funzioni, che diventeranno nuovi servizi a disposizione di altri ricercatori. Il tutto con un livello di astrazione sufficientemente alto da non rendere necessario che lo scienziato diventi un esperto di sistemi operativi e calcolatori. I servizi saranno infatti ospitati da qualcuno in grado (si spera) di gestirne sicurezza, affidabilità e scalabilità. Opportuni *client* integreranno dinamicamente i servizi distribuiti geograficamente e li pubblicheranno come nuovo servizio, realizzando quindi la separazione tra risorse e fornitore di servizi. Foster ha proseguito commentando che molto di questo è già presente nell'ultima versione del Globus Toolkit [11], lo standard *de facto* dei *middleware* per la realizzazione di griglie computazionali, che si basa su tecnologie Web Services (WS). Viene anche fornito un esempio concreto in fase di sviluppo, dato da un'applicazione che compie analisi statistiche di interesse astrofisico sulle banche dati dello

Sloan Digital Sky Survey [12], il titanico sforzo di mappatura del cielo a diverse lunghezze d'onda, che richiede un rapido accesso a centinaia di porzioni di cielo in un database di dimensioni delle decine di TeraByte, che richiede un'acquisizione dinamica di spazio disco e potenza di calcolo.

Questo nuovo paradigma deve ovviamente affrontare delle serie sfide. In primo luogo, la necessità di nuove tecnologie e abilità, riguardanti la creazione, la pubblicazione, la scoperta e quant'altro, di servizi. La conseguenza è che lo sviluppo di software diventerà probabilmente l'attività scientifica preponderante. Soprattutto sarà sempre più pressante la necessità di ulteriori e differenti infrastrutture informatiche a disposizione degli scienziati, sollevando l'ovvio problema: come è possibile che il costo di queste infrastrutture venga in qualche percentuale sostenuto da enti commerciali?

Il progetto SEPAC



Fig. 4 - Logo del Progetto SEPAC.

Il sottoscritto ha presentato una relazione sul tema "Stato e sviluppi futuri del Progetto SEPAC". Il Progetto SEPAC (South European Partnership for Advanced Computing) è nato nel 2004 come evoluzione del Progetto TAPAC [13], quando ai partner fondatori CILEA, ETH Zurigo [14] e Hewlett Packard [15] si sono uniti altri gruppi attivi nelle tecnologie avanzate, ovvero l'Università di Zurigo [16], il CSCS [17] e il consorzio SPACI [18]. Gli scopi del progetto sono:

- costruire un'infrastruttura di griglia basata su Globus Toolkit;

- analizzare lo stato dell'arte degli strumenti di griglia in un ambiente di supercalcolatori;
- testare applicazioni e interfacce selezionate in un ambiente altamente eterogeneo.

Di particolare importanza è la decisione di sostituire il sistema di informazione del Globus Toolkit, MDS, con un analogo modulo sviluppato dal gruppo SPACI dell'Università di Lecce, chiamato iGrid [19], dotato di prestazioni molto più performanti.

Una particolare attenzione è stata data alla scelta dell'interfaccia utenti più adatta, in termini di fruibilità del sistema. La scelta è anche in questo caso caduta su un applicativo sviluppato dal gruppo summenzionato dell'Università di Lecce, il portale Grid Resource Broker (GRB) [20], completamente basato su tecnologie Web Service e direttamente interfacciabile con i più comuni sistemi di code e *load balancing*. GRB è in grado di gestire workflow complessi, così come simulazioni interattive e in batch, anche con analisi parametriche.

Costruita l'infrastruttura, è stata sviluppata un'applicazione di *testbed*. La scelta è caduta su una tematica molto attuale nel campo del *drug design*, lo *screening* del *docking* di ligandi su macromolecole. E' una tematica che si presta particolarmente al Grid, in quanto lo *screening* è spesso la parte più *time consuming* nella procedura di scoperta di nuovi farmaci in silico, e allo stesso tempo è un'applicazione in cui non è necessaria la comunicazione tra i singoli task ma presenta una grande massa di dati prodotta in output da analizzare.

Il diagramma di flusso dell'applicazione in Grid è schematizzato in Fig. 5.



Fig. 5 – Diagramma di flusso dell'applicazione di test nel Grid SEPAC.

L'utente lancia l'applicazione attraverso il portale GRB. Questi interroga il sistema informativo iGrid per identificare le risorse su

cui far girare le simulazioni. GRB allora farà partire l'applicazione sulle CPU selezionate, che scaricheranno gli opportuni input dai server su cui risiedono e invieranno poi gli output altrove, per essere consolidati in un opportuno database.

Abbiamo dimostrato che il nostro progetto ha raggiunto infine un livello di produzione. Altre applicazioni in campi diversi dalla bioinformatica sono ora in fase di valutazione per nuovi *testbed*. In questo modo potremo presentarci alla comunità scientifica per dimostrare nella pratica le possibilità di queste nuove tecnologie e procedere alla creazione di una comunità di utenti.

Conclusioni

In quattro giorni è stato quindi delineato, in maniera abbastanza approfondita, lo stato di salute del mondo del calcolo ad alte prestazioni e in particolare delle tecniche di griglia computazionale.

L'impressione che si riceve è che il mondo del supercalcolo sia in un periodo di transizione, in cui le case produttrici di hardware stanno mettendo sul mercato processori sempre più potenti, grazie a un parallelismo sempre più spinto, ma che moltissimo rimanga da fare per capire come sfruttare questa potenza. Contemporaneamente, le tecnologie di griglia computazionale stanno raggiungendo una fase di consolidamento, in cui si comincia a delinearne in maniera sempre più definita vantaggi e limiti.

Bibliografia

- [1] HPC2006
URL: <http://www.hpcc.unical.it/hpc2006>
- [2] Progetto SEPAC
URL: <http://www.sepac-grid.org>
- [3] Jack Dongarra Homepage,
URL: <http://www.netlib.org/utk/people/JackDongarra/>
- [4] TOP500, URL: <http://www.top500.org>
- [5] IBM Cell
URL: <http://www.research.ibm.com/cell/>
- [6] ATI X1900XTX
URL: <http://www.ati.com/products/RadeonX1900/index.html>
- [7] Intel Woodcrest
URL: http://www.intel.com/business/enterprise/emea/eng/xeon/index.htm?ppc_cid=gg|xeon_emea_eng_wcrest|k6B26|s
- [8] Intel Montecito, URL:
<http://www.intel.com/products/processor/itanium2/index.htm>

- [9] Myricom
URL: <http://www.myri.com>
- [10] Ian foster
URL: <http://www-fp.mcs.anl.gov/~foster/>
- [11] Globus Toolkit
URL: <http://www.globus.org>
- [12] Sloan Digital Sky Survey
URL: <http://www.sdss.org>
- [13] "Il progetto Trans-Alpine Partnership for Advanced Computing", C. Arlandini, Bollettino del CILEA n.89, Ottobre 2003.
- [14] ETH Zurigo
URL: <http://www.ethz.ch/>
- [15] Hewlett Packard
URL: <http://www.hp.com>
- [16] Università di Zurigo
URL: <http://www.unizh.ch>
- [17] CSCS
URL: <http://www.cscs.ch>
- [18] Consorzio SPACI
URL: <http://www.spaci.it>
- [19] iGrid
URL: <http://sara.unile.it/~cafaro/igrd.html>
- [20] GRB
URL: <https://grb.spaci.it/cgi-bin/grb/enter>