

CCP5 Workshop: Algorithms and Architectures for Molecular Simulation

Tower Seminar Room
Daresbury Laboratory
2 December 2010

Programme

9:30 Arrival and Refreshments

10:00 Welcome

10:15 Godehard Sutmann, Jülich Supercomputer Centre: *"Fast Methods for Electrostatic Problems in Particle Simulations"*

11:00 Mike Ashworth, Daresbury Laboratory: *"Towards Exascale Resources for Molecular Simulation – Developments in Hardware and Software"*

11:45 Peter Coveney, University College London: *"Real Science at the Petascale"*

12:30 **Buffet Lunch**

13:30 Mike Payne, University of Cambridge: *"Novel Methods for Exploiting Quantum Mechanics in Atomistic Simulations"*

14:15 Martyn Guest, University of Cardiff: *"Application Performance in Molecular Simulation – Perspective, Identity and Challenges"*

15:00 Refreshments

15:15 Ian Bush, Numerical Algorithms Group Ltd: *"So What's Going On On HECToR?"*

15:45 Phil Hasnip, University of York: *"Careful planning and sneaky tricks: the parallelisation of CASTEP"*

16:30 **Close**

Fast Methods for Electrostatic Problems in Particle Simulations

Godehard Sutmann

Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)
Research Centre Jülich, Germany
g.sutmann@fz-juelich.de

The simulation of classical particle systems by means of e.g. molecular dynamics or Monte Carlo techniques, requires the evaluation of interparticle potentials or mutual forces between particles in order to explore the phase- or configuration space of the system. The interactions may be classified into short-range and long-range contributions. While short-range potentials may be evaluated very efficiently by means of neighbor list techniques, which reduce the computational complexity to $O(N)$, the long range interactions must be evaluated over all particle pair contributions in the system. This increases the complexity to $O(N^2)$, limiting very often the tractable system size to a few thousand particles.

A typical long range contribution for molecular systems is the electrostatic interaction between particles carrying partial charges - a situation which appears e.g. in polar liquids, biological molecules or ionic liquids. Progress has been made to reduce the computational work by methods, which decompose the system hierarchically and solving Poisson's equation very efficiently. An overview will be given for different methods, which were further developed or optimised at JSC and which are based on multipole-, multigrid- and tree-methods, which reduce the complexity to $O(N)$ or $O(N \log N)$. Progress on parallelization for massively parallel architectures will be reported and an outlook is given for the inclusion of the presented methods into existing complex simulation codes.

Towards Exascale Resources for Molecular Simulation – Developments in Hardware and Software

Mike Ashworth

Associate Director: High Performance Computing
Computational Science & Engineering Department
STFC Daresbury Laboratory
Daresbury Science and Innovation Campus
Warrington WA4 4AD, UK
mike.ashworth@stfc.ac.uk

High Performance Computing is currently at a turning point. The resources available for molecular simulation are heading towards the multi-Petascale and Exascale systems are forecast within the next ten years. Achieving these performance levels within reasonable cost and power budgets will require revolutionary changes to hardware and present us with substantial software challenges. Future systems are likely to feature large increases in the number of cores per chip and per node, massive increases in the number of threads of execution in the system (towards ten million) and reductions in memory per core and especially in memory bandwidth per core.

We have identified major technical challenges which need to be addressed in order to bring about a step change in current molecular simulation capabilities for grand challenge applications:

- Scalable Algorithms
- Exploitation of Multi-core
- Management of Memory Hierarchy
- Input/Output
- Fault Tolerance
- Data Management
- Visualisation

We also present STFC's plans for the Hartree Centre – a new world-leading centre for computational science and high performance computing in the UK to be located at STFC's Daresbury Science and Innovation Campus. The Hartree Centre will bring together a critical mass of academic, industrial and public sector researchers to address grand challenge problems of economic and societal importance and support a step change in STFC's knowledge exchange and economic impact activities. We will focus on the scientific challenges of developing the next generation of multi-scale, multi-physics simulation codes, the technical challenges of exploiting future multi-Petascale and Exascale systems while at the same time demonstrating socio-economic impact.

The Computational Science & Engineering Department (<http://www.cse.clrc.ac.uk/>) led from STFC's Daresbury Laboratory is a UK focus for Computational Science and Engineering, ensuring that researchers benefit from the best computational methods, and supporting them through research and collaboration, theory and software development, facilities and training. We also partner in operating and supporting the UK's national flagship HPC resource: HECToR a 360 Tflop/s Cray XT service (<http://www.hector.ac.uk/>).

Real Science at the Petascale

Peter Coveney,

Centre for Computational Science,
Chemistry Department,
University College of London,
20 Gordon Street,
WC1H 0AJ,
London,
United Kingdom
p.v.coveney@ucl.ac.uk

We present computational science research that uses petascale infrastructures to produce scientific results at an unprecedented scale and resolution. The applications cover many scientific domains, which include nano-materials science, fluid dynamics, computational biology, computational medicine and cosmology. We have run these simulations efficiently on large scale supercomputers around the world, and are now also working on multiscale simulations which contain multiple solvers to a distributed petascale infrastructure. Distributed multiscale simulations will produce science not attainable by single solvers, but they do require a high quality of wide area networking and a tightly integrated computing infrastructure, requirements which are frequently overlooked but will be show-stoppers in the era of data-driven, data-intensive computing.

Novel Methods for Exploiting Quantum Mechanics in Atomistic Simulations

Mike Payne,
TCM Group,
Cavendish Laboratory,
Cambridge.
mcp1@cam.ac.uk

In this talk I shall describe several novel quantum mechanical based techniques for atomistic simulations. These are: (i) the linear scaling density functional theory code ONETEP¹ which can be used to perform quantum mechanical calculations on systems containing many thousands of atoms on modest numbers of compute cores; (ii) the Learn-on-the-Fly² (LOTF) hybrid (or QM/MM) simulation technique in which the atoms in some regions are treated quantum mechanically while the remainder of the atoms are described using simple empirical interatomic potentials. A particular feature of the LOTF technique is that the regions in which the atoms are treated quantum mechanically can change with time; and (iii) Gaussian approximation potentials³ which do not assume a particular form for the interatomic potential but instead use a database of configurations to generate accurate atomic energies and forces. This approach can provide quantum mechanical accuracy at a fraction of the computational cost.

I shall also briefly review the field of atomistic simulations and identify the 'strengths, weaknesses, opportunities and threats' in the field and suggest that the best route forward to exploit the strengths and opportunities and overcome the weaknesses and threats is an 'Open Innovation'⁴ model for atomistic simulation software.

1. <http://www2.tcm.phy.cam.ac.uk/onetep/>
2. G. Csányi, T. Albaret, M.C. Payne and A. De Vita, *Phys.Rev.Lett.* **93** 175503 (2004)
3. A.P. Bartók, M.C. Payne, R. Kondor and G. Csányi, *Phys.Rev.Lett.* **104** 136403 (2010)
4. For an introduction see: http://en.wikipedia.org/wiki/Open_innovation.

Application Performance in Molecular Simulation – Perspective, Identity and Challenges

Martyn F. Guest
Advanced Research Computing @ Cardiff (ARCCA)
Room T.17-18 Redwood Building
King Edward VII Avenue
Cardiff CF10 3NB
Tel: +44 (0)2920 879319
email: GuestMF@cardiff.ac.uk
www.cf.ac.uk/arcca

The spectrum of major scientific challenges that can now be addressed by molecular simulation methods drives the need for the development of new or improved capabilities. Notwithstanding these capability developments, the question remains as to the optimum platforms on which to exploit these developments. It is often argued that it will be essential for molecular simulation codes to effectively utilize petascale computing platforms in order for researchers to tackle large problem sizes and to get their results quicker. There is no doubt that this presents a major challenge for the discipline.

The current generation of molecular simulation software has been shown at best to scale to thousands of processors (teraflops). However, efficiently utilizing the hardware advances and tens to hundreds of thousands of processors anticipated in the next five years will demand the development of new code architectures. Bringing these codes to the petascale platforms will be a major effort that will require a community of computational chemists, mathematicians and computer scientists to join forces. Both new capability development and increasing scalability to petaflop systems and beyond are major efforts, which combined exceed the resources and manpower of any one software development group. This provides a strong driving force to open-source to enable an extensive collaborative capability and petascale development environment which will greatly increase the developer base of the software suite.

Just as the drive to exploit petascale platforms remains an elusive goal, so commodity-based clusters have over the past two decades provided an established, cost effective solution to the provision of high performance computing for molecular simulation. Detailed cluster architecture - node and interconnect – determines the ability of such clusters to address the differing demands from the fields of capability (high-end) and capacity (mid-range) computing. We consider the performance over the past decade of a variety of commodity-based cluster systems (CS1-CS68) with that on high-end, proprietary hardware from SGI, HP, IBM and Cray. Based on simple metrics, we consider the performance of variety of simulation codes, including NWChem and GAMESS-UK, CPMD, CASTEP, DLPOLY and AMBER. We identify the challenges in scaling to large processor counts, the associated bottlenecks and provide a guide to the optimum resource that might be deployed.

Finally we question the disciplinary identity of molecular simulation as its methods, techniques and tools are increasingly exploited by the domains of materials and nanotechnology, biology and biotechnology and the neighbour sciences, especially in physics and the life sciences.

So What's Going On On HECToR?

Ian J. Bush

The Numerical Algorithms Group Ltd,
Wilkinson House,
Jordan Hill Road,
Oxford OX2 8DR,
United Kingdom.
ian.bush@nag.co.uk

Over the life time of the phase2a component of the HECToR service almost all the most used codes are written in Fortran, and almost all use MPI to express parallelism in the algorithms that they employ. So over the last decade or so it appears little has changed in how high end HPC architectures are exploited. However a closer examination of the codes shows that the situation has and is changing, driven by both a need to exploit a larger degree of parallelism and the need to exploit the hierarchical nature of modern machines. I shall discuss what new approaches are being taken and how these approaches result in more scalable code, and also what challenges these approaches give the modern HPC software developer, and I shall illustrate these with results from HECToR.

Careful Planning and Sneaky Tricks: the Parallelisation of CASTEP

Phil Hasnip

Department of Physics
University of York
York YO10 5DD
pjh503@york.ac.uk

CASTEP is a state-of-the-art materials modelling program which employs quantum mechanics to predict a wide range of chemical, mechanical and spectroscopic properties. In this talk I shall discuss the design, implementation and performance of CASTEP's hierarchical parallelisation strategy. I shall focus particularly on the challenges presented by modern HPC architectures, and recent improvements to CASTEP that aim to alleviate some of the problems.