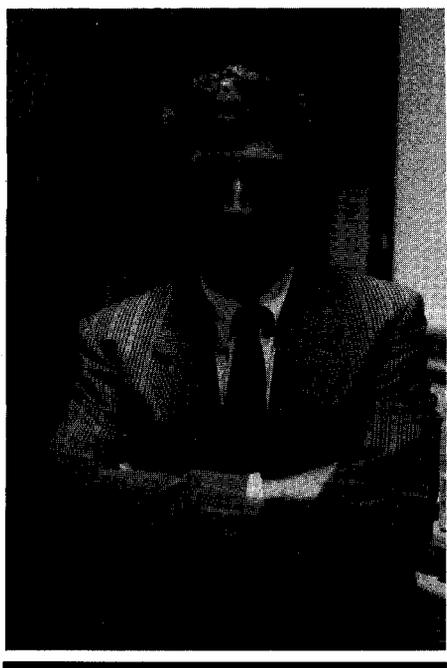


CALCOLO E SUPERCALCOLO CHIMICO IN ITALIA

CLAUDIO ZANNONI

Dipartimento di Chimica fisica ed inorganica, Università di Bologna



Claudio Zannoni è professore ordinario di Chimica fisica organica presso l'Università di Bologna. Si occupa di cristalli liquidi e di simulazioni al calcolatore. Ha fatto e fa parte dell'Editorial Board di varie riviste internazionali: *Molecular Physics, UK* (1984-1992), *Liquid Crystals, UK* (1986-1991), *J. of Fluorescence, USA* (dal 1992) e dell'*International Advisory Board* di *J. Chem. Soc., Faraday Transactions, UK* (dal 1993). È membro del Direttivo del Consorzio Interuniversitario Nazionale per la Chimica dei Materiali (INCM) e ha fatto parte (dal 1992) del Comitato di Consulenza ("Chemistry Panel of Experts") della Comunità Europea che giudica i progetti di ricerca presentati nell'ambito del progetto *Human Capital and Mobility*.

Nel pensiero dell'uomo della strada e forse anche di molti politici l'attività di calcolo è probabilmente associata più con la Matematica, la Fisica o l'Ingegneria di quanto non sia con la Chimica. In realtà le cose sono molto diverse e i chimici sono fra i più grossi utilizzatori di risorse di calcolo. Per fare un esempio le percentuali di utilizzo presso il più grande centro di calcolo italiano, il CINECA, sono: Fisica 55%; Chimica 25%; Ingegneria 10%; Astrofisica 4%; Matematica 2%. Queste percentuali variano fra i vari centri ma la presenza in prima o seconda posizione della Chimica è una costante in Italia e all'estero. Come e perché viene utilizzata questa enorme quantità di tempo calcolo? Quali sono i problemi nell'utilizzo e quale è la situazione italiana in confronto con quella degli altri paesi? Cosa è cambiato oggi rispetto al passato nell'uso dei calcolatori da parte dei chimici? Quest'ultima domanda è quella a cui è più facile rispondere. Lo sviluppo dei calcolatori, soprattutto workstation RISC (Reduced Instruction Set Computers) con istruzioni semplificate e particolarmente veloci da un lato e supercalcolatori paralleli dall'altro (vedi riquadro 1), ha cambiato lo scenario di quanto diventa fattibile come routine e di quanto è ora indispensabile per produrre ricerca competitiva. Cominciamo esaminando in quali grandi aree è concentrato l'utilizzo.

La prima è quella tradizionale della *chimica quantistica* con calcoli della geometria molecolare, dello stato elettronico fondamentale e in genere delle proprietà di singola molecola. In questo campo programmi di calcolo imponenti (molte centinaia di migliaia di righe di linguaggio FORTRAN) come Gaussian92 (Gaussian Inc.) sono stati messi a punto negli ultimi venticinque anni e sono ora disponibili per la comunità chimica anche presso i grossi centri di calcolo (vedi riquadro 2). Per molecole fino a circa cinquanta elettroni il calcolo della funzione d'onda monoelettronica dello stato fondamentale può richiedere tempi dell'ordine di una decina di minuti su un supercalcolatore e di un paio di ore su una workstation. Questo significa che è possibile effettuare calcoli a partire da principi primi *ab-initio* su molecole come il benzene usando workstation quando fino a cinque-dieci anni fa questi calcoli potevano solo essere effettuati presso grossi centri. Il problema è che il tempo di calcolo con metodi *ab-initio* varia in maniera non lineare col numero di elettroni o meglio col numero di funzioni di base usate per approssimare gli orbitali, per cui molecole complicate o di interesse biologico richiedono ancora risorse straordinarie ovvero l'uso di metodi approssimati semiempirici. Lo stesso avviene non appena si introduce il calcolo di proprietà magnetiche, ottiche, lineari e non, e di altre caratteristiche che sono d'altronde indispensabili per la previsione di proprietà importanti per molecole di cui si studia un potenziale utilizzo nell'ambito dei materiali innovativi. Naturalmente è sempre possibile effettuare su workstation calcoli semiempirici e in genere calcoli anche con metodi più moderni (es. funzionale densità) che non dipendano troppo drasticamente dal numero di elettroni. Esistono programmi, anche commerciali, dedicati al *molecular modeling*, utilizzati anche in campo farmaceutico. Per semplici molecole questi calcoli si possono ora effettuare su Personal Computer. Una possibilità è anche quella di utilizzare una workstation locale come interfaccia ai programmi situati presso il grosso centro. Questo è reso possibile dalla diffusione di *reti di trasmissione dati* che, anche in Italia, collegano almeno le principali città e la maggior parte dei centri di ricerca (v. riquadro 3). La disponibilità di reti veloci è destinata a rendere irrilevante la collocazione fisica dei centri di calcolo. Lo sviluppo di reti di trasmissione dati più efficienti è l'oggetto, negli USA, del progetto del vicepresidente Gore di "autostrade elettroniche" ("electronic data highways") ed è destinato a giocare, anche in Europa (vedi Quarto Programma Quadro della

CEE), un ruolo simile a quello che le autostrade hanno giocato nel facilitare il trasporto di persone e merci.

In sintesi, in questo settore i programmi di calcolo esistono, sono disponibili e sono sempre più utilizzati anche dal chimico non specialista. Il problema di calcolo per molecole piccole non esiste sostanzialmente più, per le grandi si tratta di disporre di risorse di supercalcolo.

La seconda area, più recente e forse ancora poco familiare in chimica, è quella delle *simulazioni al calcolatore*. In questo caso il problema non è lo studio della singola molecola ma di un insieme di molecole abbastanza numeroso da poter riprodurre le proprietà macroscopiche di un liquido o di un solido e le sue transizioni di fase. Non si tratta quindi più di andare in dettaglio sulla molecola isolata studiata in chimica quantistica (che potremmo considerare da un certo punto di vista in fase gassosa e allo zero assoluto) ma di modellare il comportamento di liquidi e di materiali organizzati a partire da più o meno semplici interazioni intermolecolari. La molecola viene rappresentata come un insieme di centri corrispondenti ad atomi o a gruppi di atomi, un po' come in un modello molecolare e l'interazione fra molecola e molecola diventa una somma di contributi provenienti dalle varie interazioni di gruppo. La simulazione del sistema molecolare all'equilibrio è realizzata con due tecniche base: metodo Monte Carlo (MC) e di dinamica molecolare (MD). Il primo serve a calcolare proprietà statiche di equilibrio mediando su configurazioni di equilibrio oppor-

tunamente generate. Il metodo MD segue invece l'evoluzione temporale del sistema integrando passo passo l'equazione del moto per tutte le molecole. È oggi praticamente possibile seguire traiettorie di sistemi dell'ordine di centinaia di picosecondi, mentre sarebbe auspicabile poter seguire fenomeni dinamici per tempi almeno cento volte più lunghi. Per ambedue le tecniche i sistemi che si riescono a studiare oggi su una workstation sono dell'ordine di alcune migliaia di centri atomici. Questi possono rappresentare di volta in volta e a seconda del problema un campione di qualche migliaio di molecole semplici (ad es. di gas raro) o una piccola proteina. Naturalmente una molecola isolata non passa da liquido a solido o a gas e per studiare transizioni di fase, ad esempio la fusione o la vaporizzazione di un liquido, è indispensabile disporre almeno di qualche centinaio di molecole. Si capisce subito che il limite ora è nel numero di molecole che si possono studiare. Per fare un esempio personale recente, una simulazione delle transizioni di fase in un insieme di mille molecole modello di cristallo liquido (R. Berardi, A. Emerson, C. Zannoni, *J. Chem. Soc. Faraday Trans.*, 1993, **89**, 4069) ha richiesto tempi di calcolo di circa settemila ore di workstation HP720 che possiamo stimare nell'ordine di quasi un migliaio di ore di supercalcolatore CRAY Y-MP). È facile immaginare che i tempi richiesti per le pur oggi possibili simulazioni di polimeri o di sistemi biologici importanti, come le proteine, richiedono il supercalcolo. Dobbiamo anche ricordare che

ALCUNE CARATTERISTICHE DI ELABORATORI PROTOTIPO

Riportiamo alcuni dati tipici di classi di calcolatori, con l'intento di dare una sensazione dell'incremento di potenzialità negli ultimi dieci anni. Non intendiamo ovviamente fornire esempi troppo specifici sia per le difficoltà di avere dati veramente omogenei sia perché per confronti esatti si dovrebbero specificare modelli commerciali, versioni e configurazioni esatte, ecc. col rischio di perdere il quadro generale che vogliamo offrire.

Ricordiamo che le unità di misura per le prestazioni sono normalmente date dal numero di operazioni con virgola mobile al secondo (flops) e che si parla quindi di megaflops (Mflops) e gigaflops (Gflops). Le prestazioni qui citate vengono riferite alla soluzione di sistemi di equazioni lineari utilizzando un programma di riferimento "benchmark", LINPACK, messo a punto da J.J. Dongarra (Oak Ridge National Laboratory USA). La memoria è espressa in byte (1 byte è costituito da 8 bit) o più comunemente oggi in milioni o miliardi di byte, cioè megabyte (Mb) e gigabyte (Gb).

VAX 11-780

È il calcolatore prototipo della passata generazione (anni 80), riportato qui per confronto. Il DEC VAX 11-780 richiedeva una stanza condizionata, aveva memoria di qualche Mb, dischi fissi da 500 Mb più o meno delle dimensioni di una lavatrice (!), velocità dell'ordine di 0.2 Mflops. Sistema operativo: VMS

Personal Computer (PC)

Un comune PC486 ha 2-16 Mb di memoria centrale, clock tipicamente da 40 a 66 MHz e disco rigido da 100-500 Mb. Le prestazioni, con coprocessore matematico, sono dell'ordine di 1 Mflop. Sistemi operativi: MS DOS, Windows, UNIX

Workstation RISC

Caratteristiche principali: Processore RISC, 32-128 Mb di memoria, dischi dell'ordine del Gb, velocità dell'ordine di 10-70 Mflops e oltre. Dimensioni simili a quelle di un PC. Sistema operativo: UNIX. Possono essere collegate in rete ethernet a "bassa velocità" (circa 1 Mbyte/sec) o in maniera più veloce con tecnologia FDDI (Fiber Distributed Data Interface) (circa 12 Mbyte/sec). Gruppi di workstation in rete possono lavorare congiuntamente (in "cluster") come una unica macchina virtuale utilizzando vari tipi di software che consentono di distribuire il calcolo, ad esempio, scambiando opportuni messaggi fra programmi in esecuzione sulle varie macchine "message passing". Fra questi sistemi quello che sta diventando uno standard di fatto è il PVM (Parallel Virtual Machine) sviluppato a Oak Ridge e disponibile liberamente, praticamente su tutti gli elaboratori attuali. Una caratteristica interessante è quella di poter far lavorare insieme anche calcolatori diversi ottenendo così un ambiente eterogeneo.

Supercomputer

La definizione di Supercomputer in termini assoluti non è semplice, vista la continua evoluzione delle prestazioni. In questo momento si intendono macchine con velocità di picco di almeno un 1 Gflop. Nel valutare le prestazioni non si deve commettere l'errore di scambiare queste cifre, che riguardano il limite di performance, con quelle ottenibili normalmente o con facilità. Molte macchine montano un numero anche elevato di processori e si avvicinano alla prestazione di picco solo quando elaborano in parallelo sulle varie componenti. I supercalcolatori più potenti sono oggi "massicciamente paralleli", con un numero di processori anche dell'ordine di molte centinaia. Quindi per un programma normale, non particolarmente ottimizzato per la macchina, le prestazioni possono essere assai inferiori. A titolo di esempio il supercomputer Cray T3D con versione con 256 processori raggiunge nei test di utilizzo prestazioni dell'ordine del 5% della potenza teorica (Saunders, 1994). È quindi particolarmente importante investire nello sviluppo di una nuova generazione di programmi di calcolo che sfruttino a pieno l'architettura parallela.

è oggi possibile accoppiare meccanica quantistica e simulazioni, ad esempio con le tecniche MD di Car-Parrinello, con tutta una serie di implicazioni affascinanti per il calcolo della struttura elettronica di materiali anche complessi.

Un'altra serie di applicazioni importanti e in via di sviluppo delle simulazioni è quella di ausilio alla interpretazione di spettri NMR di sistemi complicati come macromolecole o enzimi, contribuendo a ottenerne la struttura in fase fluida. È chiaro anche che esiste tutta una serie di problemi, che sono di interesse non solo fondamentale ma anche industriale. Un esempio è la simulazione del comportamento dei liquidi in situazioni estreme di temperatura e pressione o confinati in materiali porosi o zeoliti. Altri esempi sono la simulazione di effetti solvente, dell'auto assemblaggio (micelle, vescicole, ecc.) in miscele di oli-acqua-surfattante e, in un futuro non lontano, di reazioni in fase liquida. Non a caso in altri paesi come l'Olanda e l'Inghilterra molta ricerca in questo settore è sviluppata o finanziata dall'industria.

I problemi di calcolo a cui abbiamo accennato non sono ovviamente solo italiani! D'altra parte la ricerca in questo settore è una attività competitiva in cui i chimici italiani si trovano a confrontarsi con colleghi di altri paesi avanzati in Europa e negli USA per cui quanto si riesce a fare in Italia (e in Europa) deve essere magari non sovrabbondante quantitativamente ma di qualità eccellente in senso assolu-

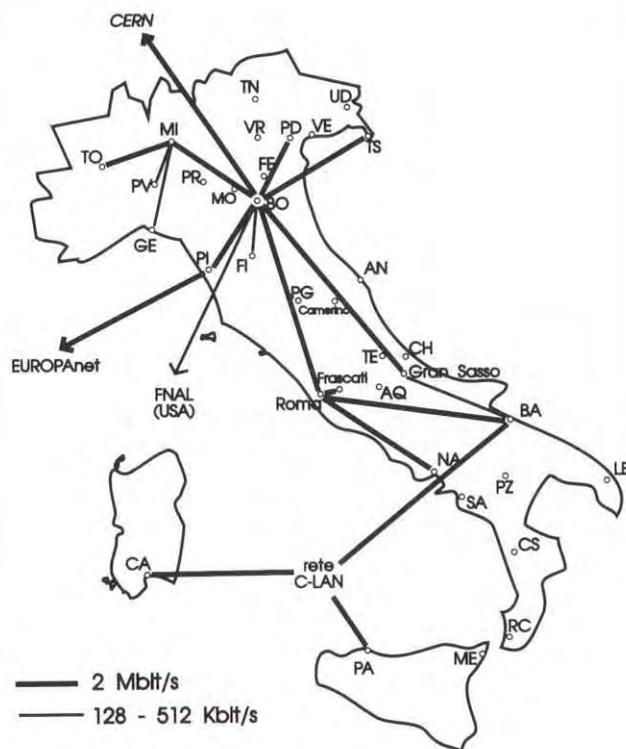
to. Dal punto di vista della competitività il settore della chimica computazionale offre al ricercatore italiano alcune prospettive interessanti e molte difficoltà. Le prospettive vengono dal buon livello della Scuola italiana e dalle possibilità di impiantare un laboratorio di calcolo competitivo con risorse, che non siano quelle destinate alla strumentazione stessa, relativamente limitate. A differenza di un laboratorio chimico o di una grossa strumentazione (es. NMR e spettrometria di massa) che richiedono locali adeguati, il rispetto di una serie di misure di sicurezza, approvvigionamenti ricorrenti ecc., nel nostro caso, potendo acquisire la strumentazione di calcolo si richiede solo una presa di corrente e un collegamento alla rete di comunicazione per arrivare al grosso centro calcolo. Aggiungendo una certa dose di capacità "computeristica", che fortunatamente non difetta neanche da noi, si può concretamente creare una struttura efficiente anche in un intorno relativamente spartano. Da questo punto di vista la tecnologia può fornire a un paese con infrastrutture carenti come l'Italia un vantaggio aggiunto. È un po' l'"effetto fax", per cui la possibilità di comunicare istantaneamente offerta dal fax porta un vantaggio maggiore a un paese con un sistema postale scadente, rispetto a paesi in cui le comunicazioni erano già in precedenza accettabili. Quindi per problemi di livello medio o medio alto dovrebbe essere favorita la diffusione di workstation, possibilmente dotando i gruppi di un minimo di

PRINCIPALI CENTRI DI CALCOLO IN ITALIA

<p>Elenchiamo brevemente i centri di calcolo più importanti a disposizione anche della comunità chimica. Non vengono riportati quindi centri di calcolo industriali o specializzati in altre attività (metereologia, astronomia ecc.). Il parco macchine installato varia continuamente e ci scusiamo quindi fin da ora per eventuali imprecisioni. La potenza delle macchine (cfr. riquadro 1) è riportata, quando disponibile.</p> <ul style="list-style-type: none"> ● CINECA, Consorzio Interuniversitario dell'Italia Nord Orientale, Bologna. Il più importante centro di calcolo italiano (Consorzio di 13 Università: Ancona, Bologna, Catania, Ferrara, Firenze, Modena, Padova, Parma, Siena, Trento, Trieste, Udine, Venezia). Dotato fino al 1993 di un Cray Y-MP8/464, 4 processori con 0.33 Gflops di picco per processore, citato nel testo. Dispone ora di: - Cray C94/2128 con potenza totale di 2 Gflops di picco - Cray T3D con 64 nodi, da 9.6 Gflops di picco (dal luglio 94) - IBM 9021/640 0.83 Gflops - DEC VAX 6000/510 0.05 Gflop - cluster di workstation HP 735 (INFN) - SGI Challenge - Utilizzo: Fisica 55%; Chimica 25 %; Ingegneria 10 %; Astrofisica 4%; Matematica 2%, - Utenza: 68% Università, Laboratori di ricerca 27%; Industrie 5% N.B.: Il CINECA è caratterizzato da un uso molto elevato da parte dei fisici per l'analisi di esperimenti di fisica delle alte energie eseguiti al CERN. Le sue statistiche di utilizzo per discipline sono quindi distorte da questo fatto. ● CNUCE, Centro Nazionale Universitario di Calcolo Elettronico, Istituto CNR, Pisa. - IBM 9121 VF 440 - IBM SP-1 (8 processori, 1Gflop di picco) che passerà a SP-2 entro il 1994 - nCUBE 2 - 6400 con 128 nodi (0.3 Gflops). L'utilizzo di questa macchina parallela (1993) è stato (mar - dic '93) : Matematica 42%, Chimica 36 %, Ingegneria 22%, Fisica 3% -varie workstation RISC IBM, Sun, DEC, HP ● CILEA, Consorzio Interuniversitario Lombardo per la Elaborazione Automatica, Milano. 7 università lombarde: Università di Milano, Politecnico di Milano, Università Cattolica 	<p>del S. Cuore, Bocconi, Pavia, Brescia e Bergamo - IBM: 3090/200 VF. - Unisys: 2200/402 - Convex: C3820 - Meta Convex parallelo su 8 ws HP735 - DEC VAX 6410</p> <ul style="list-style-type: none"> ● CRS4, Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna, Cagliari - IBM 9121-511 - Numerose workstation: 1 IBM 7013-950, 20 IBM 7013-550, 3 IBM 7013-560 e 3 DEC AXP-500 - 1TMC CM-200/8K, 1 TMC CM-5R 1IBM SP1/32, Parsytec GCel 1/32 (macchine parallele) - 1 HP 735 ● CASPUR, Consorzio per le Applicazioni di Supercalcolo per Università e Ricerca, Roma. - IBM 3090/600J VF (6 processori) da 0.8 Gflops di picco. - Cluster di 8 workstation DEC Alpha 3000/500 collegate in FDDI - Quadrics Q4 (APE) con 32 processori. Supercalcolatore parallelo ad architettura SIMD (Single Instruction Multiple Data) progettato e costruito in Italia (da INFN e ora da Alenia Spazio). Potenza di picco 1.6 Gflops. Inoltre Quadrics Q1 con 8 processori. - Varie workstation IBM, DEC, SUN ● CSATA, Tecnopolis, Bari - IBM ES9000/500 VF - DEC VAX 3900 - HP 9000/82 - Convex 4 processori - Cluster di 8 workstation HP - Meiko Surface (calcolatore parallelo) ● CSP, Centro Supercalcolo Piemonte, Torino - Cray Y-MP 2E/116 - 2 DEC VAX 6510 ● Università di Trieste - CRAY XMP/116SE ● CUC, Università di Palermo - IBM 3090/20J - DEC VAX 6510 ● CCE, Centro di Calcolo Elettronico dell'Università di Perugia, Perugia - IBM 3090 VF - Workstation IBM: 2 RISC/ 6000 530H e 560 <p>Significato delle sigle di ditte produttrici: Cray: CRAY Research; DEC: Digital Equipment Corporation; IBM: International Business Machines; HP: Hewlett Packard, SGI: Silicon Graphics, TMC: Thinking Machines Corporation</p>
--	---

Schema della rete GARR
(Gruppo Armonizzazione Reti Ricerca,
costituito da CILEA, CINECA, CNR CNUCE, CSATA,
ENEA e INFN)

La rete, inaugurata nel 1990, collega i principali centri italiani fra loro e alle reti internazionali, come mostrato in figura, dove viene anche indicata la velocità di trasmissione. Sono indicate anche le sedi collegate in rete a velocità più bassa. Bologna consiste in due nodi primari (CNAF e CINECA). Altre sedi hanno più di un nodo. La rete è collegata con il resto d'Europa e con gli USA (è indicato il Fermi Lab). In pratica è oggi possibile per la grande maggioranza dei chimici nelle Università italiane collegarsi a banche dati ad es. (consultare Chemical Abstracts), comunicare attraverso posta elettronica o Internet con colleghi in tutto il mondo.



Riquadro 3

unità di personale per la gestione.

Per i problemi che richiedono il supercalcolo ("superproblemi"), in Italia esistono fortunatamente alcuni grossi centri di calcolo e di competenze. In particolare almeno uno di questi, il CINECA si può considerare adeguato alle esigenze attuali. Ad esempio il CINECA è ora dotato di un Cray C94, che sarà affiancato fra breve da quello che rappresenta ora lo stato dell'arte del supercalcolo: un Cray T3D. È da notare che macchine di questa classe stanno per essere installate più o meno contemporaneamente anche in Gran Bretagna (Edimburgo) e in altri paesi europei. A fronte di questa disponibilità potenziale di risorse va comunque sottolineato che, a differenza dei colleghi di altri paesi, il chimico computazionale italiano deve acquistare il tempo di calcolo presso i centri. Mentre a prima vista questo può sembrare ragionevole, non lo è se pensiamo che i supercalcolatori sono un po' la "formula uno" del calcolo. Quindi far pagare il supercalcolo ai ricercatori che sono in grado di formulare problemi competitivi in senso assoluto non è più sensato che far pagare il biglietto ai piloti di formula uno! In altri paesi, in Inghilterra in particolare, il supercalcolatore è visto come una risorsa a disposizione della comunità e che deve essere utilizzato a pieno per essere veramente di beneficio al paese. I ricercatori presentano progetti che richiedono l'uso di questa risorsa e una commissione valuta e assegna il tempo calcolo (non fondi, sempre insufficienti ad acquistare il tempo calcolo.) Il problema è ancora più acuto perché se un progetto richiede veramente un supercalcolatore, deve ovviamente trattarsi di un progetto di grandi dimensioni e che non ha senso finanziare solo in piccola parte. In questo caso fornendo poche ore di tempo calcolo su un supercalcolatore a tutti invece che molte centi-

naia a progetti attentamente scelti, non si mette nessuno in condizioni di utilizzare la macchina per i problemi per cui è più appropriata (non serve far fare un giretto con la formula uno a tutti, per restare all'esempio di prima).

È anche opportuno ricordare che le risorse di supercalcolo non servono in molti casi a velocizzare la soluzione di un problema comunque affrontabile, ma a renderne praticabile la soluzione. D'altra parte, come si è già detto, il supercalcolatore non è una soluzione universale. Le linee guida di un intervento nel settore potrebbero essere articolate su tre punti:

1. Mettere a disposizione di un numero sempre più elevato di laboratori workstation RISC che possono coprire tutte le esigenze "normali". Favorire la messa in cluster di queste e lo studio di una nuova generazione di codici che sfruttino in parallelo la potenza aggregata di queste macchine (v. riquadro). Questo approccio consentirà di disporre di una struttura con potenza simile ai supercalcolatori di ieri a un costo limitato.
2. Favorire lo sviluppo di reti di calcolo veloci. Queste autostrade dati consentiranno sempre più anche l'uso dei calcolatori per accedere a grandi banche dati e/o a strumentazioni remote, consentendo anche una specializzazione dei centri di supercalcolo per tematiche o aree disciplinari invece che una copertura universale e necessariamente incompleta di esigenze geograficamente localizzate.
3. Fornire risorse di supercalcolo alla comunità chimica a livello nazionale o europeo.

In sostanza si tratta di scegliere, magari con opportune rotazioni, invece di distribuire a pioggia e si tratta di mettere anche i ricercatori italiani in grado di essere competitivi.