Institution: The University of Edinburgh/Heriot-Watt University (Maxwell Institute)

Unit of Assessment: B10, Mathematical Sciences

Title of case study: New thermostatic controls adopted by molecular dynamics software providers

1. Summary of the impact

Molecular dynamics (MD) simulations are used extensively in chemistry, biology and material sciences, placing huge demands on computer resources. Because these simulations explore the behaviour of molecules at defined ambient temperature, temperature control (thermostatting) is an essential element of MD algorithms. In a series of papers published from 2009 on, Leimkuhler (Maxwell Institute) and his collaborators developed improved numerical methods for temperature control. They proposed new algorithms and analysed their properties (such as fidelity to the dynamical model, efficiency and stability). The new algorithms have since been implemented in the world's leading MD software packages including DL-Poly, AMBER, NAMD and Accelrys's Material Studio. The research has had clear economic impact on the commercial company Accelrys by improving its product, and more broadly on the community of MD code users worldwide by providing improved simulation tools.

2. Underpinning research

In MD simulation, the system size is limited by computational considerations, yet one would like simulation parameters such as temperature and pressure to be strictly regulated so that the molecular model is relevant to the experimental conditions it is meant to mimic. This motivates the introduction in MD codes of thermal regulation mechanisms – thermostats. These are perturbations of the underlying Newtonian dynamics that enable the simulated system to sample state space (all accessible configurations of the different atoms in a protein or drug molecule, for example) in a manner that approximates experimental conditions. The shortcomings of existing thermostats led Leimkuhler (Maxwell Institute, MI) to develop stochastic-dynamical schemes which are both rigorously ergodic (meaning that they sample the entire accessible phase space) and robust. Previous investigation of the widely used Nosé-Hoover dynamics, for instance, had demonstrated that this method is not ergodic (Legoll *et al*., *Nonlinearity*, **22**, 1673, 2009). This motivated Leimkuhler to characterize the performances of different approaches used in this problem and led him to develop new efficient algorithms.

New thermostatic control. In joint work with his PhD student Noorizadeh (MI) and with Theil (Warwick), Leimkuhler showed that ergodic sampling is possible using a new thermostat mechanism that combines Nosé-Hoover dynamics with a highly degenerate (scalar) stochastic process [1]. The proof, which relies on a result of Fields medallist Lars Hörmander, establishes the regularity of the Fokker-Planck operator corresponding to the degenerate diffusion. This involves understanding the effective interactions of the stochastic process with the many physical degrees of freedom to show that the noisy process propagates into all directions and that ergodicity ensues. Subsequently, Leimkuhler and Noorizadeh joined with O. Penrose (MI) to study the 'gentleness' of various thermostats in terms of their perturbation of dynamics introduced as measured by the rate of convergence of the kinetic energy [2]. This concept was entirely undeveloped in the mathematical context and made coherent a notion which had only just been suggested by physicists (Bussi *et al.*, *J. Chem. Phys*., **126**, 014101, 2007). The gentle thermostats proposed by Leimkuhler and co-workers, termed Nosé-Hoover-Langevin (NHL), are valuable wherever measures of dynamic mobility (e.g. diffusion constants) or time-constants must be recovered from

the molecular trajectories. A follow-on project (with A. Jones, Edinburgh) has addressed the treatment of driven systems using adaptive variants of the gentle thermostats as well as Langevin dynamics [3]; this concept is relevant for molecular modelling of material defects and in connection with QM/MM algorithms which introduce artificial heating along the interface between classical and quantum models. Most recently, attention has turned to Langevin dynamics methods for use in configurational sampling, for which new integration algorithms have been obtained with high sampling accuracy [4] and these have been demonstrated to be effective for biomolecular simulation [5].

Implementation. These methods have been implemented in major software packages Materials Studio, DL-Poly, AMBER and NAMD (see section 4 below). The implementation in Materials Studio was partly carried out by Leimkuhler's PhD student Matthews who spent a residency (2012) at the Cambridge headquarters of Accelrys, the company licensing Materials Studio. Most of the methods were developed with funding from the Science and Innovation Centre for Numerical Algorithms and Intelligent Software (NAIS). The thermostatting methods (including the NHL methods and large stepsize isokinetic discretizations) and generalizations are also being implemented by Leimkuhler and his team as part of the MIST (Molecular Integration Software Toolkit) within the ExTASY (Extensible Tools for Advanced Sampling and analYsis) framework. This is a major (£2M) software initiative funded under the bi-national NSF-EPSRC Software Infrastructure for Sustained Innovation programme.

Attribution. B. Leimkuhler has been Professor of Applied Mathematics at the Maxwell Institute (MI) since 2006. His PhD students at the MI, E. Noorizadeh (graduated in 2010) and C. Matthews (graduated in 2013), contributed to the research as did O. Penrose (MI), A. Jones (Edinburgh, Physics & Astronomy) and F. Theil (Warwick).

3. References to the research

References marked with a * best indicate the quality of the research.

[1]* Leimkuhler, B., Noorizadeh, E. and Theil, F., A gentle stochastic thermostat for molecular dynamics, *J. Stat. Phys*., **135**, 261-277 (2009). http://dx.doi.org/10.1007/s10955-009-9734-0

[2]* Leimkuhler, B., Noorizadeh, E. and Penrose, O., Comparing the efficiencies of stochastic isothermal molecular dynamics method, *J. Stat. Phys*., **143**, 921-942 (2011). http://dx.doi.org/10.1007/s10955-011-0210-2

[3] Jones, A. and Leimkuhler, B., Adaptive stochastic methods for sampling driven molecular systems, *J. Chem. Phys.*, **135**, 084125, (2011). http://dx.doi.org/10.1063/1.3626941ht

[4]* Leimkuhler, B. and Matthews, C., Rational construction of stochastic numerical methods for molecular sampling, *Appl. Math. Res. Express*, **2013**, 34-56, (2013). http://dx.doi.org/10.1093/amrx/abs010

[5] Leimkuhler, B. and Matthews, C., Robust and efficient configurational molecular sampling via Langevin dynamics, *J. Chem. Phys*., **138**, 174102, (2013). http://dx.doi.org/10.1063/1.4802990

Grants:

EP/K039512/1 SI2-CHE: ExTASY: Extensible Tools for Advanced Sampling and analysis, value £550K (one part of 6 linked US and UK projects worth around £2M), 2013-2016.

EP/G036136/1: Numerical Algorithms and Intelligent Software for the Evolving HPC Platform, value £4.5M, 2009-2014.

4. Details of the impact

Improved commercial software products for molecular dynamics. The enhanced value offered by the NHL thermostats in comparison to existing approaches was immediately recognised by the commercial software company Accelrys who incorporated the techniques in their commercial code *Materials Studio*. The impact on Accelrys was achieved through an extended period of interactions from which led to a solution to the problem of 'ringing' observed with the Nosé-Hoover thermostat; this solution was subsequently used in Accelrys's software. Ringing arises when the system is initialised with data far from a correct equilibrated state. This leads to a severe oscillation in kinetic energy with poor simulation results as a consequence. Although ringing could sometimes be addressed by ad hoc approaches these were inefficient and time consuming; code developers needed a robust, systematic solution. The results published by Leimkuhler and his collaborators in [1-3] suggested that the NHL method would robustly sample the canonical distribution over a much wider range of parameters and this was subsequently verified by Matthews, Akkermans (Accelrys) and Leimkuhler who demonstrated that a working NHL implementation dramatically resolves the ringing problem in simulations of several complex molecules (a silicon system and a substantial organic molecule). As a result, Accelrys implemented NHL in its *Materials Studio* software. This implementation, carried out in collaboration with Leimkuhler and Matthews, was released in version 6.0 of *Material Studio* (Nov. 2011) and was highlighted as a valuable new feature [6]. Matthews and Leimkuhler drafted the documentation of the new method. The more recent versions of Material Studio continue to rely on the NHL thermostat. A quote from a group manager at Accelrys confirms the importance of the NHL thermostat for their product: 'During 2011 Accelrys worked closely with Leimkuhler to implement the Nosé-Hoover-Langevin thermostat within Forcite, the molecular dynamics module of Materials Studio. This thermostat was a key contribution to the product because it eliminated a widespread problem experienced by Accelrys customers, specifically the excessive time required to equilibrate a system' [7].

Accelrys's *Materials Studio* is the world's leading commercial software package for molecular simulation of materials. Accelrys had 2012 revenues of $162M, most of which comes from software licenses. Detailed breakdown of sales figures is not available, but *Materials Studio* is one of Accelrys's two primary software products and is widely used within the commercial materials sector, with thousands of installations worldwide.

Enhancements to public domain software. The majority of industrial users of MD simulation tools make use of public domain software developed by government-academic partnerships and wide impact has been achieved through the implementation of the methods in several such resources. In parallel with the Accelrys implementation, the NHL method was implemented in the DL-Poly 4.0 code (STFC Daresbury laboratory, [8]) and into AMBER, a major NSF-funded molecular software package [9]. Confirmed industrial users of these software include Sony, Samsung (DL-Poly, [8]), Pfizer, Novartis, Takeda and Dart Neuroscience (AMBER, [9]). Further implementations were carried out by Bernstein (Center for Computational Materials Science. Naval Research Laboratory, Washington, DC, USA), by Gabor Csanyi (Engineering, Cambridge University) as part of a QM/MM code, and as part of the NAMD code project [10].

This research has impacted widely on the extensive community of MD code users by reducing computational requirements. Because MD simulations use vast computational resources (30% of all CPU cycles on the NSF TeraGrid (now XSEDE) HPC system [11] and over 40% of the usage of the UK National HPC Service HECToR relate to molecular simulation [12]), algorithmic improvements lead to large gains in net computing time, with clear benefits for the accuracy, reliability and cost-effectiveness of simulations.

5. Sources to corroborate the impact

[6] What's New in Materials Studio 6.0 > Big Impact with Small Science > 'Improve temperature stability in molecular dynamics calculations with new thermostats in Forcite'; mentions NHL thermostat. http://www.maths.ed.ac.uk/~mthdat25/thermostat/whats-new

[7] The implementation of NHL thermostat in Accelrys' Materials Studio Software can be confirmed by an Accelrys Group Manager.

[8] The implementation of NHL thermostat in DL-Poly 4.0 can be confirmed by a member of the Computational Science and Engineering Department, Science and Technology Facilities Council. See also DL-Poly 4 User Manual (Sec. 3.4.6 describes the implementation of the NHL method).

[9] The implementation of adaptive thermostat within AMBER can be confirmed by a Professor at the San Diego Supercomputer Center, University of California San Diego.

[10] The implementation of Langevin thermostat within NAMD can be confirmed by a member of the Theoretical and Computational Biophysics Group, University of Illinois. See also http://www.ks.uiuc.edu/Research/namd/

[11] Statistics and examples of MD computations carried out on NSF TeraGrid are reported in http://www.teragridforum.org/mediawiki/images/d/d8/DEISA-PRACE-May2009-Towns.pdf

[12] Statistics and examples of MD computations carried out on HECToR are reported in http://www.hector.ac.uk/about-us/reports/annual/2011.pdf

Note: should links to web pages be broken, please use the website http://www.maths.ed.ac.uk/~mthdat25 to access pdf versions of the pages