Large scale electronic structure calculations in the study of the condensed phase

H.J.J. van Dam [a],*, M.F. Guest [a], P. Sherwood [a], J.M.H. Thomas [a],
J.H. van Lenthe [b], J.N.J van Lingen [b,c], C.L. Bailey [a], I.J. Bush [a]

[a] *Computational Science and Engineering Department, CCLRC Daresbury Laboratory, Warrington WA4 4AD, UK*
[b] *Theoretical Chemistry Group[1], Debye Institute, Utrecht University, 3508 CH Utrecht, The Netherlands*
[c] *Department of Inorganic Chemistry and Catalysis, Debye Institute, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands*

Available online 10 May 2006

Abstract

We consider the role that large-scale electronic structure computations can now play in the modelling of the condensed phase. To structure our analysis, we consider four distinct ways in which today's scientific targets can be re-scoped to take advantage of advances in computing resources:

1. Time to solution—performing the same calculation, with delivery of the simulation in shorter elapsed time;
2. Size—applying today's methods to a more extensive problem;
3. Accuracy—replacing current physical models with more accurate ones.
4. Sampling—simultaneously studying more chemical or conformational states.

Each of these offer some scientific rewards, but all present technological challenges and it is likely that a mixture of approaches will be needed to make the best use of capability computing. We discuss some aspects of our work in each of these areas, including replicated and distributed data parallel implementations of GAMESS-UK, and approaches incorporating multi-level parallelism. Examples of the latter include pathway optimization using replica methods and task farming approaches to global optimization problems.

We consider a typical application of GAMESS-UK to heterogeneous catalysis, and the role that large-scale DFT vibrational frequency calculations have played in the study of partial oxidation catalysts incorporating supported VO_x species.
© 2006 Elsevier B.V. All rights reserved.

Keywords: GAMESS-UK; parallel computing; QM methods; QM/MM methods; DFT

1. Introduction

The processing power available to computational science has continued to show a dramatic increase, year on year. Ab initio electronic structure methods have developed from a specialised discipline, in which the leading computers of the day were used to solve small molecule problems, to a ubiquitous tool for computational chemistry. DFT calculations on systems of tens to hundreds of atoms are now routinely performed on the same machines that are procured for word-processing. Alongside the increases in processor speed, we have seen massive increase in the scale of parallel processing,

with the result that the power of the national facilities has increased some 10,000-fold in the last 25 years. Considering the UK academic scene as an example, Daresbury Laboratory houses the current leading national resource HPCx [1], an IBM system of 1600 Power4+ CPUs, linked by a high performance switch with a sustained LINPACK performance of over 6 Tflops, soon to be upgraded to 13 Tflops. Compare this with the Cray 1S, on site in 1979, which had a sustained performance of 160 Mflops. It is not unreasonable to look ahead to the possibility of PetaFlop systems in the next 5 years [2,3]. Exploiting this level of hardware performance presents many challenges, and at Daresbury Laboratory we have, over many years, been engaged in a programme of methodological and software developments to help the UK academic community make the optimum use of such machines [4]. In this article, we describe a number of recent developments of particular relevance to the study of chemical catalysis in the condensed phase.

---

* Corresponding author. Tel.: +44 1925 603933; fax +44 1925 603100.
  E-mail address: h.j.j.vandam@dl.ac.uk (H.J.J. Dam).
[1] Affiliated to Organic Chemistry and Catalysis.

2. Opportunities for tera-scale computational chemistry

In the following sections we address the challenges and opportunities posed by the current generation of 'Tera-scale' computing resources, of which HPCx, introduced above, is a typical example. It is obvious that parallelism allows an increase in the processing power available to a computational chemistry calculation. Serial programs are limited by individual processor performance, which currently stands at around 5–10 peak Gflops. There is some scope for parallelism at the vector instruction level, but for many computational chemistry codes the scope for vectorisation is limited to speedups of around 50 [5,6]. The current generation of parallel computers, however, offers significantly greater opportunities for large scale computing. The Top 500 list [7] includes over 150 systems with processor counts between 500 and 10,000, mostly based on commodity processors with individual peak performance 2–6 Gflops, providing a sustained LINPACK performance of 1–50 Tflops. Considering future parallel systems, we note a major shift in the semiconductor industry strategy that has become readily apparent during the past 18 months, whereby increased processor throughput will be accomplished by providing more computing elements ('cores'), rather than by increasing the operating frequency. Indeed, the 5-year semiconductor industry roadmap shows a factor of two increase in operating frequency from 3.2 (2004) to 6.4 GHz (2009), whereas Intel, AMD, and others are migrating to multiple ($2\times$, $4\times$, $8\times$ …) cores within a single 'processor.' This shift will lead to increased processor throughput being accomplished through multiple cores, thereby increasing the burden of parallelization. Assuming 10 Gflops for each processing element, a potential Pflop system might thus comprise some 100,000 processing elements. The ability to deploy such massive parallelism offers a significant enhancement in the scope of all computational science disciplines. As chemistry is one of the leading consumers of cycles in most academic computing facilities, there is particular opportunity in our field. Compared to the desktop systems there is the potential to reduce time to solution by two-three orders of magnitude, potentially reducing a month long calculation to one that will run overnight. However, this benefit has proved difficult to realise in practice, particularly in the discipline of computational chemistry, where scalability beyond several hundred processors is proving quite elusive. We structure the following discussion by considering four specific ways in which large scale parallelism may be used to extend the scope of computational chemistry calculations. We illustrate the general principles with examples from our work in support of HPCx.

2.1. Reduction in time to solution

This is perhaps the most obvious way to deploy parallel processing. By porting the current code-base to parallel systems today's workload can be executed in reduced real time, and many scientific opportunities follow from this. Dynamical simulations can be run for a longer simulation times; as well as improving the statistical reliability of the time-averaged results, such developments enable the study of new classes of processes, such as conformational changes in proteins. There is also the possibility of increased interactivity in the study of chemical processes; the use of haptic devices [8] and the parameterisation of potential energy surfaces [9] clearly can take advantage of elapsed times reduced from hours to seconds or minutes.

In practical terms, the first step in such an approach will be the parallel implementation of the existing serial program. Many different techniques have been applied, but the most common approach has been that of replicated data parallelism. The same program is run on every node of the system, each maintaining the same 'replicated' data structures. Some parts of the program, for example input handling, will be performed on every node, to maintain the consistency of the data. Where possible, the computational work is divided across the processors, and inter-processor communication phases are introduced to share the results so that all processors have the same data at the start of the next phase. The memory requirements are similar, on a per-node basis, to those of the serial code. Such implementations are generally quite easy and widely attempted. From our own work, examples are the early parallelisation work on the GAMESS-UK package [10] in which a message passing harness (TCGMSG) was used. This code is still in use, now adapted to use the standardised message passing interface (MPI). In practice, this approach to parallel processing works well on small numbers of processors, typically up to 16 or 32 nodes but for larger-scale parallelism it becomes more problematic. The two main limitations are Amdahl's law [11] (which dictates that best speedup that can be obtained is related to the fraction of serial code) and the fact that as the processor count increases the global communications increase in cost, typically scaling as ln(N) with processor count N. Serial and communication phases soon come to dominate. We have addressed the Amdahl's law effect by progressively parallelising more of the code in GAMESS-UK, to do this we have introduced the Global Array (GA) tools [12,13] and PeIGS [14] tools developed at PNNL. Together they allow us to perform the matrix algebra (similarity transforms ($Q^\dagger HQ$) and diagonalisation) in parallel and provide a way to implement a memory-mapped I/O system based on a global shared memory model.

As an example of this parallel strategy, we show the performance of the Analytic DFT 2nd derivatives module of GAMESS-UK. This is not strictly a replicated data program, as the transformed integrals are held in a distributed global array (in a conventional implementation a disk file or direct re-computation strategy would be used) – however, the program as a whole is based on the replicated data approach. The diagram shows performance, in arbitrary units, for a variety of parallel systems, relative to the IBM p690 (Power 4 processor) running on 32 processors (Fig. 1).

The figure shows that speedups of around 1.7 are obtained when doubling the number of processors, leading to a progressive loss of parallel efficiency on the higher node counts. This behaviour is typical of replicated data parallel
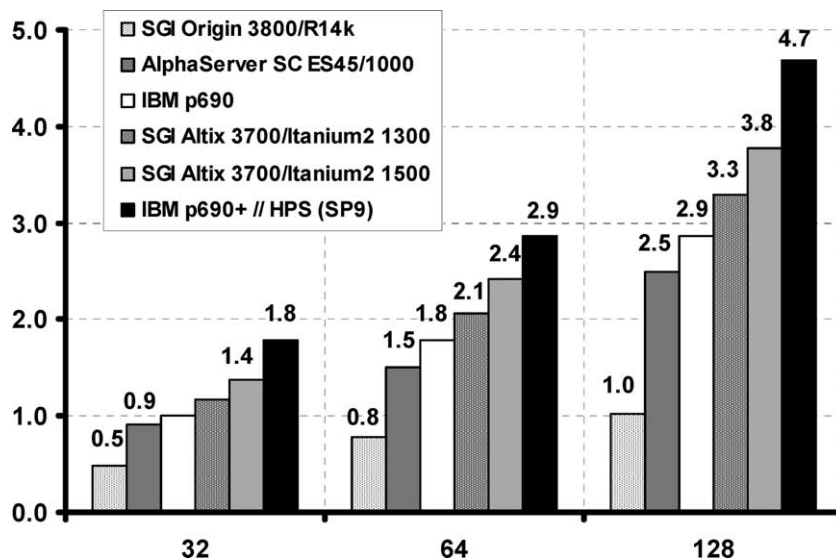
Fig. 1. Performance of the parallel DFT 2nd derivatives module of GAMESS-UK on a number of platforms.

implementations, providing useful parallelism up to perhaps 200 processors or so. For efficient use of tera-scale resources other approaches are clearly needed. The next sections outline three different lines of attack, which can be used, individually or in combination, to maximise the utility of large-scale parallel machines.

2.2. Scaling to larger problem sizes

The limitations of parallelism as a tool to simply reduce time-to-solution rest on the limited amount of work that can be performed in parallel. Scientific benefits can be realised by increasing the size of chemical systems, for example by including a larger part of a complex system in a model. The replicated data approach to parallelism outlined in Section 2.1 will generally yield increased efficiencies on large problem cases, because of the larger proportion of parallel work. However, the fact is that the memory requirements, on a given node, will increase as the problem size increases. For example, the replicated data parallel version of GAMESS-UK allows larger systems to be treated effectively, but the memory requirements limit the size of system to around 4000 basis functions on a computer with 1 GByte of memory per processor, irrespective of the number of processors used in the calculation. Such calculations are found to be scalable up to around 500 processors, at which point the SCF/gradient calculation is converged in 2000 s. In this case, the scalability limitations arise from the structure of the underlying serial program, specifically its heavy use of scratch file I/O and the costs of maintaining the replicated data structures. In some phases of the calculation, matrices which are temporarily distributed across the nodes (for example, in the parallel matrix multiply) must be re-replicated for the next phase of the calculation. It was clear that to make efficient use of the HPCx system we needed to redesign the SCF algorithm to make more effective use of the global memory. We have

adopted an approach in which most of the matrices used in the SCF scheme are distributed [15,16], while retaining a replicated data model for the Fock matrix construction step. The latter step involves the contraction of blocks of 4-index two-electron integrals, computed as required, with the density matrix P (Eq. (1)),

$$F_{\mu\nu} = H^0_{\mu\nu} + \Sigma_{\lambda\sigma} \, P_{\lambda\sigma}[(\mu\nu|\sigma\lambda) - 1/2(\mu\nu|\sigma\nu)] \tag{1}$$

Square Matrices   Integrals (computed on the Fly)

For each block of integrals, with a range of indices $\mu\nu\sigma\lambda$, we will need to access a number of elements of P and sum the result into the appropriate elements of F. In a fully distributed approach, these elements of F and/or P are likely to be held in remote memory, requiring either many 1-sided communications (which can be accomplished, for example, using the GA tools) or a complex systolic loop strategy presenting problems with load balancing. If F and P are replicated, all nodes can work independently until the construction is complete, at which point a global summation of contributions to F is required. The remainder of the SCF scheme is fully distributed, which allows a greater number of matrices to be held in core memory, thereby reducing the requirement for scratch file I/O. It is also possible to perform a sequence of distributed operations without re-replication [17]. In Fig. 2 we show the performance of this implementation, which makes use of MPI-based tools, such as ScaLAPACK. On HPCx, these were found to be significantly more efficient than the Global Arrays using the LAPI communications. This is less true with the current machine configuration as IBM's AIX 5.2 software service pack 9 has largely redressed the balance between MPI and LAPI performance. The use of replicated Fock and density matrices has one obvious drawback—some large replicated data structures are required. For a closed shell HF or DFT calculation two replicated matrices are required, and this is
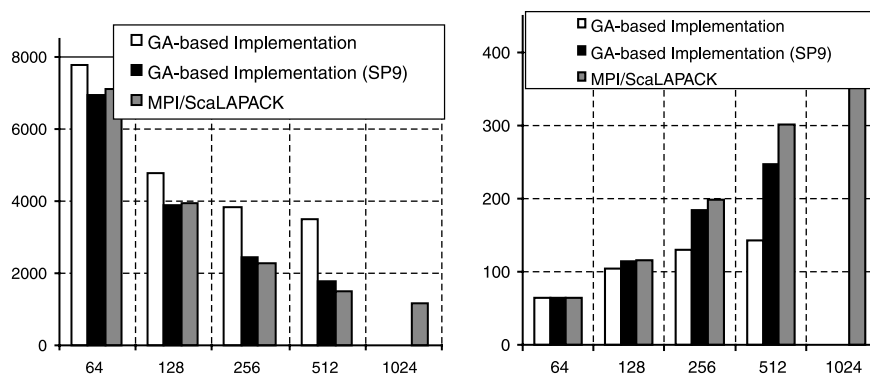
Fig. 2. (a) Time to solution and (b) Speedup for a SCF calculation on a zeolite fragment (3975 Basis functions) on the HPCx system.

doubled for UHF, giving us the ability to perform calculations of up to 12,000 basis functions on the HPCx system [17].

2.3. Enhanced accuracy

In almost all computational disciplines the choice of theoretical method is a compromise between the desired accuracy and practical time-to-solution. Parallel computing offers a way to mitigate this compromise. Choosing a more accurate method will often increase the total amount of work per computational step, opening up the possibility for increased parallel efficiency. In the realm of classical simulation, we can consider moving to forcefields incorporating polarization and/or multipole electrostatics (e.g. the DL_MULTI program—a derivative of DL_POLY). In the area of quantum mechanical methods two important options are to consider large basis sets and better treatments of electron correlation. Perhaps the best example of a sustained effort to develop scalable implementations of some of the most costly and accurate quantum chemical methods can be found within the Environmental Molecular Sciences Laboratory (EMSL) at PNNL. NWChem [18,19] is a family of electronic structure tools designed from the outset for parallel systems. It was designed and developed to be a highly efficient and portable MPP computational chemistry package and provides computational chemistry solutions that are scalable with respect to chemical system size as well as MPP hardware size. The program's architecture provides an extensible framework supporting development of new methods in computational chemistry, and has incorporated extensive tool development, such as the work on the global arrays referred to above.

Many studies of condensed phase systems now make use of hybrid QM/MM Methods. For systems with large numbers of degrees of freedom, QM/MM calculations are computationally demanding. HPC resources can enhance the value of such simulations by (i) allowing a more accurate QM Hamiltonian, e.g. DFT replacing semi-empirical methods for biomolecular simulation, or through (ii) the use of more accurate coupling between quantum and classical system, for example use of polarizable forcefields coupled to the SCF equations

2.3.1. Improving the QM methods—enzyme systems

Mixed Quantum Mechanical/Molecular Mechanical (QM/MM) methods are now routinely used within the computational biochemistry community as a way to study enzymatic reactions. The QM calculation of the active site is embedded in a simple, classical model for the environment. The need to perform calculations on many configurations, either as time-steps in a dynamical trajectory, different points on a reaction path or different configurations for the environment, has led to the widespread adoption of semi-empirical electronic structure methods. Such parameterised Hamiltonians have known failings and it would be ideal if, in the long run, more robust and flexible ab initio or DFT methods could replace semi-empirical studies in such calculations. We have been working with the developers of the CHARMM macromolecular modelling package [20] to facilitate QM/MM capabilities by providing an interface to GAMESS-UK within CHARMM [21–23].

When used for standard QM/MM geometry optimisations and molecular dynamics trajectories the parallelisation of the ab initio calculation follows the approached described above. There are limitations to the applicability of this mode of parallelisation, as the scaling shown by the small to medium scale calculations that are most useful for macromolecular QM/MM studies is intrinsically limited. There is more potential for massive parallelism if we take advantage of the fact that in such studies there will always be an interest in studying many configurations, and we provide an example below in Section 2.4.2.

2.3.2. Improving the QM/MM coupling—tools for solid state embedding

We are also working on QM/MM methods for inorganic and solid-state systems. The ChemShell package [24,25] offers an interface between quantum mechanical codes, such as GAMESS-UK, with classical models of solid-state systems using the shell model [26] such as that provided by the GULP code [27]. This combination is computationally demanding as the shell model potentials allow polarization of the classical part of the calculation, and this has to be coupled with the SCF procedure within the electronic structure calculation. This is achieved by a micro-iterative procedure, which increases the total number of SCF cycles required. The application of this
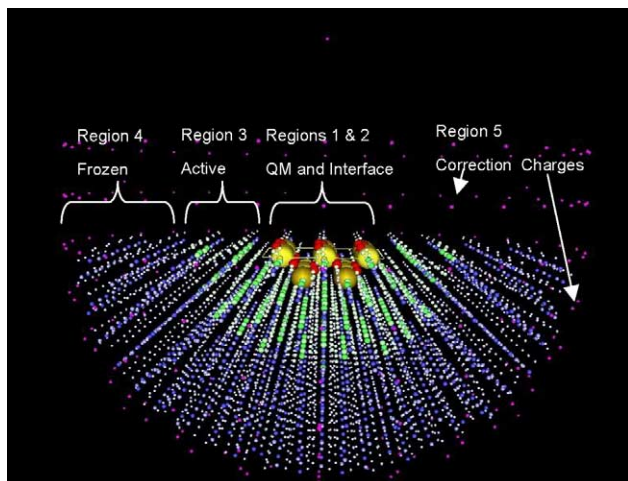
Fig. 3. Embedded cluster model for the material $FeSbO_4$.

approach to problems in heterogeneous catalysis is one of the most important applications of GAMESS-UK on the HPCx system at present.

An example of this embedding approach is shown in Fig. 3. This ionic material, $FeSbO_4$, is under study by Ricardo Grau-Crespo, Nora de Leeuw, Alexey Sokol and Richard Catlow at Birkbeck College and the Royal Institution, London. This material is thought to consist of O^{-2}, Sb^{5+} and Fe^{3+} ions, with the d^5 iron ions having sextet spin. The material as a whole is spin free, which is achieved by anti-ferromagnetic coupling between the rows of Fe^{3+} ions. The figure illustrates how the quantum mechanical cluster, classified as Region 1 and modelled at DFT/BB1K [28] level of theory, is surrounded by a multi-layer cluster model for the surface. A number of cations adjacent to the QM cluster are designated as Region 2 and carry point charges and pseudopotentials [24,29] to provide effective termination for the QM calculations. Beyond this, Region 3 is classically modelled by the shell model and both cores and shells are free to relax in the field of the QM region. Region 4 is frozen (constrained to the geometry and polarisation of a perfect, classically modelled surface) and around the exterior are a number of point charges, which are

least-squares-fitted to correct the electrostatic potential (Region 5).

To facilitate calculations on materials of this kind we have recently implemented an extension to GAMESS-UK to allow explicit specification of the atomic electron configurations used in the initial guess wavefunction. We were thus able to selectively calculate a variety of different spin distributions, two of which are shown in Fig. 4 (a) and (b). The energy difference between these two spin distributions is only 1.2 KJ at the Hartree–Fock level, and 8.7 kJ at the DFT/BB1K level.

2.4. Increased sampling

Finally, we consider the possibility of using large-scale parallel hardware to perform multiple, essentially independent calculations. We will consider two examples—the first is a trivially parallel case of simply submitting a large batch of independent jobs. Of perhaps more interest for high-end systems are the situations where a problem can be reformulated in such a way that it comprises a large number of weakly coupled computations. In some cases, there is a natural parallelism, for example, statistics for free energies, combinatorial methods, Monte Carlo (MC) Ensembles, etc. A number of such approaches are emerging, for example the replica exchange Monte Carlo methods [30], in which a number of simulations are conducted in parallel—those characterised for example by different simulation temperatures. More recently, many variations have emerged including one, in which the simulations differ by the versions of the forcefields employed, some having softer repulsive walls to allow faster exploration of configurational space [31]. A MC move involves exchanging the characteristic parameter between two simulations, with the result that conformational space is explored more effectively while still retaining the statistical distribution.

2.4.1. Quantum directed virtual evolution

The QDVE project is a collaboration with Dr Marcus Durrant (John Innes Centre, Norwich) in which a 'genetic' algorithm is used to determine an optimal catalyst for the conversion of nitrogen to hydrazine. A number of potential
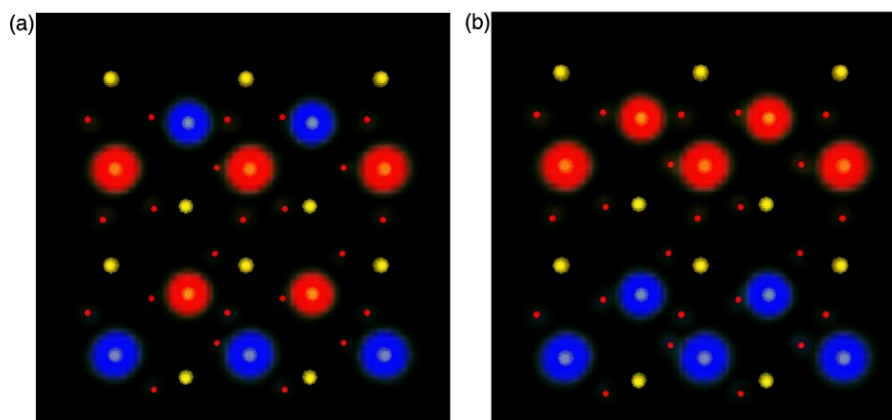

Fig. 4. The spin density of the ground state on the $FeSbO_4$ surface (a) with the oxygen and antimony ions shown in red and yellow, and the positive and negative spin densities on the iron ions shown in red and blue, (b) the spin density of an alternative electronic state.

catalytic compounds (collectively a 'generation') are created by combining a number of metal centres and ligands in several pre-defined geometries, with an associated range of spin and charge states. The energy for these catalytic complexes at specified points within the catalytic cycle are then calculated using GAMESS-UK and compared with an ideal value. A scoring system is used to select promising compounds, and a new generation of catalysts created by shuffling the characteristics of the chosen complexes (e.g. metal centre, geometry, etc.) between themselves. The new generation of catalytic compounds are submitted for calculation, and the process iterated until a catalyst of the desired efficacy has been found.

As the complexes within a generation are largely generated by random processes, their initial geometry and wavefunction may be considerably removed from their converged geometry. This places significant demands on the geometry optimiser and the SCF convergence procedure. To address some of these problems, we have made use of a module within GAMESS-UK that allows a 'user-defined' SCF convergence procedure to be used. A more robust scheme than the default one within GAMESS-UK has been developed and this has significantly improved the convergence characteristics for the complexes.

As a generation may contain several hundred complexes, the process of submitting the jobs for calculation and managing the results is not trivial. A 'taskfarming harness' has therefore been developed for GAMESS-UK to allow the multiple jobs within a generation to be handled largely as a single task on a large parallel computer. The harness is spawned across an arbitrary number of processors and the processors split into groups of a predefined size, with a single processor set aside as a server. Each group of processors runs an instance of GAMESS-UK in parallel mode. The server assigns jobs to the groups, who report back as soon as they have finished their calculation so that the server can note the result and assign the group with a new job.

Three generations of catalysts have been processed through the taskfarming harness so far, and of the eight parameters that contribute to determining the suitability of the catalysts, the best complexes already have five of these within acceptable limits.

2.4.2. The replica path and nudged elastic band methods

The replica path (RP) and nudged elastic band (NEB) methods [32,33] involve the simultaneous optimisation of a series of geometries of the reacting system, corresponding to a series of points along the reaction pathway. In the QM/MM variant of the NEB and RP [21] methods, the system is divided into three regions. The outermost part of the system is modelled classically and is *unreplicated*, meaning that all points on the pathway can share a common environment, which is optimised to a compromise configuration. If used with care, this helps to ensure that the study of the reaction is not complicated by changes in the conformation of spectator groups, or solvent rearrangements. Within the replicated region, a subset of atoms may be specified as quantum mechanical. The classical part of the system, (both replicated and non-replicated MM regions), is computed using the
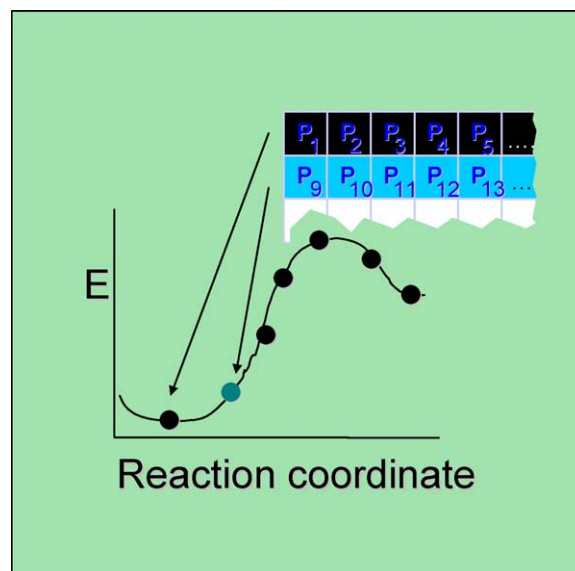

Fig. 5. Parallelisation of the Replica Path approach, showing distribution of processors across the replicas.

standard CHARMM parallel code. For the QM calculation, however, the CHARMM communication subsystem is switched such that the processors are grouped into independent sets, each set working on one of the points on the pathway. The QM/MM variants of these methods are particularly well suited to parallel execution, because they offer the possibility of multi-level parallelism. The processors are divided into groups, each group performing the QM calculation for one point along the pathway, as illustrated in Fig. 5.

The parallel nudged elastic band approach [33], using an ab initio QM/MM Hamiltonian has been tested on the chorismate/prephenate rearrangement [21] using a 6-31G* basis set for the QM region. So far we have measured a parallel speed-up of around 400 on 640 processors of HPCx.

3. Application example—cluster studies of vanadium oxide catalyst on silica

Supported vanadium oxide catalysts are widely used in industry for selective oxidation reactions. The catalysts consist of a vanadium oxide phase deposited on a high-surface area oxide support such as SiO_2 to which we will limit this discussion. Although this catalyst is widely used questions remain about the structure of the active species and the reaction mechanisms involved. These catalysts have been studied[2] with Raman, infra-red, UV–vis and EXAFS spectroscopy; the Raman and infra-red spectroscopy are of particular interest here as these can be used under typical reaction conditions, thus offering the prospect of detecting the changes of molecular structures as the reaction takes place.

---

[2] As part of an active collaboration with the Inorganic Chemistry and Catalysis Group (O.L.J. Gijzeman, and B.M. Weckhuysen) at the Debye Institute, Utrecht.
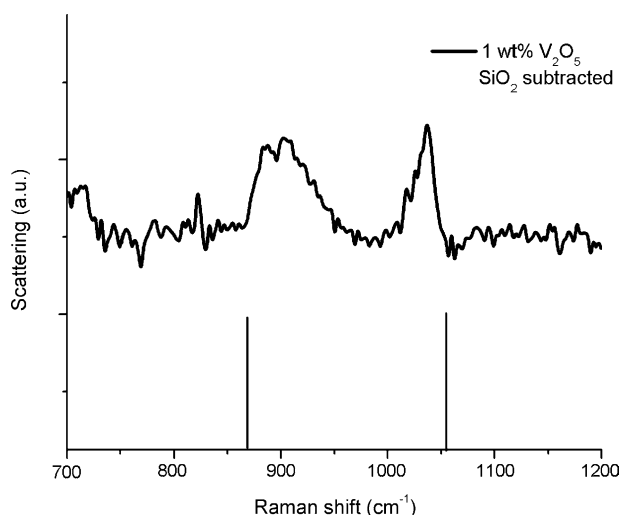
Fig. 6. Experimental Raman spectrum and the theoretical frequencies for the dominant lines for the adsorbed Vanadium species.
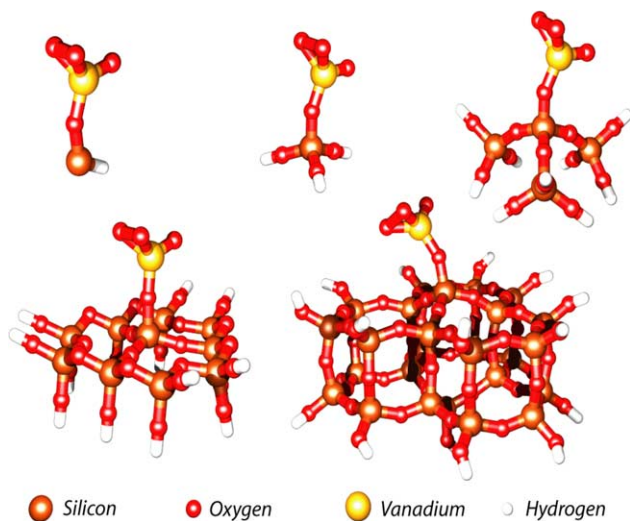

Fig. 7. The vanadium oxide species on silica support models of increasing size, labelled from the top left to the bottom right as tiny, small, medium, large, and very large.

As a reference point, the spectrum of vanadium oxide supported on silica at low loadings but with the silica spectrum [34] subtracted, is given in the upper part of Fig. 6. The spectrum is characterised by a broad line at 900 cm^{-1} and

a sharp line at 1040 cm^{-1}. The lower part gives the calculated frequencies for the dominant Raman lines.

Conventionally, it has been assumed that adsorbed VOx species adopted a pyramidal structure in which each V atom is coordinated by three surface oxygens, with a fourth oxygen present as an axially oriented V=O group. Recently, it has been shown that this spectrum, as well as the temperature dependent shifts of the observed lines, can be explained by assuming an alternative 'umbrella' model for the active species [34]. In this model, a VO$_3$ species is bound to the silica surface through a single oxygen atom. As shown in Fig. 7, it is proposed that the VO$_3$ moiety comprises one doubly bonded oxygen atom and a loosely bound oxygen molecule.

Fig. 7 also illustrates how a variety of models for the silica support have been investigated. In each case, we have optimised the geometry at the DFT level (B3LYP/LANL2DZ) and calculated analytical infrared intensities. The value of large scale computing facilities and the parallel implementation of the DFT analytical Hessian capability is that the larger clusters serve as a reference to calibrate the accuracy of the smaller models. The main bond lengths, as well as the O–O stretch and V=O stretch frequencies are collected in Table 1. These results indicate that the smaller models reproduce well the structural and vibrational properties of the larger clusters. The only aspect that varies considerably is the Si–O–V angle, which is due to the fact that that angle is quite floppy.

Based on this work, we are using the Medium (T4) cluster to study possible mechanisms for the partial oxidation of methanol to formaldehyde, and propane to propene. The availability of analytic Hessians is crucial to the location of transition states and for comparison with results of in-situ Raman and IR spectroscopy [35].

4. Summary

Hopefully the discussion above has emphasised that there are many aspects to the migration of the computational science workload onto large-scale parallel computers.

Simply transferring the current computational chemistry workload to such machines will not work. The scalability possible for the traditional methodologies and the current problem sizes will result in a job mix dominated by 32 processor jobs, and such a job mix, while scientifically productive, is more cost-effectively satisfied by other types

Table 1
Bond lengths (Å) and vibrational frequencies (cm^{-1}) obtained for the vanadium oxide species on support models of different sizes as computed and as obtained from EXAFS or Raman spectroscopy

	Tiny T1	Small T1	Medium T4	Large T10	Very large T20	Exp.
R(V–O$_s$)	1.762	1.750	1.739	1.744	1.757	1.77 (EXAFS)
R(V–O)	1.805	1.806	1.807	1.804	1.797	
R(V=O)	1.592	1.590	1.592	1.590	1.588	1.58 (EXAFS)
R(O–O)	1.540	1.538	1.540	1.541	1.544	
A(Si–O–V)	178.3	159.9	167.6	171.9	153.9	
hν (O–O)	867	866	869	867	866	915 (Raman)
hν (V=O)	1047	1055	1053	1057	1055	1030 (Raman)

of resources. If the large scale machines of today, and the PetaFlop systems of tomorrow, are to justify their existence there needs to be a change in both the scale of the problems tackled as well as the algorithms and methods chosen.

In order to maximise the efficiency of the calculation, the intrinsic ability of the faster machines to reduce time to solution can be coupled with increases in the system size, selection of state-of-the-art computational methods, and exploitation of any intrinsic parallelism. The balance of these different aspects will clearly depend on the scientific problem—it would be folly to scale up to very large systems simply to improve scalability. Taken together, however, there is the real prospect to make efficient use of thousands of processors and the challenge of the computational chemistry community is to couple ambitious problems with algorithmic innovation.

Acknowledgements

References

[1] The HPCx service, www.hpcx.ac.uk
[2] S.L. Graham, M. Snir, C.A. Patterson, N.R.C. Committee on the Future of Supercomputing, National Academies Press, Washington, DC, 2004 (ISBN 0-309-09502-6).
[3] UK/US N + N Workshop, Beyond Teraflop/s towards Petaflop/s, London, 1st–2nd October, 2002.
[4] R. Catlow, P. Burke, J. Goodfellow, D. Tildesley, M. Wilson, Philos. Trans. R. Soc. Lond. Ser. 360 (2002) 1073.
[5] V.R. Saunders, M.F. Guest, Comput. Phys. Commun. 26 (1982) 389–395.
[6] M.F. Guest, S. Wilson, in: P. Lykos, I. Shavitt, (Eds.), Supercomputers in Chemistry, 1981, p. 1.
[7] The Top 500 List, www.top500.org
[8] P. Grayson, E. Tajkhorshid, K. Schulten, Biophys. J. 85 (2003) 36–48.
[9] M. Law, Comput. Phys. Commun. 102 (1997) 252–268.
[10] M.F. Guest, R.J. Harrison, J.H. van Lenthe, L.C.H. van Corler, Theor. Chim. Acta 71 (1987) 117–148.
[11] G. Amdahl, Validity of the single processor approach to achieving large-scale computing capabilities, Proceedings of AFIPS, 1967, pp. 483–485.
[12] J. Nieplocha, R.J. Harrison, R.J. Littlefield, J. Supercomput. 10 (1996) 197–220.
[13] J. Nieplocha, R.J. Harrison, R.J. Littlefield, Global Arrays; a portable shared memory programming model for distributed memory computers, Proceedings of Supercomputing'94, 1994, pp. 340–349.
[14] G.I. Fann, R.J. Littlefield, Parallel inverse iteration with reorthogonalisation, Proceedings of the Sixth SIAM Conference on Parallel Processing for Scientific Computing, SIAM, 1993, pp. 409–413.
[15] Y. Alexeev, R.A. Kendall, M.S. Gordon, Comput. Phys. Commun. 143 (2002) 69–82.
[16] Y. Alexeev, M. Schmidt, T. Windus, M.S. Gordon, R.A. Kendall, Proceedings of Cluster 2002, IEEE International Conference on Cluster Computing, IEEE Computer Society, 2002, pp. 135–141.
[17] M.F. Guest, J.M.H. Thomas, P. Sherwood, I.J. Bush, H.J.J. van Dam, J. van Lenthe, R.W.A. Havenith, J. Kendrick, Mol. Phys. 103 (2005) 719.
[18] E. Aprà, T.L. Windus, T.P. Straatsma, E.J. Bylaska, W. de Jong, S. Hirata, M. Valiev, M. Hackler, L. Pollack, K. Kowalski, R. Harrison, M. Dupuis, D.M.A. Smith, J. Nieplocha, V, T, M. Krishnan, A.A. Auer, E. Brown, G. Cisneros, G. Fann, H. Früchtl, J. Garza, K. Hirao, R. Kendall, J. Nichols, K. Tsemekhman, K. Wolinski, J. Anchell, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Deegan, K. Dyall, D. Elwood, E. Glendening, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, T. Nakajima, S. Niu, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. van Lenthe, A. Wong, Z. Zhang, NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.6, Pacific Northwest National Laboratory, Richland, Washington 99352-0999, 2004, USA.
[19] R.A. Kendall, E. Apra, D.E. Bernholdt, E.J. Bylaska, M. Dupuis, G.I. Fann, R.J. Harrison, J.L. Ju, J.A. Nichols, J. Nieplocha, T.P. Straatsma, T.L. Windus, A.T. Wong, Comput. Phys. Commun. 128 (2000) 260–283.
[20] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, J. Comput. Chem. 4 (1983) 187–217.
[21] H.L. Woodcock, M. Hodosek, P. Sherwood, Y.S. Lee, H.F. Schaefer, B.R. Brooks, Theor. Chem. Acc. 109 (2003) 140–148.
[22] D. Das, K.P. Eurenius, E.M. Billings, P. Sherwood, D.C. Chatfield, M. Hodosek, B.R. Brooks, J. Chem. Phys. 117 (2002) 10534–10547.
[23] K.P. Eurenius, D.C. Chatfield, B.R. Brooks, M. Hodosek, Int. J. Quantum Chem. 60 (1996) 1189–1200.
[24] P. Sherwood, A.H. de Vries, M.F. Guest, G. Schreckenbach, C.R.A. Catlow, S.A. French, A.A. Sokol, S.T. Bromley, W. Thiel, A.J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S.C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer, C. Lennartz, Theochem-J. Mol. Struct. 632 (2003) 1–28.
[25] ChemShell User Manual, available online at http://www.cse.clrc.ac.uk/qcg/chemshell
[26] B.G. Dick, A.W. Overhauser, Phys. Rev. 112 (1958) 90.
[27] J.D. Gale, J. Chem. Soc.-Faraday Trans. 93 (1997) 629–637.
[28] Y. Zhao, B.J. Lynch, D.G. Truhlar, J. Phys. Chem. A 108 (2004) 2715–2719.

[29] A.A. Sokol, S.T. Bromley, S.A. French, C.R.A. Catlow, P. Sherwood, Int. J. Quantum Chem. 99 (2004) 695–712.

[30] Y. Iba, Int. J. Mod. Phys. C 12 (2001) 623–656.

[31] M.R. Wilson, J.M. Ilnytskyi, Computer Simulations of liquid crystals and polymers in: P. Pasini, C. Zannoni, S. ZŬmer (Eds.),, Kluwer, Dordecht, 2004, pp. 335–356.

[32] R. Elber, M. Karplus, Chem. Phys. Lett. 139 (1987) 375.

[33] G. Henkelman, B.P. Uberuaga, H. Jonsson, J. Chem. Phys. 113 (2000) 9901–9904.

[34] O.L.J. Gijzeman, J.N.J. van Lingen, J.H. van Lenthe, S.J. Tinnemans, D.E. Keller, B.M. Weckhuysen, Chem. Phys. Lett. 397 (2004) 277–281.

[35] L.J. Burcham, G.T. Deo, X.T. Gao, I.E. Wachs, Top. Catal. 11 (2000) 85–100.