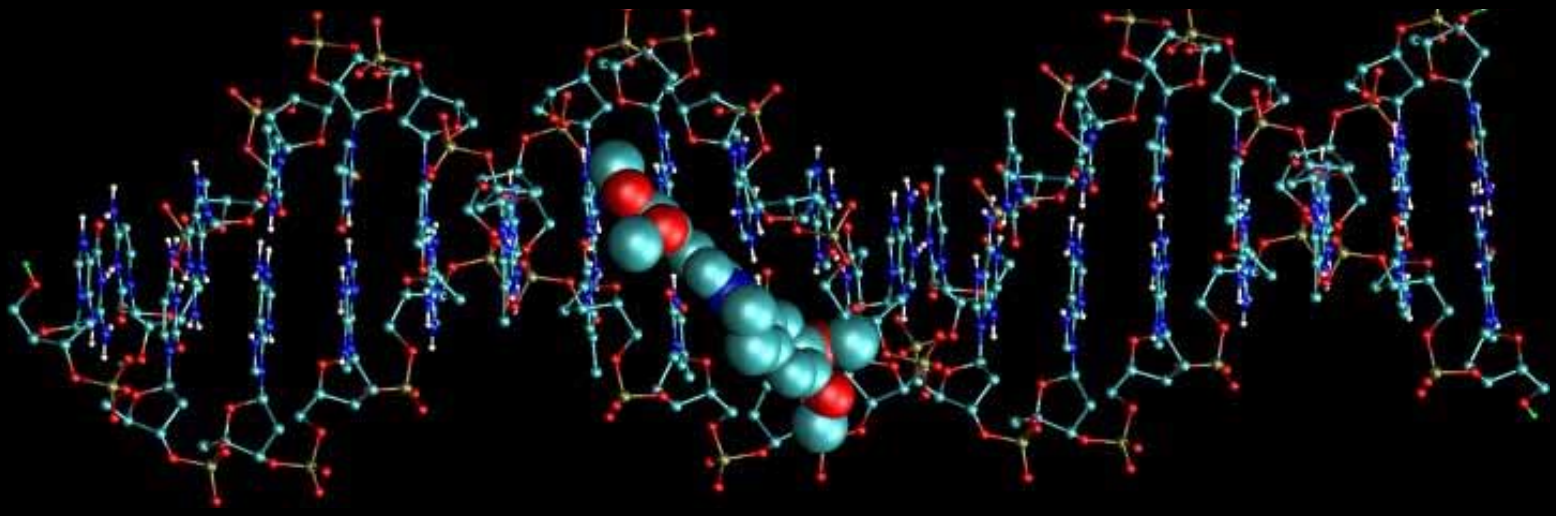
no.1 2014

PDC Center for High Performance Computing

PDC Newsletter



Erwin Laure
Director PDC and HPCViz

The PDC Newsletter is published by PDC at CSC, KTH.

PDC operates leading-edge, high-performance computers as easily-accessible national resources. These resources are primarily available for Swedish academic research and education. PDC, which is hosted by CSC, KTH, is one of the six centres in the Swedish National Infrastructure for Computing (SNIC).
Editor: Erwin Laure
Editorial Staff: Genet Edmondson
Layout: Genet Edmondson
Graphic Design: Maria Malmqvist
E-mail: pdc-newsletter@pdc.kth.se

ISSN 1401-9671

Cover

Optical biomarkers offer possibilities for early diagnosis of disease states without needing to use more harmful forms of radiation.

How to Contact PDC

Visiting address:
Teknikringen 14, "Plan 4",
KTH, Stockholm
Mail: PDC, KTH,
SE-10044 Stockholm, Sweden
E-mail: support@pdc.kth.se
www: http://www.pdc.kth.se
Phone: +46 8 790 7800
Fax: +46 8 247 784

Editorial

On the 9th of May, the board of the Swedish National Infrastructure for Computing (SNIC) approved PDC's proposal for replacing our current flagship system "Lindgren" and awarded 100 million SEK for this project. Together with a contribution of 60 million SEK from KTH and 10 million SEK from industrial partners, this adds up to an overall budget of 170 million SEK – the largest single investment in academic computing infrastructure ever in Sweden. This budget will cover both the hardware and operation of a new flagship service over four years. The service will consist of a high-end supercomputer, exceeding a performance of one petaflop, a pre- and post-processing system equipped with accelerators, and several petabytes of storage. The procurement process is currently ongoing and we expect to sign the contract with the chosen vendor in early autumn with the aim of having the system in place in early 2015. We are looking forward to an exciting new resource that will, without any doubt, provide a major step forward for the Swedish e-Science landscape.

A good example of the type of scientific research that will be performed on the new system is multiscale modelling in the materials and life sciences – a technique used, for instance, to develop biological markers for optical materials. N. Arul Murugan, Zilvinas Rinkevicius, and Hans Ågren from the Division of Theoretical Chemistry and Biology at KTH's School of Biotechnology and the Swedish e-Science Research Centre (SeRC), discuss their work on multiscale modelling in the cover article of this newsletter. In this edition we also look back at previous cover articles from the areas of fusion and fluid mechanics and provide updates on this work.

Supercomputers are sometimes compared to Formula One cars – they are on the bleeding edge of technology and using them to their full efficiency can be a difficult task that often requires special effort and skills. To support our users in exploiting our resources to their full extent and to help them optimize their codes and workflows, PDC has built up a team of second-line support application experts, people with a solid background in a scientific field and excellent supercomputing skills. Application experts are a cornerstone of SeRC's vision for e-Science and PDC's experts are employed in close collaboration with SeRC. Over the years, a great workforce has been built up at the various SNIC centres and, in this newsletter, we highlight the skills of PDC's application experts.

Apart from PDC's core activities, we are also involved in explorative projects investigating new trends such as the use of GPUs, novel storage interfaces, clouds, and the MapReduce paradigm.

You can read short updates on these activities in this edition of the newsletter.

On the European scene, the Partnership for Advanced Computing in Europe (PRACE) project provides additional high-end resources to many Swedish scientists. The PRACE consortium is just in the process of defining the second phase of the project, which will start in spring 2015. Meanwhile, it is time to analyze what PRACE has meant for Swedish users. To this end, NeIC recently performed an evaluation of the Nordic use of PRACE – the results of the evaluation are summarized in this newsletter.

Supercomputing was initially embraced by academic researchers, but industry increasingly requires supercomputing capabilities, and, most importantly, high performance computing (HPC) expertise. For many years, PDC has collaborated with industry on supercomputing issues and, to support the growing demand for the services we offer, a new business development unit has recently been created at PDC.

Finally, PDC has been engaged in organizing several high-profile events, most importantly the 2014 Exascale Applications and Software Conference (EASC2014) and the first Exascale MPI Workshop at Supercomputing 2013. These events provided excellent opportunities for the international high-performance computing community to meet and exchange ideas on how to approach the Exascale era – an era that seems to be far away even though we are likely to enter it in the next decade!

In the meantime, we are excited to be moving Sweden's e-infrastructure into the Petascale era with our new system and, while PDC will be busy in July and August evaluating the proposals for the new system, we wish all our users a relaxing and refreshing summer.

Erwin Laure, Director PDC and HPCViz

Above: PDC Open House and Pub Afternoon, 8 April 2014

In This Issue

Staff Focus

Filippia Zikou

Filippia Zikou recently joined PDC's first-line user support group where she is helping PDC users by answering questions, writing documentation and studying software performance on PDC resources. Filippia began her university studies in 2006 at the Department of Informatics and Telecommunications at the University of Athens, and in 2008 she spent a year as an Erasmus exchange student in Helsinki at the Aalto University. She was then awarded a Bachelor of Science degree from the University of Athens and currently she is doing a Masters degree in Software Engineering of Distributed Systems at KTH.

Filippia enjoys travelling. When she does not have the time to get away from the city, she goes to the gym, or enjoys walking, jogging or biking locally.

Multiscale Modelling in the Materials and Life Sciences

N. Arul Murugan, Division of Theoretical Chemistry and Biology, School of Biotechnology, KTH, Zilvinas Rinkevicius, Division of Theoretical Chemistry and Biology, School of Biotechnology, KTH, and Hans Ågren, Swedish e-Science Research Centre, KTH

Introduction

The Nobel Prize of 2013 takes chemistry into cyberspace – that was the first sentence the chairman of the Nobel Committee for Chemistry said when announcing the chemistry prize last year. The same could, however, also be said about essential parts of physics, and of the materials and life sciences. The concept of multiscale modelling [1] proposed by the Nobel laureates in the 1970s did not stir much attention at the time, partly because of the technical limitations associated with performing calculations on computers at that time, and also probably because the scientific community was not fully prepared to accept the concept as such. However, as time passed by, it became progressively more acceptable to replace the paradigm whereby one theory covered every scale of existence with the concept of bridging theories (or rather models), each of which was limited to a certain scale of length and time.

The most important variant of contemporary multiscale modelling is given by the combination of quantum mechanics and classical physics. In a sense this coalition makes it possible to combine the accuracy and rigour of the former with the applicability of the latter. It also enables us to find working approaches that can accurately address the nanoscale, which is of obvious importance for materials science (for example, nanotechnology) and also the life sciences (such as in the early stage diagnosis of diseases and drug design). In both of these fields, the two models, quantum and classical, each have shortcomings when used by themselves. Thus, joining quantum mechanics (QM) and molecular mechanics (MM) with an expedient classical force field description of atoms has become a most important and popular area of contemporary "in silico" research, which can model clusters of up to one million atoms nowadays.

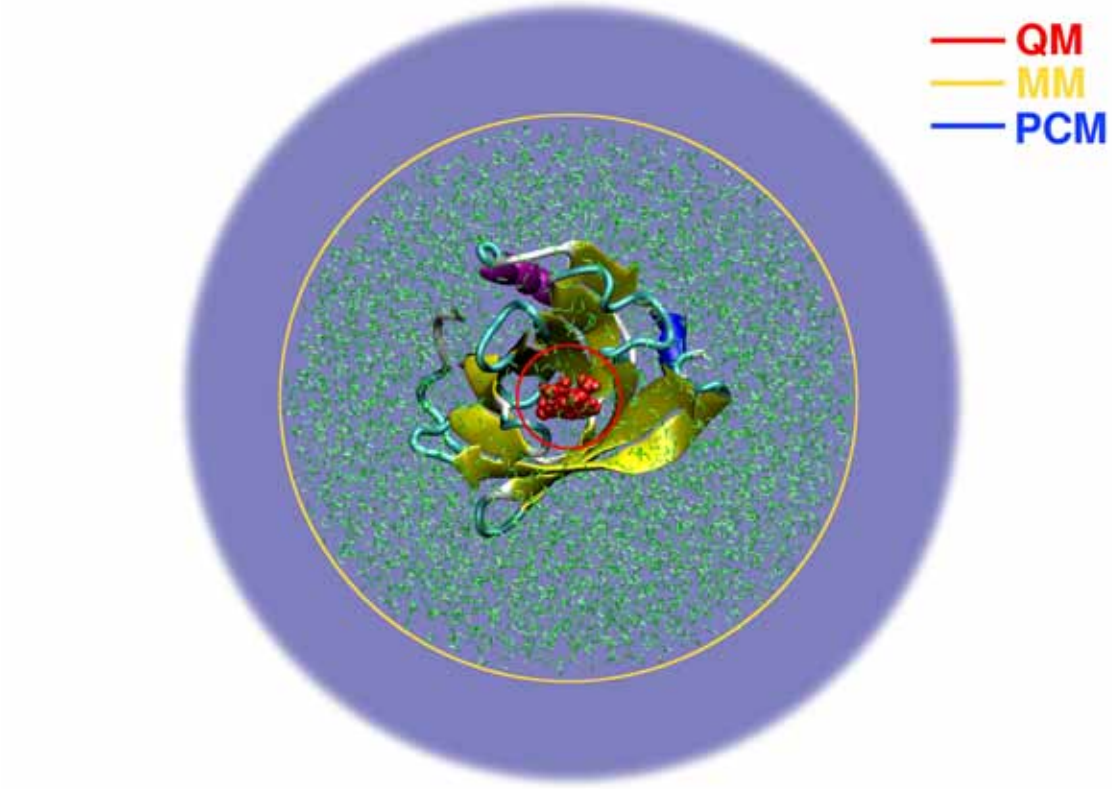
These quantum-classical approaches have made it possible to develop applications in a wide variety of applied research areas in chemistry, biotechnology, biomedicine and materials science. Thus, in addition to interpreting experiments, we have now come close to the point of true predictive modelling in which we can use computer models to design compounds that have properties that we require. E-Science has been, and will be, the active motor that promotes this development.

This article firstly reviews some work by the Division of Theoretical Chemistry and Biology at the KTH Royal Institute of Technology (KTH) in the area of multiscale modelling, which essentially concerns its "third dimension", that is, the area of general light-matter interactions and their applications to properties and spectroscopy. This includes multiscale modelling work on two- and three-level (namely QM/MM and QM/MM/PCM, where PCM stands for "polarizable continuum model"). We then report on research that concerns the development of programs including multi-QM cores in homogeneous and heterogeneous environments. Some of the applications of this research cover various types of spectroscopy, as well as the linear and non-linear properties of molecules in solution, on surfaces, in confined biological environments – such as proteins, green fluorescent proteins (GFPs), DNA, and cell membranes – and in combinations of such environments. Along with discussing the development of biomarkers, we also highlight modelling applications in the area of optical materials.

Development of Biological Markers

Biomarkers are measurable characteristics that indicate some physiological aspect of the state of an organism. They are mainly used for the non-invasive diagnosis of various diseases as they facilitate the effective and judicious selection of treatment options. However, many of the currently available imaging methods are based on MRI, PET and SPECT techniques which involve the use of biomarkers or diagnostic agents that contain a radionucleotide or heavy metals like Technetium and Gadolinium, and thus have harmful effects on the human body. Moreover, there is lack of knowledge of the structure-property relationships that could guide the design of more effective markers which would make it possible to detect diseases at even earlier stages.

We have undertaken work to find alternative diagnostic agents based on not-so-harmful one-photon and two-photon optical imaging techniques. In particular, we have been investigating various optical biomarkers that have a specific affinity to the key bio-structures (such as



Above: Pictorial representation for QM/MM/PCM approach: The biomarker is treated using QM-level theory, while the protein and solvents within a particular distance from the biomarker are described using MM-level theory and the remaining part of the system is described as a polarizable continuum.

Staff Focus

Henric Zazzi

Henric Zazzi joined PDC in January 2013 as an application expert in bioinformatics. He completed his Ph.D. in molecular biology at the Karolinska Institute in 2000, where he was involved in the Human Genome Project, and studied gene regulation of *igbp4*. Before coming to PDC, Henric worked as a bioinformatics programmer for AstraZeneca. Henric is currently involved in various bioinformatics and genetics projects and is closely collaborating with the Science for Life Laboratory, SeRC, and the Bioinformatics Infrastructure for Life Sciences (BILS).

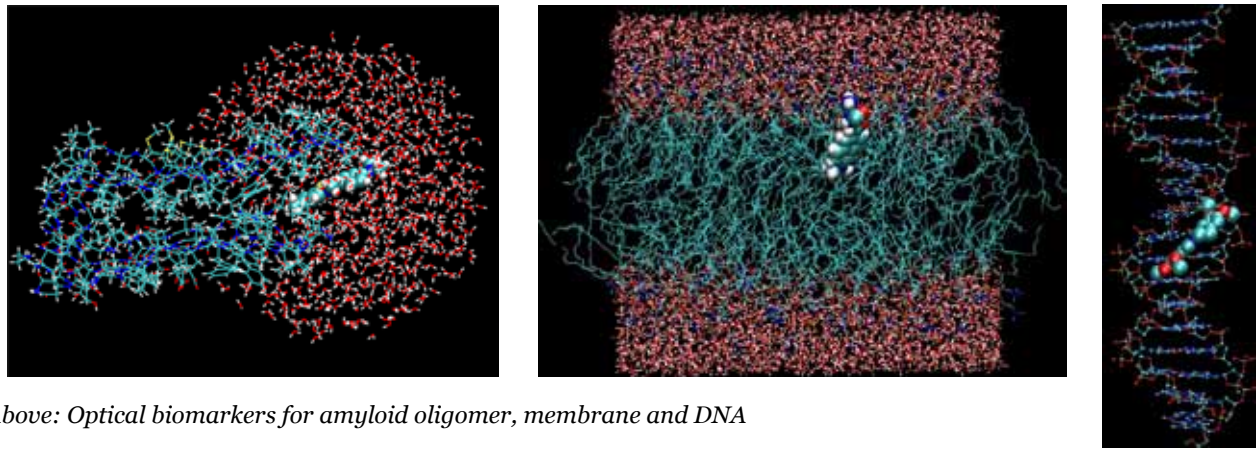
As a hobby, Henric likes to play handball and has played for Sorunda in the 2nd division in Stockholm. He also likes to play floorball (also known as indoor bandy), and is looking to join people who play on a regular basis. Besides an interest in ball sports, Henric has a rescue divers certificate and has been diving on the Great Barrier Reef in Australia. He is a competent amateur chef and likes to stand in the kitchen and experiment, which, according to his wife, is because he is not involved in practical laboratory work anymore.

peptide oligomers, fibrils, membranes, enzymes and DNA) and that can reveal information about their conformational nature, which often has a direct relation to healthy or diseased body conditions. It is worth mentioning here that the formation of amyloid and tau oligomers and fibrils in the extracellular and intracellular spaces has been designated as the cause of the so-called conformational diseases such as Alzheimer's and Parkinson's. Our research has also been concerned with biomarkers that are sensitive to different physiological parameters (like pH, and the distribution of metal ions and ionic strength within various organelles and extracellular space) as these also provide information about the condition of the human body.

The basic principle in diagnosis based on optical imaging is that the biomarkers used in the imaging process have a binding affinity with one or another of the aforementioned biomarkers and exhibit characteristic features in their absorption or emission spectra, which can be used to characterize or identify the specific conformational state of the target bio-structure. Thioflavin-T is a popular example of such a biomarker. It has characteristic absorption spectra and increased fluorescence intensity in the fibril bound state.

Our major research efforts deal with modelling the structure, dynamics, properties and binding affinity of such biomarkers in explicitly-treated solvents or bio-environments [2–5]. For computing the optical properties of the biomarkers, we rely on the most sophisticated hybrid quantum mechanics and molecular mechanics approaches that account for electrostatic and mutual polarization interactions between the biomarkers and the bio-environment subsystems. Our aim is to understand the working mechanisms of the currently available optical biomarkers, and to establish structure-property relationships for them, as well as to establish design principles for novel biomarkers for real-world diagnostic applications. In particular, we have studied a number of biomarkers that can be used to sense the dielectric properties of protein binding pockets, the presence of amyloid oligomers and fibrils, abnormalities in membrane structure, pH range and the presence of heavy metal ions. Detailed analyses have been performed to understand the changes in the structure, charge distribution, and micro-environment of the biomarkers in their bio-structure bound state, which in turn lead to the inherent specific properties of the bio-structure bound state.

A few of the biomarkers that we studied appeared to be promising, for example, two-photon biomarkers, where the images of disease-specific bio-structures can be obtained using radiation in the infrared region of the spectrum (which is less harmful and has a greater tissue-penetrating ability than, for example, optical or ultraviolet radiation). The use of two-photon biomarkers for medical



Above: Optical biomarkers for amyloid oligomer, membrane and DNA

imaging is associated with minimal health risks when compared to other currently available imaging techniques. Consequently, there is a demand for biomarkers with large two-photon absorption cross-sections which have a specific binding affinity to these disease-related bio-structures. We have recently investigated several two-photon biomarkers and have proposed a linear relationship between the electronic structure (in particular, the charge transfer) and the two-photon cross-section property relationships. This will be used to propose novel biomarkers with improved two-photon absorption characteristics. All these simulations have been carried out using the QM/MM response property module included in the DALTON suite of programs [6].

Optical Materials

We highlight our modelling work on optical materials by briefly reviewing activities in five subareas. This modelling has been of an applied nature: it has been conducted in conjunction with lab measurements, and has been made possible through external company or institutional support.

Optical power-limiting materials

A somewhat older example in this category concerns a project that began in 1998 and involved quantum modelling of optical power-limiting materials, namely materials which let light of normal intensity through but block high-intensity light (which is harmful to eyes and sensors). The methodological development has encompassed various quantum mechanical technologies for describing the coherent two-photon excitation process within the framework of wave function and density function theories. The method development has also involved models for vibronic and solvent contributions to the multi-photon excitations, as well as for pulse propagation based on the solution of the classical wave equations. Nowadays such calculations are carried out in the QM/MM framework, where the chromophores themselves constitute the QM part and the surrounding "glass" is treated with expedient force fields. Work on applications has been carried out in collaboration with (and with support from) the Swedish Defence Nanotechnology Programme through their project on Sensor Protection. This project demonstrates that sought-after materials with crucial properties can actually be predicted directly by modelling. We were able to derive metallo-porphyrin compounds and a particular platinum alkynyl complex (both of which fulfil the protection requirements) completely "in-silico". This project is also a good example of how work originally driven by pure academic curiosity - here, the development of theories and algorithms for the calculation of coherent two-photon excitation - can be useful in a technological breakthrough. For further reading, we recommend [7, 8].

Phosphorescent organic light-emitting diodes

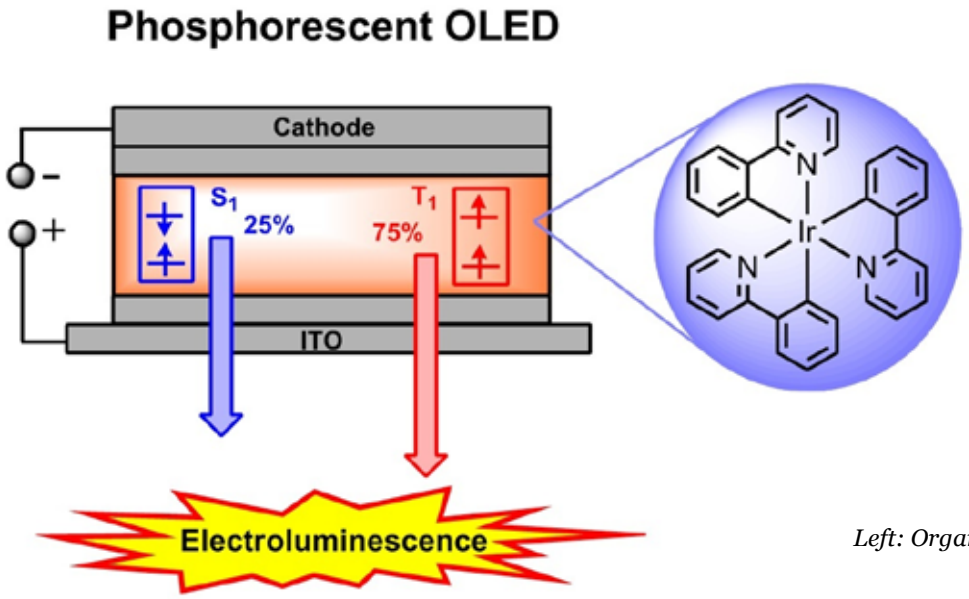
The phosphorescence process is another example of a non-linear property where we derived a unique methodology and software and which has become of unexpected practical interest. Our recent work in the area connects to organic light-

emitting device (OLED) technology, which has found numerous applications in the development of solid state lighting, flat panel displays and flexible screens. The first OLEDs were commercialized as flat displays in the form of organic polymers in which only singlet excitons emitted light. As spin-orbit coupling (SOC) effects are very weak in organic molecules, electrically generated triplet excitons (in the ratio 3:1 triplets to singlets) are not emissive. The full use of all singlet and triplet excitons, and complete internal quantum efficiency in OLEDs, are in principle possible when SOC effects provide strong mixing of spin multiplets. In fact, this occurs in complexes of heavy elements (such as Ir or Pt) with organic ligands when the phosphorescent triplet state emission harnesses all electro-generated excitons. This approach has begun to develop, and the first such phosphorescent red OLEDs have already been used in mobile phones and in OLED televisions, which are now undergoing commercialization. We have conducted a number of modelling studies with our unique phosphorescence code (as part of DALTON) and optimized a number of iridium-based complexes with respect to ligand substitution, and have found basic structure-property relations for these phosphorescent compounds with applications as OLEDs. The work is highly cited; a recent review article covering our work has appeared in [9].

All-organic electro-optical modulators

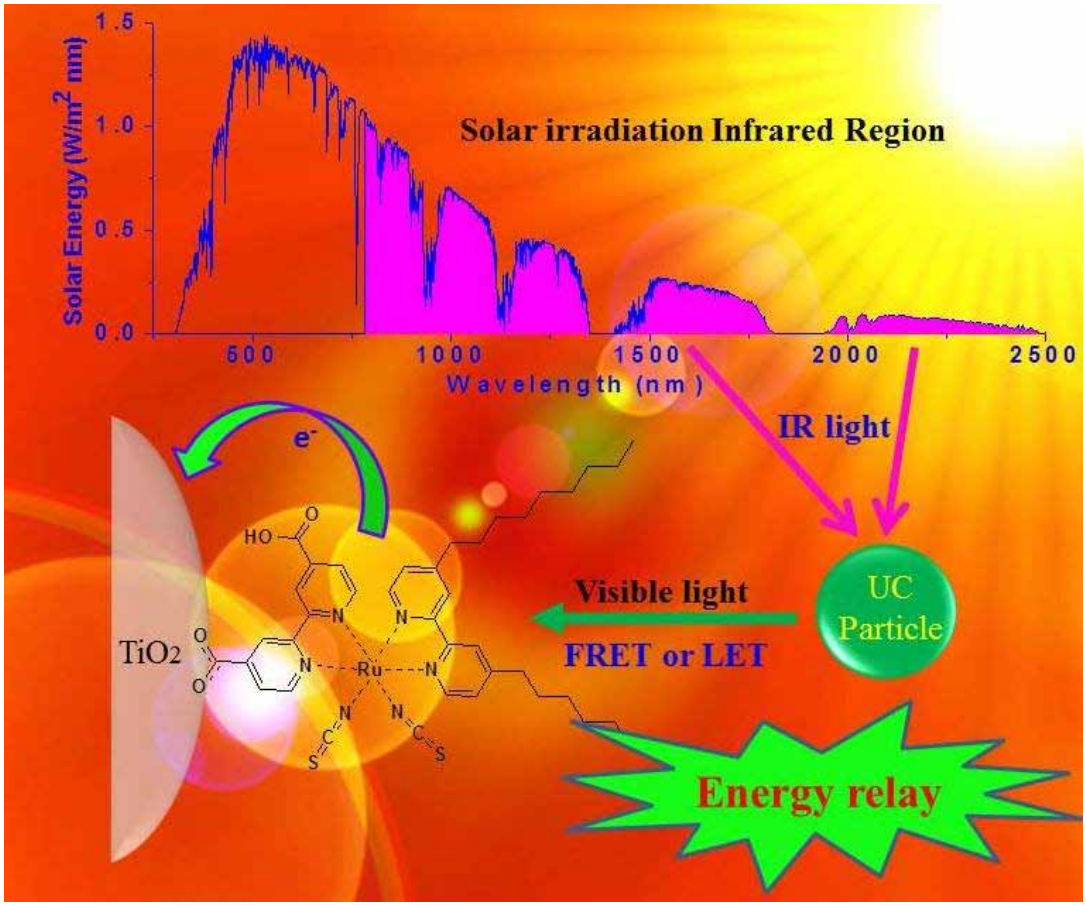
The design of electro-optical modulators is a third example that shows where our multiscale modelling of non-linear effects has been of practical use. Hybrid inorganic-organic and polymeric framework materials have an enormous chemical and structural diversity, and numerous strategies for synthesizing such materials have been reported during the past decades. New breakthrough applications based on hybrid nano-structures are now starting to emerge within the information and communications technology (ICT) sector.

We have been involved in modelling the design of the materials underlying such devices, in particular, dendrimer and organometallic-inorganic hybrid materials. This work has been undertaken in an integrated collaborative project with researchers from polymer chemistry and the Kista Photonic Research Center at KTH. As a result of this, we have proposed new materials with engineered nanostructures, which would allow us to control the properties of the substances in order to optimize them for use in electro-optical applications in photonics for ICT. We have, in particular, focused on dendrimer-chromophore complexes with unique self-poled electro-optic properties based on the second harmonic generation (SHG) effect, and which may become superior to current electro-optical modulators. In the case of these complexes, the SHG effect is



Left: Organic light-emitting device

Below: Upconversion nanocrystals for energy harvesting



obtained by treating the chromophore as a QM molecule, while the full dendrimer and water environment is handled using the expedient MM force fields. This combination of approaches makes this area very suitable for the kind of multiscale modelling we are discussing in this article. Further information on this particular topic can be found in [10].

Nanotechnology for solar energy harvesting

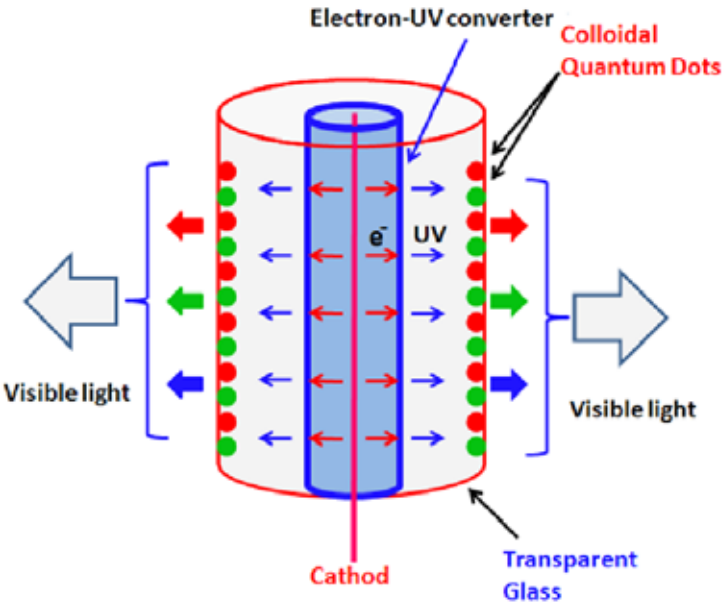
Our work in this area has been supported by the Joint Institute for Nanotechnology for Global Energy (JINGLE) which is a research centre at KTH. Here we have focused on the outstanding properties of quantum-confined nanoparticles (otherwise known as quantum dots or QDs), and modelled key processes in QD-sensitized solar cells. Through so-called "wave function engineering", we have proposed a set of design principles for composite QDs that have several semiconductor crystal shells – these are known as core-shell quantum dots. We have also investigated the effect of relieving the quantum confinement in one dimension, which gives us quantum rods, rather than dots. In addition, we have looked into several aspects that are special for QDs compared to molecular dyes, such as the role of multi-electron generation and Auger quenching effects. The use of so-called upconverting nanocrystals made up of combinations of lanthanide elements is particularly interesting. We have shown that these nanocrystals can be effective for harvesting sun light in the infrared region, where about 50% of the solar energy resides, but where no effective means of harvesting has been available hitherto. Our research has been published in a number of international journal articles, and has achieved very high visibility. It is clear that solar cells are complex entities and that modelling hardly can predict their efficiency as a whole. However, different sub-processes involved in the actions of energy harvesting can successfully be addressed, thus assisting development in this important area of renewable energy. We refer to the articles [11–13] for further details.

Stand-off resonant Raman detection of explosives

We have developed novel theories, algorithms and software (integrated in DALTON) for analysing the Raman process at resonant conditions. With the help of this software, one can use Resonant Raman detection as an approach for solving the low cross-section problem of Raman spectroscopy in standoff detection without needing to acquire any samples. The novel formulation has enabled us to define and patent an "Enhanced Resonance Raman Spectroscopy" [Patent WO 2010112531, US Prov. Patent S-8810]. We have also developed an associated software implementation of the microscopic expressions connecting molecular signatures and Raman scattering intensities, which includes conditions of electronic resonances. Adopting this so-called resonant-convergent complex polarization propagator method, we have been able to efficiently and accurately predict absolute resonant Raman intensities entirely from first principles. We have thereby made it possible to use Raman detection under resonant conditions for a broad variety of new applications, in particular, standoff detection of foreign substances at ultra-low concentrations. This approach departs substantially from conventional concepts and program designs, and, in doing so, it provides an efficient fingerprint identification method for foreign substances through their resonant Raman signals, as well as the necessary information (notably the absolute cross-sections) for designing an optimal technical system for each particular scenario of stand-off detection. With the absolute cross-sections at hand, we can precisely design the instrument, the setup, and cost, for any given concentration level that one needs to detect. This knowledge is unique in the world and has consequently been patented. The product has general applicability, but carries particular merits with respect to conventional techniques for standoff detection. We have received company support for this effort. For further details, see reference [14].

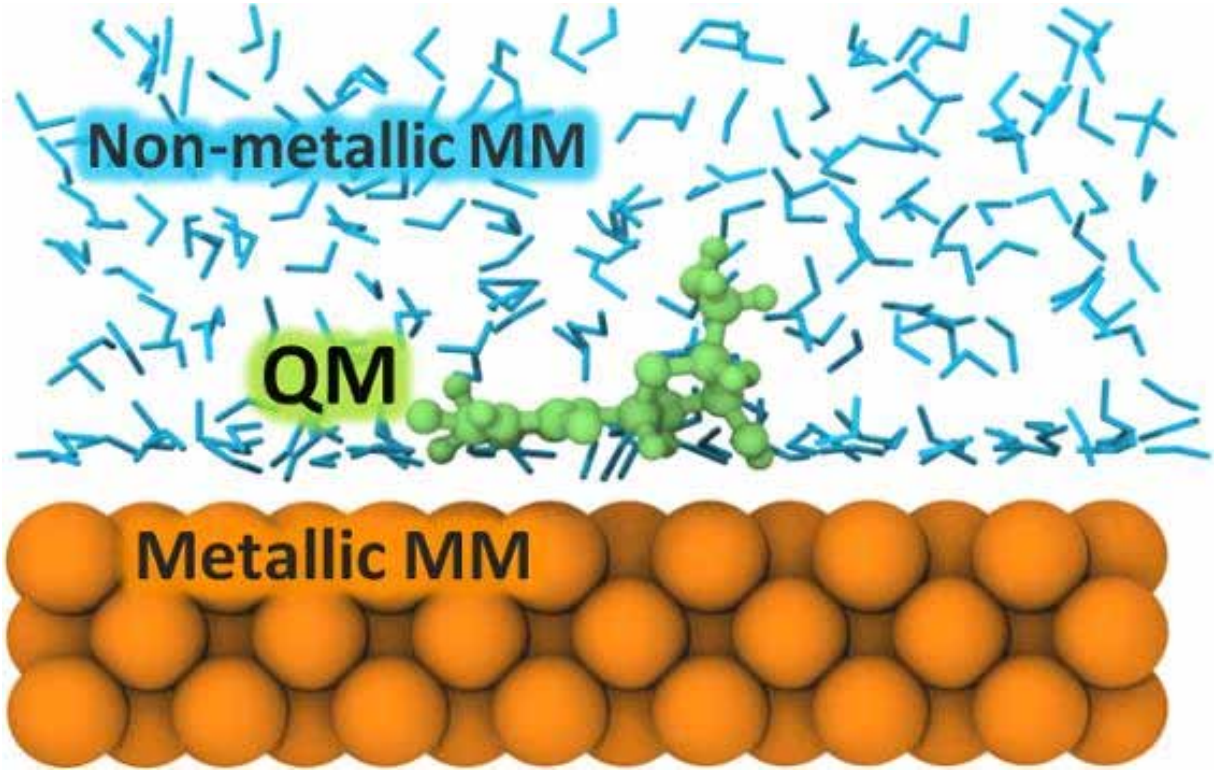
Quantum nano-materials for energy-efficient lamps

Our final example of modelling optical material is a project supported by companies and the Swedish Energy Agency that looked at the development of quantum dots as new phosphors in field emission lamps. We know that in conventional light sources, such as fluorescent tubes, electrons are emitted from a hot cathode to excite the mercury atoms that fill the glass tube so they emit UV radiation, which in turn excites the phosphors on the inner surface of the tube so they generate visible light. Thus light is produced in a two-step process: electrons → UV radiation → visible light, with mercury as an intermediate wavelength converter. In a field emission light source, electrons are emitted from a multitude of sharp tips under an intensive electric field without heating. The field emission process is an efficient way of obtaining an electron beam that passes from a solid surface to a vacuum. These field-emitted electrons excite a layer of phosphors to emit visible light. Thus the light is produced in a one-step process, electrons → visible light. A field emission light source is free of mercury, and the phosphors that are used are luminescent phosphors. In this ongoing project, we seek to develop quantum dots to replace conventional phosphors either in a one- or two-step process.



Right: Schematic structure of a Field Emission Device based on quantum dots

Below: A QM/CMM approach



E-Science Aspects

The DALTON program suite [6] is produced by a large joint project that started out developing the software specifically for quantum chemistry, and then extended the programs to include quantum-classical multiscale chemistry. The project involves researchers from Sweden, Norway and Denmark and has been running for the last 30 years. The latest release of the DALTON program, named DALTON2013, includes a QM/MM module for modelling electronic structure, and the linear and non-linear properties of molecules in complex environments (like proteins or membranes). This part of the DALTON program, along with our recently developed QM/CMM (capacitance molecular mechanics) approach [15] for heterogeneous metallic/non-metallic environments, form the key element of the multiscale modelling toolset used for modelling optical properties of biomarkers and nanomaterials. QM/MM calculations are one of the most computationally-demanding tasks encountered in quantum chemical modelling of complex systems, as they typically require several hundred electronic structure and response theory calculations in order to determine a molecular property reliably for an ensemble of system conformations. Thus, the QM/MM calculations for each conformation must be carried out as fast as possible, and a robust strategy for property sampling over conformations must be employed to optimize the usage of the high-performance computing (HPC) resources. In recent years, we have carried out extensive work on improving the performance and scalability of the DALTON program, and, for that purpose, have participated in larger network projects – like the ScalaLife FP7 project (http://www.scalalife.eu) – and in the PRACE community software initiative. Several of these code improvements – like reducing the I/O footprint during density functional theory (DFT) integration or hierarchical multilayered MPI parallelization schemes – have enabled us to improve the scalability and usability of the DALTON program on modern HPC resources. Furthermore, these improvements made it possible for us to obtain a PRACE Tier-1 allocation for our MBIOMARK research project

(which is driven by the DALTON program) in which extensive studies of magnetic biomarkers have been carried out using the QM/MM approach. These computations would have not been possible without the generous time allocation from PRACE. Furthermore, apart from enabling us to carry out computational investigations of biomarkers, the access to PRACE resources (Cray XE6 supercomputer Monte Rosa, Swiss National Supercomputing Centre, Switzerland) allowed us to test parallelization improvements in the DALTON program developed in conjunction with the ScalaLife FP7 project and the PRACE community software initiative, in which we participated as DALTON program developers. The main technical challenge for more effective usage of PRACE HPC resources for the "in silico" design of biomarkers is the development of an automated management system for computational processes. Currently, together with the MAPPER FP7 project (http://www.mapper-project.eu), we are developing an automated scheme for the smart management of computation resources during QM/MM calculations of molecular properties using integrated approaches. We hope to test this software on various PRACE resources in the near future.

References

[1] Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

[2] Murugan, N. A.; Kongsted, J.; Rinkevicius, Z.; Ågren, H. Color modeling of protein optical probes. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1107–1112.

[3] Li, X.; Rinkevicius, Z.; Kongsted, J.; Murugan, N. A.; Ågren, H. Binding Mechanism and Magnetic Properties of a Multifunctional Spin Label for Targeted EPR Imaging of Amyloid Proteins: Insight from Atomistic Simulations and First-Principles Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 47664774.

[4] Murugan, N. A.; Olsen, J. M. H.; Kongsted, J.; Rinkevicius, Z.; Aidas, K.; Ågren, H. Amyloid Fibril-Induced Structural and Spectral Modifications in the Thioflavin-T Optical Probe. *J. Phys. Chem. Lett.* **2013**, *4*, 70–77.

[5] Murugan, N. A.; Apostolov, R.; Rinkevicius, Z.; Kongsted, J.; Lindahl, E.; Ågren, H. Association Dynamics and Linear and Nonlinear Optical Properties of an N-Acetylaladanamide Probe in a POPC Membrane. *J. Am. Chem. Soc.* **2013**, *135*, 13590–13597.

[6] Aidas, K. et al. The Dalton quantum chemistry program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, 00–00, DOI: 10.1002/wcms.1172.

[7] Norman, P.; Ågren, H. First-principle quantum modeling of optical power limiting materials. *Journal of Computational and Theoretical Nanoscience* **2004**, *1*, 343–366.

[8] Baev, A.; Welinder, P.; Erlandsson, R.; Henriksson, J.; Norman, P.; Ågren, H. A Quantum Mechanical-Electrodynamical Approach to Nonlinear Properties: Applications to Optical Power Limiting with Platinum-Organic Compounds. *Journal of Nonlinear Optical Physics & Materials* **2007**, *16*, 157–169.

[9] Minaev, B.; Baryshnikov, G.; Agren, H. Principles of phosphorescent organic light emitting devices. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1719–1758.

[10] Ågren, H.; Tu, Modeling of Non-linear Optic Properties of Guest-Host Systems. *Nonlinear Optics and Quantum Optics* **2006**, *35*, 1–3.

[11] Ning, Z.; Tian, H.; Qin, H.; Zhang, Q.; Ågren, H.; Sun, L.; Fu, Y. Wave-function engineering of CdSe/CdS Core/Shell quantum dots for enhanced electron transfer to a TiO2 Substrate. *J. Phys. Chem. C* **2010**, *114*, 15184–15189.

[12] Ning, Z.; Yuan, C.; Tian, H.; Hedström, P.; Sun, L.; Ågren, H. Quantum Rod-Sensitized Solar Cells. *Chem. Sus. Chem.* **2011**, *4*, 1741–1744.

[13] Yuan, C.; Chen, G.; Prasad, P. N.; Ohulchanskyy, T. Y.; Ning, Z.; Tian, H.; Sun, L.; Ågren, H. Use of colloidal upconversion nanocrystals for energy relay solar cell light harvesting in the near-infrared region. *Journal of Materials Chemistry* **2012**, *22*, 16709–16713.

[14] Mohammed, A.; Ågren, H.; Norman, P. Time-dependent density functional theory for resonant properties: resonance enhanced Raman scattering from the complex electric-dipole polarizability. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4539–4548.

[15] Rinkevicius, Z.; Li, X.; Sandberg, J. A. R.; Mikkelsen, K. V.; Ågren, H. A Hybrid Density Functional Theory/Molecular Mechanics Approach for Linear Response Properties in Heterogeneous Environments. *J. Chem. Theory Comput.* **2014**, *10*, 989–1003.

Above: Milner arrives at PDC

Above: Hmmm... how do we get it out of the box?

Above: Unpacked and ready to be moved into position

Welcome Milner! New Neuroinformatics Computer System Installed at PDC

Gert Svensson, PDC

As we reported in the previous issue of the PDC newsletter, PDC has been procuring a new supercomputer system (for neuroinformatics research) on behalf of a consortium of researchers from KTH, the Karolinska Institute and the International Neuroinformatics Coordination Facility (INCF).

The procurement was won by Cray with an XC30 system, with the specifications listed below. The installation of the system started on the 3rd of February 2014. For those of you who wonder what it looks like to actually install a supercomputer, the photos shows the computer being delivered to PDC, followed by unloading the computer from the crate and then starting the process of installing the computer in the PDC computer hall.

The new system was fully accepted on the 27th of March after a trial period consisting of one month in normal usage. The system is named "Milner" after the British neuropsychologist Brenda Milner (http://en.wikipedia.org/wiki/Brenda_Milner). For information about using Milner, see http://www.pdc.kth.se/resources/computers/milner.

Facts about Milner

- Cray XC30 with 120 nodes
- Each node has two 2.5 GHz Intel Ivy Bridge CPUs with 10 cores
- Peak performance: 48 TFLOPS
- Total memory: 3.7 TB (32 GB per node)
- Interconnect: Cray Aries
- Dedicated Lustre file system 150 TB

Simulations of Main-Ion and Impurity Transport in Tokamak Experiments

Andreas Skyman, Department of Earth and Space Sciences, Chalmers University of Technology

In volume one of the 2013 PDC newsletter, we described our ongoing high performance computing research into modelling transport processes in the plasma in tokamaks (which are used for controlling fusion processes to generate clean energy), and understanding the effects of turbulence in the plasma (so they can be reduced, so as to produce energy more cheaply). During 2013, our main focus was on the effects of the magnetic geometry – namely, the shape of the cross-section of the fusion plasma – on the particle transport in tokamaks. It is known that elongated cross-sections have favourable stability properties, but the effects on micro-turbulent core transport are less studied, particularly in relation to impurities.

The parameters and geometry we chose to study were taken from an impurity injection experiment at the Joint European Torus, JET. For these parameters, we found from simulations with the Gyrokinetic Electromagnetic Numerical Experiment (GENE) code (see http://www.ipp.mpg.de/~fsj/gene) that the choice of magnetic geometry was very important, with a destabilisation and shift to higher wave numbers of the driving instability. By comparing the diffusivities and flux-spectra for the nonlinear simulations with those of a previous study, the effects could be attributed mainly to a reduction in the convection, and the behaviour of the lower charged impurities could be explained by the wave number dependence of the part of the convection driven by the temperature gradient.

Several additional degrees of realism were included, such as collisions, plasma rotation and the presence of background carbon at moderately high particle densities, but proper treatment of the geometry was seen to be the most important factor. Though a considerable decrease in impurity accumulation in the core was predicted for light (as well as heavy) impurities, compared to the scenarios with simplified geometries, even the most advanced scenario still showed an inward net transport of impurities. The results were published in Nuclear Fusion in January 2014 [1].

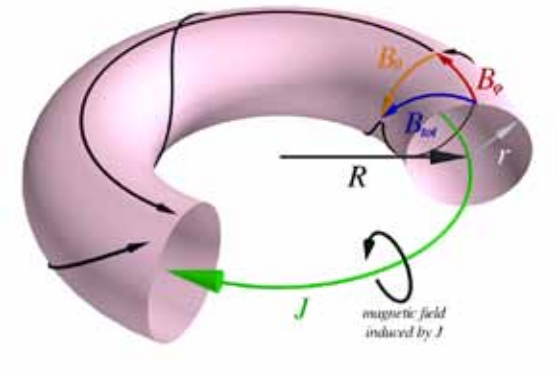
After the conclusion of this study, we shifted our focus to main-ion transport. From our studies during 2011 and 2012 [2, 3], we learned that the background density profile can greatly affect the transport of impurities. We therefore wanted to investigate the parametric dependence of this in connection with the impurity transport. Results showed that there are scenarios where a self-consistent treatment of the impurity and background transport will give significantly different results to treating the background density gradient as a parameter. This may be of particular importance for simulations of experimental data. For 2014, our main focus is on such studies, as part of a European collaboration with the GENE developers, amongst others. To this end, time has been granted on several supercomputer systems: Lindgren (PDC), JUROPA (FZJ, Germany) and HELIO (IFERC, Japan). The results from the initial study have been submitted to Physics of Plasmas for consideration.

References

[1] Skyman, A.; Fazendeiro, L.; Tegnered, D. et al. Effects of the equilibrium model on impurity transport in tokamaks. *Nuclear Fusion.* **2014.** 54 (1) s. 013009.

[2] Skyman, A.; Nordman, H.; Strand, P. Impurity transport in temperature gradient driven turbulence. *Physics of Plasmas.* **2012**. 19 (3) s. 032313.

[3] Skyman, A.; Nordman, H.; Strand, P. Particle transport in density gradient driven TE mode turbulence. *Nuclear Fusion.* **2012.** 52 (11) s. 114015.



Above:Tokamak geometry

Particles and Turbulence in a Bent Pipe

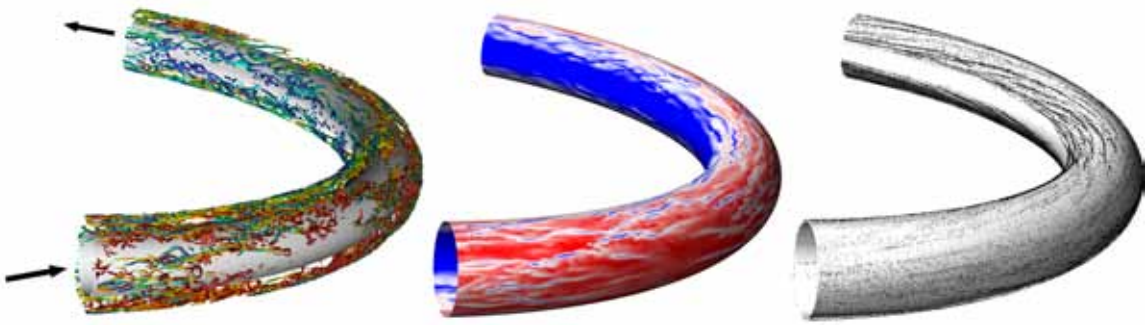
Philipp Schlatter and Azad Noorani, Swedish e-Science Research Centre (SeRC), KTH Mechanics

In volume one of the 2012 PDC newsletter, Philipp Schlatter and George El Khoury reported on large-scale simulations of the turbulent flow in straight pipes, such as oil pipelines or drinking straws. In the meantime, Azad Noorani has extended that work, and has been considering the turbulent flow in a pipe that is bent with a curvature ratio (pipe radius divided by the bend radius) of 10%. Even though this curvature appears to be relatively insignificant, the effects on the flow are profound: due to an imbalance between the (geometry-induced) centrifugal forces, strong secondary motions develop in the pipe cross-section; these motions break the rotational symmetry of the pipe flow, and lead to increased cross-sectional mixing of both mass and momentum. In fact, it turns out that these so-called Dean vortices persist for quite some distance in the flow, and are thus very relevant in many applications, be it in industry (for example, ventilation systems, piping for combustion engines, power plants and so forth), academic research (understanding of secondary motions), or turbulence modelling (considering how current models deal with comparably weak secondary flow).

The figure below illustrates several aspects of our simulations; all the data was calculated on the PDC system Lindgren using a few thousand cores running the Nek5000 code, and the visualisations were produced on the PDC post-processing machine Ellen (using VisIt). The leftmost section of pipe in the figure highlights the action of wall turbulence in a bent pipe – it shows the vortical structures coloured by the axial velocity. One can see the intricate details of turbulence. The middle panel shows a pseudo-colour representation of the heat-flux distribution along the walls. It becomes clear that the pipe surface is by no means isothermal, and there are significant fluctuations that might lead to "hot spots" and thermal fatigue in, for example, heat-exchanger applications.

In the present simulations, we have also included Lagrangian particles, namely small heavy particles that are advected by the fluid. The interesting aspect here is that these particles are affected by the turbulence, by the Dean vortices and also by their own inertia. This leads to the distribution shown in the right-hand panel. Along the side walls, organised so-called particle streaks are formed, which curve towards to inner side, which is then – quite unexpectedly – completely empty of particles. Following individual particles in time reveals that each trajectory resembles a helix around the Dean vortices. A reliable prediction of the particle concentration is crucial for many applications, for example, for pharmaceutical devices or membrane ultra-filtration systems.

Even though the illustrated simulations were performed in an idealised setting with perfect surface conditions, and uniform curvature, the results are relevant in many ways. One particularly interesting future development is the development and calibration of new turbulence models capable of capturing secondary flows. Such new models could then be included in computational fluid dynamics (CFD) software and thus improve engineering predictions in the future.



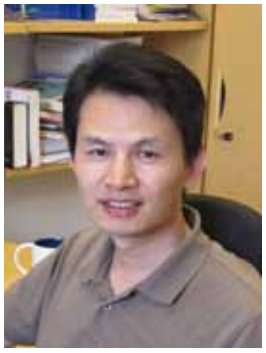
Above: (From left to right) Wall turbulence in a bent pipe, heat-flux distribution along the walls of a bent pipe, and particle concentration along the walls of a bent pipe

PDC Support: Application Experts

Genet Edmondson and Jonathan Vincent, PDC

In the previous edition of the PDC newsletter, we gave you an overview of the three different types of support that PDC offers to help people use our high-performance computing (HPC) resources, and introduced you to the lads in first-line support. (By the way, since then, those lads have been joined by a lass, Filippia Zikou – you can read about Filippia's background elsewhere in this newsletter.) Now, you may remember from the earlier article that PDC also offers second-line help in the form of application experts. These people are researchers who have expertise in a particular area of science or engineering and also have a lot of experience with parallel programming within that area. This makes them ideally suited to help when others need help with programs in specific research areas.

As you are probably aware, there are two main areas with using HPC resources where people often need help. In some cases, researchers may have code or an algorithm that is sequential, and which consequently needs to be parallelized, that is, adapted so it can run on multiple processors in parallel. It is not always obvious how to do that in the best way, so some expert advice can come in handy! Other researchers may find themselves with code that is already set up to run on a parallel computer system, but the code may not be efficient. As an example, when multiple proces-



Jing Gong

Scientific Computing

(Ph.D., Uppsala)

Previous research and work

- *industrial FEM code development*
- *computational fluid dynamics in conjunction with the Department of Mechanics, Scania, and the Department of Aeronautical and Vehicle Engineering, KTH*

Current work

- *computational fluid dynamics*
- *CRESTA and PRACE projects*

Can help PDC users with

- *computational fluid dynamics*
- *general scientific computing problems*
- *hybrid MPI/OpenACC code development*

Jonathan Vincent

Theoretical Physics

(Ph.D., Exeter)

Previous research and work

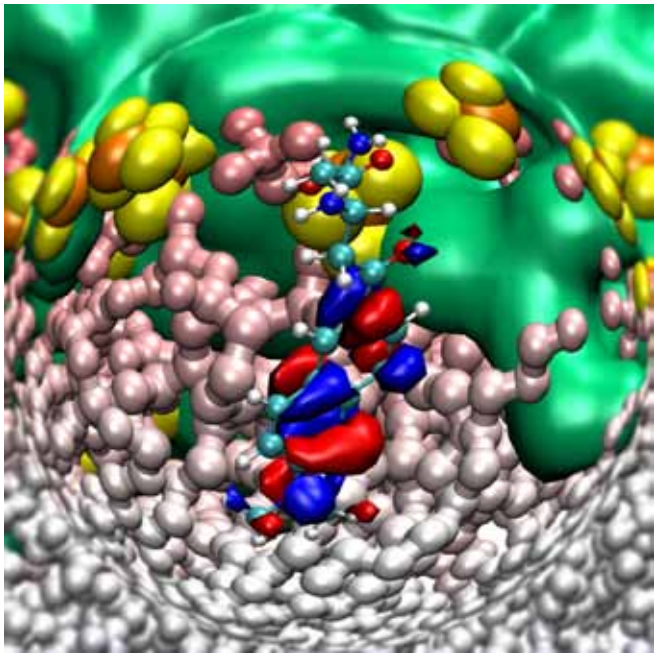
- *computational chemistry (Gothenburg, Leiden, and Uppsala)*
- *climate modelling support (UK)*

Current work

Head of PDC user support

Can help PDC users with

- *general information about using PDC resources*
- *MPI*
- *parallelisation*



Above: Hybrid QM/MM simulation of acrylin interactions with lipid membrane

sors are performing calculations all at the same time, and those results then have to be communicated to some of the other processors so that the next lot of calculations can be performed, it is possible for "bottlenecks" to occur. These are a bit like traffic jams inside the computer system, where processors end up stuck waiting for information to arrive from other processors. As there are many, many processors in HPC systems, what you can end up with is a lot of processors sitting around doing nothing much for significant amounts of time. When this happens, you are obviously not making the best use of the supercomputer system! Therefore optimizing code to take care of this (and the other potential problems that can occur when scaling problems up to run on multiple processors) can be of real benefit and may, for example, considerably increase the size of problem that can be modelled. As it is not necessarily obvious how to effectively optimize different types of code or how to optimize them to run on a particular HPC system, having someone to lend a helping hand (who has experience in the type of problem you are working on) can be a real bonus.

Rossen Apostolov

Computational Chemistry (Ph.D., Osaka)

Previous research and work

- *optimizing GROMACS MD software and porting algorithms to GPGPU processors*

Current work

- *code development of GROMACS*
- *development of formats and libraries for efficient storage of molecular trajectory data*
- *code analysis and profiling tools*
- *advanced techniques for visualization of molecular trajectories*

Can help PDC users with

- *advanced support in all areas related to molecular dynamics (MD) simulations and GROMACS*
- *best practises for software development (including building systems and version control)*
- *code analysis and profiling tools*

Radovan Bast

Computational/ Theoretical Chemistry (Ph.D., Strasbourg)

Previous research and work

- *computational/theoretical chemistry including many years contributing to the Dalton and DIRAC programs*
- *user support*

Current work

- *code optimization/parallelization in collaboration with the company PGS*
- *optimization of the Dalton code*
- *writing a new QM code for GPGPUs/accelerators*

Can help PDC users with

- *computational/theoretical chemistry*
- *relativistic electronic structure theory*
- *molecular physics*
- *scientific code development*
- *code parallelization and optimization*
- *compilation*
- *collaborative tools for code development*

So, how can PDC's application experts help you? If you need general help with optimisation and parallelisation, or with MPI (the standard message passing interface used for writing parallel programs), then Jonathan Vincent would be your first port of call. If you help with theoretical and computational chemistry, that is Radovan Bast's main area of expertise, particularly the Dalton and DIRAC programs. For the quantum and molecular dynamics side of computational chemistry, and the GROMACS program, you are best off with Rossen Apostolov. If your research is more in the area of molecular biology, bioinformatics and genetics, then Henric Zazzi can help you with his broad background in the biomedical field. However, if your problem is specifically with brains, Mikael Djurfeldt is the person to turn

Mikael Djurfeldt
Computer Science
(Ph.D., KTH)

Previous research and work
- *computational neuroscience*
- *have contributed to/directed open source software projects such as Guile, GOOPS, MUSIC and CSA*

Current work
- *research in neuroinformatics*
- *working with INCF*
- *developing MUSIC and CSA*

Can help PDC users with
- *NEST, NEURON, PyNN, MUSIC, CSA*
- *using Milner – PDC's new neuroinformatics system (https://www.pdc.kth.se/resources/computers/milner)*

Henric Zazzi
Molecular Biology
(Ph.D.,
Karolinska Institute)

Previous research and work
- *codeveloped Geisa software for genetics analysis*
- *bioinformatics research at AstraZeneca*
- *molecular medicine and genetics at Arexis*

Current work
- *developing a genetic analysis tool https://github.com/menzzana/geisa*
- *working on a proteomics tool for data extraction/integration https://github.com/percolator/pout2mzid*
- *collaborating with SNIC Cloud on a protein analysis pipeline*
- *part of SNAC working group*

Can help PDC users with
- *genetics*
- *molecular biology*
- *bioinformatics*
- *development for biological analysis tools*

to, at least in terms of computational neuroscience. (Unfortunately Mikael and his colleagues have not yet solved the problem of misplacing one's glasses or car keys as one gets older!) Finally, if your research is related to computational fluid dynamics (CFD), Jing Gong can help you to adapt commercial or open source code for CFD applications to run efficiently on HPC systems.

If you are doing research in an area that seems to lie outside those of our "resident" application experts at PDC (Please check the tables for more details about their areas!) and need help with optimising or parallelising code, we invite you to contact us in the usual way (via support@pdc.kth.se) as we have connections with many other research groups and supercomputer centres and will do our best to put you in contact with someone who can help you. For example, we maintain close connections with the application experts at the other HPC centres in Sweden that are funded by the Swedish National Infrastructure for Computing (SNIC). The photo below shows the application experts at the SNIC Application Expert meeting that was held at PDC in March 2014. All of these SNIC experts are available to help researchers using SNIC resources. You can also keep an eye on the Events section of the PDC website (http://www.pdc.kth.se/events) where we advertise various seminars and best practise meetings designed to provide you with additional help and information.

Above: Swedish application experts at SNIC Application Expert Meeting, PDC, 12-13 March 2014

SNIC GPU Project News

Michael Schliephake, PDC

The Swedish National Infrastructure for Computing, SNIC, started a three-year pilot project for the evaluation of graphics processing unit (GPU) technology in 2012. Two experimental clusters equipped with NVIDIA graphics cards have been made available to Swedish scientists through the project. This started with the establishment of the cluster Zorn at PDC in the beginning of 2012 and continued with the cluster Erik at the Lunarc Centre for Scientific and Technical Computing, Lund University, in 2013. Many research groups have used these clusters to evaluate GPU technology as well as development tools in the preparation of applications for the use of other GPU computing resources. Using this equipment, PDC offers training in GPU computing. Its accelerator programming courses for graduate students have been well attended for several years by participants from all over Sweden. Both the activities of several research groups at KTH and the support PDC is providing have been recognized by NVIDIA through the establishment of a CUDA Research Center at KTH.

Several outreach activities took place over the last six months. The SNIC GPU Roadshow visited all of the SNIC computer centres during the autumn of 2013. Jonas Lindeman from Lunarc and Michael Schliephake from PDC presented information about GPU computing in general to interested researchers and also provided details regarding the support offered by SNIC and its infrastructure. In January 2014, SNIC organised an advanced user workshop that took place at KTH, where NVIDIA and Intel presented the latest technical developments in accelerator computing together with success stories from SNIC users. These events were met with wide interest both from people who are already actively using accelerator programming and from potential new users. The increasing demand for resources that support computing with GPU accelerators has been confirmed by feedback that we were given in discussions as well as from new users who have recently been attracted to our facilities.

B2SHARE Nordic

Customized Data Sharing for Nordic Research Communities

Carl Johan Håkansson, PDC

The B2SHARE Nordic project is collaborating with research communities in the Nordic countries to provide long-term storage and sharing of research data and associated metadata by using the EUDAT B2SHARE service. The project, which is funded by the Nordic e-Infrastructure Collaboration (NeIC), started in January and will run till the end of August.

Together with the research communities, B2SHARE Nordic is developing tailor-made metadata extensions for the EUDAT B2SHARE service and for the pre-customised "premium" services that are based on the B2SHARE software but that have additional features for communities with more extensive needs.

The project is designed to encourage and promote data sharing between colleagues and research groups, as well as cross-border collaboration within and between research communities, primarily in the Nordic countries, and also more generally on European and world-wide scales. The solutions that B2SHARE Nordic implements for different research communities may serve different purposes – some solutions are, for instance, well suited to open research or crowd-sourcing as they make it possible for amateur scientists and ordinary citizens, as well as research colleagues, to access or contribute to scientific data collections. Like EUDAT's B2SHARE, the Nordic project also aims to prevent the loss of research materials that normally fall outside the bounds of institutional data management by providing a service for long-term storage with persistent identifiers.

The B2SHARE Nordic project has been making good progress since January and the first community adaptation is already online: the BioBanking and Molecular Resource Infrastructure of Sweden (BBMRI) is now using an adapted metadata extension in B2SHARE. And there is more to come, as many other research projects have already expressed interest in collaborating with B2SHARE Nordic.

SNIC Cloud Infrastructure Project

Åke Edlund, PDC and HPCViz

The SNIC Cloud Infrastructure Project started on the 1st of January 2014 and will consist of two separate year-long phases. The project is being coordinated by Åke Edlund (from PDC and HPCViz) who is working with partners from four of the SNIC centres: PDC, the High Performance Computing Center North (HPC2N) at Umeå University, the Chalmers Centre for Computational Science and Engineering (C3SE) at Chalmers University of Technology, and the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) at Uppsala University.

The goal of the project is to create a sustainable and generic SNIC Cloud infrastructure (known as an "Infrastructure as a Service" or IaaS) to form a basis for SNIC Cloud projects, and to provide the necessary structure for running project-specific "Platform as a Service" (PaaS) services. After the initial two phases, the project will be aiming to scale up the cloud infrastructure in 2016.

The SNIC Cloud infrastructure will provide an elastic resource pool, where additional resources (in the form of virtual machine hosts) can dynamically join and leave the core resource.

User groups are being added in steps, starting with users from the SNIC Galaxy project and SciLifeLab (from PDC together with UPPMAX), Systems and Synthetic Biology (from C3SE) and Forest Remote Sensing (from the Swedish University of Agricultural Science, SLU, together with HPC2N). Other research groups have been contacted with the intention of expanding to include them in the course of the project.

The requirements of these user groups guide the project when it comes to technical decisions. Projects that are added either run directly on the SNIC Cloud, or use their own cloud resources in combination with the SNIC Cloud, bearing in mind that any cloud resources that are added to the SNIC Cloud do need to conform to the technical specifications decided within the SNIC Cloud project. Completed projects will provide information regarding the resources they used (including, for example, the number and types of virtual machines, the type and format of metadata, application scripts, package recipes, and project documentation), thus ensuring that the services can be reused by other projects later on.

Above: Panel discussion about exascale computing with (from left to right) Erik Lindahl, Mark Parsons, Yutong Lu and Erwin Laure at EASC2014

EASC2014

Stefano Markidis, HPCViz

PDC organized the Exascale Applications and Software Conference that was held in Stockholm on the 2nd and 3rd of April 2014 (http://www.easc2014.se). The goal of this conference was to bring together many of the developers and researchers involved in solving the software challenges of the exascale era. The conference focused on issues of applications for exascale and the associated tools, software programming models

Above: EPiGRAM poster presentation at EASC 2014

and libraries. The keynote speakers were Yutong Lu (National University of Defense Technology, China), Mitsuhisa Sato (University of Tsukuba, Japan), Erik Lindahl (KTH Royal Institute of Technology, Sweden) and Mark Parsons (EPCC, The University of Edinburgh, UK). The conference was attended by 82 people and comprised keynote talks, regular papers and hot-topic paper presentations, and poster and panel sessions about the exascale challenges. The proceedings of the conference will be published in Springer's "Lecture Notes in Computer Science".

Above: EASC2014 conference attendants in Stockholm

Python on PDC's HPC Resources

Michael Schliephake, PDC

The programming language Python has been available since the early 1990s. It has been a much-loved language from the very beginning, perhaps because it seems to make programming work much more fun. An enthusiastic, ever-growing developer and user community has contributed so many software packages since the 1990s that one application field after the other has been conquered. Recently, numerical simulation and big data analysis have been attracting a lot of attention. Nowadays, Python is not only being used as the glue that integrates software components or makes pre- and post-processing easier, but many software packages for compute-intensive problems are actually being developed in Python. Python's popularity is also indicated by the number of workshops and "Birds of a Feather" sessions at many large high-performance computing (HPC) conferences that feature Python-related topics.

Python has been available on PDC's HPC resources for a long time. We provide a wide range of important Python packages, for example, NumPy, SciPy, and matplotlib, to name just a few. We also offer help and support for users who need to install additional Python packages for their projects.

Our resources for Python users at PDC have recently been improved even further. Two comprehensive distributions of Python software for scientific computing were made available a little while ago: Anaconda and Canopy. The Anaconda distribution provides more than 125 Python packages, and the installation of Canopy comes with about 150 packages. By using these two products, PDC has been relieved from the process of compiling and orchestrating program collections, which can be rather complex sometimes. Instead of this, we have been able to focus our efforts on measures that provide uniform installations on all our HPC systems. As a result, our users can now run their Python-based programs on all our systems in the same way – interactively and in batch jobs. Moreover, Anaconda and Canopy are available for academic users at no additional cost. This means that our users from academia can use the same Python software on their desktop and laptop computers free of charge. Having the same working environment "at home" as in the computing centre means that moving programs and data to and from HPC systems at PDC is a piece of cake.

We hope that this unified environment and the broader supply of Python packages will contribute to more efficient workflows for our current users, as well as meaning that new users will find working on our HPC resources simpler and more attractive.

HPC Provides Competitive Advantage

Leif Nordlund, PDC

Although substantial progress has been made in recent years, the industrial use of high-performance computing (HPC) is not as advanced, widespread, or effective as it could be. Many countries around the world are trying to boost innovation and competitiveness through investments in HPC, often using the experience from large national and regional centres with large-scale HPC infrastructures, combined with the know-how of scientists and application experts.

As an example of this, Chicago is about to become the home of a new "Digital Lab" with an initial budget of 320 million US dollars. This will be the USA's flagship research institute in digital manufacturing and will stimulate research and engineering growth in the Midwest. The Digital Lab, also known as the Digital Manufacturing and Design Innovation Institute, will apply computing technologies to address the manufacturing challenges faced by industry. The computing technologies used by the new institute will include mobile computing, cloud computing and high-performance computing. The goal of the research performed at the lab will be to make US industry more competitive on a global stage. The lab will offer a state-of-the art research and development facility to both large and small companies.

KTH, PDC and their Industrial Partners

The HPC experts at PDC have been working with researchers and software developers on major international industry-related projects for quite a few years now. For example, recent projects have involved simulating the human brain (http://www.humanbrainproject.eu) and calculating ways to make trucks and planes more fuel-efficient. Recently PDC started a new initiative, the Business Development Unit, which is focusing on developing PDC's capabilities to support KTH's strategic partners in the areas of large simulations and data analysis work. With more than twenty years of HPC experience at PDC, both larger and smaller organisations can benefit from the training, expert advice, and help with application tuning that PDC offers. As PDC participates in various national and international projects that involve large networks of supercomputing resources, PDC's experts can also help joint research projects to obtain access to an even broader range of HPC computing facilities and expertise than is available locally.

KTH and PDC are a key factor in making the Stockholm region more competitive by providing better support for the local science and technology economy. Working together, PDC and KTH are providing education, resources and training in simulation-based engineering science to meet the increasing need for such resources within research-focused organisations.

Above: Results from an aerodynamic simulation showing streamlines around a Scania truck equipped with a trailer

PDC Pub & Open House

This year's PDC Open House and Pub Afternoon was held on the 8th of April. Each year we invite everyone who is using our supercomputers, as well as anyone who is interested in using our resources in the future, to come and meet the PDC staff and see the facilities. Please sign up on our mailing list (see https://www.pdc.kth.se/about/mailing-lists/pdcannounce) to receive an invitation. We look forward to seeing you all next spring!

NeIC Evaluates the Nordic Use of PRACE

Lilit Axner and Michaela Barth, PDC

The Partnership for Advanced Computing in Europe (PRACE) was initially established in 2007 to pool European high-performance computing (HPC) resources and thus give researchers access to the largest European HPC research facilities. Finland, Norway and Sweden were all "founding" members of PRACE, while Denmark was one of the countries that joined PRACE later.

PRACE's "unite and conquer" approach has been very successful in establishing an e-infrastructure that provides world-class HPC resources to support science and engineering research, both in academia and industry, and the researchers of the far North have been taking full advantage of the opportunities that PRACE offers. Sweden, Finland and Norway have participated in almost all the PRACE activities since the inception of the partnership, and the activities and impact of these three Nordic countries have been substantial over the past eight years. However, as PRACE continues to develop, it is important for the Nordic countries to analyse their participation in PRACE and plan their next strategic steps. Consequently the Nordic e-Infrastructure Collaboration (NeIC) recently commissioned an independent evaluation of Nordic participation in PRACE that investigated the added value of this participation and also looked at viable future options for Nordic involvement in PRACE. The report is publicly available at https://wiki.neic.no/wiki/Public_Documents.

The evaluation committee was chaired by Dr. Rob Pennington, Deputy Director of the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign. In an interview given after delivering the findings of the committee, Dr. Pennington commented that "the evaluation committee estimates that the Nordic region's need for PRACE capabilities corresponds to roughly 25 percent of a Tier-0 system, in addition to current national capabilities". The following tables from the report demonstrate some of the ways in which the Nordic countries have been making good use of the PRACE facilities.

	Core time with Nordic PI or co-PI (in millions of hours)	Core time with Nordic PI (in millions of hours)	Average time/call (in millions of hours)	Average time for Nordic PI/call (in millions of hours)
Denmark	248	224	35	32
Finland	258	242	37	34
Norway	56	56	8	8
Sweden	277	126	40	18
Totals	839	648		

Table 1: Estimated Nordic Tier-0 usage for PRACE Regular Access Calls 2-7 based on awards to projects with Nordic principal investigators (from NeIC evaluation)

Tier-0	Denmark	Finland	Norway	Sweden	Total
Proposals submitted to Nordic PI or co-PI	13	14	2	17	46
Proposals awarded to Nordic PI or co-PI	9	8	2	8	27
Awards led by Nordic PI	7	7	2	4	20

Table 2: Nordic results for PRACE Regular Access Calls (from the NeIC evaluation)

First we look at Nordic use of the PRACE Tier-0 systems. Table 1 shows that projects with Nordic principal investigators (PIs) use a total amount of computing time that corresponds to approximately 25% of a PRACE Tier-0 system.

In general, the NeIC evaluation report showed that Sweden has been making good use of the PRACE Tier-0 resources. The latest results reinforce this view with three additional successful awards in Call 8 and many proposals expected for Call 9. These might partly be due to the additional "Nordic-wide" awareness generated by NeIC conducting the evaluation and thus bringing PRACE usage into focus more sharply. For the Tier-0 Regular Access Call 9 and onwards, we are also expecting applications from Iceland.

During the PRACE Regular Access Calls 2-7, a total of 233 awards were made. From Table 2, we can see that 12% of these awards were given to projects with Nordic principle investigators (PIs) and co-PIs, and 8.6% of the awards went to projects led by a Nordic PI.

Next we turn to Tier-1 usage. Table 3 shows that Finland, Norway and Sweden have been extensively involved in PRACE Tier-1 activities.

Note that the number of projects listed in Table 3 only gives a weak indicator of real computing time. For example, in Sweden's case, the assigned time compared to the time provided for the period of the DECI7-10 calls was 122.4% in terms of DECI standardized computing time, since there were many projects that were assigned to the Swedish Tier-1 system, even though the other half of the computational work for those projects was actually performed on other PRACE systems.

One thing that was especially remarkable in the NeIC report was the marvellous value for money that Denmark received. This was due to the fact that awards for Tier-0 access are solely based on scientific excellence, rather than the "juste retour" principle that is applied in Tier-1 calls (which means that the number of successful Danish applications for DECI was low compared to Sweden and Finland). Consequently the Danish e-Infrastructure Cooperation (DeIC) is now working to build up a national infrastructure in Denmark to answer the obvious needs of their research communities that were highlighted by the evaluation report.

The whole evaluation report should be seen in context with the Nordic High Performance Computing (NHPC) evaluation previously conducted by NeIC. Together these two evaluations provide the cornerstones for shaping the future HPC strategy in the Nordic countries!

Lilit Axner, who is the SNIC PRACE coordinator, has found the evaluation that NeIC performed on the Nordic use of PRACE to be of great benefit as it has provided an opportunity to reflect

	PRACE person months (PM) and work packages (WP)	Number of Tier-1 projects hosted on national systems	Number of projects with Nordic PIs or co-PIs awarded Tier-1 access in DECI7-10 calls
Denmark	-	-	4
Finland	340 PM on 21 WP	10	9
Norway	230 PM on 12 WP	8	2
Sweden	311 PM on 21 WP	23	16

Table 3: Nordic involvement in PRACE Tier-1 (from the NeIC evaluation)

on Swedish and Nordic involvement in PRACE and be proud of the substantial impact that the SNIC centres (and other European HPC centres) have had on the European HPC-ecosystem. The evaluation has also helped us to identify areas of congestion where SNIC needs to improve. The recommendations that PRACE received from the NeIC committee are extremely valuable and will be considered in PRACE's future commitments.

Michaela Barth, the NeIC Generic Area coordinator, agrees that the NeIC evaluation report provides many useful starting points for beneficial discussions. For example, we currently see Nordic participation in nearly all of the PRACE work packages, but in the future it is likely that we will have to specialise more and limit our participation to particular work packages. Also, if we look at one of the main conclusions of the NeIC evaluation – namely, the estimate that the Nordic countries use the equivalent of 25% of a Tier-0 system – we see that this is based on the types of systems that PRACE designated as Tier-0 during calls 2-7. Nowadays, though, we would really like to (re)define a Tier-0 system as a system that scales a magnitude higher compared to the largest available national systems; this would of course also be reflected in the cost of such a system.

Michaela had hoped for more immediate concrete reactions to the recommendations of the report, such as the Nordic countries starting to share national strategic documents and usage statistics between themselves more freely, but perhaps that was too optimistic. The current first decision in reaction to the report has been to map the scientific use case landscape in the Nordic countries to obtain a better understanding of our customers in order to create a common science strategy. NeIC is now developing a new project, code-named "Cognitus", which is setting out to do exactly this. Mapping the scientific use case landscape is certainly a wise decision, since, as Rob Pennington said, "At the end of the day, [...] scientific success should be the goal!" All infrastructure is just here to serve the purpose of supporting our scientific research. Now that NeIC is starting to perform within its collaboration projects, we can see that some of the necessary action is already underway.

Michaela concludes that it will be interesting to see how PRACE evolves in the future, particularly the business model for PRACE 2.0, as this will certainly have a big impact on the Nordic countries when it comes to making strategic decisions about how to provide the best possible services to the Nordic research community.

If you have any questions about using the PRACE resources, you are, as always, welcome to contact Lilit Axner for help, and Michaela Barth would be very happy to discuss the NeIC report if you have any queries about the report or the numbers in it.

ExaMPI13 Workshop at Supercomputing 2013

Stefano Markidis, HPCViz

The EPiGRAM project organized the first workshop on Exascale MPI (ExaMPI) at the Supercomputing 2013 conference in Denver on 22nd November 2013 (https://www.pdc.kth.se/exampi13). The goal of the workshop was to bring together developers and researchers to present and discuss innovative algorithms and concepts in message passing programming models, in particular related to MPI. More than a hundred scientists attended the event. Keynote talks were given by Torsten Hoefler, Pavan Balaji and Christian Simmendinger.

Above: Jesper Larsson Träff presenting MPI needs for large scale computing at the ExaMPI13 workshop

Staff Focus

Leif Nordlund

Leif Nordlund, a seasoned veteran in the IT business, recently joined PDC in the new role of business development manager. With a solid background in systems development for Sandvik and Skandia, Leif then worked for a number of IT companies (most notably Sun Microsystems and AMD), both in marketing and sales. His interest in high-performance computing (HPC) goes back to when he was product marketing manager for x86 servers at Sun, and he was later responsible for HPC business development in Europe for AMD. Leif has a degree in computer science from Uppsala University and enjoys alpine skiing, photography and all types of boating. Leif lives with his partner and two daughters on the island of Ekerö. Recently a boy also appeared in his house, a Siberian cat called Sam Jones.

Free AVL Fire Licenses

If you are interested in running AVL FIRE at PDC, please contact support@pdc.kth.se to request a free license - in your email please remember to include details of the research project or course where you would like to use AVL FIRE.

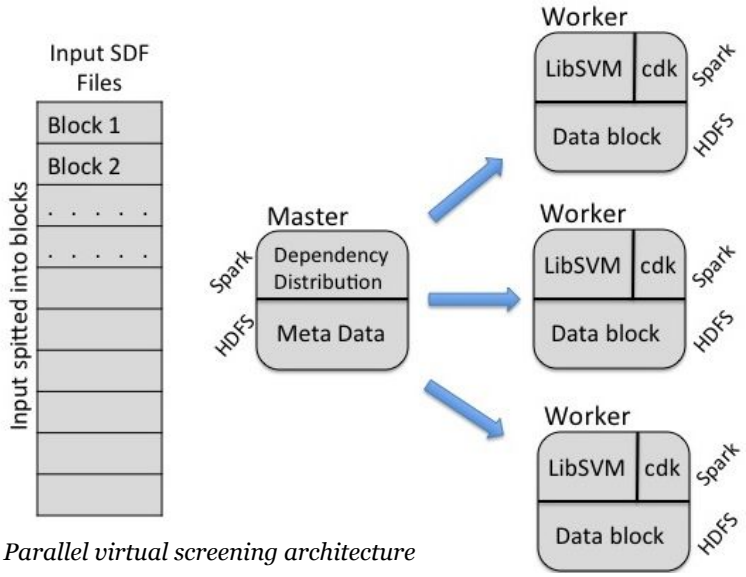
Rapid Drug Discovery: Using MapReduce for SVM-based Parallel Virtual Screening

Laeeq Ahmed, HPCViz

Introduction

Drug discovery is the process of screening large numbers of chemical libraries to find new medicines. Due to the huge size of chemical libraries, traditional screening is a time-consuming and costly activity. However, with advancements in computer technology, virtual screening can be performed using machine learning techniques for filtering large collections of chemical structures. Support vector machines (SVMs) offer one of the most famous machine learning techniques for classification and regression analysis. In this project we developed a parallel version of SVM-based virtual screening using the iterative MapReduce programming model Spark [1] to further reduce the filtering time and thus the cost of searching.

Spark is an open source implementation of MapReduce, which is used for large data analytics. In comparison to Hadoop [2], Spark is ten times faster [1] and also provides support for iterative MapReduce computations, which are crucial for many applications based on data mining, web ranking, graph processing, model fitting, and so on. Other than providing the facility of iterative MapReduce computations, Spark has new features of in-memory computing and fault tolerance using Resilient Distributed Datasets (RDDs). With in-memory computing, we can store datasets in memory and can access them far more quickly. If the size of the datasets is bigger than the available memory, the rest will be automatically stored to disk. RDDs bring innovative fault tolerance into Spark by making use of the concept of "lineage". Lineage means RDDs can remem-



Above: Parallel virtual screening architecture

ber all the transformations from which they have been derived and, on failure, can roll back to a state from which they can reproduce themselves.

Parallel Virtual Screening Architecture

Virtual screening of chemical libraries is performed using SDF (Structures-data format) files. These SDF files consist of molecule structures. For classifying these molecules, the Library for Support Vector Machines, LiBSVM [3], is used. LiBSVM creates models from example structures and later these models are used to foresee other molecule structures.

Here the Hadoop Distributed File System, HDFS [4], is used for data distribution and Spark for data processing. The datasets are divided into logical splits by the Spark master daemon. Each worker node receives its part of the work according to the logical splits as shown in the figure on the previous page. Apache Maven [5] was used to create a single Java Archive (JAR) file that consists of our application code and other dependencies like LibSVM and CDK[6]. The JAR file is then deployed to the worker nodes through SparkContext. Two processes, namely a mapper and a reducer, are running on each worker node. Each input split is provided to one mapper, thus parallelism and the number of mappers depend on the input split. The number of reducers can be changed from the driver program. After the workers complete the processing task, the result is send back to the master node.

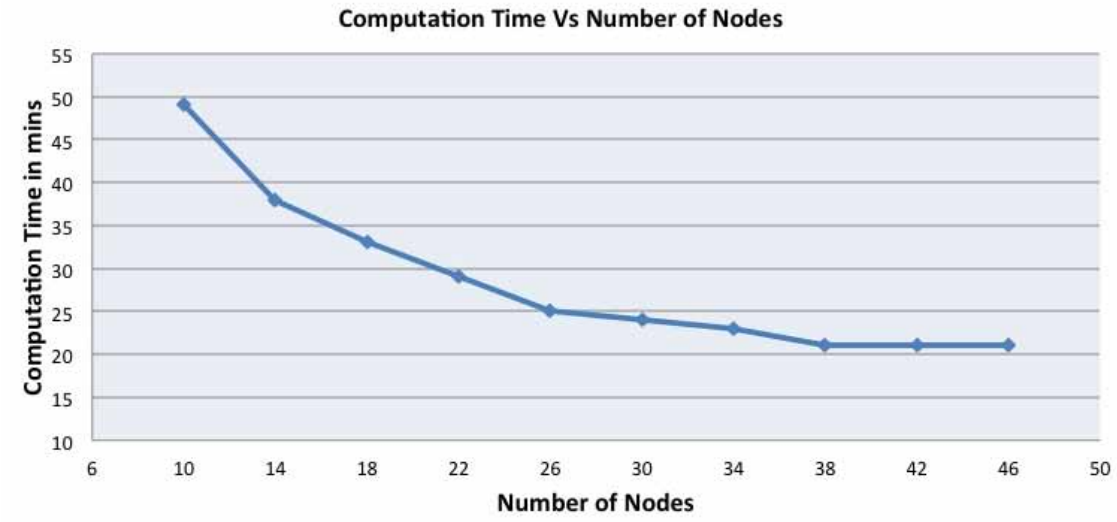
Experiments and Results

In this study, we performed two different experiments. In the first experiment we wanted to see the scaling of our MapReduce-based parallel virtual screening with an increasing number of nodes. We clearly see from the figure below that our implementation does scale well with an increasing number of nodes, though it flattens out after certain number of nodes.

In the second experiment, we wanted to see the effect of different input split sizes on the completion time of the job. We use two different input split sizes, namely 64MB and 128MB. The experiment shows that our implementation works better with the smaller input split size. With the increased input size, there is less parallelization and thus an increase completion time. This is illustrated in the figure on the next page.

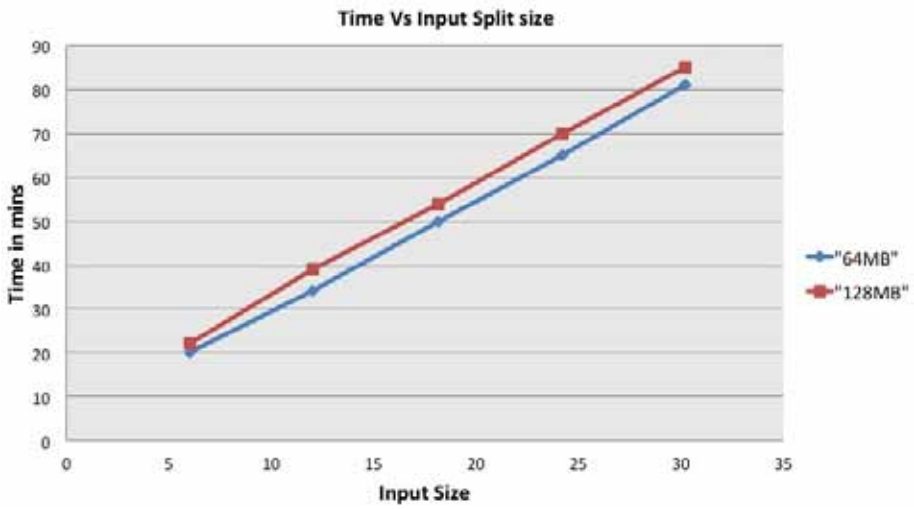
Conclusion

The study shows that virtual screening can be successfully parallelized using MapReduce to speed up the virtual screening process. Also, most of the SVM-based virtual screening is performed using MPI. Here we show that MapReduce can be a viable substitute that opens up ways to use large cloud-based infrastructures for virtual screening.



Above: Parallel virtual screening scaling

Below: Implementation behaviour with different input split sizes



References

[1] Zaharia, Matei; Chowdhury, Mosharaf; Franklin, Michael J.; Shenker, Scott; Stoica, Ion. *Spark: Cluster Computing with Working Sets*, University of California, Berkeley.

[2] Apache Hadoop, http://hadoop.apache.org

[3] Changand, C.; Lin, C.J. LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology (TIST),* **2011.** Volume 2, Issue 3, Article No. 27.

[4] HDFS Architecture Guide, http://hadoop.apache.org/docs/r1.0.4/hdfs_design.html

[5] Apache Maven Project, http://maven.apache.org

[6] Chemistry Development Kit, http://sourceforge.net/projects/cdk

PDC-Related Events

PDC Summer School 2014: Introduction to High-Performance Computing

18-29 August 2014, KTH Main Campus, Stockholm

http://www.pdc.kth.se/education/summer-school

Introduction to PDC Systems - Training Day at PDC

16 September 2014, Room 304, PDC, KTH

https://www.pdc.kth.se/events/event-repository/pdc-training-day-2014

Debugging with Allinea DDT - Training Day at PDC

25 September 2014, Room 304, PDC, KTH

https://www.pdc.kth.se/events/event-repository/ddt-training-day-2014

Introductory Course on Writing Parallel Programs using MPI

25-27 November 2014, KTH

https://www.pdc.kth.se/events/event-repository/writing-parallel-applications-using-mpi

HPC Sources

We recommend the following sources for other interesting HPC opportunities and events.

CERN
http://cerncourier.com/cws/events
http://cds.cern.ch/collection/Forthcoming%20Events?ln=en

EGI
http://www.egi.eu/about/events

HPC University
http://www.hpcuniv.org/events/current

HPCwire
http://www.hpcwire.com/events

Linux Journal
http://www.linuxjournal.com/events

PRACE
http://www.prace-ri.eu/HPC-access
http://www.prace-ri.eu/PRACE-Training-Events
http://www.prace-ri.eu/news

SNIC
http://www.snic.vr.se/news-events
http://docs.snic.se/wiki/Training

XSEDE
https://www.xsede.org/conferences-and-events

Save the date! NeIC 2015 Conference

5-8 May 2015, Finland

http://neic.nordforsk.org/neic-2015-conference

Please join us in Finland next spring and enjoy getting in touch with your Nordic colleagues while we all benefit from the dedicated workshops at the NeIC 2015 Conference!