Molecular Simulation, Vol. 32, Nos. 12–13, 15 October–15 November 2006, 945–952

Taylor & Francis
Taylor & Francis Group

Molecular dynamics in a grid computing environment: experiences using DL_POLY_3 within the *e*Minerals escience project

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§ and W. SMITH§

†Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK
‡National Institute for Environmental eScience, Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK
¶Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
§CCLRC Daresbury Laboratory, Daresbury, Warrington, Cheshire WA4 4AD, UK

(Received May 2006; in final form June 2006)

We use the example of a study of the compressibility anomaly in amorphous silica to illustrate how molecular-scale simulations can be performed using grid computing. The potential for running many simulations within a single study requires the use of new data management methods, which are discussed in this paper. The example of silica highlights the advantages of the use of grid computing for studying subtle effects.

Keywords: Amorphous silica; Grid computing; DL_POLY_3; XML; Chemical markup language; *e*Minerals

1. Introduction

With the exponential increase in the capabilities of computers with time, two possibilities for simulations are to study ever larger and more complex systems and to carry out studies with wide sweeps through the controlling parameters. A good example would be to sweep through different ensemble temperatures or pressures to study phenomena such as phase transitions. In such cases, it may frequently be desirable to run many simulations, particularly if the object of study only varies weakly with the conditions of the ensemble. In earlier times, the simulation scientist would typically be able to perform a limited number of calculations using dedicated computing facilities, one after the other, accumulating data as each simulation ends. With the availability of grid computing (described in more detail below), there is scope for the simulation scientist to accumulate much larger collections of data associated with a single study. This, of itself, means there is a need for new approaches to job and data management in simulation studies and this paper considers these issues in detail. Grid computing requires new ways of running and managing simulation studies and the work reported in this paper is taken from experience gained by the UK *e*Minerals project [1] in developing grid computing methods for molecular simulations. The focus of the discussion will be a case study of the compressibility anomaly in amorphous silica using molecular dynamics simulations with the new DL_POLY_3 code [2,3] suitably adapted for the needs of grid computing.

The outline of this paper is as follows. First, we will describe the science and the simulations to be performed, since these control the application of the methods. Then, we will describe the job submission approach, tailored for grid computing. This will be followed by a discussion of the data management issues, with some emphasis on the use of XML for data representation and the possibilities that this presents. Then, we will describe the results of the simulation study on amorphous silica, with more of an emphasis on user experiences than on the science (the main science outcomes will be described in detail elsewhere). The main aim of this paper is not to report the science *per se*, but to use the science example to show how grid computing can enable high-throughput molecular simulation studies provided that effective data management strategies are in place.

*Corresponding author. Email: mtd10@cam.ac.uk

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020600883293

2. Science case study: compressibility of amorphous silica

2.1 Science background

Silica glass is known to show a maximum in its compressibility as a function of pressure, at a pressure of around 2 GPa [4]. The compressibility is defined as

$$\beta = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)$$

where V is the volume and P is the pressure. The origin of the compressibility anomaly in silica glass has yet to be properly explored. There have been suggestions that the anomaly is associated with a low-density to high-density polyamorphic phase transition [5,6], but the evidence for such a phase transition is weak and has been disputed [7].

Our hypothesis is very simple. At high pressure, we anticipate that compressibility will decrease on increasing pressure as the natural response to compaction of the material. This is universally found to be the case. In silica glass, up to around 3–5 GPa this compaction will involve buckling of the network of corner-linked SiO_4 tetrahedra without the tetrahedra deforming significantly. At higher pressures, though, further compaction can only be accomplished if there is breaking and forming of new bonds [7,8], when the compressibility will be further reduced. The challenge is to understand what happens at low pressure. We take the limiting case of large negative pressures. In this case, the network of corner-linked SiO_4 tetrahedra will be stretched and changes in volume with changing pressure will only be possible if accompanied by deformations of the SiO_4 tetrahedra such as stretching of the stiff Si—O bond. Thus, volume changes in this limit will require higher energy changes than at pressures near 0 GPa. On the other hand, at pressures around 0 GPa, the network of corner-linked SiO_4 tetrahedra is much more flexible than in either limit [7,9] and changes in volume will not cost a lot of energy. Thus, we anticipate that the compressibility will be somewhat lower at these pressures.

2.2 Simulation details

In order to explore the behaviour under pressure, we have carried out molecular dynamics simulations using DL_POLY_3 [2,3]†. Our samples are perfect configurations of corner linked SiO_4 tetrahedra, generated from configurations of amorphous silicon produced using the WWW method. Full details of these configurations are given elsewhere [10,11]. The structures of these configurations have been validated by neutron diffraction measurements analysed by the reverse Monte Carlo method [12]. We have models containing 256, 512 and 4096 formula units; in this study, we report results obtained with 512 formula units. The configuration is shown in figure 1.

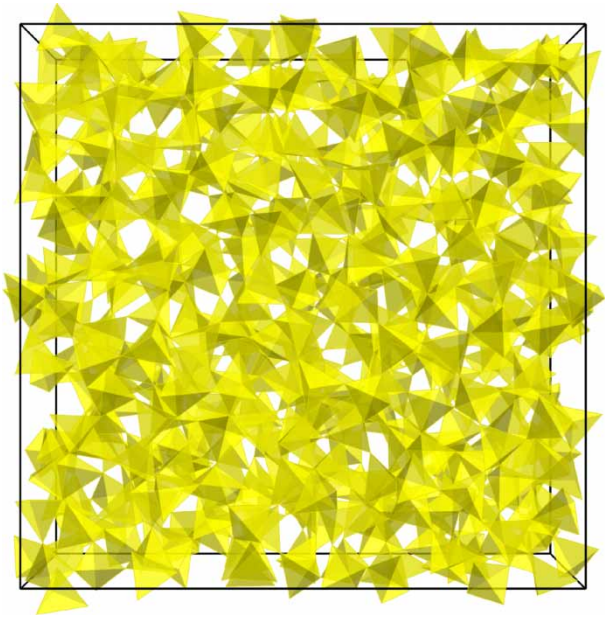

Figure 1. View of the configuration of silica used in the present study, showing the SiO_4 tetrahedra as solid units.

Our simulations were performed using two interatomic potentials, due to Tsuneyuki and co-workers [13] and to van Beest and co-workers [14]. Both were obtained by fitting potential energy functions to energy surfaces obtained by quantum mechanical calculations on silica clusters, with the van Beest model adjusted to account for effects associated with holding the atoms within bulk samples. In practice, we found that we obtained very similar results for both models and we therefore concentrate here on the results for the Tsuneyuki model.

Our simulations were performed using two ensembles: NPT (constant pressure and temperature) and NVE (constant volume and temperature). The NPT models were used to generate average volumes for different pressures and the NVE models were used to generate configurations for subsequent analysis, including the analysis of the pair distribution function reported here. All simulations were performed at a temperature of 50 K, with pressures for the NPT simulations ranging from −5 to +5 GPa to accommodate the limiting cases described above. All NPT simulations began from the same starting configuration and the NVE simulations were started from configurations generated from the output of the prior NPT simulations with the cell size set as the derived average cell size. The simulations were performed with time steps of 1 fs. Typically simulations were run with the equilibration period lasting 10 ps with a further 40 ps running for analysis.

†It should be noted that for our sample sizes, DL_POLY_2 is better optimised than DL_POLY_3, but as our development code, only DL_POLY_3 contains the capabilities described in Section 4.2.

3. Job submission and management

The simulations were performed using linux computers within grid computing environments, with the serial version of DL_POLY_3. The runs were performed using the computers of the *e*Minerals minigrid (figure 2) [15,16] and CamGrid, the campus grid at Cambridge University [17]. The minigrid is based primarily around the use of linux clusters which run PBS schedulers, as well as other computing resources and the CamGrid campus grid (which is linked to the *e*Minerals minigrid) is a group of condor pools that are flocked together to create a single resource [18]. Grid computing gives the possibility of running many jobs concurrently, thus reducing the overall turn-around time for a whole study. With the increases in power of individual processors, the capabilities of high-throughput computing are now sufficient for studies using molecular dynamics simulations. In fact, in a wide-area grid, there are increased possibilities for large numbers of concurrent calculations, but with larger numbers of jobs comes challenges associated with setting them up, managing their running and managing the data produced by the jobs.

Before describing the tools, it is important to mention some details of the grid infrastructures we are working with. The compute resources are all accessed with authentication managed by the Globus toolkit [19] and the use of X.509 digital certificates. Typically, jobs are submitted to a Globus gatekeeper in front of a job manager, whether PBS or condor. In order to submit a job, it is necessary to have Globus also installed on the user's workstation. Data transfer methods using Globus are not always ideal, not least because it is always necessary to specify exactly which files to transfer without being able to use a wildcard and it is not possible to transfer directories. Our approach, as seen in figure 2, is to incorporate the San Diego Supercomputer Center storage resource broker (SRB) [20] within our grid computing environment. The SRB is a single logical file system in which files can be distributed across several geographically distinct file systems (vaults). The user sees all files in the logical system without needing to know where the files actually are. Our approach is to upload all input files to the SRB before submission of a grid job, then upload all output files to the same directory on the SRB. Thus, the user has a complete archive of the data associated with a single job and all file transfers are handled by the SRB.

Job submission was carried out using the my_condor_submit (MCS) tool [15,16,21], which was specifically developed by one of the authors (RPB) to facilitate job submission to Globus grid computing environments with a high degree of usability as the key design requirement. As a result, the user only needs to supply a very simple script. In addition to handling job submission, MCS is also designed to enable interaction with the SRB, in order to enable downloading of files from the SRB to the actual grid resource on which the simulation is performed and the subsequent upload of output files. Many of the details of MCS are described in detail elsewhere [21].

Although MCS manages the running of jobs, there is also the challenge of preparing all the input files. This is accomplished by a separate tool developed by one of the authors (RPB) [21]. At its heart, the tool has a script that takes a template and generates a set of input files and MCS

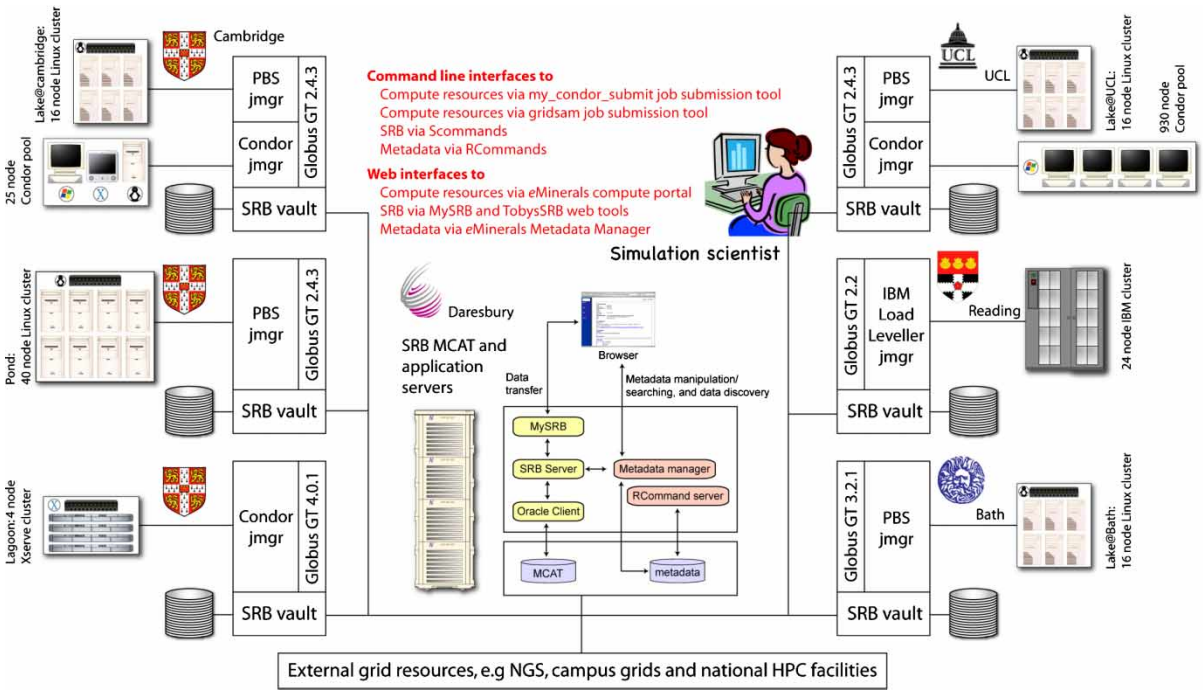

Figure 2. The eMinerals minigrid, showing a mixture of compute resources (clusters and condor pools) and the associated data management infrastructure based on the SRB.

submission scripts for each value of whatever parameter is being varied. These are all put into separate directories and the data files are uploaded into separate collections (the name given to the logical equivalent of a directory) on the SRB. Then each job is submitted independently. It should be noted that MCS performs some degree of job metascheduling subject to a set of criteria specified by the user.

4. Data management

4.1 Use of the storage resource broker

As noted above, the SRB is integral to the data management system used in this study. Each collection contains all the input files for a single DL_POLY run, with the executable being held in a separate collection.

4.2 Use of XML data representation: chemical markup language

XML is now one of the dominant standards for data representation in many applications of IT. For scientific data, it has the advantage of being able to clarify the meaning of data through a standardised syntax and use of dictionaries. This contrasts with the traditional lists of numbers generated by most simulation programs, in which meaning is provided either by header text, something simple like the names of the file and directory, or a separate document (paper or electronic).

For simulation studies, the primary advantage of XML data representation is that analysis tools can be used with output files without the scientist needing to take account of the data format. This includes extracting key pieces of information from one file to collate with similar information from other files, such as pressure and volume for construction of a volume vs. pressure curve.

For our work, we use the chemical markup language (CML) [22], which is one of the first practical implementations of XML [23,24]. DL_POLY_3 now has the ability to write CML files. A part of the CML file is reproduced in figure 3. The XML statements are typically contained within three lists:

Metadata associated with the calculation, the metadataList tags: information such as the program name and version number.
Parameter metadata, the parameterList tags: values for the parameters that control the simulation run, such as temperature, pressure, potential energy functions, cut-offs, number of time steps, potential energy parameters, etc.
Property metadata, the propertyList tags: values of various properties calculated in the simulation that would be useful for subsequent metadata searches, such as final or average energy or volume.

CML has a high degree of flexibility, by design, but there is a price to play for flexibility when it comes to developing analysis tools. Our experience suggests that simulation codes should conform to certain additional standards not imposed by CML in order to make the analysis codes more robust. We have discussed this point elsewhere [25].

One application of the use of CML is to provide a new view of the data. We have a tool called TobysSRB [26], developed by one of the authors (TOHW), which provides a web browser access to CML files stored in the SRB and converts the CML files to XHTML for viewing in an easy format in a web browser, with data plotted using scalable vector graphics (SVG, an XML language for two-dimensional graphics, which is useful for plotting graphs).

4.3 Metadata management

With the opportunity to collect many files within a simulation study, it is important to have a reliable means of providing accurate external information to each data file. Such information, which is typically called "metadata", will enable the data to retain value. For example, it will permit the scientist or a collaborator to quickly locate files they are looking for and will avoid creeping ambiguity over the provenance of files through passing time. In addition to allowing data interoperability, collation and access, the use of XML permits the automatic collection of metadata. MCS is able to extract critical pieces of information from the XML files which can be passed to the metadata database, using two tools, namely AgentX [27] and the RCommands [28], developed by two of the authors (PAC and RPT, respectively). AgentX is a library that makes it simple for simulation codes to extract information from XML data documents. It permits these documents to be queried based on terms specified in an Ontology. Mappings can be defined that relate terms in the Ontology to fragments of the data documents; such mappings have been created for DL_POLY's CML. AgentX abstracts output format the details of the data format from the users such that the complexities of dealing directly with the XML are removed. The RCommands framework consists of a metadata database and a set of commands to upload metadata to the database. The important point is that the XML/CML produced by DL_POLY_3 is used by MCS to collect metadata automatically immediately after the completion of a job, without user intervention. The metadata will include information about the program (from the *metadataList* tags), information about the control parameters (from the *parameterList* tags) and information about the final output, such as average volume (from the *propertyList* tags).

After completion of a study, the metadata can be examined using two tools developed by one of the authors (RPT) [28]. First is the set of simple unix line command (RCommands) mentioned above, which enable the metadata line tools to be listed or searched and further annotations to be added to the metadata. Second is a web interface to the metadata database to enable browsing of the metadata collection.

```
<?xml version="1.0" encoding="UTF-8"?>
<cml xmlns="http://www.xml-cml.org/schema"
 xmlns:xsd="http://www.w3.org/2001/XMLSchema"
 xmlns:dc="http://purl.org/dc/elements/1.1/title"
 xmlns:dl_poly:"http://www.cse.clrc.ac.uk/msi/software/DL_POLY/dict"
 xmlns:dl_polyUnits:"http://www.cse.clrc.ac.uk/msi/software/DL_POLY/units">

<metadataList>
 <metadata name="dc:contributor" content="I.T.Todorov &amp;amp; W.Smith"/>
 <metadata name="dc:source"
 content="cclrc/ccp5 program library package, daresbury laboratory molecular dynamics
program for large systems"/>
 <metadata name="identifier" content="DL_POLY version 3.06 / March 2006"/>
 <metadata name="systemName" content="DL_POLY : Glass 512 tetrahedra"/>
</metadataList>

<parameterList title="control parameters">
 <parameter title="simulation temperature" name="simulation temperature"
 dictRef="dl_poly:temperature">
 <scalar dataType="xsd:double" units="dl_polyUnits:K"> 5.0000E+01 </scalar>
 </parameter>
 <parameter title="simulation pressure" name="simulation pressure"
 dictRef="dl_poly:pressure">
 <scalar dataType="xsd:double" units="dl_polyUnits:katms"> -3.0000E+01 </scalar>
 </parameter>
 <parameter title="simulation length" name="selected number of timesteps"
 dictRef="dl_poly:steps">
 <scalar dataType="xsd:integer" units="dl_polyUnits:steps"> 50000 </scalar>
 </parameter>
</parameterList>

 <propertyList title="rolling averages">
  <property title="total energy" dictRef="dl_poly:eng_tot">
   <scalar dataType="xsd:double" units="dl_polyUnits:eV_mol.-1"> -2.7360E+04 </scalar>
  </property>
  <property title="volume" dictRef="dl_poly:volume">
   <scalar units="dl_polyUnits:Angstroms.3">2.2316E+04</scalar>
  </property>
 </propertyList>
 <propertyList title="execution time">
  <property title="run time">
   <scalar dataType="xsd:double" units="dl_polyUnits:s"> 17475.422 </scalar>
  </property>
 </propertyList>
</cml>
```

Figure 3. Example CML from a DL_POLY_3 run, showing the separation of markup into *metadataList*, *parameterList* and *propertyList* elements.

5. Results

5.1 Compressibility anomaly

The pressure-dependence of the volume of amorphous silica is shown in figure 4. This plot shows clearly that there is a maximum in the value of the slope around 0–1 GPa, which corresponds to the compressibility maximum. The volume curve has been fitted with a polynomial, which is used to calculate the pressure-dependence of the compressibility, figure 5.

The compressibility maximum in our simulation is at a lower pressure than found in experiment—1 GPa compared to 2 GPa, but sensitive to the polynomial fit to the volume data—which we do not consider to be a significant difference given that the interatomic potentials have not been optimised for this type of analysis. The second potential model we worked with (results not reported here)

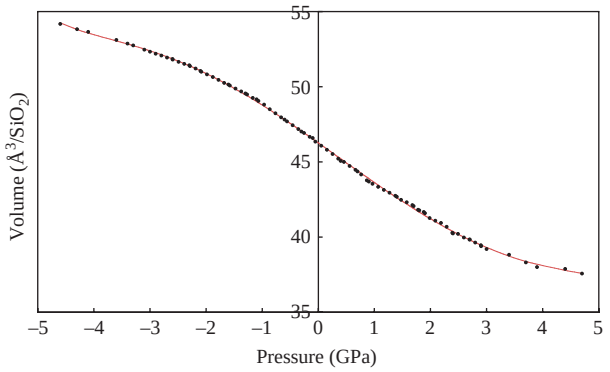

Figure 4. Pressure-dependence of the volume (per formula unit), fitted with a polynomial (curve).

gave the maximum of the compressibility at a similar pressure, the only difference being that the decrease in the value of the compressibility on changing pressure is somewhat slower.

5.2 Pair distribution functions

We have calculated the pair distribution function and we focus on the peaks associated with the nearest-neighbour Si—O, O—O and Si—Si distances. Normalised values of the peak positions, i.e. the mean interatomic distances, are plotted in figure 6 and the variances of the distributions of each distance are shown in figure 7.

Each of the mean interatomic distances increase on reducing pressure and decrease on increasing pressure. However, the mean Si—O and O—O interatomic distances vary much less than the mean Si—Si interatomic distances, the latter of which follow much more closely the variation of the sample length scale (i.e. the cube root of the volume) which is also shown in figure 6. This result demonstrates that the SiO_4 tetrahedra are barely changed by pressure other than showing a small stretch at negative pressure, as anticipated in our hypothesis (Section 2.1). On the other hand, the variation of the Si—Si mean interatomic distance suggests considerable buckling of the network of corner-linked SiO_4 tetrahedra. The point is

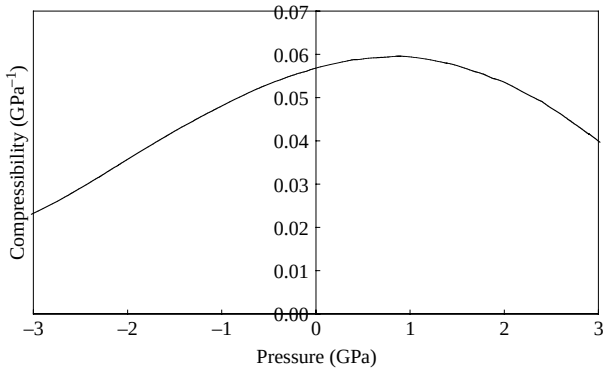

Figure 5. Compressibility of silica sample derived from the dependence of volume on pressure.

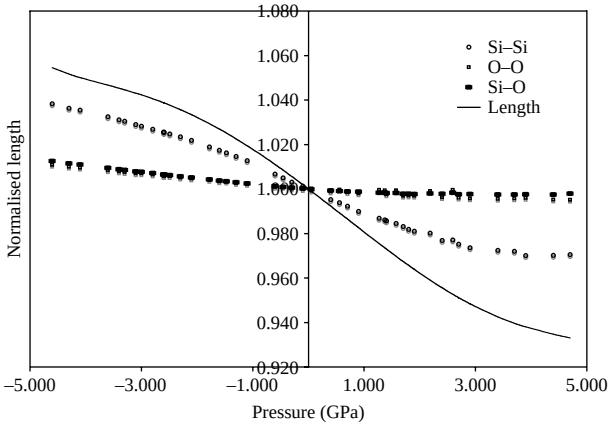

Figure 6. Normalised mean interatomic distances. The symbols are defined in the text; the smooth curve is the cube root of the volume and represents the linear size of the simulation sample.

consistent with the plots of variance, figure 7. There is considerable broadening of the Si—Si distribution on increasing pressure, much more so than for the Si—O interatomic distances. The O—O variance reflects bending of the O—Si—O bonds (a lower energy process than stretching of the Si—O bond) and the variance of the O—O distribution has a minimum around 0 GPa. The Si—O variance is much lower, but also has a minimum around 0 GPa. These variances imply that the SiO_4 tetrahedra are more uniformly in their natural unstressed state around 0 GPa, neither stretched nor compacted.

5.3 Discussion of results

The results presented above, whilst not proving the hypothesis described in Section 2.1, are fully consistent with it. The results show the following:

- that the compressibility maximum anomaly is seen clearly in simulations of amorphous silica using two different models for the interatomic potentials. We believe that this suggests that the anomaly is intrinsic to the nature of the silica network and is not merely a function of the interatomic potentials;

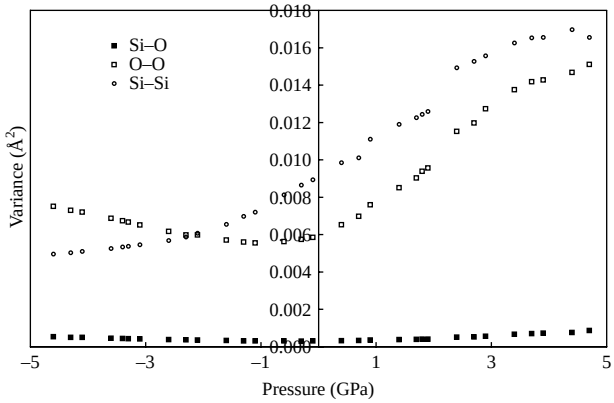

Figure 7. Variances of the mean interatomic distances; symbols are defined in the legend.

- that the primary effect on the structure caused by changing pressure is a buckling of the network of corner-linked SiO_4 tetrahedra without significant distortions of the tetrahedra. It should be noted, however, that our simulations performed at higher pressures and reported elsewhere [7,8] show that further pressure increases beyond the range covered in this study will lead to rebonding and changes in coordination numbers. This buckling, as indicated by changes in the nearest-neighbour Si—Si distance, closely follows the volume change;
- that on increasing *negative* pressure from zero pressure, there is a small stretching of the SiO_4 tetrahedra, as seen in the Si—O and O—O mean distances and increased variance of the distribution of these distances. In the limit of large negative pressure, where the sample is being stretched, some of the volume change needs to be accommodated by stretching of the tetrahedra. In the limit of large negative pressure, the Si—O bonds continue to be stretched even as the buckling of the network is reduced; and
- on increasing pressure from zero pressure, where the buckling of the network increases as seen by both the continued change of the mean Si—Si interatomic distance and the increase in the variance of the distribution of the Si—Si interatomic distances, the Si—O distance changes less than in the bond-stretching regime at negative pressures.

These results add weight to the hypothesis (Section 2.1) that the compressibility maximum arises from the fact that the high and low pressure limits have low compressibilities due to different reasons, namely compaction at high pressure and stretching at low pressure. Neither mechanism applies around zero pressure, enabling volume changes to occur with lower energy changes than at either limit.

6. Grid computing and molecular dynamics: discussion

The main point of this paper is to highlight the use of grid computing for high-throughput molecular dynamics simulations. We have briefly described the grid computing infrastructure we have used, but given a more extensive description of the data management issues because these impact more on developments of the simulation code. We are now of the opinion that simulation codes, like DL_POLY_3, should have the facility to write output XML files. Moreover, CML is proven to be one mature and useful dialect of XML for this purpose and experience of its use has led to suggestions for how CML should be implemented within simulation codes.

The big advantage of grid-enabled high-throughput simulation studies is that it is possible to collect extensive data sets derived from the raw simulation data. In the present case, our focus has been on volume and nearest-neighbour distances as detailed functions of pressure. In the present work, we have investigated the pressure-dependence of the differential of the extracted property, volume. To ensure that it is possible to obtain the maximum in this quantity with sufficient accuracy, a large number of points is useful to enable a polynomial fit to minimise the effects of statistical noise; this is best accomplished by increasing the number of data points. Because jobs are run concurrently rather than sequentially, increasing the number of data points does not give too great an increase in effort on the part of the scientist as far as running the computations is concerned. We have developed, as discussed above, tools to make the submission of jobs as part of a parameter sweep relatively easy for the scientist user.

The greatest challenge is the management of the scientific data generated by grid-enabled simulation studies. These challenges are both in accurately archiving the information associated with the data, and also with collating derived quantities. The problem of archiving data is helped both by the use of the SRB and the use of metadata within a wide metadata management infrastructure. The problem of collating data is helped by the use of XML, with tools designed specifically to extract data from many XML files.

In summary, we have solved a number of the problems that face the scientist user of molecular simulation codes within a grid computing environment, enabling studies to be performed using many more simulation runs than a single user scientist would normally be able to manage using traditional tools.

Acknowledgements

We acknowledge financial support from NERC for the *e*Minerals project.

References

[1] M.T. Dove, N.H. de Leeuw. Grid computing and molecular simulations: the vision of the *e*Minerals project. *Mol. Simul.*, **31**, 297 (2005).
[2] I.T. Todorov, W. Smith. DL_POLY_3: the CCP5 national UK code for molecular-dynamics simulations. *R. Soc. Phil. Trans.*, **362**, 1835 (2004).
[3] I.T. Todorov, W. Smith, K. Trachenko, M.T. Dove. DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism. *J. Mater. Chem.*, **16**, 1911 (2006).
[4] O.B. Tsiok, V.V. Brazhkin, A.G. Lyapin, L.G. Khvostantsev. Logarithmic kinetics of the amorphous–amorphous transformations in SiO_2 and GeO_2 glasses under high-pressure. *Phys. Rev. Lett.*, **80**, 999 (1998).
[5] L.P. Huang, J. Kieffer. Amorphous–amorphous transitions in silica glass. I. Reversible transitions and thermomechanical anomalies. *Phys. Rev. B*, **69**, Art. No. 224203 (2004).
[6] L.P. Huang, L. Duffrene, J. Kieffer. Structural transitions in silica glass: thermo-mechanical anomalies and polyamorphism. *J. Non-Cryst. Sol.*, **349**, 1 (2004).
[7] K. Trachenko, M.T. Dove. Compressibility, kinetics, and phase transition in pressurized amorphous silica. *Phys. Rev. B*, **67**, Art. No. 064107 **67** (2003).
[8] K. Trachenko, M.T. Dove. Floppy modes in silica glass under pressure. *J. Phys. Condens. Matter*, **14**, 1143 (2002).

[9] K. Trachenko, M.T. Dove. Densification of silica glass under pressure. *J. Phys. Condens. Matter*, **14**, 7449 (2002).
[10] K. Trachenko, M.T. Dove, K.D. Hammonds, M.J. Harris, V. Heine. Low energy dynamics and tunneling states in silica glass. *Phys. Rev. Lett.*, **81**, 3431 (1998).
[11] K.O. Trachenko, M.T. Dove, M.J. Harris, V. Heine. Dynamics of silica glass: two-level tunnelling states and low-energy floppy modes. *J. Phys. Condens. Matter*, **12**, 8041 (2000).
[12] M.G. Tucker, D.A. Keen, M.T. Dove, K. Trachenko. Refinement of the Si—O—Si bond angle distribution in vitreous silica. *J. Phys. Condens. Matter*, **17**, S67 (2005).
[13] S. Tsuneyuki, M. Tsukada, H. Aoki, Y. Matsui. 1st principles interatomic potential of silica applied to molecular dynamics. *Phys. Rev. Lett.*, **61**, 869 (1988).
[14] B.W.H. van Beest, G.J. Kramer, R.W. van Santen. Force fields for silicas and aluminophosphates based on *ab initio* calculations. *Phys. Rev. Lett.*, **64**, 1955 (1990).
[15] M. Calleja, L. Blanshard, R. Bruin, C. Chapman, A. Thandavan, R. Tyer, P. Wilson, V. Alexandrov, R.J. Allen, J. Brodholt, M.T. Dove, W. Emmerich, K. Kleese van Dam. Grid tool integration within the eMinerals project. *Proceedings of the UK e-Science All Hands Meeting 2004*, (ISBN 1-904425-21-6) pp. 812–817 (2004).
[16] M. Calleja, R. Bruin, M.G. Tucker, M.T. Dove, R. Tyer, L. Blanshard, K. Kleese van Dam, R.J. Allan, C. Chapman, W. Emmerich, P. Wilson, J. Brodholt, A. Thandavan, V.N. Alexandrov. Collaborative grid infrastructure for molecular simulations: the eMinerals minigrid as a prototype integrated compute and data grid. *Mol. Simul.*, **31**, 303 (2005).
[17] M. Calleja, B. Beckles, M. Keegan, M.A. Hayes, A. Parker, M.T. Dove. CamGrid: experiences in constructing a university-wide, condor-based grid at the University of Cambridge. *Proceedings of the UK e-Science All Hands Meeting 2004*, (ISBN 1-904425-21-6) pp. 173–178 (2004).
[18] D. Thain, T. Tannenbaum, M. Livny. Condor and the grid, Chapter 11. In *Grid Computing: Making The Global Infrastructure a Reality*, F. Berman, A.J.G. Hey, G. Fox (Eds.), John Wiley & Sons, New York (2003).
[19] I. Foster, C. Kesselman. Globus: a metacomputing infrastructure toolkit. *Int. J. Supercomput. Appl.*, **11**, 115 (1997).
[20] R.W. Moore, C. Baru. Virtualization services for data grids, Chapter 11. In *Grid Computing: Making The Global Infrastructure a Reality*, F. Berman, A.J.G. Hey, G. Fox (Eds.), John Wiley & Sons, New York (2003).
[21] R.P. Bruin, T.O.H. White, A.M. Walker, K.F. Austen, M.T. Dove, R.P. Tyer, P.A. Couch, I.T. Todorov. Job submission to grid computing environments. *Proceedings of All Hands*, (2006), (in press).
[22] J. Wakelin, P. Murray-Rust, S. Tyrrell, Y. Zhang, H.S. Rzepa, A. Garcia. CML tools and information flow in atomic scale simulations. *Mol. Simul.*, **31**, 315 (2005).
[23] P. Murray-Rust, H.S. Rezpa. Chemical markup, XML, and the worldwide web. 1. Basic principles. *J. Chem. Inf. Comp. Sci.*, **39**, 928 (1999).
[24] P. Murray-Rust, H.S. Rezpa. Chemical markup, XML, and the worldwide web. 1. Information objects and the CMLDOM. *J. Chem. Inf. Comp. Sci.*, **41**, 1113 (2001).
[25] T.O.H. White, P. Murray-Rust, P.A. Couch, R.P. Tyer, R.P. Bruin, M.T. Dove, I.T. Todorov, S.C. Parker. Development and use of CML in the eMinerals project. *Proceedings of All Hands* (2006), (in press).
[26] T.O.H. White, R.P. Tyer, R.P. Bruin. A lightweight scriptable web-based frontend to the SRB. *Proceedings of All Hands* (2006), (in press).
[27] P.A. Couch, P. Sherwood, S. Sufi, I.T. Todorov, R.J. Allan, P.J. Knowles, R.P. Bruin, M.T. Dove, P. Murray-Rust. Towards data integration for computational chemistry. *Proceedings of All Hands 2005*, (ISBN 1-904425-53-4) pp. 426–432 (2005).
[28] R.P. Tyer, P.A. Couch, *et al.* Automatic metadata capture and grid computing. *Proceedings of All Hands*, (2006) (in press).