Performance of Materials Science Codes (CASTEP, ONETEP) Enhanced by HECToR dCSE Team

Matt Probert, University of York
Jolyon Aarons, University of York
HECToR CSE Team, Numerical Algorithms Group Ltd (NAG)

An HPC developer from the University of York, working under NAG's Computational Science and Engineering (CSE) support service for HECToR, the UK's national academic supercomputing facility, has implemented performance improvements (in both time taken and memory used) for the geometry optimization part of the materials science codes CASTEP and ONETEP. The improvements enable larger, more complex, systems to be studied by scientists within their existing budgets.

Commenting on the dCSE project, Dr Matt Probert of the Department of Physics at the University of York said: *"[The development of the new geometry optimizer for CASTEP and ONETEP] has been a great success. The "time to science" – i.e. to do a full structural optimization in either code – has been maintained or even reduced in certain circumstances, whilst the amount of memory required has been dramatically reduced. This latter point is key, as with the move to multi-core, the amount of memory per core is going down and the old geometry optimizer consumed a large amount of non-distributed memory and was beginning to limit the size of problem that could be tackled. As a consequence, users of both CASTEP and ONETEP will now be able to tackle significantly larger systems: with the old code, there was an effective maximum size of around 3,000 atoms on HECToR Phase 3 which has now grown to 100,000 atoms. This will enable users of CASTEP to study larger structures, such as grain boundaries, and for ONETEP to tackle larger molecules, such as proteins and DNA segments. Obviously, not all CASTEP and ONETEP users on HECToR will be studying systems of this size, but there is no downside to the new minimizer: all calculations will take the same or less memory [with the exception of very small systems, which cannot be run usefully on HECToR], and converge in the same or less time than before. The new minimizer will be the default algorithm in the next release versions of both CASTEP and ONETEP."*

> The system sizes which can be studied have been increased by nearly two orders of magnitude, opening up new problem areas that can be tackled using these codes.

HECToR

HECToR is managed by EPSRC on behalf of the participating Research Councils with a mission to support capability science and engineering in UK academia. The Cray XE6 supercomputer, located at the University of Edinburgh, is managed by UoE HPCx Ltd. The CSE Support Service is provided by NAG Ltd and ensures users have access to appropriate HPC expertise to effectively exploit advanced supercomputers for their science. A critical feature of the CSE Support Service is the distributed CSE (dCSE) programme which, through lightweight peer review, delivers dedicated performance and scalability projects on specific codes in response to proposals from users. The dCSE programme now consists of over 60 focused projects complementing the traditional HPC user applications support and training also provided by NAG.

The dCSE projects completed so far have delivered outstanding examples of the cost savings and new science that can be enabled through dedicated CSE effort. The CASTEP / ONETEP project reported here adds to these success stories with a successful performance improvement.

Project Background

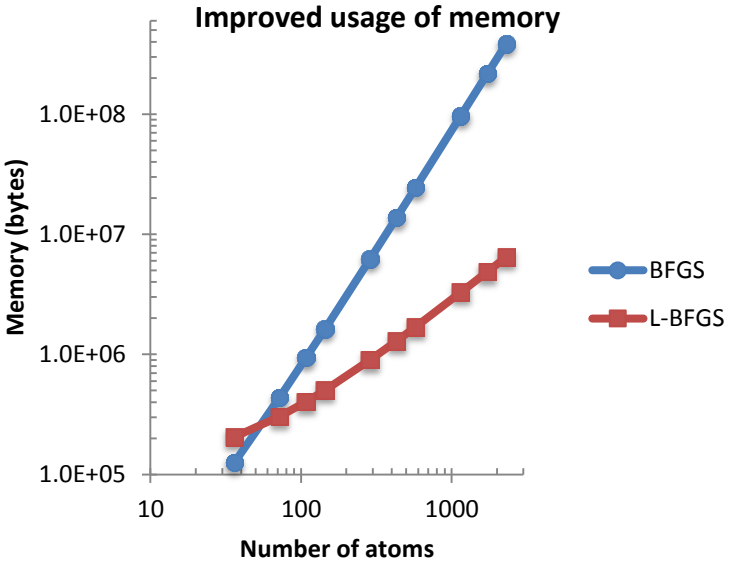
CASTEP and ONETEP are codes which use Density Functional Theory (DFT) to calculate the electronic properties of materials from first principles. They are amongst the most heavily-used applications on HECToR. A significant amount of time in these applications is devoted to geometry optimization – that is, the minimization of the system entropy with respect to atomic positions. Originally, this was done using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) non-linear optimization algorithm, which scales $O(N^2)$ with N, the number of atoms in the system under study. The goal of this dCSE project was to improve the scalability of both codes by switching to a limited-memory version of the BFGS algorithm (L-BFGS).

Matt Probert of the University of York was the Principal Investigator for the project. Jolyon Aarons of the University of York carried out the 12 person-month project, in close collaboration with the NAG CSE team.

Around 200,000 kAUs (thousands of allocation units) are used per year on HECToR to run calculations using CASTEP, ONETEP and other DFT codes.

Project Results

The original BFGS algorithm involved the explicit construction of a Hessian matrix, an operation whose memory requirements scale quadratically with system size. By contrast, L-BFGS builds the Hessian from a history of updates without having to explicitly compute or store the matrix; this results in a linear scaling of its memory requirements (see figure). L-BFGS was successfully implemented in CASTEP (and, because the geometry optimization code is shared with ONETEP, the enhanced algorithm was trivially implemented in that application too). In addition, using BFGS, the memory associated with the geometry optimization code was duplicated on every MPI image. Switching to L-BFGS means this is no longer the case, which frees up more of the hardware node's memory for use by the rest of the code. Throughout the project, opportunities to additionally optimize the existing code were taken wherever they presented themselves (for example, replacing Fortran 90 intrinsic matrix functions with their equivalents from the BLAS, which resulted in a small increase in performance).



Switching to the new algorithm reduced the amount of memory required by the codes, and changed its scaling from quadratic to linear.

Performance improvements for both time taken and memory used were demonstrated for large systems.

A full technical report can be found at http://www.hector.ac.uk/cse/distributedcse/reports/
More details for CASTEP can be found at http://www.castep.org/, and for ONETEP at http://www.onetep.org/.

For more information, please contact: HECToR CSE Team
The Numerical Algorithms Group Ltd, Wilkinson House, Jordan Hill Road, Oxford, OX2 8DR, United Kingdom
Telephone: 01865 511 245 Email: hector-cse@nag.co.uk Web: http://www.hector.ac.uk/cse/