Science carried out as part of the NW-GRID project using the eMinerals infrastructure

Jens M. H. Thomas, Rik Tyer, Rob J. Allan, Jamie M. Rintelman and Paul Sherwood
STFC Daresbury Laboratory, Warrington WA4 4AD, UK

Martin T. Dove, Kat F. Austen, Andrew M. Walker and Richard P. Bruin
Department of Earth Sciences, University of Cambridge, Downing Street
Cambridge CB2 3EQ, UK

Leon Petit
Department of Physics and Astronomy, University of Aarhus,
DK-8000 Aarhus, Denmark

Marcus C. Durrant
School of Applied Sciences, Northumbria University
Ellison Building, Newcastle Upon Tyne NE1 8ST, UK

Corresponding author: j.m.h.thomas@dl.ac.uk

Abstract

The NW-GRID is funded by the NWDA as a regional resource for the North West of England and initially involves a core of compute clusters at Daresbury Laboratory and the Universities of Lancaster, Liverpool and Manchester. The middleware infrastructure developed as part of the eMinerals project has been ported to the NW-GRID, providing a set of tools that seamlessly integrate computation, data and metadata management. This paper describes how several different science projects have exploited this infrastructure to undertake some novel science that would have been difficult, if not impossible, otherwise.

1 Introduction

The NW-GRID project [1] is a collaboration funded by the Northwest Development Agency (NWDA) between STFC Daresbury Laboratory and the Universities of Lancaster, Liverpool and Manchester, to establish a computational Grid comprising large-scale commodity computing systems coupled by a high-speed private fibre network between clusters. The project aims to be a world-class activity in the deployment and exploitation of Grid middleware technologies and to demonstrate the capabilities of the Grid in leading-edge computational science and engineering applications.

In its current incarnation, the NW-GRID consists of 192 dual-core Opteron processors at Daresbury, 96 at Lancaster, 88 at Liverpool and 50 at Manchester, giving a total of 852 processor cores. This initial setup will be expanded over the lifetime of the project.

One of the first middleware technologies to be exploited on the NW-GRID has been the eMinerals infrastructure (particularly RMCS described below), developed as part of the NERC-funded eMinerals project: Environment from the Molecular Level [2]. This project is a collaboration between the University of Cambridge (Earth Sciences, Chemistry), University College London (Earth Sciences, Chemistry, Computer Science), The Royal Institution, University of Bath (Chemistry), University of Reading (Computer Science), Birkbeck College (Chemistry) and Daresbury Computational Science and Engineering Department, and the Grid Technology Group of the e-Science Centre. It aims to develop UK e-Science/Grid capabilities for molecular simulations of environmental issues such as nuclear waste containment and transport of pollutants.

2 RMCS

RMCS (Remote My Condor Submit) [3] is the Grid computing framework developed by STFC in conjunction with other partners within the eMinerals project. The primary driver for the project was to study key environmental problems using simulations spanning many different scales of physical dimension and complexity, which entailed using many different simulation codes. Since a plethora of different codes were relevant to the project, a reasonably generic Grid computing client framework was realised. Furthermore, since the key output of the project was science, the emphasis on the technology has very much been on usability and pragmatism; responding to the requirements of users and only using cutting edge technology and computer science techniques when there was a compelling need.

This paper shows how RMCS has now been used in several different applications.

2.1 Functionality

Essentially RMCS takes a data-centric approach to Grid computing. It combines compute, data and metadata management functionality, whilst trying to abstract the users from the details as much as possible. It has also been designed to scale as efficiently as possible within the constraints placed by underlying technologies.

The emphasis on data/ metadata management in conjunction with scaling is due to the primary use case of Grid within eMinerals being parameter sweeps or ensemble runs. In this scenario, tracking dozens or hundreds of jobs and the resulting data is a non-trivial book keeping task. Parameter sweeps are now being carried out more routinely in other science areas as there is a growing realisation that this is an appropriate and effective use of Grid resources.

A Grid simulation run via RMCS system can be viewed as having the following stages:

1. The job is ”meta-scheduled” to an appropriate Grid resource. The aim here being to maximise the throughput of calculations by minimising the time spent queuing for resources. Again this facilitates parameter sweep calculations where the emphasis is primarily on job throughput. The scheduling algorithm is currently very simple, consisting of a list of all available queues on the available resources, ordered with the most recently accessed queue at the bottom. A job will be scheduled to the queue at the top of the list if it has available processors, with the queue then being placed at the bottom of the list. Although simple, this algorithm has been found to be very successful in practice.

2. Once a resource has been chosen, a temporary working directory is created within the user's home space. The input files and executables for this calculation are then downloaded from the Storage Resource Broker (SRB).

3. The job is scheduled within the machine's batch queue system.

4. After the simulation has completed, the output files are uploaded back to the SRB along with metadata describing the calculation (creating a complete audit trail for the simulation). There is currently no support for checkpointing within RMCS, so if a job has been killed due to running out of time on a queue, it can only be restarted if the individual code supports checkpointing.

5. If this simulation code has XML output, then key data can be extracted from the output file itself and stored as metadata within the database. This enables the user to view the most relevant results of a calculation without having to retrieve the file from SRB and significantly facilitates the handling of multiple output files from related calculations (e.g. the calculation of magnetic susceptibility as a function of temperature), as these parameters from all related calculations can be trivially obtained by the scientist from the metadata.

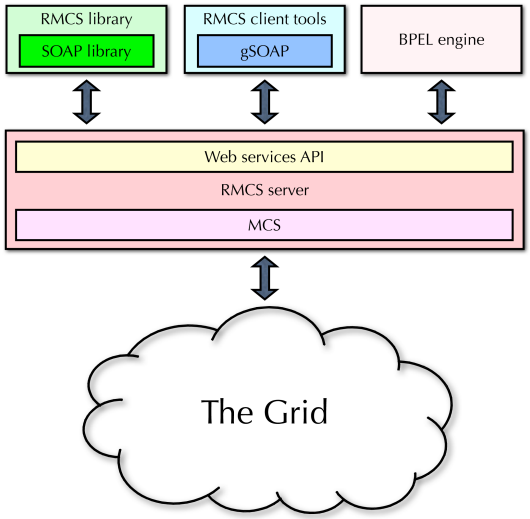

Figure 1: Architecture of RMCS.

RMCS can be viewed as a three-tier application, with the tiers being:

i. Client
ii. Application Server
iii. Grid

The use of a three-tier model introduces an application server, which allows the perennial Grid problems of client dependencies and extensive firewall configuration to be avoided. This is a common architecture for Web applications. Within RMCS, communication between tiers (i) and (ii) is via Web services, which is both firewall friendly and allows for extremely thin client interfaces. Since the client layer is mostly composed of Web service invocation code, it is much easier to integrate the functionality into existing applications, as client libraries can be written for any language that has appropriate web service client libraries. We refer to this as ”Lightweight Grid Computing”.

The middle tier of RMCS relies on both the command-line Globus tools and CondorG. Small 'fork' jobs are run using the Globus tools to query the status of the queue where the job is running, but all other tasks are submitted to the resource using CondorG.

The bottom tier of the RMCS architecture is ”the Grid”. RMCS currently relies on the following pieces of middleware:

- Globus Toolkit 2 (or equivalent within VDT or GT4);
- Access to MyProxy credential repository;
- Storage Resource Broker client;
- RCommands (if require metadata functionality);
- AgentX (if require automatic data extraction during post-processing).

The RCommands [3] are used within the post-processing stage of RMCS in order to automatically populate metadata with details of the calculation. The RCommands themselves are shell client tools and associated server side web services that provide metadata related functionality. In the context of RMCS, the RCom-

mands are used to create entries in the metadata that, at a minimum, associate the calculation with a particular piece of work and add the URL of the data files in the SRB. In addition, audit style metadata can also be stored, e.g. where the calculation ran, how long it took etc.

AgentX [4] is a powerful semantic library that provides a common logical view on data regardless of how the data is syntactically represented. This is achieved by users making queries with reference to an ontology, which are then mapped to particular data formats using descriptions of the data written in Resouce Description Framework (RDF). Using AgentX, RMCS is able to harvest key data from the simulation code output itself, assuming that it writes XML and an appropriate mapping document exists.

Using the combination of the RCommands and AgentX within RMCS, it is straightforward for the user to capture both audit-style metadata and key pieces of data from the simulation. This is especially useful within in the context of a large combinatorial parameter sweep, as the metadata can now become the primary interface for data discovery

2.2 User interface

Within the context of eMinerals, the user interface to RMCS is a set of shell commands and higher level helper scripts, which automate bulk job submission and the preparation of the input files within the SRB. However, as mentioned above, these shell tools merely invoke Web services, so this functionality can be readily integrated into existing GUIs or expert systems.

2.3 Uses of RMCS

Although RMCS has been designed within the eMinerals project and is an integral part of it, it is proving a useful tool to any scientist who wishes to manage a large project that requires numerous calculations to be run on a range of different resources and the large quantities of data generated to be processed automatically.

The remainder of this paper describes science that has been undertaken on the NW-GRID using RMCS, both within the eMinerals project and by other users of the NW-GRID.

3 *Ab Initio* Mineral Science

Andrew M. Walker.

Increasingly, methods borrowed from computational chemistry and condensed matter physics are used to tackle atomic scale problems from the mineral sciences. A particularly widely used approach is Density Functional Theory (DFT) with planewaves and pseudopotentials implemented in codes such as CASTEP[1] and VASP[2].

Bibliometrically, such research is heavily skewed towards a small number of mineral structures. This is partly because many important problems only require an understanding of relatively few simple systems but it is also due to the fact that, until very recently, only comparatively small systems could be tackled without the use of national supercomputing facilities. We investigated the performance of CASTEP on the structures of a wide range of minerals and ceramics using the NW-GRID as a (large) capacity resource. These materials were chosen to represent a range of difficult to model structures which had been afforded limited or no studies using the planewaves and pseudopotentials approach. Systems studied included andalusite, cordierite, diopside, enstatite, $Gd_2Ti_2O_7$, kyanite, scheelie, sillimanite, titanite, zircon and ZrW_2O_8.

This problem does not strictly require Grid computing, but does probe the sort of problem that distributed collections of tightly coupled clusters can now be used to solve. Each calculation required many cores to be used in parallel (most systems were of a size to make the use of 32 or 64 MPI processes efficient) and the presence of high performance Score interconnects on the NW-GRID nodes made this possible. As an aside the credential delegation offered by the gsi-enabled version of ssh certainly makes the use of more then one cluster much more straightforward to manage then the use of password protected accounts or even ssh keys.

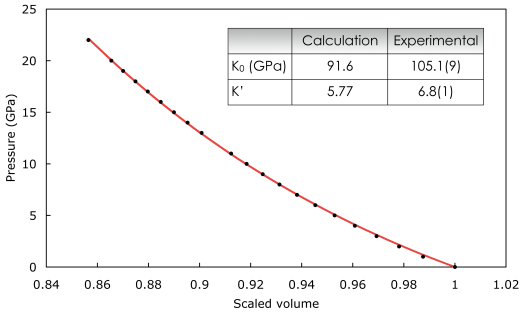

Figure 2: Equation of State

One of the first outcomes of this work is the calculation of compressibility of diopside ($CaMgSi_2O_6$) between 0 and 22 GPa, see Figure 3. To the best of our knowledge the first time this has been attempted using the generalized gradient approximation in density functional theory. The calculated structure and isothermal equation of state is in good agreement with experiment. Fitting the results to the third-order Birch-Murnaghan equation of state (Figure 2.) yields a bulk modulus (K_0) of 91.6 GPa and its derivative with respect to pressure (K') of 5.77 (experimental data from [5] give values of 105.1 and 6.8, respectively). Analysis of the variation of bond lengths with pressure shows that Si-O bonds are much less compressible then either Mg-O or Ca-O bonds and that one particular Ca-O bond is much softer then all others.

[1]CASTEP application code: `http://www.tcm.phy.cam.ac.uk/castep/`
[2]VASP application code: `http://cms.mpi.univie.ac.at/vasp/`

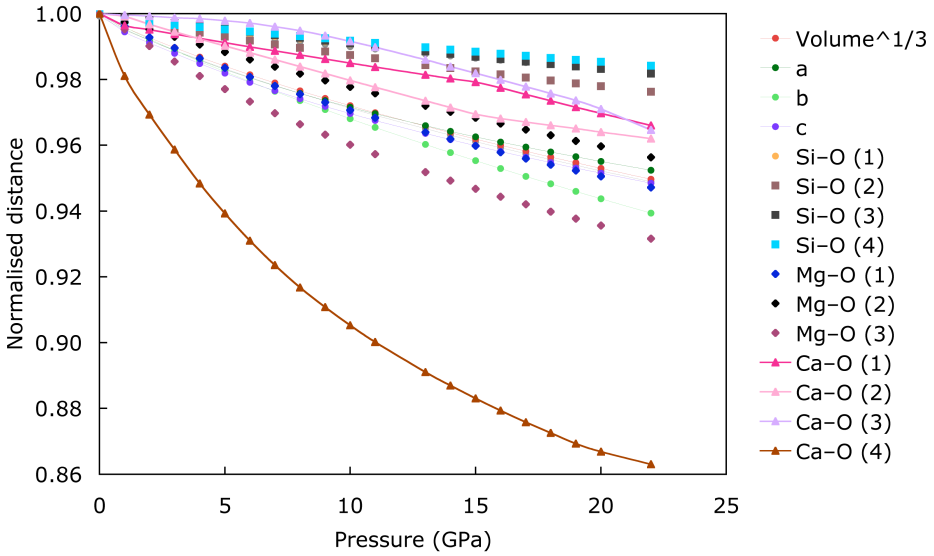

Figure 3: Compressibility of diopside ($CaMgSi_2O_6$) between 0 and 22 GPa.

4 Explaining the Equation of State of Silica Glass

Andrew M. Walker.

Experiments on silica glass (amorphous SiO_2) have revealed a number of unusual properties including negative thermal expansion and an increase in compressibility with increasing pressure (most materials increase in volume on increasing temperature and become harder to compress on increasing pressure). By performing atomic scale simulations of silica glass using molecular dynamics with parameterised potentials at a large number points in pressure (P) – temperature (T) space, we aim to explain the origin of these properties. In previous calculations [6, 7, 8] we have demonstrated that the compressibility increase, leading to a maximum at about 2 GPa, is due to the structure of the glass (specifically the structure being composed of ridged tetrahedra that are free to flex at the shared corners, Figure 4) leads to a wide range of densities where further flexing can be accommodated by flexing at the linkages without distorting tetrahedra. The previous simulations were performed at low T (50 K, giving just enough energy to probe the vibrations of the structure) and over a wide range of P (between ±5 GPa). The key aim of the current work is to probe the effect of T on the behaviour of the model, to explore the change in position of the compressibility maximum on increasing T, and to see how the mechanisms leading to the compressibility maximum and negative thermal expansion interact with each other.

Key to the study is the ability to run many simulations under slightly differing P, T conditions (we need many state points because we are interested in the derivatives of volume with respect to P and T, and for materials that are behaving in unusual ways we do not have the possibility of fitting a well behaved function to a small number of well spaced points). Along with the need to run many individual calculations (a task ideally suited to Grid resources) comes the problem of managing the torrent of output produced. Fortunately we have developed methods to do just this with the eMinerals mini-Grid and the middleware deployed on NW-GRID was compatible with our existing tools. In effect we were able to plug the NW-GRID machines into our existing compute and data Grid and massively boost the compute power of our resources without significant modification to the job submission and data management strategy.

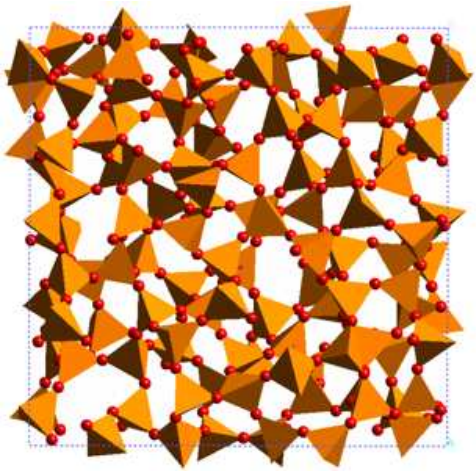

Figure 4: Snapshot of the atomic configuration used to simulate the behavior of silica glass as a function of pressure and temperature. Red spheres represent oxygen atoms which bridge between relatively rigid orange silica tetrahedra.

5 Investigation of the Torsional Behaviour of 2,2'-Dichloro Biphenyl

Kat Austen.

Poly-chlorinated biphenyl molecules are toxic, persistent, anthropogenic chemicals that cause serious environmental problems. In order to remediate environments where there is a high risk of poly-chlorinated biphenyl pollution, the molecules' behaviour in solution and their adsorption onto mineral surfaces must be understood. They are difficult to deal with experimentally, but are in need of research focus due to their high risk factor. As such, molecular simulations of the molecules are invaluable in providing information that is extremely important to remediation of the polluted environments. The structure of the biphenyl molecule is shown in Figure 5. Note the central C–C bond, around which the phenyl rings can freely rotate. One method used to characterise the behaviour of these molecules in solution is the investigation of the change in internal energy upon rotation of the rings around the central bond, both in a vacuum and in solution. The molecule chosen for this study is the 2,2'-dichloro biphenyl molecule, which has Cl substituted for H on the 2 and 2' site.

Two levels of theory have been used to study the rotation in vacuum, in order to ascertain if the torsional behaviour of the molecule is correctly described by the (computationally cheaper) DFT level of theory used in other parts of the study [9]. The other method is the more computationally expensive 2nd order Möller-Plesset (MP2) level of theory, which better describes nuances of electronic interactions.

The energy dependence on torsion angle was investigated by constraining the torsion angle, lying between C atoms 2-1-1'-2' in Fig 5, which was facilitated by the use of the zmatrix format in all input files. The GAMESS-UK code was used for the MP2 calculations, and the SIESTA code was used for the DFT calculations [10]. Both of these codes use atomic centred basis sets. For the SIESTA calculations three optimised DZP basis sets were used, whereas for the GAMESS-UK calculations two different basis sets were used for comparison: 6-311G* and CC-PVDZ.

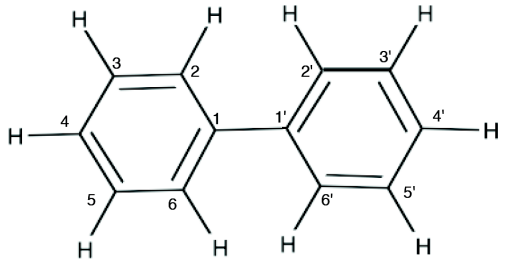

Figure 5: Biphenyl molecule. Each H atom can be substituted for a Cl atom, giving rise to 209 distinct molecules.

For each basis set, 37 calculations were performed, each at a different value of torsion angle, taken from 0° to 180° in 5° increments. The atoms involved in the torsion angle specification were held fixed, and all other atomic positions were allowed to relax. Each calculation was run in parallel on NW-GRID machines, using 16 processors each, allowing for a speedy aggregation of results, and avoiding the need for copious, cumbersome restarting.

The same workflow will not be used to investigate solvation effects on the barriers to rotation. The Direct Reaction Field (DRF) implicit solvation functionality in GAMESS-UK will be used to carry out another sweep of the torsion angle space. This method describes the effect of solvation on a solute, without requiring the atoms of the solvent to be explicitly described. The DRF calculations have the same Grid requirements and similar computational requirements to the other GAMESS-UK calculations.

Currently, molecular dynamics calculations are being carried out, using the SIESTA method, on the molecule surrounded by explicitly described water molecules. Such calculations allow comparison to be made between this and the implicit solvation model, in order to ascertain if explicit solvation is necessary in order to obtain a reliable solvation energy for the PCB molecule. Solvation energies are extremely important when trying to understand surface adsorption processes, as the strength of the molecule's interaction with the solvent impacts heavily on the likelihood of adsorption.

The molecular dynamics calculations are computationally demanding, and must be run on 32 processors. In 24 hours, the simulation will proceed by around 120 steps, which relates to 60 fs of simulation time. This type of simulation typically requires at least 3000 fs for equilibration and double that amount for production in order to collect enough data for reliable statistics. As such, the simulation progresses slowly but surely towards its conclusion.

The metascheduling aspect of the eMinerals workflow, coupled with the optimal performance of the code and system at 32 processors, means that the NW-GRID with eMinerals tools is a perfect platform for these calculations.

6 Adsorption of Polychlorinated Dibenzo-p-Dioxins (PCDDs) onto Mineral Surfaces

Kat Austen.

The adsorption of poly-chlorinated aromatic compounds onto clay and carbonate mineral surfaces has been carried out as part of the eMinerals project. There are methodological difficulties to studying the adsorption process,

as the interactions involved are weak, especially in the case of hydrophobic clay surfaces. Additionally, there are many available adsorption positions, all of which are fairly close in energy.

A traditional geometry optimisation of the molecule on the surface is unlikely to find the global minimum for adsorption. So, in order to find the correct adsorption position for each molecule, it was necessary to scan each molecule across each surface by carrying out a series of single point calculations at a sensible height from the surface ($\sim$4 Å), over 100 separate calculations for each molecule. The lowest energy position, ascertained from the scanning calculations, could then be taken and a full geometry optimisation carried out to ascertain the adsorption energy. Figure 6 shows the final relaxed geometry of a fully protonated PCDD molecule onto one of the clay surfaces studied.

The calculations comprised over 300 atoms, calculated at the DFT of theory using the SIESTA method, requiring parallel execution across 32 processors in order to run. In order to reduce exceptionally lengthy run-time for the full relaxations, 64 processors were used for these longer runs. Given the large number of parallel calculations required for this study, the work would have been prohibitively time-consuming without the utilities offered by Grid computing, and without large clusters such as the NW-GRID.

The workflow developed for the first of these calculations has been successfully applied to different mineral surfaces. Furthermore, the workflow can be used in the future to obtain a complete set of adsorption energies for additional environmentally occurring surfaces, which can be used in risk assessment and management of toxic chemical pollutants in the environment.

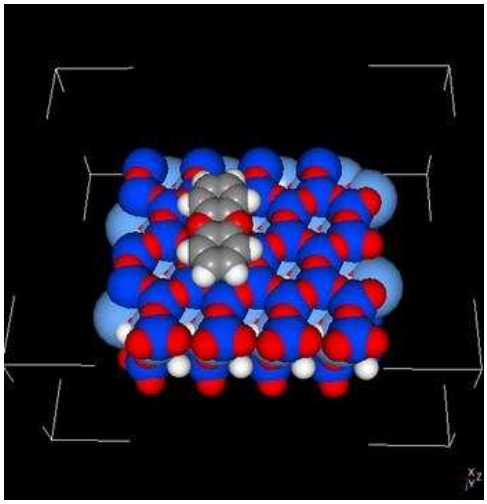

Figure 6: Adsorption of fully protonated poly-chlorinated dibenzo-p-dioxin on pyrophyllite (001) surface

7 Total Energy Investigations of Perovskite Structured Materials.

Leon Petit.

Oxide compounds based on the perovskite structure are exceedingly successful materials that find application as structural ceramics, given their capability to withstand mechanical wear and stress, and as electrical/ thermal ceramics given their excellent insulating properties. Even more important is their role in the growing field of functional materials, where the inherent electronic, magnetic, dielectric or optical properties of the material play an active role, and can be tuned to perform very specific tasks such as sensing, actuation and catalysis among others. This wealth of properties makes perovskite structured ceramics the hottest contenders for application as multifunctional materials, where two or more properties of the material can be exploited simultaneously in one single device.

Of chemical formula ABO_3, where A and B refer to almost any element in the periodic table, only few (for example $SrTiO_3$) are known to have the undistorted cubic structure. Depending on the relative sizes of the A and B ions a number of distorted structures are thus observed, one of which is depicted in Figure 7.

Of special interest to our study are the perovskites ABO_3, where A and B respectively represent a rare earth element (La, Ce, Pr, ..., Lu), and a transition metal element (Ti, V, ..., Cu). A famous representative example is $LaMnO_3$. Referring to Figure 7, the rare earth element is the black sphere situated at the center, the transition metal is represented by the yellow sphere, octahedrally coordinated by the O atoms (blue). The electronic structure of these compounds remains strongly debated, mainly due to the fact that the very successful local spin density approximation (LSD) to exchange and correlation fails to correctly describe strongly correlated electrons. Thus in $LaMnO_3$, due to the electron-electron interaction, the Mn-d electrons have a strong tendency towards remaining localized at each their site, and the band-picture description of itinerant d-electrons leads to unphysical results that are in disagreement with experiment. A number of electronic structure methods (LDA+U, DMFT) address this shortcoming by introducing the so-called Hubbard U parameter to take into account the strong correlations. The problem however with these methods is that they have little predictive capability, given their dependence on a parameter which often is derived from experiment. More specifically for the presently studied perovskite materials, in compounds such as $PrFeO_3$, on top of the U parameter for Fe, a different U parameter is needed to describe the strongly correlated f-electrons of the rare earth element Pr.

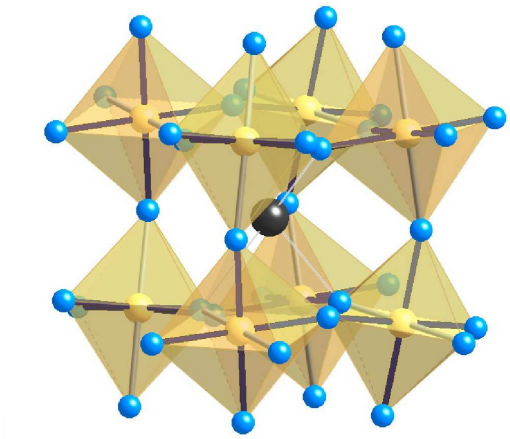

Figure 7: An example of a perovskite with the structure ABO_3: $LaMnO_3$

In the self-interaction corrected local spin-density approximation, the localized and delocalized f- and d-electrons are described on an equal footing, and the decision on whether a specific electron prefers to remain localized or take part in the bonding is determined purely on total energy considerations. The method is thus fully *ab initio* and does not require parameters. On the other hand, it implies that the total energy is minimized with respect to all possible localized/delocalized configurations of the f-d manifold. For $PrFeO_3$ for example, with Pr:[Xe]$4f^3 6s^2$, and Fe:[Ar]$3d^6 4s^2$, basically 3x6 localization/delocalization (valency) scenarios need to be considered, given that Pr can be treated with 3, 2, or 1 localized f-electrons, and Fe with 6, 5, ..,1 localized d-electrons. On top of that, for a given valency scenario with x localized d-, and y localized f-electrons, C_x^{10} x C_y^{14} combinations of orbital occupations need to be taken into consideration, in order to reflect the relative strength of crystal field splitting and exchange interaction, as well as possible gains in energy through orbital ordering. Though the resulting number of configurations can be considerably reduced, using the symmetries of the crystal and physical intuition, the remaining configuration space still poses a considerable computational challenge. Even more so given the fact that we simultaneously would like to investigate magnetic ordering (ferromagnetic versus antiferromagnetic).

A comprehensive scan of perovskite structured ABO_3 materials through the rare earth and transition metal compounds in order to establish the groundstate electronic and magnetic properties is thus a rather daunting task which nevertheless becomes possible given the possibilities of using the automation and database tools of the NW-GRID. Our preliminary tests on $LaCoO_3$ have shown that 100s of configuration runs can be efficiently prepared and executed with a straightforward presentation of the corresponding results that hugely simplifies the task of analyzing the underlying physics.

8 Quantum Directed Genetic Algorithm

Marcus Durrant and Jens Thomas.

The QDGA (Quantum-Directed Genetic Algorithm) project uses a genetic algorithm to try and determine an optimal catalyst for the conversion of nitrogen (N_2) to hydrazine (N_2H_4). Hydrazine is widely used in a range of chemical syntheses, including those for pharmaceuticals and fertilisers, and the prospect of synthesising it directly and cost-effectively from nitrogen is of great interest within a number of industries.

Within the QDGA process, a selection of potential catalysts (collectively a ”generation”) are created by randomly combining a number of metal centres and ligands in several pre-defined geometries, with an associated range of spin and charge states.

As an ”ideal” energy profile for the catalytic cycle is already known, the energy of a complex formed from the potential catalyst and the substrate at each point on the catalytic cycle can be calculated and compared with the energy of the ideal value. A scoring system is then used to select promising catalysts, and a new generation created by shuffling the characteristics of the selected complexes (e.g. metal centre, geometry etc.) between themselves.

This new generation is then submitted for calculation, and the process repeated until a catalyst that most closely reproduces the desired energy profile has been found.

As the QDGA algorithm requires 9 transition state calculations to be undertaken for each potential catalyst, sampling even a relatively small number of metal centres and ligands quickly generates a large number of calculations. DFT is used for the calculations, and as is this is a quite rigorous level of theory, the calculations have been undertaken with the GAMESS-UK *ab inito* molecular modelling code, with each calculation requiring several hours when run in parallel on 8 processors.

The computational demands of large numbers of medium-sized parallel calculations have proved particularly difficult to satisfy on local resources or on HPCx [11] (where some initial generations were processed), but the NW-GRID has proved ideal for providing access to the required resources.

The logistical issues of how to submit, monitor and parse the outputs of so many calculations was met by adopting the RMCS infrastructure from the eMinerals project. The program Monty was used to submit batches of calculations to RMCS, which in turn submitted the calculations to the NW-GRID, seamlessly handling the staging of data to and from the machines.

GAMESS-UK was modified to output XML(CML), enabling AgentX to be used within RMCS to extract and archive metadata from the calculations. The Rgem program was then used to tabulate the metadata, providing an automatic plot of the results of the runs.

When run within this infrastructure, the 384 calculations for the generation under investigation were able to be run and the results processed within a week – with

only a few hours of manpower being required for the entire process. This is a significant improvement on the several weeks that were required for previous runs, with the submission and processing of the results requiring many days.

The results of this generation are currently being processed and provide a significant advance towards the overall goal of developing a new catalyst.

9 Conclusions

The preceding reports on the science undertaken on the NW-GRID show that there is an increasing demand for large numbers of medium-sized calculations to be undertaken within computational studies. For many investigators, it is often relatively easy to gain access to large numbers of single-processor machines for several hours (e.g. via university Condor Pools). However, as soon as the calculations are required to be run in parallel, or for more than 12 hours, the Condor approach quickly becomes unusable. The alternative option of running such calculations on national or local HPC resources is also not really viable due to the limitations of batch queuing systems and the logisitics of processing the data.

Resources such as the NW-GRID provide exactly the compute resources that are required by these projects. The use of middleware such as the eMinerals framework then makes accessing these disparate resources relatively trivial, and also makes the movement and processing of the huge quantities of data produced almost transparent to the user.

The complexity of the underlying software stack, and the technical and human issues associated with reliably linking together different resources from different administrative domains, mean that the installation and mainteinence of such a system is far from trivial, and several of the projects suffered from problems engendered by this.

However, once the infrastructure is in place, the rewards and increase in productivity that can be realised would seem to more than justify the effort expended, and it can be expected that, as the middleware becomes more mature and settles down, Grid resources will increasingly become part of the everyday tools for practicing researchers.

10 Acknowledgments

Some of the work described in this paper was undertaken at the STFC e-Science Centre, Daresbury Laboratory (now STFC) Centre for e-Science. NW-GRID is funded by NWDA, the North West Development Agency.

The eMinerals project would like to acknowledge the financial support from NERC (grant numbers NER/T/S/2001/00855, NE/C515698/1 and NE/C515704/1)

References

[1] NW-GRID: the North West Grid. `http://www.nw-grid.ac.uk`

[2] eMinerals: Environment from the Molecular Level. `http://www.eminerals.org.uk`

[3] R.P. Tyer, K. Kleese van Dam, P.A. Couch, I.T. Todorov, R.P. Bruin, A.M. Walker, T.O.H. White, K.F. Austen, M.T. Dove, M.O. Blanchard *Automatic Metadata Capture and Grid Computing.* Proceedings of the UK e-Science All Hands Meeting (2006) oai:epubs.cclrc.ac.uk:work/36704

[4] P. Couch, P. Sherwood, S.A. Sufi, I. Todorov, R.J. Allan, P. Knowles, R. Bruin, M.T. Dove, P. Murray-Rust. *Towards Data Integration for Computational Chemistry* Proceedings of the UK e-Science All Hands Meeting (2006) Paper 509. oai:epubs.cclrc.ac.uk:work/35404
See also DataRepresentation Wiki: `http://www.datarepresentation.org`

[5] M. Tribaudino, M. Prencipe, M. Bruno and D. Levy *High-pressure behaviour of Ca-rich C2/c clinopyroxenes along the join diopside-enstatite ($CaMgSi_2O_6$-$Mg_2Si_2O_6$)*, Physics and Chemistry of Minerals volume 27 (2000) 656-664.

[6] A.M. Walker, M.T. Dove, L.A. Sullivan, K. Trachenko, R.P. Bruin, T.O.H. White, P. Murray-Rustm R.P. Tyer, P.A. Couch, I.T. Todorov, W. Smith, K. Kleese van Dam *Anatomy of a grid-enabled molecular simulation study: the compressibility of amorphous silica* Proceedings of the UK e-Science All Hands Meeting (2006) 653-660 ISBN 0-9553988-0-0

[7] M. T. Dove, L. A. Sullivan, A. M. Walker, R. P. Bruin, T. O. H. White, K. Trachenko, P. Murray-Rust, I. T. Todorov, R. P. Tyer, P. A. Couch, K. Kleese van Dam, W. Smith *Molecular dynamics in a grid computing environment: experiences using DL_POLY_3 within the e Minerals escience project* Molecular Simulations, 32 (2006) 949-952. doi:10.1080/08927020600883293

[8] A. M. Walker, L. A. Sullivan, K. Trachenko, R. P. Bruin, T. O. H. White, R. P. Tyer, I. T. Todorov, S. A. Wells, M. T. Dove (2007) "The origin of the compressibility anomaly in amorphous silica: a molecular dynamics study" Journal of Physics, Condensed Matter, 19, art.no. 275210. doi:10.1088/0953-8984/19/27/275210

[9] K.F. Austen, T.O.H. White, M.T. Dove, E. Artacho, R.P. Tyer *Using eScience to calibrate our tools: parameterisation of quantum mechanical calculations with Grid technologies* Proceedings of the All Hands Meeting (Nottingham, 2006)

[10] SIESTA Application. Soler, J.M., et al., Journal Of Physics-Condensed Matter 14 (11) (2002) 2745-2779

[11] HPCx Consortium: `http://www.hpcx.ac.uk`