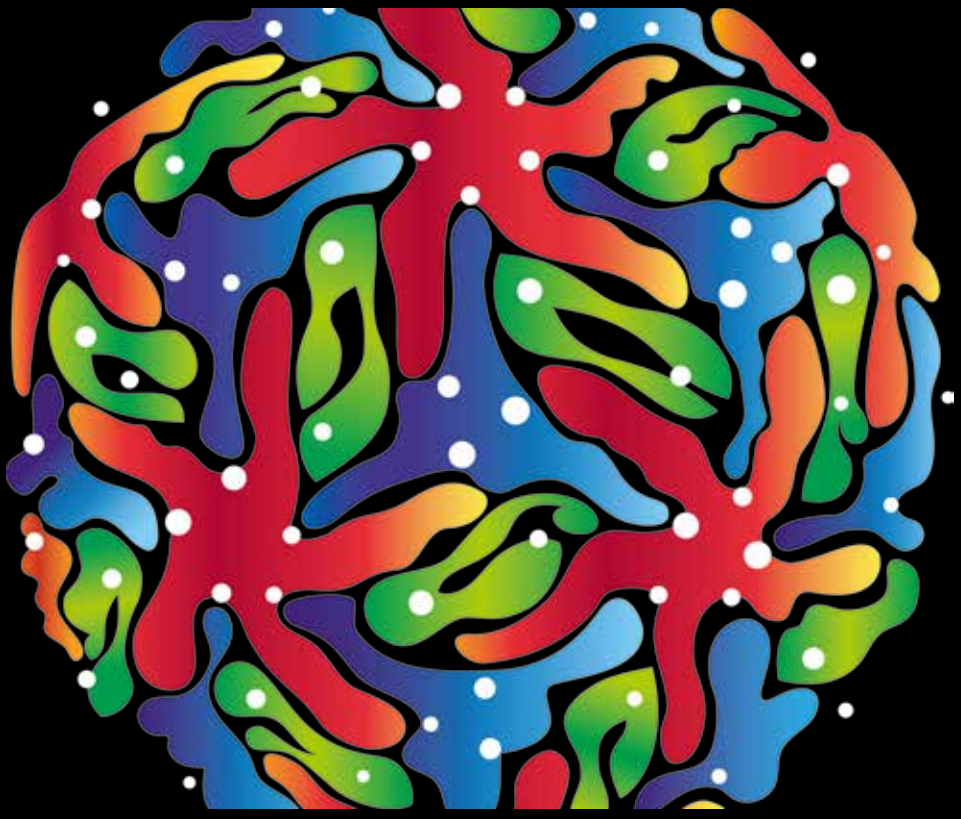
no.2 2018

PDC Center for High Performance Computing

PDC Newsletter



The PDC Newsletter is published by PDC at EECS, 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 EECS, 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: support@pdc.kth.se

ISSN 1401-9671

Cover

The cover shows a stylized image of the structure of the Zika virus (PDB ID: 5IRE). BioExcel core applications are routinely used by researchers for understanding the function of biomolecules and the development of novel drugs.

How to Contact PDC

Visiting address:
Teknikringen 14, "Plan 4",
KTH, Stockholm
Mail: PDC, KTH,
SE-100 44 Stockholm, Sweden
E-mail: support@pdc.kth.se
www: http://www.pdc.kth.se
Phone: +46 8 790 7800

Editorial

The road map towards exascale is shaping up with the formation of the EuroHPC Joint Undertaking which, after a record-breaking time of only seven months, was created by the European Commission and started operating in autumn 2018. EuroHPC is now responsible for building the European exascale infrastructure and supporting related research on exascale technologies. The first milestone for EuroHPC will be the procurement of two pre-exascale systems by 2020, which should pave the way towards European exascale systems by 2023. As time is short, the hosting entity for the pre-exascale systems will be selected in early 2019 and the procurement process will start soon after that. Almost all European countries are supportive of EuroHPC however, although Sweden has a clear ambition to contribute to the undertaking, the current political situation unfortunately means that the necessary governmental decisions cannot be made as yet. We hope that the situation will be clarified soon so that Sweden will become a full member of EuroHPC in the near future.

From the beginning, PDC has been at the forefront of the European exascale developments and has worked with different user groups to make Sweden ready for the rapidly approaching exascale era. With engineering being a key sector in Sweden, and Swedish research in computational fluid dynamics (CFD) being world leading, we have helped to push CFD towards exascale through the ExaFLOW project. After a three-year-long journey, ExaFLOW had an excellent final review and showcased its developments in three flagship runs. You can read about the details in this newsletter.

Another important effort is the BioExcel Centre of Excellence for Computational Biomolecular Research. This European flagship project is also finishing its first period and has already secured funding for a second three-year period starting in 2019. Read more about the achievements and plans in our cover article.

PDC's participation in other activities related to high performance computing is continuing too, particularly the Nordic CodeRefinery project and the European HPC-Europa3 initiative. Short updates on these can be found in this newsletter.

Earlier this year we conducted a broad PDC user survey – I hope you participated in it! Overall the results were very encouraging and showed that our efforts to improve our user services over the past couple of years have been recognized and appreciated by our users. Of course, this does not mean that we will rest on our laurels; as always, we will continue striving to improve even more. One of our

very recent improvements features a new way of interacting with you – the PDC Blog. Go to https://www.kth.se/blogs/pdc and check out tips and tricks for using PDC resources.

I hope this newsletter and the PDC Blog give you an interesting read for the coming holiday season, and I and all of the PDC staff wish you "God Jul och Gott Nytt År"!

Erwin Laure, Director PDC

BioExcel Centre of Excellence Going Strong: New Funding Secured

Rossen Apostolov, PDC

BioExcel, the leading European Centre of Excellence for Computational Biomolecular Research, was established in 2015 with support from the European H2020 program. The centre is coordinated by PDC, with the KTH GROMACS team providing the scientific lead. The consummate execution of all the centre's activities since its foundation has enabled us to secure a new round of larger funding for another three years of operation. Here we present the future plans for the even more exciting upcoming activities in the centre!

Driving Computational Biomolecular Research

The life sciences in general, and biomolecular research in particular, have grown to be one of the major users of large-scale compute infrastructures. The potential impact of, and requirements for, computing in these areas are extreme. Software in these areas has been making significant progress in the last decade, however, for applications in these areas to be able to take full advantage of the upcoming exascale systems, the research communities are facing immense challenges to co-design software and algorithms suitable for the new generations of hardware that are evolving. The life sciences are areas where European applications lead the world, however, there had not been sufficient interaction between the code developers, the researchers using the software and the organisations providing the computing resources. This had, in turn, affected the productivity of the researchers, as well as the utilization of the European HPC infrastructures. Indeed part of the initial motivation for BioExcel was to bridge the gaps between geographically diverse developers, researchers and infrastructure providers by establishing BioExcel as a focal point for interactions.

Based on the successes of the first phase of the project, BioExcel is continuing its mission to advance science and technology in the life sciences by:

- pushing the performance, efficiency, scalability, and usability of several key software packages towards exascale;

In This Issue

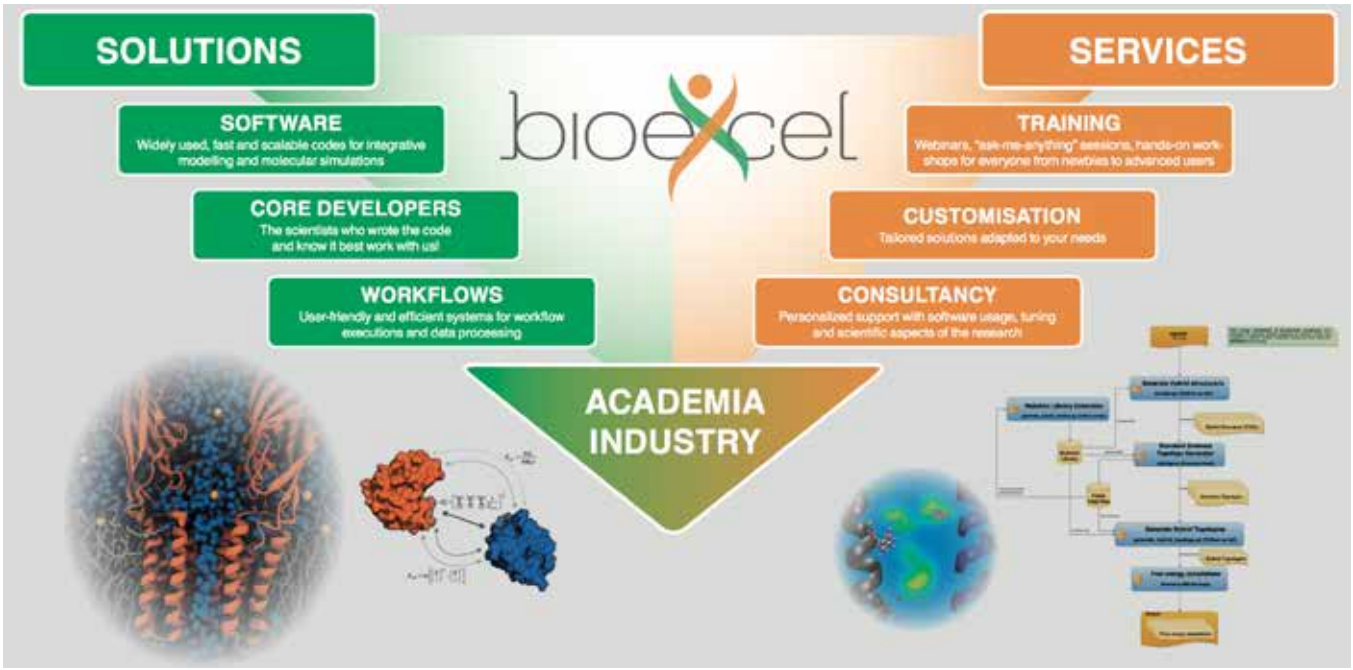
Above: PDC Summer School 2018

- improving the usability of existing applications and tools to support the convergence of high performance computing (HPC), high throughput computing (HTC), and high performance data analytics (HPDA) via the development of workflows combining HPC simulations with data management and analytics;
- significantly expanding the range of training, commercial applications and services that are offered;
- making publications, libraries, workflows, codes, and training material developed by BioExcel available via open-source and open-access principles;
- making BioExcel sustainable so researchers in academia and industry can rely on having access to BioExcel's open resources in the longer term; and
- collaborating with international initiatives to strengthen the links between the research communities and leverage on worldwide expertise.

World-Leading Exascale Biomolecular Software

BioExcel has a strong focus on getting life sciences applications ready to take advantage of the superior speed of the exascale systems that will be available soon. We host the development of several of the most widely-used European HPC codes: GROMACS for molecular dynamics simulations of biomolecular systems, HADDOCK for integrative modelling of biomolecular complexes, and the combined QM/MM capabilities of CP2K. GROMACS and CP2K are among the 12 PRACE benchmark codes – this is significant because they already sustain multi-petaflop performance – and HADDOCK is one of the most-widely used data-driven codes. To support researchers further, BioExcel has invested considerable effort in the development of related tools, such as PMX to prepare free energy calculations, POWERFIT for docking into cryo-EM maps, and RELION for the reconstruction of single-particle cryo-EM data (our GPU-accelerated cryo-EM reconstruction has been used in 31 *Nature* and 12 *Science* publications). The codes are used throughout the biotechnology, pharmaceutical and chemical industries. Several SMEs that use our software have started close collaborations with BioExcel.

BioExcel engages extensively in co-design projects with HPC vendors and system designers to improve the performance of various codes on specific types of hardware.



Above: BioExcel provides a wide range of products and services for computational biomolecular researchers.

GROMACS is one of the most highly tuned codes in the world, and is presently involved in co-design projects with AMD/OpenCL, NVIDIA/CUDA, Intel, IBM/Power9, and Arm. HADDOCK is one of the best examples of a data-driven application, for which co-design efforts focus on exploiting new storage, network and memory caching technologies in collaboration with companies such as Intel and Seagate. In upcoming projects, the QM/MM capabilities of CP2K will be improved (which will be a big advantage for the life sciences research communities) and CP2K will also be tuned for future exascale systems in collaboration with Intel, NVIDIA, and Cray amongst others. The codes will be optimised for the EuroHPC systems (including the upcoming new European accelerators). New memory and storage technologies for ensembles and high-throughput computing will be exploited, and sub-checkpointing will be employed to improve the resilience of code against hardware failures. BioExcel will also improve the scaling and usability of the codes, and make them easier to handle by developing and offering efficient workflows, distributed as container images from the BioExcel software hub.

BioExcel codes drive European business research in the life sciences. HADDOCK is currently cited in about 300 papers per year, GROMACS in about 4,000 papers per year, and CP2K in about 500 papers per year. The HADDOCK portal sees over 30,000 submissions yearly from users around the world (10 million jobs are submitted to EOSC HTC resources). The core applications are distributed as open-source and free software to ensure that they can be used freely by non-profit organisations, and also have business-friendly licensing to allow for commercial reuse and linking. The codes are already used by all of the top ten pharmaceutical companies, plus a number of SMEs, and there are several vendors selling cloud resources or hardware tailored to the codes.

Convergence of HPC/HPDA and Improved Usability

Workflows are vital for the efficient usage of exascale resources and the improved productivity of researchers. BioExcel has already developed a framework for FAIR (Findable, Accessible, Interoperable and Reusable) workflows, and a well-defined set of best practices in collaboration with ELIXIR (CWL, bio.tools, and interoperability platforms); these have been tested with pharmaceutical companies (which generated a range of scientific success stories of biological interest). BioExcel is now continuing this work by combining HPC compute engines with high performance data analytics (HPDA) and machine learning methods for the automated retrieval and deposition of data. The portfolio of platforms we support includes major ones such as CWL, KNIME, and Galaxy, as well as managers specifically designed for HPC (PyCOMPSs, Nextflow, CWLEXEC). Interoperability and reproducibility will be strengthened using

Staff Focus

Tor Kjellsson Lindblom

Tor received his Ph.D. degree in computational atomic physics from Stockholm University in December 2017. He then moved to Tokyo for a postdoctoral position at the University of Electro-Communications. His research concerns theoretical descriptions of the interaction between ultrafast laser pulses and matter, which, even for the simplest systems, may require a substantial amount of high performance computing. In Tor's case, he frequently implements hybrid OpenMP-MPI code to solve large-scale problems. In December 2018 Tor joined the team of application experts at PDC.

In his spare time, Tor enjoys high intensity training and lifting weights.

Above: PDC Summer School 2018

containers (Docker, Singularity) coupled to repositories such as BioConda, BioContainers and the BioExcel hub3. The integration of HPC-HPDA tools will support the simultaneous execution of simulations and the online analysis of the resulting data. This will, in a natural way, help to attract members of the bio- and cheminformatics research communities to HPC, since they commonly use workflows as their working environment for automation.

User-Driven Development

BioExcel's activities are guided by input from users (namely researchers using biomolecular codes) as the purpose of the centre is to make computational research in the life sciences easier to undertake and more efficient. For example, application and workflow development is being done in tandem with use cases taken from the wider academic and industrial research communities. Consequently, these use cases will become well-documented, best-practice examples of how to scale difficult problems, not only with more nodes, but also by including disruptive changes in algorithms. Ensuring that users run applications with optimal settings would double the resource efficiency for many projects. In addition, the development road maps for the core applications take into account the needs of researchers via input provided by the wider research communities.

Addressing the Skills Gap

BioExcel's mission is to enable researchers to fully exploit the power of data and computing e-infrastructures by providing support and training for both non-expert and advanced users. BioExcel pioneered an advanced training program (which was praised by reviewers) based on a competency-based training needs analysis in collaboration with the broader life sciences research community. The resulting training resources are openly available at the BioExcel Knowledge Resource Centre. BioExcel will continue to expand the training program by delivering new educational webinars and courses (face-to-face and online), as well as running best practice and knowledge-exchange workshops on cutting-edge topics and support forums focused around the core applications. BioExcel will also continue to make targeted support teams available to help researchers solve specific problems. Our educational webinars (of which 34 will have been produced by the end of 2018) have proved to be a popular and efficient way to share expert knowledge. In addition, BioExcel will set up a dedicated support group working with academic and industrial researchers in synergy with other support structures (such as PRACE HLSTs). BioExcel's online and virtual training products and services will make much-needed training available to more and more researchers all around the world as we expand the range of online tutorials for various software tools to guide researchers at their own pace. BioExcel can help researchers who are just beginning to use HPC with training on topics such as HPC readiness and introductory modules in biomolecular simulations, as well as assisting highly experienced HPC researchers with advanced training and support on specialist topics such as MD, QM/MM, and workflows. In addition, the centre provides customised training for industry researchers by arrangement.

Collaboration and Community Links

BioExcel is based on the premise that the significant developments in science are achieved by strengthening the interactions within the whole research ecosystem, consisting of software developers, academic and industrial researchers and research infrastructure providers. To facilitate that vital intra-ecosystem communication, we continue to work on forging wider connections within the life sciences research communities (both in Europe and internationally), so that the BioExcel Centre of Excellence serves as a hub, that is, a focal point for the worldwide life sciences research communities. For example, thanks to our close collaborations with the developers of other major life sciences codes (like NAMD and AMBER), we expect to be able to make these resources open to all life sciences applications in the near future. This would mean that researchers would be able to choose a code based either on their preference or its performance for a specific type of problem. BioExcel will also continue to

work with other centres of excellence, relevant EU initiatives (namely PRACE, ELIXIR, INSTRUCT, and the EOSC-Hub), and international ones (like MolSSI) to organise joint training events targeting other important codes in biomolecular research.

Long-Term Operations

In 2019 BioExcel will be establishing a commercial arm for delivering life sciences products and services to both the academic and business research communities. For example, companies in Europe, particularly in the pharmaceutical industry, would benefit significantly from having access to appropriate workflows and training. BioExcel's intellectual property plan will ensure that the products resulting from BioExcel's activities (such as code, libraries, workflows, publications, and training material) are open source and open access following the FAIR (Findable, Accessible, Interoperable and Re-usable) data principles.

All academic and industrial researchers involved in computational biomolecular research are warmly invited to contact BioExcel to find out how to take advantage of the many opportunities for support and joint activities, or to discuss new directions for collaboration. BioExcel has representatives in Finland, Germany, the Netherlands, Norway, Malta, Spain, Sweden and the UK who are happy to provide guidance and assistance with your work.

PDC Summer School 2018

Stefano Markidis, CST

The 22nd PDC Summer School ran from 13-24 August 2018 at KTH in Stockholm. The 58 students who attended the school were mainly Ph.D. and Masters students (some from Sweden and some from other countries) along with participants from industry – they learned about different aspects of high performance and parallel computing, from the foundations of computer architecture to advanced concepts (such as how to program GPUs and large scale supercomputers). The next PDC Summer School will be held in the last two weeks of August 2019.

Above: PDC Summer School participants, August 2018
Right: Summer School lectures, August 2018
Note: All Summer School photos were taken by Apostolos Vasileiadis.

Staff Focus

Mahan Tourkaman

Mahan has been working part-time at PDC for a bit over three years now, while also studying Electrical and Electronics Engineering at KTH. His main goal at PDC has been to help make processes in the Support group run more smoothly by building new internal tools and helping design new workflows. Mahan has also been responsible for the technical implementation of the current PDC website under KTH's content management system, Polopoly, which involved migrating content and the integration of external parts and features.

In his spare time, Mahan likes to practice music, learn new dishes to cook, and spend time with his family and friends.

ExaFLOW Project Helps Industry Exploit High Order CFD

Philipp Schlatter, Adam Peplinski, and Nicolas Offermans, Linné FLOW Centre and Swedish e-Science Research Centre (SeRC), KTH Mechanics, and Niclas Jansson, PDC

After three years of working on key algorithmic challenges in computational fluid dynamics (CFD), the eight consortium partners in the ExaFLOW Horizon 2020 FET (Future and Emerging Technologies) project reminisce here about what has been achieved during the lifetime of the project. With three flagship runs, ExaFLOW has managed to work on several specific CFD problems – the differences between the problems that ExaFLOW addressed serve to highlight the importance of the outcomes of the project for both industry and academia.

Fluid dynamics and turbulence are important topics for our everyday life: both from a technical/industrial point of view, and also when considering biological and geophysical systems. It turns out that a large fraction of our energy consumption is directly due to turbulent friction, that is, the dissipation of turbulent kinetic energy to heat caused by fluid viscosity. The study of turbulence is therefore an essential research discipline in today's academic and industrial arenas, and has been referred to as the last unsolved problem of classical physics.

Obtaining numerical solutions to problems involving turbulent flows, which is an important aspect of CFD, is a prime contender for simulations aiming towards higher and higher levels of computational performance. This is mainly because, in turbulent systems, a large range of scales are active, and thus for high-speed vehicles – such as aeroplanes, cars and trains – there is virtually no limit to the necessary degrees of freedom. These active modes translate directly to the size of the system to be studied via numerical simulations. Therefore, CFD applications are particularly suited for extreme parallel scaling, in other words, to reach beyond today's petascale systems towards exascale performance. Of course, these are challenging problems because, in turbulence, non-local interactions are important, which puts a heavy burden on communication between processors during the simulation.

In order to facilitate the use of accurate simulation models (that is, high-fidelity simulations) in exascale environments, ExaFLOW worked towards having three flagship runs near the end of the three-year project period. These flagship runs were designed with high industrial relevance in mind – addressing key innovation areas such as mesh adaptivity, resilience, power consumption and strong scalability. Each of the three flagship runs addressed

Staff Focus

Artem Zhmurov

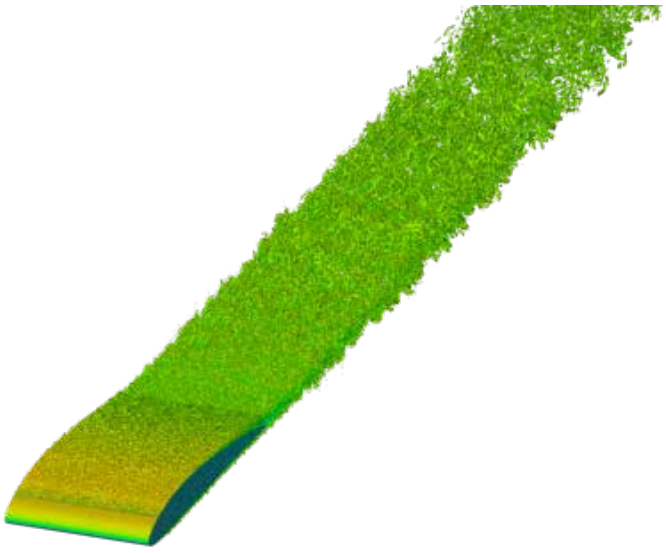
Artem graduated from the Moscow Institute of Physics and Technology (MIPT) with a Ph.D. in applied mathematics and physics, where his research projects were related to mathematical modelling in biology and physiology. Later, he obtained a Ph.D. in biochemistry from the University of Massachusetts at Lowell with the primary focus being on computational molecular biophysics. After graduating, Artem went back to MIPT to work as a postdoctoral fellow and, later, as an assistant professor. He was involved in the development of the GPU-based implementations of various coarse-grained models for simulations of large biomolecules and biomolecular complexes. He also has experience teaching computational mathematics and biomolecular simulations.

Artem joined PDC this November. His main focus will be on developing the GROMACS software package, particularly its GPU-based implementation.

specific CFD problems, thereby demonstrating the improvements made by ExaFLOW to their co-design applications: Nektar++, Nek5000, and OpenSBLI.

Flagship Run No. 1: Adaptive mesh refinement (AMR) of turbulent wall-bounded flow with Nek5000

A technique known as adaptive mesh refinement (AMR) was used to study the turbulent flow around a NACA 4412 wing section at a Reynolds number of Re = 200,000. (The NACA wing sections are mathematical models for shapes of aircraft wings that were developed to categorise the aerodynamic properties of different wing profiles, while keeping them simple and reproducible.) The framework that was used for the simulations was the Nek5000 code, which is based on the spectral element method (SEM). Using algorithms developed within ExaFLOW, the mesh that represented the wing section in the simulations could evolve dynamically depending on the estimated computational error at any given location in space and time: the simulation determined the optimal mesh on its own, thus making the design of an initial mesh much simpler. This means that the throughput time for a computational engineer (when using this approach to solve another problem of this ilk) is much faster than for other approaches used in earlier studies of this type of problem, while still complying with the specified error measures for the solution. Using adaptivity, the simulation time could be reduced by more than 50% compared to previous studies of the same case. In addition, the improved flexibility in mesh resolution makes it possible to significantly increase the domain size, while avoiding elements with high aspect ratios in the far-field. This, in turn, results in a lower sensitivity to boundary conditions and a better condition number for the numerical operators.



Above: Turbulent eddies (vortices) arising in the boundary layer around a wing profile

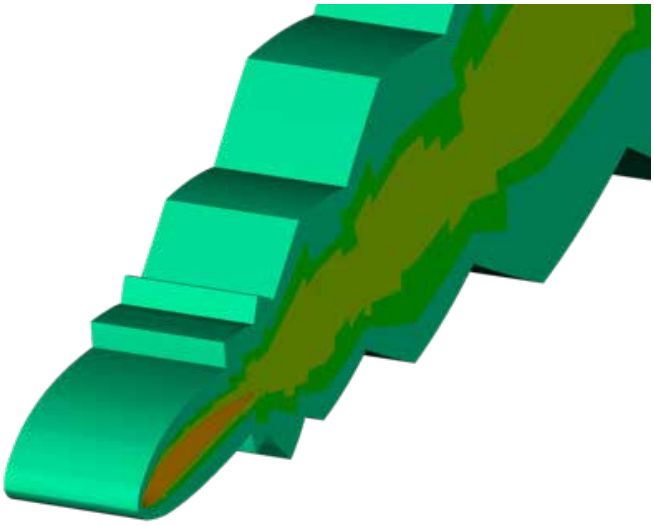
The selected error estimators for the simulations were based on the spectral properties of the discrete solution. They were shown to provide a good estimate of the local conservations of mass and momentum, while being cheap to compute.

The mesh refinement process was done via the so-called h-refinement technique, where the number of degrees of freedom can be increased locally by octree (3D) splitting of selected elements. Extending Nek5000 to incorporate such capabilities required substantial modifications in the code. In particular, these changes included the introduction of interpolation operators at the interface between fine and coarse elements, a new implementation of the preconditioners, and the use of the external libraries p4est and ParMetis for the management and partitioning of the grid, respectively.

As a result of this work the ExaFLOW partners are of the opinion that solution-aware adaptive simulation techniques will play a major role in future CFD, in particular for more complex flow situations where the physics is either not known a priori, or is changing quickly depending on outer circumstances, such as boundary conditions.

The figure to the left shows an instantaneous visualisation of the turbulent eddies (vortices) arising in the boundary layer around the wing profile. The flow is tripped at a short distance downstream of the leading edge, and then a spatially evolving pressure-gradient boundary layer develops. Downstream of the wing, a turbulent wake is established, which slowly decays due to viscous effects.

Below: Regions around a wing with refinement levels larger than 3, identified by different colours



The figure above illustrates regions of the domain with refinement levels larger than 3, identified by different colours. It can clearly be seen that the highest resolution is required close to the wing surface, and in the centre region of the turbulent wake; these are the regions where the smallest eddies (and thus the highest local Reynolds number) are expected. These regions can dynamically adapt during the course of the simulation, in case, for example, the angle of attack is changed.

Flagship Run No. 2: Efficient code generation and error indicators with OpenSBLI

The second flagship run focused on a compressible version of the NACA4412 wing profile as the actual test case, and used the finite difference code OpenSBLI. This compressible flow solver is based on automatic code generation methods developed during the course of the ExaFLOW project. The code efficiency comes from the fact that C++ code is automatically generated for the underlying architecture using a high-level language to express finite difference stencil operations. In common with the first flagship run, OpenSBLI also relied on error indicators for rapid validation of grids containing more than billions of points for simulations that were subsequently run on Europe's fastest supercomputers, showing near perfect weak scaling up to 95,000 cores, and strong scaling up to 50,000 cores.

Flagship Run No. 3: Turbulent flow around the Imperial Front Wing with Nektar++

Our final flagship run was based on the turbulent flow around a front wing of a Formula-1 racing car, a complex geometry where the challenges of high order meshing and solver robustness are evident. Utilising some of the developments that made it possible to do the first two runs in the ExaFLOW project, we were able to successfully run this geometry at the full experimental Reynolds number and with varying polynomial order in Nektar++ and compare it with state of the art finite volume large eddy simulation (LES) and time-resolved experimental data. This demonstrated significant progress in leveraging Nektar++ for the goal of moving high-order CFD from being an academic tool to an industrial one.

All the ExaFLOW partners will continue their efforts to improve the co-design applications with the aim of pushing the new algorithms to higher technology and readiness levels (TRLs) thereby making fast, accurate and resilient fluid dynamics simulations possible and accessible for both academic and industrial research and development at exascale.

For more information about ExaFLOW, visit http://www.exaflow-project.eu for links to all the reports and code developed by the project.

CodeRefinery 2.0

Thor Wikfeldt, PDC, Radovan Bast, UiT, Richard Darst, Aalto Science-IT, Max R. Eckardt, Dataku-ben-USD, Anne Fouilloux, UiO, and Stefan Negru, CSC-IT Center for Science

Since 2016, the CodeRefinery project has been teaching students and researchers how to write better code and providing research groups with software development e-infrastructure tools to support this. The project has not been holding workshops to teach people to program, but rather has been teaching research programmers the skills they are often missing. The project is funded by NeIC (the Nordic e-Infrastructure Collaboration) with co-funding from the Nordic national high performance computing (HPC) providers. After a successful first

phase, the project is now entering phase two, with three more years, a larger team, new workshops, and a renewed focus on community building.

What explains the success of the first phase of the project? Research groups often struggle with collaborating on, and maintaining, scientific software: code projects developed by Ph.D. students and postdoctoral researchers die out after the main developer moves on to another job; the inevitable discovery of bugs in scientific code can cause unnecessary panic since researchers cannot easily determine which publications may have been affected; new Ph.D. students spend months rewriting code from scratch because code inherited from former group members is undocumented and difficult to understand. CodeRefinery workshops focus on state-of-the-art tools and best practices in sustainable software development which can alleviate many of these problems. Since 2016, the project has delivered 13 three-day workshops, as well as five shorter events, across the Nordic countries to over 400 students and researchers. These workshops focus on reproducibility and managing complexity in a collaborative environment by discussing version control, code reviews, automated testing, code documentation, and reproducible workflows and notebooks. According to the CodeRefinery project manager, Radovan Bast, "We consider the CodeRefinery workshops to be a logical next step for students and researchers who have participated in Software Carpentry workshops which focus on the basics. We employ a similar teaching style with interactive exercise-driven code-along sessions, but apply the concept to more advanced work."

The next three years of the CodeRefinery project will build on the success of the previous years. More of the traditional three-day workshops will be delivered, with the hope that a self-sustaining community will start emerging with more locally organised workshops and volunteer instructors. However, the project will also – in a parallel track – develop new lesson material in order to reach new target audiences. In particular, researchers in humanities and social sciences face distinct technological challenges in making their research software and data reusable, accessible and findable. Topics such as data management, data scraping and data mining, framed around the FAIR principles for data and research output, will be taught in "data hackathons" organised around particular themes or problem types.

During the second phase of the project, universities and research institutes will be able to either request a CodeRefinery workshop or self-organise one. Requested workshops can be customized to meet the needs of the local staff by choosing which lessons will be taught. In this scenario, a local organiser is expected to assist with the administrative tasks involved in arranging a lecture room, handling registrations and advertising the workshop. It is also desirable that experienced local staff can assist as workshop helpers (and even contribute to the teaching), as this will build local competence and pave the way for future self-organised workshops. In such self-organised workshops, the open-source CodeRefinery training material can be taught by anyone who has attended and/or contributed to previous workshops.

In addition to requesting workshops, universities will be invited to join CodeRefinery as partners in the future. A partner university will be able to host 1-2 workshops per year with

experienced instructors from other locations, and in exchange the university will be expected to contribute instructors to workshops in other universities. Indeed, after phase two, CodeRefinery hopes to become a self-sustaining organisation with a light-weight coordination structure and in-kind contributions from its partners. Through this, CodeRefinery will create a lasting culture of knowledgeable researchers, peer teaching and mentoring.

The goal of the CodeRefinery project is more than just a teaching program – it is also about catalysing and building a community of Nordic Research Software Engineers (RSEs). RSEs are people who support research by their knowledge of software engineering practices. Similar communities have emerged in other countries, including the UK, the Netherlands, and Germany, and serve as an important pillar of modern computational and data-based research. CodeRefinery workshops, data hackathons and other events will provide a meeting point for local RSEs, but also provide a hub to connect with a network with RSEs across the Nordic research communities and invest in our most important resource – people.

To request a CodeRefinery workshop, ask questions about using CodeRefinery material in your own courses, get further information on becoming a CodeRefinery partner, or request access to the repository hosting platform source.coderefinery.org, please contact the CodeRefinery team through the official support line: support@coderefinery.org.

CODE REFINERY

Infrastructure on High Performance Computing

Pan-European Research

HPC Europa3

Visit from Vilnius via HPCE3

Lilit Axner, PDC

As part of the HPC-Europa3 program, PDC hosted two visitors from Lithuania during June and July 2018. Dr. Kestutis Aidas is an experienced researcher from the Institute of Chemical Physics, Faculty of Physics, at Vilnius University. Dr. Aidas and his Ph.D. student, Dovilé Lengvinaitė, wanted to visit Prof. Aatto Laaksonen at the Department of Materials and Environmental Chemistry, Stockholm University, Sweden.

Dr. Aidas had collaborated with Prof. Laaksonen previously when Dr. Aidas was a postdoctoral researcher at the KTH Royal Institute of Technology in 2011. More recently Dr. Aidas and Ms. Lengvinaitė have been working with Prof. Laaksonen on the theoretical modelling of nuclear quadrupolar coupling constants relevant in the field of nuclear magnetic resonance (NMR) relaxation using molecular dynamics (MD) simulations and combined quantum mechanics/molecular mechanics (QM/MM) approaches. When Dr. Aidas and Ms. Lengvinaitė learned about the funding that the HPC-Europa3 project provides for collaborative research visits, they considered it to be an excellent opportunity for them to continue and strengthen their collaboration with Prof. Laaksonen, since it would enable them to directly discuss matters in person and also to use the high-end computing facilities at KTH-PDC.

To model NMR properties of liquid-state molecules, one often relies on the so-called supermolecular approach. This procedure involves

geometry optimisation of the molecular complex composed of the target molecule and a few important solvent molecules, for example, those that are hydrogen-bonded to the target molecule. The entire molecular complex could also be embedded in some kind of dielectric continuum, which is used to account for bulk solvent effects. The NMR properties are then calculated for this geometry-optimised complex using the now standard gauge-including atomic orbital (GIAO) technique to remove gauge-origin dependence of the NMR shielding constants. The supermolecular approach is certainly cost-effective, and reasonably accurate results are often obtained. However, this scheme suffers from a poor description of the long-range electrostatic intermolecular interactions because the discrete nature of solvent molecules is neglected by the continuum approaches. Moreover, dynamical and temperature effects are neglected since the geometry optimised structure represents just a single point on the multidimensional potential energy surface.

These problems can be very effectively dealt with by using an integrated approach where classical MD simulations are combined with the hybrid QM/MM model for the calculation of the NMR properties. MD simulations make it possible to sample the phase space of the molecular system under proper thermodynamic conditions, whereas QM/MM approaches account for anisotropic distribution of the solvent molecules around the solute in the calculation of the electronic response properties. Moreover, a large number of solvent molecules can be easily included in the QM/MM calculation so that the effects of electrostatic interactions are converged. Mutual solute-solvent polarization interactions are accounted for by using explicitly polarizable force fields in the QM/MM calculations. Short-range Pauli repulsion can also be considered by extending the quantum mechanically treated region to include some of the solvent molecules along with the solute. However, there is a price to be paid for the increased accuracy of the model due to the fact that a large number of similar QM/MM calculations (for the molecular snapshots recorded during the MD simulation) need to be performed, typically on the order of around 100.

It was clear that, in order to achieve their goals, extensive MD simulations on the time scale of nearly 100 ns had to be performed for the ionic liquid-water mixtures, including several hundred ionic pairs while varying the number of water molecules. Thus Dr. Aidas and Ms. Lengvinaitė applied to the HPC-Europa3 third call, and were glad that their application was successful.

Above: Ms. Dovilé Lengvinaitė and Dr. Aidas Kestutis from Lithuania, as well as Dr. Ana Dobrota from Serbia, with Beskow while on a tour of the PDC computer hall during their HPC-Europa3 visit to Stockholm.

Dr. Aidas' and Ms. Lengvinaitė's visit to Sweden lasted eight weeks in total, and KTH-PDC was the host centre that provided them with accommodation, office space, and also 257,000 CPU hours of runtime on KTH-PDC's Cray XC40 system (which is called Beskow). By working at the PDC premises, they also had a great opportunity to interact with the high performance computing (HPC) experts at the centre, although the major motivating factor for their application was the potential to collaborate in person with Prof. Aatto Laaksonen, who is a world-renowned expert in classical MD simulations. They benefited from Prof. Laaksonen's experience in simulations of room temperature ionic liquids (RTIL), which is an area where they lack expertise at Vilnius University.

Furthermore, these types of calculations necessitate the extensive use of high-performance computing resources, which Dr. Aidas and Ms. Lengvinaitė had access to via KTH-PDC's facilities during their visit. This was significant as Dr. Aidas' team did not have the means to perform this important – yet computationally very costly – modelling at Vilnius University.

After the visit Dr. Aidas pointed out that the "HPC-Europa3 programme has several strong points:

1. It provides access to top class computational resources.
2. The formalities and paperwork involving the application, actual stay and report are reasonable.
3. The entire process is managed so well that one can really focus on research activities only."

Ms. Lengvinaitė reflected: "This visit was very useful to me as a young scientist, Ph.D. student. During this visit, I had the opportunity to work independently, to test different computing capabilities, to meet the challenges. Get acquainted with great scholars and get in touch."

Further information and details about applying for an HPCE3 exchange can be found at the HPCE3 website http://www.hpc-europa.eu. Dr. Lilit Axner can assist with applications.

Introducing the PDC Blog

Xin Li, PDC

We are blogging! The PDC blog is now online at https://www.kth.se/blogs/pdc. It will feature blog posts focusing on technical topics to help you to use PDC's resources more easily. For example, you can expect to find tips on making better use of high performance computing (HPC) systems, optimising workflows, and writing efficient code.

The topics we are planning to blog about can be roughly grouped into five categories:

- job scheduling with SLURM,
- environment management via Docker, Singularity and Anaconda, amongst others,
- best practices for using common software and tools, including GROMACS and VASP,
- benchmarks of software and scientific codes, and
- parallelization and performance tuning.

The PDC Blog will provide all of you who use PDC's resources in your research with an easily accessible collection of useful tips. Thus the PDC blog will serve as a complement to our PDC Support web pages, giving you additional support in your research. At the moment there are three published posts, "Getting started with SLURM", "Skip the configuration, get to the cluster: Docker way" and "Scalability: strong and weak scaling". You are welcome to subscribe to our blog to take advantage of these tips (with more to come soon)!

Above: The new PDC Blog

PDC-PRACE Workshop: "HPC Tools for the Modern Era"

Thor Wikfeldt, PDC

Basic HPC usage has not changed much in decades – we still log in to clusters using secure shell, use terminals to execute UNIX commands, submit batch jobs to the job scheduler, and so forth. But that does not mean that the HPC landscape is staying the same. In fact, new tools and novel usage patterns are constantly emerging! On the 25th and 26th of October this year, PDC ran a workshop on "HPC Tools for the Modern Era" in partnership with PRACE to teach some of these modern approaches. The four workshop modules were "HPC tips and tricks", "Singularity containers", "Jupyter Notebooks on the cluster" and "Improving your code with Arm Forge". The Arm Forge lesson focused on the debugging and profiling tools offered in Arm's HPC toolkit (These tools were formerly known as Allinea Forge.) and was taught by Conrad Hillairet who visited PDC from the Arm headquarters in Cambridge. All the other lessons were taught by the PDC application experts Xin Li, Thor Wikfeldt and Henric Zazzi. If you are interested in having a look at the lesson material, it is all available on the PDC Support pages: https://www.pdc.kth.se/support/documents/courses/prace.html.

Workshop feedback was collected after each lesson and at the end of the workshop, and overall it was highly positive. An interesting observation from the general feedback form was that many workshop participants would like to participate in more training on machine learning, big data analytics and GPU programming! This is being discussed at PDC, and more generally within the SNIC training programme, with the intention of offering workshops on these topics in 2019.

Above: PDC-PRACE Workshop, 25-26 October 2018

So watch out for a rerun of the PDC-PRACE workshop next year!

PDC Survey Results 2018

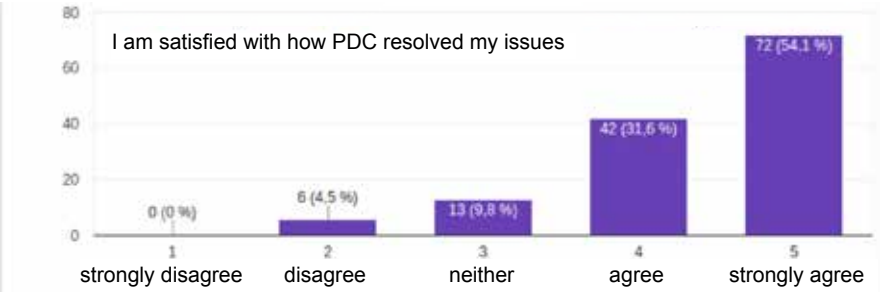
Henric Zazzi, PDC

PDC serves as an infrastructure providing computational and storage facilities for academic researchers in Sweden, and is funded by the Swedish National Infrastructure for Computing (SNIC). As such, it is vital that we take the needs of PDC's users into consideration in our acquisition processes when determining what type of hardware is necessary and also what level of support is needed. Although we do receive a number of requests from researchers relating to different aspects of using our resources, it is difficult to fathom what those users and other researchers think will be necessary (in terms of computational/storage resources) to continue their lines of research in the future.

Therefore, during the spring of 2018, we sent out a survey to all researchers using PDC's services, and also attached the same survey to recently resolved tickets in our system. (Tickets are how we track queries and requests for help from our users, so resolved tickets refer to problems that have been solved for specific PDC users, who would thus each receive an email containing the survey.) The aim was to have short and concise questions, so as to make it easy for people to fill in the questionnaire quickly and hence enable us to acquire as much useful input from our users as possible.

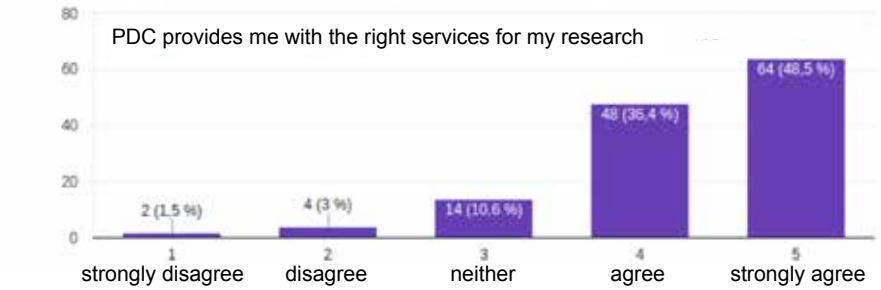
We were pleased to receive responses to our questionnaire from about 135 researchers, most of whom have been using PDC's resources for several years and were well acquainted with the system. Thanks to everyone who took the time to contribute and help us improve our services in the future!

We were glad to see that overall our users are quite happy with the level of support they receive at PDC, and that we are capable of solving their issues and they feel that communication is achieved in a timely and clear manner.



Over the last year we have invested many hours in refurbishing our support pages, so they are more structured and easier to use. It is clear that the majority of our users think that our support pages contain information that is valuable for their research, however, we are also taking various suggestions for improvements on board and will continue to work on improving the support pages in the coming year.

Users were also asked for comments on what they think PDC should provide, besides the current infrastructure and support. Several users commented that they would need more GPU-capable nodes, and more resources for their research in general. Rest assured that PDC is always working to acquire more infrastructure (in hardware terms) to meet the ever-increasing demands of research in the area of HPC! While we are pleased with the results from the survey, we continue to strive to improve our support and welcome feedback to support@pdc.kth.se.



PDC-Related Events

Deep Learning Workshop
20-21 March 2019, Stockholm
PRACE-BioExcel Seasonal School 2019
10-13 June 2019, KTH, Stockholm
PDC Summer School 2019
last two weeks of August 2019, KTH, Stockholm

Check the PDC Events calendar early in 2019 for more details about these events!

HPC Sources

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

BioExcel
http://bioexcel.eu/events

CERN
http://cerncourier.com/cws/events
https://home.cern/scientists/events/computing

EGI
https://www.egi.eu/category/events

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

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

NeIC
http://neic.nordforsk.org

PRACE
http://www.prace-ri.eu/HPC-access
http://www.training.prace-ri.eu
http://www.prace-ri.eu/events
http://www.prace-ri.eu/news

SeSE
http://sese.nu

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

XSEDE
https://www.xsede.org

Above: PDC-PRACE Workshop 2018