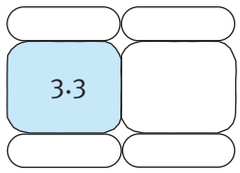
HIGH PERFORMANCE COMPUTING
STATO E PROSPETTIVE

Giovanni Erbacci

L'High Performance Computing (HPC) gioca un ruolo importante nella soluzione di problemi computazionali complessi. Il parallelismo ha affascinato i ricercatori per oltre 30 anni e oggi si assiste a un rinnovato interesse in questo settore. Nell'articolo, dopo una breve introduzione alle scienze computazionali e ad alcune applicazioni scientifiche di frontiera che si affrontano con strumenti HPC, viene presentata l'evoluzione delle architetture per il supercalcolo e se ne evidenziano le prospettive. Infine si sottolinea l'importanza dell'HPC come tecnologia abilitante a supporto della ricerca.



1. INTRODUZIONE ALLE SCIENZE COMPUTAZIONALI

Nella ricerca scientifica la formulazione di nuove teorie non può prescindere da una rigorosa formulazione di modelli matematici e da una loro verifica e sperimentazione basata sempre più spesso su simulazioni numeriche condotte al calcolatore.
Tali simulazioni, che si avvalgono di strumenti informatici avanzati, permettono di indagare sistemi fisici complessi e di estendere le teorie fondamentali della scienza moderna.
La simulazione numerica su calcolatore permette di estendere o di ridurre a piacere la scala del tempo e dello spazio per arrivare a rappresentare fenomeni molto grandi (come avviene per esempio in meteo-climatologia, in astrofisica e in geofisica) o molto piccoli (come nel caso della sperimentazione di nuovi farmaci, nella genomica, nella progettazione e validazione di dispositivi elettronici) oppure fenomeni complessi (fisica delle particelle elementari, dinamica dei fluidi), ma anche fenomeni pericolosi o costosi da gestire con metodi tradizionali (come per esempio la simulazione di guasti di impianti industriali critici, la crash analysis e così via).
Si parla di scienze computazionali per intendere quelle discipline scientifiche che utilizzano in modo sistematico metodi matematici e strumenti informatici avanzati nella loro metodologia di indagine.
L'affermazione delle *scienze computazionali* come metodologia distinta dell'indagine scientifica, affiancata alla teoria e alla sperimentazione (Figura 1), è anche il risultato della crescita continua e sostanziale delle prestazioni dei sistemi di supercalcolo. Oggi le scienze computazionali sono in grado di indirizzare i problemi scientifici più complessi e permettono di indagare fenomeni non immaginabili anche solo un decennio fa, proprio grazie alla crescita di potenza che ha caratterizzato i sistemi di elaborazione in questi anni.
L'importanza di questa metodologia è stata sancita ufficialmente nel 1998 quando venne assegnato il premio Nobel per la chimica a John A. Pople, per il contributo dato allo svi-

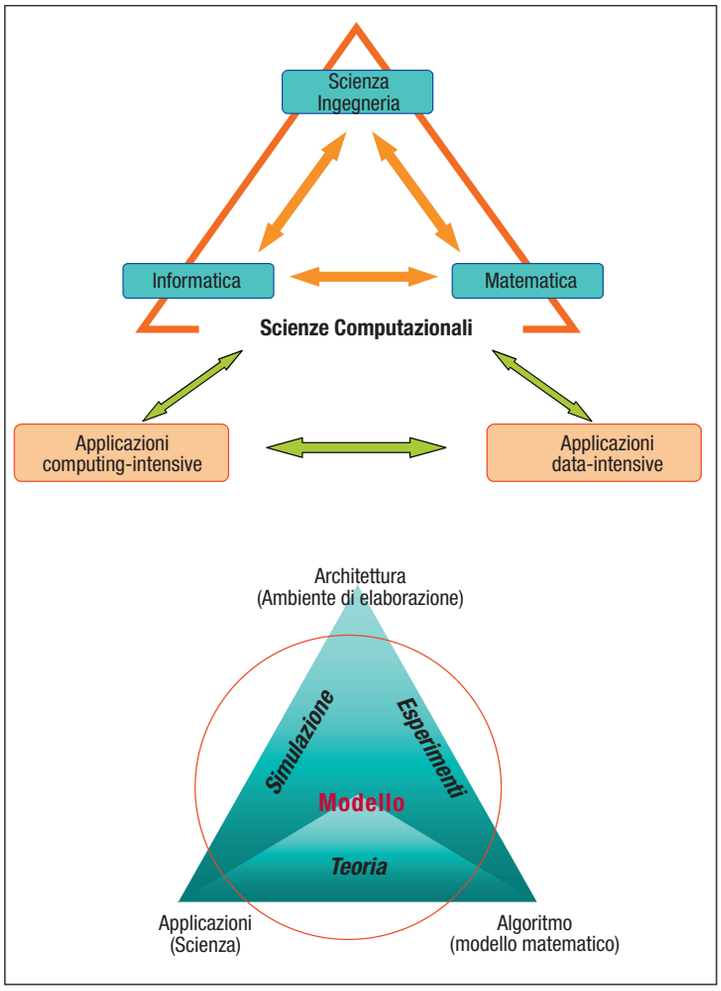


FIGURA 1

Relazioni tra scienze computazionali, informatica, matematica e applicazioni

luppo di metodi computazionali nella chimica quantistica [1].

Così come le epoche scientifiche sono definite in termini di crescita di un ordine di grandezza della capacità di osservazione sperimentale (per esempio in termini di risoluzione degli strumenti) altrettanto le epoche computazionali sono definite in termini di crescita di ordini di grandezza nelle prestazioni dei calcolatori, sia in termini di potenza di elaborazione che di capacità di memoria.

A metà degli anni '70, quando fecero la loro comparsa i primi supercomputer vettoriali, la potenza dei supercomputer più potenti era dell'ordine dei Mflop/s (10^6 operazioni in virgola mobile al secondo) [2]. Alla fine degli anni '90 si arrivò al traguardo dei Gflop/s (10^9 operazioni al secondo), attualmente siamo nell'era dei Tflop/s (10^{12}) e prima del 2010 saranno disponibili sistemi in grado di erogare una potenza sostenuta di oltre un Pflop/s (10^{15} operazioni al secondo).

Il primo supercomputer di rilevanza per la comunità scientifica è stato il *Cray-1*, progettato da Seymour Cray. Il primo esemplare, installato al Los Alamos National Laboratory (USA) nel 1976 al costo di 8,8 milioni di dollari, aveva un processore vettoriale con una frequenza di 80 MHz ed era dotato di una memoria centrale di 8 Mbyte. La potenza di picco era di 160 Mflop/s e in un particolare kernel computazionale raggiunse una prestazione sostenuta record di 133 Mflop/s.

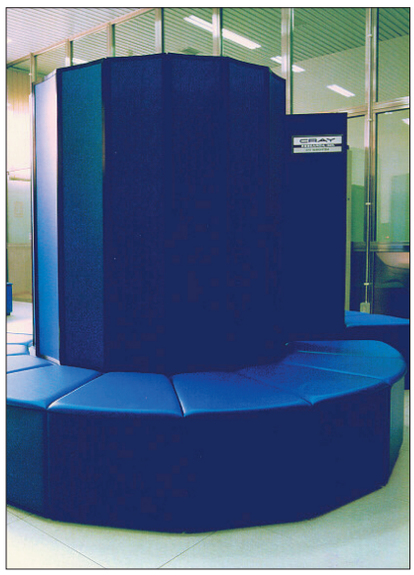
Oggi il calcolatore più potente è installato a poca distanza, al Lawrence Livermore National Laboratory. Si tratta di un sistema IBM BlueGene/L con 212.992 processori e 73 Tbyte di memoria centrale. La potenza di picco è 596 Tflop/s, mentre quella sostenuta raggiunge i 478 Tflop/s.

In un arco temporale di 30 anni si è assistito a un incremento in termini di prestazioni di oltre un fattore 10^6, sia in termini di prestazioni del sistema che di capacità di memoria.

La complessità dei processori, in termini di densità di transistori continua a crescere e, anche se con qualche rallentamento, ancora segue la legge di Moore[1]. L'incremento di potenza è dovuto alle prestazioni del singolo processore ma, soprattutto, alla capacità di integrare più processori in un'architettura scalabile dove la potenza computazionale della macchina può essere espansa con l'aggiunta di più moduli.

In Italia, il primo supercomputer fu installato al Centro Interuniversitario CINECA nel 1984. Era un supercomputer vettoriale *Cray* X-MP 12 con una potenza di picco di 160 MFlop/s e una memoria di 16 Mbyte. Attualmente, il supercomputer più potente in Italia è installato sempre in CINECA: si tratta di un cluster linux con 5120 processori e una potenza di picco di 61 TFlop/s (Figura 2).

[1] Gordon Moore, co-fondatore di Intel, predisse nel 1965 che la densità dei transistori nei circuiti integrati sarebbe raddoppiata circa ogni 18 mesi. Moore nel 1975 modificò poi tale previsione in 24 mesi.



A

B

FIGURA 2

A) il Supercalcolatore Cray X-MP48 installato in CINECA a metà degli anni '80 (4 CPU vettoriali e 64 Mbyte di RAM, potenza di picco 0.94 GFlop/s. B) il supercomputer più potente oggi in Italia, installato sempre in CINECA (Cluster Linux con 5120 core Xeon a 3.0 GHz, 20 Tbyte di RAM e una potenza di picco di 61.4 TFlop/s)

2. APPLICAZIONI SCIENTIFICHE E PROBLEMI DI FRONTIERA

Molte applicazioni scientifiche simulano un fenomeno fisico nel dominio bi-dimensionale (2D) o tri-dimensionale (3D). Il fenomeno viene studiato su un certo dominio temporale per osservarne l'evoluzione, per esempio movimenti o cambiamenti nelle forme e delle proprietà degli oggetti. Per tradurre il fenomeno fisico in un modello matematico si usano sistemi, spesso accoppiati, di equazioni differenziali che non si risolvono analiticamente e richiedono uno schema di soluzione numerica che implica calcoli pesanti su sistemi HPC.

Il problema deve essere quindi discretizzato nello spazio e nel tempo. La discretizzazione spaziale significa che il dominio è sostituito da una griglia (o mesh) 2D o 3D e per ogni punto di questa griglia si individuano un certo numero di parametri rilevanti per la computazione (per esempio, temperatura, velocità, pressione ecc.). La discretizzazione spaziale è spesso regolare, nel senso che la distanza tra i punti della mesh è costante su tutto il dominio e in tutte le dimensioni. Tipicamente, più è piccola la distanza, più accurata risulta la computazione, ma questo comporta un costo in quanto aumenta sia il tempo di calcolo che la memoria richiesta. Se certe aree del dominio fisico richiedono una maggiore accuratezza nella computazione, si introducono griglie irregolari con distanze differenti tra i punti della mesh.

La discretizzazione temporale significa che la simulazione ripete i calcoli sui diversi elementi della mesh per una sequenza di passi temporali, che varia a seconda del fenomeno da simulare: dai femto-secondi, per esempio, per le simulazioni di dinamica molecolare, agli anni, per esempio, per la simulazione delle variazioni climatiche.

Pertanto, il tempo totale per la simulazione diventa non solo proporzionale al numero di punti nella mesh ma anche al numero dei passi temporali.

Fino a un decennio fa, solo le discipline scientifiche classiche, come la chimica e la fisica, si avvalevano dei supercalcolatori per indirizzare problemi computazionali puntuali ma, nel corso degli ultimi anni, le simulazioni numeriche sono diventate sempre più una metodologia di indagine comune nei più diversi campi scientifici e indirizzano problemi

su larga scala, in modo globale. Alcuni esempi riguardano la simulazione biomolecolare applicata a proteine e ad altri sistemi complessi (Figura 3); la turbolenza in fluidodinamica; il sequenziamento in genomica; la simulazione globale dell'ecosistema geofisico terrestre ottenuto integrando diversi sottosistemi (atmosfera, oceani, flussi di calore, riscaldamento chimico, magnetismo terrestre); la simulazione di sistemi e processi biologici, multi-componenti, fino ad arrivare alla possibilità di descrivere interi organismi e addirittura popolazioni; le nanotecnologie, che affrontano lo studio di nano-dispositivi con specifiche funzionalità [3].
Per comprendere meglio la complessità dei problemi che si possono affrontare, introduciamo con maggior dettaglio due applicazioni in un settore non tradizionale dell'HPC come quello della Scienza delle Terra.

2.1. Terremoti virtuali

Negli ultimi anni la teoria e le applicazioni della propagazione delle onde acustiche hanno permesso di indirizzare nuove metodologie computazionali in campi quali la sismologia, l'oceanografia, la meteorologia l'acustica, ma anche l'ingegneria, le scienze dei materiali le scienze mediche ecc..
La simulazione del fenomeno tridimensionale completo della propagazione delle onde su strutture realistiche, con un adeguato livello di dettaglio, permette di studiare sistemi complessi in campi diversi e a diverse scale come, per esempio, la prospezione geofisica a diverse scale (analisi e gestione dei giacimenti) o il monitoraggio del rischio sismico e vulcanico.
Nella sismologia computazionale, metodi di tomografia ad alta risoluzione permettono di simulare la propagazione delle onde sismiche a varie scale, da quella locale (dimensione di una città) a quella globale, includendo man mano fenomeni complessi quali l'evoluzione della sorgente sismica e l'interazione con le strutture geologiche. Questo permette una valutazione del rischio sismico in ambiti specifici fino allo studio della struttura interna della terra.
Questi problemi diventano vere e proprie sfide computazionali quando si aumenta la complessità dei modelli coinvolti poiché in tutti i metodi attualmente in uso (differenze finite, elementi finiti, metodi spettrali ecc.) la risoluzione è funzione della scala introdotta. Inoltre, all'interno dell'analisi del rischio sismico, assume un peso considerevole lo studio probabilistico degli effetti locali, che comporta la definizione di tutti i possibili, o quantomeno probabili, scenari di rischio; questi metodi richiedono usualmente diverse centinaia di simulazioni per raggiungere un buon grado di attendibilità.
Una sfida ulteriore è rappresentata dalla grande quantità di dati coinvolti; un esempio concreto della dimensione del problema è dato dalla modellazione del bacino di Los Angeles, il cui scenario completo in quattro dimensioni è dell'ordine di 40 Tbytes.
Progetti di questo tipo, che riguardano modelli realistici, richiedono risorse importanti sia in termini di innovazione algoritmica (per migliorare la complessità computazionale in funzione dell'accuratezza) che di supercalcolo. Esempi notevoli sono i tre grandi progetti attivi al SCEC, (*Southern California Earthquake Center*): TeraShake per la simulazione della propagazione delle onde sismiche, CyberShake, una piattaforma computazionale per modellare la forma d'onda 3D e sviluppare le curve probabilistiche di rischio sismico di prossima generazione, ed infine DynaShake per la simulazio-

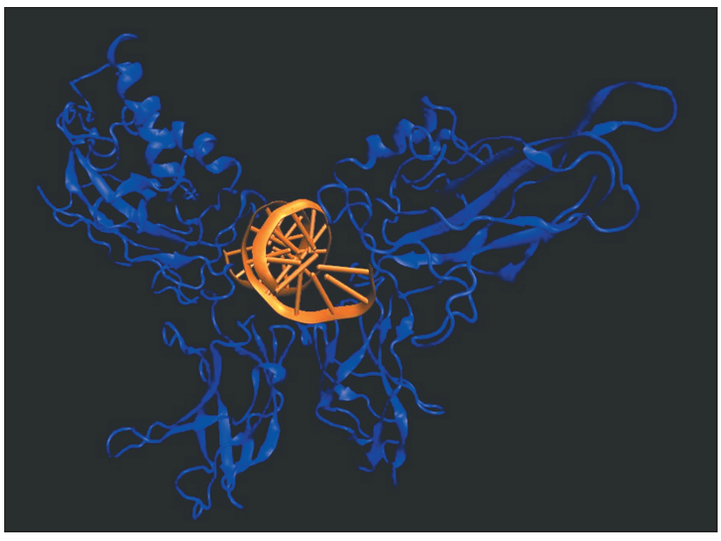


FIGURA 3

Simulazione della dinamica molecolare di un complesso proteico (blu) con il DNA (arancio). La simulazione per un sistema di circa 150.000 atomi, che dura 13 ns, ha richiesto circa 10.000 h di CPU su un cluster Linux

ne delle rotture dinamiche a livello sismico e la parametrizzazione cinematica degli epicentri[2]. L'esecuzione di questi progetti richiede potenze di calcolo considerevoli quali l'utilizzo di un sistema IBM BlueGene a 40.000 processori.

2.2. Vulcani simulati

Un esempio più vicino a noi è rappresentato dal progetto Exploris[3]. Nell'ambito di questo progetto è stato formulato il modello matematico 3D delle eruzioni esplosive e della dispersione di ceneri vulcaniche. Per la prima volta al mondo sono state realizzate le simulazioni tridimensionali dei processi di dispersione delle ceneri nell'atmosfera, del collasso della colonna vulcanica e della formazione di colate piroclastiche lungo le pendici del vulcano [4].
Tali modelli, utilizzando notevoli potenze computazionali, hanno consentito di rappresentare i fenomeni vulcanici in modo molto accurato nonché di quantificare meglio le azioni pericolose ad essi associate. Assieme alla temperatura, le simulazioni hanno permesso di studiare anche molte altre grandezze che caratterizzano l'eruzione, come la densità dei piroclasti e dei gas, la pressione, la velocità e la direzione del flusso piroclastico.
Le simulazioni più accurate hanno avuto come soggetto l'eruzione esplosiva più probabile per il Vesuvio, per la quale è stato utilizzato un dominio di simulazione comprendente un'area di 12 km di lato (Vesuvio più territorio circostante) e che si estende fino alla quota di 8 km; il dominio è stato discretizzato su una griglia cartesiana di dimensione 200^3 a risoluzione variabile (dai 20 m in prossimità del cratere a 100 metri per le celle più lontane).
L'eruzione è stata seguita per 30 min di tempo reale utilizzando una discretizzazione temporale di 0.1 s. Per portare a compimento questa simulazione sono stati utilizzati 450 processori del sistema IBM SP Power5 del CINECA per un totale di 180 h (80.000 h/cpu), equivalenti a 10 anni di calcolo utilizzando un solo processore [5].
Le simulazioni su grande scala, come quelle sopra citate, producono un'enorme mole di dati, circa 0.5 Tbyte per ogni simulazione, che devono essere analizzati e visualizzati per essere comprensibili agli esperti. Per la visualizzazione sono stati realizzati due strumenti: uno dedicato all'analisi quantitativa, che si interfaccia direttamente con i dati della simulazione (pressione, temperatura, concentrazione e velocità delle particelle e dei gas); l'altro, dedicato alla analisi qualitativa ed alla integrazione con altre tipologie di dati provenienti da sorgenti diverse. Tale strumento consente di integrare in un'unica visualizzazione: i dati sui flussi piroclastici provenienti dalla simulazione; un modello fotorealistico del suolo (*orthofoto*) e il suo profilo topografico (DEM - *Digital Elevation Model*); e le informazioni geografico-urbanistiche (GIS). Grazie a questa integrazione è possibile, ad esempio, visualizzare contemporaneamente la densità abitativa o la rete viaria, sovrapposta alla temperatura del flusso piroclastico, come si vede nella figura 4.

3. GRANDI QUANTITÀ DI DATI

Come mostrano gli esempi precedenti, oggi le simulazioni riguardano problemi reali complessi che vengono affrontati globalmente con approcci multidisciplinari. È distante ormai il caso in cui lo studio della aerodinamica di un veicolo non poteva essere gestito in modo globale perché la simulazione era troppo onerosa (sia in termini di tempi di calcolo che di memoria richiesta) e si analizzava solo una metà del veicolo, sfruttando la probabile simmetria dell'altra metà.
Oggigiorno, si possono affrontare simulazioni complesse sia perchè sono presenti sistemi di supercalcolo potenti, ma anche perché sofisticati dispositivi di acquisizione dati quali sensori, radar, satelliti, TAC, scan-

[2] http://www.scec.org/

[3] EXPLOsive Eruption RISk and Decision Support for EU Populations Threatened by Volcanoes. È un progetto triennale finanziato dalla Comunità Europea e coordinato dalla sezione di Pisa dell'Istituto Nazionale di Geofisica e Vulcanologia. CINECA ha collaborato nell'attività di sviluppo dei modelli numerici e nell'integrazione e visualizzazione dei dati. Tutta l'attività computazionale si è svolta sul sistema IBM SP Power 5 di CINECA.

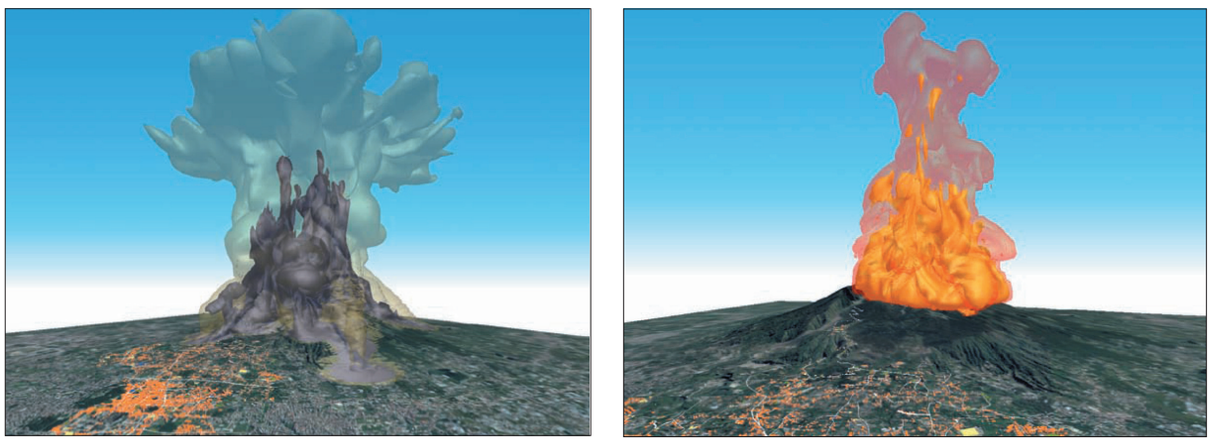


FIGURA 4

Il Vesuvio pochi minuti dopo l'inizio dell'eruzione virtuale; due diverse gradazioni di rosso evidenziano le isosuperfici della temperatura, a 100 (superficie esterna) e 350 (superficie interna) gradi Celsius. Le isosuperfici grigie rappresentano due diversi livelli di concentrazione del materiale eruttivo

ner, micro-arrays[4] ecc. producono con facilità enormi quantità di dati, indispensabili per la formulazione di modelli più rigorosi e complessi.

Le simulazioni numeriche, che si appoggiano su modelli complessi e dati di input consistenti, a loro volta producono moli di dati altrettanto grandi che spesso richiedono una post-elaborazione al fine di poter recuperare tutta l'informazione utile, e che vanno conservati per successive elaborazioni.

La produzione di dati sta subendo una grande accelerazione e il volume di dati prodotti cresce sia come dimensioni che in termini di complessità: si stima che il volume di dati prodotti a livello mondiale raddoppi ogni anno. Occorre quindi studiare nuove metodologie per organizzare, manipolare ed analizzare i dati, nonché strumenti per garantirne l'accesso in modo efficiente, ed ottimizzare così il processo di produzione della conoscenza.

In molti campi della ricerca si assiste quindi a una forte spinta verso la creazione di infrastrutture capaci di garantire la massima diffusione e fruibilità delle informazioni. Per esempio, la comunità internazionale astrofisica da alcuni anni lavora alla creazione del *Virtual Observatory* (VO), definito come: "*an enabling and coordinating entity to foster the development of tools, protocols, and collaborations necessary to realize the full scientific potential of astronomical databases in the coming decade*" [7].

I sistemi di supercalcolo non possono quindi limitarsi ad offrire solo potenza di calcolo, ma diventano vere e proprie infrastrutture in grado di fornire una serie di servizi complementari per la gestione e la fruizione efficace dei dati prodotti, limitando la movimentazione degli stessi su reti spesso troppo lente o inaffidabili.

Inoltre è sempre più importante tenere se-

[4] I micro-arrays, o matrici ad alta densità, sono una tecnica che, sfruttando le caratteristiche peculiari della doppia elica del DNA, ha aperto di fatto la strada alla possibilità di analizzare i profili di espressione genica di un intero organismo. In generale, un esperimento di analisi dei profili di espressione fornisce come risultato una matrice di dati in cui le righe rappresentano i geni monitorati e le colonne corrispondono alle diverse condizioni sperimentali, quali punti temporali, condizioni fisiologiche, tessuti. Ogni elemento della matrice rappresenta quindi il livello di espressione di un particolare gene in uno specifico stato fisiologico.
La gestione e l'interpretazione dei dati generati dalle matrici ad alta densità rappresenta un aspetto fondamentale di questa tecnologia. Infatti, queste matrici diventano sempre più grandi e richiedono tecniche di analisi statistiche avanzate, quali il data mining, che impegnano risorse di calcolo e di memorizzazione importanti. Nel caso dei profili di espressione genica, le tecniche di data mining rappresentano un utile strumento per identificare ed isolare particolari pattern di espressione che di fatto rappresentano delle vere e proprie impronte digitali genetiche di un determinato stato fisiologico [6].

parata l'attività di produzione dei dati da quella più vasta di analisi. Occorre quindi strutturare i dati prodotti attraverso metadati efficaci in grado di descrivere i dati stessi, secondo metodologie standard, in modo da poterli accedere, analizzare e visualizzare attraverso strumenti software appropriati, ma al tempo stesso poterli condividere efficientemente nell'ambito di discipline scientifiche diverse.

4. ARCHITETTURE PER IL SUPERCALCOLO

L'introduzione reale dei sistemi di supercalcolo avvenne alla fine degli anni '70, quando i primi sistemi vettoriali divennero un supporto concreto ed efficace alla comunità scientifica per realizzare simulazioni numeriche importanti. Da allora l'evoluzione di tali sistemi è stata rapida e impetuosa sia in termini di potenza che di innovazione architetturale e il loro utilizzo si è radicato in differenti contesti applicativi.
Le tappe di questa evoluzione architetturale hanno riguardato dapprima i sistemi vettoriali mono-processore (come per esempio, il Cray 1, e il CDC Cyber 203) poi, a partire dai primi anni '80, i sistemi vettoriali multi-processore a memoria condivisa. Si trattava di sistemi SMP (*Symmetric Multi Processor*) dove il singolo processore aveva funzionalità vettoriali, come, per esempio, i sistemi Cray della serie X-MP e Y-MP, i sistemi NEC della serie SX e, successivamente, i sistemi Cray C90 e i sistemi IBM della serie 3090 dotati di "vector facilities". Nella seconda metà degli anni '80 comparvero i primi sistemi paralleli a memoria distribuita, come per esempio, l'Intel *iPSC/1* che utilizzava microprocessori Intel 80826 interconnessi da una rete *Ethernet* secondo una topologia ipercubo o la *Connection Machine* CM2 della *Thinking Machine Corporation* [8].
A partire dai primi anni '90 i sistemi *Massively Parallel Processors* (MPP) si sono imposti sul mercato HPC. La CM5, il Cray T3D e il Cray T3E sono gli esempi di maggior successo.
Da allora si è consolidata la presenza di sistemi HPC costituiti da *cluster* di nodi SMP basati su processori RISC, o microprocessori *off the shelf* X86-32 e X86-64, interconnessi da reti veloci. Queste architetture, seguendo la rapida evoluzione tecnologica dei microprocessori oggi caratterizzano la maggior parte dei sistemi HPC sul mercato. La distinzione tra sistemi convenzionali e supercomputer diviene molto più sfumata. Un sistema convenzionale, opportunamente "clusterizzato", può diventare di fatto un supercomputer. Nella attuale classifica della Top 500[5] il numero di sistemi basato su processori *commodity* rappresenta ormai oltre l'80% del totale.
Le ragioni della rapida affermazione dei cluster nel mondo HPC sono da ricercarsi nella disponibilità di componenti *commodity* di prestazioni elevate e a costi contenuti. Questo ha fatto sì che a parità di potenza di picco un cluster potesse costare fino a 10 volte meno di un supercalcolatore tradizionale. Una tale differenza rende più difficile giustificare la scelta di supercalcolatori tradizionali, anche a fronte dalla maggior robustezza ed affidabilità di questi ultimi.

4.1. Classificazione delle architetture parallele

La maggior parte dei sistemi paralleli attuali ricade nella classe MIMD (*Multiple Instruction stream Multiple Data stream*) secondo la classica tassonomia di Flynn [9], che distingue i sistemi paralleli in base al flusso delle istruzioni e a quello dei dati interno alla architettura stessa. La classe MIMD racchiude sistemi paralleli costituiti da un insieme di processori indipendenti, ciascuno dei quali è governato da una propria unità di controllo.
Una modalità forse più utile di classificare le architetture parallele è in base al loro modello di memoria: sistemi a memoria condivisa e sistemi a memoria distribuita, come schematizzato nella figura 5. Nei sistemi a memoria condivisa i processori condividono la memoria centrale e vedono un singolo spa-

[5] La classifica top 500 (www.top500.org) fornisce la graduatoria dei 500 calcolatori più potenti a livello mondiale. La misura viene fatta in base alle prestazioni raggiunte nella soluzione di un sistema lineare denso Ax = b, ricorrendo alla libreria LINPACK (www.netlib.org). Tale classifica viene pubblicata due volte all'anno a partire dal 1993.

zio degli indirizzi. I sistemi a memoria condivisa si distinguono a loro volta in sistemi SMP (*Symmetric Multi Processors*), detti anche sistemi UMA (*Uniform Memory Access*), dove l'accesso alla memoria è uniforme per ogni processore, e sistemi NUMA (*Non-Uniform Memory Access*) per i quali il tempo di accesso alla memoria dipende dalla locazione locale o remota dei dati.
Nei sistemi a memoria distribuita, i processori sono connessi tramiti una rete di interconnessione, ogni processore può indirizzare direttamente solo la propria memoria locale e occorre un protocollo a scambio di messaggi per scambiare informazioni tra un processore e l'altro. Tali sistemi sono detti sistemi NORMA (*No Remote Memory Access*). Ormai da oltre un decennio, i sistemi a memoria distribuita non interconnettono più il singolo processore ma nodi, costituiti da sistemi SMP: i nodi, spesso costituiti da processori *off-the-shelf* e memoria DRAM *commodity*, sono interconnessi da reti più o meno performanti, anch'esse spesso *commodity* [10].
A questi modelli architetturali corrispondono altrettanti modelli di programmazione parallela: il *modello a memoria condivisa*, che sfrutta la condivisione della memoria per lo scambio delle informazioni tra i diversi thread che operano sui processori del sistema parallelo; il *modello a memoria locale*, dove lo scambio di informazioni e la sincronizzazione tra i processi viene gestita per mezzo del paradigma *message passing*. In entrambi i casi la programmazione avviene per mezzo di linguaggi ad alto livello di tipo procedurale o ad oggetti (C, C++, Fortran90) con il supporto di API specifiche per la gestione del parallelismo: OpenMP[6] per il modello a memoria condivisa e MPI[7] per quello a memoria locale [2, 11].

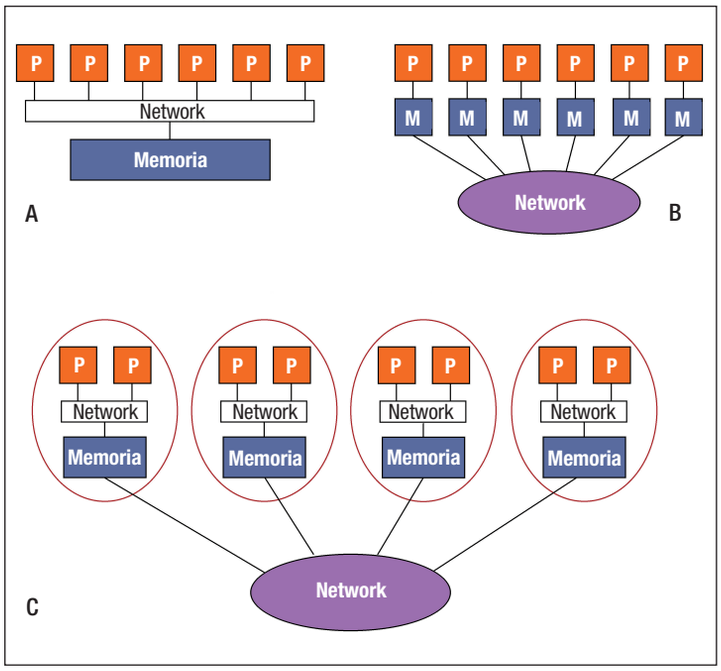


FIGURA 5

Schema architetturale dei sistemi paralleli.
***A**) Architettura a memoria condivisa: tutti i processori indirizzano l'intera memoria (singolo spazio degli indirizzi). **B**) Architettura a memoria distribuita: ogni processore indirizza solo la propria memoria locale; per scambiare dati tra i processori occorre adottare un paradigma a scambio di messaggi.*
***C**) Architettura a memoria distribuita costituita da nodi SMP: all'interno del nodo si può adottare un paradigma a memoria condivisa mentre per comunicare all'esterno del nodo occorre un paradigma a scambio di messaggi*

4.2. Trend dei sistemi HPC

Se si considera l'andamento della classifica Top500 dal 1993, anno della sua prima pubblicazione, ad oggi, si traggono utili informazioni circa la rapida evoluzione dei sistemi per il supercalcolo. Nel 1993, il supercomputer al primo posto raggiungeva 59,7 Gflop/s mentre quello al 500° posto raggiungeva i 422 Mflop/s; nel 2007 quello al primo posto raggiunge i 478 TFlop/s contro i 5.9 Tflop/s di quello all'ultimo posto, con un incremento in termini di prestazioni di circa un fattore 8.000 per il sistema al primo posto e di oltre 13.000 per quello in fondo alla classifica.
Dal 2005 entrano nella graduatoria solo sistemi con una potenza superiore al Tflop/s e nel 2007 il sistema al 500° posto ha una potenza di 5.9 Tflop/s. Questo sistema nella lista precedente (giugno 2007) sarebbe stato al 255° posto e ben 245 sistemi della lista precedente risultano troppo poco potenti per avere ancora un posto nella lista attuale.
La figura 6 mostra l'andamento negli anni della potenza effettiva, rispettivamente del sistema al 1° posto e di quello al 500° della Top500. Nel grafico è riportata anche la potenza effettiva, negli anni, del sistema italiano più potente: solo sistemi installati in

[6] http://www.openmp.org

[7] http://www.mcs.anl.gov/mpi/
http://www.mpi-forum.org/

CINECA hanno guadagnato questa posizione nella Top 500.
Nella lista di novembre 2007, su 500 sistemi, ben 354 usano processori Intel (70.8%), 78 sistemi (15,6%) usano processori della famiglia AMD Opteron e 61 sistemi (12,2%) usano processori IBM Power.
I sistemi cluster rappresentano l'architettura più comune nella Top500 con una presenza di ben 406 sistemi. Per quanto riguarda le reti di interconnessione la tecnologia Infiniband[8] accresce la sua presenza con una presenza di 121 sistemi, mentre Gigabit Ethernet è ancora la tecnologia di interconnessione più utilizzata (270 sistemi).
Gli Stati Uniti continuano ad essere i leader di questa classifica con 284 sistemi installati, contro i 149 sistemi dell'Europa. Ma un nuovo trend geografico si impone in modo significativo in questi ultimi anni: la crescita dei sistemi nel continente Asiatico che vede oggi 50 sistemi, di cui 20 sono in Giappone, 10 in Cina e 9 in India.
Tra le nazioni europee, il primo posto è ricoperto dall'Inghilterra con ben 48 sistemi installati, seguita dalla Germania con 31 sistemi. L'Italia è presente con soli 6 sistemi.
Il primo di questi è al 48° posto ed è installato in CINECA. Si tratta di un sistema cluster linux BladeCenter Xeon dual core (5120 core a 3.0 GHz) con una memoria RAM complessiva di 20 Tbyte e una rete di interconnessione full bisection Infiniband 4X.

4.3. Evoluzione della tecnologia dei microprocessori

Lo sviluppo dei microprocessori mostra alcune linee di tendenza pressoché indipendenti dal singolo costruttore e dal disegno stesso del processore che valgono sia per processori

FIGURA 6

Andamento (in scala logaritmica) delle prestazioni dei sistemi nella lista Top 500 dall'inizio della pubblicazione della classifica (Giugno 1993) ad oggi, rispettivamente del sistema al 1° posto (linea azzurra), del sistema al 500° posto (linea arancio) e del sistema più potente installato in Italia (linea verde). Tutti i sistemi italiani sono stati installati in CINECA

[8] www.infinibandta.org

CISC (*Complex Instruction Set Computer*), RISC (*Reduced Instruction Set Computer*), EPIC (*Explicity Parallel Instruction Computing*) e anche per processori *custom* (vettoriali) [10].
La prima e più importante tendenza è quella relativa al mantenimento del processo di miniaturizzazione basato su tecnologia al silicio per i prossimi 5-6 anni. In particolare i processori si avviano a una nuova generazione basata sul processo tecnologico con scala litografica a 65 nm, contro i 90 nm dei processori attuali. È già prevista la costruzione di linee di produzione con processi tecnologici a 45 nm e 32 nm. Questo consentirà ai produttori di rimanere sulla curva delineata dalla legge di Moore che prevede un raddoppio della densità dei circuiti ogni 18-24 mesi, che è esattamente l'incremento che si ha nel passaggio da una scala di integrazione alla successiva.
Per quel che riguarda le frequenze di clock non ci dovremmo invece aspettare incrementi spettacolari come quelli a cui abbiamo assistito negli scorsi anni, questo soprattutto per i problemi connessi al calore sviluppato dalle correnti parassite che, con la tecnologia attuale, scalano col quadrato della frequenza. Per questo i produttori stanno investendo molto nella ricerca di materiali ad alta costante dielettrica (k) per diminuire le correnti parassite.
Come conseguenza della impossibilità di elevare il clock oltre un certo livello i produttori, per aumentare la potenza allo stesso passo dell'integrazione (raddoppio ogni 18-24 mesi) si sono orientati verso la produzione di microprocessori *multi-core*, ovvero contenenti più CPU. Questa è ormai una tendenza consolidata, che al momento ha determinato la comparsa di microprocessori *dual-core* e *quad-core*, e che nel prossimo futuro porterà ad avere sul mercato microprocessori con un numero sempre più elevato di core [10].
Questo fatto combinato con l'altrettanto consolidata tendenza ad avere sul mercato sistemi SMP con 2, 4, 8 microprocessori, farà crescere nei sistemi di prossima generazione il numero di CPU per nodo, con conseguente stress del sottosistema di memoria, che risulterà sempre più critico per le prestazioni del sistema globale.
Risulta quindi naturale che vi sia anche una tendenza generale a dotare i sistemi di memoria cache di terzo livello (L3) di dimensioni generose e condivisa tra i processori SMP. Mentre la cache di secondo livello (L2) sarà condivisa a livello di microprocessore e quella di primo livello (L1) a livello di singolo core. L'organizzazione gerarchica della memoria sarà quindi sempre più marcata e saranno presenti almeno 5 livelli: Registri, L1, L2, L3, RAM. La gestione degli accessi alla memoria dovrà essere efficiente affinché le applicazioni riescano a sfruttare appieno la potenza del processore.
È opportuno sottolineare come la linea di processori RISC Power di IBM abbia anticipato tutte queste tendenze, essendo stata la prima ad introdurre processori multi-core e cache di livello 3.
Infine è ormai definitivo il passaggio di tutte le architetture a microprocessori all'indirizzamento a 64 bit, che consente di aumentare la quantità di RAM indirizzabile dal singolo processore.
La tendenza ad aumentare il numero di core all'interno dei microprocessori, è documentata dalla disponibilità di un processore Intel Xeon e di un processore AMD Opteron *quad-core* per singolo chip, entrambi prodotti con un processo tecnologico a 65 nm. Osserviamo che la disponibilità di processori *quad-core*, compatibili con i *dual-core*, fa sì che a parità di volume fisico occupato, si possa raddoppiare la potenza di elaborazione di un cluster pre-esistente, semplicemente sostituendo i processori.
Altre novità per i sistemi ad alte prestazioni sono l'evoluzione del processore Power IBM e il Cell processor.
Il Power 6 IBM si presenta come la naturale evoluzione del Power 5 e sarà presentato inizialmente solo come dual core, e con un clock molto elevato (4.7 GHz). Introdurre solo un dual core, in questo caso può essere un vantaggio perchè permetterà di avere un miglior rapporto potenza/banda di memoria [12]. Poco si sa ancora del Power 7 successore del Power 6. Potrebbe uscire tra un paio di anni e potrebbe essere *pin compatible* con altri processori già presenti sul mercato in modo da permettere l'integrazione di sistemi disomogenei.
Il Cell processor IBM (derivato dal processore della PlayStation di terza generazione) presenta un disegno architetturale innovativo (*Syner-*

gic Processor)[9]. In breve, l'unità di elaborazione è costituita da un processore Power e da 8 co-processori specializzati per operazioni floating-point, integrati in una architettura SIMD in cui tutti i co-processori eseguono la stessa istruzione su dati diversi. Tale architettura si presta al calcolo tecnico scientifico, e sulle operazioni a singola precisione il processore può dimostrare una potenza prossima ai 250 Gflop/s, mentre sulla doppia precisione la stima è di 25 Gflop/s. Sebbene molto promettente, questo tipo di processore potrà, nel breve periodo, raggiungere prestazioni non adeguate alle sue potenzialità perché richiede una sostanziale revisione delle applicazioni e del supporto dei compilatori.

4.4. Evoluzione delle reti di interconnessione

Le reti di interconnessione dei sistemi di calcolo ad alte prestazioni devono essere all'altezza dei processori che collegano, e quindi un calcolatore parallelo deve essere corredato da una rete di interconnessione adeguata in termini di *bandwidth* (banda passante) e *latenza* [10].
Al giorno d'oggi esistono varie reti di interconnessione utilizzate per i sistemi HPC, che possiamo dividere in reti proprietarie, e quindi di uso esclusivo del fornitore dei sistemi e spesso integrate con il sistema stesso, e in reti standard *off-the-shelf*, che i fornitori integrano nelle loro soluzioni spesso utilizzando componenti di fornitori terzi.
Alcuni esempi di reti di interconnessione proprietarie sono:

- l'*High Performance Switch* (HPS) - Federation adottato da IBM nei sistemi della serie SP Power. HPS è uno *switch flat* che collega i processori direttamente, senza passare attraverso hardware addizionale come l'interconnessione PCI;
- la rete toroidale 3D dei sistemi paralleli Cray XT4, composta da router proprietari denominati *SeaStar2*[10]. I processori AMD Opteron di XT4 sono collegati ai SeaStar2, tramite HyperTransport a 6.4 GB/s e, in prospettiva, HyperTransport 3.0;
- le reti del sistema Blue Gene/L: Il sistema è composto da nodi biprocessore (PowerPC 440 e coprocessore matematico) interconnessi da 3 reti di comunicazione adibite al calcolo, con caratteristiche e velocità differenti: rete toroidale tridimensionale (per le comunicazioni punto-punto), rete collettiva ad albero binario (supporto hardware per le operazioni collettive) e una rete barrier di pura sincronizzazione.

Da tempo sono presenti sul mercato reti di interconnessione standard *off-the-shelf* come Myrinet, Qsnet, SCI Dolphin, ma recentemente grande aspettativa è riposta sulle tecnologie Infiniband[11] e 10 Gbit Ethernet (GbE), tanto che le case produttrici di tali reti offrono prodotti basati su queste tecnologie.
In futuro sarà probabilmente definito un protocollo per la 40-GbE o per la 100-GbE, ma il dato vero è che queste reti difficilmente si presteranno nell'immediato futuro come reti di interconnessione dei cluster per il calcolo ad alte prestazioni: il loro obiettivo primario è quello di sostituire la GbE in ambito enterprise nelle situazioni in cui questa non risulti sufficiente.
InfiniBand (IB) è stata creata come standard "open" per il supporto ad una architettura di I/O ad alte prestazioni. Lo standard nasce quindi orientato all'I/O, ma l'interconnessione oggi giorno è largamente utilizzata come rete di connessione tra nodi di calcolo per la realizzazione di sistemi paralleli. Il vantaggio di essere "open" risiede nel fatto che è possibile utilizzare in modo cooperativo hardware e software di produttori diversi all'interno della stessa rete, così come avviene per altri dispositivi (Ethernet, SCSI, Fiber Channel ecc.).
Lo standard IB corrente supporta le bandwidth 1X (2.5 Gbit/s), 4X (10 Gbit/s) e 12X (30 Gbit/s). Inoltre lo standard prevede anche tramissioni DDR e QDR che permettono di raddoppiare e quadruplicare tali valori. Quindi lo standard prevede di fatto reti di trasmissione in grado di trasmettere fino a 120 Gbit/s.
Quanto previsto dallo standard è correntemente implementato principalmente da 4 fornitori: Mellanox, Cisco, Silver Storm e Voltaire. Le implementazioni si differenziano per vari fattori quali: bandwidth supportata, numero di porte dello switch, tipo di interconnessione (rame o fibra ottica), tipo e caratteristiche dell'Host Channel Adapter.

[9] http://www.research.ibm.com/cell/

[10] http://www.cray.com/

[11] http://www.infinibandta.org/

Per esempio, lo switch monolitico full bisectional Cisco, permette il collegamento di 288 porte 4X DDR, e una full bisection bandwidth di 11.52 Tbit/s. Con sistemi di questo tipo è possibile realizzare cluster composti da migliaia di nodi di calcolo, la cui dimensione è limitata principalmente dalla massima lunghezza dei cavi (ovvero dal costo e dalla stabilità, visto che nel caso di collegamenti in fibra ottica, non ci sono limitazioni importanti relativamente alla lunghezza dei cavi).

4.5. Evoluzione delle architetture per il supercalcolo

L'affermazione delle architetture *cluster* basate prevalentemente su componenti quali processori e rete *off the shelf* è sempre più radicata in ambito HPC. Tali sistemi saranno caratterizzati da prestazioni sempre più elevate, in funzione della disponibilità di sistemi multi-core che caratterizzeranno i processori dell'immediato futuro.
Accanto a queste architetture, emergono anche architetture di supercalcolo *special purpose*, e architetture eterogenee che indirizzano meglio il capability computing. Il *capability computing*, contrapposto al *capacity computing*, indirizza quei problemi *challenge* che richiedono architetture potenti e ben bilanciate, in termini di potenza del processore, capacità di memoria e ampiezza della banda di interconnessione.
Un esempio di architettura special purpose di successo è il sistema IBM BluGene/P[12] (BG/P) successore di BluGene/L inizialmente progettato per lo studio del problema del folding delle proteine, poi commercializzato anche in altri ambiti applicativi. La caratteristica principale di questo sistema è quella di aver spinto il parallelismo interno all'estremo, ben oltre quello ottenibile con i cluster tradizionali, contenendo contemporaneamente il consumo energetico e l'occupazione fisica dello spazio. A differenza dei cluster i nodi di BG/P non ospitano un sistema operativo "completo" e sono collegati da più reti dedicate a diverse funzioni, tutto per aumentare il sincronismo a livello applicativo. I principali punti deboli di BG/P sono sia la scarsa disponibilità di memoria sul singolo nodo che la difficoltà di adattare le applicazioni alle caratteristiche della macchina.
BG/P, grazie all'utilizzo del processore PowerPC 450 (quad-core), raggiungerà il Tflop/s di picco, rispetto ai 360 Gflop/s di BG/L. L'evoluzione è rappresentata da BG/Q che, nel 2010, potrebbe adottare architettura Power o Cell e raggiungere i 10 Tflop/s.
Un'altra architettura di rilievo per il supercalcolo è quella offerta da Cray con i sistemi XT5, in grado di tenere assieme tecnologia di microprocessore general purpose (processore AMD Opteron quad-core) e tecnologia di rete dedicata.
Il Cray XT5 è il punto attuale di una *roadmap* Cray più ampia e tesa a produrre sistemi in grado di raggiungere il Pflop/s sostenuto su applicazioni reali, prima del 2010. Il primo passo in questa direzione è la piattaforma selezionata dal U.S. Department of Energy's (DOE) per l'*Oak Ridge National Laboratory* (ORNL) che, basata sull'evoluzione XT5, raggiungerà il Pflop/s di picco entro il 2009. Sarà un sistema con 23936 processori AMD quad-core a 2.8 GHz (95744 cores).
Cray, con il progetto Cascade[13], sta progettando una nuova generazione di tecnologia HPC che realizza il modello di *adaptive supercomputing*. Tale modello fa perno attorno al concetto di architettura adattiva che integra in un unico sistema multi-architetturale tecnologie di elaborazione scalari, vettoriali, multithreading e di computing riconfigurabili (FPGA[14]). Il progetto Cascade prevede anche

[12] http://www-03.ibm.com/systems/deepcomputing/bluegene/

[13] http://www.cascade.cray.com

[14] Un FPGA (*Field Programmable Gate Array*) è un circuito integrato *general-purpose* che può essere riprogrammato, anche dopo essere stato inserito in un sistema. La programmazione consiste nel caricare un programma di configurazione chiamato *bitstream* in un'apposita memoria statica. Come il codice binario per un processore, il *bitstream* è prodotto da strumenti di compilazione che traducono le astrazioni di alto livello definite dal progettista.
Un FPGA si presenta come un *array bidimensionale* e configurabile di risorse utili ad implementare una vasta gamma di funzioni aritmetiche e logiche quali ricerca, ordinamento, elaborazione di segnali, manipolazione di immagini, crittografia, correzione degli errori, codifica/decodifica, generazione di numeri casuali ecc.. Un FPGA può essere utilizzato come co-processore per accelerare le applicazioni o parti di esse.

lo sviluppo di un adeguato corredo software che spazia dai sistemi operativi ai compilatori agli ambienti di sviluppo. L'obiettivo ambizioso è quello di realizzare supercomputer che si possono adattare alle applicazioni, anziché forzare le applicazioni ad adattarsi al supercomputer.
Cascade è la risposta Cray all'HPCS *program* lanciato dal DARPA[15] per finanziare lo sviluppo di una nuova generazione di supercomputer di classe Pflop/s, facili da programmare e in grado di gestire un'ampia gamma di applicazioni (general purpose).
Solo Cray e IBM hanno raggiunto l'ultima fase del programma HPCS e nel novembre 2006 sono state finanziate rispettivamente con 250 e 245 milioni di dollari per arrivare a produrre sistemi Pflop/s con tali caratteristiche, entro il 2010.

5. VERSO NUOVE SFIDE

Come si è visto, la disponibilità di sistemi paralleli multi-core con un elevato numero di processori è realtà e ci si aspetta che il numero di core per chip raddoppi ad ogni generazione di processori. Questi sistemi pongono sfide importanti per il programmatore. Il *multi-core* non è una nuova versione di sistema SMP e i *core* non si possono considerare del tutto indipendenti ai fini della programmazione infatti i *core* condividono sul chip risorse in modo diverso da come fanno i processori SMP tradizionali. Questa situazione sarà ancora più complessa in un prossimo futuro quando altre componenti non standard verranno integrate a livello architetturale, come tipi di core differenti e acceleratori hardware quali GPU[16] e FPGA. Bisognerà quindi introdurre tecniche innovative per gestire il problema dell'accesso ottimale ai dati, nelle diverse gerarchie di memoria che i sistemi *multi-core* inevitabilmente adotteranno.
Occorrerà introdurre nuove tecniche di programmazione per scalare su numeri elevati di processori senza perdere di efficienza negli algoritmi. Uno degli aspetti critici nell'utilizzo di risorse computazionali avanzate è la scalabilità che caratterizza gli algoritmi numerici. Questo aspetto riguarda non solo la scalabilità in termini di tempo di esecuzione in funzione del numero di processori ma anche l'interdipendenza tra il tempo di esecuzione, le richieste di memoria e la dimensione computazionale del problema.
Per almeno due decenni i programmatori in ambito HPC si erano abituati al fatto che ad ogni nuova generazione di microprocessori corrispondeva un aumento delle prestazioni del codice pre-esistente anche senza pesanti ottimizzazioni. Ora invece siamo di fronte a una nuova sfida che vede il programmatore stimolato a fare un salto innovativo nello studio di nuovi algoritmi e di metodologie capaci di sfruttare adeguatamente le nuove architetture parallele per risolvere problemi non pensabili anche solo un po' di anni fa, nei campi applicativi più diversi.
Di seguito sottolineiamo alcuni aspetti che occorre affrontare in questo processo innovativo:

- È importante semplificare il processo di sviluppo di applicazioni in grado di raggiungere alte prestazioni sulle nuove architetture, e sviluppare nuovi approcci e modelli di programmazione sufficientemente flessibili e adattivi per le nuove architetture. In questo senso si muove il progetto DARPA per la realizzazione del Petaflop/s computer entro il 2010, che prevede anche un'attività di ricerca e sviluppo per la realizzazione di un linguaggio parallelo per le scienze computazionali. Cray, IBM e SUN in collaborazione con partner accademici stanno rispondendo a questa richiesta con lo sviluppo di 3 nuovi linguaggi paralleli, chiamati rispettivamente Chapel [13], X10[17] e Fortress[18]. I 3 linguaggi, di tipo *object-oriented*, si basano tutti sul modello

[15] http://www.highproductivity.org

[16] Graphical Processing Unit. Il modello G80 prodotto da NVIDIA ha 128 processing units ed è in grado di raggiungere 500 GFlop/s.

[17] http://www.research.ibm.com/x10/

[18] http://research.sun.com/projects/plrg/faq/index.html

di programmazione parallela dinamico con *Partitioned Global Address Space* (PGAS)[19] derivato da *Co-Array Fortran* (CAF)[20] e *Unified Parallel C* (UPC)[21]. È prevedibile che questi linguaggi si fondano in uno unico o, per lo meno, che si possa definire un supporto di basso livello comune.

❑ Anche se è presto per dire se l'introduzione di questi linguaggi avrà un impatto significativo e stabile sulla comunità scientifica, occorre predisporre un'adeguata attività di formazione e supporto che vada verso l'introduzione di metodologie *object oriented* e più scalabili di MPI se si vogliono realizzare applicazioni parallele per i supercomputer del prossimo futuro, caratterizzate da un'alta scalabilità. Uno strumento importante in questa direzione è rappresentato dal paradigma Charm++[22], una libreria parallela *object oriented* costruita sopra C++ e portabile su piattaforme eterogenee.

❑ Per ottenere un alto livello di performance sarà poi necessario cambiare la struttura interna delle applicazioni e sperimentare nuovi algoritmi che tengano conto di architetture specifiche e scalabili. Solo un numero limitato di metodologie numeriche sono fondamentali in ambito scientifico e ingegneristico: algebra lineare su matrici densea e sparse, metodi spettrali, metodi N-body, griglie strutturate e non strutturate, metodi Monte-Carlo. È importante predisporre metodologie scientifiche ottimizzate e che contemplino la scalabilità di questi metodi su sistemi con migliaia di processori.

❑ Infine va ricordato l'aspetto della *fault tolerance*: il *mean-time-to-failure* (MTTF) di sistemi con un numero molto elevato di processori diventa per forza di cose molto più basso rispetto ai sistemi HPC più tradizionali, diventa pertanto indispensabile pensare ad applicazioni in grado di gestire elementi di fault tolerance in modo semplice.

6. UNO SGUARDO ALLO SCENARIO EUROPEO

L'HPC nella sua accezione più avanzata (*capability computing*) si basa su sistemi caratterizzati da un elevato numero di processori, grande capacità di memoria e un'alta banda di interconnessione con una bassa latenza. La disponibilità di sistemi innovativi di questa classe è indispensabile per accrescere le competitività nelle diverse aree della scienza e della ricerca in Europa, tant'è che l'ESFRI[23], il Forum Strategico sulle Infrastrutture di Ricerca in Europa, recentemente, ha identificato l'HPC come una priorità strategica per l'Europa [14].

L'Unione europea ha fatto proprio questa priorità e nell'ambito del VII programma quadro (2007-2012) prevede di supportare l'installazione di alcuni sistemi HPC di classe Pflop/s (Tier 0). Tali sistemi verranno poi integrati con i sistemi HPC presenti nelle singole Nazioni; questi sistemi, che ci si aspetta essere dell'ordine delle centinaia di Tflop/s, costituiscono il cosiddetto Tier 1 e a loro volta saranno integrati con sistemi HPC locali o regionali, meno potenti (Tier 2), secondo un modello a piramide che definisce un preciso eco-sistema HPC a supporto della comunità scientifica europea. Il progetto PRACE[24] (*Partnership for Advanced Computing in Europe*) finanziato dalla Unione europea ha l'obiettivo d studiare la creazione di tale sistema HPC pan-europeo e individuare alcuni centri di classe Pflop/s. Questi sistemi rappresenteranno il cosiddetto Tier 0 dell'eco-sistema HPC europeo.

Invece il progetto DEISA[25] (*Distributed European Infrastructure for Supercomputing Applications*), sempre finanziato dalla Comunità europea, ha l'obiettivo di consolidare il Tier 1 tramite l'integrazione degli 11 maggiori centri HPC a livello europeo.

[19] http://www.cs.berkeley.edu/~kamil/papers/pasco07.pdf

[20] http://www.hipersoft.rice.edu/caf/

[21] http://upc.lbl.gov/

[22] http://charm.cs.uiuc.edu/

[23] http://cordis.europa.eu/esfri/

[24] http://www.csc.fi/english/pages/prace

[25] http://www.deisa.org

DEISA gestisce un'iniziativa di *Estreme Computing* che consiste nell'erogare importanti quantità di risorse computazionali e di supporto specialistico alle comunità scientifiche dei Paesi che partecipano al progetto. Tali risorse consentono di realizzare ricerche di punta con un grosso impatto innovativo ed eccellenza scientifica e difficili da realizzare senza l'infrastruttura DEISA.
Infine va sottolineato HPC-Europa[26], un altro progetto europeo che supporta l'accesso dei ricercatori a sei tra le maggiori infrastrutture di supercalcolo a livello europeo, con l'obiettivo di fornire in modo integrato il supporto ai ricercatori coinvolti nelle attività computazionali che necessitano di importanti servizi HPC. Il servizio viene erogato a largo spettro sia in termini di accesso alle infrastrutture HPC europee, ma anche in termini di fruizione di ambienti computazionali avanzati per consentire ai ricercatori europei di rimanere competitivi con gli altri gruppi di ricerca a livello mondiale.
È fondamentale che anche l'Italia come sistema Paese possa giocare un ruolo importante in questo contesto, predisponendo un ambiente di supercalcolo in linea con quelli che le diverse nazioni europee stanno predisponendo a supporto delle loro comunità scientifiche. Sistemi HPC con potenze dell'ordine delle centinaia di Tflop/s per sistema sono già disponibili (o lo diverranno nel corso del 2008) in Francia, Spagna, Inghilterra, Germania, Norvegia, Svezia e Finlandia. Solo l'integrazione di sistemi a questo livello di potenza, e che evolvono nel tempo, potrà contribuire in modo decisivo a innalzare il livello competitivo degli scienziati europei.

7. CONCLUSIONI

L'evoluzione della ricerca scientifica comporta l'utilizzo di strumenti computazionali complessi che si avvalgono di sistemi di elaborazione avanzati ed in rapida evoluzione. Dispositivi per la memorizzazione, l'archiviazione e l'analisi di grosse mole di dati e strumenti innovativi per la visualizzazione dei risultati sono il necessario complemento ai sistemi per il calcolo ad alte prestazioni se si vogliono affrontare in modo adeguato i problemi di frontiera posti oggi dalla scienza, ma anche dalla vita di tutti i giorni.
Il miglioramento delle prestazioni dei sistemi di supercalcolo si ottiene ormai con l'integrazione di più processori sullo stesso chip. Già oggi i cluster HPC impiegano componenti *quad-core* e la presenza di sistemi *multi-core* (con decine di processori per chip) diverrà sempre più concreta nell'immediato futuro. Inoltre, una nuova classe di architetture ad elevate prestazioni è in fase di studio con l'obiettivo di aumentarne l'efficienza e la scalabilità. Tali architetture sfruttano il computing eterogeneo, integrando processori con funzionalità differenti come FPGAs, *Processing Unit Grafiche* (GPU), acceleratori SIMD e quant'altro, per minimizzare il tempo di esecuzione. I programmatori HPC e i ricercatori computazionali sono quindi chiamati a un salto innovativo che comporta l'utilizzo di nuovi algoritmi, paradigmi di programmazione parallela efficienti, librerie e strumenti di supporto per lo sfruttamento ottimale di architetture *multi-core* ed eterogenee altamente parallele.
Già oggi sono fruibili supercomputer con una potenza del centinaio di Tflop/s e, entro il 2010, la disponibilità di sistemi capaci di raggiungere il Pflop/s sarà reale.
Le maggiori potenze industriali nel mondo, Stati Uniti, Estremo Oriente, Europa, stanno concentrando i loro sforzi tecnologici ed economici per garantire al loro sistema di ricerca e sviluppo la disponibilità di infrastrutture per il calcolo in grado di reggere il confronto competitivo.
Anche a livello italiano è di fondamentale importanza, da parte del sistema Paese, sostenere i ricercatori nei loro bisogni computazionali avanzati, sia in termini di infrastrutture HPC, in linea con quelle degli altri Paesi europei, che di supporto e di competenze. Solo in questo modo si potranno affrontare e risolvere le sfide scientifiche del nostro tempo, per le quali esistono in Italia capacità e potenzialità di indubbio valore, e generare nuova conoscenza in grado di attraversare i confini disciplinari tradizionali.

[26] http://www.hpc-europa.org

Bibliografia

[1] Pople John A.: *Quantum Chemical Models, Nobel Lecture*. December 8, 1998.

[2] Grama A., Gupta A., Karypis G., Kumar V.: *An Introduction to Parallel Computing*. Design and Analysis of Algorithms, 2nd edition, Addison-Wesley, 2003.

[3] Dongarra J., Fostr I., Fox G., Gropp W., Kennedy K., Torczon L., White A.: *Sourcebook of Parallel Computing*. Morgan Kaufmann Publishers, 2003.

[4] Neri A., Esposti Ongaro T., Menconi G., De' Michieli Vitturi M., Cavazzoni C., Erbacci G., Baxter P.J.: *4D simulation of explosive eruption dynamics at Vesuvius*. Geophys. Res. Lett., 34, L04309, doi:10.1029/2006GL028597, 2007.

[5] Esposti Ongaro T., Neri A., Cavazzoni C., Erbacci G., Salvetti M.V.: Parallel multiphase flow code for the 3D simulation of explosive volcanic eruptions. *Parallel Computing*, Vol. 33, August 2007, p. 541-560.

[6] Pevzner P.A.: *Computational Molecular Biology. An algorithmic approach*. The MIT press, 2000.

[7] *The US National Virtual Observatory White Paper*. In Virtual Observatories of the Future, ASP Conference Proceedings, Vol. 225. Edited by Brunner R.J., Djorgovski S.G., Szalay A.S.. San Francisco: Astronomical Society of the Pacific, ISBN: 1-58381-057-9, 2001.

[8] Hoffmann E.: Evoluzione e prospettive nell'High Performance Computing. *Mondo Digitale*, n. 1, marzo 2003, p. 51-65.

[9] Flynn M.J.; Some computer organizations and their effectiveness. *IEEE Transaction on Computers*, Vol. C-21, n. 9, 1972.

[10] Hennessy J., Patterson D.: *Computer architecture: A quantitative approach*. 4th edition, Morgan Kauffman, 2007.

[11] Quinn M.J.: *Parallel Programming in C with MPI and OpenMP*. Mc Graw Hill, 2004.

[12] Mehta H. (Guest Editor): IBM POWER 6 Microprocessor Technology. *IBM Journal of Research and Development*, Vol. 51, November, 2007.

[13] Chamberlain B.L., Callahan D., Zima H.P.: Parallel Programmability and the Chapel Language. *International Journal of High Performance Computing Applications*, Vol. 21, n. 3, 291-312, DOI: 10.1177/1094342007078442, 2007.

[14] *ESFRI European Roadmap for Research Infrastructures*. Report 2006; European Community, Luxembourg, 2006, ISBN 92-79-02694-1.

Giovanni Erbacci si è laureato in Informatica presso l'Università di Pisa, dal 1999 coordina il gruppo Supercalcolo del CINECA e partecipa a diversi progetti europei nel settore dell'HPC. E' professore a contratto di architetture e programmazione parallela presso l'Università di Ferrara. I suoi interessi principali riguardano le architetture parallele, lo studio e la realizzazione di algoritmi paralleli efficienti, nonché la valutazione delle prestazioni dei sistemi e dei programmi paralleli. Giovanni Erbacci è autore o co-autore di oltre 40 articoli pubblicati su riviste e atti di convegni, ed è membro dell'ACM.
E-mail: g.erbacci@cineca.it