



UNIVERSITÀ FEDERICO II DI NAPOLI

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

Dottorato di Ricerca in Scienze Chimiche
XX Ciclo

NUOVE STRATEGIE COMPUTAZIONALI
PER LO STUDIO DI PROPRIETÀ CONFORMAZIONALI
E VIBRAZIONALI DI SISTEMI MOLECOLARI
DI INTERESSE BIOLOGICO

Giovanni Morelli

Tutore

Prof. Orlando Crescenzi

Coordinatore

Prof. Aldo Vitagliano

Anno Accademico 2007-2008

Abstract for:

New computational strategies to study conformational and vibrational properties of biological molecular systems

Giovanni Morelli

PhD Thesis

The main goal of this PhD thesis is to develop new computational strategies to study the conformational and vibrational properties of biological systems.

As a matter of fact, the definition of reliable computational approaches based on ab-initio calculations to study complex biological systems is, by far, one of the most intriguing topic because concentrate chemical problems, architectural solutions and performances as well as software code optimizations. For the sake of readability, I separated the problems in five Chapters and one Appendix.

In the Chapter 1, a reliable and computationally efficient approach has been proposed for the theoretical evaluation of spectroscopic observables of molecules in solution. Such a challenging problem requires that solvent effects and thermal fluctuations of the molecular conformations should be properly taken into account. Our procedure has been schematically divided into three successive steps: 1) the parameterization of an effective classical force field for the solute intramolecular and solute/solvent intermolecular interactions, 2) the sampling of the configurational space of the system via Molecular Dynamics (MD) simulations using the mean field (MF) model, and 3) the quantum mechanic (QM) computation of spectroscopic parameters on a representative set of molecular configurations. This computational approach has been illustrated by considering the acetone molecule in aqueous solution as a test system, which has been extensively studied both experimentally and theoretically by different groups. First, we have shown that classical MD based on a reliable force field compares favorably with Car-Parrinello Molecular Dynamics (CPMD) simulations, as shown by acetone/water radial distribution functions (RDFs) and hydrogen-bond distributions, and can be used effectively to obtain extended sampling of the solute/solvent configurations. In addition, it was necessary to modify the standard OPLS-AA force field for acetone by including oxygen lone pairs and to improve some geometrical features, such as the C=O stretching and out-of plane mode, by an accurate reparameterization. Indeed, force field-based MD could provide more reliable molecular geometries, and indirectly spectroscopic properties as seen for ^{13}C and ^{17}O NMR shielding constants, with respect to standard CPMD. Besides, the MF model has been shown to be an effective methodology for simulating solute/solvent systems, and is particularly well-suited to post ab initio calculations of spectroscopic parameters. Finally, it was shown how a mixed quantum mechanic/molecular mechanic (QM/MM) approach can accurately reproduce experimental spectro-

scopic observables of acetone, such as NMR shielding constants and $n \rightarrow \pi^*$ excitation energy, provided that an appropriate level of theory and a suitable number of solvent molecules are treated at the QM level along with the solute. The present approach has the advantage of being highly flexible concerning the QM model used for the post-MD calculations, and was found to be equivalent to more expensive full QM calculations as well as to QM calculations using the polarizable continuum model (PCM) model, as shown for the case of acetone ^{17}O NMR chemical shifts. Moreover, due to a much longer simulation timescale, our procedure has provided results with a significantly smaller statistical error for the computed UV transition energies than that for previous results based on full QM and QM/MM molecular dynamics simulations. The approach described in this Chapter is currently applied to study systematically the spectroscopic properties of a variety of molecular systems in solution, from small to large molecules.

In the Chapter 2, the excited electronic states of the anion radical of *para*-benzoquinone were studied by time dependent density functional theory (TD-DFT) including bulk solvent effects by the PCM model as a further example of applicability of multi-scale approach. The computed vertical excitation energies for the first four low-lying doublet states are in good agreement with previous post-Hartree–Fock computations. Geometry optimization of excited states and inclusion of solvent effects lead to a remarkable agreement between computed adiabatic transition energies and experimental band maxima. Together with their specific interest, the results point out the reliability of TD-DFT/PCM approach for valence excitations and the need to take geometry relaxation and solvent effects into the proper account for a meaningful comparison between computed and experimental absorption spectra.

In the Chapter 3 are sorted out the basis for the computation on distributed wide area network, well known as *grid computing*, which will be the technical solution adopted in the Chapter 4.

In fact, in the Chapter 4 is described the implementation of a quantum chemistry code on a service-oriented grid where the computational workflow is spawned over multiple, geographically distributed sites. The application porting over the grid and its extension as a web service over local and wide area networks is fully outlined. A description of the procedures developed for this experiment is given in full detail and the applicability of the methodology to similar codes in this scientific area is assessed. The new developed, grid-aware, application has been tested by performing a comprehensive set of benchmarks including quantum mechanical calculations on the water molecule. The obtained results of the benchmarks are reported and a full comparison with respect to the parallel execution of the same tests using the standard code is discussed in detail. Note that, thanks to the ongoing development of effective TD-DFT approaches and their analytical gradients both for isolated molecules and for condensed phases, the range of application of the above approaches is being extended to the evaluation of Franck–Condon factors and other vibronic couplings of paramount relevance in several branches of molecular spectroscopy. By using a semi-

classical approach, the model can be used also to compute reaction rates including anharmonic and tunnelling effects. A nice example of a such application is reported for the case of Zinc-porphyrin complexes to test the numerical stability of all procedures.

Last but not least, there is an important Appendix A, which accurately described the configuration of the Bionap and Avogadro cluster machines of the University of Naples and Padova, respectively, which I seted up during my PhD period.

Indice

Introduzione	4
1 Calcolo di proprietà spettroscopiche in soluzione	9
1.1 Introduzione	9
1.2 Il modello multiscala	10
1.3 Approccio integrato QM/MM per il calcolo di proprietà spettroscopiche in soluzione	12
1.4 Metodo e dettagli computazionali	15
1.4.1 Parametri del campo di forza	15
1.4.2 Dinamica Molecolare Classica	18
1.4.3 Calcolo QM delle Proprietà Spettroscopiche	18
1.5 Risultati e Considerazioni	19
1.5.1 Parametrizzazione del campo di forza	19
1.5.2 Dinamica Molecolare	22
1.5.3 Parametri spettroscopici	26
2 Calcolo TD-DFT/PCM di proprietà spettroscopiche	33
2.1 Introduzione	33
2.2 Il metodo	34
2.3 Risultati	35
3 Il calcolo distribuito in chimica	44
3.1 Introduzione	44

INDICE

3.2	I cluster di calcolatori	46
3.3	Il grid computing	47
3.3.1	Grid: i servizi	50
3.3.2	Aspetti generali sulla sicurezza delle griglie	54
3.4	Il ciclo di vita di un job nella griglia	56
3.4.1	User Interface e Logging and Bookkeeping	57
3.4.2	Resource broker	58
4	Calcolo numerico di frequenze su griglia computazionale	60
4.1	Introduzione	60
4.2	GaussianGrid con la tecnologia WS	61
4.3	Distribuzione dei test paralleli ed analisi delle prestazioni	64
4.4	GaussianGrid presso <i>LSDM</i>	67
4.5	Algoritmo per il calcolo delle frequenze	68
4.6	Interfaccia web per GaussianGrid	70
4.7	Test sul complesso porfirina-zinco	73
5	Conclusioni	75
	Appendice	77
A	I “<i>nodr</i>” di Napoli e Padova	78
A.1	Breve panoramica sulle architetture di calcolatori	78
A.2	Il nodo di Napoli: il cluster “ <i>Bionap</i> ”	79
A.2.1	Collocazione dei componenti	81
A.2.2	Configurazione dello spazio disco	82
A.2.3	Configurazione di rete	85
A.2.4	Benchmarks MPI	85
A.2.5	Il software per il calcolo chimico	86
A.3	Il nodo di Padova: Il cluster “ <i>Avogadro</i> ”	89
A.3.1	Collocazione dei componenti	89

INDICE

A.3.2	Configurazione dello spazio disco sul nodo di management DL380	90
A.3.3	Configurazione dello spazio disco	91
A.3.3.1	Spazio disco locale sui nodi del cluster	91
A.3.3.2	Spazio disco condiviso	92
A.3.4	Backup e cloning con la CMU	92
A.3.5	Configurazione delle connessioni di rete	93
A.3.6	Installazione e configurazione dell'ambiente di sviluppo	95
A.3.7	Installazione del software per la chimica computazionale	95
A.3.8	Configurazione dell'ambiente utente	96
A.3.9	Benchmarks MPI	96
Bibliografia		100

Introduzione

Il principale obiettivo di questa tesi di dottorato è lo sviluppo di nuove strategie computazionali applicate a sistemi molecolari complessi di interesse biologico. In realtà, lo sviluppo di queste tecniche va ben oltre questo personale lavoro di tre anni ma si intreccia con uno degli scopi della chimica computazionale “tout court” vale a dire razionalizzare e/o predire proprietà macroscopiche di sistemi complessi partendo dalla determinazione della struttura a livello molecolare. Come vedremo nel prosieguo di questo lavoro affrontare questo problema significa scontrarsi/incontrarsi con campi disciplinari che possono apparire, almeno a prima vista, lontani da quelli più tradizionalmente consoni ai chimici. Infatti, malgrado la crescita esponenziale della potenza dei moderni calcolatori e lo sviluppo di software adeguati per la modellizzazione molecolare, restano ancora enormi aree di incertezza intorno a fenomeni “critici” per la costruzione di un quadro completo delle proprietà su “scala macroscopica”. Il problema fondamentale alla base di queste difficoltà è che le proprietà della materia a livello macroscopico sono intimamente connesse alle struttura elettronica. Tuttavia, i metodi computazionali in grado di affrontare uno studio sulle scale atomiche (spaziali e temporali) non sono assolutamente, o perlomeno non facilmente estendibili alla scala macroscopica. Il più credibile tentativo di colmare questo “gap dimensionale” è costituito dal *metodo multiscala*. In questo approccio si cerca di rendere i risultati ottenuti ad un livello di scala, per esempio atomico, accessibili al livello di scala superiore e così via, fino ad arrivare a livelli di

scala macroscopici. In questo modo si cerca di trasferire “step by step” l’informazione acquisita su scala elettronica verso quella macroscopica. Risulta evidente che la possibilità di “espandere” la trattazione ab-initio verso scale di tempi e lunghezze sempre maggiori, oppure trattare sistemi complessi con un sempre maggior numero di atomi (si considerino ad es. molecole biologiche o materiali polimerici) è intimamente legata allo sviluppo di algoritmi capaci di ridurre l’ordine di scalabilità con le dimensioni del sistema, alle potenze dei calcolatori e, per ultimo, a nuovi paradigmi di calcolo.

Con l’avvento dei personal computer, la diffusione dei prodotti tecnologici su larga scala e la conseguente crescita esponenziale della potenza di calcolo la chimica computazionale ha potuto esplorare ambiti un tempo inimmaginabili fornendo interpretazioni strutturali e meccanicistiche; nella chimica pura, nella chimica biologica, nelle scienze dei materiali solo per citarne alcuni. L’incremento nella potenza di calcolo in realtà ha fatto molto di più che offrire macchine per eseguire conti velocemente. Al fianco di queste risorse si sono infatti parallelamente sviluppati negli anni sofisticati linguaggi e paradigmi di programmazione (la programmazione orientata agli oggetti ne è un esempio) che oltre a fornire la possibilità di estrarre l’enorme potenza di calcolo messa a disposizione dalle nuove macchine, hanno consentito di cominciare a pensare al calcolo numerico computazionale come ad una “possibilità di fare esperimenti”: il calcolo quindi nella sua complessità “hardware e software” costituisce un vero e proprio strumento da laboratorio con il quale il chimico, proprio come con gli strumenti tradizionali, può progettare nuove sostanze, nuovi materiali, sintetizzare farmaci, interpretare meccanismi di reazione ...

Ovviamente come tutti gli strumenti anche quello computazionale deve essere conosciuto per poterne estrarre gran parte delle potenzialità. È logico chiedersi allora: cosa significa per il chimico conoscere i sistemi di calcolo? La risposta non è affatto banale, ma almeno il tentativo di formularne una è necessario per fornire il quadro generale di questa tesi. La possibilità per un chimico analitico o organico, di far uso di uno spettrometro NMR non

richiede ovviamente che egli conosca nei minimi dettagli i principi teorici della risonanza magnetica, che sappia risolvere le equazioni di Bloch ed interpretarne i risultati. Tanto meno è necessario che conosca i dettagli tecnici e tecnologici con i quali sono stati realizzati gli strumenti di cui fa uso. Tuttavia per avvicinarsi all'uso dell'NMR il chimico ha "studiato" i principi fisici che regolano il fenomeno della risonanza magnetica e nel processo di "avvicinamento", di "presa di confidenza" con la strumentazione vera e propria, scopre ed acquisisce anche un certo numero di nozioni tecniche necessarie per un utilizzo corretto dell'apparato sperimentale: questo schema è, sotto certi aspetti, affine anche per il chimico computazionale malgrado gli elementi siano personal computer, cluster di calcolatori, linguaggi di programmazione, modelli chimici teorici.

Al fianco del chimico computazionale negli ultimi anni sono comparse nuove figure nello scenario del calcolo il cui profilo è ad alto contenuto tecnologico. Questo processo di ampliamento ed integrazione di competenze tecnico/informatiche nel bagaglio professionale del chimico computazionale è ancora in corso ed è guidato in particolare dalle nuove frontiere del calcolo, in particolare del calcolo distribuito geograficamente : *il grid computing*. Malgrado chi esegue calcoli di chimica non debba necessariamente essere un esperto di *High Performance Computing* (HPC) o di reti di calcolatori, nella comunità dei chimici computazionali ci si è presto resi conto della necessità di far partecipare anche i chimici al processo evolutivo dei sistemi di calcolo per la chimica. Ovviamente le architetture computazionali non si evolvono per la chimica in particolare, ma quest'ultima è in grado di approfittare delle nuove tecnologie e delle soluzioni di calcolo distribuito per i propri scopi. In questo processo di approvvigionamento di risorse e soluzioni per la chimica diventa sempre più fondamentale la partecipazione del chimico come figura in grado di adattare gli apparati di calcolo alle proprie necessità. Quest'ultimo aspetto è forse il più importante per comprendere il ruolo che il chimico informatico riveste nella realizzazione e nell'uso di risorse di calcolo. Le nuo-

ve tecnologie computazionali sono sempre più potenti ed anche sempre più flessibili: in questo modo è possibile realizzare soluzioni commerciali *general purpose*¹ adattabili a contesti produttivi, commerciali e di ricerca. Questa flessibilità nell'adattamento delle risorse al contesto di applicazione se da un lato consente alle case costruttrici di contenere i costi allargando il bacino di vendita, dall'altro implica la presenza di risorse umane in grado di configurare il sistema per fornire le prestazioni migliori per un particolare contesto di produzione: una stessa risorsa di calcolo, a prescindere dai dettagli, sarà configurata in modo diverso a seconda che venga collocata in una banca per offrire servizi oppure in un ente di ricerca per eseguire calcoli.

Nell'ambito della chimica computazionale non si può prescindere dalla conoscenza delle architetture di calcolo e dei loro principali schemi di configurazione se si vogliono raggiungere le alte prestazioni che tali soluzioni mettono a disposizione. Nel contesto del calcolo distribuito su griglia diventa addirittura necessario che il chimico computazionale partecipi sia alla realizzazione che al mantenimento della stessa: il grid è uno strumento di condivisione di risorse, ma anche di competenze, ed una comunità che vuole condividere le proprie conoscenze deve necessariamente collaborare a tale processo. In conclusione la chimica computazionale, soprattutto quella che si sta affacciando ai nuovi sviluppi tecnologici, almeno in parte dovrà essere costruita e sviluppata da chimici che con l'ausilio delle competenze tecnologiche, potranno sfruttare al meglio le risorse messe a disposizione dalle nuove soluzioni di supercalcolo.

Questa lunga seppur necessaria premessa, oltre ad offrire il mio personale punto di vista sull'argomento, serve per preparare il lettore al fatto che in questo lavoro di tesi ho dovuto affrontare problematiche molto diverse tra loro che spaziano dai metodi computazionali applicati alle previsioni di proprietà spettroscopiche, alle architetture di calcolo distribuito, fino alla configurazione hardware delle macchine. Ho tentato di rendere più agevole la

¹Con questo termine si indicano soluzioni tecnologiche applicabili a più settori.

lettura, con una netta separazione degli argomenti anche quando strettamente interconnessi. Per questo motivo ho discusso gli approcci computazionali integrati per il calcolo delle proprietà spettroscopiche di molecole organiche in soluzione nei Capitoli 1 e 2. Nel Capitolo 3 ho trattato il tema del calcolo distribuito in chimica discutendo le linee generali del grid computing. Nel Capitolo 4 presento l'implementazione del calcolo numerico su grid di proprietà molecolari, in particolare della derivata seconda dell'energia e quindi delle frequenze di vibrazione: in questo capitolo quindi, discuto il *mapping* del metodo multiscala sulle architetture distribuite presentate nel capitolo precedente. Infine nel Capitolo 5 ho riportato le principali conclusioni di questo lavoro. Ho riportato qui di seguito l'elenco delle pubblicazioni prodotte nel corso del lavoro di tesi; in Appendice A viene inoltre descritta la configurazione completa delle due principali risorse di calcolo da me messe a punto durante il periodo di tesi.

1. N. Sanna, T. Castrignano, P. D'Onorio De Meo, D. Carrabino, A. Grandi, G. Morelli, P. Caruso, V. Barone, *Gaussian grid: a computational chemistry experiment over a web service-oriented grid*, Theoretical Chemistry Accounts 117, 1145-1152 (2007)
2. V. Barone, R. Improta, G. Morelli, F. Santoro, *UV-vis spectra of p-benzoquinone anion radical in solution by a TD-DFT/PCM approach*, Theoretical Chemistry Accounts 118, 143-148 (2007)
3. M. Pavone, G. Brancato, G. Morelli, V. Barone, *Spectroscopic properties in liquid phase: combining ab initio calculations and classical molecular dynamics*, ChemPhysChem 7, 148-156 (2006)
4. M. Pavone, O. Crescenzi, G. Morelli, N. Rega, V. Barone, *Solvent Effects on the UV ($n \rightarrow \pi^*$) and NMR (^{17}O) Spectra of Acetone in Aqueous Solution: Development and Validation of a Modified AMBER Force Field for an Integrated MD/DFT/PCM Approach*, Theoretical Chemistry Accounts 116, 456-461 (2006)

Capitolo 1

Calcolo di proprietà spettroscopiche in soluzione

1.1 Introduzione

Gli strumenti di cui la chimica computazionale si serve per indagare la struttura e le proprietà delle molecole e della loro aggregazione a costituire la materia sono sempre più numerosi. Ognuno di essi poggia su determinati principi classici o quantistici a seconda del livello di approssimazione con il quale si vuole studiare il soggetto molecolare compreso eventualmente l'ambiente nel quale esso è immerso (fase gassosa, soluzione, stato solido, ...). È chiaro che la descrizione quantistica riproduce più fedelmente la realtà chimica del sistema di quanto possa fare quella classica a scapito però dell'efficienza computazionale: la soluzione (integrazione) delle equazioni classiche è infatti decisamente meno onerosa dei calcoli ab-initio previsti dall'integrazione dell'equazione di Schrödinger per sistemi atomici o molecolari. Tuttavia in molti casi l'approccio quantistico alla descrizione di un sistema molecolare pur comportando un alto costo computazionale non aggiunge molta più informazione di quanto non possa fornirne quello classico; in particolare questo accade quando il comportamento chimico del sistema è definito essenzial-

mente da effetti medi. Se consideriamo per esempio la condizione chimica nella quale si trova un soluto in una soluzione, è ragionevole pensare che rispetto alla maggior parte delle molecole del solvente la molecola solvatata sperimenterà una interazione che dal punto di vista computazionale può essere assimilata ad un effetto medio. Per contro “il senso chimico” suggerisce che le molecole di solvente “prossime” a quella di soluto possano dar conto di un’anisotropia dell’interazione: in questo caso un livello di trattazione a “pochi corpi” è in grado di fornire una descrizione migliore di quella media. Da queste considerazioni emerge la possibilità per il chimico computazionale di combinare diverse tecniche e metodi di calcolo per ottimizzare le risorse di calcolo a sua disposizione. Durante lo svolgimento di questa tesi, ho partecipato all’applicazione di questa tecnica ad un caso specifico: il calcolo dei parametri spettroscopici UV ed NMR per la molecola di acetone in acqua: grazie a questo lavoro, da un lato ho appreso la tecnica dell’approccio multiscala lavorando su uno specifico caso e dall’altro ho potuto verificare le configurazioni hardware e software dei sistemi di calcolo realizzate durante lo svolgimento del mio dottorato.

1.2 Il modello multiscala

Molti problemi chimici sono caratterizzati dalla necessità di connettere aspetti atomici e subatomici come la rottura e la formazione di legami o il trasferimento di carica, ad altri su scala macroscopica, ad esempio le proprietà di *bulk*. Le forze intermolecolari si estendono su distanze dell’ordine dei nanometri mentre le strutture microscopiche di alcune fasi possono estendersi per decine di nanometri fino ad arrivare alle scale dei micron, dei millimetri ed oltre a seconda della natura molecolare del sistema. Allo stesso modo, la scala dei tempi per la descrizione dei processi dinamici può estendersi dai femto ai millisecondi fino ad arrivare, in alcuni casi, ai minuti ed ore. Esempi di sistemi che esibiscono proprietà su scale così ampie di distanze e tempi

sono sicuramente i polimeri, i sistemi colloidali e quelli biologici.

La descrizione di questi problemi richiede un'approccio multiscala: al momento non esiste infatti un unico algoritmo in grado di descrivere un sistema in modo trasversale su tutte le scale spaziali e temporali.

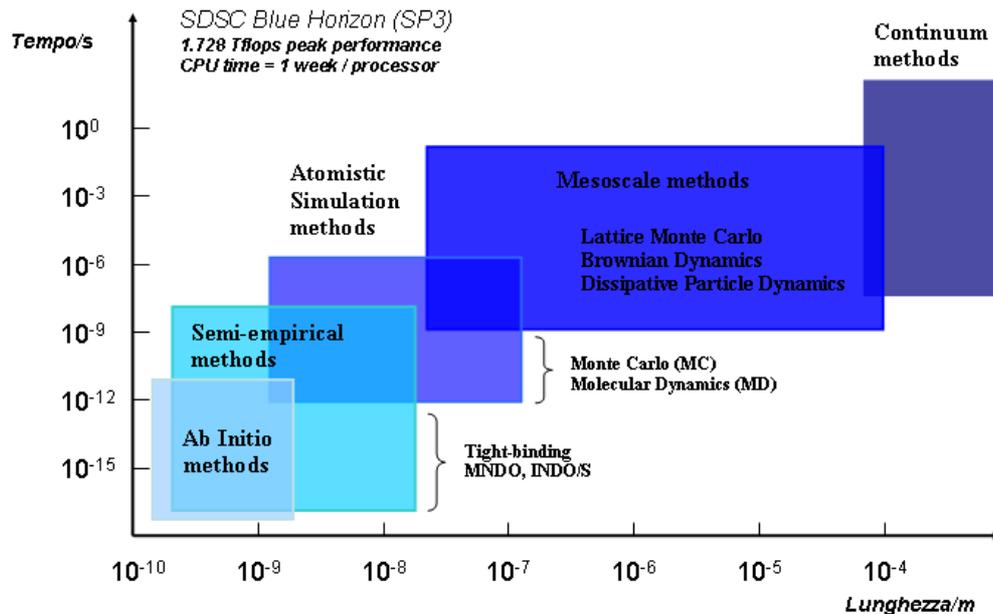


Figura 1.1: Il metodo multiscala in chimica.

In figura 1.1 sono riportati i gruppi principali dei metodi computazionali, collocati nelle varie regioni spaziali e temporali compatibili con le attuali potenze di calcolo. In particolare la figura 1.1 è stata costruita riferendosi ad una macchina Blue Horizon (IBM SP3) da 1.728 Tflops¹. Nel passaggio dalla regione degli algoritmi *ab initio* a quella del continuo debbono essere introdotte molte approssimazioni: i calcoli sulle scale macroscopiche possono essere eseguiti in tempi ragionevoli a scapito dell'accuratezze dei risultati. La struttura elettronica si perde oltre le simulazioni atomistiche, mentre i det-

¹Informazioni ottenibili all'indirizzo web <http://www.npaci.edu/BlueHorizon/>

tagli atomistici scompaiono a livello di meso scala. La sfida computazionale nell'approccio multiscala risiede nel trovare i collegamenti tra i vari metodi nei vari livelli fino a coprire l'intero range di spazio e tempo che parte dal microscopico mondo atomico e subatomico ed arriva alla materia percepita sulle scale macroscopiche. Di seguito descriverò un'applicazione di questo metodo in ambito chimico, che integra gli approcci quantistici e classici per la determinazione delle proprietà spettroscopiche di molecole in soluzione di cui mi sono servito nella mia ricerca.

1.3 Approccio integrato QM/MM per il calcolo di proprietà spettroscopiche in soluzione

Il modello chimico preso in esame è costituito dall'acetone in acqua. Si tratta di un tipico esempio di sistema soluto/solvente nel quale la molecola di soluto è spettroscopicamente attiva: un tale sistema è rappresentativo di una vasta gamma di sistemi molecolari in fase condensata. Per tali sistemi la risposta del soluto ad una sollecitazione esterna come quella prodotta da un campo elettromagnetico per misure spettroscopiche, costituisce un fenomeno fisico molto complesso da descrivere. A tale complessità forniscono il loro contributo da un lato le interazioni tra il soluto ed il solvente e dall'altro le fluttuazioni strutturali sia del soluto stesso che delle molecole di solvente. Dal punto di vista teorico, il problema della riproduzione accurata dei parametri spettroscopici in soluzione può essere decomposta in due parti principali:

1. Costruzione di un sottoinsieme rappresentativo dello spazio delle configurazioni del sistema soluto/solvente;
2. Calcolo dei parametri spettroscopici, con elevati livelli di teoria, sui singoli frame e media sui valori ottenuti.

Il calcolo delle quantità spettroscopiche sui singoli frame può essere realizzato con metodi *post-Hartree-Fock* (HF) o basandosi sulla teoria del funzionale della densità (DFT), soprattutto quando entrano in gioco proprietà del secondo ordine [1, 3]. In entrambi i metodi gli effetti dell'ambiente nel quale il soluto è immerso possono essere tenuti in conto o inserendo esplicitamente le molecole di solvente nella trattazione quantistica (QM) [4] con il conseguente aumento del costo computazionale oppure, in maniera computazionalmente più efficiente, considerando il soluto immerso in un dielettrico continuo polarizzabile (PCM) [5, 6, 7]. La soluzione qui usata prevede l'adozione di entrambi i metodi avendo cura di applicare gli stessi in diverse regioni spaziali intorno al soluto: adottando in prossimità dello stesso la trattazione esplicita (QM) e considerando il resto del bulk come un mezzo continuo polarizzabile. Meno dispendiosa è, in linea di principio, la costruzione dell'insieme di configurazioni su cui eseguire il calcolo ab-initio. Una possibilità è costituita dalla Dinamica Molecolare ab initio del tipo Car-Parrinello soprattutto quando sono richiesti alti livelli di accuratezza [10]. Ovviamente, come per tutti i metodi ab-initio, l'elevato costo computazionale rende questa scelta applicabile solo a sistemi di piccole dimensioni e a tempi di simulazione molto brevi. Allo stesso tempo, i modelli di tipo QM generalmente adottati in tecniche di MD ab-initio come, ad esempio, il DFT con funzionali basati sull'approssimazione di gradiente generalizzato (GGA), possono non essere in grado di riprodurre in modo accurato le proprietà strutturali ed energetiche dei sistemi molecolari: la conseguenza di tutto ciò è costituita da un considerevole scarto tra i valori spettroscopici sperimentali e quelli calcolati [9, 11, 12]. Sul versante opposto, i metodi di Dinamica Classica offrono tempi computazionali ridotti al prezzo di livelli di approssimazione elevati implicitamente presenti nella derivazione dei campi di forza adottati: non di meno la simulazione classica fornisce la possibilità di eseguire dinamiche rappresentative (estese) per la costruzione di medie statistiche. Di seguito verrà discusso il metodo che prevede l'adozione di entrambi gli approcci, classico ed ab-initio.

Malgrado tale metodo ha una lunga tradizione applicativa [8, 9, 11, 13], di seguito riportiamo alcuni aspetti che caratterizzano nello specifico il modello qui adottato:

1. Parametrizzazione del campo di forze per il sistema soluto/solvente da usare per simulazioni a tempi lunghi in Dinamica Molecolare classica. I parametri possono essere presi da modelli già esistenti come quelli presenti in AMBER [15], OPLS [14] o CHARMM [16] oppure derivati, se necessario, da calcoli di tipo QM. In particolare i potenziali intermolecolari possono essere convenientemente tarati per riprodurre la distribuzione radiale soluto/solvente a partire da simulazioni di Dinamica Molecolare ad-initio su tempi brevi.
2. Adozione del metodo *Mean Field* [17] (MF) per la trattazione degli effetti di solvatazione: in breve, in tale metodo il soluto è posto al centro di una cavità sferica immersa in un mezzo dielettrico continuo e solvatato da un certo numero di molecole di solvente trattate esplicitamente.
3. Il calcolo QM delle proprietà spettroscopiche sui vari frame è eseguito tenendo conto degli effetti di solvatazione provenienti dal mezzo continuo ed in modo esplicito (QM) dalle poche molecole prossime al soluto attraverso quindi un uso concorrente ed integrato di calcoli classici e quantistici a realizzare il così detto metodo QM/MM.

Il modello del MF riproduce molto bene le proprietà strutturali e termodinamiche del bulk acquoso [17] così come quelle di soluzioni acquose di semplici soluti, come mette in evidenza il confronto con i metodi che fanno uso di condizioni periodiche al contorno (PBC). Tale approccio si è dimostrato inoltre adatto a metodi integrati di tipo QM/MM o QM puri nel caso di simulazioni di liquidi puri e soluzioni consentendo l'uso di basi localizzate al posto delle onde piane molto usate per tali sistemi. È importante sottolineare che l'approccio adoperato in questo lavoro per il calcolo delle proprietà spettroscopiche a posteriori sui diversi frame della dinamica, è da considerarsi una

differenza metodologica rispetto ad altre procedure che calcolano le medesime quantità “on the fly” [18] oppure mediante tecniche perturbative [19]. Vale la pena sottolineare che in tale procedura prevede una serie di aggiustamenti dipendenti dal sistema molecolare in studio e riguardanti soprattutto la parametrizzazione del campo di forze e la conseguente convergenza delle proprietà spettroscopiche oltre a quelle geometriche. In particolare nel lavoro riguardante la molecola di acetone in acqua qui trattato, è stata studiata la transizione elettronica $n \rightarrow \pi^*$ e la *nuclear magnetic shielding constant* del ^{13}C ed ^{17}O del gruppo carbonilico sia in fase gassosa che in soluzione: tale sistema ha manifestato una forte dipendenza delle sue proprietà spettroscopiche sia dalle componenti intra che intermolecolari dei campi di forza. I passi fondamentali per la trattazione di questo sistema sono stati innanzi tutto la costruzione di un opportuno campo di forze (classico) contenente entrambi i contributi di interazione inter ed intramolecolari: per tale taratura è stata adottata una dinamica molecolare ab initio Car-Parrinello. Successivamente un gran numero di campionamenti sono stati effettuati dalla dinamica classica, sia in fase gassosa che in soluzione a distanza di circa 1 ns ciascuno: su tali *snapshot* sono stati calcolati i vari parametri (vedi schema in figura 1.6).

1.4 Metodo e dettagli computazionali

1.4.1 Parametri del campo di forza

Come anticipato nella descrizione generale del metodo, per ottenere una dinamica molecolare classica rappresentativa si è fatto uso di un potenziale costruito su base quantistica lavorando sia sulla molecola di acetone che sul cluster acetone-(H₂O)₂ a diversi livelli di teoria per ottenere un potenziale di interazione in grado di descrivere entrambi i contributi intra ed intermolecolari. In particolare è stato testato il funzionale Perdew-Burke-Ernzerhof (PBE) [22] e la sua controparte ibrida HF-DFT nota come PBE0 [23] per la taratura dei termini intramolecolari. Come ci si aspettava [11, 12, 21], l’in-

clusione del termine di scambio HF nella formulazione del funzionale fornisce un migliore accordo con i dati strutturali sperimentali. I termini intermolecolari tra la molecola di acetone e le molecole di acqua sono stati ottimizzati con dinamiche Car-Parrinello: ne discuteremo più avanti i dettagli. In tutti i casi in cui sono stati eseguiti calcoli QM si è fatto uso del codice Gaussian03 [24]. In tabella 1.1 sono riportati i parametri usati per la dinamica MM mentre in figura 1.2 sono riportati i parametri strutturali della molecola di acetone al minimo di energia conformazionale. In particolare, lo stretching del legame C=O ed il modo fuori dal piano dell'acetone sono stati "fittati" su dati QM, con un potenziale di Morse ed armonico rispettivamente, mentre tutte le altre costanti di forza sono state ottenute a partire dal campo di forza OPLS-AA [14] ed i valori di equilibrio sono stati scalati per accordare i parametri strutturali a quelli ottenuti con i calcoli QM.

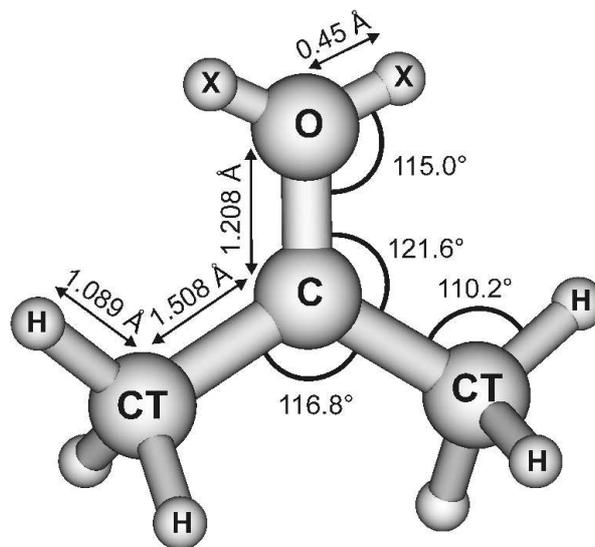


Figura 1.2: Struttura e tipo di atomi nella molecola di acetone

Proprietà spettroscopiche in soluzione

Interazioni di legame			Interazioni di legame		
Parametro	Valore	Unità	Parametro	Valore	Unità
$D(CO)$	810.82	$kJmol^{-1}$	q_0	0.0	
$B(CO)$	2.2636	Å^{-1}	q_X	-30.23	
$r_0(CO)$	1.2080 (1.2170)	Å	q_C	0.5825	
$k_b(CC)$	2652.66	$kJmol^{-1}\text{Å}^{-2}$	q_{CT}	-03204	
$r_0(CC)$	1.5088	Å	q_H	0.1105	
$k_b(CH)$	2845.12	$kJmol^{-1}\text{Å}^{-2}$	$\epsilon_{O,H}$	0.4697	$kJmol^{-1}$
$r_0(CH)$	1.0895	Å	$\sigma_{O,H}$	2.9500	Å
$k_\theta(OCC)$	669.44	$kJmol^{-1}rad^{-2}$	$\epsilon_{CT,H}$	0.09310	$kJmol^{-1}$
$\theta_0(OCC)$	121.79	deg	$\sigma_{CT,H}$	3.0000	Å
$k_\theta(CCC)$	527.18	$kJmol^{-1}rad^{-2}$	$\epsilon_{H,H}$	0.12552	$kJmol^{-1}$
$\theta_0(CCC)$	116.43	deg	$\sigma_{H,H}$	2.5000	Å
$k_\theta(HCC)$	418.40	$kJmol^{-1}rad^{-2}$	$\epsilon_{Ow,O}$	0.76427	$kJmol^{-1}$
$\theta_0(HCC)$	110.26	deg	$\sigma_{Ow,O}$	3.2980	Å
$k_\theta(HCH)$	292.88	$kJmol^{-1}rad^{-2}$	$\epsilon_{Ow,C}$	0.53438	$kJmol^{-1}$
$\theta_0(HCH)$	108.62	deg	$\sigma_{Ow,C}$	3.4580	Å
k_ζ	115.0	$kJmol^{-1}rad^{-2}$	$\epsilon_{Ow,CT}$	0.42367	$kJmol^{-1}$
ζ_0	0.0	deg	$\sigma_{Ow,CT}$	3.3330	Å
$k_\phi(OCCH)$	1.6736	$kJmol^{-1}$	$\sigma_{Ow,H}$	0.28564	$kJmol^{-1}$
$n(OCCH)$	1		$\sigma_{Ow,H}$	2.5830	Å
$\phi_0(OCCH)$	0.0	deg			
$k_\phi(OCCH)$	0.16736	$kJmol^{-1}$			
$n(OCCH)$	3				
$\phi_0(OCCH)$	180.0	deg			

Tabella 1.1: Parametri per il *force-field* per la molecola di acetone ed il sistema acetone/acqua.

I due atomi *dummy* indicati con la lettera “X” in figura 1.2 sono stati introdotti per rappresentare i *lone pairs* presenti sull’atomo di ossigeno e garantire la direzionalità del legame idrogeno durante la dinamica. Per la costruzione delle cariche parziali sulla molecola di acetone si è fatto ricorso al *fitting* su una griglia di punti, di un potenziale elettrostatico ab initio, in accordo con la procedura del *restrained electrostatic potential* (RESP) [25].

Il modello MM adoperato per trattare le molecole di acqua in soluzione e quelle del cluster è il popolare *simple point charge* (SPC) [26]. Infine una versione modificata del software *GROMACS* [27] è stata usata per eseguire tutti i calcoli di tipo MM e MD.

1.4.2 Dinamica Molecolare Classica

La dinamica molecolare classica per l'acetone in fase gassosa ed in soluzione è stata eseguita in un insieme canonico alla temperatura di 300 K: in entrambi i casi l'intero tempo di simulazione è stato di 1 ns compreso l'intervallo di stabilizzazione iniziale di 400 ps. La simulazione in fase condensata è stata eseguita come accennato nell'introduzione, facendo uso del modello MF i cui dettagli sono descritti in [17]. In breve tale metodo consiste nella simulazione esplicita di un sistema molecolare all'interno di una cavità di un mezzo continuo in grado di riprodurre gli effetti di interazione a lungo raggio con l'ambiente; ciò consente di evitare l'uso delle condizioni periodiche al contorno che, in principio, possono introdurre indesiderati effetti di correlazione. Nello specifico, il metodo MF include sia contributi elettrostatici (*reaction field*) che di dispersione-repulsione. In particolare, nel lavoro svolto per l'acetone, questo è stato solvatato con 111 molecole di acqua SPC (*Simple Point Charge*) e posto al centro della cavità immersa nel mezzo continuo. Il moto roto-traslazionale del soluto è stato rimosso durante la dinamica facendo ricorso al metodo descritto in [28]: tutti i calcoli sono stati eseguiti in doppia precisione facendo uso di una versione modificata di GROMACS [27] che include il modello MF e la termostatazione isocinetica Gaussiana [29].

1.4.3 Calcolo QM delle Proprietà Spettroscopiche

Tutti i calcoli di tipo QM sono stati eseguiti con il codice Gaussian03 [24] e le energie di eccitazione elettronica sono state calcolate con il metodo TD-DFT [1] (*time dependent-DFT*) con l'uso del funzionale PBE0 [23] e l'uso

combinato di basi 6-311++G(2d,2p) sugli atomi del carbonile e 6-31+G(d,p) sui rimanenti atomi del soluto e delle molecole di acqua (d'ora in poi indicherò tale base [6-311++G(2d,2p), 6-31+G(d,p)]) [30]: la consistenza di tale scelta è stata provata in altri recenti lavori nei quali sono stati calcolati i parametri UV [20]. Lo stesso set di base è stato adottato per il calcolo dei parametri di risonanza magnetica (in particolare il tensore σ), con il formalismo GIAO (*gauge-including atomic orbital*) [2] al livello di teoria PBE0 [23] ed MP2 [31]. A questo punto, sono stati eseguiti i calcoli QM su un numero di frame necessari per la convergenza dei risultati.

1.5 Risultati e Considerazioni

1.5.1 Parametrizzazione del campo di forza

La struttura dell'acetone in fase gassosa è nota sperimentalmente grazie a tecniche come la spettroscopia di diffrazione elettronica [32] e verrà usata in questo lavoro come riferimento per testare l'accuratezza dei risultati computazionali sia ab initio che di dinamica molecolare. Innanzitutto sono stati eseguiti calcoli QM a diversi livelli di teoria al fine di trovare la migliore combinazione di metodologia ab initio e basis set: a tale scopo si sono testati i livelli teorici PBE [22] ed HF-DFT (PBE0) [23] e per quest'ultimo si sono verificati diversi basis set a partire dal doppi e triplo zeta della serie di Pople [30] al triplo zeta della serie di Dunning [33]. I risultati ottenuti sono mostrati in tabella 1.2. Dall'analisi dei risultati emerge una sostanziale differenza nei valori ottenuti con i livelli PBE e PBE0 a parità di basis set 6-31+G(d,p): il funzionale ibrido migliora rispetto a quello standard l'accuratezza delle distanze di legame senza alterare i valori degli angoli di legame. Tra le funzioni di base utilizzati col funzionale PBE0, le triple-zeta della serie di Pople con funzioni diffuse e di polarizzazione sugli atomi di carbonio ed ossigeno, funzioni di polarizzazione su quelli di idrogeno, forniscono parametri geometrici consistenti con insiemi di base più estesi come cc-pVTZ ed

Aug-cc-pVTZ e con i valori sperimentali, a meno della distanza di legame C-H.

	C=O	C-C	C-H	C-C-C	C-C=O
PBE/6-31 +G(d,p)	1.229	1.520	1.102	116.6	121.7
PBE0/6-31 +G(d,p)	1.214	1.510	1.094	116.6	121.7
PBE0/6-311 +G(d,p)	1.208	1.509	1.090	116.4	121.8
PBE0/cc-pVTZ	1.206	1.507	1.091	116.4	121.8
PBE0/Aug-cc-pVTZ	1.207	1.506	1.091	116.4	121.8
MM	1.208	1.508	1.089	116.8	121.6
sperimentale	1.210	1.507	1.076	116.7	121.7
$\sigma_{sp.}$	0.003	0.002	0.006	0.3	

Tabella 1.2: Parametri geometrici per la molecola di acetone ottenuti a diversi livelli computazionali e rispettivi valori sperimentali: le distanze di legame sono in Å, gli angoli in gradi.

Tale valore è comunque affetto a livello sperimentale da un consistente errore per cui i valori ottenuti col calcolo sono da considerarsi soddisfacenti. A partire da queste considerazioni si è individuato quale miglior compromesso tra accuratezza ed efficienza il livello PBE0/6-311+G(d,p), e tutti i calcoli QM per la parametrizzazione del campo di forza sono stati eseguiti con questo schema.

In tabella 1.2 sono riportati anche i parametri strutturali ottenuti con il calcolo MM: l'accordo con quelli QM e sperimentali è molto buono. Come anticipato, il calcolo QM è stato sfruttato per fittare sui valori ottenuti un potenziale di tipo Morse per lo stretching del doppio legame C=O

$$V_{bond}(r) = D[1 - \exp(-B(r - r_0))]^2 \quad (1.1)$$

dove D , B e r_0 sono i parametri che compaiono nel potenziale. Tale forma

funzionale che descrive lo stretching del legame C=O è stata usata per tenere in considerazione il carattere anarmonico del moto. Allo stesso modo, ma con un potenziale armonico, è stato costruito il termine di interazione per descrivere l'oscillazione fuori dal piano dell'acetone:

$$V_{dihedral}(\zeta) = k_{\zeta}(\zeta - \zeta_0)^2 \quad (1.2)$$

dove ζ è l'angolo diedro e k_{ζ} è la costante di forza corrispondente (vedi tabella 1.1 per i valori). Il corretto campionamento lungo questi due termini di interazione si è dimostrato cruciale per la corretta determinazione dei parametri spettroscopici. Come già anticipato, nella fase di parametrizzazione del campo di forza sono stati aggiunti due atomi fittizi per simulare gli effetti dei due *lone pairs* sull'atomo di ossigeno carbonilico al fine di riprodurre il minimo energetico nella struttura acetone-(H_2O)₂: l'ottimizzazione delle geometrie del legame idrogeno corrispondente ha fornito una lunghezza di legame O...X di 0.45 Å ed un angolo del legame C=O-X di 115 gradi (vedi tabella 1.3 e figura 1.2). È da notare che i campi di forza convenzionali come quelli presenti in OPLS, CHARMM e GROMACS non includono *lone pairs* legati all'ossigeno carbonilico per cui sono del tutto inadeguati a descrivere la direzionalità dei legami idrogeno. Nel passaggio dalla fase gassosa a quella in soluzione, è stato necessario scalare la distanza di equilibrio del legame C=O per tenere in conto gli effetti di solvatazione: tali modificazioni strutturali sono indotte da riarrangiamenti nella densità elettronica e non possono essere ricostruiti con termini di campo di forza classici. Per questa ragione si sono usati per la dinamica molecolare due diversi valori per la distanza di equilibrio del legame C=O a seconda della fase in cui si è svolta la simulazione. È inoltre interessante notare come il solvente abbia un effetto trascurabile sulla lunghezza del legame C=O già oltre le due molecole di acqua più vicine; nella terza colonna della tabella 1.3 è mostrato come l'inclusione del PCM nei calcoli QM relativi al cluster acetone-(H_2O)₂ non modifichi in modo significativo (≈ 0.003 Å) la distanza del legame carbonilico. Questo risultato è in accor-

do con quello ottenuto a livello MM ($r_{C=O} = 1.220 \text{ \AA}$ per acetone-(H_2O)₂ e $r_{C=O} = 1.222 \text{ \AA}$ per acetone in soluzione, vedi tabella 1.4).

	MM	QM	QM+PCM
$O...H_w$	1.91	1.91	1.88
$O...O_w$	2.91	2.85	2.85
$C = O...H_w$	115.4	117.2	126.4
$O...O_w...H_w$	1.8	12.2	0.2
$C = O$	1.211 (1.220)	1.221	1.224

Tabella 1.3: Parametri geometrici del cluster acetone- H_2O valutati con ottimizzazioni QM ed MM. I calcoli QM sono stati eseguiti a livello PBE0/6-311+G(d,p) mentre in quelli MM si è fatto ricorso ai parametri in tabella 1.1. Gli atomi appartenenti alle molecole di acqua sono etichettati con il pedice w . Le distanze di legame sono in \AA e gli angoli in gradi.

Osserviamo inoltre, che i parametri relativi all'interazione intermolecolare acetone/acqua, ad esempio il potenziale di *Lennard-Jones* tra gli ossigeni delle due rispettive molecole, sono stati "aggiustati" in modo da riprodurre le caratteristiche strutturali attestate con simulazioni CPMD [21].

1.5.2 Dinamica Molecolare

Una volta determinati i potenziali di interazione inter ed intramolecolare possiamo occuparci della dinamica molecolare. Il modello di dinamica adottato è stato quello canonico ad una temperatura di 300 K sia in fase gassosa che in soluzione acquosa. Dalla dinamica abbiamo estratto un gran numero di configurazioni che rappresentano lo "spazio configurazionale" sul quale sono state mediate le proprietà. In questa fase di "taratura" sono stati confrontati i risultati della dinamica classica (MM) con quelli basati sul metodo Car-Parrinello (CPMD). Le simulazioni CPMD sono molto onerose per cui

avvengono per tempi non lunghi e quindi i confronti con i risultati MM sono, nella maggioranza dei casi, soltanto qualitativi.

	C=O	C-C	C-H	C-C-C	C-C=O
gas MM	1.209(0.002)	1.514(0.003)	1.089(0.02)	116.9(1.1)	121.4(0.6)
gas CPMD	1.23(0.02)	1.52(0.04)	1.11(0.04)	116.9(5.3)	121.4(4.3)
soluzione MM	1.222(0.002)	1.511(0.002)	1.090(0.001)	115.7(0.6)	121.5(0.5)
soluzione CPMD	1.25(0.01)	1.50(0.03)	1.11(0.02)	117.4(3.1)	121.0(3.4)

Tabella 1.4: Parametri geometrici medi derivanti dalla dinamica in fase gassosa ed in soluzione. Le distanze sono in Å e gli angoli in gradi.

In tabella 1.4 sono riportati i valori medi dei parametri strutturali ottenuti con i metodi MM e CPMD sia in fase gassosa che in soluzione: appare evidente che, solo il legame C=O diviene significativamente più lungo passando dalla fase gassosa a quella liquida (in virtù dei già citati effetti di interazione con il solvente). La maggior differenza predittiva nei parametri strutturali tra i metodi MM e CPMD è quindi in tale legame: come discusso in precedenza nell'ambito del metodo MM la descrizione del gruppo carbonilico è stata affidata alla parametrizzazione contenente un funzionale DFT ottimizzato (PBE0) che fornisce, in ultima analisi, risultati migliori rispetto al CPMD. Un'altra considerazione importante riguarda il legame idrogeno tra il gruppo carbonilico e le molecole di acqua, sia in termini strutturali (distanza e disposizione spaziale) che quantitativi (numero medio di legami). Per discutere gli aspetti relativi al legame idrogeno consideriamo la funzione di distribuzione radiale $O_{acetone} - O_{acqua}$ e $O_{acetone} - H_{acqua}$ ottenute con i metodi MM e CPMD. Come si può vedere dalla figura 1.3 entrambi i metodi forniscono la stessa distanza per il primo picco (oltre che la stessa altezza) mentre a distante maggiori gli scostamenti tra i due metodi diventano significativi: il confronto in queste regioni perde tuttavia di significato visto il basso peso statistico della dinamica CP.

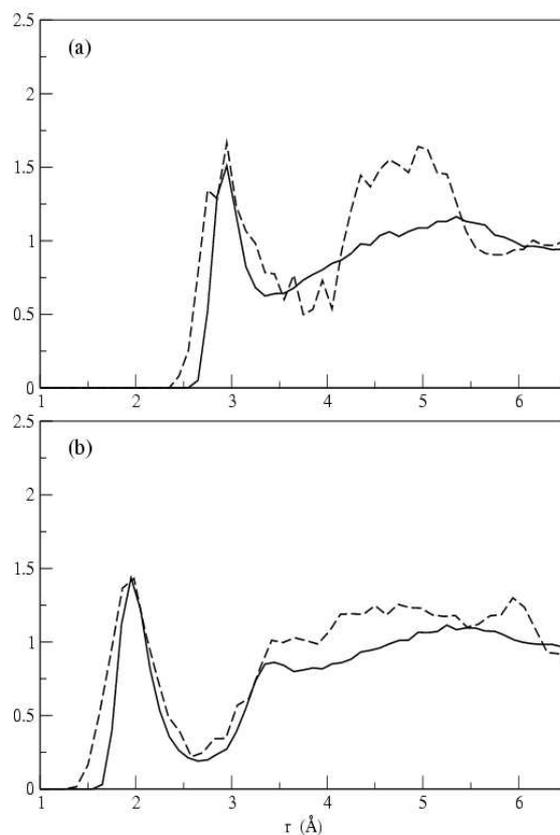


Figura 1.3: Funzioni di distribuzione radiale per: $O_{acetone}-O_{acqua}$ (a) e $O_{acetone}-H_{acqua}$ (b); tratto continuo MM, tratteggiato Car-Parrinello.

Un'altro parametro importante per i legami idrogeno tra acetone ed acqua è costituito dal loro numero, ovvero dalla distribuzione percentuale fra assetti in cui l'ossigeno carbonilico è coinvolto in uno, due o tre legami idrogeno. La tabella 1.5 fornisce i dati ottenuti con i due metodi, che mostrano un evidente accordo. Da notare come in entrambi casi un significativo numero di configurazioni prevedono la formazione di uno e tre legami idrogeno con una leggera preferenza per quest'ultimo caso. Un'ultima considerazione riguarda l'importanza dell'introduzione dei *lone pairs* posti sull'atomo di ossigeno

carbonilico: la loro presenza si ripercuote sulla struttura dei legami idrogeno facendo sì che le molecole di acqua si dispongano sul piano molecolare dell'acetone e non sfericamente intorno ad essa (figura 1.4).

	CPMD	MM
numero di legami H	2.1	2.0
% di 1 legame H	17	15
% di 2 legame H	52	66
% di 3 legame H	31	19
O...H _W	1.99 Å	2.06 Å

Tabella 1.5: Caratteristiche principali del legame idrogeno acetone-acqua.

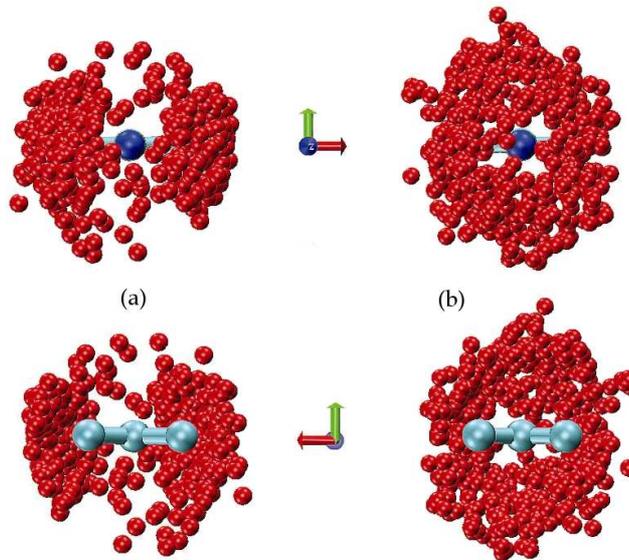


Figura 1.4: Distribuzione delle molecole di acqua coinvolte nel legame idrogeno con il gruppo C=O dell'acetone: (a) in presenza dei *lone pairs*, (b) in assenza dei *lone pairs*.

1.5.3 Parametri spettroscopici

In questa sezione sono discussi i valori dei parametri spettroscopici, in particolare la transizione elettronica $n \rightarrow \pi^*$ e la costante di schermo nucleare per ^{13}C e ^{17}O , nel passaggio dal sistema in fase gassosa a quello in soluzione acquosa.

	σ_{iso}	σ_{iso}
0 QM+MM	-258	-249
2 QM+MM	-253	-243
3 QM+MM	-253	-244
5 QM+MM	-252	-241
0 QM+PCM	-299	-289
2 QM+PCM	-261	-243
3 QM+PCM	-258	-245
0 QM	-362	-349
2 QM		-283
3 QM		-280
5 QM		-260
35 QM		-235
55 QM		-239

Tabella 1.6: Convergenza dei parametri NMR in funzione del numero di molecole di acqua trattate esplicitamente a livello QM. La costante di schermo isotropica σ_{iso} per ^{17}O è stata calcolata a livello PBE0/[6-311++G(2d,2p), 6-31+G(d,p)] per la parte quantistica; le molecole di acqua nel calcolo MM sono state trattate nello schema SPC.

In accordo con il metodo generale precedentemente descritto, sono state estratte un gran numero di configurazioni dalla traiettoria ottenuta con l'evoluzione classica del sistema e su queste sono state mediate le proprietà spettroscopiche di interesse ottenute su ciascuna configurazione. In funzione

delle proprietà d'interesse e del sistema molecolare è opportuno adottare particolari scelte come quella del livello di teoria nel metodo QM, la dimensione della regione da trattare a livello QM nel calcolo misto QM/MM, il numero di configurazioni necessarie al raggiungimento della convergenza. Come già discusso nella sezione relativa al metodo, il calcolo dell'energia della transizione UV è stato eseguito con un approccio TD-DFT mentre il calcolo delle costanti di schermo è stato eseguito con il metodo GIAO-DFT; si è fatto uso delle basi PBE0/[6-311++G(2d,2p), 6-31+G(d,p)].

	$\Delta E_{(n \rightarrow \pi^*)} [eV]$
0 QM+MM	4.56
2 QM+MM	4.57
3 QM+MM	4.57
5 QM+MM	4.57

Tabella 1.7: Energia di eccitazione nella transizione $n \rightarrow \pi^*$ calcolata a livello PBE0/[6-311++G(2d,2p), 6-31 +G(d,p)] in funzione del numero di molecole di acqua trattate esplicitamente a livello QM. Nel livello MM le molecole di acqua sono trattate come SPC ; la media è calcolata su 100 configurazioni.

Un certo numero di calcoli preliminari sono stati eseguiti sul sistema acetone/acqua per valutare il numero di molecole di acqua da trattare a livello QM mentre le restanti sono state assimilate a cariche puntiformi secondo il modello SPC. Nelle tabelle 1.6 e 1.7 sono riportati i valori ottenuti per per la costante di schermo dell' ^{17}O e l'energia di eccitazione per la transizione $n \rightarrow \pi^*$ rispettivamente, calcolati trattando 0,2,3 e 5 molecole di acqua a livello QM. Inoltre in tabella 1.6 sono riportati, per confronto, i risultati ottenuti applicando il modello del PCM direttamente alla regione QM (xQM+PCM), evitando quindi l'inclusione di cariche puntiformi, e per ultimo i valori ottenuti escludendo la regione del continuo (xQM). Dall'analisi della convergenza per la costante di schermo nel caso in cui non venga con-

siderata la solvatazione (ultima colonna della tabella 1.6) risulta evidente l'importanza degli effetti a lungo raggio (a prescindere dal fatto che questi siano inclusi per mezzo di un modello continuo (PCM) o per trattazione esplicita con il modello a cariche puntiformi (MM)). Il calcolo della costante di schermo è stato eseguito a livello QM sul cluster acetone- $(H_2O)_{55}$ su una specifica configurazione ed il valore così ottenuto (-239 ppm) non risulta molto diverso da quello ottenuto per lo stesso cluster in cui il numero di molecole di acqua nella regione QM è stato fissato a cinque e le rimanenti 50 sono state trattate a livello MM (-244 ppm), mettendo in evidenza ancora una volta la bontà del metodo. L'inclusione degli effetti di solvatazione sia a livello PCM che MM ha fatto convergere la costante di schermo verso lo stesso valore, a meno di deviazioni di natura statistica, riducendo drasticamente il peso computazionale. Risulta inoltre evidente che in generale con questo approccio è necessario considerare un numero molto limitato di molecole di acqua nella regione QM per raggiungere risultati soddisfacenti: in questo caso il numero è pari alle molecole di acqua coinvolte nella formazione di legami idrogeno con l'ossigeno del gruppo carbonilico. Per quanto riguarda la dipendenza dagli effetti di solvatazione dei parametri UV emerge, dall'analisi dei valori riportati in tabella 1.7 che le energie di transizione non dipendono in modo significativo dal numero di molecole di acqua trattate esplicitamente, a differenza di quanto avviene per le costanti di schermo nucleare. Del resto nell'ultimo caso il contributo quantistico all'interazione soluto-solvente è cruciale mentre il trattamento perturbativo [19], che tiene conto solo degli effetti elettrostatici, è completamente inadeguato. A partire da queste considerazioni, tutti i calcoli delle proprietà spettroscopiche sono stati eseguiti considerando due molecole di acqua a livello QM e tutte le altre sono state descritte come cariche puntiformi più un campo di reazione. Per validare ulteriormente i valori dei parametri spettroscopici così ottenuti, gli stessi sono stati calcolati mediando i risultati provenienti dalla dinamica molecolare su un insieme di configurazioni scorrelate. In figura 1.5 è riportata la conver-

genza dei parametri spettroscopici in funzione del numero di configurazioni su cui sono state costruite le medie sia in fase gassosa che in soluzione.

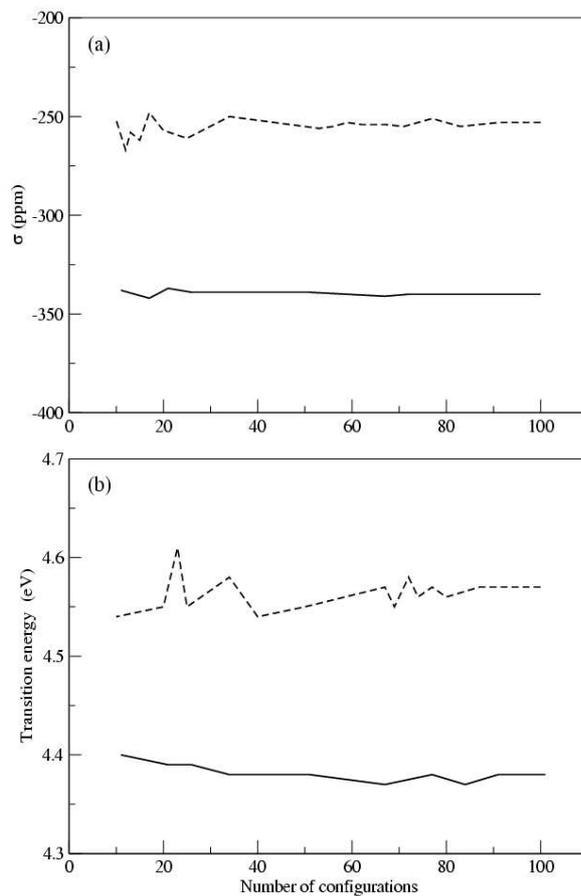


Figura 1.5: Andamento del valor medio della costante di schermo per ^{17}O (a) e del valor medio dell'energia di eccitazione $n \rightarrow \pi^*$ (b) in funzione del numero di configurazioni. I valori sono stati ottenuti a livello PBE0/[6-311++G(2d,2p), 6-31+G(d,p)]. I dati in soluzione (linea tratteggiata) sono stati ottenuti includendo due molecole di acqua nello strato QM; il tratto continuo si riferisce alla fase gassosa.

Risulta chiaro come si raggiunga la convergenza con un numero basso (dell'ordine delle decine) di configurazioni. Visto il contenuto peso computazionale per questo tipo di calcolo, i valori medi sono stati calcolati su un insieme di 100 configurazioni estratte ad intervalli regolari di tempo dalla traiettoria classica. In tabella 1.8 sono riportati i valori delle costanti di schermo

		PBE0	MP2	Exp.
^{13}C	gas	-23	-9	
	soluzione	-42	-28	
	$\Delta_{\text{solv.}}$	-19	-19	-18.9
^{17}O	gas	-340	-265	
	soluzione	-253	-200	
	$\Delta_{\text{solv.}}$	87	65	75.5

Tabella 1.8: Costante isotropica di schermo per gli atomi del gruppo carbonilico dell'acetone in fase gassosa ed in soluzione acquosa. I calcoli a livello PBE0 ed MP2 sono stati eseguiti con set di base [6-311 ++G(2d,2p), 6-31 +G(d,p)]. I calcoli in soluzione sono stati eseguiti considerando nello strato QM due molecole di acqua. In entrambe le fasi le medie sono state eseguite su 100 configurazioni.

isotropiche per il ^{13}C e ^{17}O del gruppo carbonilico nella molecola di acetone per la fase gassosa ed in soluzione per due livelli di teoria, PBE0 ed MP2. Per quanto riguarda il ^{13}C , la variazione della costante di schermo nel passaggio tra le due fasi, non mostra differenze dipendenti dal livello di teoria adottato; tale risultato inoltre è in ottimo accordo con quello sperimentale. Per contro il valore ottenuto per ^{17}O risulta meno accurato: come risulta dai valori riportati in tabella 1.8 il PBE0 sovrastima l'effetto solvente rispetto al valore sperimentale mentre l'MP2 fornisce un valore inferiore. Tuttavia uno scostamento di 10 *ppm* dei valori ottenuti va considerato accettabile tanto più che non sono stati introdotti fattori di scala o correzioni basate su con-

siderazioni di natura empirica. Vale inoltre la pena sottolineare che i valori calcolati al livello PBE0 per ^{13}C ed ^{17}O dalla dinamica CPMD sono rispettivamente di -23.1 e 105 ppm; il significativo scostamento si può ricondurre alla sovrastima nella lunghezza del legame C=O. Per quanto riguarda l'energia della transizione UV $n \rightarrow \pi^*$ infine, vi è un ottimo accordo tra il valore ottenuto in fase gassosa (4.38eV contro 4.5eV ottenuto in [34]) e soprattutto per lo shift nel passaggio in soluzione (0.19eV contro 0.2eV in [34]).

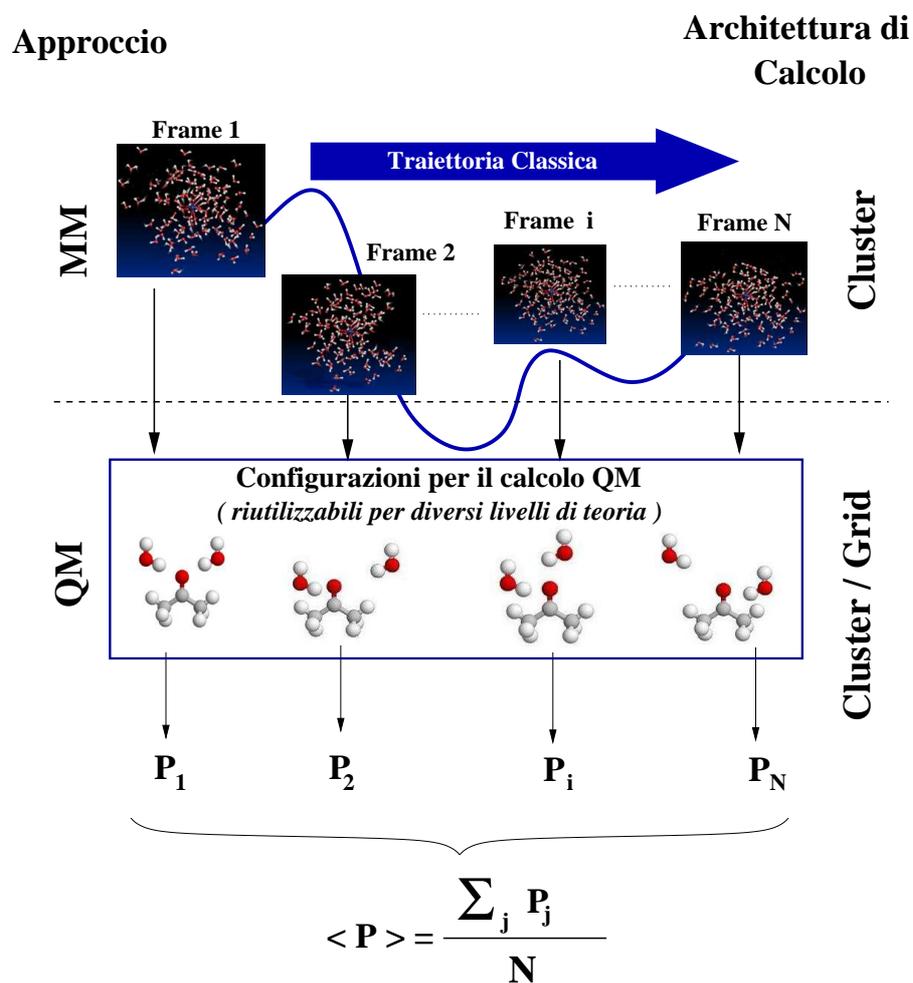


Figura 1.6: Schema del metodo di approccio integrato QM/MM/PCM.

Capitolo 2

Calcolo TD-DFT/PCM di proprietà spettroscopiche

2.1 Introduzione

I chinoni ed i loro prodotti di riduzione mono e bieletronica (semichinoni ed idrochinoni rispettivamente) sono molecole di grande interesse biologico in virtù soprattutto del ruolo che essi svolgono nei processi di trasferimento elettronico (ET) e nei processi naturali quali la respirazione e la fotosintesi [35].

Il *p*-benzochinone (*p*-BQ) è diventato la molecola chiave per lo studio della struttura elettronica e delle funzioni svolte dai chinoni in generale. La struttura elettronica dello stato eccitato dell'anione di tale molecola è costituita da un denso numero di livelli energetici i quali giocano un ruolo fondamentale proprio nei meccanismi di ET e rendono conto della capacità di tali specie di catturare elettroni nei sistemi naturali [36]. Lo spettro del radicale anionico del *p*-benzochinone (*p*-BQ⁻) è stato determinato in vari solventi [37, 38] ma una sua completa descrizione non è ancora stata eseguita: per questo motivo sono stati condotti una serie di studi quantomeccanici sugli stati elettronici eccitati di tale specie (i quali hanno un tempo di vita molto breve) con i

metodi *post-Hartree-Fock* più sofisticati come CASPT2 [39] o SAC-CI [40]. Malgrado i risultati conseguiti nel caso di specie isolate in fase gassosa siano incoraggianti, il passaggio a sistemi chinonici di maggiori dimensioni e/o in fase condensata richiede lo sviluppo di metodi meno dispendiosi dal punto di vista del calcolo. Negli ultimi anni si è affermato in chimica il metodo del Funzionale della Densità DFT [41] e la sua versione dipendente dal tempo TD-DFT [42] per lo studio di struttura e proprietà di sistemi di grandi dimensioni. Come già discusso nel Capitolo 1 lo studio di questi sistemi può essere condotto includendo anche gli effetti di solvatazione attraverso, per esempio, modelli quali il PCM [43]. A partire da queste premesse, si rende necessario testare l'applicazione di tali metodi su sistemi relativamente semplici per valutare la possibilità di passaggio a molecole di interesse biologico di maggiori dimensioni. In questo senso gli stati eccitati di radicali liberi organici costituiscono un ottimo punto di partenza: malgrado infatti molti studi siano stati condotti su queste specie [47], ancora poco si conosce circa la possibilità di trattare con il modello TD-DFT/PCM tali sistemi. A tale scopo, di seguito illustrerò l'applicazione di tale approccio al caso specifico del p-BQ⁻ mostrando come l'energia di eccitazione verticale in fase gassosa sia molto vicina a quella ottenuta con il metodo CASPT2 e come i dati spettroscopici di assorbimento e fluorescenza in fase condensata siano in ottimo accordo con quelli sperimentali.

2.2 Il metodo

Tutti i calcoli qui presentati sono stati eseguiti con la versione di sviluppo del codice Gaussian03 [48] usando PBE0 come funzionale ibrido della densità [23], diversi basis set della serie di Pople [49] e l'approccio di Kohn-Sham (UKS) per i sistemi open-shell. Gli effetti di solvatazione sono stati inclusi per mezzo del modello PCM [43, 50] nel quale il solvente è rappresentato attraverso un mezzo dielettrico continuo polarizzabile. L'ottimizzazione della

geometria è stata realizzata a livello PBE0/6-31G(d) in fase gassosa e a livello PCM/PBE0/6-31G(d) in soluzione facendo ricorso per gli stati elettronici eccitati, alla recente implementazione del calcolo dei gradienti analitici nel modello TD-DFT e PCM/TD-DFT [45] presente, al momento, solo nella versione di sviluppo del codice Gaussian03. Le energie elettroniche verticali ed adiabatiche di transizione sono state calcolate con diversi sets di base usando, in soluzione, il metodo della risposta lineare di non equilibrio per l'approccio PCM/TD-DFT [51].

2.3 Risultati

Come passo preliminare, è stata valutata la possibilità di adoperare il funzionale PBE0 per la descrizione dello stato fondamentale del p-BQ⁻ e del suo corrispondente neutro P-BQ. I principali parametri geometrici (vedi fig. 2.1) ottimizzati a livello PBE0/6-31G(d) per lo stato fondamentale della specie neutra, sono molto vicini a quelli sperimentali ottenuti con diffrazione elettronica [52] come riportato in tabella 2.1. Il funzionale PBE0 fornisce inoltre una stima molto accurata della deformazione della geometria associata all'introduzione di un elettrone nel p-BQ, fornendo un incremento di circa 0.04 Å per il legame C1-O1, di 0.03 Å per C2-C3 ed una contrazione di 0.035 Å per C1-C2. Questi valori sono in buon accordo con quelli ottenuti al livello CASSCF [39] che con quelli ottenuti dagli esperimenti di risonanza Raman [37] i quali riportano un incremento della frequenza di vibrazione per lo stretching del C1-C2 ed una diminuzione per i modi associati al C1-O1 e C2-C3. Tali risultati possono essere razionalizzati sulla base del carattere di legame/anti-legame dell'orbitale del p-BQ⁻ che formalmente contiene l'elettrone spaiato (6π in fig.2.2).

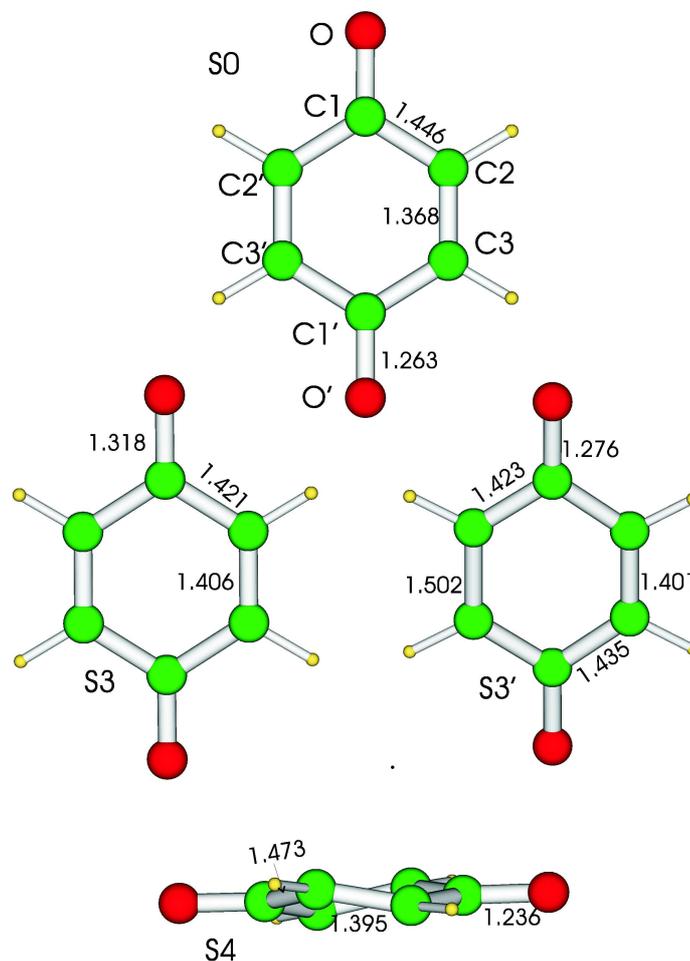


Figura 2.1: Strutture geometriche ottimizzate per lo stato fondamentale e primo stato eccitato del $p\text{-BQ}^-$: i valori riportati in parentesi si riferiscono al $p\text{-BQ}$.

Parametro	PBE0/6-31G(d)	Sperimentale
C1-O	1.228 Å	1.225 Å
C1-C2	1.481 Å	1.481 Å
C2-C3	1.340 Å	1.344 Å

Tabella 2.1: Parametri geometrici del $p\text{-BQ}$.

Il calcolo PBE0 infine ha fornito anche una buona descrizione della densità elettronica del p-BQ^- . In effetti anche i dati relativi al tensore g il quale dipende in modo critico proprio dalla densità di spin in prossimità dei nuclei, sono in eccellente accordo con quelli sperimentali [53]. Consideriamo ora il fatto che gli elettroni spaiati nello stato fondamentale del p-BQ^- hanno formalmente *spin up* (da qui in poi indicato con \uparrow) e ricordiamo che nel formalismo *unrestricted* \uparrow e \downarrow sono spin-orbitali diversi.

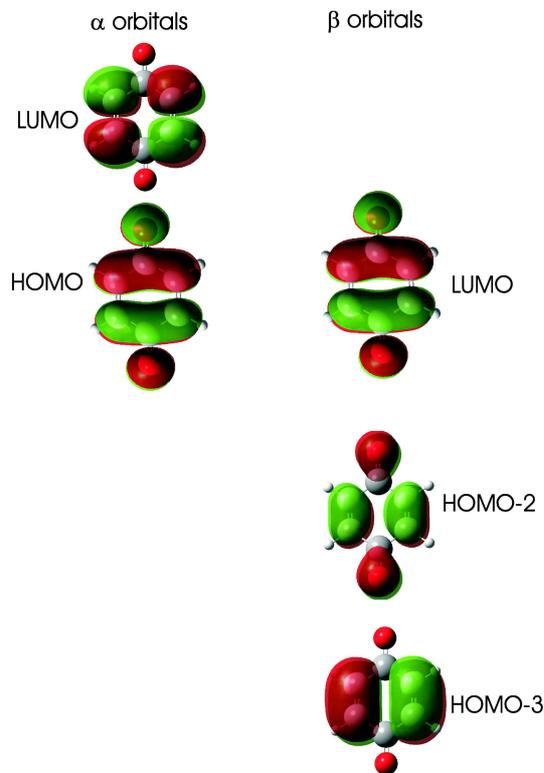


Figura 2.2: Rappresentazione schematica degli orbitali π e π^* coinvolti nelle transizioni elettroniche del p-BQ^- nelle rappresentazioni dei gruppi D_{2h} e C_{2v} ($\alpha \equiv \uparrow$; $\beta \equiv \downarrow$).

In tale scenario il SOMO (*Singly Occupied Molecular Orbital*) è lo spin-orbitale $6\pi \uparrow$ mentre gli spin-orbitali di frontiera che descrivono i primi cinque stati di eccitazione a qualsiasi livello computazionale sono quelli vuoti $7\pi \uparrow$

e $6\pi \downarrow$ e quelli occupati $3n \downarrow$, $4n \downarrow$, $4\pi \downarrow$ e $5\pi \downarrow$. Per etichettare le transizioni verticali, inoltre, possiamo far riferimento gruppo D_{2h} cui appartiene lo stato fondamentale 1^2B_{2g} (D0). I primi due stati eccitati D1 (1^2B_{2u}) e D2 (2^2B_{3g}), dominati dalle transizioni $4n \downarrow \rightarrow 6\pi \downarrow$ e $3n \downarrow \rightarrow 6\pi \downarrow$ rispettivamente, hanno entrambi carattere n/π^* e quindi, in principio, vietate. La coppia successiva di stati eccitati [D3 (1^2B_{3u}) e D4 (1^2A_u)] derivano dalle transizioni $5\pi \downarrow \rightarrow 6\pi \downarrow$ e $6\pi \uparrow \rightarrow 7\pi \uparrow$ rispettivamente permesse e del tipo $\pi \rightarrow \pi^*$. Lo stato successivo D5 (1^2B_{1g}) corrisponde ad una transizione $\pi \rightarrow \pi^*$ ($4\pi \downarrow \rightarrow 6\pi \downarrow$) ed infine D6 (2^2B_{3u}) dominata dalla transizione $6\pi \uparrow \rightarrow 8\pi \uparrow$ del tipo $\pi \rightarrow \pi^*$. Gli spin-orbitali KS π e π^* sopra discussi, sono raffigurati in figura 2.2. La natura anionica della specie in studio suggerisce l'importanza delle funzioni diffuse nella descrizione della struttura elettronica come del resto appare dai risultati mostrati in tabella 2.2. Infatti le energie di transizione calcolate con la base 6-31G(d) sono significativamente diverse dalle altre nelle quali l'estensione della base fino alla 6-31+G(d,p) mette in evidenza un netto miglioramento: d'ora in poi tutti i risultati saranno quindi riferiti alla base 6-31+G(d,p).

Le energie verticali di eccitazione (VEEs) in fase gassosa calcolate con il TD-DFT per le transizioni precedenti e la descrizione dello stato eccitato sono in buon accordo con quanto ottenuto con i calcoli CASPT2 e SAC-CI (ai quali d'ora in poi ci riferiremo come *post Hartree-Fock*). In particolare dall'analisi dei risultati ottenuti, emerge la presenza di due stati eccitati con tempo di vita molto breve (1^2B_{2u} e 1^2B_{3g}) intorno a 2.3 eV, di due ulteriori stati (1^2B_{3u} e 1^2A_u) intorno a 3 eV e per ultimo un stato (2^2B_{3u}) con forte assorbimento intorno a 4 eV. Da un punto di vista quantitativo, le differenze tra i risultati ottenuti con il TD-DFT ed i metodi *post-Hartree-Fock* sono minori di quelle registrate tra il CASPT2 e SAC-CI. L'unico punto significativo di discrepanza tra il TD-DFT ed il CASPT2 riguarda il valore della VEE relativo alla transizione $6\pi \uparrow \rightarrow 8\pi \uparrow$ che risulta di circa 0.5 eV più alta rispetto al metodo post HF. Al momento il valore misurato sperimentalmente in so-

(A) Energie di eccitazione verticale (in eV) del p-BQ⁻ in fase gassosa con TD-DFT, CASPT2 e SAC-CI

Stato ^a	Transizione	6-31G(d) ^a	6-31+G(d,p) ^{a,b}	6-311+G(2p,2d) ^{a,b}	CASPT2 ^c	SAC-CI ^d
D1(1 ² B _{2u})	4n ↓ → 6π ↓	2.23(0.00)	2.30(0.00)	2.31(0.00)	2.23(0.00)	2.44
D2(1 ² B _{3g})	3n ↓ → 6π ↓	2.27(0.00)	2.35(0.00)	2.36(0.00)	2.25(0.00)	2.38
D3(1 ² B _{3u})	5π ↓ → 6π ↓	3.11(0.07)	3.09(0.08)	3.09(0.07)	2.80(0.05)	2.71
D4(1 ² A _u)	6π ↑ → 7π ↑	3.26(0.05)	3.10(0.05)	3.05(0.05)	2.82(0.10)	3.50
D5(1 ² B _{1g})	4π ↓ → 6π ↓	3.42(0.00)	3.32(0.00)	3.32(0.00)	3.25(0.00)	3.75
D6(2 ² B _{3u})	6π ↑ → 8π ↑	5.28(0.23)	4.18(0.00)	4.11(0.11)	3.56(0.32)	3.74

(B) Energie delle transizioni attive (in eV) del p-BQ⁻ in soluzione con PCM/TD-PBE0 e sperimentali

Stato	Transizione	6-31G(d) ^e	6-31+G(d,p) ^e	6-311+G(2p,2d) ^e	Exp. ^f
CCl₄					
D3(1 ² B _{3u})	5π ↓ → 6π ↓	3.08(0.10)	3.06(0.13)	3.06(0.12)	
D4(1 ² A _u)	6π ↑ → 6π ↑	3.29(0.07)	3.19(0.07)	3.13(0.07)	
D6(2 ² B _{3u})	6π ↑ → 6π ↑	5.18(0.29)	4.15(0.14)	4.08(0.13)	
DMSO					
D3(1 ² B _{3u})	5π ↓ → 6π ↓	3.08(0.10)	3.08(0.13)	3.08(0.12)	2.73
D4(1 ² A _u)	6π ↑ → 6π ↑	3.40(0.07)	3.35(0.07)	3.28(0.06)	≈ 3.10 ^g
D6(2 ² B _{3u})	6π ↑ → 6π ↑	5.18(0.23)	4.16(0.14)	4.09(0.15)	3.82

^aLe etichette di simmetria si riferiscono al gruppo puntuale D_{2h} .^bCalcolo single point TD-PBE0 sulla geometria ottimizzata con PBE0/6-31G(d,p).^cDati riportati in [39].^dDati riportati in [40].^eCalcolo single point PCM/TD-PBE0 sulla geometria ottimizzata PCM/PBE0/6-31G(d,p).^fDati riportati in [36].^gShoulder a ≈ 400 nm (in [38]).Tabella 2.2: Energie di eccitazione in fase gassosa ed in soluzione per il p-BQ⁻. In parentesi sono riportate le intensità.

luzione di DMSO risulta intermedio tra quelli CASPT2 e TD-DFT ottenuti in fase gassosa (tabella 2.2). L'accuratezza del calcolo TD-PBE0 costituisce un risultato di notevole importanza soprattutto in previsione di applicazioni a sistemi di grandi dimensioni vista la scalabilità del metodo rispetto a quelli post HF. Naturalmente la scelta di un funzionale ibrido (PBE0) e l'assenza di trasferimento di carica o transizioni Rydberg contribuiscono a sostenere questo successo, ma rimane il considerevole fatto di aver riprodotto lo spettro di valenza di un sistema *open shell* con grande accuratezza. Anche l'ottimizzazione delle geometrie degli stati eccitati porta a risultati in accordo con quelli CASSCF e sperimentali. L'analisi delle geometrie mette in evidenza che il passaggio allo stato eccitato D4 porta ad un leggero incremento dei legami C1-C2 e C2-C3 insieme ad una piccola diminuzione del C1-O1 rispetto allo stato fondamentale. Si registra inoltre un allargamento degli angoli C2-C1-C2' (C3-C1-C3') e C2-C3-H (C2'-C3'-H). In stretta analogia con i risultati CASSCF [39] i cambiamenti geometrici indotti dalla transizione elettronica D0→D4 (1^2A_u) possono essere descritti in termini di *ring breathing*. A tale proposito, un risultato importante ottenuto con il TD-DFT prevede per lo stato D4 una geometria non planare dell'anello: risultato indirettamente fornito anche a livello CASSCF dalla presenza di una frequenza immaginaria associata ad una vibrazione normale *fuori dal piano*. La piccola distorsione fuori dal piano può servire a ridurre la tensione generata dall'apertura degli angoli C2-C1-C2' e C3-C1-C3' nel nuovo stato. Quando i calcoli sono eseguiti fissando la simmetria D_{2h} sia il TD-DFT che il CASSCF portano ai medesimi risultati riguardo l'ottimizzazione della geometria anche per lo stato D3 (1^2B_{3u}) il quale, rispetto allo stato fondamentale, è caratterizzato da un notevole allungamento del legame C1-O1 e C2-C3, in accordo con il forte carattere di antilegame dello spin orbitale $6\pi \downarrow$. Dato che la lunghezza del legame C1-C2 diminuisce, l'anello del chinone diventa simile a quello benzenico con la distanza del legame carbonio-carbonio di circa 1.4 Å.

Tuttavia se rimuoviamo il vincolo di conservazione della simmetria, l'ot-

timizzazione porta a risultati diversi. Quando la simmetria D_{2h} non è un vincolo, gli stati D3 ($5\pi \downarrow \rightarrow 6\pi \downarrow$) e D5 ($4\pi \downarrow \rightarrow 6\pi \downarrow$) possono interagire se passiamo ad un set di base più esteso. Così se usiamo la 6-31+G(d,p) l'ottimizzazione dello stato D3 ($5\pi \downarrow \rightarrow 6\pi \downarrow$) converge ad un “nuovo” stato eccitato (D3' in fig.2.1) caratterizzato da un minimo di simmetria C_{2v} e sostenuto da entrambe le transizioni $4\pi \downarrow \rightarrow 6\pi \downarrow$ e $5\pi \downarrow \rightarrow 6\pi \downarrow$ (vedi fig.2.3).

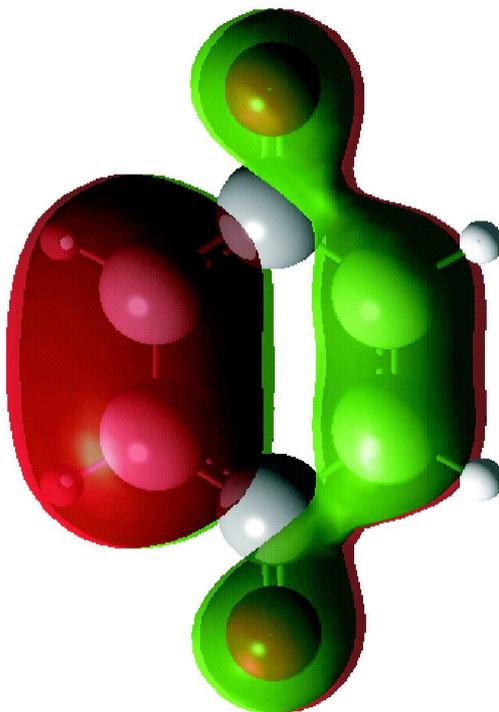


Figura 2.3: Rappresentazione dell'orbitale π risultante dalla combinazione degli spin-orbitali 4π e 5π quando la simmetria D_{2h} è ridotta a C_{2v} (struttura D3' in fig.2.2).

La geometria corrispondente al nuovo stato eccitato D3' è planare ma è caratterizzata da una distanza del legame C2-C3 minore rispetto a quella C2'-C3' il cui valore è prossimo a quello di un legame semplice carbonio-carbonio.

A questo punto passiamo ad analizzare i risultati ottenuti in soluzione: al fine di mettere in evidenza comportamenti generali, sono stati eseguiti calcoli con due solventi caratterizzati da polarizzabilità molto diverse, come il CCl_4 ed il DMSO (i dati sperimentali per il confronto sono disponibili solo per quest'ultimo). L'ottimizzazione della geometria in soluzione ha portato a strutture molto simili a quelle ottenute in fase gassosa e riportate in fig.2.1. Tuttavia i risultati riportati nella tabella 2.2 mostrano che lo *shift* indotto dal solvente sulle transizioni elettroniche non è trascurabile e che solo includendo tale spostamento è possibile un buon accordo con i dati sperimentali. In particolare si registra un piccolo *blue-shift* quando la polarità del solvente aumenta in accordo con quanto osservato sperimentalmente (vedi nota della Tabella 2 in [39]) Osserviamo inoltre come l'energia della transizione $4\pi \downarrow \rightarrow 6\pi \downarrow$ diminuisce con la polarità del solvente spostando così la banda corrispondente nello spettro UV. Tale effetto può essere interpretato sulla base del diverso valore del momento di dipolo totale nei due differenti stati eccitati. Infatti malgrado il momento di dipolo dello stato D3 è nullo per simmetria, quello relativo allo stato D3' vale circa 1.5 Debye ; questo spiega la relativa stabilizzazione di tale stato in DMSO. In ultimo in tabella 2.3 sono riportati i valori calcolati della transizione verticale al fine di confrontarli con i corrispondenti massimi di banda degli spettri di fluorescenza sperimentali (ove disponibili).

Stato	Transizione	6-31G(d) ^a	6-31+G(d,p) ^a	6-311+G(2p,2d) ^a	Exp.
Assorbimento in fase gassosa					
D3'	$5\pi \downarrow \rightarrow 6\pi \downarrow + 4\pi \downarrow \rightarrow 6\pi \downarrow$	2.91(0.11)	2.84(0.13)	2.88(0.13)	
D4	$6\pi \uparrow \rightarrow 7\pi \uparrow$	3.01(0.04)	2.82(0.04)	2.79(0.04)	
Assorbimento in CCl ₄					
D3'	$5\pi \downarrow \rightarrow 6\pi \downarrow + 4\pi \downarrow \rightarrow 6\pi \downarrow$	2.85(0.15)	2.76(0.20)	2.80(0.20)	
D4	$6\pi \uparrow \rightarrow 7\pi \uparrow$	3.05(0.06)	2.90(0.05)	2.87(0.05)	
Assorbimento in DMSO					
D3'	$5\pi \downarrow \rightarrow 6\pi \downarrow + 4\pi \downarrow \rightarrow 6\pi \downarrow$	2.78(0.04)	2.64(0.03)	2.67(0.03)	2.73 ^b
D4	$6\pi \uparrow \rightarrow 7\pi \uparrow$	3.15(0.07)	3.06(0.05)	3.02(0.05)	$\approx 3.10^c$
Emissione in fase gassosa					
D3'	$5\pi \downarrow \rightarrow 6\pi \downarrow + 4\pi \downarrow \rightarrow 6\pi \downarrow$	2.70(0.11)	2.67(0.13)	2.67(0.13)	
D4	$6\pi \uparrow \rightarrow 7\pi \uparrow$	2.51(0.04)	2.31(0.04)	2.28(0.04)	
Emissione in CCl ₄					
D3'	$5\pi \downarrow \rightarrow 6\pi \downarrow + 4\pi \downarrow \rightarrow 6\pi \downarrow$	2.60(0.15)	2.56(0.20)	2.57(0.20)	
D4	$6\pi \uparrow \rightarrow 7\pi \uparrow$	2.55(0.06)	2.40(0.05)	2.36(0.05)	
Emissione in DMSO					
D3'	$5\pi \downarrow \rightarrow 6\pi \downarrow + 4\pi \downarrow \rightarrow 6\pi \downarrow$	2.31(0.04)	2.18(0.03)	2.18(0.03)	
D4	$6\pi \uparrow \rightarrow 7\pi \uparrow$	2.63(0.06)	2.53(0.05)	2.48(0.05)	

^aCalcolo TD-PBE0 con geometria ottimizzata PCM/PBE0/6-31G(d,p).

^bDati riportati in [36].

^cShoulder a ≈ 400 nm (in [38]).

Tabella 2.3: Energie di assorbimento ed emissione (in eV) nella transizione verticale 0-0 dei due stati D3' e D4 del p-BQ⁻ in soluzione, calcolati a livello PCM/TD-DFT con valori sperimentali (dove presenti) riferiti al massimo di banda. In parentesi sono riportate le intensità.

Capitolo 3

Il calcolo distribuito in chimica

3.1 Introduzione

La chimica ha una forte connotazione applicativa e come tutte le scienze di questo tipo non può prescindere da strumenti ed apparati sperimentali: la sintesi organica non potrebbe esistere senza il corredo fatto di distillatori, estrattori e molto altro mentre non è possibile pensare di fare chimica analitica senza spettrometri, NMR, gascromatografi e via dicendo. Allo stesso modo non esisterebbe la chimica computazionale senza i calcolatori elettronici. In un certo senso quindi la storia di questa materia, della chimica al computer, è strettamente legata a quella dei calcolatori. Anche oggi ciò che riusciamo a produrre in questo ambito di ricerca è legato alle potenze di calcolo di cui disponiamo e quando pensiamo agli scenari futuri rivolgiamo in qualche misura lo sguardo anche agli sviluppi tecnologici dell'informatica. Una delle soluzioni computazionali ancora oggi di riferimento per il chimico computazionale è l'architettura cluster: a tale proposito presenterò due realizzazioni di questo tipo alle quali ho lavorato nella fase di installazione, configurazione e *tuning*¹ presso il Dipartimento di Chimica dell'Università di Napoli e di Padova: per

¹Con questo termine si intende l'insieme di operazioni hardware e software volte ad ottimizzare il rendimento di un calcolatore.

entrambi, in Appendice A, sono descritte le loro principali caratteristiche ed i principali passaggi di installazione/configurazione da me eseguiti. Per quanto possibile cercherò di rendere conto di come sia fondamentale che la configurazione di apparati di calcolo ad alte prestazioni sia assistita anche da chimici computazionali e di come per tali figure sia fondamentale avere una discreta competenza tecnologica. Gli apparati per l'HPC infatti, non sono generalmente concepiti dai produttori per questo scopo, per ovvie ragioni di mercato; in altre parole i vendors cercano di realizzare soluzioni hardware il più possibile *general purpose* in modo da estendere le loro offerte in più ambiti commerciali. Questo aspetto rende le macchine finali molto flessibili in termini di adattamento alle diverse esigenze: tale possibilità di adattamento si traduce in una vasta gamma di configurazioni sia a livello hardware che software.

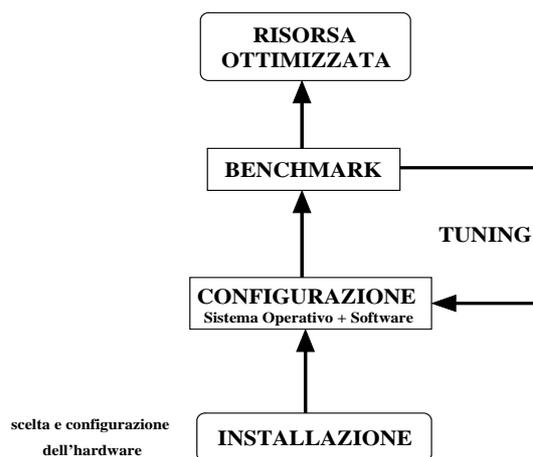


Figura 3.1: Ottimizzazione di una risorsa di calcolo.

Il chimico computazionale deve essere in grado di fornire il suo contributo al raggiungimento dell'ottimizzazione dell'ambiente di calcolo in entrambi gli ambiti, hardware e software. Nel caso dell'hardware per esempio, può intervenire nella scelta del sistema di interconnessione tra i vari nodi di un cluster

(Ethernet, Fibre Channel, Myrinet, InfiniBand, etc . . .) oppure nella scelta dei sistema di IO su disco (SCSI, FiberChannel, SATA, IDE, . . .) cercando di volgere a suo vantaggio il rapporto prestazioni/prezzo. Nell'ambito del software, invece, i benchmark, ovvero una serie di calcoli di riferimento rispetto ai quali misurare le prestazioni delle macchine, gli permetteranno di realizzare il tuning della risorsa: in altri termini la natura teorica del problema numerico da un lato e le caratteristiche architettoniche dell'altro, consentono al chimico computazionale di intervenire sui parametri di configurazione sia per il sistema operativo che per il software utilizzato. In figura 3.1 è riportato il ciclo di ottimizzazione descritto.

3.2 I cluster di calcolatori

I cluster di calcolatori sono la principale forma di architettura *distributed memory* (vedi appendice A). In pratica sono costituiti da un gruppo di calcolatori completi (ovvero capaci di funzionare in modo autonomo al di fuori del cluster) interconnessi in modo tale da lavorare insieme come una risorsa di calcolo unificata, che può creare all'utente l'illusione di avere a che fare con un'unica macchina. La possibilità di sfruttare questo tipo di architettura dipende dalla capacità dei codici di usare la memoria distribuita attraverso protocolli come MPI (Message Passing Interface). Parte della mia attività di dottorato ha riguardato l'installazione, la configurazione e l'ottimizzazione di apparati cluster orientati al calcolo chimico. Inoltre il *setup* di queste apparecchiature è stato realizzato pensando al loro inserimento nelle griglie computazionali. In Appendice A ho riportato il lavoro da me svolto sui due principali apparati di calcolo del progetto *VILLAGE*² : il cluster “*Bionap*” presso il laboratorio LSDM³ di Napoli ed “*Avogadro*” presso il Dipartimento di Chimica dell'Università di Padova.

²<http://village.unina.it>

³<http://lsdm.campusgrid.unina.it>

3.3 Il grid computing

In questa sezione descriverò l'architettura grid avendo cura di non entrare troppo in dettagli tecnici e specialistici superflui in questo contesto ma discutendo gli aspetti generali. Partiamo dalla seguente definizione di “griglia computazionale”: *“insieme molto ampio e dinamico di entità distinte e geograficamente distribuite che condividono risorse eterogenee per la soluzione coordinata di problemi”*. Di seguito analizzeremo i termini fondamentali presenti nella definizione appena data, per mettere in evidenza gli aspetti essenziali su cui si sviluppa il grid computing.

1. ... **entità** ...

Le entità sono i costituenti fondamentali di una griglia: il termine generico sta ad indicare che queste sono una collezione di hardware, software, infrastrutture e soprattutto risorse umane. In generale le entità sono rappresentate da Università, Istituti di ricerca, aziende, privati,...

Le entità per conto loro appartengono a “domini” distinti per cui in esse vi sono amministratori, strumenti e politiche di gestione delle risorse assolutamente autonomi. Direttamente legato al concetto di “entità” vi è quello di *Virtual Organization* (VO) che in questo contesto può essere definito come un insieme dinamico di entità che condividono risorse. La condivisione può avere:

- diversi scopi (la realizzazione di un esperimento numerico, l'analisi di una grande quantità di dati,...);
- diversa durata;
- diversa dimensione.

Di fondamentale importanza è il fatto che la stessa entità può appartenere a diverse *Virtual Organizations* (VOs) simultaneamente.

2. ... **ampio e dinamico** ... Le dimensioni delle VO e quindi il numero

di entità in esse contenute, è dinamico: nuove entità possono aggiungersi mentre altre possono uscire in un contesto (quello delle VOs) in cui non esistono relazioni di fiducia tra le entità stesse.

3. ...che **condividono** risorse ...

La condivisione è presupposto fondamentale per la costituzione ed il funzionamento di una griglia: ogni entità mette a disposizione della stessa un insieme di risorse locali ed utilizza quelle offerte combinandole in modo opportuno per risolvere un determinato problema. In questo senso ogni entità agisce sia da client che da server come nel modello *peer to peer*.

4. ...condividono **risorse eterogenee** per ...

L'eterogeneità in una griglia si estende sia ai **dispositivi fisici** (non solo computers) come:

- risorse computazionali;
- data bases;
- software repositories;
- storage;
- risorse speciali: sensori, apparati di misura sperimentali, ...;
- sistemi operativi;
- ambienti di sviluppo;
- software applicativo.

5. ...**la soluzione coordinata di problemi**.

Il raggiungimento di obiettivi, come la realizzazione di un esperimento numerico, o l'analisi di una serie di dati di grandi dimensioni richiede grandi sforzi computazionali sempre più inimmaginabili su un singolo calcolatore (anche supercomputers o cluster): il grid computing offre l'infrastruttura di risorse (entità) volte al raggiungimento degli stessi.

Dopo aver discusso gli aspetti concettuali presenti nella definizione di grid sopra fornita, discutiamo quelli più strettamente architetturali ponendo l'attenzione sul ruolo svolta dal così detto *middleware*: a tale scopo faremo riferimento alle componenti presenti nella figura 3.2.

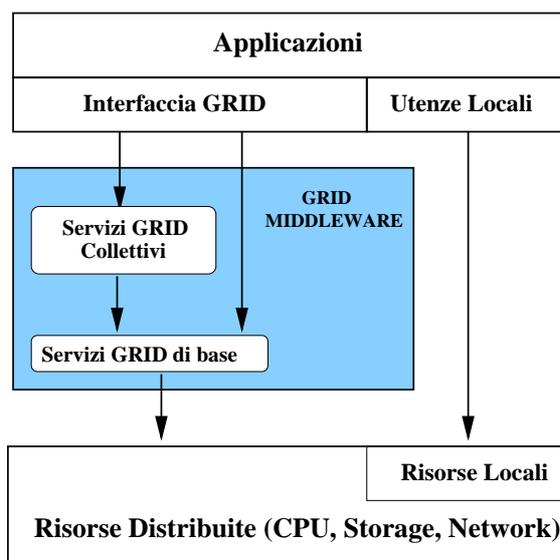


Figura 3.2: Il *middleware* grid.

A prescindere dal grid computing l'ottenimento di un risultato computazionale presuppone la disponibilità di un'applicazione ed un insieme di risorse hardware su cui eseguire la stessa: nel caso della singola workstation (e con opportune estensioni nei cluster) è il Sistema Operativo (S.O.) che fornisce lo strato intermedio per consentire tale accesso. In figura 3.2 è stato indicato complessivamente con il box "Utenti Locali" la collezione di applicazioni che con il supporto del S.O. attingono risorse hardware (CPU, memoria, spazio disco) necessarie alla loro esecuzione direttamente dalle "Risorse Locali" che, come evidenziato dalla figura 1, possono a loro volta concorrere a popolare l'insieme delle risorse distribuite. Per accedere a queste ultime (discuteremo tra poco le modalità di accesso) il S.O. da solo non è più sufficiente: è qui

che si innesta la scelta architetturale grid la quale prevede una interfaccia attraverso la quale si accede ad un livello intermedio di supporto al S.O., che consente alle applicazioni di raggiungere le risorse condivise collocando, come appare dalla figura 3.2, insiemi di servizi collettivi e di base proprio tra applicazioni e risorse. È l'insieme di questi "servizi" che costituisce il cuore di tale architettura. A questo punto risulterà chiaro il riferimento al *middleware* grid come ad una "architettura orientata ai servizi"; servizi che consentono ad applicazioni di accedere a risorse distribuite su qualsiasi scala (locale, metropolitana, geografica, ...)

3.3.1 Grid: i servizi

Passiamo a questo punto a descrivere i servizi principali che sono alla base di qualsiasi *middleware* di griglia. Partendo dal concetto di "servizio" precedentemente introdotto riportiamo in figura 3.3

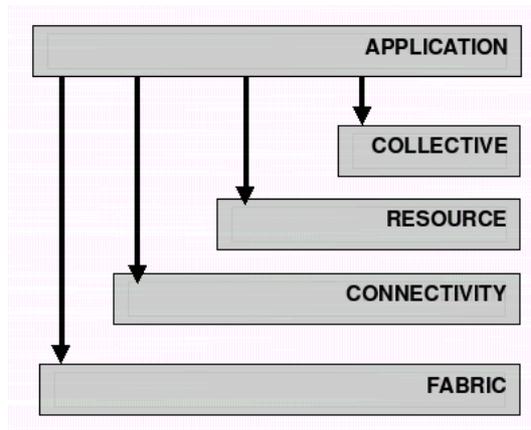


Figura 3.3: L'architettura grid

il così detto modello architetturale a clessidra per descrivere gli aspetti "operativi" dei servizi nei vari livelli: passeremo quindi da una descrizione strutturale dei servizi, ad una funzionale. Un aspetto fondamentale del mo-

dello in figura 3.3 riguarda l'accessibilità del livello applicazioni a tutti quelli sottostanti in maniera autonoma ed indipendente: riportiamo ora brevemente il ruolo svolto dai vari livelli.

1. LIVELLO APPLICATION; è il livello entro il quale sono collocati tutti i servizi (e più in generale le risorse) che consentono all'utente di sfruttare le risorse di griglia: in particolare in tale livello si trovano tutti i servizi orientati alla gestione dell'accesso alla griglia ed alla gestione della stessa (è chiaro a questo punto il senso della nota fatta in precedenza sul fatto che tale livello possa accedere a "tutti" quelli sottostanti).
2. LIVELLO COLLECTIVE; si occupa (attraverso i servizi ed i protocolli che vi appartengono) di "coordinare le risorse" svolgendo la ricerca di risorse con certe proprietà, di co-allocarle (allocazione di più risorse), di replicare dati. In particolare appartengono a tale livello servizi come:
 - Workload Manager
È il regista della sottomissione attraverso meccanismi di schedulazione ad alto livello (che tengono conto cioè dello stato attuale delle risorse nella griglia)
 - Directory Service
È responsabile della visibilità delle risorse nell'ambito delle VO di appartenenza.
 - Monitoring
Recupera tutte le informazioni sullo stato della griglia e ne consente la pubblicazione.
 - Diagnostic
È il servizio che offre la possibilità di svolgere diagnosi sullo stato dei vari componenti.

3. LIVELLO RESOURCE; questo livello si preoccupa della gestione della risorsa locale curando in particolare la gestione della sua allocazione, dell'accesso e del trasferimento dati e dello stato della stessa.
4. LIVELLO CONNECTIVITY; in questo livello si trovano l'insieme dei protocolli di comunicazione e di sicurezza necessari ad implementare lo scambio di dati tra le risorse.
5. LIVELLO FABRIC; trovano collocazione in questo livello tutte le risorse da condividere in griglia: secondo la terminologia introdotta in precedenza, il livello fabric è una collezione di "entità" fisiche e/o logiche.

Presenteremo ora più in dettaglio i servizi specifici su cui è costruita l'architettura grid discutendo in linea generale la specificità delle loro azioni e la loro interoperabilità. A partire dal modello a clessidra possiamo classificare i servizi come appartenenti ad uno dei tre seguenti livelli:

1. COLLECTIVE SERVICE
2. RESOURCE SERVICE
3. CONNECTIVITY SERVICE

Vediamo quali servizi appartengono ai gruppi sopra indicati e quali compiti svolgono.

1. COLLECTIVE SERVICE
 - GSI (Grid Security Infrastructure): si occupa di garantire l'accesso sicuro alle risorse implementando meccanismi basati sull'uso di certificati digitali x.509.
 - Servizi di Network: vi appartengono i protocolli di rete standard che consentono l'accesso alle diverse entità.

2. RESOURCE SERVICE

- Gatekeeper: è il servizio che si occupa sia dell'interfacciamento tra le risorse computazionali ed i servizi di griglia collettivi, che della "mappatura" degli utenti di griglia muniti di certificato, in utenti locali al fine di consentire agli stessi di eseguire job sulla risorsa: la mappatura segue la regola del *Single Sign On* (SSO).
- Local Resource Manager System: è il servizio che si occupa del sistema di code per l'accesso alle risorse locali e della gestione delle relative policy e priorità. Appartengono a tale categoria anche i sistemi per la pubblicazione dello stato di utilizzo di CPU e memoria delle risorse locali al fine di consentire il *Brokering* al livello collective.
- Monitoring: si occupa propriamente della pubblicazione di informazioni generali di carico ed utilizzo delle risorse.
- Storage: è costituito dalla collezione di servizi e protocolli per gestire il trasferimento di dati, l'interfaccia con i cataloghi collettivi e la consultazione dei dati presenti.

3. CONNECTIVITY SERVICE

- VOMS (*Virtual Organization Membership Service*): consente una gestione avanzata degli utenti attraverso l'accesso trasversale rispetto a più VO, alle informazioni sugli stessi consentendo così il superamento delle costrizioni gerarchiche presenti in LDAP (Lightweight Directory Access Protocol).
- Servizi di gestione del carico: si preoccupano dello scheduling dei task sulla base dell'analisi delle risorse disponibili da un lato e dei criteri di ottimizzazione imposti dagli amministratori da l'altro: il tutto in un contesto di totale trasparenza per l'utente finale.

- **Data Management:** è il servizio che si occupa della nomenclatura dei file a livello collettivo al fine di renderli univocamente identificabili, della loro creazione ed eliminazione, della loro replica, se necessario, su più sistemi di storage (preoccupandosi di preservarne la coerenza) e dell'accesso in remoto evitando il trasferimento tra postazioni.
- **Servizi di Monitoring:** l'insieme dei servizi di monitoring propriamente detti preposti alla raccolta di informazioni ed alla loro pubblicazione: tali servizi sono indispensabili soprattutto ai fini dell'amministrazione della griglia consentendo di intervenire tempestivamente per rimuovere situazioni di fault o gestire problemi di traffico.
- **Information Index:** è il servizio che raccoglie le informazioni sulle risorse disponibili preoccupandosi con continuità di scoprirne di nuove e di aggiornare l'elenco di quelle presenti rendendosi sempre disponibile alle interrogazioni del workload management.

3.3.2 Aspetti generali sulla sicurezza delle griglie

La sicurezza sulla griglia è basata sullo standard PKI X.509 ovvero il sistema *Public Key Infrastructure* con certificazione digitale (X.509) della chiave pubblica. Tale modello è nato prima del grid computing per rispondere ai requisiti fondamentali in tema di comunicazioni sicure tra cui:

1. certezza della conformità dell'informazione scambiata;
2. certezza sulla autenticità di sorgente e destinazione dell'informazione;
3. certezza della riservatezza dei dati che compongono l'informazione;

La sicurezza e l'affidabilità in ambito grid (come in tutti i contesti informatici) fanno riferimento a tutti gli strumenti necessari a garantire due processi (o fasi) fondamentali:

1. L'autenticazione;
2. L'autorizzazione.

Durante la prima fase il sistema deve trovare la risposta alla seguente domanda: chi è questo utente ? Nella seconda invece, la domanda è : a quali risorse ha accesso questo utente ? La fase di autenticazione è sostenuta dalle *Certification Authorities* (CA), le quali dopo aver verificato l'identità della persona con metodi "tradizionali", rilasciano un certificato digitale personale che costituisce la carta d'identità digitale dell'utente nella griglia. Ad ogni richiesta di risorsa l'utente deve allegare una copia del proprio certificato per comprovare la propria identità. Le CA sono configurate per rispondere a proprie politiche e procedure di comportamento che prevedono anche l'esistenza di liste di altre CA di cui "fidarsi" oppure no. Ogni CA inoltre pubblica una *Certificate Revocation List* (CRL) ovvero una lista con i certificati che sono stati compromessi e quindi revocati. Data la necessità di esibire il proprio certificato digitale ad ogni richiesta di risorsa grid, l'architettura prevede la creazione di un *proxy* il quale include l'identità dell'utente, una chiave privata (specifica del *proxy*) e la data di scadenza: in questo modo l'utente delega la propria identità al job per un periodo di tempo limitato. L'autorizzazione all'uso delle risorse prevede che l'utente abbia già superato la fase di autenticazione e che abbia registrato il proprio certificato con una o più *Virtual Organizations* (VO). La gestione vera e propria dell'autorizzazione ad accedere alle risorse viene fatta come già sottolineato dal servizio VOMS il quale oltre a "propagare" le credenziali di autenticazione (il certificato digitale) diffonde le informazioni relative alla VO di appartenenza e al ruolo dell'utente in essa: in questo senso VOMS consente una gestione molto raffinata dell'accesso alle risorse su griglia.

3.4 Il ciclo di vita di un job nella griglia

Per fornire un quadro generale sulla funzionalità dei vari componenti grid fin qui discussi, illustrerò i passaggi cui è soggetto un job mandato in esecuzione su un griglia e, per mezzo di rappresentazioni schematiche, il ruolo svolto dai vari servizi e dalle risorse che la compongono: in questo modo gran parte dei termini presentati potranno trovare una più razionale collocazione all'interno di una architettura così articolata.

Il soggetto della rappresentazione è quindi costituito dal “ciclo di vita” di un job a partire dalla sottomissione alla griglia fino alla consegna dell'output: in questo modo avremo modo di spiegare il lavoro svolto dalle singole componenti e gli aspetti “relazionali” che costituiscono l'essenza della loro interoperabilità e del grid computing stesso. Con riferimento alla figura 3.4 analizziamo i flussi relazionali tra componenti/servizi che concorrono alla esecuzione di un job in ambiente grid: descriveremo l'intero ciclo che parte dalla sottomissione di un input fino alla consegna del corrispondente output isolando di volta in volta un “sottogruppo” di componenti presenti nella figura 3.4 discutendone proprietà e funzioni.

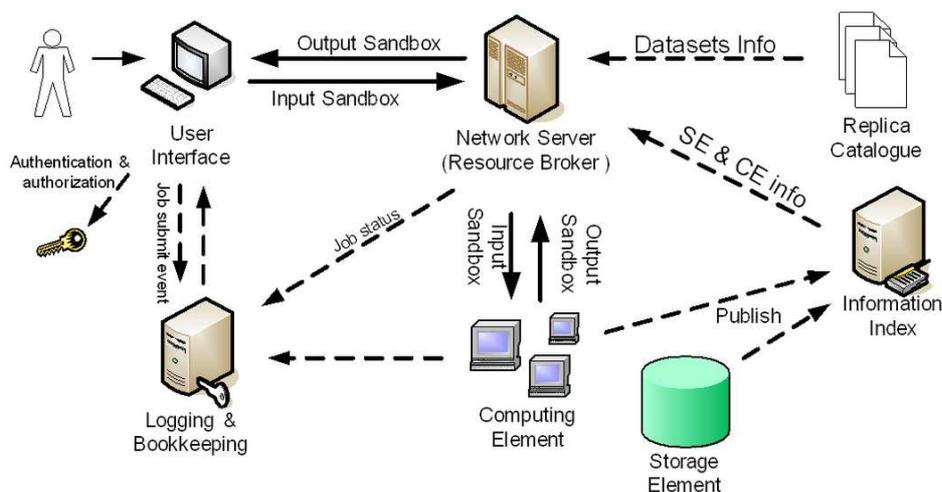


Figura 3.4: Interazioni tra le varie componenti *grid* durante l'esecuzione di un job.

3.4.1 User Interface e Logging and Bookkeeping

Nella figura 3.5 è riportato il sottogruppo di componenti rispetto a alle quali l'utente interagisce non appena deve accedere alla griglia. In particolare la *User Interface* (UI) rappresenta il componente client di accesso che può a seconda dei casi essere costituito da una interfaccia web o più semplicemente da una *Command Line Interface* (CLI). La seconda componente presente in questo sottogruppo si occupa della gestione dell'autenticazione nonché della verifica dei diritti e delle credenziali: tutti questi compiti possono far riferimento per esempio, ad un server VOMS. L'ultimo componente in questo sotto gruppo è rappresentato dal *Logging & Bookkeeping* (LB) il quale

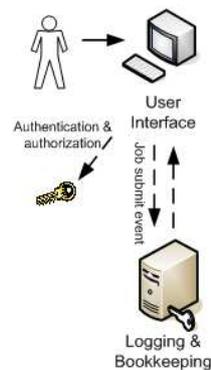


Figura 3.5: User Interface

tiene traccia di tutte le azioni compiute dall'utente e dagli altri servizi.

3.4.2 Resource broker

Quando l'utente deve sottomettere un job descrive lo stesso tramite il *Job Description Language* (JDL) ed invia la richiesta di sottomissione al *Resource Broker* (RB) "agganciando" alla richiesta una serie di file (*Input Sandbox*).

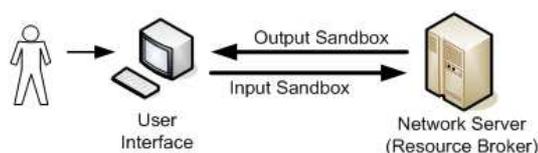


Figura 3.6: Resource Broker.

Nel documento JDL l'utente si identifica tramite *proxy*, specifica i requisiti necessari all'esecuzione, definisce una funzione di qualità (rank) e specifica se il job deve essere eseguito in prossimità di alcuni file (*InputData*) oppure no. Il *Resource Broker* accetta la richiesta di

sottomissione (fig.3.6), analizza il documento JDL e decide a quale risorsa inviare la richiesta nell'ambito di una lista in suo possesso: per svolgere queste funzioni si avvale dell'*Information Index* e del *Replica Catalogue* come indicato dal sottogruppo in figura 3.7.

Al termine dell'esecuzione sarà ancora il RB a raccogliere l'output prodotto e a mantenerlo in locale in attesa che l'utente inoltri la richiesta di recupero (*Output Sandbox*, fig.3.8). Per decidere quale sia la risorsa a cui inviare il job il RB si avvale di un algoritmo di *matchmaking*. Una volta eseguito il *matchmaking*, il RB decide a quale *Computing Element* (CE) inviare il job insieme all'*Input Sandbox*: il CE accoda la richiesta e dopo l'esecuzione produce l'output che verrà inviato di nuovo al RB (*Output Sanbox*). A "run time" il job potrà accedere ai vari *Storage Element* su cui sono salvati i file necessari alla sua esecuzione (fig.3.9). Il CE comunica al LB ogni variazione sullo stato del job

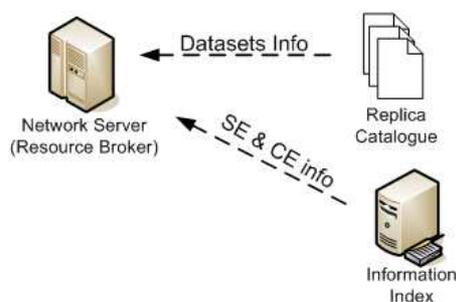


Figura 3.7: Replica Catalogue.

per cui l'utente potrà risalire a tale informazione interrogando lo stesso LB e richiedere così l'output al termine dell'intero ciclo.

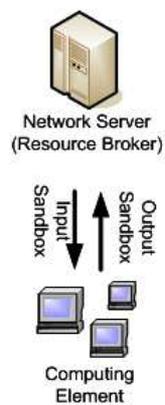


Figura 3.8: Consegna dell'output.

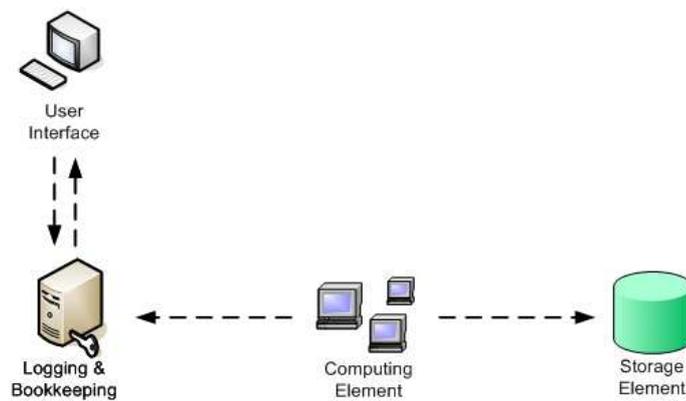


Figura 3.9: Loggin & Bookkeeping.

Capitolo 4

Calcolo numerico di frequenze su griglia computazionale

4.1 Introduzione

In questo capitolo è descritta l'implementazione del software grid oriented GaussianGrid. Le implementazioni di questo sistema alle quali ho partecipato durante il mio dottorato, sono due e fanno uso di due diverse soluzioni tecnologiche. La prima è stata realizzata in collaborazione con il centro di supercalcolo di Roma *CASPUR*¹ ed è basata sulla tecnologia dei *web service* mentre la seconda, sviluppata nel laboratorio *LSDM* di Napoli presso cui ho svolto il mio dottorato, è basata sulla tecnologia web ordinaria (senza *web service*) e *middleware* Globus e gLite. È stata sviluppata prima l'applicazione “*service-oriented*” in collaborazione con il *CASPUR* in quanto questa soluzione fornisce una infrastruttura di base (una sorta di contenitore di servizi) ed una serie di servizi (i *web service appunto*) che svolgono determinate azioni. Nel nostro caso quindi, non ci siamo preoccupati di ricostruire l'intero *middleware* (il contenitore) ma ci siamo concentrati sulla implementazione del servizio vero e proprio: in questo modo abbiamo diminuito i

¹<http://www.caspur.it>

tempi di realizzazione e ci siamo concentrati sulle prestazioni della soluzione algoritmica.

Dall'analisi dei risultati ottenuti con questo strumento e che illustrerò nei prossimi paragrafi, è emerso che la soluzione di calcolo distribuito per la determinazione di *proprietà di griglia* attraverso il codice Gaussian03 è "scalabile": nel seguito spiegherò il significato di questa affermazione e quello di *proprietà di griglia*. A partire da questi risultati ho realizzato presso *LSDM* una nuova versione di GaussianGrid che rispetto a quella sviluppata in collaborazione con il *CASPUR*, non è basata sulla tecnologia dei *web service* ma sul linguaggio di scripting PHP e finalizza i risultati del calcolo distribuito nella determinazione delle frequenze vibrazionali, implementando il *post processing* non presente nella prima versione.

Nelle prossime sezioni descriverò le due implementazioni di GaussianGrid e, per quella sviluppata presso *LSDM*, l'applicazione per il calcolo numerico delle frequenze di vibrazione.

4.2 GaussianGrid con la tecnologia WS

La storia dell'informatica degli ultimi anni ha avuto come protagonista indiscusso il networking, ovvero l'insieme di tecnologie hardware e software per connettere due o più computer. In particolare sul fronte software hanno fatto la loro comparsa i così detti *web service*: per quanto riguarda il livello applicativo al quale siamo interessati in questa tesi, possiamo considerarli come una serie di servizi fruibili attraverso il web e disponibili, per altri servizi, attraverso un protocollo di comunicazione. Questa tecnologia è alla base dell'infrastruttura di GaussianGrid sviluppato con il *CASPUR* e presentata in fig. 4.1.

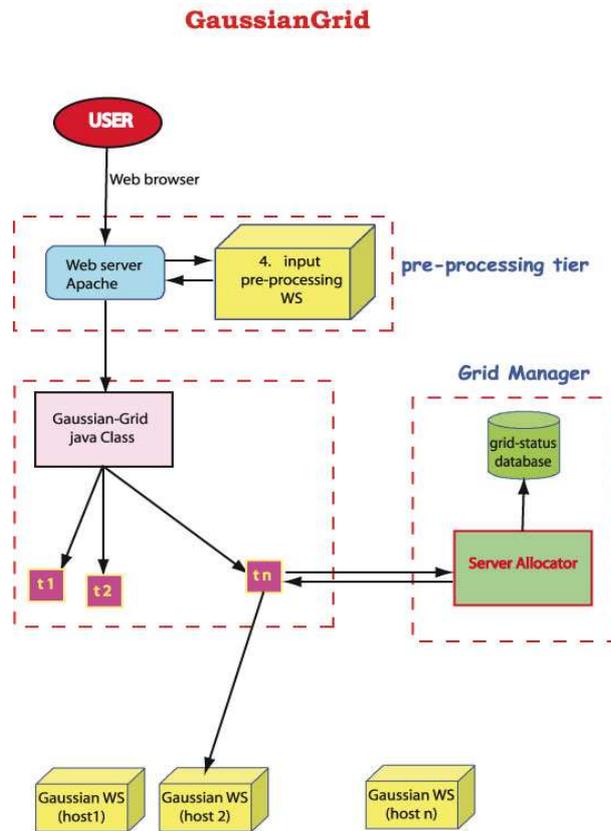


Figura 4.1: L'architettura GaussianGrid

L'architettura sviluppata prevede tre livelli ciascuno dei quali svolge un determinato compito:

1. pre-processing;
2. task execution;
3. data collector:

in fig. 4.1 sono riportati in dettaglio le funzionalità dei primi due. Il primo step nel flusso di esecuzione (a livello di pre-processing) è implementato per mezzo di una interfaccia web grazie alla quale un *web service* interagisce con

l'utente per realizzare le fasi di autenticazione, data input e configurazione del job da sottomettere. Successivamente il WS costruisce la collezione di input Gaussian03 a partire dai dati forniti dall'utente: tutta questa fase è gestita a livello del server http su cui è in esecuzione il WS e tutti i dati sono locali alla stessa risorsa. La seconda parte nel processo di esecuzione invoca una *classe java* [62] responsabile della generazione dei processi di esecuzione per ognuno degli input Gaussian03. Tale classe dopo aver contattato il *grid manager* invia ai nodi della griglia lo script di esecuzione dei job. In tale architettura questa fase è sicuramente la più critica rispetto alle prestazioni complessive in quanto dalla risposta del *Server Allocator* dipende il *load balancing* dei vari job su tutta la griglia. L'ultima fase del processo prevede la raccolta dei vari output ed il loro utilizzo. In questa implementazione di GaussianGrid non è stata realizzata nessuna applicazione di post processing: come anticipato in precedenza in questo lavoro ci siamo concentrati sulla fattibilità del progetto e dell'analisi delle prestazioni e dopo la verifica positiva delle stesse abbiamo realizzato presso il laboratorio *LSDM* la versione di GaussianGrid con l'applicazione di post processing per il calcolo numerico delle frequenze di vibrazione di cui parlerò più avanti. Il progetto di calcolo grid prevede la generazione e la distribuzione di vari job Gaussian03 nei quali vengono calcolate proprietà molecolari (energia, gradiente, momento di dipolo, etc ...) su una griglia cartesiana ottenuta muovendo tutti gli atomi della molecola in tutte le direzioni di una stessa quantità uno alla volta. A partire da questi dati è possibile realizzare diverse applicazioni numeriche indipendentemente dall'architettura grid: in questo lavoro ci siamo quindi concentrati sui primi due livelli per studiare la risposta in termini di prestazioni dell'intera architettura, argomento che tratterò nella prossima sezione.

4.3 Distribuzione dei test paralleli ed analisi delle prestazioni

Per misurare le prestazioni di GaussianGrid abbiamo scelto il calcolo QM per la molecola d'acqua in due livelli di teoria/base, PCM/HF/6-31G* e B3LYP/6-31G* su una griglia cartesiana di $6N$ punti. Sui vari input Gaussian03 abbiamo di volta in volta modificato il parametro $Nproc$ in modo da calcolare l'efficienza del parallelismo intrinseco del codice Gaussian03 per metterla a confronto con quella della classe java che si occupa della distribuzione dei vari job sulla griglia. Le misure sono state eseguite impiegando due nodi AMD bi-processore e due nodi AMD quadri-processore interconnessi da una rete LAN. Le tabelle 4.1 e 4.2 riportano il tempo di esecuzione dei singoli processi (in secondi) per *diverse combinazioni di distribuzione degli stessi sulle varie CPU della grid* oltre ai tempi relativi ai job G03 ed alla classe java responsabile dell'allocazione sui vari nodi. I dati riportati in tab. 4.1 mostrano che l'efficienza del codice G03 è leggermente influenzata dall'esecuzione su griglia rispetto all'esecuzione "libera" (di circa il 10%). Questi risultati sono in linea con le performance attese nel caso di job molto veloci dove il ritardo di tempo dovuto all'allocazione su griglia (1-2 secondi a seconda del numero di processi allocati) influenza molto l'*overall elapsed time*. Tuttavia anche in questo caso limite (esecuzione di job molto veloci) l'efficienza del G03 sulla griglia passa dal 95% (4 processi su 4 macchine) al 98% (12 processi su 4 macchine). Si intuisce dai dati che l'*overhead* dovuto all'allocazione dei jobs su griglia è quasi costante e che dipende dal numero di nodi coinvolti oltre che dai processi. Infatti ripetendo lo stesso set di misure con job Gaussian più pesanti (vedi tab. 4.2) troviamo il solito overhead di 1-2 secondi sui singoli processi *ma in questo caso l'efficienza nel caso di coinvolgimento di tutti i nodi è molto vicina al 99%*. In tabella 4.2 è stato riportato per completezza anche lo SpeedUp del codice G03 che risulta essere simile a quello del codice fuori dalla griglia (3.2 per 4 CPU) dimostrando che l'implementazione

su griglia di Gaussian03 comporta solo un piccolo overhead sul parallelismo intrinseco del codice.

Machine	T_{Eff}	$4xN_1$	$4xN_2$	$2xN_4$	$2x4N_1$	$4x2N_1$	$2x2N_1+2x4N_1$
P2_1	Java (s)	21.4	15.3			22.9	22.7
	G03 (s)	20.3	13.7			20.1	20.2
	%Eff	94.8	89.5			87.8	90.0
P2_2	Java (s)	21.5	15.2			23.1	22.7
	G03 (s)	20.5	13.8			20.3	20.3
	%Eff	95.3	90.8			87.8	91.0
P4_1	Java (s)	22.3	15.5	13.3	22.2	23.1	22.4
	G03 (s)	21.4	14.5	12.1	20.5	21.1	20.6
	%Eff	96.0	93.5	91.0	92.3	91.3	92.0
P4_2	Java (s)	22.0	15.5	13.4	22.2	22.8	22.5
	G03 (s)	21.0	14.5	12.4	20.6	21.2	20.6
	%Eff	95.5	93.5	92.5	92.8	93.0	91.6
Overall	(s)	22.4	15.6	13.5	22.9	23.2	22.9
	%Eff(G03)	94.9	94.5	97.6	99.5	94.8	98.1
	%Eff/Java)	95.5	97.4	98.5	96.9	98.3	97.8

Tabella 4.1: Tempi di calcolo single point (in secondi) ed efficienza (%), HF/6-31G* con PCM . Le sigle P2_ indicano i nodi biprocessore mentre con P4_ indichiamo i quadriprocessori. La notazione $4xN_1$ indica l'esecuzione di 4 processi Gaussian con $Nproc = 1$, $4xN_2$ indica 4 processi con $Nproc = 2$ e così via. Con Overall sono indicate le misure effettuate coinvolgendo tutti i nodi possibili con il numero di processi.

Questo risultato insieme all'alta stabilità del sistema distribuito, giustifica l'implementazione del calcolo distribuito dove centinaia di jobs vengono gestiti per essere eseguiti su centinaia di CPU. L'estensione di tale procedura a calcoli su larga scala può trovare impedimento solo nella disponibilità di griglie computazionali di grandi dimensioni. In questo tipo di calcolo QM su griglia si realizza quello che in ambito informatico è indicato come *embarrassingly parallel applications*, ovvero applicazioni distribuite nelle quali

Calcolo numerico di frequenze

le comunicazioni tra i processi sui vari nodi di calcolo non sono necessarie e i tempi coinvolti nel pre e post processing sono trascurabili.

Machine	T_{Eff}	Overall best performance			Overall average performance		
		$4xN_1$	$4xN_2$	$2xN_4$	$4xN_1$	$4xN_2$	$2xN_4$
P2_1	Java (s)	308.0	159.1		310.5	159.1	
	G03 (s)	306.9	158.0		308.9	158.0	
	%Eff	99.6	99.3		99.6	99.3	
P2_2	Java (s)	307.7	159.9		310.9	159.6	
	G03 (s)	306.6	159.6		309.8	158.5	
	%Eff	99.6	99.8		99.6	99.3	
P4_1	Java (s)	308.7	166.2	95.9	310.3	168.2	97.7
	G03 (s)	307.6	165.2	94.6	309.3	167.2	96.6
	%Eff	99.6	99.4	98.6	99.7	99.4	98.9
P4_2	Java (s)	308.4	165.2	96.0	309.6	164.3	96.7
	G03 (s)	307.2	164.1	94.9	308.5	163.3	95.3
	%Eff	99.6	99.3	98.8	99.6	99.4	98.6
Overall	(s)	308.8	166.3	96.1	311.0	168.3	97.8
	%Eff(Java)	99.6	95.7	99.8	99.5	94.5	98.9
	SpeedUp	1.00	1.86	3.21	1.00	1.85	3.18

Tabella 4.2: Calcolo single point di energia (B3LYP/6-31G(3df,2p)); tempi di calcolo (in secondi), efficienza (%) e SpeedUp.

Per fornire un quadro più completo delle possibilità offerte dal *porting* di Gaussian03 su griglia, è stato realizzato il calcolo QM delle funzioni GTOs (Gaussian Type Orbitals) e loro derivate su una griglia cartesiana 3D (un cubo). Il codice che abbiamo adoperato per confrontare le prestazioni con quelle di GaussianGrid è SCELlib [75] il quale consente di valutare le funzioni GTO e le loro derivate intorno ad un centro di espansione. È stato scelto questo codice per l'omogeneità dell'algoritmo e per il fatto di disporre della curva di SpeedUp su una macchina parallela SMP fino a 32 CPU (IBM p691).

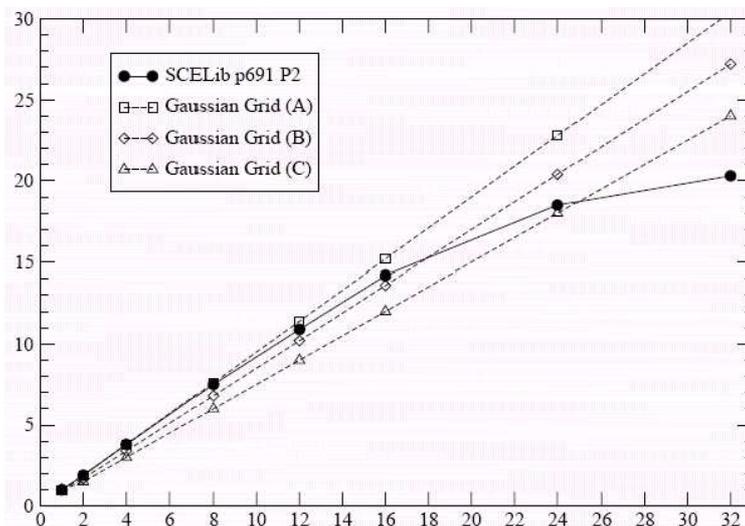


Figura 4.2: Curva di scalabilità per GaussianGrid per il mapping di funzioni GTO su una griglia cartesiana 3D. I dati si riferiscono ad una efficienza del codice su grid pari al 95% (A), 85% (B) e 75% (C) confrontata con le prestazioni del codice parallelo SCELib su IBM p691 (vedi tabella 9 in Ref. [75]).

Dalla fig. 4.2 risulta evidente che anche nella condizione di minima efficienza, il calcolo distribuito risulta più vantaggioso rispetto al calcolo su macchina shared-memory.

4.4 GaussianGrid presso *LSDM*

In questa sezione discuterò l'implementazione del progetto GaussianGrid presso il laboratorio *LSDM* che, come anticipato, non si serve della tecnologia dei WS ma del tradizionale linguaggio di web scripting PHP. A differenza del progetto sviluppato in collaborazione con il *CASPUR*, in questa versione è stata implementata una applicazione di *post processing*: il calcolo numerico delle frequenze di vibrazione. A tale proposito, è importante sottolineare che nell'ambito dei calcoli TD-DFT sono disponibili i gradienti analitici nella versione di sviluppo del codice Gaussian03 [45] ma non le derivate seconde

dell'energia, calcolabili solo numericamente: il calcolo su grid di tale proprietà costituisce quindi, (in modo esclusivamente numerico) una vera alternativa. Recentemente sono stati sviluppati vari codici per la determinazione delle frequenze vibrazionali [63], per il calcolo di proprietà molecolari medie [64, 65] e funzioni termodinamiche [66] in grado di andare oltre l'approssimazione armonica. In tutti questi casi è necessario disporre di proprietà su griglia insieme alle loro derivate (in alcuni casi fino al terzo ordine) per l'applicazione dei rispettivi metodi: è chiaro quindi il campo di applicabilità del metodo qui proposto. Se consideriamo poi gli approcci semi-classici [67], il modello del calcolo di proprietà su griglia può essere usato per calcolare *reaction rates* con l'inclusione di effetti anarmonici e di *tunneling* [63]: le analisi dei diversi percorsi reattivi (IRC, QST2, etc...) possono essere inviate ai vari nodi della griglia computazionale. Un altro settore di utilizzazione riguarda le applicazioni che richiedono il calcolo dell'energia e/o proprietà in diversi punti dello spazio k , ad esempio nel campo dei polimeri [68, 69] e delle superfici [70, 71]. Infine, il calcolo distribuito può essere estremamente utile nei metodi di dinamica molecolare del tipo Born-Oppenheimer o Lagrangiani estesi (CPMD [72, 60], ADMP [73, 74]).

4.5 Algoritmo per il calcolo delle frequenze

Il calcolo numerico delle frequenze di vibrazione a livello QM prevede la determinazione della matrice *Hessiana* e la sua successiva diagonalizzazione: gli autovalori costituiscono le frequenze di vibrazione e gli autovettori i modi normali: gli elementi di tale matrice sono dati dalla derivata seconda dell'energia rispetto alle tre coordinate cartesiane di ciascun atomo, per un totale di $3N \times 3N$ valori. Per calcolare numericamente gli elementi dell'Hessiano è possibile applicare il metodo delle differenze finite; sarà necessario applicare le formule per le derivate seconde se partiamo dall'energia, mentre sarà sufficiente calcolare le derivate prime se disponiamo dal gradiente.

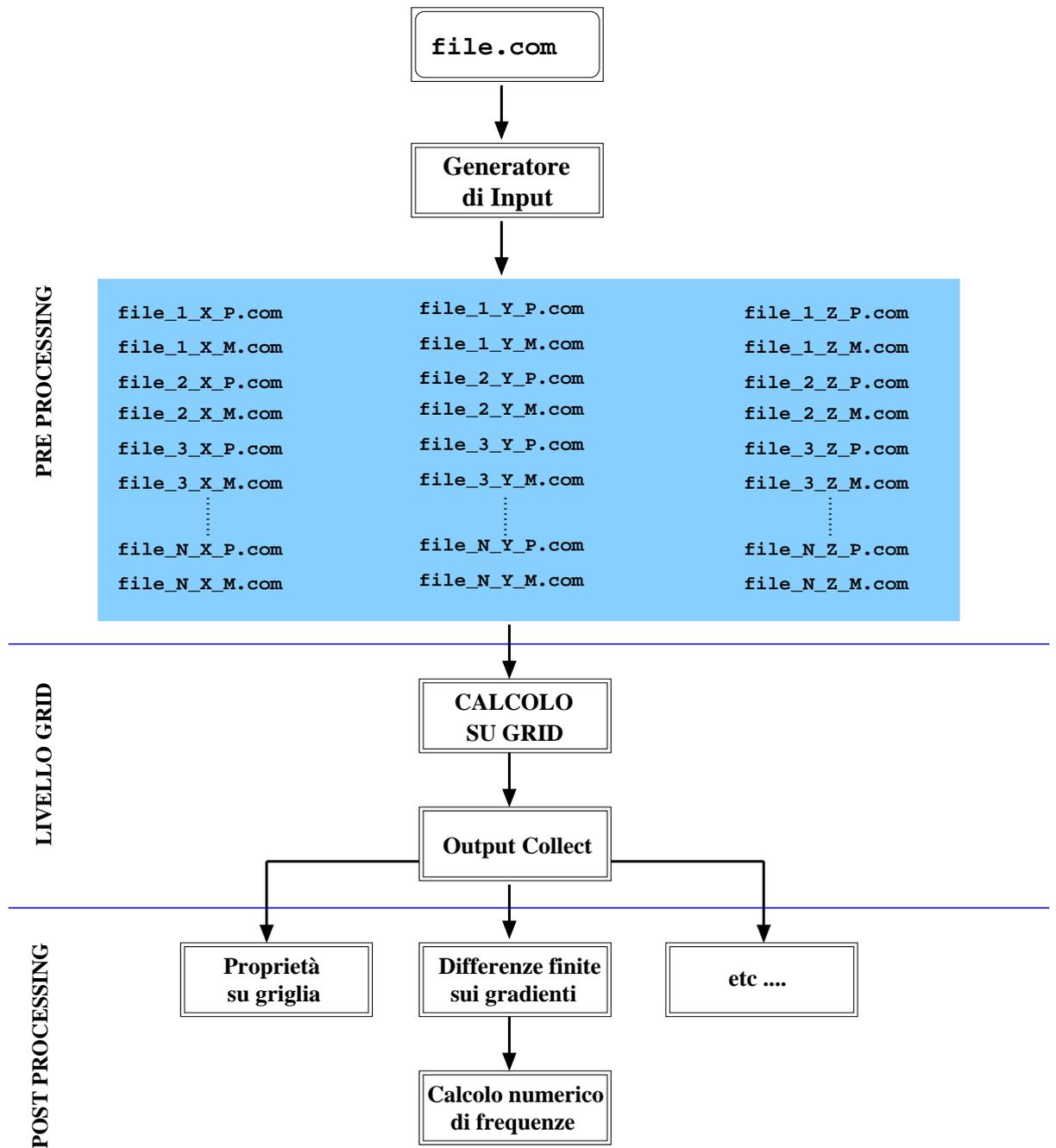


Figura 4.3: Diagramma di flusso di GaussianGrid presso *LSDM*.

Come visto nel Capitolo 2, l'applicazione del metodo TD-DFT consente il calcolo di proprietà di stati eccitati: al momento il codice *Gaussian03* può calcolare in modo analitico solo i gradienti dell'energia mentre per la matrice Hessiana è previsto solo il metodo numerico. *Da questo punto di vista, la realizzazione di una procedura numerica per il calcolo delle frequenze di vibrazione diventa concorrenziale con quella del Gaussian03.* In fig. 4.3 è rappresentato il diagramma di flusso che a partire dal singolo input Gaussian03 fornisce la collezione di output ottenuti deformando la geometria iniziale dentro una griglia cartesiana di $6N$ punti: rispetto al GaussianGrid sviluppato con i WS non vi è fino a questo punto, nessuna differenza a livello procedurale. Dopo aver generato gli output single point alle varie geometrie è possibile processare i file di output per costruire la matrice Hessiana: per tale scopo ho scritto un codice in *Fortran90* integrato nella procedura automatizzata su griglia. Nella prossima sezione discuterò l'applicazione del metodo ad un sistema porfirina-zinco mettendo in luce anche l'interfaccia web (*User Interface, UI*) che ho realizzato per consentire una interazione *user friendly* con l'utente.

4.6 Interfaccia web per GaussianGrid

Per accedere alla procedura per il calcolo numerico di frequenze ho realizzato una interfaccia web² sfruttando la piattaforma di sviluppo LAMP (Linux, Apache, MySQL e PHP) integrata dal gestore di contenuti Open Source *Drupal*³. In figura 4.4 è riportata la schermata nella quale l'utente inserisce l'input Gaussian e tutti i parametri per il calcolo di griglia. In questa fase è possibile scegliere lungo quale direzione spostare gli atomi dalla loro posizione di equilibrio e per quanti punti: dalle equazioni del metodo alle differenze

²Al momento in fase di testing e raggiungibile all'indirizzo <http://village.unina.it/en/GaussianGrid>

³<http://www.drupal.org>

finite sappiamo che possiamo calcolare le derivate di ordine superiore in un dato punto, all'aumentare del numero di punti intorno ad esso.

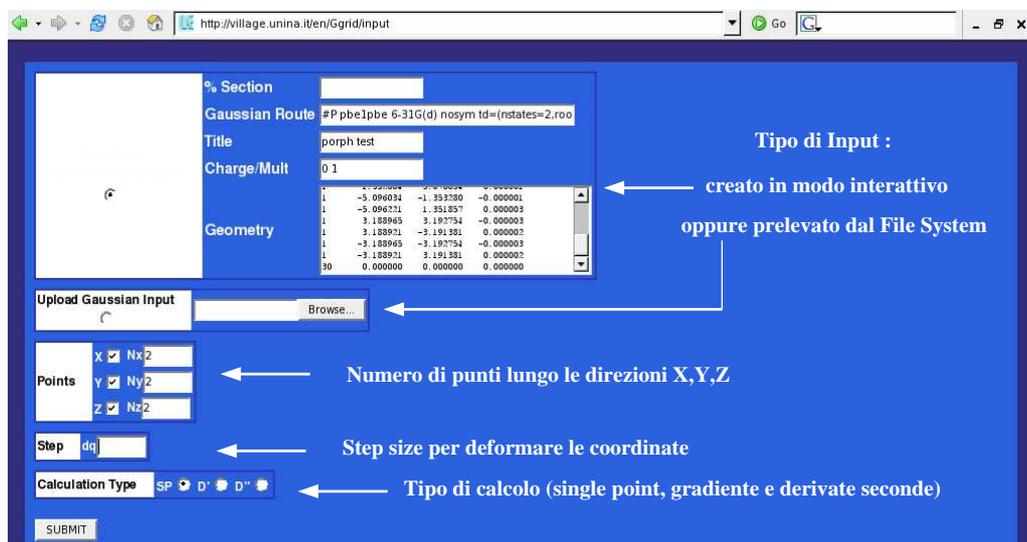


Figura 4.4: Creazione dell'input sul GaussianGrid

Nel caso del complesso porfirina zinco qui trattato ho considerato due soli punti in quanto interessa la derivata prima del gradiente (calcolato in modo analitico da *Gaussian03*) ovvero la derivata seconda dell'energia. Una volta terminata la fase di immissione dei dati attraverso l'interfaccia web, e prima di distribuire il calcolo sulla griglia, viene mostrato il riassunto di quanto inserito nello step precedente: come mostrato in figura 4.5 viene riportato da un lato l'input gaussian testuale mentre sul lato destro viene mostrata la struttura molecolare attraverso l'applicazione java *Jmol*. Per rendere più *user friendly* l'interfaccia, ho aggiunto la possibilità di modificare in modo interattivo la visualizzazione stessa, attraverso le funzione disponibili in *Jmol*. Dopo aver constatato la correttezza dell'input il calcolo può essere affidato alla griglia secondo il flusso riportato in figura 4.3. Una volta disponibili

tutti gli output il post processing eseguirà il calcolo della matrice Hessiana e la sua digonalizzazione.

The figure displays two screenshots of the Jmol interactive viewer interface, showing the input data for a Gaussian calculation and the resulting molecular structure visualization.

Top Screenshot: Shows the Gaussian input file content and the Jmol viewer with a blue background. The input file includes the title "Input Gaussian fornito" and a list of atomic coordinates. The Jmol viewer displays the molecular structure with a blue background and includes controls for "atoms" (off, 20%, 100%), "spin", and "axes". The background is set to "orange".

Bottom Screenshot: Shows the same Gaussian input file content and Jmol viewer interface, but with a different background color. The background is set to "skyblue".

Gaussian Input:

```

Gaussian Input :
#P pbe1pbe 6-31G(d) nosym td=(nstates=2,root=1) force
porph test
Input Gaussian fornito
0 1
7 0.000000 2.028194 0.000000
6 1.096124 2.851802 0.000000
6 2.424719 2.420883 -0.000002
6 2.867304 1.096056 -0.000001
7 2.045785 0.000383 0.000000
6 2.866447 -1.095231 0.000001
6 2.424111 -2.419694 0.000002
6 1.096468 -2.850377 0.000001
7 0.000000 -2.028194 0.000000
6 -1.096124 -2.851802 0.000000
6 -2.424719 -2.420883 -0.000002
6 -2.867304 -1.096056 -0.000001
7 -2.045785 -0.000383 0.000000
6 -2.866447 -1.095231 0.000001
6 -2.424111 2.419694 0.000002
6 -1.096468 2.850377 0.000001
6 0.681641 4.229014 -0.000001
6 -0.682006 4.238774 -0.000001
6 4.246589 -0.680525 0.000002
6 4.246513 0.681921 -0.000001
6 -0.681641 -4.229014 -0.000001
6 0.682006 -4.238774 -0.000001
6 -4.246513 -0.681921 -0.000001
6 -4.246589 0.680525 0.000002
1 1.351773 5.079561 -0.000001
1 -1.352884 5.078854 -0.000001
1 5.096221 -1.351857 0.000003
1 5.096034 1.352380 -0.000001
1 -1.351773 -5.079561 -0.000001
1 1.352884 -5.078854 -0.000001
1 -5.096221 -1.351857 -0.000003
1 5.188921 3.192754 -0.000003
1 3.188921 -5.191381 0.000002
1 -3.188921 3.191381 0.000002
10 0.000000 0.000000 0.000000
    
```

Gaussian Grid Parameters:

```

Gaussian Grid Parameters
check_x = 1 ; Nx = 2
check_y = 1 ; Ny = 2
check_z = 1 ; Nz = 2
delta = ; ctype = 1
    
```

Visualizzatore Jmol interattivo:

atoms off 20% 100% spin axes

background orange

Figura 4.5: Riassunto dell'input inserito

4.7 Test sul complesso porfirina-zinco

Per testare l'infrastruttura e l'algoritmo di calcolo è stato scelto il sistema molecolare presentato in fig 4.6 costituito da un'unità porfirinica coordinata ad un atomo di zinco per un totale di 37 atomi. In figura 4.7 ho riportato l'output relativo al calcolo numerico delle frequenze di vibrazione per il complesso porfirina-zinco direttamente con Gaussian03 e quello nel quale vengono calcolati solo i gradienti in modo analitico: è proprio questo secondo tipo di calcolo che viene diffuso sulla griglia e che nel post processing consente la costruzione della matrice Hessiana.

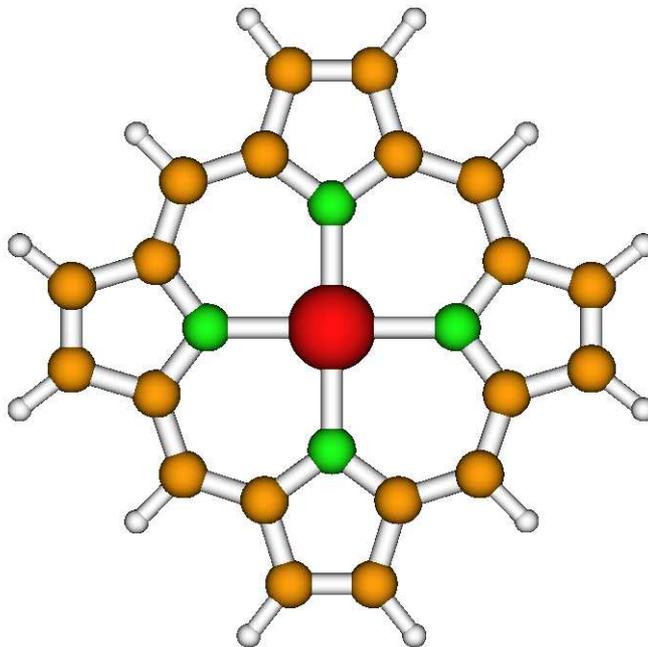


Figura 4.6: Complesso Porfirina-Zinco

Come risulta evidente vi è una notevole contrazione del tempo di calcolo: è ovvio che se la griglia computazionale avesse la dimensione del numero di

job (nel caso del complesso porfirinico $2 \times 3 \times 37 = 222$) il tempo di calcolo su griglia sarebbe pari a quello del singolo job (a meno di latenze trascurabili). Il calcolo su griglia fornisce gli stessi valori ottenuti (a meno di trascurabili errori di arrotondamento) direttamente dal Gaussian per le 105 frequenze di vibrazione ($3N-6$ per $N=37$).

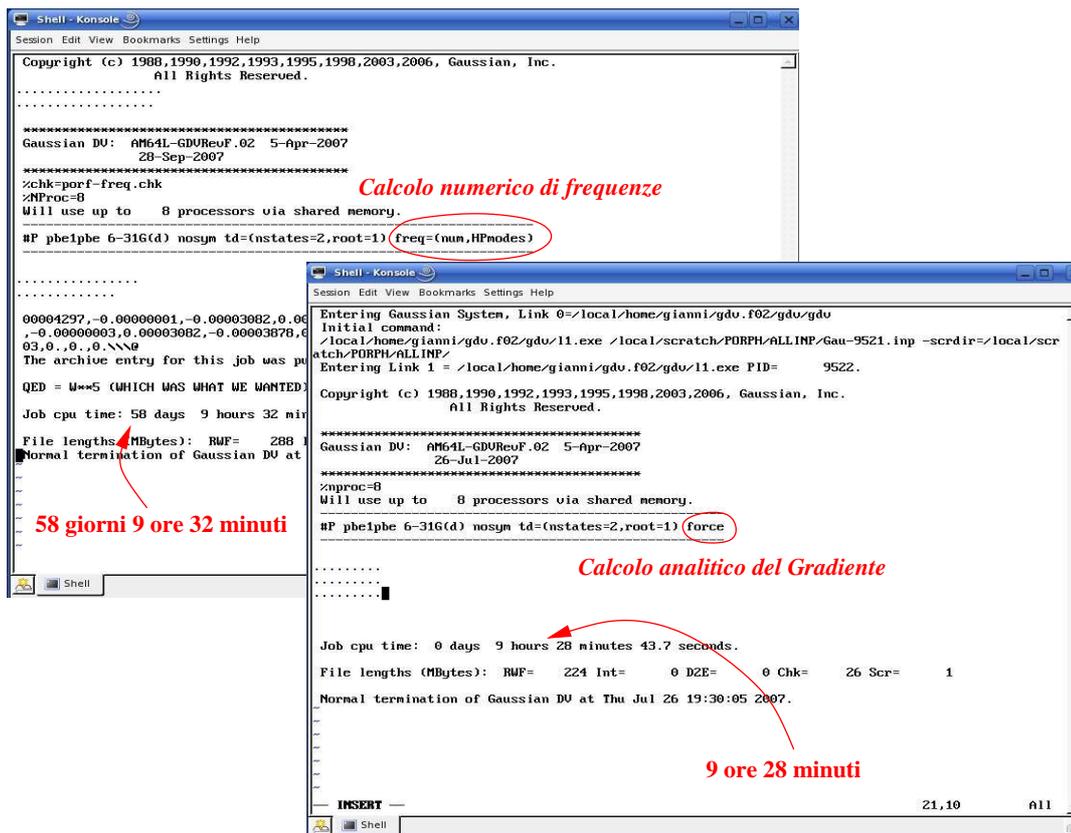


Figura 4.7: Confronto dei tempi di calcolo tra calcolo numerico di frequenze e gradiente analitico con Gaussian03 (in entrambi i casi TD-DFT).

Capitolo 5

Conclusioni

In questa tesi di dottorato ho svolto lo studio di alcune proprietà molecolari utilizzando diversi approcci computazionali, fra i quali il più rappresentativo è quello multiscala: nel Capitolo 1 ho applicato tale approccio nel caso del calcolo delle proprietà spettroscopiche in soluzione di una molecola organica attraverso l'integrazione di metodi QM ed MM. La procedura proposta consiste in tre passaggi:

1. parametrizzazione del campo di forza;
2. campionamento dello spazio configurazionale generato attraverso la dinamica classica;
3. calcolo QM delle proprietà medie sulle configurazioni estratte.

La procedura sopra descritta è stata applicata al caso specifico della molecola di acetone in acqua. Il lavoro svolto ha messo in evidenza che la dinamica classica fornisce risultati confrontabili con quelli forniti da simulazione CPMD e che dalla stessa è possibile ricavare uno spazio configurazionale molto esteso. Per ottenere risultati in accordo con quelli sperimentali è stato necessario modificare il campo di forza standard OPLS-AA introducendo i *lone pairs* sull'ossigeno carbonilico, riparametrizzare lo *stretching* del legame C=O ed il modo di vibrazione fuori dal piano. In effetti la dinamica molecolare ottenuta

con il campo riparametrizzato ha fornito una corretta geometria molecolare ed indirettamente parametri spettroscopici migliori di quelli ottenuti con il metodo Car-Parrinello il quale, come noto, soffre alcune limitazioni connesse con i funzionali DFT utilizzabili. È stato dimostrato inoltre che l'approccio integrato QM/MM può riprodurre in modo accurato parametri spettroscopici come la costante di schermo NMR e l'energia di eccitazione per la transizione $n \rightarrow \pi^*$ in soluzione, purchè sia scelto un livello di teoria opportuno ed un numero minimo di molecole di solvente vengano trattate in modo esplicito a livello QM insieme al soluto. Tale approccio ha il notevole vantaggio di poter affrontare il calcolo QM sulle varie configurazioni a diversi livelli di teoria senza la necessità di dover eseguire di nuovo la dinamica. Tale soluzione si è dimostrata equivalente ai calcoli *full QM*, come dimostrato esplicitamente dal calcolo del *chemical shift* per l' ^{17}O . Inoltre, grazie ad una più estesa scala dei tempi nella simulazione MD, la procedura integrata ha consentito di ridurre l'errore statistico associato alle energie di transizione UV, rispetto a precedenti calcoli QM [11, 21] e QM/MM [9].

Nel Capitolo 2 ho discusso l'approccio TD-DFT allo studio delle transizioni elettroniche nel caso del radicale anione del *p*-benzochinone in soluzione. Le energie verticali di eccitazione per i primi quattro stati di doppietto sono in accordo con quelle calcolate con i metodi *post-Hartree-Fock*. Per avere un buon accordo con i dati sperimentali (massimo della banda di fluorescenza) inoltre, è stato fondamentale ottimizzare la geometria degli stati eccitati senza imporre vincoli sulla conservazione della simmetria. Anche in questo tipo di calcolo gli effetti di solvatazione sono stati presi in considerazione attraverso il modello PCM, mostrando ancora una volta l'efficienza dei modelli "misti".

Il Capitolo 3 tratta gli aspetti tecnologici del mio lavoro di dottorato. In esso vengono illustrati i principi generali del calcolo distribuito attraverso la presentazione dei cluster di calcolatori (di cui viene fornita una descrizione più dettagliata in Appendice A) e del grid computing. Ovviamente per

necessità di sintesi tale argomento viene presentato in maniera compressa.

Nel Capitolo 4 ho presentato lo sviluppo di un'applicazione chimica distribuita: *GaussianGrid*. Per tale applicazione sono state realizzate due versioni: una in collaborazione con il *CASPUR* basata sulla tecnologia dei *web services* e l'altra, interamente realizzata presso *LSDM*, basata su tecnologie web tradizionali. L'applicazione consente di calcolare proprietà molecolari su una griglia cartesiana, e di disporne nella fase di *post processing* per le diverse applicazioni: in particolare nel caso da me realizzato, la proprietà calcolata su griglia è rappresentata dalla matrice delle derivate seconde dell'energia (Hessiano), data la quale è possibile ricavare le frequenze di vibrazione. Tale procedura è stata applicata al calcolo TD-DFT sul complesso porfirina-zinco: per tale livello di teoria il metodo numerico per il calcolo delle frequenze vibrazionali è l'unica scelta possibile, per cui il trasferimento su griglia computazionale diventa una reale possibilità per aumentare la scalabilità parallela del metodo.

In Appendice A infine, ho presentato il lavoro svolto in ambito strettamente tecnologico, sulle due principali risorse di calcolo presso il gruppo *LSDM* di Napoli ed il Dipartimento di Chimica di Padova; i cluster *Bionap* ed *Avogadro* rispettivamente.

Appendice A

I “*nodi*” di Napoli e Padova

A.1 Breve panoramica sulle architetture di calcolatori

In questa appendice descriverò brevemente l'attività tecnologica svolta sulle due principali risorse di calcolo presso il gruppo *LSDM* di Napoli ed il Dipartimento di Chimica di Padova, durante il dottorato di ricerca. Prima, però, ritengo necessario fornire gli elementi essenziali relativi alla nomenclatura delle architetture di calcolatori, al fine di fornire un *vocabolario minimo* per leggere il resto dell'appendice e le sezioni tecnologiche della tesi. Consideriamo a tale proposito la figura A.1. In essa ho riportato, nella parte superiore, la così detta *tassonomia di Flynn* e in quella inferiore le corrispondenti implementazioni hardware dei sovrastanti modelli architetturali (in modo semplificato). Il primo blocco SISD costituisce il modello di base per i monoprocessori mentre quello SIMD è il modello con il quale sono stati realizzati i processori vettoriali, capaci di eseguire le stesse operazioni su più dati (si pensi alla somma di due vettori). Del terzo blocco non esistono, ad oggi, realizzazioni hardware. L'ultimo modello, il MIMD, ha il maggior numero di implementazioni hardware, le quali si differenziano per l'organizzazione della memoria. Se questa è condivisa rispetto ai processori che vi accedono abbiamo le architetture *Shared Memory*. Un ulteriore livello di differenziazione per queste architetture si ha rispetto al modo di accedere alla memoria condivisa: se l'accesso è uniforme si hanno le architetture SMP, mentre se le CPU possono accedere ai dati in memoria in modo non uniforme abbiamo le *NUMA architecture*. I cluster costituiscono infine, la realizzazione pratica più famosa del modello a memoria distribuita nel quale

le memorie (e le CPU) lavorano in maniera concorrente grazie ad un sistema di comunicazione esterno (Fast Ethernet, Fiber Channell, Infiniband, ...). Ovviamente i cluster possono a loro volta essere costituiti da una collezione di nodi SMP e/o NUMA realizzando così un sistema ad architettura mista: le due macchine discusse nel seguito di questa appendice, appartengono a questa famiglia.

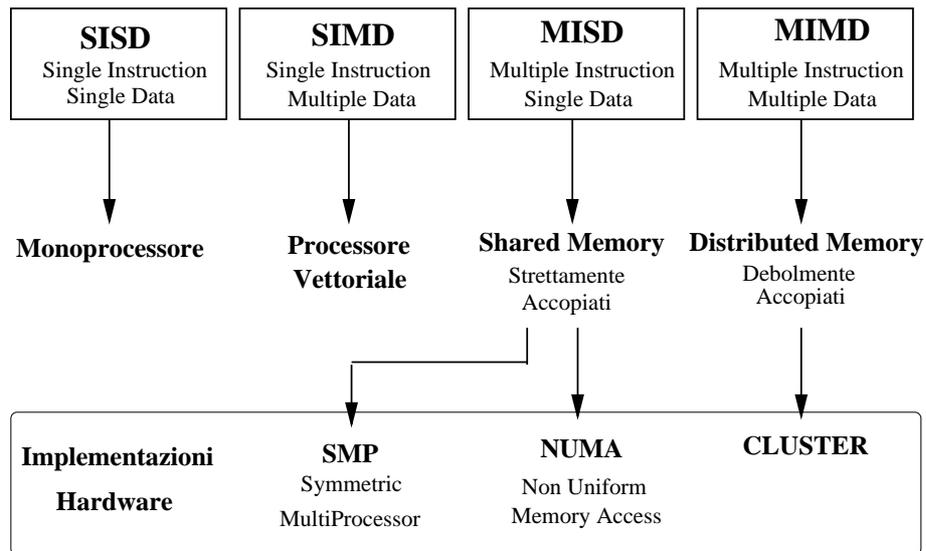


Figura A.1: Le architetture di calcolatori

A.2 Il nodo di Napoli: il cluster “*Bionap*”

Il centro del Dipartimento di Chimica di Napoli dove ho svolto la mia attività di dottorato, dispone di diverse risorse per l’High Performance Computing (HPC). Gli apparati più significativi in termini di prestazioni sono costituiti da due cluster: uno formato da 10 nodi quadri-processore HP ES45 e l’altro da 21 nodi Opteron di cui 5 quadri processori dual core. Accanto a questi due principali componenti vi sono altre risorse dedicate allo sviluppo ed al calcolo dedicato tra cui 4 nodi bi-pro DS20 2 nodi bi-pro Itanium2 ed 1 nodo quadri-pro Itanium2. Su tutti i cluster ed i nodi singoli sono installati i software di calcolo quantistico e di dinamica molecolare. Di seguito descriverò l’architettura e la configurazione del cluster *Bionap* ed i benchmark usati per valutarne le prestazioni.

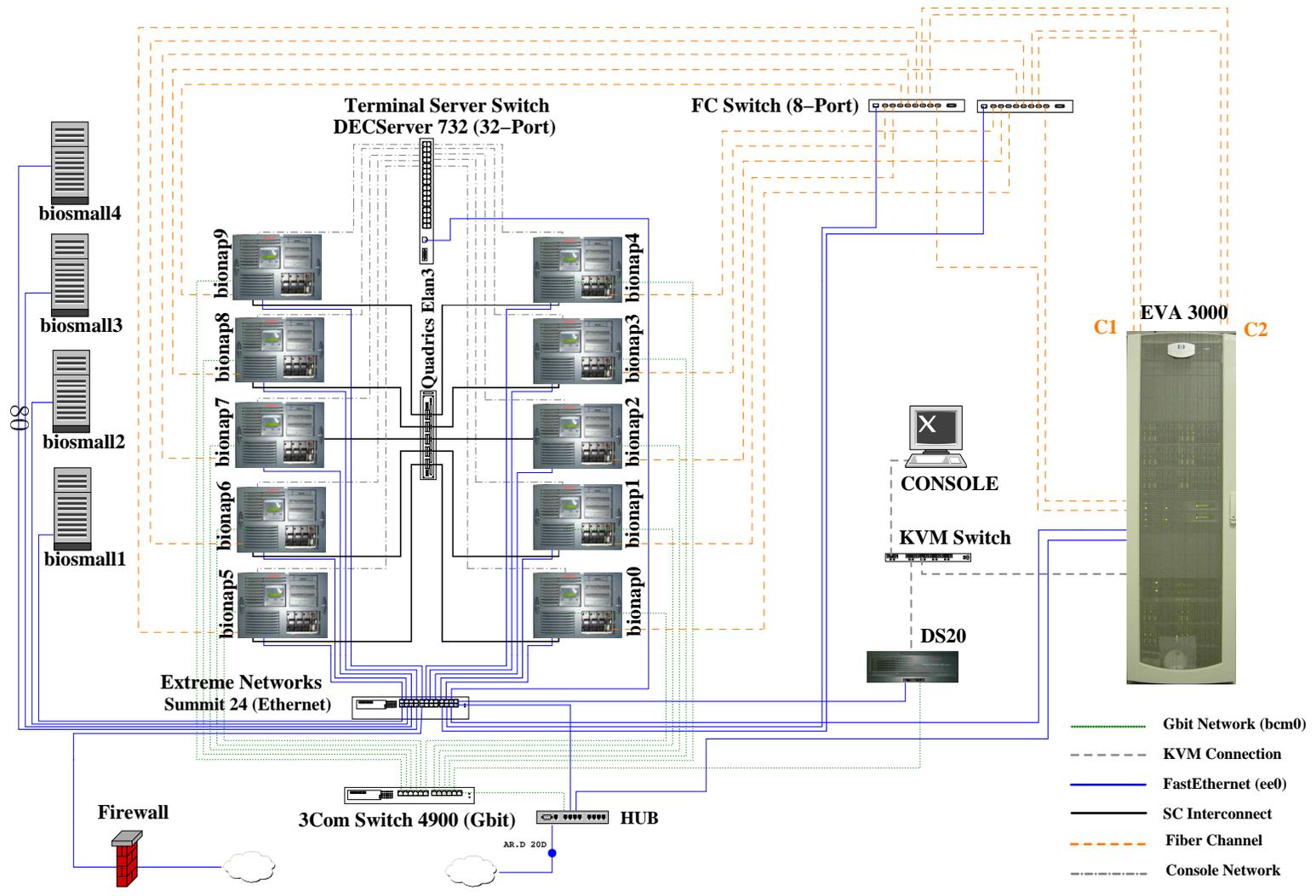


Figura A.2: Schema delle commissioni fisiche del cluster *Bionap* e dei nodi DS20E.

A.2.1 Collocazione dei componenti

I **10 nodi HP-ES45** del cluster sono disposti in due rack del tipo *Standard RET-MA 2.0 M tall racks* . In un terzo rack dello stesso tipo sono state collocate le seguenti unità:

- 1 AlphaServer DS20 (Management Server)
- 1 AlphaServer SC Interconnect Switch (16 porte)
- 1 Terminal Server
- 2 Switch Fiber Channel (8 Porte)
- 1 Switch Gbit-Rame (12 Porte)
- 1 Switch Fast Ethernet (24 Porte)
- 1 Switch KVM
- 1 Console
- 1 Sistema UPS

L'intero sistema dispone di un'unità di storage esterna **HP-EVA300**, collocata in un ulteriore rack e composta da:

- 1 Server Xeon (Windows 2000 server),
- 2 dischi di sistema SCSI Ultra 320 10k da 72 Gb,
- 2 Controller Fiber Channel,
- 2 bay, in cui sono alloggiati 14 dischi Fiber Channel 10k da 146 Gb (per un totale di 2 TB circa di spazio disco)

Le caratteristiche hardware dei Nodi ES45 sono:

- 4 CPU ev68 @ 1250 MHz (16 MB ECC)
- 8 Gb RAM
- 1 scheda Gigabit Ethernet 10/100/1000
- 1 scheda ELAN
- 2 schede Fast Ethernet 10/100

- **1 scheda Fiber Channel 2 Gb/s**

Su tutti i nodi è installato *Tru64 UNIX V5.1b*, *HP AlphaServer SC system software 2.5* compreso il software per il *Cluster File System*, il *Parallel File System* e la piattaforma *LSF*. In figura A.2 è riportato lo schema di interconnessione dei componenti del cluster *Bionap* e dei 4 DS20E.

A.2.2 Configurazione dello spazio disco

- **Spazio locale**

Ogni nodo è fornito di 4 dischi **SCSI 320 da 72 GB** ad eccezione del primo che ne ha 5. Lo spazio associato ai dischi locali è configurato secondo lo schema presentato in tabella A.3, nella quale è riportato, per ogni nodo, la mappatura dei dischi ed il loro utilizzo. Su ogni nodo sono presenti due *partizioni di boot*: quella ordinaria e quella di emergenza (*alternate boot*). Sugli stessi dischi è presente anche una partizione di swap per uno spazio complessivo di 32 GB per nodo.

Due dischi concorrono alla realizzazione di un *filedomain#fileset* su cui sono montate le aree scratch locali. Sul quinto disco del primo nodo è installata una copia locale del S.O.

- **Spazio condiviso**

Nello storage EVA3000 sono presenti n. 14 dischi **Fiber Channel 10k da 146 GB** (spindle) posti su n. 2 contenitori (Shelf disco tipo 4310R) connessi a due controller FC con 2 GB di memoria cache ciascuno. *Dai 14 spindle sono stati ricavati 6 virtual disk.*: in figura A.4 sono riportate sia la configurazione dei virtual disk al livello di storage (EVA3000) che la corrispondente *mappatura* a livello di S.O. sui nodi del cluster.

	bionap0	bionap1	bionap2	bionap3	bionap4	bionap5	bionap6	bionap7	bionap8	bionap9
Boot	dsk0	dsk17	dsk25	dsk13	dsk29	dsk21	dsk9	dsk37	dsk33	dsk41
Swap										
Swap										
Alt. Boot	dsk1	dsk18	dsk26	dsk14	dsk30	dsk22	dsk10	dsk38	dsk34	dsk42
scratch	dsk3	dsk19	dsk27	dsk15	dsk31	dsk23	dsk11	dsk39	dsk35	dsk43
	dsk4	dsk20	dsk28	dsk16	dsk32	dsk24	dsk12	dsk40	dsk36	dsk44
Tru64	dsk2									
	member1	member2	member3	member4	member5	member6	member7	member8	member9	member10

Figura A.3: Spazio disco su nodi del cluster *Bionap*

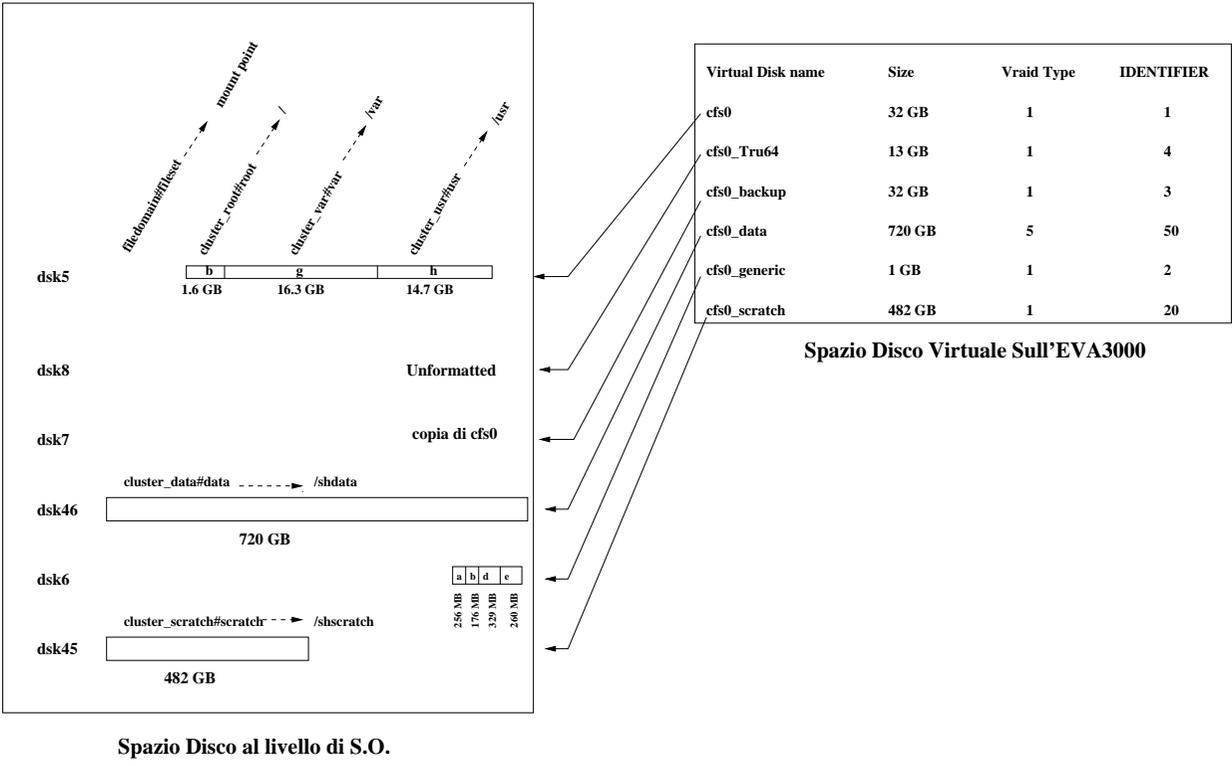


Figura A.4: Spazio disco condiviso sullo storage EVA3000 e mappatura a livello di S.O.

A.2.3 Configurazione di rete

I cluster HP AlphaServer SC prevedono il supporto per i *domini multipli*: nel caso del cluster in esame, abbiamo un unico dominio costituito da 10 nodi. All'interno del cluster sono presenti le reti riportate nella tabella A.1: esse sono previste dalla configurazione standard, il *Cluster Interconnect*, il *System Interconnect* la *management network*, ed una rete esterna *Gbit-Rame*.

Rete	Suffisso IP	Componente
Cluster Interconnect	-icso	AlphaServer SC
System Interconnect	-eip0	AlphaServer SC
Management Network	-ext0	Fast Ethernet
External Network	-ext1	Gbit-Rame

Tabella A.1: Elenco delle connessioni di rete

A.2.4 Benchmarks MPI

I benchmark MPI sono stati realizzati per misurare la banda passante e la latenza dell'architettura SMP e del Cluster Interconnect (Distributed Memory). In figura A.5 **A**₁ e **A**₂ sono riportati i valori di banda in MB/s su architettura **SMP** (due CPU su singolo nodo) mentre in figura A.5 **B** gli stessi dati si riferiscono all'utilizzo del Cluster Interconnect SC. Il codice usato per questo benchmark, realizza la trasmissione/ricezione (send/receive) di un pacchetto con data size variabile fornendo per tale trasferimento, il picco dell'ampiezza di banda: la latenza corrisponde al valore di banda passante quando la dimensione del pacchetto è nulla. I grafici in figura A.5 sono riportati anche nelle figure A.14 e A.15. dove vengono messi a confronto con quelli ottenuti con il cluster *Avogadro* di Padova.

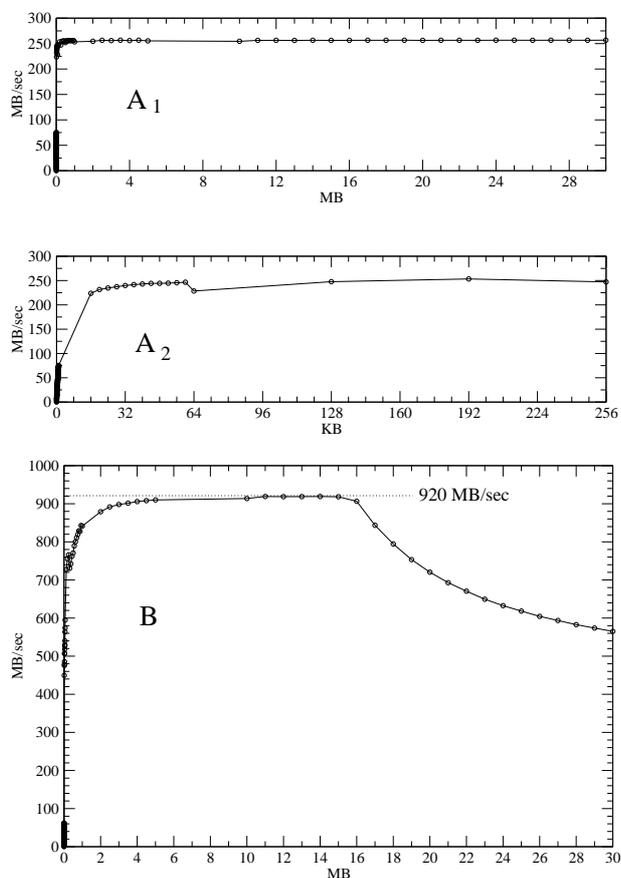


Figura A.5: Benchmark in Distributed Memory (A₁ ed A₂) e Shared Memory (B) sul cluster *Bionap*. I grafici mostrano l'andamento della banda (in MB/sec) in funzione della dimensione del *data size* del pacchetto trasferito.

A.2.5 Il software per il calcolo chimico

Sul cluster *Bionap* sono stati installati i codici di chimica computazionale *Gaussian03*, *Gromacs* ed *Amber*. La compilazione dei vari codici è stata ottimizzata per l'architettura *Alpha ev68* e le varie versioni sono state testate su vari sistemi molecolari: qui riporto i benchmark relativi ad *Amber* e *Gromacs* per i sistemi molecolari nelle figure A.7, e A.6. I risultati sono riportati nelle figure A.8 e A.9.

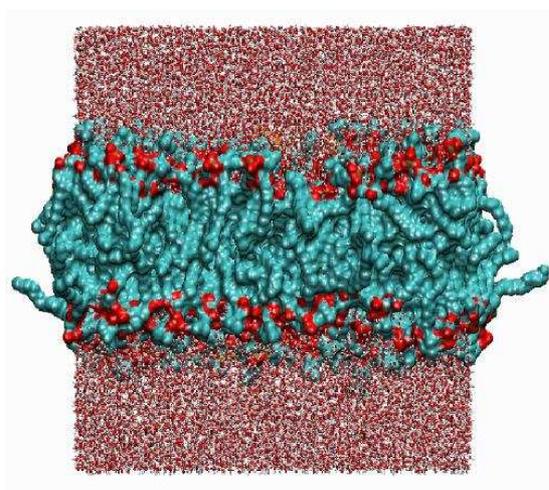


Figura A.6: Sistema molecolare per il benchmark *GROMACS*. Il sistema è costituito da una **Membrana Lipidica idratata**. In totale sono presenti **1024** molecole di DPPC, **23552** molecole di acqua, **121856** atomi.

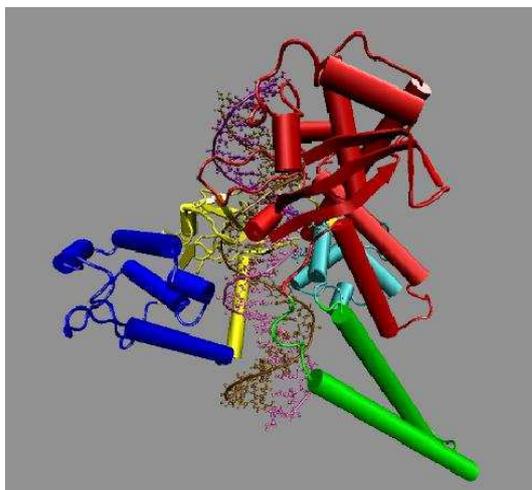


Figura A.7: Sistema molecolare per i Benchmark *AMBER*. Il sistema è costituito dal **Complesso Topoisomerasi I-DNA**. In esso sono presenti **563** residui, **22** basi di DNA, **28313** molecole d'acqua, **95778** atomi totali.

Appendice

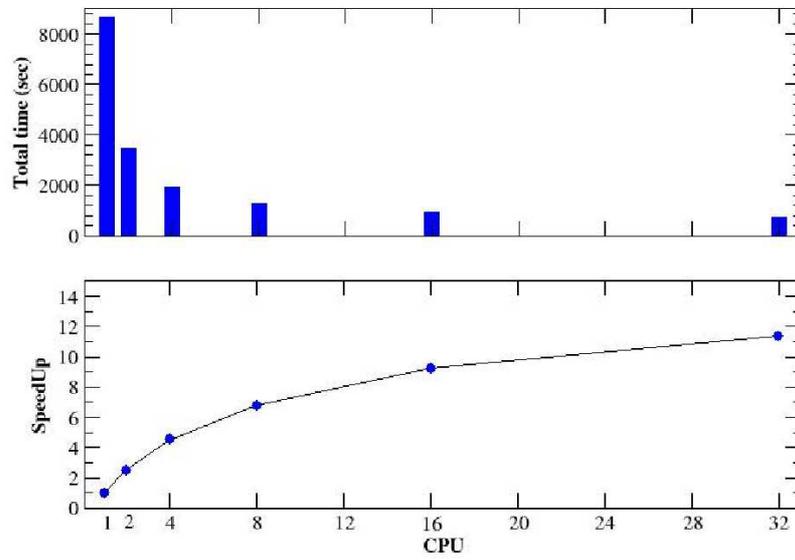


Figura A.8: Total time e speedup nel benchmark AMBER.

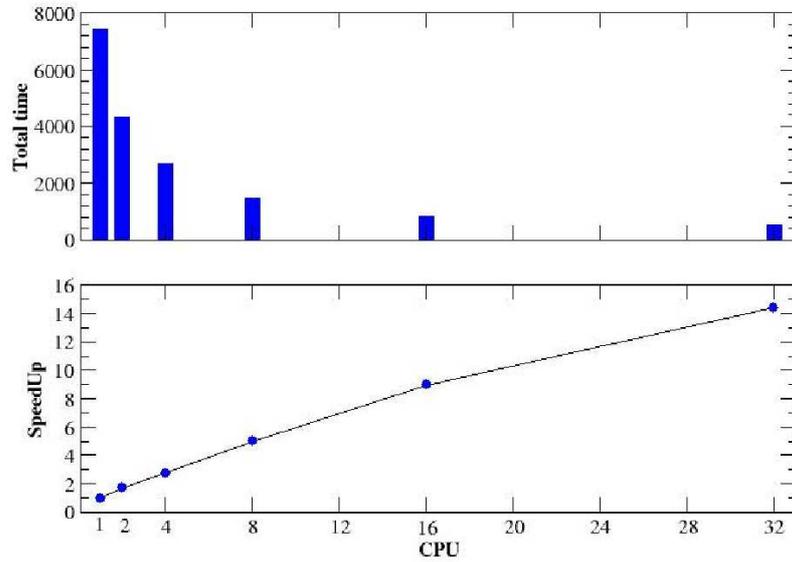


Figura A.9: Total time e speedup nel benchmark GROMACS.

A.3 Il nodo di Padova: Il cluster “*Avogadro*”

In questa sezione discuterò gli aspetti più importanti relativi all’installazione ed alla configurazione del cluster *Avogadro* presso il Dipartimento di Chimica dell’Università di Padova e costituito da 5 enclosure HP BladeSystem C7000 in cui sono disposti 71 nodi HP BL460c G1. Il S.O. adottato è *RedHat Linux Enterprise 4* installato sia sul server che sui singoli nodi. Per l’installazione/gestione del cluster è stata adottata la soluzione *HP Cluster Management Utility v3.0 (CMU)*. Di seguito verranno illustrati i seguenti argomenti:

- collocazione dei componenti;
- configurazione dello spazio disco sul DL380 (Management Node);
- configurazione dello spazio disco locale al primo nodo del cluster;
- backup del primo nodo e clonaggio del cluster con la CMU;
- configurazione dello spazio disco condiviso;
- configurazione delle connessioni di rete compresa quella InfiniBand (IB);
- installazione e configurazione dell’ambiente di sviluppo: compilatori PGI, Intel e parallel environment;
- installazione e configurazione di software di interesse chimico: in particolare *Gaussian03* e *AMBER9*;
- configurazione dell’ambiente utente;
- benchmark per l’analisi delle prestazioni dell’ambiente per il calcolo distribuito nelle implementazioni OpenMPI e SCALI.

A.3.1 Collocazione dei componenti

Il cluster in oggetto è assemblato in due rack *HP 10000 G2 da 42U* dove sono collocati i 5 enclosure C7000: due nel primo rack e 3 nel secondo come mostrato in figura A.10. All’interno degli enclosure sono disposti 16 blade (l’ultimo enclosure al momento è riempito per metà). Nei due rack trovano inoltre alloggio le seguenti unità:

- **1 DL380 G5 (Management Node)**
- **2 Switch di secondo livello IB Voltaire (24 porte)**
- **1 Switch Gbit (24 Porte)**

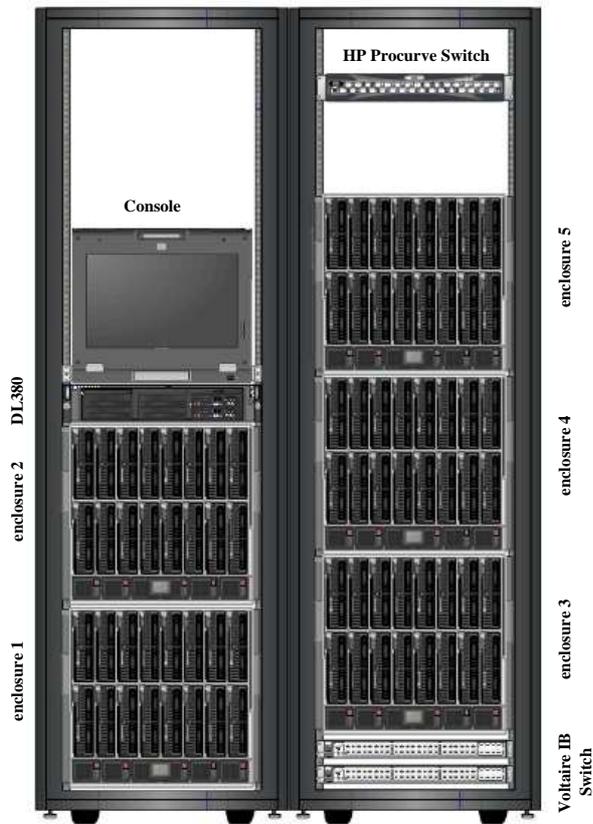


Figura A.10: Vista frontale del cluster *Avogadro*

La configurazione HP BladeSystem prevede la presenza di switch di interconnessione di primo livello in ogni enclosure: in particolare nella configurazione attuale sono presenti per ogni enclosure due switch IB da 4 porte ciascuno ed uno switch *Gigabit HP GbE2c* dove è attestata la rete di sistema per il traffico NFS.

A.3.2 Configurazione dello spazio disco sul nodo di management DL380

Sul nodo di management DL380 sono presenti 8 dischi SAS da 73 GB di cui 7 configurati in RAID 5 ed uno lasciato come disco “*spare*” pronto a subentrare nel RAID 5 in caso di rottura di uno dei dischi. Lo spazio disco risultante è stato partizionato come mostrato dall’output del comando *df*:

```
[root@avogadro-srv ~]# df -h
```

Appendice

Filesystem	Size	Used	Avail	Use%	Mounted on
/dev/cciss/c0d0p2	50G	9.5G	38G	21%	/
/dev/cciss/c0d0p1	251M	26M	212M	11%	/boot
none	2.0G	0	2.0G	0%	/dev/shm
/dev/cciss/c0d0p4	345G	6.4G	321G	2%	/local

Nel mount point */local* sono state create le directory */local/home* e */local/share* a cui sono “linkate” rispettivamente */home* e */share* come mostrato di seguito:

```
[root@avogadro-srv ~]# ls -l /home
lrwxrwxrwx  1 root root 11 Feb 21 17:45 /home -> /local/home
[root@avogadro-srv ~]# ls -l /share
lrwxrwxrwx  1 root root 12 Feb 23 20:47 /share -> /local/share
```

Le directory */home* e */share* sono esportate via NFS sui nodi del cluster con le opzioni mostrate di seguito:

```
[root@avogadro-srv ~]# cat /etc/exports
# CMU
/opt/cmu/ntbt/rp          *(sync,ro,no_all_squash,no_root_squash)
# CMU
/home    *(rw,sync,no_root_squash)
/share  *(rw,sync,no_root_squash)
```

A.3.3 Configurazione dello spazio disco

A.3.3.1 Spazio disco locale sui nodi del cluster

Le caratteristiche principali dei nodi BL460c G1 sono:

- **2 CPU Xeon@2.66GHz dual core**
- **4 Gb RAM**
- **1 scheda Gigabit Ethernet 10/10/1000**
- **1 mezzanine card IB DDR 4x**
- **2 dischi SAS da 73 GB**

Per quanto riguarda lo spazio disco locale abbiamo adottato la seguente configurazione sul primo nodo del cluster che si è poi propagata a tutti gli altri grazie alla procedura di cloning prevista dalla CMU: ogni nodo è fornito di 2 dischi **SAS 73 GB** che si è scelto di configurare in RAID 0 per uno spazio complessivo disponibile di 136 GB. Riporto di seguito l'output del comando *df* per fornire i dettagli di configurazione dello spazio disco locale insieme ai relativi mount point.

```
[root@avogadro-01 ~]# df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/cciss/c0d0p2 9.9G  3.2G  6.3G  34% /
/dev/cciss/c0d0p1 122M   12M  105M  10% /boot
none            2.0G    0    2.0G   0% /dev/shm
/dev/cciss/c0d0p3 120G  129M  114G   1% /local
192.168.16.1:/home 345G  6.4G  321G   2% /local/home
192.168.16.1:/share 345G  6.4G  321G   2% /local/share
```

Su ogni nodo è presente la directory `/scratch` che è un link a `/local/scratch`

```
[gianni@avogadro-01 ~]$ ls -l /scratch
lrwxrwxrwx 1 root root 14 Mar 30 18:15 /scratch -> /local/scratch
```

Questa directory verrà usata come area temporanea per i vari codici di calcolo tra cui *Gaussian03*.

A.3.3.2 Spazio disco condiviso

Come detto in precedenza sul nodo di management è presente uno spazio disco complessivo di circa 400 GB. Le directory `/home` e `/share` sono esportate via NFS e montate come `/local/home` e `/local/share` sui singoli nodi come si deduce dalle ultime due righe del comando `df` appena mostrato. Per tali mount point sono stati creati i seguenti link sui nodi:

```
[root@avogadro-01 ~]# ls -l /home
lrwxrwxrwx 1 root root 11 Mar 30 18:15 home -> /local/home
[root@avogadro-01 ~]# ls -l /share
lrwxrwxrwx 1 root root 12 Mar 30 18:15 share -> /local/share
```

A.3.4 Backup e cloning con la CMU

Sul nodo di management è stata installata la CMU nel percorso standard per cui tutto il pacchetto è presente nella directory `/opt/cmu`. Dopo l'installazione sono state eseguite varie prove per testare il Backup ed il Cloning per giungere alla creazione di un *logical group* denominato **REDHAT_ES4_UP4** cui appartengono tutti i nodi del cluster dopo la clonazione: la situazione di insieme attraverso il monitoring della CMU è visibile nella figura A.11.

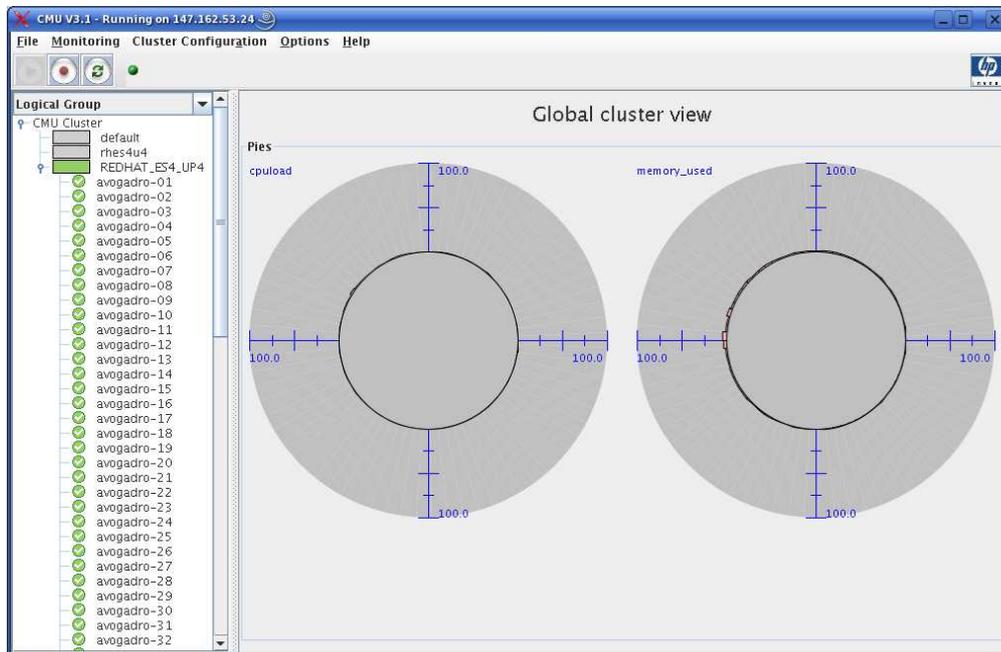


Figura A.11: Vista del cluster *Avogadro* dalla CMU.

A.3.5 Configurazione delle connessioni di rete

Il cluster *Avogadro* dispone di più connessioni di rete la cui organizzazione è riportata in tabella A.2 distinguendo tra nodi di calcolo e management node. Sono presenti le tre reti previste dalla configurazione standard, il *Cluster Interconnect*, il *System Interconnect* la *Management Network*, ed una rete esterna: ad eccezione della rete su cui è implementato il Cluster Interconnect l'infrastruttura hardware è su *Gbit*. Il Cluster Interconnect sul quale viene implementata la comunicazione internodo per il calcolo parallelo è realizzato con tecnologia Infiniband. In figura A.12 è riportato lo schema di interconnessione delle varie reti sul cluster.

Appendice

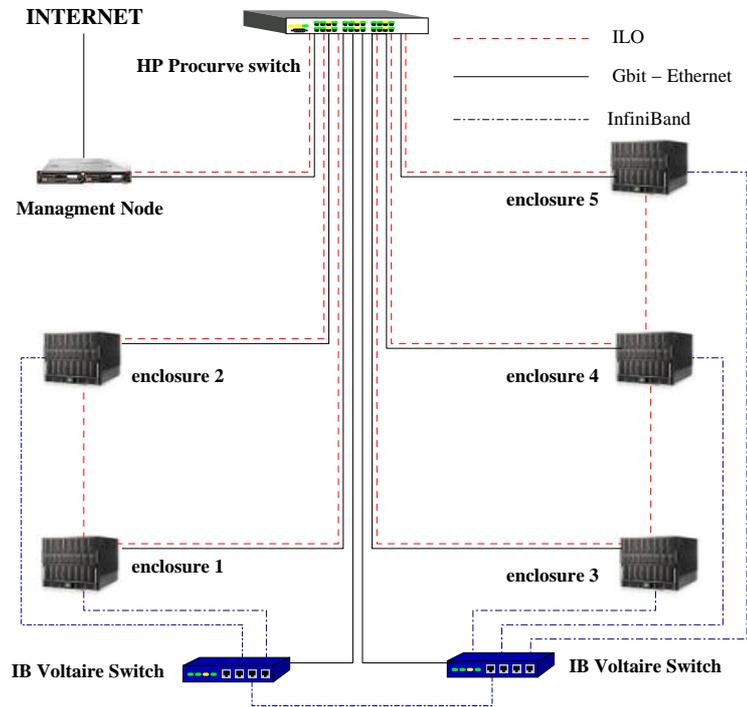


Figura A.12: Schema fisico di interconnessione

Tabella A.2: Elenco delle connessioni di rete

sui nodi		
Rete	Device a livello S.O.	Tipo
Cluster Interconnect	ib0	InfiniBand
System Network (NFS)	eth1	Gigabit
Management Network (CMU)		ILO

sul managment node		
Rete	Device a livello S.O.	Tipo
Management Network (CMU)	eth0	Gigabit
System Network (NFS)	eth1	Gigabit
External Network	eth2	Gigabit

A.3.6 Installazione e configurazione dell'ambiente di sviluppo

Sul cluster *Avogadro* sono stati installati i compilatori Intel, PGI le librerie MKL per lo sviluppo. La tabella A.3 riporta le rispettive versioni e path di installazione.

Tabella A.3: Compilatori sul cluster *Avogadro*

COMPILATORE	VERSIONE	PATH
Intel	9.1	/share/intel/fce(cce)(idbe)/9.1.040
PGI	7.0-2	/share/pgi/linux86-64/7.0-2
MKL	9.0	/share/intel/mkl/9.0/lib/em64t

Per quanto riguarda la configurazione dell'ambiente per il calcolo parallelo con MPI nel file di configurazione della *tsh* (shell di default per gli utenti) è stato inserita la seguente istruzione

```
source /opt/scali/etc/scampivars.csh
```

attraverso la quale vengono impostate le variabili di ambiente per lo SCALI MPI tra cui:

```
PATH=/opt/scali/bin:/opt/scali/sbin
```

```
MPI_HOME=/opt/scali
```

```
LD_LIBRARY_PATH=/opt/scali/lib64:/opt/scali/lib
```

a queste variabili dovrà riferirsi l'utente nel caso di compilazioni di codici che necessitano del supporto MPI.

A.3.7 Installazione del software per la chimica computazionale

Al momento i pacchetti per il calcolo chimico installati su *Avogadro* sono *Gaussian03* ed *Amber*: versioni e path sono mostrati in tabella A.4.

Tabella A.4: Software chimico sul cluster *Avogadro*

CODICE	VERSIONE	PATH
Gaussian03	D.02	/share/g03
Amber	9.0	/share/amber9

A.3.8 Configurazione dell'ambiente utente

Sul management node del cluster *Avogadro* è stato configurato un server NIS cui afferisce il dominio “*Avogadro*” per cui l'autenticazione è centralizzata. Di default la shell per gli utenti è la *tsh* per cui l'ambiente è configurato a partire dal file */etc/csh.cshrc*. Di seguito riporto le linee relative alla configurazione dell'ambiente di sviluppo

```
# SCALI ENV.
source /opt/scali/etc/scampivars.csh
# INTEL COMPILER ENV.
setenv PATH /share/intel/idbe/9.1.045/bin:
          /share/intel/fce/9.1.040/bin:
          /share/intel/cce/9.1.045/bin:$PATH
# INTEL MKL ENV.
setenv MKL_LIB /share/intel/mkl/9.0/lib/em64t
# SHARE BIN (Gaussian03, \ldots )
setenv PATH /share/bin:$PATH
```

A.3.9 Benchmarks MPI

La configurazione della rete InfiniBand (IB) è stata eseguita dopo l'installazione dei driver necessari al supporto hardware. A tale scopo è stato installato il pacchetto OFED-1.1¹ L'installazione è stata eseguita in modo standard compresa la scelta del “path” che per ogni nodo vede tutti i pacchetti OFED depositati nella directory */usr/local/ofed*. Una volta installato il supporto IB è stato possibile sfruttare la configurazione *IPoIB* (IP over InfiniBand) attraverso la quale ad ogni interfaccia IB è stato possibile associare un indirizzo IP e configurare la rete IB stessa. Sulla rete di interconnessione è stato installato l'ambiente di calcolo distribuito MPI in due versioni: open-source e commerciale. La prima è costituita dalla versione OpenMPI fornita dal pacchetto OFED mentre la seconda è quella sviluppata da SCALI. Per entrambi è stato eseguito il test preliminare PingPong fornito dalla suite *PALLAS*² per i benchmark MPI. Dall'analisi di questo test preliminare è emerso chiaramente che l'uso ottimale dell'hardware IB presente si realizza con l'ambiente parallelo SCALI per cui la scelta finale è ricaduta sull'adozione della soluzione commerciale. In figura A.13 sono riportati i valori di banda in MB/s

¹<http://www.openfabrics.org>

²Vedi a proposito: <http://www.open-mpi.org/community/lists/devel/2005/08/0176.php>

Il gruppo Pallas è stato assorbito da Intel. L'ultima versione stabile dei benchmark non è la PMB-2.2 ma la PMB-2.3 di Intel scaricabile da: <http://www.intel.com/cd/software/products/asm-na/eng/cluster/mpi/219848.htm>.

su architettura **SMP Shared Memory** usando OpenMPI mentre in figura A.14 gli stessi dati si riferiscono all'utilizzo del software SCALI: in entrambi è riportato l'andamento dello stesso test sul nodo di Napoli (vedi le caption delle figure). Il codice usato per questo benchmark, realizza la trasmissione/ricezione (send/receive) di un pacchetto con data size variabile fornendo per tale trasferimento, il picco dell'ampiezza di banda.

Dal confronto delle figure A.13 e A.14 emerge chiaramente che l'ambiente parallelo offerto dalla soluzione commerciale SCALI è molto più performante di quella OpenMPI e soprattutto più stabile: per questo motivo si è deciso di configurare l'ambiente parallelo con quest'ultimo. Da notare in figura A.15 le notevoli performance di interconnessione dei nodi offerte dalla rete InfiniBand in collaborazione con l'ambiente SCALI. Dai dati riportati nelle figure A.14 e A.15 è inoltre possibile verificare l'incremento di prestazioni dovuto all'evoluzione tecnologica nell'arco di tre anni. In particolare la banda passante mostra un incremento di oltre il 30% su SMP (figura A.14) mentre da quelli in figura A.15 (InfiniBand vs. Elan3, Distributed Memory) l'incremento va oltre il 300% !

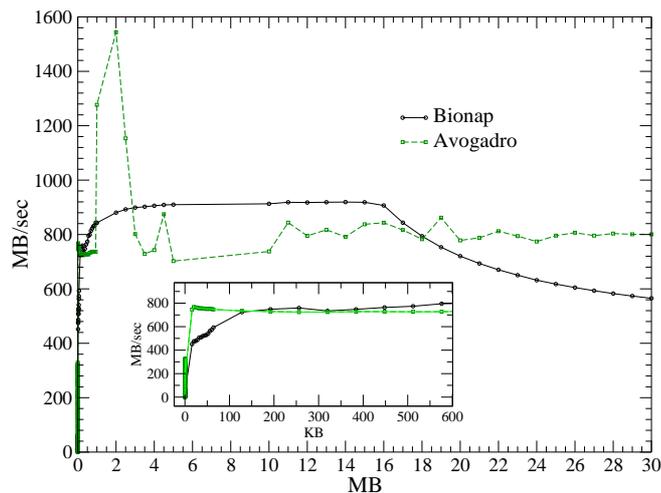


Figura A.13: Benchmark PALLAS PingPong con OpenMPI (forniti con il pacchetto ofed) in Shared Memory. Per confronto è riportato lo stesso test eseguito su un nodo AlphaServer SC ev68@1250MHz (*Bionap*) presso il Dipartimento di Chimica dell'Università di Napoli (<http://lsdm.campusgrid.unina.it>)

Appendice

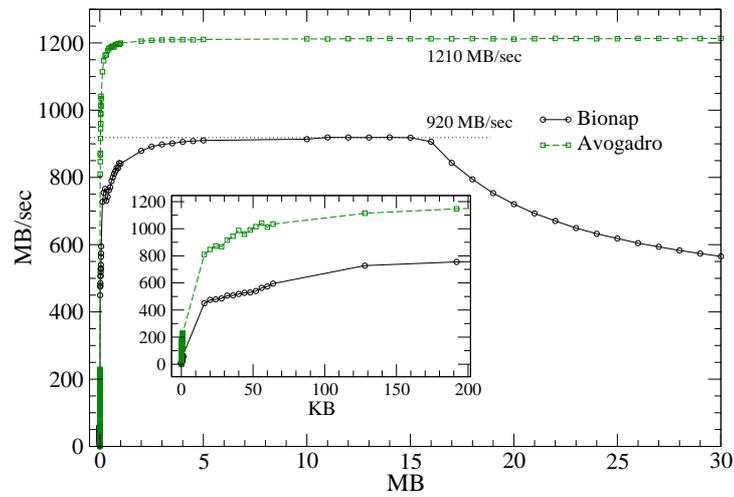


Figura A.14: Benchmark PALLAS PingPong con l'ambiente SCALI (<http://www.scali.com>) in Shared Memory. Anche qui è riportato il test eseguito su AlphaServer

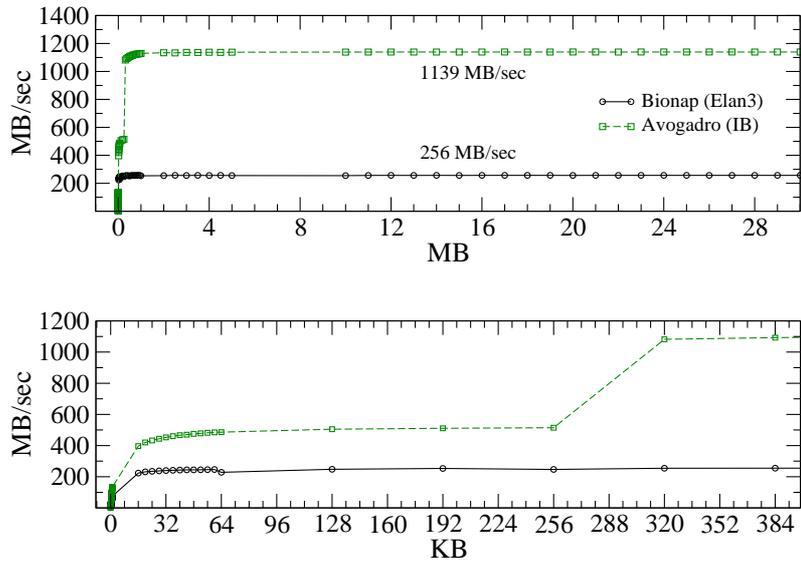


Figura A.15: Benchmark PALLAS PingPong con l'ambiente SCALI (<http://www.scali.com>) in Distributed memory. Anche qui è riportato il test eseguito su AlphaServer

Appendice

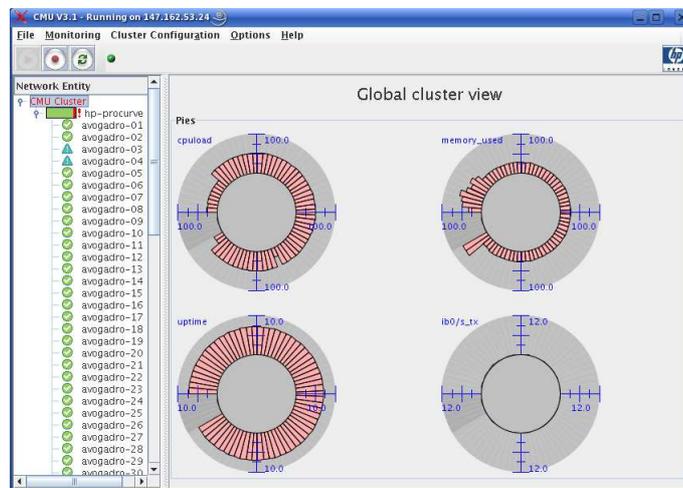


Figura A.16: Monitoring delle risorse con la CMU durante un job MPI. Lo screenshot è stato eseguito quando sul cluster non erano stati ancora clonati gli ultimi 6 nodi come mostrato dal settore circolare in grigio scuro nei vari grafici. Il job MPI in questione coinvolge 65 nodi.

Bibliografia

- [1] (a) E. K. U. Gross, W. Kohn, *Phys. Rev. Lett.* **1985**, 55, 2850. (b) R. E. Stratmann, G. E. Scuseria, M. J. Frish, *J. Chem. Phys.* **1998**, 109, 8218. (c) M. E. Casida in *Recent Advances In Density Functional Methods (part I)*, **1994**, pp. 155-192 ed. D. P. Chong, *World Scientific, Singapore*.
- [2] (a) K. Wolinski, J. F. Hilton, P. Pulay, *J. Am. Chem. Soc.* **1990**, 112, 8251. (b) J. R. Cheeseman, G. W. Trucks, T. A. Keith, M. J. Frish, *J. Chem. Phys.* **1998**, 104, 5497.
- [3] (a) O. Crescenzi, G. Correale, A. Bolognese, V. Piscopo, M. Parrilli, V. Barone, *Org. Biomol. Chem.* **2004**, 1, 1577.
- [4] (a) S. Iwata, K. Morokuma, *J. Am. Chem. Soc.* **1975**, 97, 966. (b) M. Kraus, S. P. Webb, *J. Chem. Phys.* 1997, 107, 5771.
- [5] S. Miertus, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, 55, 117.
- [6] (a) M. Cossi, G. Scalmani, N. Rega, V. Barone, *J. Chem. Phys.* **2002**, 117, 43. (b) G. Scalmani, V. Barone, K. N. Kudin, C. S. Pomelli, G. E. Scuseria, M. J. Frisch, *Theor. Chem. Acc.* **2004**, 111, 90. (c) M. Cossi, V. Barone, *J. Phys. Chem. A* **2000**, 104, 10614.
- [7] V. Barone, M. Cossi, J. Tomasi, *J. Chem. Phys.* **1997**, 107, 3210.
- [8] (a) S. Canuto, K. Coutinho, M. Zerner, *J. Chem. Phys.* **2000**, 112, 7293. (b) K. Coutinho, S. Canuto, *J. Chem. Phys.* **2000**, 113, 9132. (c) K. Coutinho, S. Canuto, *Theochem* **2003**, 632, 235.
- [9] (a) U. F. Rohrig, I. Frank, J. Hutter, A. Laio, J. VandeVondele, U. Rothlisberger, *ChemPhys. Chem.* **2003**, 4, 1177. (b) M. Sulpizi, P. Carloni, J. Hutter, U. Rothlisberger, *Phys. Chem. Chem. Phys.* **2003**, 5, 4798. (c) M. Sulpizi, U. F. Rohrig, J. Hutter, U. Rothlisberger, *Int. J. Quantum Chem.* **2005**, 101, 671.
- [10] R. Car, M. Parrinello, *Phys. Rev. Lett.* **1985**, 55, 2471.

BIBLIOGRAFIA

- [11] L. Bernasconi, M. Sprik, J. Hutter, *J. Chem. Phys.* **2003**, 119, 12417.
- [12] (a) C. Adamo, A. di Matteo, V. Barone, *Adv. Quantum Chem.* **1999**, 36, 45.
(b) M. Pavone, V. Barone, I. Ciofini, C. Adamo, *J. Chem. Phys.* **2004**, 120, 9167.
- [13] (a) D.J.Searles, H.Huber, *Encyclopedia of Nuclear Magnetic Resonance, vol.9 Advances in NMR*, pp. 215-226, John Wiley & Sons, Chichester, **2002**. (b) D.J.Searles, H.Huber in *Calculation of NMR and EPR Parameters, Theory and Applications*, ed. M.Kaup, M.BÄ¼hl, V.G.Malkin, *Wiley-VCH, Weinheim*, **2004**, pp.175-189.
- [14] D. Kony, W. Damm, S. Stoll, W. F. Van Gunsteren, *J. Comput Chem.* **2002**, 23, 1416.
- [15] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Jr. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, 117, 5179.
- [16] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comp. Chem.* 1983; 4, 187.
- [17] G. Brancato, A. Di Nola, V. Barone, A. Amadei, *J. Chem. Phys.* **2005**, 122, 154109.
- [18] (a) D. Sebastiani, M. Parrinello, *J. Phys. Chem. A* **2001**, 105, 1951. (b) M. Buhl, *J. Chem. Phys. A* **2002**, 106, 10505.
- [19] (a) M. Aschi, R. Spezia, A. Di Nola, A. Amadei, *Chem. Phys. Lett.* **2001**, 344, 374. (b) R. Spezia, M. Aschi, A. Di Nola, A. Amadei, *Chem. Phys. Lett.* **2002**, 365, 450. (c) A. Amadei, F. Marinelli, M. D'Abramo, M. D'Alessandro, M. Anselmi, A. di Nola, M. Aschi, *J. Chem. Phys.* **2005**, 122, 124506.
- [20] K.Aidas, J.Kongsted, A.Osted, K.V.Mikkelsen, *J.Phys.Chem. A* ASAP article, DOI: 10.1021/jp0527094.
- [21] O. Crescenzi, M. Pavone, F. De Angelis, V. Barone, *J. Phys. Chem. B* **2005**, 109, 445.
- [22] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* 1996, 77, 3865.
- [23] C. Adamo, V. Barone, *J. Chem. Phys.* 1999, 110, 6158.
- [24] M. J. Frisch et al., *Gaussian03, revision C.02*, Gaussian, Inc., Pittsburgh, PA, **2004**.

BIBLIOGRAFIA

- [25] P. Cieplak, W. D. Cornell, C. Bayly, P. A. Kollman, *J. Comput. Chem.* **1995**, 16, 1357.
- [26] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gusteren, J. Hermans, in *Intermolecular Forces*, ed. B. Pullman (Reidel, Dordrecht, 1981).
- [27] D. V. der Spoel et al., *Gromacs User Manual version 3, Nijenborgh 4*, 9747 AG Groningen, The Netherlands, **2001**.
- [28] A. Amadei, G. Chillemi, M. A. Ceruso, A. Grottesi, A. Di Nola, *J. Chem. Phys.* **2000**, 112, 9.
- [29] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids (Oxford University Press, Oxford, 1987)*.
- [30] M. M. Francl, W. J. Pietro, W. J. S. Hehre, J. Binkley, M. S. Gordon, D. J. DeFrees, J. A. Pople, *J. Chem. Phys.* **1982**, 77, 3654.
- [31] C. Moller, M. S. Plesset, *Phys. Rev.* **1934**, 46, 618.
- [32] R. L. Hilderbrandt, A. L. Andreassen, S. H. Bauer, *J. Phys. Chem.* 1970, 74, 1586.
- [33] T. H. Jr. Dunning, *J. Chem. Phys.* 1989, 90, 1007.
- [34] (a) N. S. Bayliss, E. G. McRae, *J. Phys. Chem.* **1954**, 58, 1006. (b) A. Balasubramanian, C. N. R. Rao, *Spectrochim. Acta* **1962**, 18, 1337. (c) N. S. Bayliss, G. Wills-Johnson, *Spectrochim. Acta A* **1968**, 24, 551.
- [35] Patai S **1974** Chemistry of Quinoid compounds. *Interscience*, New York
- [36] Schiedt J, Weinkauff R **1999** *J. Chem. Phys.* 110:304
- [37] Zhao X, Imahori H, Zhan C-G, Sakata Y, Iwata S, Kitagawa T **1997** *J. Phys. Chem. A* 101:622
- [38] Zhao X, Kitagawa T **1998** *J. Raman Spec.* 29:773
- [39] Pou-Amrigo R, Serrano-Andres L, Merchà M, Ortì E, Forsberg N **2000** *J. Am. Chem. Soc.* 122:6067
- [40] Honda Y, Hada M, Ehara M, Nakatsuji H **2002** *J. Phys. Chem. A* 106:3838
- [41] Koch W, Holthausen MC **2001** A chemist's guide to density functional theory, 2nd ed. *Wiley-VCH*, Weinheim

BIBLIOGRAFIA

- [42] Marques MAL, Gross EKV **2004** *A. Rev. Phys. Chem.* 55:427
- [43] Tomasi J, Mennucci B, Cammi R **2005** *Chem. Rev.* 105:2999
- [44] Dapprich S, Komaromi I, Byun KS, Morokuma K, Frisch MJ **1999** *Theochem* 461, 1
- [45] Scalmani G, Frisch MJ, Mennucci B, Tomasi J, Cammi R, Barone V **2006** *J. Chem. Phys.* 124:094107
- [46] Improta R, Barone V, Scalmani G, Frisch MJ **2006** *J. Chem. Phys.* 125:054103
- [47] Adamo C, Barone V (**1999**) *Chem. Phys Lett.* 314:152
- [48] Frisch MJ et al. **2003** Gaussian 03, Revision C.02. *Gaussian, Inc.*, Wallingford
- [49] Foresman JB, Frisch AE (**1996**) Exploring chemistry with electronic structure methods, 2nd edn. *Gaussian Inc.*, Pittsburgh
- [50] Cossi M, Scalmani G, Rega N, Barone V **2002** *J. Chem. Phys.* 117:43
- [51] Cossi M, Barone V **2001** *J. Chem. Phys.* 115:4708
- [52] Hagen K, Hedberg K (**1973**) *J. Chem. Phys.* 59:158
- [53] Mattar SM **2004** *J. Phys. Chem. B* 108:9449
- [54] Drew A, Weisman JL, Head-Gordon M **2003** *J. Chem. Phys.* 119:2943
- [55] Della Sala F, Görling A **2001** *J. Chem. Phys.* 115:5718
- [56] Yanai T, Tew DP, Handy N **2004** *Chem. Phys Lett.* 393:51
- [57] Dierksen M, Grimme SS **2005** *J. Chem. Phys.* 122:244101
- [58] Gustavsson T, Banyasz A, Lazzarotto E, Markovitsi D, Scalmani G, Frisch MJ, Improta R, Barone V **2006** *J. Am. Chem. Soc.* 128:607
- [59] Improta R, Barone V, Santoro F **2007** **Angew Chem** 46:405
- [60] Pavone M, Cimino P, De Angelis F, Barone V **2006** *J. Am. Chem. Soc.* 128:4338
- [61] Barone V, Polimeno A **2006** *Phys Chem Chem Phys* 8:4609
- [62] <http://java.sun.com> and references therein.
- [63] Barone V (**2005**) *J Chem Phys* 122:014108.

BIBLIOGRAFIA

- [64] Barone V, Viglione R, (2005) *J Chem Phys.* 123:234304.
- [65] Barone V, Carbonniere P, Pouchan C (2005) *J Chem Phys* 122:224304.
- [66] Barone V (2004) *J. Chem Phys* 120:3059.
- [67] Miller WH, Hernandez R, Handy NC, Jayatilaka D, Willets A (1990) *Chem Phys Lett* 172: 62.
- [68] D'Amore M., Talartico G, Barone V (2006) *J. Am Chem Soc* 128:1099.
- [69] Improta R, Barone V, Kudin KN, Scuseria GE (2001) *J Chem Phys* 114:2541.
- [70] Cantele G, Trani F, Ninno D, Cossi M, Barone V (2006) *J Phys C* 18:2349.
- [71] Festa G, Cossi M, Barone V, Cantele G, Ninno D, Iadonisi G (2005) *J Chem Phys* 122:184714.
- [72] Car R, Parrinello M (1985) *Phys Rev Lett* 55:2471.
- [73] Schlegel HB et al. (2001) *J Chem Phys* 114:9758.
- [74] Brancato G, Rega N, Barone V (2006) *J Chem Phys*, in press.
- [75] Sanna N, Morelli G (2004) *Comp Phys Comm* 162:51.