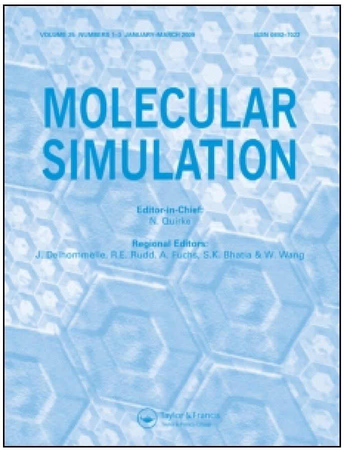

Molecular Simulation

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

A Parallel Quantum Mechanical MD Simulation of Liquids

Fredrik Hedman[a]; Aatto Laaksonen[b]
[a] Center for Parallel Computers (PDC), Royal Institute of Technology, Stockholm, Sweden [b] Department of Physcial Chemistry, Arrhenius Laboratory, Stockholm University, Stockholm, Sweden

To cite this Article Hedman, Fredrik and Laaksonen, Aatto(1998) 'A Parallel Quantum Mechanical MD Simulation of Liquids', Molecular Simulation, 20: 5, 265 — 284

To link to this Article: DOI: 10.1080/08927029808022037

URL: http://dx.doi.org/10.1080/08927029808022037

Molecular Simulation, 1998, Vol. 20, pp. 265–284
Reprints available directly from the publisher
Photocopying permitted by license only

A PARALLEL QUANTUM MECHANICAL MD SIMULATION OF LIQUIDS*

FREDRIK HEDMAN[a] and AATTO LAAKSONEN[b]

[a]*Center for Parallel Computers (PDC), Royal Institute of Technology, S-100 44 Stockholm, Sweden;*
[b]*Department of Physical Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden*

(Received July 1997; Accepted July 1997)

A parallel implementation to perform quantum molecular dynamics simulations of liquids on Born-Oppenheimer surfaces is evaluated. The MD method allows the use of atomic energy gradient forces at an arbitrary level of quantum chemical methodology and electronic state. The computational scheme is very simple to implement, although still expensive to use. The method can easily be developed further for studies which are beyond the capabilities of current classical simulations, e.g., simple chemical reactions or other chemical processes involving excited electronic states or radicals. In a preliminary study, the approach is applied to simulate liquid water (with all internal degrees of freedom included) at the semi-empirical AM1 molecular orbital (MO), and at the *ab initio* SCF-MO Hartree-Fock level. Performance, scaling and load balancing properties of the parallel scheme are discussed. Timings of an implementation on an IBM SP-2 are presented. To test the approach, the simulation results are compared with those obtained from corresponding classical simulations of water, from simulations using the Car-Parrinello method and to experimental radial distribution functions.

Keywords: First principles molecular dynamics simulation; MD simulation; parallel computers; quantum mechanics; liquid water

Physics and Astronomy Classification: 02.70.Ns; 34.20.Gj; 61.20.Gy; 61.20.Ja; 61.25.Em; 82.20.Wt

---

*Partly presented at the 10th Nordic Symposium on Computer Simulations, June 28–July 1, 1996, Tallinn, Estonia.

1. INTRODUCTION

Computer simulations, such as, molecular dynamics (MD) and Monte Carlo (MC) have over the last two decades become very important tools to study many-particle systems and nearly every possible aspect of condensed matter [1, 2]. The necessary computational power for this development has been made possible by a combination of more efficient simulation methods and the rapid development of the microchip technology [3]. Parallel computer architectures are boosting the development further [4] and the increased computing power opens up several paths to improve the accuracy and reliability of molecular simulations. Examples of these are:

- The size of the simulated systems can be increased to enable the studies of large molecular systems of biological relevance or collective effects like phase transitions, but also to eliminate possible artifacts, resulting from periodicity. For a simulation containing N interaction sites the computational cost scales as $O(N^2)$. In practice this cost can often be reduced to $O(N \log N)$ or even $O(N)$ [5].
- The length of the simulations can be extended to cover larger regions of the coordinate or conformational phase space and to follow slow dynamic processes. The computing time in this case will increase linearly with the number of time-steps in the simulation.
- More detailed models can be used by introducing more realistic interaction potentials closer to fundamental principles of molecular physics. Examples of currently used sophisticated models are so called polarizable interaction potentials [6] and distributed multipoles [7]. With these models we are still in the domain of classical effective pair potentials, so even though the cost of calculating each interaction is greater, the simulation cost is still of order $O(N \log N)$ to $O(N)$.
- Going a step further and introducing electrons through quantum mechanical interactions between some or all particles in the simulation should quite dramatically increase the accuracy of the simulation, but the computational cost will also rise sharply. The increased computational demand will to a large extent depend on the level of sophistication of the quantum model (Semi-emprical, Hartree-Fock, DFT, MP2, etc).

In this preliminary work we investigate the potential of a general quantum mechanical MD simulation scheme for liquid systems with periodic boundary conditions based on first principles, parallellised for multiprocessor computers. It can accommodate an arbitrary level of quantum chemical sophistication to treat the chemical character in molecular systems, including processes through excited electronic states, electron, proton and hydrogen transfer and systems involving molecular radicals – all inaccessible to classical MD simulation methods. The parallellization of the computational scheme is made in a short interface routine, connecting the two software packages. The objectives of the paper are:

- to describe a simple scheme to perform first-principle MD simulations in which the forces are calculated "in flight" as gradients of the variational energy expressions [8, 9] rather than Hellman-Feynman forces. (Section 2)
- to apply the method to simulate liquid water allowing all the internal degrees of freedom and compare the results with corresponding classical results, Car-Parrinello [10] calculations and experimental radial distribution functions for water [11]. (Section 3)
- to discuss and design an efficient parallel computational scheme to perform the calculations and also present one implementation strategy that requires very little programming effort by using already existing conventional program packages as building blocks. (Section 4)

2. QUANTUM MECHANICAL FORCE FIELDS

Force fields are clearly the main hurdle in every attempt to extend the MD computer simulation methods to studies of chemical processes, simply because the present models cannot treat phenomena such as dissociation, electron transfer or excited states. Fundamental quantum mechanics must be introduced into the simulation methods to replace the simple empirical interaction models. Unfortunately, to solve the time-dependent Schrödinger equation for condensed systems of some reasonable size is still far beyond current computational capacities.

One possible strategy is to treat a limited, chemically interesting area of the studied system quantum mechanically, while treating the neighborhood classically, and combining the two domains by mixing the interactions at the borders. This approach was pioneered by Warshel [12], Kollman [13], Field [14] and Gao and their coworkers [15] and also developed further in our group [16, 17, 18, 19]. The critical part in using these methods is to describe the cross interactions between the quantum mechanical part and the molecular mechanical part, since no simple coupling exists between classical and quantum mechanics.

Car and Parrinello [20] developed the first *ab initio* type of MD simulation method describing the electronic structure using density-functional theory (DFT) within the local-density approximation [21, 22]. A limitation of DFT methods is that they deal only with ground states of molecular systems. In the Car-Parrinello (CP) method forces are calculated via the Hellman-Feynman theorem [23, 24]. Development of super-soft pseudo-potentials by Vanderbilt [25], and their incorporation in the CP method [26], was a major improvement to make the method more powerful. The CP method is still mainly used in its original formulation, based on pseudo-potentials and plane wave basis sets. The CP method and its applications have recently been reviewed by Galli and Pasquarallo in [27]. The reader is referred to this article for details of the method and its variations and applications. An excellent introduction to CP Molecular Dynamics, from "a point of view of a molecular physicist" is given by Remler and Madden [28].

A different approach to a quantum mechanical MD method, in which the forces are calculated as gradients of the variational energy expressions [8, 9], can be constructed by completely *replacing* the classical intra- and intermolecular force calculations in a classical MD program by the corresponding quantum mechanical force calculations. To test the viability of this approach, while trying to minimize the programming work, we have used a modified version of the computer simulation program "McMoldyn" [29], as the classical starting point and interfaced it to a standard quantum chemistry package, "Gaussian94" [30]. For our purposes, any other quantum chemistry package that contain calculations of atomic forces could just as well be used.

Our point of departure implements the classical MD method using the neighborlist technique (NL) and periodic boundary conditions using the minimum image criteria. We reuse this algorithmic structure but replace the classical pair-wise force calculations with a quantum mechanical (QM) calculation. The force calculation then becomes a many-body interaction calculation. To continue with periodic boundary conditions we must treat each central molecule in every NL in the same way. This implies that for each molecule in the simulation we get one QM calculation with its current NL included and minimum image criteria applied. Using this approach all participating molecules are treated in the same way resulting in N cluster calculations per time-step in a simulation containing N molecules. The total energy is the sum of the potential energy of the nuclear positions and the electrons plus the kinetic energy of the electrons and the nuclei. The kinetic energy of the nuclei is solved using classical molecular dynamics. QM calculations are carried out to a same accuracy at each time step to ensure conservation of the energy. Following this approach, every time step contains the following tasks:

1. The MD program generates input files for the quantum chemistry program. The coordinates of the molecules, currently kept in the neighbor lists, are used. The molecules in each neighbor list become "clusters", which are input into a QM calculation. With N molecules in the simulation cell, N quantum mechanical cluster calculations are carried out in each time step.
2. From each cluster calculation we take the forces acting on the central atom in the cluster. After having looped through all N clusters and calculated the forces on each molecule, we perform a numerical time integration and update the position and velocity of each atom.

In addition to the sophistication level of the QM method employed, the cost of a time step will depend on the size of the system (N) and the number of functions (L) in the basis set used. Particularly, since the cluster of molecules on which we calculate cannot be expected to contain any symmetries to reduce the number of operations, the computational complexity becomes $O(N * L^4)$. When we increase either of N or L, this is obviously going to become a very time consuming method, making it natural to turn to parallel computers for more compute power.

3. METHOD TEST AND VALIDATION

The number of published computer simulations on water or aqueous solutions can be counted in hundreds at present. Still, the theoretical modeling of water remains a great challenge, depending on the many anomalous properties it exhibits [31]. Many properties of liquid water can be both measured experimentally and reproduced reasonably well in classical computer simulations using empirical effective pair potentials. A few examples of such properties are the (neutron diffraction) liquid structure [11, 32] the diffusion coefficients [33] and the (NMR) reorientational correlation times [34]. A system of water molecules constitutes a "well-known" system to test and validate the method described in Section 2.

Setting up the Water Simulation

Comparing simulations of a small and large classical systems of water, one finds minor differences. We assume that the same should hold true for small and large QM simulations. A system of liquid water is prepared using classical MD. A cubic box is filled with 32 water molecules to a density of 1.0 g/cm^3, giving a box length of 9.86 Å. Water molecules in the classical simulation are described using the flexible SPC potential model [35, 36]. Periodic boundary conditions and minimum image convention are applied. The system is equilibrated classically at 300 K. Temperature is maintained constant using the Nóse-Hoover method [37, 38]. The time step is chosen to 0.2 fs in order to cover various types of fast motions of the water molecules while treating all the internal degrees of freedom, including bond-stretching and angle bending.

The neighbor-list technique is used to calculate the intermolecular interactions. For each of the 32 water molecules, when half the box length is used as a cut-off radius, there is a spherical volume around its centre-of-mass (COM), containing 18–22 neighbors. The varying number molecules kept in the neighbor-lists is simply due to local density fluctuations of the COM points, on which the cut-off is applied. After an initial classical simulation, the Newtonian forces are exchanged to corresponding quantum mechanical forces, supplied by the interfaced quantum chemistry software.

Other simulations of liquid water with quantum mechanical ingredients have appeared in literature recently, most of which have been carried out using hybrid type QM/MM methods that dissolve one single QM water in a bath of classical water molecules [39, 40, 41]. The QM molecule is kept rigid in these works by constraining its geometry to correspond to that of the classical molecules. This allows the use of a longer time step, thus saving the expensive computing time.

Sprik *et al.* [42] test three frequently used gradient-corrected density functionals (B, BP and BLYP) in an *ab initio* MD simulation of liquid water using a system of 32 molecules. Ortega *et al.* [43] have developed a "first-principles simplified" MD method to simulate 216 water molecules at different physical conditions.

MD Simulation at the AM1 Level

After a classical MD simulation was carried out to equilibrate the system of liquid water, the AM1 force field was turned on and the system equilibrated during 2000 steps. It took about 800 steps for the water molecules to adjust into the new interactions, after which, temperature and energies were fluctuating around their mean values. The simulation was continued another 3000 steps. The total length of the simulation using the AM1 force field corresponds to 1.0 picoseconds. The use of AM1 method also serves as a soft intermediate between the classical MM force field and the *ab initio* Hartree-Fock force field.

MD Simulation at the Hartree-Fock Level[1]

The final configuration from the AM1 simulation was taken as an initial configuration for a run with a Hartree-Fock force field. Again, the water system was first equilibrated for 2000 steps in order to get it adapted into a new environment of interactions. In Hartree-Fock calculations, the Fock matrix is diagonalized iteratively until a specified threshold, normally based either on the energy or on the electron density, is reached. The energy of the whole system of water molecules in the simulation cell is assumed to be conserved by solving the HF equations to the same accuracy during each cluster calculation. This resulted in some fluctuations in the numbers of iterations to reach convergence. As in the case of the AM1 model, all runs had to be carried out without use of symmetry. In an attempt to minimize the number of iterations, all the clusters from the previous time step were kept in check point files, from which the final wavefunctions were taken as trial wavefunctions for the new cluster calculations. The basis set we used was the limited STO-3G. The number of time steps of the simulation was the same as in using the AM1 model.

Liquid Structure of Water

In Figure 1, are the $g(r_{OO})$ radial distribution functions shown from both quantum mechanical simulations together with the RDF curve from a classical MD simulation using the flexible SPC model [35, 36] and the experimental distribution function [11]. For a consistent comparison, the classical simulation set up was identical to the quantum mechanical ones.

Compared to the experimental curve, all the simulations (both the classical and the two quantum mechanical) give the first maximum position shifted to a closer distance. Compared to the $g(r_{OO})$ function from a "full-scale" classical simulation with a larger number of molecules using same potential model [36], there are no essential differences between the results. This would indicate that the liquid near structure of water can be described reasonably well using as few as 32 water molecules. The position of $g(r_{OO})$

---

[1] To our knowledge, this is the first MD simulation of liquid water carried out at the Hartree-Fock level.

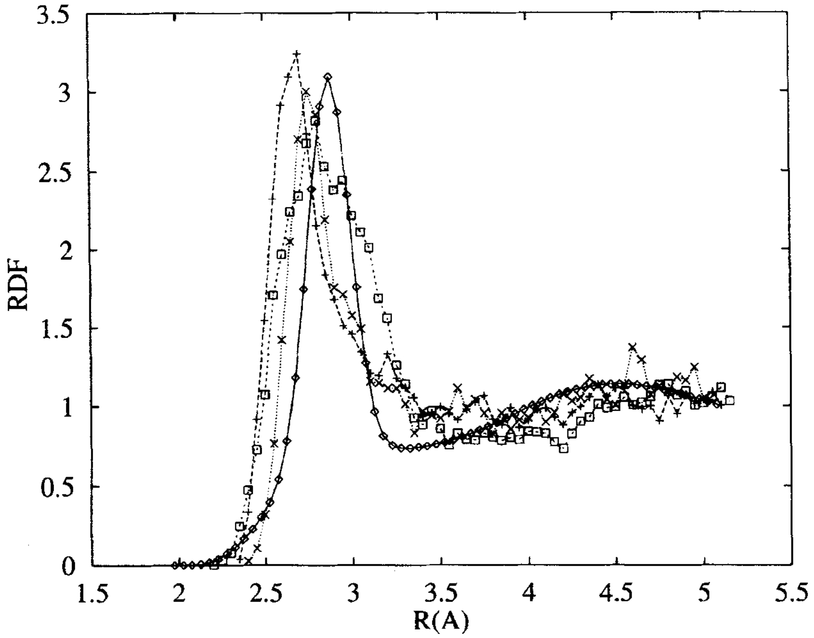

FIGURE 1 Radial distribution function $g(r_{OO})$. Experimental [11] (◇), classical MD (×), AM1(□), HF (+).

from the AM1 simulation appears to be closer to the experimental curve than the STO-3G curve but having a lower peak height.

Figure 2 contains the corresponding $g(r_{OH})$ curves. The striking feature here is that the AM1 model completely fails to describe the hydrogen-bonds. This comes as a surprise, since the AM1 model was claimed to be superior to the MNDO model in being able to reproduce hydrogen-bonds [44]. The same observation was made in the work by Tunon *et al.* [40]. The HF curve is much more shifted to closer distances than the classical $g(r)$, giving too short H-bond distance. However, the height of the second peak is in better agreement with the experimental curve.

The hydrogen–hydrogen radial distribution functions are shown in Figure 3. Again, the AM1 model fails the reproduce the characteristic shape of $g(r_{HH})$ for water. The flexible SPC model and the HF STO-3G both give too high intensities in comparison with the experimental curve. Again, the HF curve is shifted too much to closer distances.

It is of interest to compare the radial distribution functions obtained here with those reported by Laasonen *et al.* [10] with a box length of 9.6 Å. Laasonen *et al.*, report two sets of RDF curves for heavy water. The first set

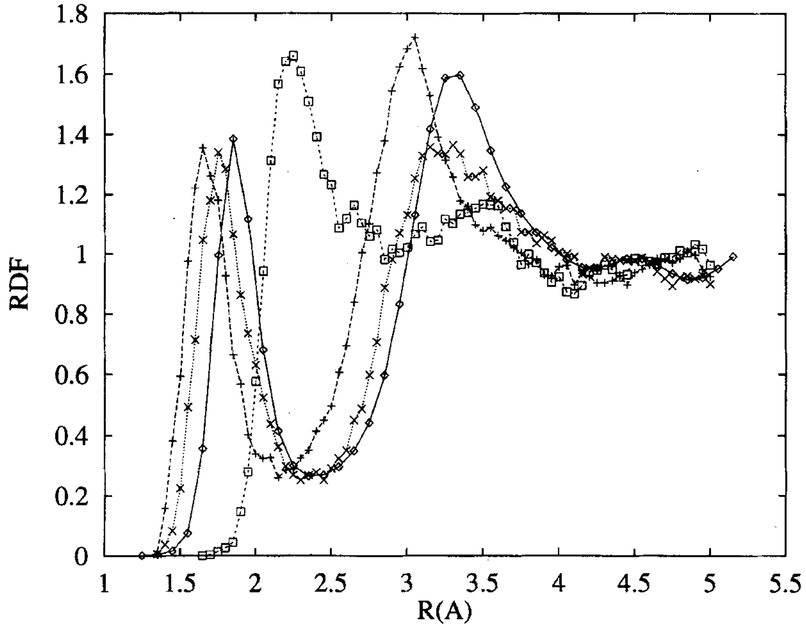

FIGURE 2 Radial distribution function $g(r_{OH})$. Experimental [11] (◇), classical MD (×), AM1 (□), HF (+).

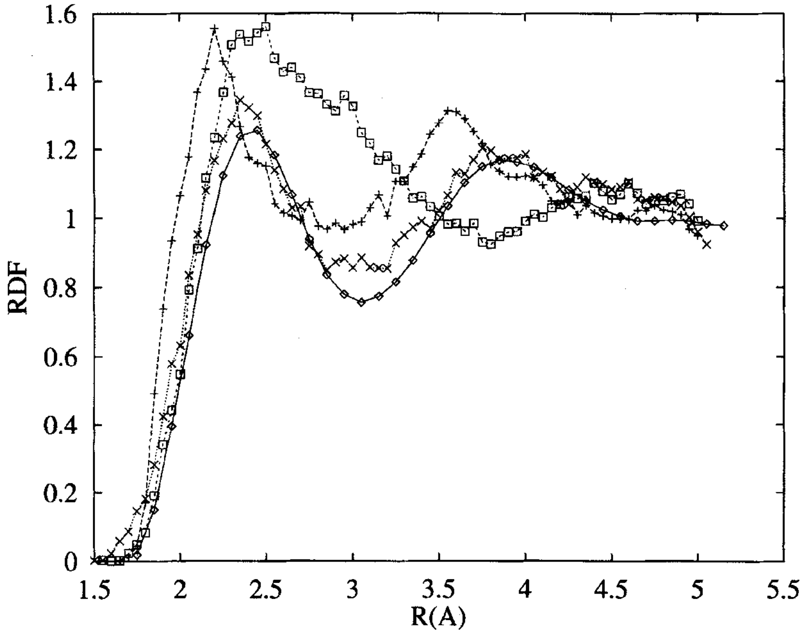

FIGURE 3 Radial distribution function $g(r_{HH})$. Experimental [11] (◇), classical MD (×), AM1 (□), HF (+).

is calculated with full Becke-Perdew corrections. After taking the initial configuration from a classical MD simulation with TIP4P potential model [45] they equilibrated for 0.5 ps and collected results during 1.5 ps. Interestingly, their results appear to be quantitatively very much similar to ours, at the Hartree-Fock level. They also notice that all the curves shifted to closer distances. Compared to normal water, the heavy water is slightly more structured due to the mass effect. The other set is carried out using a different gradient correction in order to enhance the exchange effects by omitting the Perdew term from the general gradient correction and recalculating the pseudopotential. This gave a considerable improvement to the RDF results.

Dynamical Results

Figure 4 contains mean square displacement (MSD) curves of the centre-of-mass of water using three different types of water models: classical, AM1 and Hartree-Fock STO-3G. These curves can be used to calculate the translational diffusion coefficients. All the three curves give comparable slopes and should give similar diffusion coefficients.

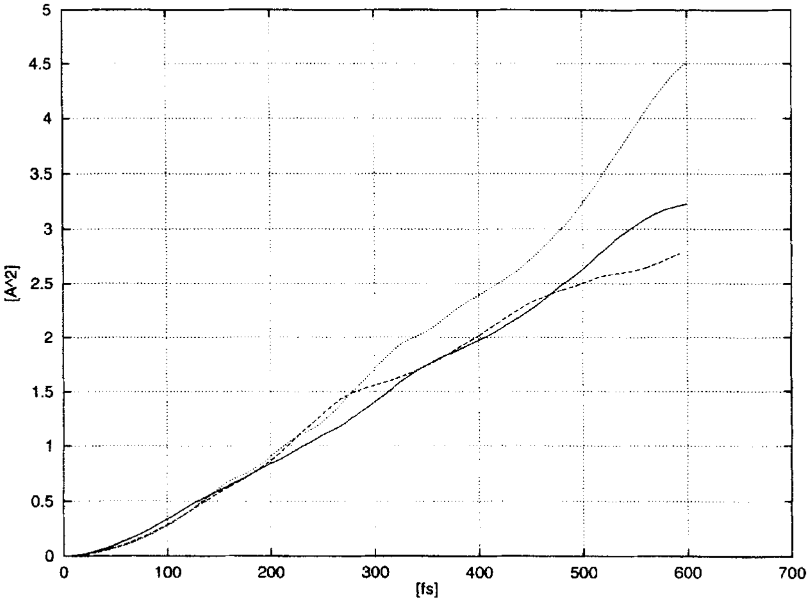

FIGURE 4 Mean square displacements for the centre-of-mass of water. AM1 model (dashed), Hartree-Fock STO-3G (dotted), classical MD (solid).

The flexible SPC model by Toukan and Rahman [36] is shown to give diffusion coefficients in good agreement with the experimental values reported in [46]. Therefore, using the classical curve as a reference, the translational dynamics should be described reasonably well when using either the AM1 or the Hartree-Fock STO-3G models.

Also, the reorientational correlation functions were calculated from the simulation data. The time scale of reorientation is, however, far beyond the time scale covered in the present simulation. Only, the initial decay of these autocorrelation functions could be calculated. All the three models gave comparable results (not shown).

4. PARALLEL CONSIDERATIONS AND IMPLEMENTATION

Parallel computers are an obvious choice to solve very CPU intensive tasks [47]. MD simulations are naturally parallel since, during each time step, interactions on molecules can be calculated and accumulated independently of other molecules. An overwhelmingly large fraction of the total CPU simulation time is spent on calculating these interactions. It can be differently split between processors depending on the character of interactions between particles (short-range/long-range) and on the size of the simulated system (typically the number of interacting sites) [48, 49]. Following the two first alternatives discussed in the beginning of Section 1, we have previously developed massively parallel methods for large scale applications [50, 51], with many more particles than processors. The method outlined in Section 2 follow the third alternative of Section 1. It differs in that the number of tasks and processors are approximately the same and that the character of force calculation is very much different than in a classical MD scheme.

During each time step, the quantum chemical calculations of molecular clusters are, besides very CPU demanding, also completely independent of each other and can be done concurrently. Also the integration of forces can be done locally to obtain the new positions and velocities for the atoms. The amount of data to be communicated is limited only to the atomic coordinates, velocities of each molecule and a few molecular properties. The inter-processor communication should practically not take any time at all in comparison to the time the nodes are busy doing the force calculations. At first sight, this problem appears to be an ideal and straightforward problem to parallelize. We have a number of large-grain tasks that only need to communicate very small amounts of data, and during each time step we only need one synchronization point. This problem naturally fits the Replicated Data (RD) [48, 49] parallelization.

However, we cannot use more than N processors and load balancing becomes a problem if there are variations in the time it takes to complete a cluster calculation task. With only one task per processor the most time consuming task in every time step sets the pace, and this means that efficiency will decrease. Since the problem complexity is $O(N * L^4)$ it is clear that in order to efficiently use more processors than there are particles we must also be able to parallelize cluster calculations (the L dimension). This implies that we need software that will allow us to vary the number of processors used for a cluster calculation as well as the number of particles assigned to a set of processors. When parallelizing along both the N and L problem dimensions, or when running on an inhomogeneous machine resource, a strategy using task farming should also be a flexible and reasonably efficient approach. In the present study we have only parallelized over the number of molecules, the main reason being that the version of the quantum chemistry package, available to us, was serial.

Implementation on an SP2

The parallel computer used in the current work is the IBM Power Parallel SP, commonly called the SP2. The SP2 has a MIMD architecture: the interconnect is a high performance switch with each node being a standard POWER2 RS/6000 processor with local memory. The SP2 installation we have used contains two different types of nodes: "thin nodes" and "wide nodes", often called T2 and W1 respectively. Compared to T2 nodes, the W1 nodes have twice the memory bandwidth and data cache. In the machine used W1 nodes had four times more physical memory than T2 nodes. Details of the hardware configuration used can be found in Tables I and II.

We started from a conventional MD program, added code to evenly distribute the cluster calculations to different processors and also interfaced it to a standard quantum chemistry package (see Section 2). The parallel scheme is implemented in such a way that the program will automatically accommodate for the number of available nodes at program start-up. The Unix system call *system* is used to start the cluster calculations from the MD code. The MD program is idle while the quantum mechanical forces are calculated.

We have used the User Space (US) communication subsystem and the high-performance switch. The parallel communication library used was

TABLE I Technical data for the nodes of the 96 node (80 T2 and 16 W1) SP2 used

SP2 Nodes		*T2*	*W1*
peak performance	[MFlop/s]	267	267
frequency	[MHz]	66.7	66.7
scratch disk	[GByte]	2.0	2.0
memory	[MByte]	128	512
b/w to memory	[GByte/s]	1.0	2.1
data cache	[KByte]	128	256
instruction cache	[KByte]	32	32

TABLE II Technical data for the high-performance switch

bandwidth	[MByte/s]	35
latency	[μs]	40
adaptors		TB2

IBM's MPL. To be able to use both US communication and *system* we had to explicitly tell the compiler to use the extended memory model. Details of the software configuration can be found in Table III and an MPL message passing pseudo MD code is given in Appendix A.

Timing Results on an SP2

To obtain figures on performance and scalability, and to exclude cold-start effects, we have made a number of runs for some combinations of type of processor, number of processors and quantum mechanical calculation method. Timings were done using *timef* which gives wall clock elapsed time. Times reported in Table IV are the fastest runs of each type measured on processor 0. Also, to minimize external interference, we made sure that once a node was allocated it was ours exclusively.

Performance

For the semi-empirical AM1 scheme, the time per time step and molecule was around 5 seconds for W1 and around 7 seconds for T2 nodes. Using the Hartree-Fock SCF MO scheme with the STO-3G basis set the calculation

TABLE III Details of software and compiler options used

operating system	AIX 4.1.4
parallel software	PSSP 2.1
message passing library	MPL
compiler	mpxlf
compiler options	-O2-qarch = pwr2-bmaxdata:0x40000000

TABLE IV Timings from runs with different number W1 and T2 nodes. Values are given in seconds per time step per particles per node. The AM1 and the HF quantum mechanical cluster calculation methods have been tried with and without checkpointing

# nodes	*AM1*		*HF*	
W1	*chkpt*	*no chkpt*	*chkpt*	*no chkpt*
1	4.6	5.1	44.8	45.7
2	4.6	5.1	57.3	58.1
4	4.5	5.1	59.2	59.9
8	4.6	5.2	58.7	59.6

# nodes	*AM1*		*HF*	
T2	*chkpt*	*no chkpt*	*chkpt*	*no chkpt*
1	6.2	5.3	137.6	138.0
2	6.4	5.4	135.8	135.2
4	6.7	5.7	140.1	140.6
8	6.7	5.9	143.3	144.6
16	7.7	7.0	151.0	155.1
32	10.6	10.0	164.5	167.1

times per molecule and time step went up by around a factor 12 for wide nodes and around a factor 20 for thin nodes. See Table IV.

When running g94 we have used the same "Default.Route" for both T2 and W1 nodes: 95 MByte memory and 2 GByte temporary disk. It is interesting to see that processors that are otherwise almost identical, except for bandwidth to memory, differ so distinctly. The gap in speed between T2 and W1 nodes should mainly be attributed to this difference.

Scaling Properties

In Table V we have calculated the speed-up for varying number of nodes and processors. The speed-up achieved is respectable for almost all cases.

TABLE V Speed-up from runs with different number of W1 and T2 nodes

# nodes	*AM1*		*HF*	
W1	*chkpt*	*no chkpt*	*chkpt*	*no chkpt*
1	1.0	1.0	1.0	1.0
2	2.0	2.0	1.6	1.6
4	4.0	4.0	3.0	3.0
8	7.9	7.8	6.1	6.1

# nodes	*AM1*		*HF*	
T2	*chkpt*	*no chkpt*	*chkpt*	*no chkpt*
1	1.0	1.0	1.0	1.0
2	1.9	2.0	2.0	2.0
4	3.7	3.7	3.9	3.9
8	7.4	7.2	7.7	7.6
16	12.8	12.1	14.6	14.2
32	18.8	17.0	26.8	26.4

See Table V and Figure 5. The most outstanding exception comes from running the AM1 model on 32 processors. The explanation for the poor speed-up in this case comes from the fact that the calculation times of each cluster calculation are short and the variations in calculation time are relatively large. In general, since we are using a static assignment of molecules to processors, we cannot expect speed-ups to be perfectly linear. Using a dynamic assignment based on previous calculation times for each molecule one could devise a load balancing scheme that would be slightly more efficient. However, when weighing the increased code complexity against what can be gained in efficiency, it is hard to motivate at this stage.

5. SUMMARY

This communication shows that it is a relatively simple matter to set up a software to perform computer simulations using quantum mechanical force fields on Born-Oppenheimer surfaces. Although it is possible to do it at an

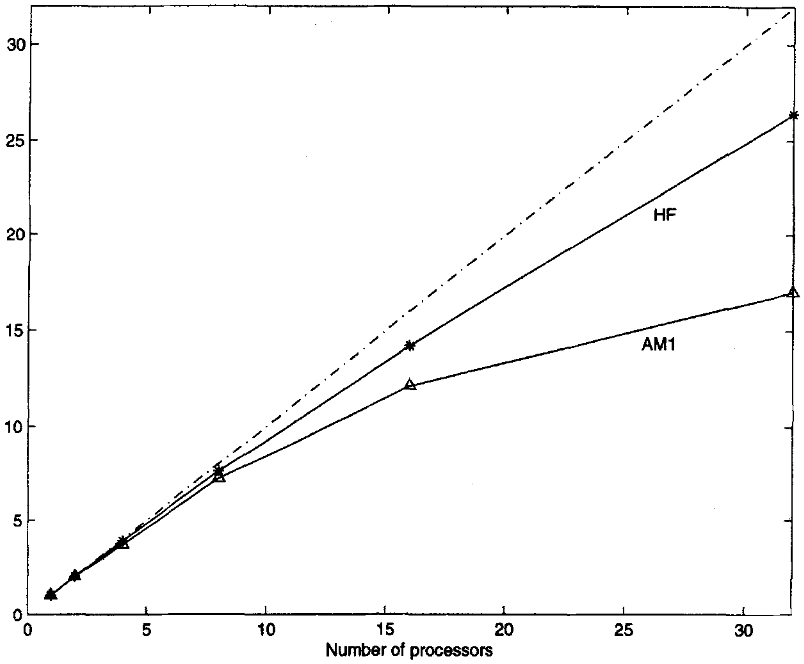

FIGURE 5 Speed-up plot for T2 nodes without checkpointing. AM1 model (Δ), Hartree-Fock STO-3G(∗).

arbitrary level of sophistication, it is still computationally very expensive. Because of the large CPU demand, we have only used the low-end methods of quantum chemical methodology and a limited basis set. Still, it is encouraging to observe that the structure and dynamics of liquid water can already be described reasonably well. Knowing the capacity of high quality quantum chemical single point calculations, it should be possible to carry out very accurate MD simulations some time in future. For example, the too close distances, now observed in the RDF curves should be corrected by increasing the size of the basis set.

From a pure computational point of view, the problem is ideal to be solved on parallel computers, involving heavy tasks of local computations in different nodes, followed by a very small amount of communication. The scaling properties are good. Load balancing, however, is more tricky because of the fluctuating character of local particle densities in a liquid. The amount of computation for the clusters, as in our case, vary even if the number of electrons is kept constant. This is because of fluctuating atom–atom distances, requiring sometimes more iterations in order to converge to a same accuracy. These fluctuations are not easy to predict making the optimization of the load balancing difficult. This becomes obvious from the performance figures in Table IV in cases where the number of nodes is equal to the number of clusters.

In the forthcoming paper we address various properties which are not accessible in simulations using atom–atom effective pair potentials. These include three dimensional distributions of electron densities and electric potential as well as induced molecular multipole moments and polarizabilities.

Acknowledgements

This work has been supported by the Swedish Natural Science Research Council (NFR).

A PSEUDO CODE IN FORTRAN AND MPL

A pseudo code for a message passing approach to first principles molecular dynamics given in Fortran/MPL. Only the vital parts for the parallelisation are kept for the sake of simplicity. Some elementary knowledge of the structure of standard MD programs is required in order to implement the scheme.

```
  PROGRAM QMD
C
C ...
C
    CALL MP_ENVIRON(np, myid)
    CALL MP_TASK_QUERY(nbuf, 4,3)
    IDG = nbuf(4)
C
C Initialize positions (SX, SY, SZ) and
C velocities (VX, VY, VZ) for all NSTOT atoms.
C
    LENMP = 8*NSTOT       !(8 bytes→double precision)
    CALL MP_SYNC(IDG)
    CALL MP_BCAST (SX, LENMP, O, IDG)

C
C     The same for SY, SZ, VX, VY, VZ
C
    DO NSTEP = 1, NSTEPS
      DO I = 1+myid, NR_MOLECULES, np
C
C Calculate forces on atoms of molecule I using neighbor lists.
C Integrate forces to temporary sparse arrays:
C SPX, SPY, SPZ, VPX, VPY, VPZ.
C
    END DO
C
C Wait for all processors to finish and add up the changes in positions and
C velocities accumulated in each processor.
C
  CALL MP_SYNC(IDG)
  CALL MP_COMBINE(SPX, SX, D_VADD, IDG)
C
C     The same for SY, SZ, VX, VY, VZ
C
    END DO
```

```
C
C ...
C
      STOP
      END
```

References

[1] Allen, M. P. and Tildesley, D. J. *The Computer Simulation of Liquids*, Oxford: Clarendon Press, 1987.

[2] *Computer Simulation in Chemical Physics*, Allen, M. P. and Tildesley, D. J. eds., NATO ASI Series, Kluwer Academic Publishers, Dordrecht, 1993.

[3] Hennessy, J. L. and Patterson, D. A. *Computer Architecture a Quantitative Approach*, Morgan Kaufmann Publishers Inc., San Mateo, California, 1990.

[4] Hwang, K, *Advanced Computer Architecture: Parallelism, Scalability, Programmability*, McGraw-Hill Inc., 1993.

[5] *Many-body Tree Methods in Physics*, Susanne Pfalzner and Paul Gibbson, Cambridge University Press, 1996.

[6] Zhu, S.-B., Yao, S., Zhu, J.-B., Singh, S. and Robinson, G. W. (1991). "A flexible/polarizable simple point charge water model", *J. Phys. Chem.*, **95**, 6211–6217 and references therein.

[7] Stone, A. J. and Alderton, M. (1985). "Distributed multipole analysis. Methods and applications", *Mol. Phys.*, **56**, 1047–1064.

[8] Pulay, P. (1969). "*Ab initio* calculation of force constants and equilibrium geometries in polyatomic molecules. I Theory", *Mol. Phys.*, **17**, 197–204.

[9] Pulay, P. (1987). "Analytical derivative methods in quantum chemistry" in *Ab initio methods in Quantum Chemistry, II*, 241–287 (Ed. K. P. Lawley) J. Wiley.

[10] Laasonen, K., Sprik, M., Parrinello, M. and Car (1993). "*Ab initio*, liquid water", *J. Chem. Phys.*, **99**, 9080–9089.

[11] Soper, A. K. and Philips, M. G. (1986). "A determination of the structure of water at 25°C", *Chem. Phys.*, **107**, 47–60.

[12] Åqvist, J. and Warshel, A. (1993). "Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches", *Chem. Rev.*, **93**, 2523–2544 and references therein.

[13] Alagona, G., Desmeules, P., Ghio, C. and Kollman, P. A. (1984). "Quantum mechanical and molecular mechanical studies on a model for dihydroxyacetone phosphate–glyceraldehyde phosphate isomerization catalyzed by triosephosphate isomerase (TIM)", *J. Am. Chem. Soc.*, **106**, 3623–3632.

[14] Field, M. J., Bash, P. A. and Karplus, M. (1990). "A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations", *J. Comp. Chem.*, 700–733.

[15] Gao, J. and Xia, X. (1992). "A priori evaluation of aqueous polarisation effects through Monte Carlo QM-MM simulations", *Science*, **258**, 631–635.

[16] Laaksonen, A., Wang, J. and Boyd, R. J. (1995). Internal motion and flexibility of benzene. A molecular dynamics study", *Chem. Phys. Lett.*, **241**, 380–386.

[17] Wang, J., Laaksonen, A., Eriksson, L. A. and Boyd, R. J. (1995). "Internal motion and tunneling rates of CH_4^+ and CD_4^+", *J. Chem. Phys.*, **103**, 8166–8176.

[18] Wang, J., Boyd, R. J. and Laaksonen, A. (1996). "A Hybrid quantum mechanical force field molecular dynamics simulation of liquid methanol. Vibrational frequency shifts as a probe of the quantum mechanical molecular mechanical coupling", *J. Chem. Phys.*, **104**, 7261–7269.

[19] Eriksson, L. A. and Laaksonen, A. (1996). "Hybrid DFT-MD simulations of geometry and hyperfine structure of the CCH radical in argon matrices at low temperature", *J. Chem. Phys.*, **105**, 8195–8203.

[20] Car, R. and Parrinello, M. (1985). "Unified approach for molecular dynamics and density-functional theory", *Phys. Rev. Lett.*, **55**, 2471–2474.

[21] Hohenberg, P. and Kohn, W. (1964). "Inhomogeneous electron gas", *Phys. Rev.*, **136**, B 864–871.

[22] Kohn, W. and Sahm, L. J. (1963). "Self-consistent equations including exchange and correlation effects", *Phys. Rev.*, **140**, A 1133–1138.

[23] Hellman, J. *Einführung in die quantenchemie*, Leipzig, Deuticke and Co., 1937.

[24] Feynman, R. P. (1939). "Forces in molecules", *Phys. Rev.*, **56**, 340–343.

[25] Vanderbilt, D. (1990). "Soft self-consistent pseudopotentials in a generalized eigenvalue formalism", *Phys. Rev. B.*, **41**, 7892–7895.

[26] Laasonen, K., Car, R., Lee, C. and Vanderbilt, D. (1991). "Implementation of ultrasoft pseudopotentials in *ab initio* molecular dynamics", *Phys. Rev. B*, **43**, 6796–6799.

[27] Galli, G. and Pasquarallo, A. "First principles molecular dynamics" in *Computer Simulations in Chemical Physics* (Eds. M. P. Allen and D. J. Tildesley), NATO ASI Kluwer, Dordrecht, 1993.

[28] Remler, D. K. and Madden, P. A. (1990). "Molecular dynamics without effective potentials via the Car-Parrinello approach", *Mol. Phys.*, **70**, 921–966.

[29] Laaksonen, A. (1986). "Computer simulation package for liquids and solids with polar interactions. I Mc-MOLDYN/H_2O: Aqueous systems", *Comp. Phys. Commun.*, **42**, 271–300.

[30] Gaussian 94 (Revision B2), Frisch, M. J., Trucks, G. W., Schlegel, H. B., Gill, P. M. W., Johnson, B. G., Robb, M. A., Cheeseman, J. R., Keith, T. A., Peterson, G. A., Montgomery, J. A., Raghavachari, K., Al-Laham, M. A., Zakrewski, V. G., Ortiz, J. V., Foresman, J. B., Cioslowski, J., Stefanov, B. B., Nanaykkara, A., Challacombe, M., Peng, C. Y., Ayala, P. Y., Chen, W., Wong, M. W., Andres, J. L., Replogle, E. S., Gomperts, R., Martin, R. L., Fox, D. J., Binkley, J. S., Defrees, D. J., Baker, J., Stewart, J. P., Head-Gordon, M., Gonzales, C. and Pople, J. A. (1995). Gaussian, Inc., Pittsburgh, PA.

[31] Eisenberg, D. and Kauzmann, W. (1969). *The Structure and Properties of Water*, Oxford University Press, Oxford.

[32] Narten, A. H. and Levy, H. A. (1971). "Liquid water: molecular correlation functions from X-ray diffraction", *J. Chem. Phys.*, **55**, 2263–2269.

[33] Krynicki, K., Green, C. D. and Sawyer, D. W. (1978)."Pressure and temperature dependence of self-diffusion in water", *Faraday Discuss. Chem. Soc.*, **66**, 199–208.

[34] Krynicki, K. (1966). "Proton spin-lattice relaxation in pure water between 0°C and 100°C", *Physica*, **32**, 167–178.

[35] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. and Hermans, J. "Interaction models for water in relation to protein hydration" in *Intermolecular Forces*, 331–342, (Ed. B. Pullman), Reidel Dordrecht.

[36] Toukan, K. and Rahman, A. (1981). "Molecular-dynamics of atomic motions in water", *Phys. Rev. B*, **31**, 2643–2648.

[37] Nóse, S. (1984). "A molecular dynamics method for simulations in the canonical ensemble", *Mol. Phys.*, **52**, 255–268.

[38] Hoover, W. G. (1985). "Canonical dynamics: equilibrium phase-space distributions", *Phys. Rev. A*, **31**, 1695–1697.

[39] Wei, D. and Salahub, D. R. (1994). "A combined density functional and molecular dynamics simulation of a quantum water molecule in aqueous solution", *Chem. Phys. Lett.*, **224**, 291–296.

[40] Tunon, I., Martins-Costa, M. T. C., Millot, C., Ruiz-Lopez, M. F. and Rivail, J. L. (1996). "A coupled density functional-molecular mechanics Monte Carlo simulation method: The water molecule in liquid water", *J. Comp. Chem.*, **17**, 19–29.

[41] Moriarty, N. W. and Karlström, G. (1996). " Electronic polarization of a water molecule in water. A combined quantum mechanical and statistical mechanical treatment", *J. Phys. Chem.*, **100**, 17791–17796.

[42] Sprik, M., Hutter, H. and Parrinello, M. (1996). "*Ab initio* molecular dynamics simulation of liquid water: comparison of three gradient-corrected density functionals", *J. Chem. Phys.*, **105**, 1142–1152.

[43] Ortega, J., Lewis, J. P. and Sankey, O. F. (1997). "First principles simulations of fluid water: the radial distribution function", *J. Chem. Phys.*, **106**, 3696–3701.

[44] Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. and Stewart, J. J. P. (1985). "AM1: A new general purpose quantum mechanical molecular model", *J. Am. Chem. Soc.*, **107**, 3902–3909.

[45] Jorgensen, W. L. and Madura, J. D. (1986). "Temperature and size dependence for Monte Carlo simulations of TIP4P water", *Mol. Phys.*, **56**, 1381–1392.

[46] Anderson, J., Ullo, J. J. and Yip, S. (1987). "Molecular dynamics simulation of dielectric properties of water", *J. Chem. Phys.*, **87**, 1726–1732.

[47] Forster, I. T. *Designing and Building Parallel Programs*, Addison-Wesley, Reading, Mass., 1995.

[48] Fincham, D. (1987). "Parallel computers and molecular simulation", *Molecular Simulation*, **1**, 1–45.

[49] Smith, W. (1991). *Comput. Phys. Comm.*, "Molecular dynamics on hybercube parallel computers", **62**, 229–248 (1991).

[50] Hedman, F. and Laaksonen, A. (1993). "Data-parallel large-scale molecular dynamics for liquids", *Int. J. Quant. Chem.*, **46**, 27–38.

[51] Hedman, F. and Laaksonen, A. (1995). "A data-parallel molecular dynamics method for liquids with coulombic interactions", *Molecular Simulation*, **14**, 235–244.